Integrated facility location and capacity planning under uncertainty

Isabel Correia[1] · Teresa Melo[2]

Abstract

We address a multi-period facility location problem with two customer segments having distinct service requirements. While customers in one segment receive preferred service, customers in the other segment accept delayed deliveries as long as lateness does not exceed a pre-specified threshold. The objective is to define a schedule for facility deployment and capacity scalability that satisfies all customer demands at minimum cost. Facilities can have their capacities adjusted over the planning horizon through incrementally increasing or reducing the number of modular units they hold. These two features, capacity expansion and capacity contraction, can help substantially improve the flexibility in responding to demand changes. Future customer demands are assumed to be unknown. We propose two different frameworks for planning capacity decisions and present a two-stage stochastic model for each one of them. While in the first model decisions related to capacity scalability are modeled as first-stage decisions, in the second model, capacity adjustments are deferred to the second stage. We develop the extensive forms of the associated stochastic programs for the case of demand uncertainty being captured by a finite set of scenarios. Additional inequalities are proposed to enhance the original formulations. An extensive computational study with randomly generated instances shows that the proposed enhancements are very useful. Specifically, 97.5% of the instances can be solved to optimality in much shorter computing times. Important insights are also provided into the impact of the two different frameworks for planning capacity adjustments on the facility network configuration and its total cost.

Keywords Facility location · Dynamic capacity adjustment · Delivery lateness · Demand uncertainty · Valid inequalities

Mathematics Subject Classification 90B80 · 90C11 · 90C15

✉ Teresa Melo
teresa.melo@htwsaar.de

1 Centro de Matemática e Aplicações (CMA), Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa , 2829-516 Caparica, Portugal

2 Business School, Saarland University of Applied Sciences, 66123 Saarbrücken, Germany

1 Introduction

Location analysis provides a framework for simultaneously finding sites for facilities and assigning spatially distributed demand points to these facilities to optimize some measurable criterion (e.g., cost, profit). Since market conditions and customer service requirements evolve over time, it is also relevant to determine the optimal timing for opening new facilities, in addition to selecting their best locations (Nickel and Saldanha-da-Gama 2019). Furthermore, to be better prepared to market changes (e.g., demand fluctuations), the decision process should also comprise capacity sizing decisions. The latter should specify a schedule for capacity acquisition, capacity expansion, and capacity contraction at the selected locations, in the course of the planning horizon.

Generally, facility location research relies on the assumption that future conditions can be predicted with a reasonable degree of accuracy. This has given rise to the development of deterministic models, which are by far the most common. Even though such models have the advantage of often being amenable to mathematical analysis, the mathematical clarity comes at the cost of decreased realism. In practice, many factors are uncertain (e.g., demands, costs, resource availability), thereby motivating the development of stochastic models (Correia and Saldanha-da-Gama 2019). Ignoring uncertainty entails the risk that once the true nature of the uncertain parameters reveals itself, adjustments in the facility network configuration may be necessary that are time-consuming and require a prohibitively high expense.

In this paper, we consider a variant of the multi-period capacitated facility location problem in which some data are uncertain. We assume that there is a finite set of customers with demand for a single product family over a multi-period finite planning horizon, and a finite set of potential locations for the facilities that will offer the commodity. The goal is to choose facility locations, determine a schedule for facility opening and capacity adjustment (i.e., expansion and contraction), and to decide which customers to serve and from which facilities. A distinctive feature of our problem is that customers are sensitive to delivery lead times. This means that some customers require their demands to be met on time (i.e., these customers impose a zero delivery lead time), whereas the remaining customers accept delayed deliveries as long as lateness does not exceed a pre-specified threshold. This form of customer segmentation can be encountered in a variety of industries, such as retailing (Duran et al. 2008; Li et al. 2015) and aftermarket services (Alvarez et al. 2015; Hung et al. 2012; Wang et al. 2002). Naturally, firms that follow this type of customer segmentation also adopt a price differentiation policy as an incentive for customers to accept longer delivery lead times. Future demands are not known a priori. Instead, they are assumed to be described by a finite discrete random vector, following a probability distribution that is known in advance (e.g., estimated using historical data). Under the assumption that the decision-maker follows a risk neutral strategy, the objective is to minimize the total expected cost. The latter includes fixed costs for opening and operating facilities, fixed costs for capacity scaling, and variable processing, distribution, and tardiness penalty costs. We note that our problem arises in the context of facility sizing decisions being reversible in the medium term. This is the case, for example, when on-demand warehousing is used, which is an emerging trend that allows relatively short commitment horizons for lease contracts (Unnu 2020; Pazour and Unnu 2018). Technological innovations in the chemical process industry also provide flexible production facilities whose capacities are easily redeployed through adding or removing modular units (Becker et al. 2019). Therefore, the three planning levels are integrated through considering strategic decisions (involving facility location), tactical decisions (regarding capacity scalability), and operational decisions (pertaining to demand allocation). Accordingly, facility

location and capacity sizing decisions can only be made at selected time periods, whereas demand allocation decisions can be made at any time period of the planning horizon.

A variant of our problem was studied by Correia and Melo (2017), who have shown that this is already a challenging NP-hard problem when all parameters are assumed to be known with certainty. To the best of the authors' knowledge, our work is the first attempt to understand the additional complexity that arises from incorporating stochasticity into the underlying difficult-to-solve deterministic facility location and capacity scalability problem in a multi-period setting. For this reason, we focus on demand uncertainty and defer the consideration of additional uncertain parameters.

Given that demand uncertainty is captured by a finite set of scenarios with known probabilities, we follow a scenario-based approach and formulate two-stage stochastic programs for two different strategies regarding capacity sizing. In both models, the first-stage problem determines the opening schedule for facilities and their initial capacities over the entire planning horizon before uncertainty about demand is revealed. Additionally, operational decisions are captured in the second-stage problem by specifying the commodity flow between the selected locations and the customers. Hence, distribution decisions are made after the demand uncertainty has been resolved. The proposed stochastic programs differ in the way capacity scalability decisions are treated. In the first model, we follow the approach usually taken in the literature (Correia and Saldanha-da-Gama 2019; Govindan et al. 2017; Snyder 2006), and assume that these decisions have a strategic nature. Hence, we include them in the set of ex ante or first-stage decisions. By contrast, in the second model, decisions on the timing and sizing of capacity adjustments are assumed to have a tactical nature. In this case, they become ex post or second-stage decisions. This setting is suitable when, e.g. on-demand warehousing services are available, thus enabling the capacities of operating facilities to be upgraded and downgraded on a medium-term basis (Unnu 2020; Pazour and Unnu 2018).

The contributions of the present work are summarized as follows. (1) We propose two alternative modeling strategies for a new facility location and capacity planning problem with uncertain demand. (2) For each strategy, we formulate the *Deterministic Equivalent Program* and develop valid inequalities that significantly improve the polyhedral description of the feasible region. (3) We evaluate the potential benefit obtained from solving the stochastic programs instead of their deterministic counterparts. (4) We assess the validity of the proposed two-stage stochastic mixed-integer linear programming (MILP) models by reporting and analyzing the results of an extensive computational study using a general-purpose optimization solver. In particular, we provide insights into the impact of considering capacity planning decisions as second-stage (tactical) decisions as opposed to the typical approach of treating them as first-stage (strategic) decisions. Without the support of the two models proposed in this paper it would otherwise be difficult for stakeholders (typically from higher management) to perceive the trade-offs associated with different strategies for facility configuration.

The remainder of this paper is organized as follows. Section 2 provides an overview of the relevant literature. In Sect. 3, we formally describe our problem and present two different stochastic programs along with the corresponding deterministic equivalent forms. Several sets of valid inequalities are developed in Sect. 4 to enhance the original formulations and, therefore, make them more amenable to standard off-the-shelf optimization software. Computational results are reported in Sect. 5 and the proposed formulations are compared using various metrics. Finally, Sect. 6 includes a summary of our findings and gives directions for future research.

2 Literature review

In this section, we review selected research articles on facility location. Our aim is not to provide an exhaustive review of this field but to analyze the extent to which the features captured by our models have been addressed in the literature. Accordingly, special emphasis will be given to stochastic and multi-period location problems.

Different sources of uncertainty have been considered, either individually or simultaneously, in the literature dedicated to stochastic facility location, with demand uncertainty being the most often considered factor. Early contributions addressing uncertain demand include the works by Louveaux (1986) and Laporte et al. (1994). Over the past decades, the stochastic dimension in facility location modeling has evolved towards different areas, e.g., hub location (Alumur et al. 2012; Contreras et al. 2011; Correia et al. 2018; Rahmati and Neghabi 2021). From an application perspective, recent advances in stochastic modeling have been reported, e.g., in humanitarian logistics (Döyen et al. 2012; Grass et al. 2020; Kınay et al. 2018), and supply chain network design (Ghafarimoghadam et al. 2019; Govindan et al. 2017; Mohamed et al. 2020).

While static (or single-period) stochastic location problems have attracted increasing attention in the literature, there has been limited research on the multi-period stochastic counterpart. Albareda-Sambola et al. (2013) study a multi-period facility location problem with various sources of uncertainty. Marín et al. (2018) propose a stochastic programming modeling framework for a general class of covering location problems and develop a Lagrangian relaxation-based heuristic to tackle large-scale instances. In the context of humanitarian logistics, Kim et al. (2019) investigate the problem of finding optimal locations for drone facilities taking into account that flight distance is uncertain due to battery consumption being affected by weather conditions. Khodaparasti et al. (2018) address the problem of designing a network of nursing homes under uncertain demand and a budget constraint on each time period. Marković et al. (2017) develop a multi-stage stochastic program for the problem of locating law enforcement facilities on a road network to intercept stochastic vehicle flows that try to avoid these facilities. In Mohamed et al. (2020), a multi-stage stochastic program is proposed to locate intermediate facilities in a two-echelon distribution network facing uncertain demand.

Capacity planning decisions are often incorporated into facility location problems (Owen and Daskin 1998; Verter and Dincer 1995). In particular, the gradual adjustment of capacities over a planning horizon is an important driver for improving flexibility within a facility network, a feature that is captured in this paper. Ideally, the provision of capacity should be in line with actual demand needs. Depending on the application context, different strategies can be adopted along the planning horizon, such as capacity expansion (Correia et al. 2013; Cortinhal et al. 2015), capacity expansion combined with capacity reduction (Antunes and Peeters 2001; Wilhelm et al. 2013), relocation of capacities to different locations (Melo et al. 2006), and temporary facility closing and reopening (Dias et al. 2007; Jena et al. 2016). Among these strategies, capacity expansion is the predominant policy in a stochastic setting. Early contributions in this area focus on determining a schedule for capacity acquisition and expansion for a set of facilities already operating at fixed locations (Ahmed and Garcia 2004; Ahmed et al. 2003). Any continuous amount of capacity can be installed. By contrast, Alonso-Ayuso et al. (2003) assume a finite set of capacity expansion levels in a multi-commodity production system. For a multi-echelon logistics network, Aghezzaf (2005) propose a decomposition-based algorithm to determine a facility location and capacity expansion schedule that is robust to the uncertain realization of customer demand. In

the problem studied by Hernández et al. (2012), the location and size of prisons are to be determined along with a schedule for capacity upgrading at both existing and new facilities under stochastic demand. Recently, a stochastic modeling framework was proposed for a multi-period hub location problem with uncertain demand (Correia et al. 2018). In addition to finding the location of hub facilities and setting their initial capacities, an expansion plan is also determined. This entails deciding on the time period and the number of additional modular units to be installed in a hub facility. Even though modular capacities are relevant in many contexts, e.g. to represent modular equipment (Delmelle et al. 2014; Gourdin and Klopfenstein 2008), they have been hardly considered in the literature. This feature is captured in our work in conjunction with capacity expansion and capacity reduction. To the best of our knowledge, there is only one other study that has addressed both capacity upgrading and downgrading in a stochastic facility location setting, namely Zhuge et al. (2016). In their two-stage stochastic program, first-stage decisions include locating new distribution centers (DCs) in the first time period and selecting their initial capacities from a set of three options. The second-stage decisions define the timing for capacity expansions and contractions as well as the flow of multiple products. Uncertainty in future demand and the budget available for replacing capacity levels are considered. Small-sized problems using aggregated data from an automotive logistics firm are solved by a Lagrangian-based heuristic. The only other works that also model capacity scalability decisions as second-stage decisions, the former being limited to capacity expansion decisions, are Correia et al. (2018) for a hub location problem and Heckmann (2016) for a facility location problem with disruptive events.

Despite considerable progress on stochastic location modeling, there are still a number of relevant features which have not been addressed so far. The present paper contributes to filling this gap by proposing two stochastic models for location and capacity planning in a multi-period context with service-differentiated customer segments.

3 Problem statement and stochastic formulations

The stochastic multi-period facility location and capacity planning problem addressed in this paper builds upon a deterministic variant studied in Correia and Melo (2017). For convenience, we restate the assumptions that are common to the deterministic and stochastic settings, and introduce additional information pertaining to the stochastic problem. Subsequently, two stochastic formulations are proposed that differ in the framework under which capacity planning decisions can be made.

A planning horizon comprising a finite number of time periods is considered along with a set of customers with demands for a single commodity in each period. Customers are differentiated according to their service requirements. Specifically, some customers require their demands to be met on time, whereas the remaining customers accept delayed deliveries as long as lateness does not exceed a pre-specified threshold. In this case, an order can be split over multiple time periods for the same customer. Typically, customers belonging to the first segment are willing to pay higher prices to have their demands satisfied in the same period in which they occur. To customers in the second segment, discount prices are often offered in exchange for longer response times. In our modeling framework, this scheme is translated into a tardiness penalty cost that is incurred per unit of backorder and per period. Customer service-differentiation offers greater flexibility to a firm in designing and managing its facility network.

A set of potential locations where facilities can be established is assumed to be available at the beginning of the planning horizon. When a facility is opened, its initial capacity level must also be specified. The latter is expressed by the number of modular units that are installed, all having the same size. The adoption of a base modular unit is common in different application areas, such as in the public sector (Delmelle et al. 2014) and in telecommunications (Gourdin and Klopfenstein 2008). Over the planning horizon, the capacity of an operating facility can be adjusted to increase its responsiveness to changes in the level of customer demand. Two forms of capacity scalability are considered: capacity expansion and capacity contraction. This means that one or several modular units can be added to or removed from a facility. Naturally, these options are mutually exclusive for the same facility at the same time period. A limit on the total number of modular units that a facility can hold at any time is pre-specified. Due to the sizeable investment associated with opening facilities, locations cannot be temporarily closed and reopened.

The objective is to find a schedule for the deployment of facilities and the expansion and contraction of their capacities to satisfy customer demands at minimal total cost. Different time scales are considered for decision-making. The capital-intensive decisions involving facility location and capacity planning decisions can be made at the beginning of selected time periods, hereafter termed *design periods*. All other decisions concerning the processing of the commodity at the open facilities and the allocation of customer demands can be made at any time period. This setting is also adopted by other authors, e.g., Mohamed et al. (2020) and Bashiri et al. (2012). Uncertainty about future demand is described by a finite random vector whose joint probability distribution is assumed to be foreknown. The finitely many possible realizations of this random vector are called *scenarios*. Specifically, each scenario describes all customer demands over all periods of the planning horizon.

Before detailing the two stochastic programming models, we first introduce the notation that is common to both formulations.

3.1 Notation

Sets

T Set of time periods along the planning horizon.
T_L Subset of *design* time periods at which location and capacity planning decisions can be made ($T_L \subset T$).
I Set of potential facility sites.
J^0 Set of customers requiring timely deliveries (i.e., their demands must be satisfied in the same time periods in which they occur).
J^1 Set of customers that tolerate delayed demand fulfillment.
J Set of all customers ($J = J^0 \cup J^1$; $J^0 \cap J^1 = \emptyset$).

Let $\ell = 1$, resp. $\ell_{\max} = \max\{\ell \in T_L\}$, be the first, resp. last, design period in which decisions can be made on opening facilities at potential locations and adjusting their capacities through adding or removing modular units. The size of a facility at time period $\ell \in T_L$ is the outcome of the design and scalability decisions taken until that period. For $\ell < \ell_{\max}$, this size remains unchanged over all intermediate periods between two consecutive design periods, ℓ and ℓ' ($\ell < \ell'$). For $\ell = \ell_{\max}$, the capacity of the facility is maintained until the last period of the planning horizon. Let $\phi(\ell)$ denote the last time period between two consecutive design

periods ℓ and ℓ'. This parameter is defined as follows:

$$\phi(\ell) = \begin{cases} \max\{t \in T : t < \ell'\} & \text{if } \ell < \ell_{\max}, \\ |T| & \text{if } \ell = \ell_{\max}. \end{cases} \tag{1}$$

Deterministic parameters

Q Capacity of one modular unit.

n_i Maximum number of modular units that can be available at location i at any time period ($i \in I$).

ρ_j Maximum allowed delay (expressed by the total number of time periods) to satisfy the demand of customer j ($j \in J^1$).

FO_{ik}^ℓ Fixed cost of opening facility i with k modular units at design period ℓ ($i \in I$; $k = 1, \ldots, n_i$; $\ell \in T_L$).

FE_{ik}^ℓ Fixed cost of adding k modular units to the capacity of facility i at design period ℓ ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L\backslash\{1\}$).

FR_{ik}^ℓ Fixed cost of removing k modular units from the capacity of facility i at design period ℓ ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L\backslash\{1\}$).

M_{ik}^t Fixed cost of operating facility i with k modular units at time period t ($i \in I$; $k = 1, \ldots, n_i$; $t \in T$).

o_{ik}^t Unit processing cost charged by facility i operating with k modular units at time period t ($i \in I$; $k = 1, \ldots, n_i$; $t \in T$).

c_{ij}^t Unit distribution cost from facility i to customer j at time period t ($i \in I$; $j \in J$; $t \in T$).

$p_j^{tt'}$ Unit tardiness penalty cost for satisfying demand of customer j in period t' that was originally placed in period t ($j \in J^1$; $t \in T$; $t' = t, t+1, \ldots, \min\{t + \rho_j, |T|\}$); in particular, $p_j^{tt'} = 0$ for $t' = t$.

The customer-specified time lag for demand satisfaction ρ_j ($j \in J^1$) imposes that an order placed by customer j for time period t ($t \in T$) must be filled over periods $t, t+1, \ldots, t+\rho_j$. In case $t + \rho_j > |T|$, then the last delivery must occur at period $|T|$. This condition ensures that demand cannot be carried over to periods beyond the planning horizon. If $\rho_j = 0$ for every customer $j \in J^1$ then our problem reduces to the classical setting in multi-period facility location.

As shown in Appendix B, some cost parameters capture economies of scale, namely all fixed costs (FO_{ik}^ℓ, FE_{ik}^ℓ, FR_{ik}^ℓ, M_{ik}^t) and the variable processing costs (o_{ik}^t).

Stochastic parameters:

$d_j^t(\xi)$ Random variable representing the demand of customer j at time period t ($j \in J$; $t \in T$).

Ξ Finite discrete random vector with $\Xi = \left((d_j^t)_{j \in J;\, t \in T}\right)$ and $\Xi \subset \mathbb{R}_+^{|J| \times |T|}$.

It is assumed that the joint probability distribution of the random vector Ξ is foreknown.

3.2 A stochastic formulation

Given the problem assumptions, a natural approach to the decision-making process is to develop a two-stage stochastic programming model, where the facility location and capacity planning decisions are made in the first stage, and the remaining processing and distribution decisions are deferred to the second stage. In other words, the first-stage (or ex ante) decisions are associated with the definition of a schedule for location, capacity acquisition, and capacity

adjustment to be implemented over the entire planning horizon, before a realization of the demand becomes known. This results from the strategic nature assumed for these decisions as they have a long-term impact. The following binary variables represent the first-stage decisions.

$z_{ik}^\ell = 1$ if a new facility is established with k modular units at potential location i
at time period ℓ, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i$; $\ell \in T_L$).   (2)

$e_{ik}^\ell = 1$ if the capacity of a facility in location i is expanded with k modular units at
time period ℓ, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L\backslash\{1\}$).   (3)

$r_{ik}^\ell = 1$ if the capacity of a facility in location i is reduced by removing k modular
units at time period ℓ, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L\backslash\{1\}$).   (4)

$v_{ik}^\ell = 1$ if a facility is operated at location i with k modular units at time period
ℓ, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i$; $\ell \in T_L$).   (5)

After customer demand is observed, additional decisions are made by specifying the quantities of product to be processed at the operating facilities and distributed to customers over the planning horizon. Therefore, second-stage (or ex post) decisions are represented by the following continuous variables:

$x_{ij}^t(\xi)$: Amount of product distributed from facility i to customer j at time period
t ($i \in I$; $j \in J^0$; $t \in T$).   (6)

$y_{ij}^{tt'}(\xi)$: Amount of product distributed from facility i to customer j at time period
t' to (partially) satisfy demand of period t ($i \in I$; $j \in J^1$; $t \in T$; $t' = t$,
$t + 1, \ldots, \min\{t + \rho_j, |T|\}$).   (7)

$w_{ik}^t(\xi)$: Total quantity of product handled by facility i operating with k modular
units at time period t ($i \in I$; $k = 1, \ldots, n_i$; $t \in T$).   (8)

The implicit representation of the two-stage stochastic program is as follows:

$$\text{Min} \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} FO_{ik}^\ell z_{ik}^\ell + \sum_{\ell \in T_L\backslash\{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FE_{ik}^\ell e_{ik}^\ell +$$

$$\sum_{\ell \in T_L\backslash\{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FR_{ik}^\ell r_{ik}^\ell + \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} v_{ik}^\ell \sum_{t=\ell}^{\phi(\ell)} M_{ik}^t + \mathbb{E}_\Xi [Q(z, e, r, v, \xi)] \quad (9)$$

subject to

$$\sum_{\ell \in T_L} \sum_{k=1}^{n_i} z_{ik}^\ell \le 1 \quad i \in I \quad (10)$$

$$\sum_{k=1}^{n_i} v_{ik}^\ell = \sum_{\ell' \in T_L: \ell' \le \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \quad i \in I, \ \ell \in T_L \quad (11)$$

$$\sum_{k=1}^{n_i-1} e_{ik}^\ell + \sum_{k=1}^{n_i-1} r_{ik}^\ell \le \sum_{\ell' \in T_L: \ell' < \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \quad i \in I, \ \ell \in T_L\backslash\{1\} \quad (12)$$

$$\sum_{k=1}^{n_i} k\, v_{ik}^{\ell} = \sum_{\ell' \in T_L:\, \ell' \leq \ell} \sum_{k=1}^{n_i} k\, z_{ik}^{\ell'}$$

$$+ \sum_{\ell' \in T_L \setminus \{1\}:\, \ell' \leq \ell} \sum_{k=1}^{n_i-1} k\left(e_{ik}^{\ell'} - r_{ik}^{\ell'}\right) \quad i \in I,\, \ell \in T_L \tag{13}$$

$$z_{ik}^{\ell} \in \{0, 1\} \quad i \in I,\, k = 1, \ldots, n_i,\, \ell \in T_L \tag{14}$$

$$v_{ik}^{\ell} \in \{0, 1\} \quad i \in I,\, k = 1, \ldots, n_i,\, \ell \in T_L \tag{15}$$

$$e_{ik}^{\ell},\, r_{ik}^{\ell} \in \{0, 1\} \quad i \in I,\, k = 1, \ldots, n_i - 1,$$
$$\ell \in T_L \setminus \{1\} \tag{16}$$

with $\mathbb{E}_{\Xi}\left[Q(z, e, r, v, \xi)\right]$ denoting the recourse function, which represents the expected value of the second-stage problem. The optimal value of this problem, $Q(z, e, r, v, \xi)$, is given by

$$Q(z, e, r, v, \xi) = \text{Min} \sum_{t \in T} \sum_{i \in I} \sum_{k=1}^{n_i} o_{ik}^t\, w_{ik}^t(\xi) + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^0} c_{ij}^t\, x_{ij}^t(\xi) +$$

$$\sum_{t \in T} \sum_{i \in I} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j,\, |T|\}} \left(c_{ij}^{t'} + p_j^{tt'}\right) y_{ij}^{tt'}(\xi) \tag{17}$$

subject to

$$\sum_{i \in I} x_{ij}^t(\xi) = d_j^t(\xi) \quad j \in J^0,\, t \in T \tag{18}$$

$$\sum_{i \in I} \sum_{t'=t}^{\min\{t+\rho_j,\, |T|\}} y_{ij}^{tt'}(\xi) = d_j^t(\xi) \quad j \in J^1,\, t \in T \tag{19}$$

$$w_{ik}^t(\xi) \leq k\, Q\, v_{ik}^{\ell} \quad i \in I,\, k = 1, \ldots, n_i,\, t \in T,$$
$$\ell = \max\{\ell' \in T_L : \ell' \leq t\} \tag{20}$$

$$\sum_{k=1}^{n_i} w_{ik}^t(\xi) = \sum_{j \in J^0} x_{ij}^t(\xi) +$$

$$\sum_{j \in J^1} \sum_{t'=\max\{1,\, t-\rho_j\}}^{t} y_{ij}^{t't}(\xi) \qquad i \in I,\, t \in T \tag{21}$$

$$x_{ij}^t(\xi) \geq 0 \quad i \in I,\, j \in J^0,\, t \in T \tag{22}$$

$$y_{ij}^{tt'}(\xi) \geq 0 \quad i \in I,\, j \in J^1,\, t \in T,$$
$$t' = t, \ldots, \min\{t + \rho_j,\, |T|\} \tag{23}$$

$$w_{ik}^t(\xi) \geq 0 \quad i \in I,\, k = 1, \ldots, n_i,\, t \in T \tag{24}$$

In the first-stage problem (9)–(16), the objective is to minimize the sum of fixed costs and expected processing and demand allocation costs. The fixed costs account for facility location, capacity expansion, capacity contraction, and facility operating costs. By including the expected total cost in (9), a neutral approach to risk is assumed, which is a common criterion to address risk in decision-making problems, often leading to computationally tractable

models (Birge and Louveaux 2011). Constraints (10) ensure that at most one facility can be opened at a potential location over the planning horizon. Equalities (11) impose that facilities can only be operated provided they have previously been established. Constraints (12) guarantee that an open facility can be subjected to at most one type of capacity adjustment (either expansion or contraction) at each design period. The number of available modular units at open facilities is defined by equalities (13) for each design period. This number is the outcome of the sizing choice made when the facility was opened, and the number of modular units that were added or removed afterward. Finally, constraints (14)–(16) set the binary conditions on the first-stage (strategic) variables.

Every particular realization of customer demand ξ of Ξ yields a second-stage LP-model (17)–(24). Observe that in this model, the design variables z_{ik}^ℓ, v_{ik}^ℓ, e_{ik}^ℓ, and r_{ik}^ℓ take on fixed values. The objective function (17) minimizes the sum of processing costs at operating facilities, distribution costs to customers, and tardiness penalty costs for orders delivered with delay to customer segment J^1. Demand satisfaction is imposed by constraints (18) and (19). Capacity constraints are enforced by inequalities (20). Equalities (21) state that the product flow leaving a facility at a given time period is split into deliveries to priority customers and to customers that tolerate late shipments. Non-negativity conditions on the second-stage (operational) variables are set by constraints (22)–(24). Observe that we are facing a stochastic problem with fixed recourse, since the coefficients of the second-stage variables (i.e., the unit cost parameters in (17) and the capacity of a modular unit in (20)) are all known in advance.

Recall that we assume that the random variable ξ has a finite support, i.e., ξ is defined by a finite probability distribution. Accordingly, let $s \in S$ be the index of the possible realizations of ξ (called *scenarios*), where S is a finite set, and let π_s represent the associated (positive) probabilities such that $\sum_{s \in S} \pi_s = 1$. In this case, $\mathbb{E}_\Xi \left[Q(z, e, r, v, \xi) \right] = \sum_{s \in S} \pi_s \, Q(z, e, r, v, \xi_s)$. It follows that we can rewrite formulation (9)–(16) as a MILP model by defining scenario-indexed demands and recourse variables. The former are given by d_{js}^t ($t \in T$; $j \in J$; $s \in S$), while the latter replace the second-stage variables (6)–(8) and are represented by

x_{ijs}^t : Amount of product distributed from facility i to customer j at time period

t under scenario s ($i \in I$; $j \in J^0$; $t \in T$; $s \in S$). (25)

$y_{ijs}^{tt'}$: Amount of product distributed from facility i to customer j at time period

t' to (partially) satisfy demand of period t under scenario s ($i \in I$; $j \in J^1$;

$t \in T$; $t' = t, t+1, \ldots, \min\{t + \rho_j, |T|\}$; $s \in S$). (26)

w_{iks}^t : Total quantity of product handled by facility i operating with k modular

units at time period t under scenario s ($i \in I$; $k = 1, \ldots, n_i$; $t \in T$;

$s \in S$). (27)

Using the above redefinition of the recourse variables, we obtain the so-called *deterministic equivalent* MILP model, or extensive form, hereafter denoted ($\mathbf{P}_{DE-1}$).

$$\text{Min} \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} FO_{ik}^\ell \, z_{ik}^\ell + \sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FE_{ik}^\ell \, e_{ik}^\ell$$

$$+ \sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FR_{ik}^\ell \, r_{ik}^\ell + \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} v_{ik}^\ell \sum_{t=\ell}^{\phi(\ell)} M_{ik}^t$$

$$+ \sum_{s \in S} \pi_s \left[\sum_{t \in T} \sum_{i \in I} \sum_{k=1}^{n_i} o_{ik}^t w_{iks}^t + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^0} c_{ij}^t x_{ijs}^t \right.$$

$$\left. + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j,\, |T|\}} \left(c_{ij}^{t'} + p_j^{tt'} \right) y_{ijs}^{tt'} \right] \tag{28}$$

subject to

(10)−(16)

$$\sum_{i \in I} x_{ijs}^t = d_{js}^t \quad j \in J^0,\, t \in T,\, s \in S \tag{29}$$

$$\sum_{i \in I} \sum_{t'=t}^{\min\{t+\rho_j,\,|T|\}} y_{ijs}^{tt'} = d_{js}^t \quad j \in J^1,\, t \in T,\, s \in S \tag{30}$$

$$w_{iks}^t \le k\, Q\, v_{ik}^\ell \quad i \in I,\, k = 1, \ldots, n_i,\, t \in T,$$
$$\ell = \max\{\ell' \in T_L : \ell' \le t\},\, s \in S \tag{31}$$

$$\sum_{k=1}^{n_i} w_{iks}^t = \sum_{j \in J^0} x_{ijs}^t + \sum_{j \in J^1} \sum_{t'=\max\{1,\, t-\rho_j\}}^{t} y_{ijs}^{t't} \quad i \in I,\, t \in T,\, s \in S \tag{32}$$

$$x_{ijs}^t \ge 0 \quad i \in I,\, j \in J^0,\, t \in T,\, s \in S \tag{33}$$

$$y_{ijs}^{tt'} \ge 0 \quad i \in I,\, j \in J^1,\, t \in T,$$
$$t' = t, \ldots, \min\{t + \rho_j, |T|\},\, s \in S \tag{34}$$

$$w_{iks}^t \ge 0 \quad i \in I,\, k = 1, \ldots, n_i,\, t \in T,\, s \in S. \tag{35}$$

We note that the extensive form of the stochastic program becomes very large even when the random variable ξ has a moderate number of possible realizations. Since all the strategic decisions made in the first stage do not depend on the realization of the random demand process, and so the location and capacity scalability schedule determined for the entire planning horizon is the same for all scenarios, formulation ($\mathbf{P}_{DE-1}$) satisfies the non-anticipativity principle. For this reason, ($\mathbf{P}_{DE-1}$) will be called the *scenario-independent* location and capacity scalability model.

3.3 A stochastic formulation for an alternative strategy

Discrepancies between the capacity of a firm and the demands of its customers result in inefficiencies, either in underutilized resources or dissatisfied customers. The ability of a firm to adjust the capacity of its resources in response to varying customer demand is one of the key factors in decreasing these discrepancies and achieving competitiveness. Therefore, an alternative strategy to the settings considered in the previous section is to assume that capacity adjustment decisions have a tactical nature, and thus can be deferred to the second-stage problem. In this strategy, the first-stage problem consists of defining a schedule for locating facilities and setting their initial capacities over the planning horizon. Accordingly, the strategic facility location decisions are the *here-and-now* decisions. The second-stage problem addresses the *wait-and-see* (tactical and operational) decisions by prescribing a scheme for adjusting the capacities of operating facilities (through expansion or contraction)

at the design periods, and specifying the product flows from facilities to customers over all time periods. In contrast to the approach presented in Sect. 3.2, the capacity adjustment measures depend now on the realization of the uncertain customer demand. As mentioned earlier, the resulting two-stage stochastic program is suitable for those cases in which sizing decisions can be reverted in the medium-term. This arises, for example, in the context of leasing or renting space and equipment. On-demand warehousing is an emerging trend that enables firms to adapt quickly to variable demand and cost conditions through deploying storage space at different locations, for different volumes, in a dynamic fashion (Unnu 2020; Pazour and Unnu 2018). In particular, e-commerce retailers profit from this pay-per-use model. Another setting that also allows flexible facility reconfiguration arises in the chemical process industry through modular plants whose capacities can be expanded, contracted or even relocated over a short planning horizon (Becker et al. 2019). Malladi et al. (2020) and Faugère et al. (2020) also report on different logistics systems that rely on modular production capacity that can be redeployable in the short term.

To formulate a two-stage stochastic program for this alternative strategy, the binary location variables (2) and the continuous flow variables (6), (7), and (8) are used. Moreover, the binary variables associated with the sizing decisions, i.e., (3), (4), and (5), are replaced by $\widetilde{e}_{ik}^{\ell}(\xi)$, $\widetilde{r}_{ik}^{\ell}(\xi)$, and $\widetilde{v}_{ik}^{\ell}(\xi)$, respectively. Each one of these new variables depends on the random vector ξ ruling future customer demands. The total expected cost to be minimized is given by $\sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} FO_{ik}^{\ell} z_{ik}^{\ell} + \mathbb{E}_{\Xi}[Q'(z, \xi)]$ under constraints (10) and (14). The optimal value $Q'(z, \xi)$ of the second-stage problem is determined by

$$
\begin{aligned}
Q'(z, \xi) = \text{Min} \quad & \sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FE_{ik}^{\ell} \widetilde{e}_{ik}^{\ell}(\xi) + \sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i-1} FR_{ik}^{\ell} \widetilde{r}_{ik}^{\ell}(\xi) \\
& + \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} \widetilde{v}_{ik}^{\ell}(\xi) \sum_{t=\ell}^{\phi(\ell)} M_{ik}^t + \sum_{t \in T} \sum_{i \in I} \sum_{k=1}^{n_i} o_{ik}^t w_{ik}^t(\xi) \\
& + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^0} c_{ij}^t x_{ij}^t(\xi) + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j, |T|\}} \left(c_{ij}^{t'} + p_j^{tt'} \right) y_{ij}^{tt'}(\xi)
\end{aligned}
$$

subject to   (36)

$$(18), (19), (21) - (24)$$

$$
\sum_{k=1}^{n_i} \widetilde{v}_{ik}^{\ell}(\xi) = \sum_{\ell' \in T_L : \ell' \leq \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \quad i \in I, \ \ell \in T_L \tag{37}
$$

$$
\sum_{k=1}^{n_i-1} \widetilde{e}_{ik}^{\ell}(\xi) + \sum_{k=1}^{n_i-1} \widetilde{r}_{ik}^{\ell}(\xi) \leq \sum_{\ell' \in T_L : \ell' < \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \qquad i \in I, \ \ell \in T_L \backslash \{1\} \tag{38}
$$

$$
\begin{aligned}
\sum_{k=1}^{n_i} k \, \widetilde{v}_{ik}^{\ell}(\xi) = & \sum_{\ell' \in T_L : \ell' \leq \ell} \sum_{k=1}^{n_i} k \, z_{ik}^{\ell'} \\
& + \sum_{\ell' \in T_L \backslash \{1\} : \ell' \leq \ell} \sum_{k=1}^{n_i-1} k \left(\widetilde{e}_{ik}^{\ell'}(\xi) - \widetilde{r}_{ik}^{\ell'}(\xi) \right) \qquad i \in I, \ \ell \in T_L
\end{aligned} \tag{39}
$$

$$
w_{ik}^t(\xi) \leq k \, Q \, \widetilde{v}_{ik}^{\ell}(\xi) \quad i \in I, \ k = 1, \ldots, n_i, \ t \in T,
$$

$$\ell = \max\{\ell' \in T_L : \ell' \leq t\} \tag{40}$$

$$\widetilde{v}_{ik}^{\ell}(\xi) \in \{0, 1\} \quad i \in I, \; k = 1, \ldots, n_i, \; \ell \in T_L \tag{41}$$

$$\widetilde{e}_{ik}^{\ell}(\xi), \; \widetilde{r}_{ik}^{\ell}(\xi) \in \{0, 1\} \quad i \in I, \; k = 1, \ldots, n_i - 1,$$
$$\ell \in T_L \backslash \{1\}. \tag{42}$$

Due to capacity adjustment decisions being now recourse decisions, the new constraints (37)–(39) are the counterpart of constraints (11)–(13). Likewise, the capacity constraints (40) replace inequalities (20).

By imposing the same conditions on the random demand vector ξ as in Sect. 3.2, we can rewrite the recourse function as a function of all scenarios, i.e., $\mathbb{E}_\Xi[Q'(z, \xi)] = \sum_{s \in S} \pi_s \, Q'(z, \xi_s)$. Accordingly, the set of scenario-dependent flow variables (25)–(27) is extended with the new scenario-indexed capacity scalability variables, $\widetilde{v}_{iks}^{\ell}$, $\widetilde{e}_{iks}^{\ell}$, and $\widetilde{r}_{iks}^{\ell}$. The definition of these variables and the resulting deterministic equivalent MILP formulation, hereafter denoted ($\mathbf{P}_{DE-2}$), are detailed in Appendix A. Since all capacity adjustment decisions depend on the realizations of the random parameters, ($\mathbf{P}_{DE-2}$) will be called the *scenario-dependent* capacity scalability model. ($\mathbf{P}_{DE-2}$) defines a large-scale MILP model with a significant number of binary variables and constraints, even for a small-sized scenario set.

4 Additional inequalities

In an attempt to improve the polyhedral description of the set of feasible solutions of models ($\mathbf{P}_{DE-1}$) and ($\mathbf{P}_{DE-2}$), we develop in this section several sets of additional inequalities. The proposed enhancements may contribute to more problem instances and with larger sizes being solved to (near-)optimality with general-purpose optimization software within reasonable computing time.

4.1 Inequalities for model (P_{DE-1})

Recall from (28)–(35) that in the extensive form of the deterministic equivalent model ($\mathbf{P}_{DE-1}$), all facility location and capacity-related decisions are made in the first stage, before uncertainty about future demand is disclosed. Next, we successively show how the information conveyed by the various scenarios can be used in this case to derive a lower bound on the total number of facilities that must be open at a particular design period.

Let us assume, without loss of generality, that the length of the planning horizon is a multiple of the number of design periods, and let us define $\sigma = |T|/|T_L|$. Hence, parameter σ is integer-valued and gives the total number of periods between two consecutive design periods. Since all demand requirements of the preferred customer segment J^0 must be serviced without delay in every scenario, they are used to set minimum capacity requirements between two consecutive design periods. The latter are given by

$$D^{\ell} = \max \left\{ \sum_{j \in J^0} d_{js}^t : s \in S; \; t = \ell, \ldots, \phi(\ell) \right\},$$

for every $\ell \in T_L$ and $\phi(\ell)$ defined in (1). Hence, the following inequalities must hold for every design period:

$$Q \sum_{i \in I} \sum_{k=1}^{n_i} k \, v_{ik}^\ell \geq D^\ell, \qquad \ell \in T_L.$$

Dividing the above inequalities by the size of a modular unit, Q, and taking into account that the left-hand side must be integer-valued, it follows that

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, v_{ik}^\ell \geq \left\lceil \frac{D^\ell}{Q} \right\rceil, \qquad \ell \in T_L. \tag{43}$$

Furthermore, for a particular design period ℓ ($\ell \in T_L$), let $\widetilde{D}^\ell$ denote the largest minimum demand quantity over all scenarios that has to be covered from period ℓ through period $\phi(\ell)$.

$$\widetilde{D}^\ell = \begin{cases} \max\limits_{s \in S} \left\{ \sum\limits_{j \in J^0} \sum\limits_{t=\ell}^{\phi(\ell)} d_{js}^t + \sum\limits_{j \in J^1} \sum\limits_{t=\ell}^{\phi(\ell)-\rho_j} d_{js}^t \right\} & \text{if } \ell < \ell_{max}, \\[2em] \max\limits_{s \in S} \left\{ \sum\limits_{j \in J} \sum\limits_{t=\ell}^{|T|} d_{js}^t \right\} & \text{if } \ell = \ell_{max}. \end{cases}$$

The above expression includes all orders from customer segment J^0 as well as the minimum quantity of demand that must be delivered to customers J^1 accepting late shipments. Since demand cannot be lost, in the last time interval (which comprises all periods from ℓ_{max} through $|T|$), the orders of both customer segments must be completely satisfied. Accordingly, the following inequalities ensure that facilities open at time period ℓ (and therefore at all subsequent periods up to period $\phi(\ell)$) must have sufficient capacity to serve the demand $\widetilde{D}^\ell$:

$$\sigma Q \sum_{i \in I} \sum_{k=1}^{n_i} k \, v_{ik}^\ell \geq \widetilde{D}^\ell, \qquad \ell \in T_L.$$

Dividing these inequalities, first by Q and then by the time lag σ, we obtain

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, v_{ik}^\ell \geq \left\lceil \frac{\left\lceil \frac{\widetilde{D}^\ell}{Q} \right\rceil}{\sigma} \right\rceil, \qquad \ell \in T_L. \tag{44}$$

Combining inequalities (43) and (44), and defining

$$M^\ell = \max \left\{ \left\lceil \frac{D^\ell}{Q} \right\rceil, \left\lceil \frac{\left\lceil \frac{\widetilde{D}^\ell}{Q} \right\rceil}{\sigma} \right\rceil \right\}$$

yields

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, v_{ik}^\ell \geq M^\ell, \qquad \ell \in T_L.$$

By applying the Chvátal–Gomory rounding procedure a finite number of times p ($p \geq 1$) to the above inequalities (Wolsey 1998), we obtain

$$\sum_{i \in I} \sum_{k=1}^{n_i} \left\lceil \frac{k}{p} \right\rceil v_{ik}^\ell \geq \left\lceil \frac{M^\ell}{p} \right\rceil, \qquad \ell \in T_L, \ p = 1, \ldots, \max_{i \in I}\{n_i\}. \tag{45}$$

The reasoning followed to derive (45) is itself a proof of the following result:

Proposition 1 *The set of inequalities* (45) *is valid for formulation* $(\mathbf{P}_{DE-1})$.

The enhanced formulation is called hereafter $(\mathbf{P}_{DE-1}^{+})$ for the scenario-independent location and capacity scalability model.

4.2 Inequalities for model $(\mathbf{P}_{DE-2})$

For the extensive form of the deterministic model given by (52)–(58) (see Appendix A), we develop additional inequalities that serve a purpose similar to (45). In this case, the variables $\widetilde{v}_{iks}^{\ell}$ (cf. (51)) representing the number of modular units available in a location at time period ℓ ($\ell \in T_L$) are scenario-indexed. Hence, the minimum quantity of demand that must be satisfied in each period between two consecutive design periods is scenario dependent and defined by

$$\Delta_s^{\ell} = \max \left\{ \sum_{j \in J^0} d_{js}^t : t = \ell, \ldots, \phi(\ell) \right\}, \qquad \ell \in T_L, \ s \in S.$$

Given a scenario s and a design period ℓ, it follows that the minimum number of open facilities must satisfy the following conditions:

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, \widetilde{v}_{iks}^{\ell} \geq \left\lceil \frac{\Delta_s^{\ell}}{Q} \right\rceil, \qquad \ell \in T_L, \ s \in S. \tag{46}$$

The above inequalities are the counterpart of (43) for the scenario-dependent capacity scalability model.

Additionally, let $\widetilde{\Delta}_s^{\ell}$ denote the minimum total demand requirements to be satisfied in scenario s for all periods from ℓ through $\phi(\ell)$, with $\ell \in T_L$.

$$\widetilde{\Delta}_s^{\ell} = \begin{cases} \displaystyle\sum_{j \in J^0} \sum_{t=\ell}^{\phi(\ell)} d_{js}^t + \sum_{j \in J^1} \sum_{t=\ell}^{\phi(\ell)-\rho_j} d_{js}^t & \text{if } \ell < \ell_{max}, \\[3ex] \displaystyle\sum_{j \in J} \sum_{t=\ell}^{|T|} d_{js}^t & \text{if } \ell = \ell_{max}. \end{cases}$$

Therefore, the following inequalities must also hold

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, \widetilde{v}_{iks}^{\ell} \geq \left\lceil \frac{\left\lceil \frac{\widetilde{\Delta}_s^{\ell}}{Q} \right\rceil}{\sigma} \right\rceil, \qquad \ell \in T_L, \ s \in S, \tag{47}$$

and using inequalities (46) and (47) yields the following relations:

$$\sum_{i \in I} \sum_{k=1}^{n_i} k \, \widetilde{v}_{iks}^{\ell} \geq \widetilde{M}_s^{\ell}, \qquad \ell \in T_L, \ s \in S,$$

where the lower bound $\widetilde{M}_s^{\ell}$ is defined as the maximum of the two parameters on the right-hand sides of inequalities (46) and (47). Finally, we apply once again the Chvátal–Gomory rounding method to the above inequalities and obtain the following result:

Table 1 Parameter values

Parameter	Value	Parameter	Value
$|T|$	12, 24	$|T_L|$	3
$|J|$	20, 40	$|J^0|$	$\beta^J \, |J|$ with $\beta^J \in \{0.25, \ 0.5, \ 0.75\}$
$|I|$	$0.25 \, |J|$	ρ	1, 2, 3
$|S|$	5	π_s	$1/5 \ (s \in S)$

Proposition 2 *The set of inequalities*

$$\sum_{i \in I} \sum_{k=1}^{n_i} \left\lceil \frac{k}{p} \right\rceil \widetilde{v}_{iks}^{\ell} \geq \left\lceil \frac{\widetilde{M}_s^{\ell}}{p} \right\rceil, \qquad \ell \in T_L, \ s \in S, \ p = 1, \ldots, \max_{i \in I}\{n_i\} \tag{48}$$

is valid for formulation $(\mathbf{P}_{DE-2})$.

The steps presented above serve as proof for this proposition. If for a particular design period ℓ and scenario s there exist several inequalities (48), all having the same right-hand side for different values of p, then we only need to consider the inequality with the strongest left-hand side. A similar remark also applies to (45). The enhanced formulation for the scenario-dependent capacity scalability case is called hereafter $(\mathbf{P}_{DE-2}^+)$.

5 Computational study

In this section, we present the results of an extensive computational study. The computational experiments were guided by five objectives, namely: (i) to evaluate the usefulness of using a general-purpose MILP solver to identify optimal or near-optimal solutions to the original deterministic equivalent models within reasonable computing time; (ii) to analyze the effectiveness of the proposed additional inequalities; (iii) to compare the two strategies for deciding on the timing and sizing of capacity adjustments; (iv) to discuss relevant insights into the characteristics of the (near-)optimal solutions identified; and (v) to assess the benefits of using a stochastic programming approach for the problem under study.

5.1 Characteristics of test instances

Since benchmark data sets are not available for the problem at hand, a set of instances was randomly generated. The size of each instance is mainly dictated by the length of the planning horizon and the total number of customers according to the choices specified in Table 1. Planning horizons with 12 and 24 time periods are considered. In both cases, there are three design periods for making location and capacity scalability decisions. Specifically, for $|T| = 12$ (resp. $|T| = 24$), these opportunities occur at the beginning of periods 1, 5, and 9 (resp. 1, 9, and 17).

Two different sizes for the customer set are assumed and three different partitions of J are considered. For example, a set of 40 customers can be partitioned into $|J^0| = 10$ and $|J^1| = 30$, $|J^0| = |J^1| = 20$, and $|J^0| = 30$ and $|J^1| = 10$. In all cases, customers in segment J^1 specify the same maximum number of periods for late deliveries, i.e., $\rho_j = \rho$ for all $j \in J^1$. Instances with $\rho = 1$, $\rho = 2$, and $\rho = 3$ were generated.

In total, five scenarios are considered that reflect different variations in demand. This choice is mainly driven by the very large size of our models, as displayed in Table 2, which is impacted by the number of periods and the number of customers selected. This difficulty is reflected in other similar research in stochastic facility location as well, thus limiting the total number of scenarios (Aghezzaf (2005); Alumur et al. (2012); Correia et al. (2018); Diglio et al. (2020); Heckmann (2016); Marín et al. (2018); Mete and Zabinsky (2010); Moreno et al. (2018)).

To model different trends in demand, the following continuous uniform distributions are considered: (i) $\mathcal{U}[10, 100]$, (ii) $\mathcal{U}[100, 200]$, and (iii) $\mathcal{U}[200, 300]$. Motivated by the works of Fernández et al. (2019) and Hinojosa et al. (2014) for a stochastic fixed-charge transportation problem, we have combined these stochastic distributions in different ways to create five scenarios. In scenario 1, all customer demands are generated from interval (i) in the first time period; in scenario 2, all demand levels are taken from interval (ii) in the first period; in scenario 3, interval (iii) is used to generate all demands in period 1; in scenarios 4 and 5, a demand level is initially chosen at random for each customer, among the three intervals. This sets the choice for period $t = 1$. Each interval has an equal probability of being selected. In every scenario s, the demands of customer j in all periods $t > 1$ are calculated as $\beta_{js}^t d_{js}^{t-1}$ with $\beta_{js}^t \in \mathcal{U}[0.8, 1.2]$. This scheme creates demand fluctuations over the planning horizon that will call for adjustments in the capacity of the facilities. We notice that these scenarios differ significantly from each other, originating a high demand variability, and thus making them a suitable choice in a setting with uncertain demand.

The capacity of a modular unit is defined as a function of the demand scenarios, namely $Q = \frac{1}{n_i |I|} \mathcal{U}[3, 4] \frac{\sum_{j \in J} \sum_{t \in T} \max_{s \in S} d_{js}^t}{|T|}$, with $n_i = 4$ for every $i \in I$. Appendix B provides further details on the procedure for randomly generating all cost parameters. Five instances were randomly generated for each combination of the parameters shown in Table 1, giving rise to a total of 180 different instances. Table 2 summarizes the sizes of these instances according to the formulation chosen. In particular, the number of binary variables increases significantly in the scenario-dependent capacity scalability model, namely more than 360%. Together with the considerable number of continuous variables, especially for the instances with 40 customers, we notice that the models' dimensions are challenging in a stochastic MILP context. With the valid inequalities proposed in Sect. 4, the total number of constraints slightly increases in both enhanced formulations.

5.2 Numerical results

Formulations ($\mathbf{P}_{DE-1}$) and ($\mathbf{P}_{DE-2}$) along with their enhancements were coded in C++ using IBM ILOG Concert Technology and run with IBM ILOG CPLEX 12.6. All experiments were performed on a workstation with a multi-core Intel Xeon E5-2650V3 processor (2.3 GHz, 10 cores), 32 GB RAM, and running the Ubuntu operating system (64-bit). Due to the strategic nature of the problem that we study, fast solution times are not of paramount importance. Therefore, a limit of 10 hours of CPU time was set for each solver run. Furthermore, CPLEX was used with default settings under a deterministic parallel mode.

Regarding goal (i) listed at the beginning of Sect. 5, we evaluate the effectiveness of using CPLEX to identify (near-)optimal solutions to the test instances for the proposed original formulations, ($\mathbf{P}_{DE-1}$) and ($\mathbf{P}_{DE-2}$). To do so, we compare the optimality gaps and the computing times reported by CPLEX. Table 3 summarizes the relevant information according to the size of the customer set, while Table 4 provides information from a different perspective, namely for varying values for the maximum delivery delay tolerated by customer

Table 2 Average size of the test instances for formulations $(\mathbf{P}_{DE-1})$ and $(\mathbf{P}_{DE-2})$

$|J|$	$|T|$	$(\mathbf{P}_{DE-1})$
		    # binary variables
			$(\mathbf{P}_{DE-2})$
			    # binary variables
				$(\mathbf{P}_{DE-1}), (\mathbf{P}_{DE-2})$
				    # continuous variables
							$(\mathbf{P}_{DE-1})$
							    # const.
								$(\mathbf{P}_{DE-2})$
								    # const.

				$\rho = 1$	$\rho = 2$	$\rho = 3$		
20	12	180	660	10,200	13,200	16,200	2745	2905
	24	180	660	20,400	26,400	32,400	5445	5605
40	12	360	1320	38,400	50,400	62,400	5490	5810
	24	360	1320	76,800	100,800	124,800	10,890	11,210

segment J^1. Column 1 gives the parameter selected ($|J|$ in Table 3 and ρ in Table 4). Columns 2–5 report the number of instances solved to optimality (# opt sol.) and the number of instances not solved to proven optimality within the specified time limit (# non-opt sol.) for all models (i.e., the original formulations and their enhancements). For those instances *not* solved optimally, the minimum, average, and maximum optimality gaps achieved by CPLEX after 10 h of computing time are given for each formulation in columns 7–10 (MIP gap (%) $= (z^{UB} - z^{LB})/z^{UB} \times 100\%$, with z^{UB} denoting the objective value of the best feasible solution available and z^{LB} representing the best lower bound identified during the branch-and-cut procedure). Columns 11–14 display the minimum, average, and maximum computing times, in seconds (MIP CPU (s)). The last row refers to all 180 instances. The best (average) values with respect to the evaluation criteria are given in bold type.

Table 3 shows that the capability of CPLEX to identify an optimal solution within the given time limit is affected by the formulation that is used. Regarding formulations ($\mathbf{P}_{DE-1}$) and ($\mathbf{P}_{DE-2}$), optimality is achieved in all test instances with 20 customers, but deteriorates as the total number of customers grows to 40. Not surprisingly, the larger formulation ($\mathbf{P}_{DE-2}$) poses an additional challenge to CPLEX. In particular, for this formulation, optimal solutions are only available for less than half of the instances with 40 customers (44.4% or 40/90), whereas 80% (72/90) of the instances with the same number of customers could be solved to optimality with model ($\mathbf{P}_{DE-1}$). These findings are also supported by the growth observed in the computing times. Interestingly, for those instances that could not be solved to proven optimality, the associated gaps are, on average, relatively small (cf. columns 7 and 9 in Table 3). The challenges posed by formulation ($\mathbf{P}_{DE-2}$) result in slightly larger gaps compared to formulation ($\mathbf{P}_{DE-1}$).

Table 3 also reveals the positive effect of the proposed valid inequalities (45) and (48), an aspect that was mentioned at the beginning of Sect. 5 as goal (ii) to be investigated. Specifically, the additional constraints prove to be very useful since they allow the identification of optimal solutions for a significantly larger number of problem instances in shorter computing times. Adding inequalities (45) to formulation ($\mathbf{P}_{DE-1}$) results in obtaining optimal solutions to all the instances. Regarding the enhanced model ($\mathbf{P}_{DE-2}^+$), and despite the greater size of this formulation, 95% (171/180) of all instances are solved to optimality. In this case, the valid inequalities (48) prove to be very effective for the more challenging instances with 40 customers (90% or 81/90 optimal solutions become available).

Table 4 shows that increasing the maximum lead time for customers accepting late shipments results in slightly more challenging problems for CPLEX. This fact is evidenced by the larger number of instances not solved to proven optimality as the value of ρ grows. We remark that increasing ρ brings more flexibility to designing the facility network and adjusting the capacities of the operating facilities over the planning horizon. Interestingly, the quality of the best solutions returned by CPLEX upon reaching the time limit does not seem to be considerably affected by parameter ρ. The enhanced formulations clearly succeed in solving more instances to optimality and in providing lower optimality gaps under a substantial reduction of the computational burden.

As evidenced by the significantly shorter computing times of the tighter formulations ($\mathbf{P}_{DE-1}^+$) and ($\mathbf{P}_{DE-2}^+$) (cf. last column in Tables 3, 4), the exploration of the branch-and-cut tree generated by CPLEX is less time consuming, despite the greater number of constraints. However, additional computational effort may be required to solve the associated linear relaxations. This feature is present in solving the LP-relaxation of model ($\mathbf{P}_{DE-2}^+$) but not for the LP-relaxation of model ($\mathbf{P}_{DE-1}^+$), as displayed in Tables 5 and 6. These tables report in the last four columns the minimum, average, and maximum solution times of the

Table 3 Optimality gaps and CPU times for all formulations according to the total number of customers

$|J|$	# opt sol./# non-opt sol.					MIP gap (%)*				MIP CPU (s)			
	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^+_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^+_{DE-2})$		$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^+_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^+_{DE-2})$	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^+_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^+_{DE-2})$
20	90/0	90/0	90/0	90/0	Min	–	–	–	–	11.0	3.3	13.9	5.3
					Avg	–	–	–	–	187.2	38.1	318.1	52.8
					Max	–	–	–	–	2132.8	168.8	8956.9	540.1
40	72/18	90/0	40/50	81/9	Min	0.77	–	0.61	0.85	51.7	38.4	183.1	72.6
					Avg	4.40	–	5.14	4.80	12,902.9	1004.8	25,556.2	7174.2
					Max	7.85	–	9.92	8.09	**	16,642.9	**	**
All	162/18	**180/0**	130/50	**171/9**	Avg	4.40	–	5.14	**4.80**	6545.0	**521.5**	12,937.2	**3613.5**

*MIP gap for instances not solved to optimality within 10 h; **time limit (10 h) reached

Table 4 Optimality gaps and CPU times for all formulations under different values for the maximum delivery delay, ρ

ρ		# opt sol./# non-opt sol.				MIP gap (%)*				MIP CPU (s)			
		$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^{+}_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^{+}_{DE-2})$	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^{+}_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^{+}_{DE-2})$	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^{+}_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^{+}_{DE-2})$
1	Min	55/5	60/0	46/14	59/1	0.77	–	0.61	0.85	11.0	3.3	18.9	5.3
	Avg					3.83	–	4.97	0.85	5767.5	191.2	10,510.2	1519.0
	Max					5.26	–	8.58	0.85	**	1846.6	**	**
2	Min	53/7	60/0	43/17	57/3	2.35	–	2.56	2.31	14.3	6.5	20.0	6.3
	Avg					4.76	–	5.43	5.32	6854.9	401.5	13,734.5	3679.2
	Max					7.85	–	9.92	8.09	**	4239.4	**	**
3	Min	54/6	60/0	41/19	55/5	2.75	–	1.19	3.44	24.3	5.4	13.9	7.0
	Avg					4.44	–	5.01	5.27	7012.8	971.7	14,566.8	5642.5
	Max					7.16	–	7.59	7.14	**	16,642.9	**	**
All	Avg	162/18	**180/0**	130/50	**171/9**	4.40	-	5.14	**4.80**	6545.0	**521.5**	12,937.2	**3613.5**

*MIP gap for instances not solved to optimality within 10 h; **time limit (10 h) reached

linear relaxation (LP CPU (s)) of all models, for different values of parameter $|J|$ (Table 5) and parameter ρ (Table 6). In addition, columns 3–6 show the minimum, average, and maximum integrality gaps (LP gap (%)). The latter give the relative percent deviation between the objective value of the best feasible solution available (z^{UB}) and the lower bound provided by the linear relaxation. The best average values over all 180 instances are highlighted in bold type in the last row of these tables.

Tables 5 and 6 indicate that the linear relaxations of the enhanced formulations provide substantially better lower bounds, and as a result lower integrality gaps, as opposed to the original models. Even though the number of available optimal solutions decreases with a growing size of J, and accordingly the quality of z^{UB} declines, the average integrality gap becomes significantly smaller when inequalities (45) and (48) are added to the associated formulations. By contrast, the magnitude of the maximum allowed tardiness for orders in customer segment J^1 does not seem to have a major effect on the average integrality gap in any of the four formulations, as shown in Table 6.

Finally, we note that good integrality gaps such as the ones we have obtained with the enhanced formulations can be very useful in evaluating the quality of a feasible solution produced, for example, by a tailored heuristic method. In particular, this feature is relevant for problems with large customer sets since in this case it becomes increasingly difficult for a state-of-the-art optimization solver to achieve optimality within acceptable time. From a computational viewpoint, the optimal solution to the linear relaxation is inexpensive (< 200 seconds for the enhanced models, see Tables 5 and 6).

5.3 Deployment of facility location and capacity scalability strategies

In this section, we focus on goals (iii) and (iv) stated at the beginning of Sect. 5, and compare the characteristics of the two strategies proposed for locating facilities and adjusting their capacities over the planning horizon. For this purpose, Table 7 gives the minimum, average, and maximum relative deviation of the objective values of the best solutions reported by CPLEX for the enhanced formulations. Specifically, columns 2–4 (resp. 6–8) display the value of $(z^+_{DE-1} - z^+_{DE-2})/z^+_{DE-1} \times 100\%$ according to the size of the customer set (resp. the maximum delivery delay). We denote by z^+_{DE-1} and z^+_{DE-2} the best objective values identified by CPLEX to formulation ($\mathbf{P}^+_{DE-1}$) and ($\mathbf{P}^+_{DE-2}$), respectively, within the given time limit. Each row associated with the size of a customer set aggregates the results for 90 instances, while each row associated with parameter ρ aggregates the results for 60 instances. Since the relative deviations presented in this table are all positive, they indicate that cost savings can be achieved with the scenario-dependent capacity scalability model (strategy 2) over the scenario-independent location and scalability model (strategy 1). This finding suggests that modeling capacity adjustment decisions as second-stage variables in formulation ($\mathbf{P}^+_{DE-2}$) allows a more agile response to temporal changes in demand. In other words, these decisions can be made taking into account the realization of demand. Therefore, in general, the resulting schedule for facility opening and capacity adjustment is cheaper compared to the case in which all design and capacity planning decisions must be made in the first stage. In the latter case, such decisions do not depend on the particular realization of the random demand process.

Over all instances, strategy 2 allows for up to 10.92% reduction in total cost, with an average reduction of 4.5%. Even when an instance is not solved to optimality with formulation ($\mathbf{P}^+_{DE-2}$) and an optimal solution is available to formulation ($\mathbf{P}^+_{DE-1}$), strategy 2 always

Table 5 Integrality gaps and solution times of the LP relaxation of all formulations according to the total number of customers

$|J|$		LP gap (%)				LP CPU (s)			
		$(\mathbf{P}_{DE-1})$	$(\mathbf{P}_{DE-1}^{+})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}_{DE-2}^{+})$	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}_{DE-1}^{+})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}_{DE-2}^{+})$
20	Min	8.29	0.76	8.38	1.21	0.2	0.2	0.6	0.3
	Avg	14.92	3.86	15.92	3.61	5.1	7.1	2.0	1.4
	Max	21.15	7.14	22.90	7.87	40.0	62.3	10.5	4.1
40	Min	2.40	1.29	2.70	1.52	3.3	4.7	4.9	5.7
	Avg	6.22	3.11	6.87	3.29	22.1	27.9	16.0	10.8
	Max	11.59	4.60	11.70	9.77	100.8	192.5	57.6	29.5
All	Avg	10.57	**3.49**	11.40	**3.45**	**13.6**	17.5	9.0	**6.1**

Table 6 Integrality gaps and solution times of the LP relaxation of all formulations under different values for the maximum delivery delay, ρ

ρ		LP gap (%)				LP CPU (s)			
		$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^+_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^+_{DE-2})$	$(\mathbf{P}_{DE-1})$	$(\mathbf{P}^+_{DE-1})$	$(\mathbf{P}_{DE-2})$	$(\mathbf{P}^+_{DE-2})$
1	Min	2.40	0.76	3.00	1.21	0.4	0.2	0.6	0.5
	Avg	11.21	3.05	11.81	2.85	8.4	9.7	5.8	5.4
	Max	21.15	7.12	22.90	5.60	35.4	37.8	31.1	15.3
2	Min	2.40	1.30	3.38	1.76	0.2	0.2	0.6	0.3
	Avg	10.58	3.63	11.54	3.50	14.4	12.1	8.4	7.5
	Max	19.63	6.78	22.54	9.77	72.4	49.9	35.6	29.5
3	Min	2.40	1.29	2.70	1.37	0.6	0.7	1.1	0.5
	Avg	9.91	3.78	10.85	4.00	18.1	30.7	12.8	5.4
	Max	19.63	7.14	22.21	8.39	100.8	192.5	57.6	15.3
All	Avg	**10.57**	**3.49**	**11.40**	**3.45**	**13.6**	**17.5**	**9.0**	**6.1**

Table 7 Comparison between the best objective values of formulations ($\mathbf{P}^+_{DE-1}$) and ($\mathbf{P}^+_{DE-2}$) for different parameters

$|J|$	% Cost reduction of ($\mathbf{P}^+_{DE-2}$) over ($\mathbf{P}^+_{DE-1}$)			ρ	% Cost reduction of ($\mathbf{P}^+_{DE-2}$) over ($\mathbf{P}^+_{DE-1}$)		
	Min	Avg	Max		Min	Avg	Max
20	0.55	3.86	10.92	1	0.55	3.47	7.03
40	1.50	5.06	10.60	2	0.64	4.50	8.82
				3	0.68	5.41	10.92

has lower total cost than strategy 1. This particular case occurs in 9 instances (5% of the instances).

To further analyze the differences between the two strategies, Table 8 provides additional information about the selected facilities for each combination of potential facilities ($|I|$), number of customers ($|J|$), and value of ρ (columns 1–3). Average results for strategy 1, resp. strategy 2, are summarized in columns 4–7, resp. columns 8–11. Since adjustments in the number of modular units at open facilities are second-stage decisions in strategy 2, the average values shown in columns 8–10 are taken over all scenarios. The average delayed demand is also obtained over all scenarios for both strategies (columns 7 and 11).

Not surprisingly, the total number of open facilities is, on average, similar in both strategies. This is explained by facility opening decisions being modeled as first-stage variables in both approaches. Furthermore, decisions on upgrading and downgrading the capacities of operating facilities are driven by the type of strategy used. When such choices are included in the first-stage (strategy 1), there is a tendency to provide more capacity through installing additional modular units. Capacity contraction hardly ever occurs. By contrast, capacity adjustments are more noticeable in strategy 2, especially with regard to removing modular units. The additional cost incurred by reducing the capacity of a facility is offset by a lower maintenance cost for that facility. As expected, strategy 2 offers greater flexibility to respond to demand changes over the planning horizon, thereby yielding network configurations with lower total cost (cf. Table 7).

In general, and independently of the strategy followed to determine a schedule for facility opening and capacity adjustment, the percentage of late deliveries increases with the maximum allowed delay. This feature is more evident in the instances with 20 customers.

5.4 The relevance of the stochastic setting

This section focuses on goal (v) of the computational study, namely on the evaluation of the benefits of the stochastic programming approach. This is a relevant issue because of the additional complexity posed by the stochastic setting compared to its deterministic counterpart. Even though there are no standard 'robust measures' for assessing the relevance of using a stochastic model instead of a (simpler) deterministic model (Birge and Louveaux 2011), two metrics are often used. The first is the *expected value of perfect information* (EVPI), while the second is the *value of the stochastic solution* (VSS). Many studies in stochastic facility location have used these metrics (Correia et al. 2018; Diglio et al. 2020; Heckmann 2016; Hinojosa et al. 2014; Marín et al. 2018; Moreno et al. 2018). We have calculated the EVPI and VSS to measure the benefit of working with the stochastic programs ($\mathbf{P}^+_{DE-1}$) and ($\mathbf{P}^+_{DE-2}$).

As discussed by Birge and Louveaux (2011), EVPI and VSS are distinct indicators that assess different types of uncertainty. The EVPI represents the maximum amount that the

Table 8 Number of selected facilities and demand served with delay to customer segment J^1 in strategy 1 (formulation ($\mathbf{P}^+_{DE-1}$)) and strategy 2 (formulation ($\mathbf{P}^+_{DE-2}$))

$|I|$	$|J|$	ρ	Scenario-independent capacity adjustments ($\mathbf{P}^+_{DE-1}$)				Scenario-dependent capacity adjustments ($\mathbf{P}^+_{DE-2}$)			
			Avg no. of open facilities	Avg no. of facilities with cap. adjustments		Avg delayed demand (%)	Avg no. of open facilities	Avg no. of facilities with cap. adjustments		Avg delayed demand (%)
				Expansion	Contraction			Expansion	Contraction	
5	20	1	2.00	0.30	0.07	0.45	2.00	0.08	0.73	0.47
		2	2.00	0.47	0.03	1.74	2.00	0.15	0.56	1.60
		3	2.00	0.63	0.00	3.08	2.00	0.27	0.48	3.00
10	40	1	3.03	0.83	0.00	0.46	3.03	0.17	1.33	0.38
		2	3.03	1.00	0.00	0.86	3.07	0.24	1.13	0.76
		3	3.03	1.33	0.00	1.53	3.07	0.33	0.91	1.24
All instances			2.52	0.76	0.02	1.35	2.53	0.21	0.86	1.25

decision-maker would be willing to pay in return for complete information about the future. It is defined as the difference between the optimal value of the stochastic program (in our case, z_{DE-1}^+ or z_{DE-2}^+) and the optimal value of the so-called 'wait-and-see' (WS) problem. For each demand scenario s, formulation ($\mathbf{P}_{DE-1}^+$) (or ($\mathbf{P}_{DE-2}^+$)) is solved under the assumption that the scenario occurs with probability 1. In this way, we obtain the optimal decisions that should be made if it were possible to fully predict future demands. Let $z_{DE-1,s}^+$ (resp. $z_{DE-2,s}^+$) denote the optimal value of the deterministic problem associated with a particular scenario $s \in S$ and strategy 1 (resp. strategy 2). It follows that the optimal value of the WS problem is given by $\sum_{s \in S} \pi_s z_{DE-1,s}^+$ for strategy 1. A similar calculation is performed for strategy 2.

The VSS measures the difference between the optimal objective value of the so-called 'expected value problem' (EEV) and the optimal objective value of the stochastic program. The former is obtained by creating a single scenario in which all random variables are replaced by their expected values, and then determining the optimal values of the first-stage variables for that single scenario. Next, the expected cost over all scenarios using those first-stage variable values is calculated. In the case of our problem, the average scenario would be constructed by taking the average demand of every customer in each time period, $d_j^t = \sum_{s \in S} \pi_s d_{js}^t$ ($j \in J$; $t \in T$). Unfortunately, the associated values of the first-stage variables may have no feasible completion in the second-stage, thus resulting in an infeasible stochastic problem, either for model ($\mathbf{P}_{DE-1}^+$) or model ($\mathbf{P}_{DE-2}^+$), or even both. To overcome this difficulty, a 'reference scenario' is often used (Birge and Louveaux 2011). Typically, the worst-case scenario, i.e. the scenario with the highest total demand level for all periods is adopted. Due to the features of our data, it is not possible to identify such scenario (recall that the demand of every customer may decline or grow between two consecutive periods, and in scenarios 4 and 5, a random choice of the interval from which demand is generated takes place for each customer and $t = 1$). Hence, a so-called 'modified reference scenario' was determined by taking $d_j^t = \max_{s \in S} \left\{ d_{js}^t \right\}$ for every $j \in J$ and $t \in T$. Summarizing, the EVPI measures the value of knowing the future with certainty, while the VSS indicates how good the optimal solution to the EEV is as an approximation to the optimal solution to the stochastic program.

Tables 9 and 10 report the average relative values of EVPI (column 5) and modified VSS (column 7) for strategy 1 and strategy 2, respectively. The relative values are calculated by dividing each metric by the optimal objective value of the associated stochastic problem. Column 4 indicates the number of instances for which optimal solutions are available and thus were used for calculating these average relative gaps. In addition, columns 6 and 8 give the average total computing time required to solve the WS and EEV problems to optimality, respectively. The EVPI and modified VSS gaps reported in the tables indicate that it is appropriate to consider a stochastic approach rather than its deterministic counterpart. For the scenario-independent location and capacity scalability model (strategy 1), the relative EVPI gap ranges from 12.59% to 28.83%, with an average of 21.12%. Concerning the scenario-dependent capacity scalability model (strategy 2), slightly smaller EVPI gaps are obtained (average: 17.35%) that vary between 9.83% and 25.83%. In both cases, these relatively large values reveal that having access to perfect information about future customer demands would be quite valuable. This is an indication of a high degree of uncertainty, which, in turn, demonstrates the usefulness of considering a stochastic approach. Interestingly, the average EVPI gap does not seem to be greatly affected neither by the size of the customer set nor by the maximum delivery delay tolerated by customer segment J^1.
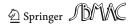
Table 9 EVPI and modified VSS gaps for the scenario-independent location and capacity scalability model

$|I|$	$|J|$	ρ	# opt sol. $(\mathbf{P}_{DE-1}^{+})$	Metric 1		Metric 2	
				Avg EVPI gap (%)	Avg total CPU (s)	Avg modified VSS gap (%)	Avg total CPU (s)
5	20	1	30	21.65	127.5	5.43	11.0
		2	30	21.24	169.2	5.93	12.3
		3	30	20.81	191.8	5.89	15.9
10	40	1	30	21.01	186.0	9.25	174.3
		2	30	21.11	241.3	8.98	271.7
		3	30	20.90	343.5	9.06	579.7
All instances			180	21.12	209.9	7.42	177.5

Table 10 EVPI and modified VSS gaps for the scenario-dependent capacity scalability model

$|I|$	$|J|$	ρ	# opt sol. $(\mathbf{P}^+_{DE-2})$	Metric 1		Metric 2	
				Avg EVPI gap (%)	Avg total CPU (s)	Avg modified VSS gap (%)	Avg total CPU (s)
5	20	1	30	19.43	76.3	1.27	21.7
		2	30	17.97	38.7	2.42	18.5
		3	30	16.77	32.2	2.44	16.5
10	40	1	29	17.69	64.9	7.27	425.5
		2	27	16.67	104.0	7.05	350.0
		3	25	15.16	118.3	8.19	752.6
All instances			171	17.35	70.6	4.62	251.6

For the modified VSS gap, we observe that it gradually increases with growing value of ρ. Recall that this metric measures how far away the value of the optimal solution of a deterministic simplified version of the problem is from the optimal value of a stochastic problem. By increasing ρ, more opportunities arise to satisfy the demand of some customers with delay, which, in turn, enable a lower investment in designing the facility network (e.g., through reducing the need to establish more facilities). This feature becomes even more relevant in a stochastic setting, whereas the quality of the optimal solution to the deterministic EEV problem deteriorates. In other words, the approximation of the stochastic model by one where expected values are used is poor. Finally, and as expected, the computational burden required to optimally solve the WS and EEV problems is considerably lower than solving the stochastic programming models.

6 Conclusions

We have studied a novel multi-period facility location and capacity planning problem under uncertainty taking into account the sensitivity of each individual customer to delivery lead times. In addition to determining a schedule for opening facilities and setting their initial number of modules, this new problem also allows for two different ways to adjust capacity over the planning horizon. Specifically, capacities can be increased incrementally by installing additional modules or decreased by removing existing modular units. For the case of uncertain future customer demand, we have proposed two stochastic models that differ in the framework in which capacity adjustments can be planned. The first two-stage stochastic model follows the natural approach of defining all decisions related to the design of the facility network as first-stage decisions. Demand allocation decisions are deferred to the second stage. This model is suitable to real-world settings where opening facilities and sizing their capacities are time-consuming and require substantial capital investment. The second model presents an alternative strategy which is particularly relevant when capacity adjustment decisions have a tactical nature, e.g., when short commitment horizons for lease or rental contracts are offered. In this case, first-stage decisions include the selection of locations to open facilities and setting their initial capacities. The phased capacity expansion and contraction of facilities along with demand allocation represent the recourse decisions. Naturally, it is up to the decision-maker to decide which of the approaches should be followed depending on the problem at hand. Under the assumption that demand uncertainty can be captured by a finite set of scenarios, we have presented the extensive form of the deterministic equivalent for the two stochastic programs. Additionally, valid inequalities were developed to enhance the resulting large-scale formulations.

Numerical experiments with randomly generated instances indicate that the proposed additional inequalities substantially increase the capability of an off-the-shelf MILP solver to identify optimal solutions in much shorter computing times. Moreover, improved solution quality is also achieved for instances for which the specified time limit is reached without guaranteed optimality. Deferring the decisions on capacity expansion and capacity contraction to the second-stage problem results in network configurations with noticeable lower total cost compared to the approach where such decisions belong to the first-stage problem. This cost advantage arises due to the possibility of adapting the network configuration to particular demand realizations. However, this approach leads to considerably larger models, which, in turn, require higher computational burden. Our comparative analysis enables the decision-maker to understand the opportunities that arise from different strategies and helps

raise awareness of the potential benefits of conducting this type of study. Despite the average number of open facilities being similar in the two approaches, decisions on upgrading or downgrading the capacities of operating facilities differ. In particular, greater flexibility to respond to demand fluctuations is displayed by the second strategy. Thus, a firm having access to e.g. on-demand warehousing services will be aware of the benefits to be gained. Finally, we have also shown that it is worth adopting a modeling framework which captures the uncertainty in future customer demand despite the additional complexity.

From a practical viewpoint, the NP-hard problems that we study pose significant challenges as they involve a very large number of variables and constraints. Hence, obtaining an optimal solution within reasonable computing time with a general-purpose optimization solver becomes extremely difficult for large-sized problems. This characteristic is also encountered in many other multi-period facility location problems under uncertainty (e.g., Correia et al. (2018)), and calls for the development of specially tailored decomposition approaches and heuristic methods. In this case, the enhancements proposed in Sect. 4 would also be helpful as they sharpen the linear relaxation. The LP-bounds would also be useful to evaluate the quality of the solutions returned by a heuristic procedure. Moreover, it would be interesting to investigate the application of a sampling-based approach to handle problems with a very large number of scenarios. In further work, the proposed models could also be extended by considering the uncertainty present in other parameters (e.g., costs) along with the unknown demand. Finally, another line of research could be directed toward the development of a multi-stage stochastic framework to enable the revision of the planning decisions as more information regarding the uncertain data is revealed. Even though a multi-stage model yields a better characterization of the dynamic planning process, solution techniques remain computationally challenging (Bakker et al. 2020). However, recent developments in stochastic nested decomposition techniques seem to be a promising venue for future research (Zou et al. 2019).

Appendix A: Deterministic equivalent formulation of the scenario-dependent capacity scalability problem

For the scenario-dependent capacity scalability problem introduced in Sect. 3.3, the extensive form of the deterministic equivalent model of the two-stage stochastic program uses the continuous flow variables (25)–(27) and the following (scenario-indexed) binary variables

that define the capacity adjustment decisions over the planning horizon:

$\widetilde{e}_{iks}^{\ell} = 1$ if the capacity of a facility in location i is expanded with k modular units at time period ℓ under scenario s, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L \backslash \{1\}$; $s \in S$).   (49)

$\widetilde{r}_{iks}^{\ell} = 1$ if the capacity of a facility in location i is reduced by removing k modular units at time period ℓ under scenario s, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i - 1$; $\ell \in T_L \backslash \{1\}$; $s \in S$).   (50)

$\widetilde{v}_{iks}^{\ell} = 1$ if a facility is operated at location i with k modular units at time period ℓ under scenario $s \in S$, 0 otherwise ($i \in I$; $k = 1, \ldots, n_i$; $\ell \in T_L$; $s \in S$).   (51)

Formulation ($\mathbf{P}_{DE-2}$) is given by

$$
\text{Min} \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} FO_{ik}^{\ell} z_{ik}^{\ell} + \sum_{s \in S} \pi_s \left[\sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i - 1} FE_{ik}^{\ell} \widetilde{e}_{iks}^{\ell} \right.
$$

$$
+ \sum_{\ell \in T_L \backslash \{1\}} \sum_{i \in I} \sum_{k=1}^{n_i - 1} FR_{ik}^{\ell} \widetilde{r}_{iks}^{\ell} + \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k=1}^{n_i} \widetilde{v}_{iks}^{\ell} \sum_{t=\ell}^{\phi(\ell)} M_{ik}^t
$$

$$
+ \sum_{t \in T} \sum_{i \in I} \sum_{k=1}^{n_i} o_{ik}^t w_{iks}^t + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^0} c_{ij}^t x_{ijs}^t
$$

$$
\left. + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j, |T|\}} \left(c_{ij}^{t'} + p_j^{tt'} \right) y_{ijs}^{tt'} \right]
$$

subject to   (52)

(10), (14), (29), (30), (32) − (35)

$$
\sum_{k=1}^{n_i} \widetilde{v}_{iks}^{\ell} = \sum_{\ell' \in T_L: \ell' \leq \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \quad i \in I, \ell \in T_L, s \in S \tag{53}
$$

$$
\sum_{k=1}^{n_i - 1} \widetilde{e}_{iks}^{\ell} + \sum_{k=1}^{n_i - 1} \widetilde{r}_{iks}^{\ell} \leq \sum_{\ell' \in T_L: \ell' < \ell} \sum_{k=1}^{n_i} z_{ik}^{\ell'} \quad i \in I, \ell \in T_L \backslash \{1\}, s \in S \tag{54}
$$

$$
\sum_{k=1}^{n_i} k \, \widetilde{v}_{iks}^{\ell} = \sum_{\ell' \in T_L: \ell' \leq \ell} \sum_{k=1}^{n_i} k \, z_{ik}^{\ell'}
$$

$$
+ \sum_{\ell' \in T_L \backslash \{1\}: \ell' \leq \ell} \sum_{k=1}^{n_i - 1} k \left(\widetilde{e}_{iks}^{\ell'} - \widetilde{r}_{iks}^{\ell'} \right) \quad i \in I, \ell \in T_L, s \in S \tag{55}
$$

$$
w_{iks}^t \leq k \, Q \, \widetilde{v}_{iks}^{\ell} \quad i \in I, k = 1, \ldots, n_i, t \in T,
$$
$$
\ell = \max\{\ell' \in T_L : \ell' \leq t\}, s \in S \tag{56}
$$

$$
\widetilde{v}_{iks}^{\ell} \in \{0, 1\} \quad i \in I, k = 1, \ldots, n_i, \ell \in T_L, s \in S \tag{57}
$$

$$
\widetilde{e}_{iks}^{\ell}, \widetilde{r}_{iks}^{\ell} \in \{0, 1\} \quad i \in I, k = 1, \ldots, n_i - 1,
$$
$$
\ell \in T_L \backslash \{1\}, s \in S \tag{58}
$$

Appendix B: Cost generation

The cost generation scheme used in our test instances is adapted from a deterministic variant of our problem, Correia and Melo (2017). The generation of the variable costs relies on two real random numbers, β_1 and β_2, in the range 1.01 and 1.03 (details are given next). For $i \in I$ and $j \in J$, the unit variable distribution costs are generated as follows:

$$
\begin{aligned}
c_{ij}^1 &= \mathcal{U}[5, 10], \\
c_{ij}^t &= c_{ij}^1 && \text{for } t = 2, \ldots, \sigma, \\
c_{ij}^t &= \beta_1 \, c_{ij}^{12} && \text{for } t = 1 + \sigma, \ldots, 2\sigma, \\
c_{ij}^t &= \beta_2 \, c_{ij}^{24} && \text{for } t = 1 + 2\sigma, \ldots, |T|.
\end{aligned}
$$

Recall that $\sigma = |T|/|T_L|$ gives the total number of periods between two consecutive design periods. Since $|T_L| = 3$, the planning horizon is divided into three sections. For $|T| = 12$ (resp. $|T| = 24$), each section spans 4 (resp. 8) time periods. The unit distribution costs are assumed to be constant in all time periods within a section. However, the costs increase between 1% and 3% from one section to the next. This type of pattern is also present in the unit variable processing costs. The latter reflect economies of scale by considering the number of modular units that are available at a particular location. For $i \in I$, we set

$$
\begin{aligned}
o_{ik}^1 &= 100 / \sqrt{Q} && \text{for } t = 1, \ldots, \sigma, \\
o_{ik}^t &= 0.9 \, o_{ik-1}^t && \text{for } t = 2, \ldots, \sigma; \; k = 2, \ldots, n_i, \\
o_{ik}^t &= \beta_1 \, o_{ik}^\sigma && \text{for } t = 1 + \sigma, \ldots, 2\sigma; \; k = 2, \ldots, n_i, \\
o_{ik}^t &= \beta_2 \, o_{ik}^{2\sigma} && \text{for } t = 1 + 2\sigma, \ldots, |T|.
\end{aligned}
$$

Recall that at most four modular units may be available at a location, i.e. $n_i = 4$ $(i \in I)$. Regarding the generation of the fixed costs, the following procedure was employed.

- At the first design period, $\ell = 1$, the fixed cost of opening facility i $(i \in I)$ with k modular units $(k = 1, \ldots, n_i)$ is defined by

$$
FO_{ik}^1 = \alpha_i + \gamma_i \sqrt{kQ}, \quad \alpha_i \in \mathcal{U}[500, 1000]; \; \gamma_i \in \mathcal{U}[4000, 6000].
$$

  In the remaining strategic periods $\ell \in T_L \backslash \{1\}$, we set $FO_{ik}^\ell = \mathcal{U}[1.01, 1.03] \, FO_{ik}^{\ell-1}$ for every $k = 1, \ldots, n_i$. Once again, the capacity size of a facility at the time it is established also affects the above costs.

- All remaining fixed costs at location i $(i \in I)$ represent a given percentage of the fixed opening costs:

$$
\begin{aligned}
FE_{ik}^\ell &= 0.25 \, FO_{ik}^\ell && k = 1, \ldots, n_i - 1; \; \ell \in T_L \backslash \{1\}, \\
FR_{ik}^\ell &= 0.10 \, FO_{ik}^\ell && k = 1, \ldots, n_i - 1; \; \ell \in T_L \backslash \{1\}, \\
M_{ik}^t &= 0.05 \, FO_{ik}^\ell && k = 1, \ldots, n_i; \; \ell \in T_L; \; t = \ell, \ldots, t'; \; t' \in T : t' < \ell + 1.
\end{aligned}
$$

Finally, we adapt the scheme proposed in Correia and Melo (2017) to generate the unit tardiness penalty costs for late deliveries to customers $j \in J^1$ as follows:

$$
p_j^{tt'} = 0.01 \, \theta_j^t \, (t' - t)^2 \quad \text{for } t \in T \text{ and } t' = t, \ldots, \min\{t + \rho_j, |T|\},
$$

with

$$
\theta_j^t = \frac{\displaystyle\sum_{i \in I} \sum_{k=1}^{n_i} M_{ik}^t}{TD \, |I| \displaystyle\sum_{i \in I} n_i} + \frac{\displaystyle\sum_{i \in I} c_{ij}^t}{|I|} + \frac{\displaystyle\sum_{i \in I} \sum_{k=1}^{n_i} o_{ik}^t}{|I| \displaystyle\sum_{i \in I} n_i},
$$

and TD denoting the maximum quantity demanded across all scenarios. Therefore, $TD = \sum_{j \in J} \sum_{t \in T} \max_{s \in S} d_{js}^t$. Observe that the unit tardiness penalty cost is a function of the average fixed facility operating costs, and the average variable distribution and processing costs.

References

Aghezzaf E (2005) Capacity planning and warehouse location in supply chains with uncertain demands. J Oper Res Soc 56:453–462

Ahmed S, Garcia R (2004) Dynamic capacity acquisition and assignment under uncertainty. Ann Oper Res 124:267–283

Ahmed S, King AJ, Parija G (2003) A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. J Global Optim 26:3–24

Albareda-Sambola M, Alonso-Ayuso A, Escudero LF, Fernández E, Pizarro C (2013) Fix-and-relax coordination for a multi-period location-allocation problem under uncertainty. Comput Oper Res 40:2878–2892

Alonso-Ayuso A, Escudero LF, Garín A, Ortuño MT, Pérez G (2003) An approach for strategic supply chain planning under uncertainty based on stochastic 0–1 programming. J Glob Optim 26:97–124

Alumur SA, Nickel S, Saldanha-da-Gama F (2012) Hub location under uncertainty. Transp Res Part B: Methodol 46:529–543

Alvarez EM, van der Heijden MC, Zijm WHM (2015) Service differentiation in spare parts supply through dedicated stocks. Ann Oper Res 231:283–303

Antunes A, Peeters D (2001) On solving complex multi-period location models using simulated annealing. Eur J Oper Res 130:190–201

Bakker H, Dunke F, Nickel S (2020) A structuring review on multi-stage optimization under uncertainty: Aligning concepts from theory and practice. Omega 96:102080

Bashiri M, Badri H, Talebi J (2012) A new approach to tactical and strategic planning in production-distribution networks. Appl Math Model 36:1703–1717

Becker T, Lier S, Werners B (2019) Value of modular production concepts in future chemical industry production networks. Eur J Oper Res 276:957–970

Birge JR, Louveaux FV (2011) Introduction to Stochastic Programming, 2nd edn. Springer Series in Operations Research and Financial Engineering. Springer, New York, Dordrecht, Heidelberg, London

Contreras I, Cordeau J-F, Laporte G (2011) Stochastic uncapacitated hub location. Eur J Oper Res 212:518–528

Correia I, Melo T (2017) A multi-period facility location problem with modular capacity adjustments and flexible demand fulfillment. Comput Ind Eng 255:729–746

Correia I, Saldanha-da-Gama F (2019) Facility location under uncertainty. In: Laporte G, Nickel S, Saldanha-da-Gama F (eds), Location science, chapter 8, 2nd edn. Springer, Heidelberg, pp. 185–213

Correia I, Melo T, Saldanha-da-Gama F (2013) Comparing classical performance measures for a multi-period, two-echelon supply chain network design problem with sizing decisions. Comput Ind Eng 64:366–380

Correia I, Nickel S, Saldanha-da-Gama F (2018) A stochastic multi-period capacitated multiple allocation hub location problem: formulation and inequalities. Omega 74:122–134

Cortinhal MJ, Lopes MJ, Melo MT (2015) Dynamic design and re-design of multi-echelon, multi-product logistics networks with outsourcing opportunities: a computational study. Comput Ind Eng 90:118–131

Delmelle EM, Thill J-C, Peeters D, Thomas I (2014) A multi-period capacitated school location problem with modular equipment and closest assignment considerations. J Geogr Syst 16:263–286

Dias J, Captivo ME, Clímaco J (2007) Dynamic location problems with discrete expansion and reduction sizes of available capacities. Investig. Oper. 27:107–130

Diglio A, Nickel S, Saldanha-da-Gama F (2020) Towards a stochastic programming modeling framework for districting. Ann Oper Res 292:249–285

Döyen A, Aras N, Barbarosoğlu G (2012) A two-echelon stochastic facility location model for humanitarian relief logistics. Oper Res Lett 6:1123–1145

Duran S, Liu T, Simchi-Levi D, Swann JL (2008) Policies utilizing tactical inventory for service-differentiated customers. Oper Res Lett 36:259–264

Faugère L, Klibi W, White III C, Montreuil B (2020) Dynamic pooled capacity deployment for urban parcel logistics. arXiv:2007.11270

Fernández E, Hinojosa Y, Puerto J, Saldanha-da-Gama F (2019) New algorithmic framework for conditional value at risk: application to stochastic fixed-charge transportation. Eur J Oper Res 277:215–226

Ghafarimoghadam A, Ghayebloo S, Pishvaee MS (2019) A fuzzy-budgeted robust optimization model for joint network design-pricing problem in a forward-reverse supply chain: the viewpoint of third-party logistics. Comp Appl Math 38:194

Gourdin E, Klopfenstein O (2008) Multi-period capacitated location with modular equipments. Comput Oper Res 35:661–682

Govindan K, Fattahi M, Keyvanshokooh E (2017) Supply chain network design under uncertainty: a comprehensive review and future research directions. Eur J Oper Res 263:108–141

Grass E, Fischer K, Rams A (2020) An accelerated L-shaped method for solving two-stage stochastic programs in disaster management. Ann Oper Res 284:557–582

Heckmann I (2016) Towards supply chain risk analytics, chapter 12, pages 313–361. Springer Gabler, Wiesbaden

Hernández P, Alonso-Ayuso A, Bravo F, Escudero LF, Guignard M, Marianov V, Weintraub A (2012) A branch-and-cluster coordination scheme for selecting prison facility sites under uncertainty. Comput Oper Res 39:2232–2241

Hinojosa Y, Puerto J, Saldanha-da-Gama F (2014) A two-stage stochastic transportation problem with fixed handling costs and a priori selection of the distribution channels. TOP 22:1123–1147

Hung H-C, Chew EP, Lee LH, Liu S (2012) Dynamic inventory rationing for systems with multiple demand classes and general demand processes. Int J Prod Econ 139:351–358

Jena SD, Cordeau J-F, Gendron B (2016) Solving a dynamic facility location problem with partial closing and reopening. Comput Oper Res 67:143–154

Khodaparasti S, Bruni ME, Beraldi P, Maleki HR, Jahedi S (2018) A multi-period location-allocation model for nursing home network planning under uncertainty. Oper Res Health Care 18:4–15

Kim D, Lee K, Moon I (2019) Stochastic facility location model for drones considering uncertain flight distance. Ann Oper Res 283:1283–1302

Kınay ÖB, Kara BY, Saldanha-da-Gama F, Correia I (2018) Modeling the shelter site location problem using chance constraints: a case study for Istanbul. Eur J Oper Res 270:132–145

Laporte G, Louveaux FV, van Hamme L (1994) Exact solution to a location problem with stochastic demands. Transp Sci 28:95–103

Li Z, Lu Q, Talebian M (2015) Online versus bricks-and-mortar retailing: a comparison of price, assortment and delivery time. Int J Prod Res 53:3823–3835

Louveaux FV (1986) Discrete stochastic location models. Ann Oper Res 6:23–34

Malladi SS, Erera AL, White CC III (2020) A dynamic mobile production capacity and inventory control problem. IISE Trans 52:926–943

Marín A, Martínez-Merino LI, Rodríguez-Chia AM, Saldanha-da-Gama F (2018) Multi-period stochastic covering location problems: modeling framework and solution approach. Eur J Oper Res 268:432–449

Marković N, Ryzho IO, Schonfeld P (2017) Evasive flow capture: a multi-period stochastic facility location problem with independent demand. Eur J Oper Res 257:687–703

Melo MT, Nickel S, Saldanha-da-Gama F (2006) Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Comput Oper Res 33:181–208

Mete HO, Zabinsky ZB (2010) Stochastic optimization of medical supply location and distribution in disaster management. Int J Prod Econ 126:76–84

Mohamed IB, Klibi W, Vanderbeck F (2020) Designing a two-echelon distribution network under demand uncertainty. Eur J Oper Res 280:102–123

Moreno A, Allem D, Ferreira D, Clark A (2018) An effective two-stage stochastic multi-trip location-transportation model with social concerns in relief supply chains. Eur J Oper Res 269:1050–1071

Nickel S, Saldanha-da-Gama F (2019) Multi-period facility location. In: Laporte G, Nickel S, Saldanha-da-Gama F (eds) Location science, chap 11, 2nd edn. Springer, Heidelberg, pp 303–326

Owen SH, Daskin MS (1998) Strategic facility location: a review. Eur J Oper Res 111:423–447

Pazour JA, Unnu K (2018) On the unique features and benefits of on-demand distribution models. In: Ballot E, Carrano AL, Ellis K, Ferrell W, Porter JD, Reed D (eds) 15th IMHRC proceedings - progress in material handling research: 2018, Savannah, Georgia (USA). Available at: https://digitalcommons.georgiasouthern.edu/pmhr_2018/13/. Accessed 15 Jan 2021

Rahmati R, Neghabi H (2021) Adjustable robust balanced hub location problem with uncertain transportation cost. Comput Appl Math 40:14

Snyder L (2006) Facility location under uncertainty: a review. IIE Trans 38:537–554

Unnu K (2020) Optimization models and frameworks for on-demand warehousing systems. PhD thesis, Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York

Verter V, Dincer MC (1995) Facility location and capacity acquisition: an integrated approach. Nav Res Logist 42:1141–1160

Wang Y, Cohen MA, Zheng Y-S (2002) Differentiating customer service on the basis of delivery lead-times. IIE Trans 34:979–489

Wilhelm W, Han X, Lee C (2013) Computational comparison of two formulations for dynamic supply chain reconfiguration with capacity expansion and contraction. Comput Oper Res 40:2340–2356

Wolsey LA (1998) Integer programming. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York

Zhuge D, Yu S, Zhen L, Wang W (2016) Multi-period distribution center location and scale decision in supply chain network. Comput Ind Eng 101:216–226

Zou J, Ahmed S, Sun XA (2019) Stochastic dual dynamic integer programming. Math Program 175:461–502

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.