European Journal of Operational Research 000 (2016) 1–18

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Multi-period capacitated facility location under delayed demand satisfaction

Isabel Correia[a,b], Teresa Melo[c,*]

[a] Centro de Matemática e Aplicações (CMA), Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), P 2829-516, Caparica, Portugal
[b] Departamento de Matemática, Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa (UNL), P 2829-516, Caparica, Portugal
[c] Business School, Saarland University of Applied Sciences, D 66123, Saarbrücken, Germany

A R T I C L E I N F O

Article history:
Received 4 August 2015
Accepted 16 June 2016
Available online xxx

Keywords:
Location
Multi-period
Capacity choice
Delivery lateness
MILP models

A B S T R A C T

We address an extension of the classical multi-period facility location problem in which customers are sensitive to delivery lead times. Accordingly, two customer segments are considered. The first segment comprises customers that require timely demand satisfaction, whereas customers accepting delayed deliveries make up the second segment. Each customer belonging to the latter segment specifies a maximum delivery time. A tardiness penalty is incurred to each unit of demand that is not satisfied on time. In the problem that we study, a network is already in place with a number of facilities being operated at fixed locations. The network can be expanded by establishing new facilities at a finite set of potential sites and selecting their capacity levels from a set of available discrete sizes. In addition, existing facilities may be closed over the time horizon. Two mixed-integer linear programming formulations are proposed to re-design the network at minimum cost and a theoretical comparison of their linear relaxations is provided. We also extend the mathematical models to the case in which each customer accepting delayed demand satisfaction requires late shipments to occur at most once over the delivery lead time. To gain insight into how challenging these problems are to solve, a computational study is performed with randomly generated instances and using a general-purpose solver. Useful insights are derived from analyzing the impact of different delivery lead time restrictions on the network structure and cost.

1. Introduction

In this paper, we consider a company that operates a set of facilities at fixed locations to fulfil customer demand requirements for a given product. Customers are classified on the basis of their sensitivity to delivery lead times. In particular, two customer segments are considered. The first segment comprises customers requiring timely demand satisfaction, whereas customers accepting delayed deliveries make up the second segment. Each customer belonging to the latter segment specifies a desired maximum delivery lead time. Changing market and business conditions, frequently in conjunction with increased cost pressure and service requirements, compel the company to restructure its network of facilities. To this end, the network configuration can gradually change over a multi-period horizon through opening new facilities and closing initially existing facilities. Locations for new facilities are chosen from a finite set of candidate sites. Moreover, at each potential site, different capacity levels are also available for selection. The objective is to determine the optimal network configuration over the planning horizon so as to minimize the sum of fixed and variable costs. The former are associated with strategic decisions and comprise fixed costs for facility siting and operation, for capacity acquisition and for closing existing facilities. Variable costs for processing and distributing the product to the customers along with tardiness penalty costs for delayed demand satisfaction are also incurred. These costs are associated with tactical decisions.

The problem that we study is motivated by the online retail and manufacturing industries (e.g. computer and mobile telephone manufacturers) who often adopt different pricing policies for their products, thereby offering price incentives to customers in exchange for longer order lead times (Agatz, Fleischmann, & van Nunen, 2008). Multiple customer segments that are differentiated on the basis of delivery lead times and price options can also be encountered in other business environments (see e.g. Hung, Chew, Lee, & Liu, 2012). For example, Wang, Cohen, and Zheng (2002) report the case of a semiconductor equipment manufacturer that provides a two-class service policy. Customers with emergency demand for repairable service parts form the first class. These customers pay a premium price to have their returned defective parts repaired immediately. Non-emergency service is provided to the

* Corresponding author.
E-mail address: teresa.melo@htwsaar.de (T. Melo).

http://dx.doi.org/10.1016/j.ejor.2016.06.039

second class of customers who pay a lower price for a longer repair time. Cheong, Bhatnagar, and Graves (2005) describe a similar case in the textile dye industry where textile mills with small capacities favor shorter delivery lead times from their suppliers as opposed to large textile mills that can keep higher stock levels and, therefore, can cope with longer order fulfilment lead times. Wu and Wu (2015) describe this type of strategy as "demand postponement" because the company decides upon the actual delivery time for orders committed by customers who are less sensitive to lead times.

On the one hand, lower unit prices may result in decreased total revenue for the company but, on the other hand, the company will have to invest in increasing capacity to guarantee a shorter delivery time to those customers who are willing to pay a premium price. Hence, one of the main research objectives of our work is to investigate the interplay between location and capacity investment (or disinvestment) decisions and delivery time decisions in a facility system operated by a company. The coordination of location, capacity acquisition, distribution and demand fulfilment decisions has the potential to improve system efficiencies. In particular, retail and manufacturing environments can benefit from the insights provided by our research work.

As will be shown in Section 2, our study is the first to embed customer segments having distinct sensitivity to delivery lead times in a multi-period facility location setting. This new feature is combined with different time scales for strategic and tactical decisions. This means that tactical decisions regarding the commodity flow from operating facilities to customers may take place in any time period, whereas strategic location decisions can only be made over a subset of the time periods of the planning horizon. Furthermore, the decision space is extended with strategic facility sizing decisions, an aspect that is not often encountered in the literature. In fact, most facility location models consider capacity as an exogenous factor. However, from an application point of view, capacity is often purchased in the form of equipment which is only available at a few discrete sizes. Capacity choices incur specific fixed installation costs that are subject to economies of scale. Hence, a further research objective of this study is to add three new dimensions (customer segments with distinct service requirements, different decision time scales and multiple capacity choices) to the classical multi-period facility location problem which is known to be an NP-hard problem. From a computational standpoint, this results in a challenging problem for which the possibility of solving large-scale instances to optimality within acceptable time is rather limited. In such cases, one often resorts to heuristic methods to obtain feasible solutions. However, to be able to measure the quality of such solutions it is of paramount importance to have (good) lower bounds for the problem. Therefore, the third research objective of our work is to develop different mathematical formulations and to compare them in terms of the LP-relaxation bound they provide. Additional inequalities are also proposed in an attempt to strengthen these bounds. Finally, we conduct an extensive computational study to obtain managerial insights that illustrate the far-reaching implications of delivery lead time restrictions on the network structure and its cost. Without the support of the models developed in this paper it would otherwise be difficult to obtain most of these insights. Given the typically high investment costs and the limited reversibility of strategic decisions, it is essential for stakeholders to perceive the impact of such decisions on overall system performance.

The remainder of this article is organized as follows. Section 2 summarizes the relevant literature. In Section 3, we develop two mixed-integer linear programming (MILP) formulations for the problem under study and present a theoretical comparison of their linear relaxations. In particular, we also consider the special case in which customers accepting late deliveries wish to receive single shipments even if they arrive with some delay. In other words, partial, late deliveries are not allowed for such customers. In Section 4, additional inequalities are proposed to enhance the original formulations. Section 5 reports and discusses the results of an extensive computational study using general-purpose optimization software. Finally, in Section 6, conclusions are provided and directions for future research are identified.

2. Literature review

Discrete facility location models are typically concerned with determining the number, location and capacities of facilities that should be established to serve the demands of a set of spatially distributed customers with least total cost. This field of location analysis has been an active and rich research area over the past decades. A wide variety of applications have emerged in many contexts such as strategic logistics planning (see e.g. Alumur, Kara, & Melo, 2015, chap. 16 and Melo, Nickel, & Saldanha da Gama, 2009) and telecommunications (see e.g. Fortz, 2015, chap. 20), just to name a few.

Most discrete location models ensure the satisfaction of customer demands by imposing distance and/or time limits as service level requirements. In contrast, location problems with flexibility regarding demand fulfilment have received much less attention. The case of unfilled demand can be treated either with the lost sales assumption or with the backorder assumption. The former situation applies to contexts in which satisfying all customer demands may not be economically attractive due to high investment costs on establishing new facilities with appropriate capacities. In a static setting, Alumur et al. (2015) describe a generic model for a facility location problem arising in logistics network design that includes this feature, while Correia, Melo, and Saldanha-da-Gama (2013) address this issue in the design of a two-echelon production–distribution network over multiple time periods. The models developed by Badri, Bashiri, and Hejazi (2013), Bashiri, Badri, and Talebi (2012), Canel and Khumawala (1996) and Sousa, Shah, and Papageorgiou (2008) also allow lost sales over a dynamic horizon. In the previous studies (Badri et al., 2013; Bashiri et al., 2012; Canel and Khumawala, 1996; Correia et al., 2013; Sousa et al., 2008), strategic location and tactical logistics decisions are made under a profit maximization objective. In addition, Correia et al. (2013) also investigate their problem from a cost minimization perspective with additional constraints enforcing a minimum rate for demand fulfilment. For a number of test instances with small and moderate sizes, the MILP formulations proposed in Alumur et al. (2015); Bashiri et al. (2012); Canel and Khumawala (1996); Correia et al. (2013); Sousa et al. (2008) could be solved to optimality by a commercial MILP solver within acceptable time. Badri et al. (2013) developed a Lagrangian-based heuristic through dualizing a set of constraints that limit the expenditures for opening new facilities and expanding the capacity at existing locations.

The lost sales assumption is also present in the problem addressed by Altiparmak, Gen, Lin, and Paksoy (2006) through the maximization of the overall fraction of demand that is delivered to customers. This objective is integrated with the minimization of the total cost of designing and operating a multi-stage network and the maximization of the capacity utilization of facilities. These three objectives are combined into a single-objective function by building a weighted sum and feasible solutions are determined with a genetic algorithm. The latter solution methodology was also adopted by Lieckens and Vandaele (2007) for a facility location problem arising in reverse logistics with stochastic lead times for processing and moving used products. In this case, a fraction of the returned products may not be collected and demand for reused products may be only partially met. Cheong et al. (2005) follow a different approach to deal with lost sales in an uncapacitated two-echelon distribution network. To this end, each customer is

Table 1
Classification of facility location models.

	Planning horizon (S/D)[a]	Decoupled time scales	Objective function (C/P/M)[b]	Demand fulfilment flexibility: Lost sales	Demand fulfilment flexibility: Backorders	Delivery lead times	Facility location (O/C)[c]	Facility sizing: Capacity type (U/F/M)[d]	Facility sizing: Capacity changes (E/R)[e]	Modeling approach (D/S)[f]
Altiparmak et al. (2006)	S		M	✓			O	F		D
Alumur et al. (2015)	S		C	✓			O/C	F	E	D
Badri et al. (2013)	D	✓	P	✓			O	M	E	D
Bashiri et al. (2012)	D	✓	P	✓			O	M	E	D
Canel and Khumawala (1996)	D		P	✓			O	F		D
Cheong et al. (2005)	S		C	✓		✓	O	U		D
Correia et al. (2013)	D		C/P	✓			O	M	E	D
Gebennini et al. (2009)	D		C		✓	✓	O	U		D
Lieckens and Vandaele (2007)	S		P	✓			O	M		S
Meisel et al. (2015)	S		C		✓	✓	O	F		D
Shen (2006)	S		P	✓			O	F		D/S
Sousa et al. (2008)	D		P	✓			O	F		D
Wilhelm et al. (2005)	D	✓	P		✓		O	F		D
Wilhelm et al. (2013)	D		C		✓		O	M	E/R	D
New models (cf. Section 3)	**D**	✓	**C**		✓	✓	**O/C**	**M**		**D**

[a] S: Static (single period), D: Dynamic (multi-period).
[b] C: Cost minimization, P: Profit or net income maximization, M: Multiple objectives.
[c] O: Open new facilities, C: Close initially existing facilities.
[d] U: Uncapacitated, F: Fixed capacity limit, M: Modular capacities.
[e] E: Capacity expansion, R: Capacity contraction (including closing new facilities).
[f] D: Deterministic model, S: Stochastic model.

offered a short and a long delivery lead time and his demand requirements for a single product decline as the lead time increases. Fixed costs for locating one central warehouse and several local warehouses as well as variable costs for inventory holding and product distribution are considered along with lost sales costs. The proposed MILP formulation is solved for randomly generated instances involving five customers. Contrary to the aforementioned works, which allow partial satisfaction of the demand of a particular customer, in the location-inventory problem addressed by Shen (2006), the demand of each customer for a single product is either completely satisfied or not. This type of decision is impacted by the customer's reserve price (i.e. the highest price the customer is willing to pay) and the total price charged by the company. The latter comprises the selling price and the delivery cost, and both factors depend on the location, distribution and inventory decisions made. If the total price is higher than the customer's reserve price then the company will lose the customer. The problem is solved with a branch-and-price algorithm.

Compared to the lost sales case, the backorder option has received much less attention in the literature dedicated to facility location. In the problem studied by Gebennini, Gamberini, and Manzini (2009), all customers tolerate a delay of at most one period for demand fulfilment. Moreover, a constant backorder cost representing a late-delivery penalty is considered in each period of the planning horizon for each customer. While this case may occur, it has a limited domain of applicability since different customers have different sensitivity to delays and the impact of the latter in terms of costs may also not be the same for every customer.

Wilhelm et al. (2005) address the design of a multi-echelon production–distribution network over a multi-period horizon in an international context. Backorder costs are incurred to demand requirements that cannot be satisfied on time. Contrary to Gebennini et al. (2009), no time limit is imposed on the delay for satisfying the demand of a particular customer. Although lateness can be restrained through setting high backorder costs, the delay experienced by a customer in this case may still be well beyond the maximum due date the customer is willing to tolerate. Recently, Wilhelm, Han, and Lee (2013) have adopted the same setting for customer demand satisfaction in a multi-echelon network involving suppliers, plants, distribution centers (DCs) and customer zones. Over a planning horizon, new DCs can be established at candidate sites and their initial capacities have to be selected from a set of capacity alternatives. Subsequently, capacity adjustments at a DC location are possible through capacity expansion, capacity contraction or even closing the facility. An upper bound is specified on the total number of expansions and contractions that a DC may undergo over the time horizon. Moreover, the reconfiguration of each DC is also constrained by the available budget. This setting is especially appropriate to deal with demand fluctuations and when frequent changes in facility sizes are possible due to leasing or renting capacity.

Compared to the previous studies, Meisel, Rei, Gendreau, and Bierwirth (2015) adopt a broader view of customer sensitivity to delivery lead times. In their three-echelon production–distribution system, a maximum lead time is promised to each customer. Production and transportation times determine the actual lead time for satisfying the demand requirements of a customer. However, backorder costs for late shipments are not incurred, meaning that customers are insensitive to lateness in demand fulfilment, an assumption that has limited practical application.

In order to relate the surveyed literature on facility location to the problem studied in this paper, a classification of the aforementioned works is given in Table 1. This table is not intended to provide an exhaustive list of the features addressed in this section but rather to illustrate the extent to which our mathematical models compare to the models that have appeared in the literature. In particular, Table 1 gives emphasis to the three new dimensions that are addressed by our multi-period location models: different time scales for strategic and tactical decisions (column 3), demand fulfilment mode (columns 5–6) and customer segments having distinct delivery lead times (column 7), and multiple capacity choices (column 9) in conjunction with facility location decisions. The latter are classified according to the nature of the location project, namely if a "greenfield" approach is used (i.e. a new network is to be designed by establishing new facilities) or if a network is already in place which can be restructured through opening new facilities and/or closing initially existing facilities (column 8). The last row of Table 1 highlights the main features of the models to be discussed in Section 3. It can be seen that various features are integrated to capture the trade-off between the potential for increased timely demand fulfilment against the costs of re-designing and operating an existing network of facilities.

The economic benefits of demand fulfilment flexibility have been widely studied in the inventory management literature (see e.g. Hung et al., 2012 and references therein) and, to a lesser extent, in the vehicle routing literature (see e.g. Albareda-Sambola, Fernandez, & Laporte, 2014; Archetti, Jabali, & Speranza, 2015; Athanasopoulos & Minis, 2011, chap. 11; and Wen, Cordeau, Laporte, & Larsen, 2010). In contrast, there is still a noticeable lack of published research in discrete facility location, including the joint problem of locating facilities and designing vehicle routes (see the recent reviews by Drexl & Schneider, 2015 and Prodhon & Prins, 2015). This paper aims at giving a first contribution toward filling an important gap in this area.

3. Mathematical formulations

In this section, we propose and discuss two MILP models for the multi-period facility location problem with delayed demand satisfaction and multiple capacity levels (MFLPDDSM). Furthermore, we address a particular case of this problem in which the demand of a customer must be delivered as a single shipment when the customer tolerates late deliveries. Two MILP formulations will also be developed for this case.

All models rely on the following assumptions:

- A company operates a set of facilities at fixed locations with given capacities. A single product (or product family) is processed at these facilities to be delivered to customers (or customer zones).
- Due to changes in the distribution of customer demand, the existing facility system may no longer provide adequate service. When this situation occurs, the company may wish to close some of its existing facilities and open some new facilities to better serve its customers.
- Prior to the network re-design project, the company has selected a set of candidate sites where new facilities can be established. At each candidate site, a set of discrete capacity levels has been identified.
- A planning horizon with a finite number of discrete time periods is considered. Strategic decisions related to opening new facilities, closing initially existing facilities and installing capacity levels at new sites can be made at selected time periods, called *strategic periods*. The latter represent a subset of the planning horizon. In contrast, product distribution tactical decisions can be made at every time period. These periods are also called *tactical periods*.
- If an initially existing facility is closed, it cannot be reopened. If a new facility is established at a candidate site, a capacity level has to be selected and installed. In this case, further capacity changes are not allowed over the planning horizon and the facility must remain in activity until the end of the time horizon.
- The company differentiates its customers on the basis of their sensitivity to delivery lead times. Customers who receive preferred service have a zero demand lead time, i.e. their demands are satisfied on time. These customers typically contribute most to the company's profit. In contrast, customers who are not averse to waiting for their demand requirements to be filled specify a maximum allowed delay. These customers are compensated with a substantially lower price which is translated into a tardiness penalty cost for delayed deliveries to reflect the negative impact on the company's profit margin.
- It is assumed that all relevant input data (costs, capacities and other parameters) were collected using e.g. appropriate forecasting methods and company-specific business analyzes.

Furthermore, we consider that all sites are company-owned, space and equipment cannot be leased and operations are not subcontracted. These assumptions are meaningful in the context of a product that requires expensive equipment and highly skilled workforce. The semiconductor equipment manufacturer reported by Wang et al. (2002) is such an example. In this case, location and capacity acquisition decisions involve sizeable investments (e.g. to build facilities, to purchase equipment, to qualify workforce, etc.). Therefore, new facilities are expected to remain operable with the initially installed capacity for an extended time period. Partial closing or even reopening of facilities are not viable options in this context.

We now introduce the notation that will be used hereafter.

Facilities, customers and planning horizon:

I^e	Set of existing facilities at the beginning of the planning horizon
I^n	Set of candidate sites for locating new facilities
I	Set of all facility locations, $I = I^e \cup I^n$
K_i	Set of discrete capacity levels that can be installed at candidate site i ($i \in I^n$)
K_i	Capacity type of initially existing facility i, $K_i = \{1\}$ ($i \in I^e$)
J^0	Set of customers whose demands must be satisfied on time
J^1	Set of customers that may experience delayed demand satisfaction
J	Set of all customers, $J = J^0 \cup J^1$, $J^0 \cap J^1 = \emptyset$
T	Set of discrete time periods
T_L	Set of *strategic time periods* in which location and capacity acquisition decisions can be made, $T_L \subset T$

We denote by $\ell = 1$, resp. ℓ_{max}, the first, resp. last, strategic time period in which decisions related to opening/closing facilities and installing capacity levels at candidate sites can be taken. Tactical decisions involve the quantities to be shipped between facilities and customers at each time period. Hence, different time scales are considered in a similar way as followed by Albareda-Sambola, Fernandez, and Nickel (2012) for a facility location-routing problem and by Badri et al. (2013) and Bashiri et al. (2012) in the context of logistics network design.

Capacity and demand parameters:

Q_{ik}	Capacity of level k that can be installed at candidate site i ($i \in I^n; k \in K_i$)
Q_{i1}	Capacity of existing facility i ($i \in I^e$) at the beginning of the time horizon
d_j^t	Demand of customer j in time period t ($j \in J; t \in T$)
ρ_j	Maximum allowed delay (in number of time periods) to satisfy the demand of customer j ($j \in J$)

Given the above definition of the tolerated delay in demand fulfilment, the two categories of customers previously introduced correspond to $J^0 = \{j \in J : \rho_j = 0\}$ and $J^1 = \{j \in J : \rho_j > 0\}$. In particular, the demand of customer $j \in J^1$ in time period $t \in T$ must be filled within ρ_j time periods, i.e. over periods $t, \ldots, \min\{t + \rho_j, |T|\}$. Hence, demand satisfaction cannot be carried over to future periods beyond the planning horizon.

If a new facility is established at candidate location $i \in I^n$ then a capacity level has to be selected from the set of available discrete sizes K_i. We assume that the latter are sorted in non-decreasing order, that is, $Q_{i1} < Q_{i2} < \ldots < Q_{i|K_i|}$.

Fixed and variable cost rates:

FO^{ℓ}_{ik} Fixed cost of opening a new facility at candidate site i with capacity level k at the beginning of time period ℓ ($i \in I^n; k \in K_i; \ell \in T_L$)

FC^{ℓ}_{i1} Fixed cost of closing the initially existing facility i at the end of time period ℓ ($i \in I^e; \ell \in T_L$)

M^t_{ik} Fixed maintenance cost incurred by operating facility i with capacity level k in time period t ($i \in I; k \in K_i; t \in T$)

c^t_{ij} Cost of distributing one unit of product from facility i to customer j in time period t ($i \in I; j \in J; t \in T$)

o^t_{ik} Cost of processing one unit of product at facility i with capacity level k in time period t ($i \in I; k \in K_i; t \in T$)

$p^{tt'}_j$ Tardiness penalty cost for satisfying one unit of demand of customer j in period t' that was originally demanded in period t ($j \in J^1;\ t \in T;\ t' = t, t+1, \ldots, \min\{t+\rho_j,\ |T|\}$); in particular, for $t' = t$, the tardiness penalty cost is equal to zero

Economies of scale favoring large capacity levels are present. They are reflected in the fixed costs of opening (FO^{ℓ}_{ik}) and maintaining (M^t_{ik}) a facility. The fixed cost of closing an initially existing facility (FC^{ℓ}_{i1}) also takes its size into account. In addition, the variable cost of processing the product at a facility (o^t_{ik}) is subject to economies of scale. By combining the fixed facility and maintenance costs over an appropriate number of time periods, the following aggregated cost parameters are obtained:

$$F^{\ell}_{ik} = \begin{cases} FO^{\ell}_{ik} + \sum\limits_{t=\ell}^{|T|} M^t_{ik} & \text{for } i \in I^n;\ k \in K_i;\ \ell \in T_L \\ FC^{\ell}_{i1} + \sum\limits_{t=1}^{\ell} M^t_{i1} & \text{for } i \in I^e;\ \ell \in T_L \end{cases} \tag{1}$$

Observe that for a new facility $i \in I^n$, F^{ℓ}_{ik} represents the total cost of establishing the facility at the beginning of the strategic period $\ell \in T_L$ and operating it until the end of the time horizon. In a similar way, for an initially existing facility $i \in I^e$, F^{ℓ}_{ik} gives the total cost of operating the facility until the end of time period $\ell \in T_L$, the time point at which the facility is removed, and the fixed closing cost. We note that the earliest moment in time for closing an initially existing facility is at the end of the first time period.

3.1. Mixed-integer linear programming model

A natural formulation of the MFLPDDSM relies on binary variables to represent strategic facility location and capacity acquisition decisions as follows:

$$z^{\ell}_{ik} = 1 \text{ if a new facility is established at candidate location } i \text{ with capacity level } k \text{ at the beginning of time period } \ell, 0 \text{ otherwise} (i \in I^n;\ k \in K_i;\ \ell \in T_L) \tag{2}$$

$$z^{\ell}_{i1} = 1 \text{ if the initially existing facility } i \text{ is closed at the end of time period } \ell, 0 \text{ otherwise} (i \in I^e;\ \ell \in T_L). \tag{3}$$

Observe that if a new facility is opened in period ℓ then it will operate in periods $\ell, \ldots, |T|$. Analogously, if an initially existing facility is removed at the end of period ℓ then it operates in periods $1, \ldots, \ell$. In addition, the formulation also includes two sets of continuous variables that prescribe tactical decisions:

$$x^t_{ijk} : \text{Amount of product distributed from facility } i \text{ with capacity level } k \text{ to customer } j \text{ in time period } t\ (i \in I;\ k \in K_i;\ j \in J^0;\ t \in T) \tag{4}$$

$$y^{tt'}_{ijk} : \text{Amount of product distributed from facility } i \text{ with capacity level } k \text{ to customer } j \text{ in time period } t' \text{ to (partially) satisfy demand of period } t\ (i \in I;\ k \in K_i;\ j \in J^1;\ t \in T;\ t' = t, t+1, \ldots, \min\{t+\rho_j,\ |T|\}). \tag{5}$$

We denote by (**P**) the following MILP formulation for the MFLPDDSM:

$$\text{Min} \quad \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k \in K_i} F^{\ell}_{ik} z^{\ell}_{ik} + \sum_{t \in T} \sum_{i \in I^e} M^t_{i1} \left(1 - \sum_{\ell \in T_L} z^{\ell}_{i1}\right) + \sum_{t \in T} \sum_{i \in I} \sum_{k \in K_i} \sum_{j \in J^0} \left(c^t_{ij} + o^t_{ik}\right) x^t_{ijk} + \sum_{t \in T} \sum_{i \in I} \sum_{k \in K_i} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j,\ |T|\}} \left(p^{tt'}_j + c^{t'}_{ij} + o^{t'}_{ik}\right) y^{tt'}_{ijk} \tag{6}$$

s.t.

$$\sum_{i \in I} \sum_{k \in K_i} x^t_{ijk} = d^t_j \qquad j \in J^0,\ t \in T \tag{7}$$

$$\sum_{i \in I} \sum_{k \in K_i} \sum_{t'=t}^{\min\{t+\rho_j,\ |T|\}} y^{tt'}_{ijk} = d^t_j \qquad j \in J^1,\ t \in T \tag{8}$$

$$\sum_{\ell \in T_L} \sum_{k \in K_i} z^{\ell}_{ik} \le 1 \qquad i \in I \tag{9}$$

$$\sum_{j \in J^0} x^t_{ijk} + \sum_{j \in J^1} \sum_{t'=\max\{1,\ t-\rho_j\}}^{t} y^{t't}_{ijk} \le Q_{ik} \sum_{\ell \in T_L:\ \ell \le t} z^{\ell}_{ik} \qquad i \in I^n,\ k \in K_i,\ t \in T \tag{10}$$

$$\sum_{j \in J^0} x^t_{ij1} + \sum_{j \in J^1} \sum_{t'=\max\{1,\ t-\rho_j\}}^{t} y^{t't}_{ij1} \le Q_{i1} \left(1 - \sum_{\ell \in T_L:\ \ell < t} z^{\ell}_{i1}\right) \qquad i \in I^e,\ t \in T \tag{11}$$

$$x^t_{ijk} \ge 0 \qquad i \in I,\ j \in J^0, k \in K_i,\ t \in T \tag{12}$$

$$y^{tt'}_{ijk} \ge 0 \qquad i \in I,\ j \in J^1, k \in K_i,\ t \in T,\ t' = t, \ldots, \min\{t+\rho_j, |T|\} \tag{13}$$

$$z^{\ell}_{ik} \in \{0, 1\} \qquad i \in I,\ k \in K_i,\ \ell \in T_L. \tag{14}$$

The objective function (6) minimizes the sum of the fixed and variable costs. The former include the costs incurred for establishing new facilities and installing capacity levels at the new sites, removing initially existing facilities, and maintaining facilities in those periods in which they are operated (recall (1)). Variable costs account for processing and shipping the product to customers along with the tardiness costs resulting from delayed deliveries. Constraints (7), resp. (8), guarantee the satisfaction of the demand over the time horizon for customer segment J^0, resp. J^1. For each candidate site $i \in I^n$, constraints (9) impose that at most one new facility can be established with a given capacity level over the time horizon. Constraints (9) also allow each initially existing facility $i \in I^e$ to be closed at most once throughout the planning horizon. Inequalities (10), resp. (11), are capacity constraints for new, resp. existing, facilities. Observe that since an existing facility can only be closed at the end of a given time period, say ℓ, its capacity is not available in any subsequent period. This is described in (11) by considering all strategic periods $\ell \in T_L$ such that $\ell < t$ for every $t \in T$. In contrast, if a new facility is established in time period t then its capacity also becomes available in the same period. Therefore, in constraints (10) we consider all periods $\ell \in T_L$ such that $\ell \le t$ for every $t \in T$. Finally, constraints (12)–(14) state non-negativity and binary conditions.

The formulation that we propose covers multiple situations. In particular, it generalizes the classical multi-period uncapacitated facility location problem (MUFLP). The latter corresponds to setting $T = T_L$, $J = J^0$, $J^1 = \emptyset$, $I^e = \emptyset$, $|K_i| = 1$, and $Q_{i1} = \infty$ $(i \in I)$. Since the MUFLP is an NP-hard problem (see e.g. Jacobsen, 1990, chap. 4), the MFLPDDSM is also NP-hard. If $J = J^0$ then model (**P**) reduces to a classical case in multi-period capacitated facility location in which all customers must have their demands satisfied on time. If $J = J^1$, the opposite case is captured, namely all customers accept a delay in product delivery. In the event that $J^0 \neq \emptyset$ and $J^1 \neq \emptyset$, an intermediate situation is modeled by (**P**). In particular, this variant ensures that all important customers for the company (i.e. the members of set J^0) receive preferred service. Another distinctive feature of our model, that results from considering different time scales for strategic and tactical decisions, is the extended length of the time horizon compared to classical multi-period location problems where typically only instances with a reduced number of time periods can be solved exactly within acceptable computing times. As it will be shown in Section 3.4, this characteristic has a significant impact on the overall size of the model. This has prompted us to develop an alternative formulation in an attempt to reduce the size and difficulty of the resulting problem.

3.2. Alternative formulation

The mathematical formulation (**P**) uses four-index and five-index flow variables (x^t_{ijk} and $y^{tt'}_{ijk}$), making the model computationally expensive to solve. An alternative formulation of the MFLPDDSM is to replace these variables by the following variables:

r^t_{ij} : Total quantity of product shipped from facility i to customer j in time period t $(i \in I;\ j \in J^0;\ t \in T)$ (15)

$s^{tt'}_{ij}$: Amount of product distributed from facility i to customer j in time period t' to (partially) satisfy demand of period $t\,(i \in I;\ j \in J^1;\ t \in T; t' = t, \ldots, \min\{t + \rho_j,\ |T|\})$ (16)

w^t_{ik} : Total quantity of product that is shipped from facility i with capacity level k in time period $t\,(i \in I;\ k \in K_i;\ t \in T)$. (17)

The relationship between the new variables and the original flow variables (4)–(5) is given by an appropriate aggregation of the latter as follows:

$$r^t_{ij} = \sum_{k \in K_i} x^t_{ijk} \qquad i \in I,\ j \in J^0,\ t \in T \tag{18}$$

$$s^{tt'}_{ij} = \sum_{k \in K_i} y^{tt'}_{ijk} \qquad i \in I,\ j \in J^1,\ t \in T,\ t' = t, \ldots, \min\{t + \rho_j,\ |T|\} \tag{19}$$

$$w^t_{ik} = \sum_{j \in J^0} x^t_{ijk} + \sum_{j \in J^1} \sum_{t'=\max\{1,\, t-\rho_j\}}^{t} y^{t't}_{ijk} \qquad i \in I,\ k \in K_i,\ t \in T. \tag{20}$$

Under the transformations (18)–(20), the following formulation, denoted ($\mathbf{P}_a$), is obtained:

$$\text{Min} \quad \sum_{\ell \in T_L} \sum_{i \in I} \sum_{k \in K_i} F^\ell_{ik} z^\ell_{ik} + \sum_{t \in T} \sum_{i \in I^e} M^t_{i1} \left(1 - \sum_{\ell \in T_L} z^\ell_{i1}\right) + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^0} c^t_{ij} r^t_{ij} + \sum_{t \in T} \sum_{i \in I} \sum_{k \in K_i} o^t_{ik} w^t_{ik} + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j,\, |T|\}} \left(p^{tt'}_j + c^{t'}_{ij}\right) s^{tt'}_{ij} \tag{21}$$

s.t.

(9), (14)

$$\sum_{i \in I} r^t_{ij} = d^t_j \qquad j \in J^0,\ t \in T \tag{7'}$$

$$\sum_{i \in I} \sum_{t'=t}^{\min\{t+\rho_j,\, |T|\}} s^{tt'}_{ij} = d^t_j \qquad j \in J^1,\ t \in T \tag{8'}$$

$$w^t_{ik} \le Q_{ik} \sum_{\ell \in T_L:\, \ell \le t} z^\ell_{ik} \qquad i \in I^n,\ k \in K_i,\ t \in T \tag{10'}$$

$$w^t_{i1} \le Q_{i1} \left(1 - \sum_{\ell \in T_L:\, \ell < t} z^\ell_{i1}\right) \qquad i \in I^e,\ t \in T \tag{11'}$$

$$\sum_{k \in K_i} w^t_{ik} = \sum_{j \in J^0} r^t_{ij} + \sum_{j \in J^1} \sum_{t'=\max\{1,t-\rho_j\}}^{t} s^{t't}_{ij} \quad i \in I,\ t \in T \tag{22}$$

$$r^t_{ij} \ge 0 \quad i \in I, j \in J^0, t \in T \tag{23}$$

$$s^{tt'}_{ij} \ge 0 \quad i \in I,\ j \in J^1,\ t \in T,\ t' = t, \ldots, \min\{|T|, t + \rho_j\} \tag{24}$$

$$w^t_{ik} \ge 0 \quad i \in I,\ k \in K_i,\ t \in T. \tag{25}$$

The original demand satisfaction constraints are replaced by equalities (7′) and (8′), while the capacity constraints are now imposed by conditions (10′) and (11′). The new set of constraints (22) link the newly defined continuous variables. They state that the total product outflow from a facility in a given time period is split into deliveries to customers with high service requirements (the first term on the right-hand side) and deliveries to customers accepting delays in demand satisfaction (the last term on the right-hand side). Finally, non-negativity and binary conditions are given by (14) and (23)–(25).

3.3. The single shipment case

For the customer segment J^1, formulations (**P**) and ($\mathbf{P}_a$) allow an order to be split over multiple periods of time for the same customer. However, in some cases, the customer may prefer to receive a single shipment even if it arrives with some delay. For the customer, the cost of handling a single shipment is proportionally less than the cost of processing several deliveries belonging to a particular order. To model this requirement, we introduce the following binary variables for every $j \in J^1$, $t \in T$ and $t' = t, \ldots, \min\{t + \rho_j,\ |T|\}$:

$$v^{tt'}_j = \begin{cases} 1 & \text{if all the demand of customer } j \text{ in period } t \\ & \quad \text{is delivered in period } t' \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

Table 2
Number of decision variables in the proposed formulations.

Problem	Formulation	Number of decision variables	
		Binary	Continuous
MFLPDDSM	$(\mathbf{P})$	$O(\lvert T_L\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert))$	$O(\lvert T\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert)\cdot(\lvert J^0\rvert + \bar{\rho}\cdot\lvert J^1\rvert))$
	$(\mathbf{P}_a)$	$O(\lvert T_L\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert))$	$O(\lvert T\rvert \cdot (\lvert I\rvert\cdot\lvert J^0\rvert + \bar{\rho}\cdot\lvert I\rvert\cdot\lvert J^1\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert))$
MFLPDDSM-S	$(\mathbf{Q})$	$O(\lvert T_L\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert) + \bar{\rho}\cdot\lvert J^1\rvert\cdot\lvert T\rvert)$	$O(\lvert T\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert)\cdot(\lvert J^0\rvert + \bar{\rho}\cdot\lvert J^1\rvert))$
	$(\mathbf{Q}_a)$	$O(\lvert T_L\rvert \cdot (\bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert) + \bar{\rho}\cdot\lvert J^1\rvert\cdot\lvert T\rvert)$	$O(\lvert T\rvert \cdot (\lvert I\rvert\cdot\lvert J^0\rvert + \bar{\rho}\cdot\lvert I\rvert\cdot\lvert J^1\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert))$

Table 3
Number of constraints in the proposed formulations.

Problem	Formulation	Number of constraints
MFLPDDSM	$(\mathbf{P})$	$O(\lvert T\rvert \cdot (\lvert J\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert) + \lvert I\rvert)$
	$(\mathbf{P}_a)$	$O(\lvert T\rvert \cdot (\lvert I\rvert + \lvert J\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert) + \lvert I\rvert)$
MFLPDDSM-S	$(\mathbf{Q})$	$O(\lvert T\rvert \cdot (\lvert J\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert + \bar{\rho}\cdot\lvert J^1\rvert) + \lvert I\rvert)$
	$(\mathbf{Q}_a)$	$O(\lvert T\rvert \cdot (\lvert I\rvert + \lvert J\rvert + \bar{k}\cdot\lvert I^n\rvert + \lvert I^e\rvert + \bar{\rho}\cdot\lvert J^1\rvert) + \lvert I\rvert)$

We denote the single shipment case by MFLPDDSM-S and adapt formulations $(\mathbf{P})$ and $(\mathbf{P}_a)$ accordingly. In the first case, the demand satisfaction constraints (8) for customers that accept late deliveries are replaced by the following three sets:

$$\sum_{i\in I}\sum_{k\in K_i} y_{ijk}^{tt'} = d_j^t v_j^{tt'} \qquad j\in J^1,\ t\in T,\ t'=t,\ldots,\min\{t+\rho_j,\ |T|\} \tag{27}$$

$$\sum_{t'=t}^{\min\{t+\rho_j,\ |T|\}} v_j^{tt'} = 1 \qquad j\in J^1,\ t\in T \tag{28}$$

$$v_j^{tt'} \in \{0,1\} \qquad j\in J^1,\ t\in T,\ t'=t,\ldots,\min\{t+\rho_j,\ |T|\}. \tag{29}$$

Let us denote by $(\mathbf{Q})$ the MILP model defined by (6)–(7), (9)–(14), (27)–(29).

The counterpart of formulation $(\mathbf{P}_a)$ is obtained by replacing constraints (8′) by

$$\sum_{i\in I} s_{ij}^{tt'} = d_j^t v_j^{tt'} \qquad j\in J^1,\ t\in T,\ t'=t,\ldots,\min\{t+\rho_j,\ |T|\}. \tag{30}$$

As a result, the new MILP formulation, $(\mathbf{Q}_a)$, has the objective function (21) and constraints (7′), (9), (10′), (11′), (14), (22)–(25), (28)–(30).

3.4. Comparison of formulations

Let $\bar{k}$ be the largest number of capacity levels that are available, that is, $\bar{k} = \max_{i\in I}\{|K_i|\}$. Moreover, we define $\bar{\rho} = 1 + \max_{j\in J^1}\{\rho_j\}$. This is the maximum time span between order placement and order delivery for a customer belonging to segment J^1. Recall that the demand of such a customer occurring in period t must be delivered over periods $t,\ldots,\min\{t+\rho_j,\ |T|\}$. Table 2 displays the size of the proposed formulations with respect to the number of their variables, whereas Table 3 reports on the number of constraints.

In the MFLPDDSM, there is a significant reduction in the total number of continuous variables and a marginal increase in the number of constraints (namely, $|I|\cdot|T|$) in model $(\mathbf{P}_a)$ compared to $(\mathbf{P})$. In fact, if n denotes the number of continuous variables in the latter formulation then $(\mathbf{P}_a)$ has only $n/\bar{k} + \bar{k}\cdot|I|\cdot|T|$ such variables. The requirement of single shipments for customers accepting delayed demand satisfaction results in a considerable increase in the number of binary variables, namely in $\bar{\rho}\cdot|J^1|\cdot|T|$ additional variables. The number of constraints also increases by the same factor. The alternative formulation $(\mathbf{Q}_a)$ benefits from the same variable reduction as $(\mathbf{P}_a)$.

Let $(\bar{\mathbf{P}})$, resp. $(\bar{\mathbf{P}}_a)$, denote the linear relaxation of formulation $(\mathbf{P})$, resp. $(\mathbf{P}_a)$. The following result states that formulations $(\bar{\mathbf{P}})$ and $(\bar{\mathbf{P}}_a)$ are equally strong in terms of the lower bounds they provide. We denote by $v(\cdot)$ the optimal objective function value of a model.

Theorem. $v(\bar{\mathbf{P}}) = v(\bar{\mathbf{P}}_a)$

Proof. See Appendix A. □

In spite of the above result, our computational experience (see Section 5) indicates that the different sizes of the formulations have a significant impact from a computational standpoint. In particular, formulation $(\mathbf{P}_a)$ outperforms formulation $(\mathbf{P})$.

4. Additional inequalities

In this section, we develop two groups of valid inequalities in an attempt to strengthen the bounds of the linear relaxations of the mathematical models introduced in the previous section. Both groups of inequalities strengthen the setting of the strategic decisions (i.e. the binary variables) by imposing a lower bound for the total number of facilities that must operate during given time periods. The first group of inequalities is associated with the strategic periods $\ell\in T_L$, whereas the second group concerns those time periods immediately after $\ell\in T_L$.

4.1. Inequalities involving strategic time periods

Denoting by R^ℓ the minimum number of facilities that must operate in time period $\ell\in T_L$, the following inequalities can be added to the MILP formulations:

$$|I^e| + \sum_{i\in I^n}\sum_{k\in K_i} z_{ik}^1 \geq R^1 \tag{31}$$

$$\sum_{i\in I^e}\left(1 - \sum_{\ell'\in T_L:\ \ell'<\ell} z_{i1}^{\ell'}\right) + \sum_{i\in I^n}\sum_{k\in K_i}\sum_{\ell'\in T_L:\ \ell'\leq\ell} z_{ik}^{\ell'} \geq R^\ell \qquad \ell\in T_L\setminus\{1\}. \tag{32}$$

Condition (31) is established for the first period in the time horizon by assuming, without loss of generality, that the latter corresponds to the first opportunity to make location and capacity acquisition decisions. Inequalities (32) apply to the remaining strategic periods. In both cases, the left-hand side represents the total number of facilities operating at time period ℓ. Since the decision to remove any initially existing facility takes place at the end of a strategic period, all existing locations are available in period $\ell=1$.

In order to determine an appropriate value for the lower bound R^ℓ, we need to consider the minimum quantity D^ℓ that must be delivered in time period ℓ. This corresponds to the total demand requirements of the preferred customer segment:

$$D^\ell = \sum_{j\in J^0} d_j^\ell \qquad \ell\in T_L.$$

At the beginning of the time horizon, the total capacity available in the facility system is equal to $\sum_{i\in I^e} Q_{i1}$. Hence, if $D^1 \le \sum_{i\in I^e} Q_{i1}$ then $R^1 = |I^e|$, meaning that the minimum demand requirements can be met by the capacity provided by the initially existing facilities. However, in the case that $D^1 > \sum_{i\in I^e} Q_{i1}$, it is necessary to expand the system through opening new facilities. For this purpose, we consider the largest capacity level that can be installed in each candidate location $i \in I^n$ and build a sequence with these capacity sizes $Q_{i|K_i|}$ sorted by non-increasing order. Let us denote this sequence by $\widetilde{Q}_{[1]} \ge \widetilde{Q}_{[2]} \ge \ldots \ge \widetilde{Q}_{[|I^n|]}$. After having identified the number m of required capacity levels such that the following inequalities hold

$$\sum_{i=1}^{m-1} \widetilde{Q}_{[i]} < D^1 - \sum_{i\in I^e} Q_{i1} \le \sum_{i=1}^{m} \widetilde{Q}_{[i]}$$

we set $R^1 = |I^e| + m$. By restricting the selection of capacity levels to the largest sizes that could be installed, we ensure that R^1 is indeed a lower bound for the actual number of facilities that must operate in the first period.

In all strategic periods ℓ other than the first, a similar sequence of capacity levels needs to be created in order to determine an appropriate value for R^ℓ. The only difference lies in the fact that the capacities of all initially existing facilities must also be considered in addition to the largest capacity levels that can be installed at candidate sites. With the sorted sequence of capacity sizes thus obtained (i.e. $\widetilde{Q}_{[1]} \ge \widetilde{Q}_{[2]} \ge \ldots \ge \widetilde{Q}_{[|I^n|+|I^e|]}$), we identify the minimum number of facilities that must be available in time period ℓ by employing similar conditions as above. In other words, for every $\ell \in T_L\backslash\{1\}$ we take $R^\ell = m$ such that

$$\sum_{i=1}^{m-1} \widetilde{Q}_{[i]} < D^\ell \le \sum_{i=1}^{m} \widetilde{Q}_{[i]}.$$

4.2. Inequalities involving selected tactical time periods

We now derive a second group of valid inequalities that serve a purpose similar to (31)–(32) but apply to those time periods immediately after the strategic decisions are made. Observe that the status of each facility remains unchanged over all intermediate periods between two consecutive strategic periods ℓ and $\ell' \in T_L$. Therefore, at the end of period ℓ, all strategic decisions made until then define the subset of facilities that are open over $\{\ell+1, \ldots, \ell'-1\}$ along with their capacities. These facilities will have to serve at least part of the demand occurring over the intermediate periods, namely all orders placed by customers $j \in J^0$ as well as those orders placed by customers $j \in J^1$ that must be satisfied before the next strategic period, even if some of these customers may experience delivery delays.

To illustrate this case, suppose that the time horizon spans 3 years with each year being divided into 12 months. Moreover, location and capacity acquisition decisions can be taken at the beginning of each year. Therefore, $|T| = 36$, $T_L = \{1, 13, 25\}$ and $\ell_{max} = 25$. Let us assume that the maximum allowed delay ρ_j is the same for all customers $j \in J^1$ and that it is equal to two periods. For example, over the time interval covering the tactical periods $t = 2, \ldots, 12$, all demands of the preferred customer segment must be filled. In addition, the orders placed by customers $j \in J^1$ in periods $t = 2, \ldots, 10$ must also be served. Demand requirements of these customers for periods $t = 11$ and $t = 12$ could be satisfied in periods beyond this time interval through using capacity that would only become available in the next strategic period $\ell = 13$. Therefore, these demands do not have to be considered. This line of reasoning also applies to the intermediate periods $t = 14, \ldots, 24$ but not to the last tactical time interval. In the latter case, all demand requirements of the customers that accept delayed deliveries must also be met.

Let us assume without loss of generality that $|T|$ is a multiple of $|T_L|$ and let us denote the number of time periods between two consecutive strategic periods by $\tau = |T|/|T_L|$. In the above example, $\tau = 12$. The minimum quantity of demand that must be served between two strategic time periods is determined as follows:

$$\overline{D}^{\ell+1} = \begin{cases} \sum_{j\in J^0} \sum_{t=\ell+1}^{\ell+\tau-1} d_j^t + \sum_{j\in J^1} \sum_{t=\ell+1}^{\ell+\tau-\rho_j-1} d_j^t & \text{if } \ell \in T_L \text{ and } \ell < \ell_{max} \\ \sum_{j\in J} \sum_{t=\ell+1}^{|T|} d_j^t & \text{if } \ell = \ell_{max} \end{cases}$$

We also need to sort the capacity levels of all facilities $i \in I$ according to the same procedure presented in Section 4.1 and use the ordered sequence to identify the minimum number of facilities that must operate in time period $\ell+1$ ($\ell \in T_L$) as follows:

$$\sum_{i=1}^{m-1} (\tau-1)\,\widetilde{Q}_{[i]} < \overline{D}^{\ell+1} \le \sum_{i=1}^{m} (\tau-1)\,\widetilde{Q}_{[i]}.$$

Clearly, $R^{\ell+1} = m$ for every $\ell \in T_L$. Finally, the following inequalities are valid for the mathematical formulations introduced in Section 3:

$$\sum_{i\in I^e} \left(1 - \sum_{\ell'\in T_L:\, \ell'\le \ell} z_{i1}^{\ell'}\right) + \sum_{i\in I^n} \sum_{k\in K_i} \sum_{\ell'\in T_L:\, \ell'\le\ell} z_{ik}^{\ell'} \ge R^{\ell+1} \qquad \ell \in T_L. \tag{33}$$

Inequalities (31)–(33) are computationally inexpensive since in total only $2\,|T_L|$ constraints need to be added to each formulation. The resulting MILP models are denoted by ($\mathbf{P}^+$), ($\mathbf{P}_a^+$), ($\mathbf{Q}^+$) and ($\mathbf{Q}_a^+$).

5. Computational study

In this section, we describe the data generation scheme developed for our computational experiments followed by a discussion of the numerical results.

5.1. Description of test instances

Since the problems studied in this paper are new, benchmark instances are not available and consequently we randomly generated a set of test instances. The size of each instance is mainly dictated by the length of the planning horizon and the total number of customers as shown in Table 4.

Three capacity levels are considered at each candidate location representing small, medium and large sizes. The planning horizon spans 3 years, each comprising 12-month periods, which yields in total 36 time periods. Location decisions may be taken either once a year ($|T_L| = 3$) or once every 6 months ($|T_L| = 6$). In the first case, the strategic location periods are $T_L = \{1, 13, 25\}$ and in the second case, $T_L = \{1, 7, 13, 19, 25, 31\}$. Observe that depending on the interpretation given to a time period, different time frames can be

Table 4
Cardinality of index sets.

Parameter	Value
$\|J\|$	$100, 150$
$\|J^0\|$	$\lceil \beta^J \|J\| \rceil$ with $\beta^J \in \{0.25, 0.5, 0.75\}$
$\|I\|$	$0.1\,\|J\|$
$\|I^n\|$	$0.8\,\|I\|$
$\|K_i\|$	$3\,(i \in I^n)$
$\|T\|$	36
$\|T_L\|$	$3, 6$

Table 5
Demand and capacity parameters.

Parameter	Value	
d_j^1	$U[20, 100]$	
d_j^t	$\beta_d^t\, d_j^{t-1}$	$t = 2, \ldots, \lvert T\rvert;\ \beta_d^t \in U[0.95,\ 1.05]$
ρ_j	1, 2, 3	$j \in J^1$
$Q_{i\lvert K_i\rvert}$	$\frac{1}{\lvert I\rvert}\, U[2,\ 3]\, \frac{\sum_{j\in J}\sum_{t\in T} d_j^t}{\lvert T\rvert}$	$i \in I^n$
Q_{ik}	$0.7\, Q_{i,k+1}$	$i \in I^n;\ k \in K_i\backslash\{\lvert K_i\rvert\}$

captured. For example, if a time period represents a 2-month or a 3-month period then $|T| = 36$ corresponds to a time frame of 6 and 9 years, respectively.

In what follows, we denote by $U[a, b]$ the generation of random numbers over the range $[a, b]$ according to a continuous uniform distribution. Table 5 describes how customer demands and the sizes of the capacity levels are obtained.

We note that both downward and upward demand fluctuations are possible over the time horizon. In fact, changes are allowed up to ±5 percent between two consecutive time periods. In the customer segment J^1, lateness in demand fulfilment ranges from one to three time periods. Regarding the capacity choices, the small ($k = 1$), resp. medium ($k = 2$), size corresponds to 49 percent, resp. 70 percent, of the largest size. Observe that the latter depends on the mean total demand per period. To determine the initial capacity of each existing facility, also three capacity levels are generated according to the procedure in Table 5. One of these capacities is then selected at random.

The generation of the variable costs relies on two random numbers, β_1 and β_2, both belonging to $U[1.01, 1.03]$ (details are given next).

- For $i \in I$ and $j \in J$, the variable distribution costs are set according to

$$\begin{array}{ll} c_{ij}^1 = U[5, 10] & \\ c_{ij}^t = c_{ij}^1 & t = 2, \ldots, 12 \\ c_{ij}^t = \beta_1\, c_{ij}^{12} & t = 13, \ldots, 24 \\ c_{ij}^t = \beta_2\, c_{ij}^{24} & t = 25, \ldots, 36. \end{array}$$

  It is assumed that the distribution costs are constant over 1 year (12 periods) but they increase between 1 percent and 3 percent from 1 year to the next.
- The variable processing costs at the facilities are generated in order to reflect economies of scale by considering the available capacity levels. Hence, the larger the capacity size, the lower the corresponding processing cost per unit of product. For $i \in I$, we set

$$\begin{array}{ll} o_{i1}^1 = 100 / \sqrt{Q_{i1}} & t = 1, \ldots, 12 \\ o_{ik}^t = 0.9\, o_{ik-1}^t & t = 1, \ldots, 12;\ k \in K_i \setminus \{1\} \\ o_{ik}^t = \beta_1\, o_{ik}^{12} & t = 13, \ldots, 24;\ k \in K_i \\ o_{ik}^t = \beta_2\, o_{ik}^{24} & t = 25, \ldots, 36;\ k \in K_i. \end{array}$$

  Cost fluctuations follow a pattern similar to that of the variable distribution costs.

To obtain the fixed facility costs given in (1), we describe next how the various components are generated.

- The fixed costs of opening new facilities at candidate sites $i \in I^n$ reflect economies of scale by taking into account the three capacity levels. In the first time period, these costs are set according to

$$FO_{i1}^1 = \alpha_i + \gamma_i \sqrt{Q_{i1}} \qquad \alpha_i \in [0, 1000];\ \gamma_i \in [6000, 6500]$$

$$FO_{ik}^1 = 1.01\left(FO_{ik-1}^1 + o_{ik-1}^1\, Q_{ik-1}\right) - o_{ik}^1\, Q_{ik-1} \qquad k \in K_i \setminus \{1\}.$$

  In the remaining strategic periods $\ell \in T_L\backslash\{1\}$, we take $FO_{ik}^\ell = \beta_{|T_L|}\, FO_{ik}^{\ell-1}$ for every $k \in K_i$. For $|T_L| = 3$, we set $\beta_{|T_L|} = U[1.01,\ 1.03]$ and for $|T_L| = 6$, we consider $\beta_{|T_L|} = U[\sqrt{1.01},\ \sqrt{1.03}\,]$.
- A similar scheme is used to generate the fixed closing costs. Recall that the capacity of an initially existing facility $i \in I^e$ is randomly chosen among three sizes (small, medium and large, see Table 5). For these capacity levels $k \in \{1, 2, 3\}$, auxiliary fixed costs FO_{ik}^ℓ are determined for the existing facility i through applying the above procedure. The fixed closing cost of that facility corresponds to 20 percent of the associated auxiliary cost, that is, $FC_{i1}^\ell = 0.2\, FO_{is}^\ell$ ($\ell \in T_L$) with s denoting the randomly selected capacity size ($s \in \{1, 2, 3\}$). Facility closing costs account, for example, for indemnity payments due to termination of employment and work contracts, expenditures due to inventory and equipment transfers from the closed sites, payments incurred by disposal activities, and workforce relocation and retraining expenditures.
- Facility maintenance costs correspond to 5 percent of the associated opening/auxiliary costs. On the one hand, if a new facility is established at candidate site $i \in I^n$ with size $k \in K_i$ in time period $\ell \in T_L$ then the maintenance cost $MC_{ik}^t = 0.05\, FO_{ik}^\ell$ is incurred in every period $t = \ell, \ldots, |T|$. On the other hand, closing an initially existing facility $i \in I^e$ in period $\ell \in T_L$ incurs maintenance costs $MC_{i1}^t = 0.05\, FO_{is}^\ell$ in all periods $t = 1, \ldots, \ell$ (recall that s denotes the capacity size that was randomly selected for facility i, $s \in \{1, 2, 3\}$).

Finally, tardiness penalty costs for orders delivered with delay to customers $j \in J^1$ result from combining the average maintenance, distribution and processing costs in the following way:

$$p_j^{tt'} = 0.1\, \theta_j^t\, (t' - t)^2 \qquad t \in T;\ t' = t, \ldots, \min\{t + \rho_j,\ |T|\}) \tag{34}$$

with

$$\theta_j^t = \frac{\sum_{i\in I}\sum_{k\in K_i} M_{ik}^t}{TD_t\, |I|\, \sum_{i\in I} |K_i|} + \frac{\sum_{i\in I} c_{ij}^t}{|I|} + \frac{\sum_{i\in I}\sum_{k\in K_i} o_{ik}^t}{|I|\, \sum_{i\in I} |K_i|}$$

and TD_t denoting the total quantity demanded in period t, that is, $TD_t = \sum_{j\in J} d_j^t$. The above scheme was motivated by a procedure used by Albareda-Sambola et al. (2010) in the context of a facility location problem with lost sales.

A sample of representative test instances was obtained with the above procedure. For each choice of the total number of customers according to Table 4, six test instances were randomly generated, thus yielding a total of 36 instances. Each one of these instances was considered with four values for the maximum allowed delay in demand fulfilment, namely 0, 1, 2 and 3. All test instances are feasible since sufficient capacity is available to satisfy the demand requirements of all customers. Moreover, all cost parameters associated with capacity levels reflect economies of scale.

5.2. Numerical results

The formulations including their enhancements were implemented in C++ using IBM ILOG Concert Technology and solved with IBM ILOG CPLEX 12.3. All experiments were conducted on a PC with a 3.4 gigahertz Intel Core i7-2600K processor, 8 gigabyte RAM and running Windows 7 (64-bit). A limit of 10 hours of CPU time was set for each instance. CPLEX was used with default settings, as it is typically the case in practice, making full use of its MIP heuristics in an attempt to find high quality solutions at an early stage of the branch-and-cut algorithm. Finally, a deterministic parallel mode was selected to ensure that multiple runs with the same instance reproduce the same solution path and results.

Table 6
Sizes of test instances with formulations (**P**) and ($\mathbf{P}_a$) for the MFLPDDSM under different values of the maximum delivery delay.

ρ		No. of binary variables			No. of continuous variables			No. of constraints		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
0	(**P**)	78	146.3	234	93600	152100	210600	4546	5682.5	6819
	($\mathbf{P}_a$)	78	146.3	234	36936	59670	82404	4906	6132.5	7359
1	(**P**)	78	146.3	234	116350	225810	363480	4546	5682.5	6819
	($\mathbf{P}_a$)	78	146.3	234	45686	88020	141204	4906	6132.5	7359
2	(**P**)	78	146.3	234	138450	297414	511992	4546	5682.5	6819
	($\mathbf{P}_a$)	78	146.3	234	54186	115560	198324	4906	6132.5	7359
3	(**P**)	78	146.3	234	159900	366912	656136	4546	5682.5	6819
	($\mathbf{P}_a$)	78	146.3	234	62436	142290	253764	4906	6132.5	7359

Table 7
Summary of results for the MFLPDDSM under different values of the maximum delivery delay; *instances not solved to optimality within 10 hours.

ρ	No. of opt sol./No. of non-opt. sol.				MIP gap (percent)*			LP gap (percent)		MIP CPU (seconds)			LP CPU (seconds)		
	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)		(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**), ($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)
0	33/3	**34**/2	**34**/2	Min	0.38	0.43	0.59	1.98	1.32	55.43	21.40	9.28	0.67	0.33	0.53
				Avg	0.84	**0.44**	1.05	3.82	**2.94**	7021.07	4957.04	**3612.76**	1.84	**0.80**	1.52
				Max	1.13	0.44	1.50	10.99	6.12	36000.00	36000.00	36000.00	3.93	1.36	3.43
1	23/13	**30**/6	**30**/6	Min	0.26	0.27	0.78	0.65	0.65	89.89	23.09	11.90	3.57	1.19	0.98
				Avg	1.21	**1.04**	1.22	2.93	**2.85**	15743.01	11014.94	**10851.84**	12.64	**3.18**	3.44
				Max	2.77	1.73	1.74	5.65	5.65	36000.00	36000.00	36000.00	28.59	6.01	6.19
2	24/12	**29**/7	**29**/7	Min	1.06	0.39	0.58	0.65	0.65	56.75	32.20	19.91	5.23	1.58	1.62
				Avg	1.61	**1.43**	1.52	2.85	**2.78**	15796.96	**12309.21**	12624.36	18.16	**4.26**	4.48
				Max	3.09	2.69	2.79	5.65	5.65	36000.00	36000.00	36000.00	43.82	8.99	10.11
3	22/14	**27**/9	**27**/9	Min	1.18	0.57	0.60	0.65	0.65	96.60	39.52	35.65	6.55	1.86	2.00
				Avg	1.77	1.58	**1.46**	2.84	**2.83**	16762.89	**13555.61**	13569.86	23.06	**5.17**	5.40
				Max	3.49	4.61	2.89	5.65	5.65	36000.00	36000.00	36000.00	67.06	12.09	12.42
All	102/42	**120**/24	**120**/24	Avg	1.49	**1.31**	1.38	3.11	**2.85**	13830.98	**10459.20**	10614.52	13.92	**3.35**	3.71

*5.2.1. Formulations (**P**), ($\mathbf{P}_a$) and ($\mathbf{P}^+{}_a$)*

Table 6 summarizes the sizes of the test instances with formulations (**P**) and ($\mathbf{P}_a$) for different choices of the maximum allowed delay in demand fulfilment. The latter is denoted by ρ in the first column. Each value of ρ indicates that the demands of all customers in segment J^1 must be satisfied within the same number of time periods ($\rho_j = \rho,\ j \in J^1$). In particular, $\rho = 0$ corresponds to the classical setting in multi-period facility location in which all customer demand must be served without delays. In this case, $J = J^0$ and $J^1 = \emptyset$.

The number of binary variables **z** is identical in both formulations. As discussed in Section 3.4, instances solved with formulation ($\mathbf{P}_a$) contain significantly fewer continuous variables than formulation (**P**). In fact, the number of continuous variables in ($\mathbf{P}_a$) is on average only 39 percent of the corresponding number in model (**P**). However, instances run with formulation ($\mathbf{P}_a$) have on average 7.9 percent more constraints. The introduction of the additional inequalities described in Section 4 has little impact on the model sizes and therefore, the corresponding information is omitted from the table. Instances with $|T_L| = 3$, resp. $|T_L| = 6$, strategic periods have 6, resp. 12, more constraints.

To evaluate the effectiveness of the proposed formulations (**P**), ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$), we compare them by means of their LP-relaxation bounds, the time necessary to solve the LP-relaxations, the optimality gaps reported by CPLEX and the computing times required to solve the instances. Table 7 summarizes the results obtained for different values of the maximum allowed delay in demand fulfilment, ρ. For each one of the three formulations, the table reports the number of instances that were solved to optimality (No. of opt. sol.) as well as the number of instances for which the optimal solution was not identified within the given time limit of 10 hours (No. of non-opt. sol.). For the latter instances, we also present the minimum, average and maximum optimality gaps as reported by CPLEX for each type of formulation (MIP gap $= (z^{UB} - z^{LB})/z^{UB} \times 100$ percent with z^{UB} denoting the optimal objective value or the value of the best feasible solution and z^{LB} representing the best lower bound). In addition, the relative percentage deviation between the objective value of the best feasible solution available and the LP-relaxation bound (z^{LP}) is presented in columns 9 and 10 (LP gap $= (z^{UB} - z^{LP})/z^{UB} \times 100$ percent). Since the LP-relaxations of models (**P**) and ($\mathbf{P}_a$) provide the same lower bound, the latter is shown in column 9. The minimum, average and maximum computing times (in seconds) to solve the instances for each one of the formulations are displayed in columns 11–13 (MIP CPU). Finally, the minimum, average and maximum computing times (also in seconds) required by the LP-relaxations are given in the last three columns (LP CPU). The last row of Table 7 reports the average values over all instances (in total 144). For each choice of the parameter ρ, the best average result is highlighted by boldface. Detailed results are given in Tables 16–19 in Appendix B for each value of the parameter ρ.

A closer look at Table 7 indicates that the proposed alternative formulations ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$) perform consistently better than formulation (**P**). Not only more instances were solved to optimality but also substantially shorter computing times were obtained with models ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$). Furthermore, the optimality gaps decreased for those instances in which the pre-specified time limit was attained. Hence, although formulations (**P**) and ($\mathbf{P}_a$) have the same linear relaxation values, ($\mathbf{P}_a$) clearly outperforms (**P**). The lower bound provided by the LP-relaxation of formulation ($\mathbf{P}_a^+$) is slightly better, as expected. This is achieved at the cost of marginally higher computing times compared to formulation ($\mathbf{P}_a$).

Table 8
Sizes of instances with formulation ($\mathbf{Q}_a^+$) for the MFLPDDSM-S under different values of the maximum delivery delay.

ρ	No. of binary var.			No. of continuous var.			No. of constraints		
	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
1	1853	4571.9	8186	45686	88020	141204	6687	10567.2	15323
2	2703	6691.3	11994	54186	115560	198324	7537	12686.5	19131
3	3528	8748.3	15690	62436	142290	253764	8362	14743.5	22827

Table 9
Summary of results with formulation ($\mathbf{Q}_a^+$) for the MFLPDDSM-S under different values of the maximum delivery delay; *instances not solved to optimality within 10 hours.

ρ	No. of opt. sol./ no. of non-opt. sol./ no. of out of memory		Gap (percent) MIP*	Gap (percent) LP	CPU (seconds) MIP	CPU (seconds) LP
1	13/21/2	Min	0.01	1.44	183.50	1.33
		Avg	2.20	3.61	28636.71	4.71
		Max	6.15	7.50	36000.00	9.06
2	3/32/1	Min	0.01	1.28	956.07	2.53
		Avg	1.85	3.48	33259.83	7.33
		Max	4.36	6.67	36000.00	17.11
3	1/33/2	Min	0.10	1.01	25266.11	2.90
		Avg	1.97	3.43	35684.30	9.04
		Max	4.60	8.56	36000.00	19.67
All	17/86/5	Avg	1.98	3.51	32534.06	7.03

Table 10
Percentage of demand splitting for the MFLPDDSM-S.

Customer segment	ρ	No. of operating facilities		
		1	2	3
J^0	1	97.51	2.46	0.03
	2	98.11	1.88	0.01
	3	97.98	2.01	0.01
J^1	1	97.36	2.61	0.03
	2	97.78	2.21	0.01
	3	97.66	2.33	0.01

The results further reveal that increasing the maximum delivery lead time for customers accepting delayed shipments yields more challenging instances. This feature is reflected in the growing number of instances that could not be solved to optimality within the given time limit and also in the increasing average integrality gaps. Interestingly, the LP-bounds seem to be insensitive to the parameter ρ. This aspect is very important, especially regarding formulation ($\mathbf{P}_a^+$). The tight linear relaxation bounds provided by this model (on average, 2.85 percent) indicate that the LP-relaxation of ($\mathbf{P}_a^+$) could be used to evaluate the quality of feasible solutions obtained, for example, by means of a tailored heuristic method. In addition, tight LP-bounds can also be very helpful to accelerate the solution process when a branch-and-bound or branch-and-cut method has been especially designed to solve the problem to optimality. The differences in the computing times of formulations ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$) are not significant for $\rho \geq 1$. The largest variation occurs in those instances where all customers must have their demands satisfied without delays ($\rho = 0$). In these cases, the overall computational effort is considerably smaller with model ($\mathbf{P}_a^+$) than with model ($\mathbf{P}_a$).

5.2.2. Numerical results for the single shipment case

Since the proposed alternative models ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$) have shown to be clearly superior in terms of overall computational performance, we decided to study the MFLPDDSM-S by evaluating the effectiveness of formulation ($\mathbf{Q}_a^+$) (recall that for the MFLPDDSM the best LP-relaxation bound is obtained with the additional constraints introduced in Section 4). As shown in Table 8, the set of binary variables increases by a factor ranging from 31.3 ($\rho = 1$) to 59.8 ($\rho = 3$) compared to the problem variant where late deliveries can be split over several time periods. Formulations ($\mathbf{P}_a$), ($\mathbf{P}_a^+$), ($\mathbf{Q}_a$), and ($\mathbf{Q}_a^+$) have the same number of continuous variables (recall Table 2). In contrast, formulation ($\mathbf{Q}_a^+$) has roughly twice as many constraints as model ($\mathbf{P}_a^+$). Hence, the problem instances result in large-scale model representations.

Table 9 summarizes the results obtained under the requirement of single shipments for customers accepting late deliveries and for maximum delivery lead times ranging from one to three periods. The structure of the table is similar to that of Table 7 but it only includes information about a single formulation, namely ($\mathbf{Q}_a^+$). Since CPLEX has run out of memory in some of the instances, we also display in column 2 the number of instances exhibiting this characteristic. Given the large size of the MILPs, it is not surprising that CPLEX's ability to prove optimality within the given time limit greatly declines. This occurs with 79.6 percent of the instances and thus, the computing times increase significantly compared to the MFLPDDSM. Nevertheless, despite the large number of variables and constraints, the integrality gaps are rather small, ranging from 0.01 percent to 6.15 percent. Furthermore, the maximum allowed delay in demand fulfilment does not seem to have a significant impact on the MIP gap although the size of the instances increases considerably as the value of ρ grows (recall Table 8). The linear relaxation also provides very good lower bounds, even if the LP gaps are slightly larger compared to those given in Table 7. This results from the fact that proven optimal solutions are only available for 15.7 percent of the instances. The major difference lies in the increasing computing times of the LP-relaxation but for such large MILPs this does not pose a limitation in practice, since all LP-relaxations required less than 20 seconds to be solved to optimality. One important drawback is that no feasible solution could be identified in 4.6 percent of the instances (5 out of 108) due to insufficient memory of the computer used. In this case, the importance of obtaining good LP-bounds becomes even greater in order to be able to evaluate, for example, the quality of upper bounds produced by a heuristic method.

We also investigated the extent to which demand splitting occurs in the best solutions identified by CPLEX. Observe that in the single shipment variant, the demand of every customer is allowed to be served by multiple facilities in each time period. Table 10 illustrates the presence of this feature in the solutions obtained. For the instances associated with a given value of the maximum allowed delivery delay, we determined for each customer segment the number of demands that are satisfied by a single facility, two facilities and three facilities (a larger number of sources were not observed in any solution). We express the calculated values as a percentage of the total number of demands within a customer segment over the planning horizon. It can be seen that about 98 percent of all customer demands are supplied from a single location. Interestingly, this feature is present across all customer segments and all values of the parameter ρ. So even though single-sourcing requirements are not explicitly enforced, the problem structure is the main driver for assigning almost every demand to a single facility.

Table 11
Performance of formulations (**P**), ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$) for the MFLPDDSM as a function of the total number of customers and the total number of strategic periods; *instances not solved to optimality within 10 hours.

Parameter		No. of opt. sol./no. of non-opt. sol.				MIP gap (percent)*			LP gap (percent)		MIP CPU (seconds)			LP CPU (seconds)		
Symbol	Value	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)		(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**), ($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)	(**P**)	($\mathbf{P}_a$)	($\mathbf{P}_a^+$)
$\|J\|$	100	66/6	**72**/0	**72**/0	Min	0.26			1.12	1.12	55.43	21.40	9.28	0.67	0.33	0.53
					Avg	1.35			3.83	**3.38**	6324.73	3573.37	3642.28	6.61	1.79	1.96
					Max	1.85			10.99	6.12	36000.00	24218.42	29400.54	17.94	3.99	3.85
	150	36/36	**48**/24	**48**/24	Min	0.38	0.27	0.58	0.65	0.65	454.38	93.16	82.46	1.33	0.76	1.11
					Avg	1.51	1.31	1.38	2.39	**2.32**	21337.23	17345.03	17586.77	21.24	4.92	5.46
					Max	3.49	4.61	2.89	4.39	4.39	36000.00	36000.00	36000.00	67.06	12.09	12.42
$\|T_L\|$	3	56/16	**62**/10	**62**/10	Min	0.48	0.27	0.78	0.92	0.92	55.43	23.74	9.28	0.67	0.33	0.53
					Avg	1.36	1.06	1.16	3.08	**2.80**	10836.81	8393.59	8343.03	13.41	3.25	3.60
					Max	2.32	1.49	1.93	10.99	6.12	36000.00	36000.00	36000.00	67.06	12.09	12.42
	6	46/26	**58**/14	**58**/14	Min	0.26	0.39	0.58	0.65	0.65	56.75	21.40	16.05	0.94	0.44	0.78
					Avg	1.57	1.49	1.54	3.14	**2.89**	16825.15	12524.81	12886.02	14.44	3.46	3.82
					Max	3.49	4.61	2.89	7.38	5.65	36000.00	36000.00	36000.00	57.11	12.06	12.18

Table 12
Performance of formulation ($\mathbf{Q}_a^+$) for the MFLPDDSM-S as a function of the total number of customers and the total number of strategic periods; *instances not solved to optimality within 10 hours.

Parameter		No. of opt. sol./		Gap (percent)		CPU (seconds)	
Symbol	Value	no. of non-opt. sol./ no. of out of memory		MIP*	LP	MIP	LP
$\|J\|$	100	13/37/4	Min	0.01	1.44	183.50	1.33
			Avg	2.20	4.14	29401.48	3.78
			Max	4.60	8.56	36000.00	7.19
	150	4/49/1	Min	0.01	1.01	19675.67	2.23
			Avg	1.82	2.91	35489.33	10.08
			Max	6.15	7.50	36000.00	19.67
$\|T_L\|$	3	12/38/4	Min	0.01	1.01	183.50	1.33
			Avg	1.80	3.34	30995.63	6.81
			Max	6.15	7.50	36000.00	18.58
	6	5/48/1	Min	0.01	1.11	1747.23	1.90
			Avg	2.13	3.67	33985.41	7.22
			Max	4.60	8.56	36000.00	19.67

5.2.3. Analysis of selected parameters

In this section, we investigate how sensitive our models are to given input parameters. Specifically, we evaluate the results obtained as a function of the total number of customers and the number of strategic periods in the planning horizon over the range of maximum delivery lead times. Tables 11 and 12 present summary statistics for the standard problem and the single shipment case, respectively.

As expected, increasing the total number of customers results in significantly more challenging instances for both the MFLPDDSM and the MFLPDDSM-S. All instances of the MFLPDDSM with 100 customers were solved to optimality with formulations ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$), while the optimal solution was identified in only two-thirds of the instances with 150 customers. In the latter case, the average computing time is almost five times larger. Moreover, formulations ($\mathbf{P}_a$) and ($\mathbf{P}_a^+$) are clearly superior to model (**P**).

Increasing the total number of strategic time periods for making location and capacity acquisition decisions results in significantly more binary variables (recall Table 2), especially in the single shipment case. Therefore, it is not surprising that instances with six strategic periods require more CPU time compared to instances with only three strategic periods. A similar effect is also observed in terms of the total number of instances solved to optimality by CPLEX. An interesting feature is that the LP-relaxation bounds provided by formulations ($\mathbf{P}_a^+$) and ($\mathbf{Q}_a^+$) are strong, regardless of the cardinality of the sets J and T_L.

Furthermore, we also assessed the performance of two formulations, ($\mathbf{P}_a^+$) and ($\mathbf{Q}_a^+$), for three different levels of the tardiness penalty costs: low, medium and high. Large costs were generated according to (34), medium costs correspond to halving (34) and low costs are represented by dividing (34) by a factor of four. Table 13 summarizes the results for instances where customers belonging to segment J^1 tolerate delivery delays of at most three time periods. It is interesting to note that the results provided by CPLEX are relatively insensitive to the magnitude of the tardiness penalty costs. However, the latter affect the extent to which late shipments occur, in particular in the MFLPDDSM, as shown in the last column of Table 13. As expected, lower tardiness costs result in a larger average percentage of the total demand that is covered with delay.

5.2.4. Managerial insights

To further gain insight into the characteristics of the optimal or near-optimal solutions identified by CPLEX, Table 14 reports the contribution of various cost components to the overall cost for each problem type and different values of the maximum allowed delivery delay. The results shown in the table refer to averages determined with respect to the best feasible solution available for each instance. Fixed facility cost rates are given separately for new and initially existing facilities according to (1). In addition, the average processing (o_{ik}^t), distribution (c_{ij}^t) and tardiness penalty ($p_j^{tt'}$) cost rates are also displayed. The last column of Table 14 presents the average percentage of total demand that is satisfied with delay.

For the test instances of the MFLPDDSM, the effect of enforcing timely demand fulfilment can be compared to the opposite

Table 13
Effect of tardiness penalty costs when the maximum delivery delay is equal to three periods ($\rho = 3$); *instances not solved to optimality within 10 hours.

Problem, model	Tardiness penalty costs	No. of opt. sol./ no. of non-opt. sol./ no. of out of memory	Avg gap (percent)		Avg CPU (seconds)		Percent delayed demand
			MIP*	LP	MIP	LP	
MFLPDDSM, ($\mathbf{P}_a^+$)	low	26/10/0	1.65	2.87	14386.28	6.10	18.0
	medium	26/10/0	1.44	2.85	14470.83	5.80	13.3
	high	27/9/0	1.46	2.83	13569.86	5.40	8.6
MFLPDDSM-S, ($\mathbf{Q}_a^+$)	low	1/34/1	2.54	4.04	35970.94	10.08	8.3
	medium	1/35/0	2.20	3.64	35128.52	9.79	4.6
	high	1/33/2	1.97	3.43	35684.30	9.04	3.1

Table 14
Average cost rates under different values of the maximum delivery delay.

Problem	ρ	Percent of total cost							Percent delayed
		New facilities		Existing facilities				Tardiness	Demand
		Opening cost	Maint. cost	Closing cost	Maint. cost	Processing cost	Distribution cost	penalty cost	
MFLPDDSM	0	14.3	25.5	0.7	11.3	16.4	31.7	–	–
	1	13.7	23.8	0.6	11.7	16.9	33.0	0.2	6.1
	2	13.8	23.9	0.6	11.3	16.9	33.0	0.4	8.8
	3	13.8	23.9	0.6	11.4	16.9	33.0	0.4	8.6
MFLPDDSM-S	1	13.8	24.2	0.6	11.8	16.8	32.6	0.1	3.3
	2	13.9	24.5	0.7	11.3	16.7	32.8	0.1	2.9
	3	13.8	24.0	0.6	11.7	16.7	32.9	0.2	3.1

Table 15
Capacity utilization rates and average number of new facilities and closed facilities.

Problem	ρ	Capacity usage (percent)						No. of facilities	
		New facilities			Existing facilities			New	Closed
		Min	Avg	Max	Min	Avg	Max		
MFLPDDSM	0	73.3	90.6	98.1	63.6	86.5	100.0	4.0	1.4
	1	87.4	96.0	100.0	73.7	92.7	100.0	3.7	1.3
	2	86.8	95.9	100.0	83.5	94.8	100.0	3.7	1.3
	3	86.8	95.8	100.0	83.5	94.8	100.0	3.7	1.3
MFLPDDSM-S	1	87.9	94.9	99.9	73.2	90.3	100.0	3.8	1.3
	2	88.9	95.1	99.9	77.4	91.1	100.0	3.8	1.4
	3	87.7	95.0	100.0	67.4	91.2	100.0	3.7	1.4

situation in which some customers accept late shipments. The former case ($\rho = 0$) calls for a larger investment on establishing new facilities, choosing appropriate capacity levels for the new locations and on closing initially existing facilities (see also columns 9 and 10 in Table 15). This investment arises from increasing capacity needs in order to be able to meet all customer demands without delays. In contrast, when a subset of the customers may experience delayed demand satisfaction, less capacity is installed which results in lower investment spending on facilities. Nevertheless, as shown in columns 3–6 of Table 14, this reduction is not substantial due to the negative impact of late deliveries. This can be seen by the low contribution of the tardiness penalty cost to the total cost (0.2–0.4 percent) and by the small percentage of delayed demand. Naturally, when the delivery time limit increases, also more demand is satisfied with some delay. The typical trade-off between the fixed facility costs and the variable processing and distribution costs is also illustrated in Table 14. The higher investment is partly abated by lower expenditures through selecting facilities that are less expensive and/or closer to the demand markets. Our results also indicate that the average total cost decreases as the maximum allowed delay increases. Compared to the classical case ($\rho = 0$), the cost reduction is on average 3.04 percent and 3.15 percent for $\rho = 1$ and $\rho = 2$, respectively. When the delivery delay grows to three periods, the average overall cost of a solution is only slightly smaller than for $\rho = 2$.

Regarding the MFLPDDSM-S, since late deliveries cannot be split over several time periods, fewer delays in demand satisfaction occur. However, the required adjustments in the network configuration do not yield substantially different fixed and variable cost rates from those in the MFLPDDSM. Observe that the single shipment case lies in between two extreme situations, namely one in which timely demand fulfilment is enforced and another for which the satisfaction of customer demand may be split over at most four periods (in the case of $\rho = 3$). It is also important to note that the results reported in Table 14 include many instances of the MFLPDDSM-S for which the optimal solution is not available.

From a managerial perspective, an additional important aspect to be investigated is the capacity utilization level at the facilities. This metric gives insight into the overall slack capacity and, therefore, it is an important indicator of whether it is possible to process larger product amounts without incurring the expensive costs of establishing new facilities and installing additional capacity. Table 15 provides information on this metric for both types of problems. Furthermore, it also presents the mean number of new

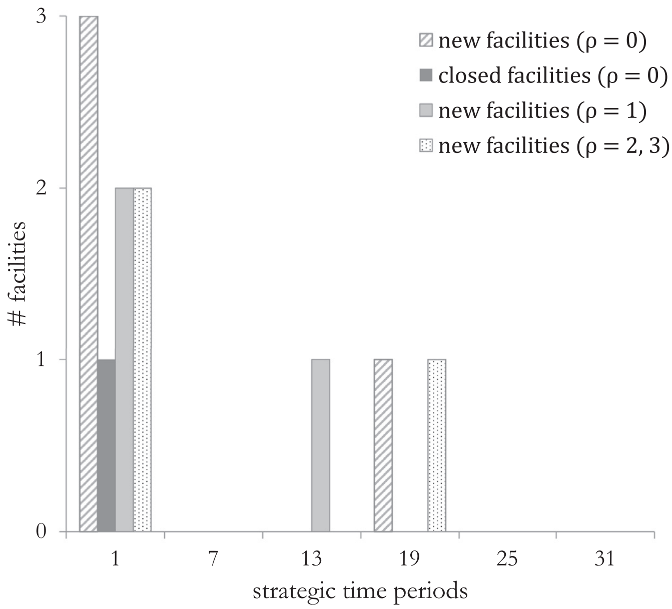

Fig. 1. Optimal location decisions in an instance of the MFLPDDSM under different maximum delivery delays.

facilities that are opened and the mean number of initially existing facilities that are closed over each set of instances. For the MFLPDDSM, the impact of a positive maximum allowed delivery delay is reflected in a higher capacity usage. This results from operating fewer facilities as highlighted by the last two columns of Table 15. Hence, by allowing delays in demand satisfaction cost benefits can be achieved (cf. Table 14) through the acquisition of the required capacity. As expected, this feature is also present in the MFLPDDSM-S but at a slightly lower level. In contrast, timely demand satisfaction ($\rho = 0$) yields more slack capacity in a larger network of facilities.

To further illustrate the impact of the maximum allowed delivery delay on location decisions, we selected a representative instance of the MFLPDDSM with 100 customers that was solved to optimality by CPLEX for each value of the parameter ρ. At the beginning of the planning horizon, two facilities are in service and eight candidate sites for establishing new facilities are available. When lateness in demand fulfilment is permitted (i.e. $\rho > 0$), the customers are equally distributed between the two segments (i.e. $|J^0| = |J^1| = 50$). Fig. 1 shows the number of facilities that are opened and closed in each of the six strategic periods over a 36-period horizon.

The optimal schedules of facility openings and closures displayed in the figure are in line with the findings presented in Tables 14 and 15. In particular, it can be seen that the enforcement of timely customer deliveries results in closing one existing facility at the end of the first time period and opening a total of four new facilities, three of them directly at $t = 1$ and the fourth one in the second half of the time horizon ($t = 19$). In contrast, when late shipments are allowed, all initially existing facilities are maintained and only three new facilities are established. This results in a lower investment in the reconfiguration of the facility system compared with the problem variant with $\rho = 0$. More specifically, the expenditures decrease by 6.8 percent and 9.2 percent for the case $\rho = 1$ and $\rho \in \{2, 3\}$, respectively. This reduction in investment spending comes at the price of satisfying 2.8 percent and 10.4 percent of the total demand with lateness, respectively.

It is also interesting to note that the decision to open one of the new facilities takes place earlier when $\rho = 1$. Finally, increasing the value of ρ from 2 to 3 does not affect the optimal location choices and the optimal solutions coincide. This result provides useful insight into the sensitivity of the system configuration to the maximum allowed delivery delay.

6. Conclusions

In this paper, we introduced an extension of the classical multi-period facility location problem by considering customer segments with distinct sensitivity to delivery lead times and by incorporating different time scales for strategic and tactical decisions into the time horizon. For each candidate site for locating a new facility, it is assumed that a set of discrete capacity levels is available. A variant of the problem was also studied in which customers accepting delayed demand satisfaction require late shipments to occur at most once over the delivery lead time. We proposed two mathematical programming formulations for each problem and developed additional inequalities to strengthen their linear relaxations. A theoretical comparison between the models without additional inequalities showed that they are equally strong in terms of the lower bounds provided by their linear relaxations. However, using randomly generated test instances, our computational experiments with a state-of-the-art MILP solver demonstrated the superiority of one of the formulations over the other. Furthermore, for medium-sized test instances, high quality solutions could be identified by the optimization solver in acceptable computing times. In our empirical study, additional insights were gained by analyzing several characteristics of the best solutions obtained. In particular, our analysis illustrated the far-reaching implications of the delivery lead time for customers accepting delayed shipments with respect to the overall cost and the capacity usage of the operating facilities.

Several further research opportunities can be identified. In view of our numerical results, a future line of research would be to develop a heuristic to find good quality solutions for large problem instances within reasonable computing times. This would be particularly important for the more difficult problem variant where single shipments are imposed. Another line of research could be directed toward the development of more comprehensive multi-period facility location models by including multiple commodities, considering several facility layers and allowing for the dynamic reconfiguration of the facility system (e.g. through capacity expansion and contraction, partial closing and reopening of facilities). These aspects play an important role in the context of supply chain network design. Furthermore, at the time facility location decisions must be made, many input parameters for long-term planning are inherently uncertain (e.g. costs, demand). Hence, focus could also be given to the design of a stochastic model to explicitly account for the uncertainty associated with future conditions.

Acknowledgments

This work was partially supported by FCT—Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (CMA/FCT/UNL). This support is gratefully acknowledged. The authors would like to thank four anonymous reviewers for their valuable comments and insights that helped improving the manuscript.

Appendix A. Theoretical result

We show that the LP-relaxations of formulations ($\mathbf{P}$) and ($\mathbf{P}_a$) provide the same lower bound.

Theorem. $v(\overline{\mathbf{P}}) = v(\overline{\mathbf{P}}_a)$

Proof. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ be a feasible solution to $(\overline{\mathbf{P}})$. Using relations (18)–(20), it is easy to construct a feasible solution to $(\overline{\mathbf{P}}_a)$.

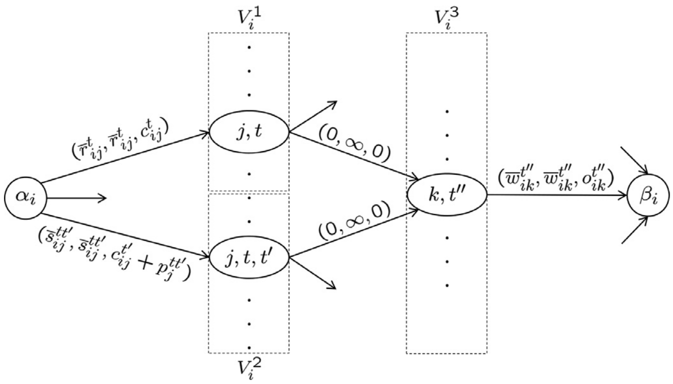

Fig. 2. General structure of the network $N_i = (V_i, E_i)$ associated with $i \in I$.

Moreover, both solutions have the same objective function value. Therefore, $v(\overline{\mathbf{P}}_a) \le v(\overline{\mathbf{P}})$.

Let us now consider any feasible solution $(\overline{\mathbf{r}}, \overline{\mathbf{s}}, \overline{\mathbf{w}}, \overline{\mathbf{z}})$ to $(\overline{\mathbf{P}}_a)$. We will show that $v(\overline{\mathbf{P}}) \le v(\overline{\mathbf{P}}_a)$. To this end, we describe next how this solution can be used to obtain a feasible solution to $(\overline{\mathbf{P}})$. In particular, the values of variables $\overline{x}_{ijk}^t$ $(j \in J^0; k \in K_i; t \in T)$ and $\overline{y}_{ijk}^{tt'}(j \in J^1;\ k \in K_i;\ t \in T,\ t' = t, \ldots, \min\{t + \rho_j, |T|\})$ will be determined by solving a sequence of *single commodity minimum cost flow problems*, one for each facility $i \in I$ in a directed multi-layer network $N_i = (V_i, E_i)$ defined by the set of vertices V_i and the set of arcs E_i $(E_i \subset V_i \times V_i)$ (see Fig. 2). V_i includes a source α_i, a sink β_i, and three types of transshipment vertices denoted by V_i^1, V_i^2 and V_i^3, distributed across two layers. For each arc $(u, v) \in E_i$, let $\underline{m}_{uv}$, resp. $\overline{m}_{uv}$, be the lower, resp. upper, capacity bound and let c'_{uv} be a non-negative cost per unit of flow.

The first layer in network N_i $(i \in I)$ consists of arcs pointing from the source α_i to transshipment vertices in V_i^1 and V_i^2. The latter are associated respectively with variables $\overline{\mathbf{r}}$ and $\overline{\mathbf{s}}$ as follows. For each variable $\overline{r}_{ij}^t$ taking a positive value, a vertex is considered which is denoted by the pair (j, t) in Fig. 2 $(j \in J^0; t \in T)$. If $\overline{r}_{ij}^t = 0$ for a given j and t then we set $\overline{x}_{ijk}^t = 0$ for every $k \in K_i$. Hence, there are at most $|J^0| \cdot |T|$ vertices in V_i^1 corresponding to variables $\overline{\mathbf{r}}$. Each arc $\alpha_i \to u$ $(u \in V_i^1)$ has capacity $\underline{m}_{\alpha_i u} = \overline{m}_{\alpha_i u} = \overline{r}_{ij}^t$ and unit cost $c'_{\alpha_i u} = c_{ij}^t$. Transshipment vertices associated with all positive variables $\overline{s}_{ij}^{tt'}$ make up the set V_i^2. Such vertices are displayed in Fig. 2 as triples (j, t, t') for $j \in J^1, t \in T$ and $t' = t, \ldots, \min\{t + \rho_j, |T|\}$. An arc from the source α_i to vertex $u \in V_i^2$ has capacity $\underline{m}_{\alpha_i u} = \overline{m}_{\alpha_i u} = \overline{s}_{ij}^{tt'}$ and unit cost $c'_{\alpha_i u} = c_{ij}^{t'} + p_j^{tt'}$. Again, if $\overline{s}_{ij}^{tt'} = 0$ for a given j, t and t' then we consider $\overline{y}_{ijk}^{tt'} = 0$ for every $k \in K_i$ and it is not necessary to include the corresponding vertex in V_i^2.

Regarding V_i^3, there are in total $|K_i| \cdot |T|$ such vertices that are depicted in Fig. 2 by the pairs (k, t'') with $k \in K_i$ and $t'' \in T$. As we will show later, they are related to variables $\overline{\mathbf{w}}$. From any vertex $(j, t) \in V_i^1$, we draw in total $|K_i|$ arcs, one to each vertex $(k, t'') \in V_i^3$ such that $k \in K_i$ and $t = t''$. These arcs are associated with timely deliveries to customers $j \in J^0$ in time period t from facility i with capacity size k. Furthermore, each vertex $(j, t, t') \in V_i^2$ is connected with $|K_i|$ vertices $(k, t'') \in V_i^3$ such that $k \in K_i$ and $t' = t''$. In this case, the arcs represent late-deliveries to customers $j \in J^1$ in time period t' from facility i with capacity level k. An arc $u \to v$ with $u \in V_i^1 \cup V_i^2$ and $v \in V_i^3$ has a lower bound capacity $\underline{m}_{uv} = 0$, an upper bound capacity $\overline{m}_{uv} = \infty$ and a cost per unit of flow $c'_{uv} = 0$ (see Fig. 2).

Finally, all vertices $v \in V_i^3$ are linked to the sink β_i. As shown in Fig. 2, each arc $v \to \beta_i$ with $v = (k, t'')$ has the following characteristics: $\underline{m}_{v\beta_i} = \overline{m}_{v\beta_i} = \overline{w}_{ik}^{t''}$ and $c'_{v\beta_i} = o_{ik}^{t''}$.

According to the structure of the network N_i, the total amount of flow generated by the source α_i for a given facility $i \in I$ is given by

$$b(\alpha_i) = \sum_{t \in T} \sum_{j \in J^0} \overline{r}_{ij}^t + \sum_{t \in T} \sum_{j \in J^1} \sum_{t'=t}^{\min\{t+\rho_j, |T|\}} \overline{s}_{ij}^{tt'}$$

whereas the sink β_i has the total inflow

$$b(\beta_i) = \sum_{t \in T} \sum_{k \in K_i} \overline{w}_{ik}^t$$

Since $(\overline{\mathbf{r}}, \overline{\mathbf{s}}, \overline{\mathbf{w}}, \overline{\mathbf{z}})$ is a feasible solution to $(\overline{\mathbf{P}}_a)$, constraints (22) hold, and thus $b(\alpha_i) = b(\beta_i)$. This means that each minimum cost network flow problem is feasible. The optimal solutions to these $|I|$ problems contain the values of the variables $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ of model $(\overline{\mathbf{P}})$. For each network associated with $i \in I$, the flow passing through the arc $u \to v$ such that $u = (j, t) \in V_i^1$ and $v = (k, t) \in V_i^3$ sets the value of variable $\overline{x}_{ijk}^t$. Moreover, the flow in the arc $u \to v$ with $u = (j, t, t') \in V_i^2$ and $v = (k, t') \in V_i^3$ sets the value of variable $\overline{y}_{ijk}^{tt'}$.

Due to the structure of each network N_i $(i \in I)$, it is now easy to establish a relationship between variables $\overline{\mathbf{r}}$ and $\overline{\mathbf{x}}$. For $u = (j, t) \in V_i^1$ it follows that $\overline{r}_{ij}^t = \sum_{k \in K_i} \overline{x}_{ijk}^t$. In a similar way, we can show that variables $\overline{\mathbf{s}}$ and $\overline{\mathbf{y}}$ are linked by considering the inflows and outflows of vertices V_i^2. For $v = (j, t, t') \in V_i^2$, it follows that $\overline{s}_{ij}^{tt'} = \sum_{k \in K_i} \overline{y}_{ijk}^{tt'}$. Finally, the relationship between variables $\overline{\mathbf{r}}$, $\overline{\mathbf{x}}$ and $\overline{\mathbf{w}}$ is determined by considering the inflows and outflows of vertices V_i^3. Observe that each vertex $v = (k, t'') \in V_i^3$ receives flow from vertices $u = (j, t) \in V_i^1$ for $j \in J^0$ and $t = t''$, and from vertices $u = (j, t, t') \in V_i^2$ such that $j \in J^1$, $t = \max\{1, t'' - \rho_j\}$ and $t' = t''$. Hence, the total inflow to vertex $v \in V_i^3$ is determined by $\sum_{j \in J^0} \overline{x}_{ijk}^{t''} + \sum_{j \in J^1} \sum_{t=\max\{1, t''-\rho_j\}}^{t''} \overline{y}_{ijk}^{tt''}$ and this equals the outflow,

Table 16
Average results for timely demand satisfaction ($\rho = 0$, $J^1 = \emptyset$, $J = J^0$); *instances not solved to optimality within 10 hours.

Formulation	$\lvert J\rvert$	$\lvert I^e\rvert$	$\lvert I^n\rvert$	No. of opt. sol./ no. of non-opt. sol.	Gap (percent) MIP*	Gap (percent) LP	CPU (seconds) MIP	CPU (seconds) LP
$(\mathbf{P})$	100	2	8	18/0		4.67	1061.59	1.10
	150	3	12	15/3	0.84	2.97	12980.54	2.58
All				**33/3**	**0.84**	**3.82**	**7021.07**	**1.84**
$(\mathbf{P}_a)$	100	2	8	18/0		4.67	808.00	0.47
	150	3	12	16/2	0.44	2.97	9106.08	1.14
All				**34/2**	**0.44**	**3.82**	**4957.04**	**0.80**
$(\mathbf{P}_a^+)$	100	2	8	18/0		3.16	939.19	0.98
	150	3	12	16/2	1.05	2.71	9884.91	2.06
All				**34/2**	**1.05**	**2.94**	**5412.05**	**1.52**

Please cite this article as: I. Correia, T. Melo, Multi-period capacitated facility location under delayed demand satisfaction, European Journal of Operational Research (2016), http://dx.doi.org/10.1016/j.ejor.2016.06.039

Table 17
Average results for $\rho_j = 1$ $(j \in J^1)$; *instances not solved to optimality within 10 hours.

Formulation	$\lvert J^0\rvert$	$\lvert J^1\rvert$	$\lvert I^e\rvert$	$\lvert I^n\rvert$	No. of opt. sol./ no. of non-opt. sol./ no. of out of memory	Gap (percent) MIP*	Gap (percent) LP	CPU (seconds) MIP	CPU (seconds) LP
(**P**)	25	75	2	8	5/1/0	0.26	3.89	11641.21	6.87
	50	50	2	8	5/1/0	1.42	3.76	8557.84	6.01
	75	25	2	8	6/0/0		3.24	1705.52	5.43
	38	112	3	12	1/5/0	1.60	2.21	30075.73	24.76
	75	75	3	12	3/3/0	0.98	2.22	20686.23	19.84
	113	37	3	12	3/3/0	1.05	2.27	21791.55	12.93
All					**23/13/0**	**1.21**	**2.93**	**15743.01**	**12.64**
($\mathbf{P}_a$)	25	75	2	8	6/0/0		3.89	4775.39	2.14
	50	50	2	8	6/0/0		3.76	5574.62	1.81
	75	25	2	8	6/0/0		3.24	927.03	1.42
	38	112	3	12	2/4/0	1.25	2.21	27258.47	5.65
	75	75	3	12	5/1/0	0.97	2.22	12374.46	4.53
	113	37	3	12	5/1/0	0.27	2.27	15179.65	3.55
All					**30/6/0**	**1.04**	**2.93**	**11014.94**	**3.18**
($\mathbf{P}_a^+$)	25	75	2	8	6/0/0		3.68	4879.97	1.91
	50	50	2	8	6/0/0		3.48	4316.44	1.89
	75	25	2	8	6/0/0		3.24	722.84	1.75
	38	112	3	12	2/4/0	1.36	2.21	27034.19	5.97
	75	75	3	12	5/1/0	1.12	2.22	12283.55	4.91
	113	37	3	12	5/1/0	0.78	2.25	15874.04	4.20
All					**30/6/0**	**1.22**	**2.85**	**10851.84**	**3.44**
($\mathbf{Q}_a^+$)	25	75	2	8	1/5/0	1.83	4.02	30030.58	3.10
	50	50	2	8	4/1/1	1.95	3.52	17554.63	2.90
	75	25	2	8	4/2/0	2.51	4.02	20126.19	2.39
	38	112	3	12	1/5/0	2.12	3.08	35973.54	8.31
	75	75	3	12	2/4/0	3.52	3.87	31876.65	6.61
	113	37	3	12	1/4/1	1.33	3.05	35566.66	4.66
All					**13/21/2**	**2.18**	**3.60**	**28636.71**	**4.71**

Table 18
Average results for $\rho_j = 2$ $(j \in J^1)$; *instances not solved to optimality within 10 hours.

Formulation	$\lvert J^0\rvert$	$\lvert J^1\rvert$	$\lvert I^e\rvert$	$\lvert I^n\rvert$	No. of opt. sol./ no. of non-opt. sol./ no. of out of memory	Gap (percent) MIP*	Gap (percent) LP	CPU (seconds) MIP	CPU (seconds) LP
(**P**)	25	75	2	8	5/1/0	1.42	3.89	13979.35	10.22
	50	50	2	8	5/1/0	1.67	3.62	9332.14	8.68
	75	25	2	8	6/0/0		3.05	1988.60	6.55
	38	112	3	12	2/4/0	2.15	2.35	27814.79	38.88
	75	75	3	12	3/3/0	1.18	2.14	22519.29	26.74
	113	37	3	12	3/3/0	1.36	2.03	19147.64	17.87
All					**24/12/0**	**1.61**	**2.85**	**15796.96**	**18.16**
($\mathbf{P}_a$)	25	75	2	8	6/0/0		3.89	7891.36	2.66
	50	50	2	8	6/0/0		3.62	4611.84	2.32
	75	25	2	8	6/0/0		3.05	789.88	1.73
	38	112	3	12	2/4/0	1.78	2.35	29229.80	8.50
	75	75	3	12	5/1/0	1.21	2.14	15588.77	6.10
	113	37	3	12	4/2/0	0.87	2.03	15743.65	4.27
All					**29/7/0**	**1.43**	**2.85**	**12309.21**	**4.26**
($\mathbf{P}_a^+$)	25	75	2	8	6/0/0		3.68	7667.03	2.61
	50	50	2	8	6/0/0		3.45	5805.24	2.32
	75	25	2	8	6/0/0		3.02	1165.21	2.00
	38	112	3	12	2/4/0	1.90	2.35	29621.76	8.81
	75	75	3	12	5/1/0	1.11	2.14	15018.36	6.46
	113	37	3	12	4/2/0	0.96	2.03	16468.54	4.70
All					**29/7/0**	**1.52**	**2.78**	**12624.36**	**4.48**
($\mathbf{Q}_a^+$)	25	75	2	8	1/5/0	3.05	4.71	30159.35	4.91
	50	50	2	8	2/4/0	2.00	3.77	25856.35	3.88
	75	25	2	8	0/5/1	1.29	4.18	36000.00	2.88
	38	112	3	12	0/6/0	2.54	3.38	36000.00	14.59
	75	75	3	12	0/6/0	1.23	2.50	36000.00	10.04
	113	37	3	12	0/6/0	1.16	2.46	36000.00	6.97
All					**3/32/1**	**1.85**	**3.48**	**33259.833**	**7.33**

Table 19
Average results for $\rho_j = 3$ ($j \in J^1$); *instances not solved to optimality within 10 hours.

Formulation	$\|J^0\|$	$\|J^1\|$	$\|I^e\|$	$\|I^n\|$	No. of opt. sol./ no. of non-opt. sol./ no. of out of memory	Gap (percent)		CPU (seconds)	
						MIP*	LP	MIP	LP
($\mathbf{P}$)	25	75	2	8	5/1/0	1.50	3.89	13750.74	13.66
	50	50	2	8	5/1/0	1.85	3.62	9721.33	10.71
	75	25	2	8	6/0/0		3.05	2035.32	7.94
	38	112	3	12	1/5/0	2.10	2.35	30117.90	52.31
	75	75	3	12	2/4/0	1.46	2.15	25282.64	33.46
	113	37	3	12	3/3/0	1.71	1.98	19669.38	20.28
All					**22/14/0**	**1.77**	**2.84**	**16762.89**	**23.06**
($\mathbf{P}_a$)	25	75	2	8	6/0/0		3.89	8805.16	3.33
	50	50	2	8	6/0/0		3.62	5781.67	2.75
	75	25	2	8	6/0/0		3.05	1299.46	1.97
	38	112	3	12	1/5/0	2.10	2.35	30027.29	10.45
	75	75	3	12	5/1/0	1.38	2.15	16940.32	7.52
	113	37	3	12	3/3/0	0.80	1.98	18479.75	5.01
All					**27/9/0**	**1.58**	**2.84**	**13555.61**	**5.17**
($\mathbf{P}_a^+$)	25	75	2	8	6/0/0		3.89	9078.29	3.21
	50	50	2	8	6/0/0		3.62	6432.30	2.79
	75	25	2	8	6/0/0		3.02	822.50	2.12
	38	112	3	12	1/5/0	1.84	2.35	30025.68	11.37
	75	75	3	12	5/1/0	1.38	2.15	16635.81	7.57
	113	37	3	12	3/3/0	0.84	1.98	18424.55	5.35
All					**27/9/0**	**1.46**	**2.83**	**13569.86**	**5.40**
($\mathbf{Q}_a^+$)	25	75	2	8	1/5/0	3.19	4.91	34211.02	5.60
	50	50	2	8	0/6/0	2.22	4.10	36000.00	4.69
	75	25	2	8	0/4/2	1.57	3.85	36000.00	3.32
	38	112	3	12	0/6/0	2.24	3.11	36000.00	17.76
	75	75	3	12	0/6/0	1.43	2.48	36000.00	12.60
	113	37	3	12	0/6/0	1.24	2.30	36000.00	8.34
All					**1/33/2**	**1.97**	**3.43**	**35684.30**	**9.04**

$\overline{w}_{ik}^{t''}$. We have just shown that relations (18)–(20) hold which justifies the feasibility of variables $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$. Regarding the location variables $\bar{\mathbf{z}}$, it is clear that their values in the solution to $(\overline{\mathbf{P}}_a)$ coincide with the values of the corresponding variables in $(\overline{\mathbf{P}})$.

We have shown how to construct a feasible solution to $(\overline{\mathbf{P}})$ from any feasible solution to $(\overline{\mathbf{P}}_a)$. It is straightforward to verify that both solutions have the same cost since the optimal flow in each network N_i $(i \in I)$ has total cost $\sum_{t\in T}\sum_{j\in J^0} c_{ij}^t \overline{r}_{ij}^t + \sum_{t\in T}\sum_{j\in J^1}\sum_{t'=t}^{\min\{t+\rho_j,\,|T|\}} (c_{ij}^{t'} + p_j^{tt'})\overline{s}_{ij}^{tt'} + \sum_{t\in T}\sum_{k\in K_i} o_{ik}^t\,\overline{w}_{ik}^t$ We conclude that $v(\overline{\mathbf{P}}) \le v(\overline{\mathbf{P}}_a)$.

We also draw attention to the fact that for a given feasible solution $(\bar{\mathbf{r}}, \bar{\mathbf{s}}, \overline{\mathbf{w}}, \bar{\mathbf{z}})$ to $(\overline{\mathbf{P}}_a)$, the system of linear equations (18)–(20) does not have a unique solution (observe that there are more unknowns than equations). Accordingly, there is not a one-to-one correspondence between the feasible solutions of problems $(\overline{\mathbf{P}})$ and $(\overline{\mathbf{P}}_a)$. □

Appendix B. Detailed results

Tables 16–19 report additional results for the various choices of the maximum allowed delay in demand fulfilment and for all formulations that were tested. In particular, the total number of customers in each segment ($|J^0|$, $|J^1|$), the number of initially existing facilities ($|I^e|$) and the number of candidate locations for new facilities ($|I^n|$) are shown. For $\rho = 0$ (Table 16), each row represents 18 instances, whereas for $\rho > 0$ (Tables 17–19), the results of six instances are summarized in each row.

References

Agatz, N., Fleischmann, M., & van Nunen, J. (2008). E-fulfillment and multi-channel distribution–A review. *European Journal of Operational Research, 187*, 339–356.

Albareda-Sambola, M., Alonso-Ayuso, A., Escudero, L., Fernandéz, E., Hinojosa, Y., & Pizarro-Romero, C. (2010). A computational comparison of several formulations for the multi-period incremental service facility location problem. *TOP, 18*, 62–80.

Albareda-Sambola, M., Fernandez, E., & Laporte, G. (2014). The dynamic multiperiod vehicle routing problem with probabilistic information. *Computers & Operations Research, 48*, 31–39.

Albareda-Sambola, M., Fernandez, E., & Nickel, S. (2012). Multiperiod location-routing with decoupled time scales. *European Journal of Operational Research, 217*, 248–258.

Altiparmak, F., Gen, M., Lin, L., & Paksoy, T. (2006). A genetic algorithm approach for multi-objective optimization of supply chain networks. *Computers & Industrial Engineering, 51*, 197–216.

Alumur, S., Kara, B., & Melo, M. (2015). Location and logistics. In G. Laporte, S. Nickel, & F. Saldanha da Gama (Eds.), *Location science* (pp. 419–441). Heidelberg: Springer.

Archetti, C., Jabali, O., & Speranza, M. (2015). Multi-period vehicle routing problem with due dates. *Computers & Operations Research, 61*, 122–134.

Athanasopoulos, T., & Minis, I. (2011). Multi-period routing in hybrid courier operations. In I. Minis, V. Zeimpekis, G. Dounias, & N. Ampazis (Eds.), *Supply chain optimization, design, and management: Advances and intelligent methods* (pp. 232–251). New York: Business Science Reference.

Badri, H., Bashiri, M., & Hejazi, T. (2013). Integrated strategic and tactical planning in a supply chain network design with a heuristic solution method. *Computers & Operations Research, 40*, 1143–1154.

Bashiri, M., Badri, H., & Talebi, J. (2012). A new approach to tactical and strategic planning in production–distribution networks. *Applied Mathematical Modelling, 36*, 1703–1717.

Canel, C., & Khumawala, B. (1996). A mixed-integer programming approach for the international facilities location problem. *International Journal of Operations & Production Management, 4*, 49–68.

Cheong, M., Bhatnagar, R., & Graves, S. (2005). Logistics network design with differentiated delivery lead-time: Benefits and insights. In *Proceedings of 2005 SMA conference, Singapore* (pp. 1–20).

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. *Computers & Industrial Engineering, 64*, 366–380.

Drexl, M., & Schneider, M. (2015). A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research, 241*, 283–308.

Fortz, B. (2015). Location problems in telecommunications. In G. Laporte, S. Nickel, & F. Saldanha da Gama (Eds.), *Location science* (pp. 537–554). Heidelberg: Springer.

Gebennini, E., Gamberini, R., & Manzini, R. (2009). An integrated production–distribution model for the dynamic location and allocation problem with safety stock optimization. *International Journal of Production Economics, 122*, 286–304.

Hung, H.-C., Chew, E., Lee, L., & Liu, S. (2012). Dynamic inventory rationing for systems with multiple demand classes and general demand processes. *International Journal of Production Economics, 139*, 351–358.

Jacobsen, S. (1990). Multiperiod capacitated location models. In P. Mirchandani, & R. Francis (Eds.), *Discrete location theory*. New York: Wiley.

Lieckens, K., & Vandaele, N. (2007). Reverse logistics network design with stochastic lead times. *Computers & Operations Research, 34*, 395–416.

Meisel, F., Rei, W., Gendreau, M., & Bierwirth, C. (2015). Designing supply networks under maximum customer order lead times. *IIE Transactions*. doi:10.1080/0740817X.2015.1110267.

Melo, M., Nickel, S., & Saldanha da Gama, F. (2009). Facility location and supply chain management—A review. *European Journal of Operational Research, 196*, 401–412.

Prodhon, C., & Prins, C. (2015). A survey of recent research on location-routing problems. *European Journal of Operational Research, 238*, 1–17.

Shen, Z.-J. (2006). A profit-maximizing supply chain network design model with demand choice flexibility. *Operations Research Letters, 34*, 673–682.

Sousa, R., Shah, N., & Papageorgiou, L. (2008). Supply chain design and multilevel planning—An industrial case. *Computers & Chemical Engineering, 32*, 2643–2663.

Wang, Y., Cohen, M., & Zheng, Y.-S. (2002). Differentiating customer service on the basis of delivery lead-times. *IIE Transactions, 34*, 979–989.

Wen, M., Cordeau, J.-F., Laporte, G., & Larsen, J. (2010). The dynamic multi-period vehicle routing problem. *Computers & Operations Research, 37*, 1615–1623.

Wilhelm, W., Han, X., & Lee, C. (2013). Computational comparison of two formulations for dynamic supply chain reconfiguration with capacity expansion and contraction. *Computers & Operations Research, 40*, 2340–2356.

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

Wu, Z., & Wu, J. (2015). Price discount and capacity planning under demand postponement with opaque selling. *Decision Support Systems, 76*, 24–34.