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

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:

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

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:

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

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

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:

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

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 $\ell \leq t$ 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)

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

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

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

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

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

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

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

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

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 $t' = t, \ldots, \min\{t+\rho_j,\ |T|\}$:

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