POLITECNICO DI TORINO
Repository ISTITUZIONALE

A bi-level approach for last-mile delivery with multiple satellites

Original
A bi-level approach for last-mile delivery with multiple satellites / Bruni, Maria Elena; Khodaparasti, Sara; Perboli, Guido. - In: TRANSPORTATION RESEARCH. PART C, EMERGING TECHNOLOGIES. - ISSN 0968-090X. - STAMPA. - 160:(2024), pp. 1-24. [10.1016/j.trc.2024.104495]

Availability:
This version is available at: 11583/2985833 since: 2024-02-19T10:39:48Z

Publisher:
Elsevier

Published
DOI:10.1016/j.trc.2024.104495

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

17 August 2024

Transportation Research Part C 160 (2024) 104495

Contents lists available at ScienceDirect

Transportation Research Part C

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

A bi-level approach for last-mile delivery with multiple satellites

Maria Elena Bruni [a,c,*], Sara Khodaparasti [b], Guido Perboli [b,c]

[a] Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende, 87036, Cosenza, Italy
[b] Department of Management and Production Engineering, Polytechnic University of Turin, Turin, 10129, Italy
[c] CIRRELT, Montreal, Canada

ARTICLE INFO

Keywords:
Last-mile delivery
Dynamic tariffs
Collaborative logistics
Bi-level optimization

ABSTRACT

Last-mile delivery is regarded as an essential, yet challenging problem in city logistics. One of the most common initiatives, implemented to streamline and support last-mile activities, are satellite depots. These intermediate logistics facilities are used by companies in urban areas to decouple last-mile activities from the rest of the distribution chain. Establishing a business model that considers different stakeholders' interests and balances the economic and operational dimensions, is still a challenge. In this paper, we introduce a novel problem that broadly covers such a setting, where the delivery to customers is managed through satellite depots. The interplay and the hierarchical relation between the problem agents are modeled in a bi-level framework. Two mathematical models and an exact solution approach, properly customized for our problem, are presented. To assess the validity of the proposed formulations and the efficiency of the solution approach, we conduct an extensive set of computational experiments on benchmark instances. In addition, we present managerial insights for a case study on parcel delivery in Turin, Italy.

1. Introduction

Urban distribution poses significant challenges in the field of logistics due to the increasing congestion effects and negative environmental impacts associated with it. The urgent need for a sustainable transition in the logistics sector further accentuates these challenges. Last-mile logistics plays a crucial role in this transition, not only because of its significant economic impact but also due to its socio-environmental implications. The projected global growth of last-mile deliveries by up to 78% by 2030 (World Economic Forum, 2020) will exacerbate these challenges in the near future. Therefore, it becomes imperative to explore new technologies and business models that can reshape last-mile logistics (Crainic et al., 2020, 2018). The development of successful strategies requires a clear vision to ensure high service levels, low delivery costs, and operational profitability.

To face these challenges, city logistics systems are implementing consolidation and coordination principles based on the collaboration among the different actors (Crainic et al., 2021). In particular, the involved agents are considering joining forces, by sharing facilities and services for last-mile delivery: the case of Hershey and Ferrero sharing warehousing facilities and transport services is only one example of collaborative supply chain operations in real-world logistics applications (Kane Logistics, 2011). Engagement in collaborative last-mile operations is not only limited to multi-billion big companies; local businesses striving to remain active in the competitive market may take the advantage of cooperative strategies by sharing city logistic facilities, called satellites. Envisioned benefits of the use of shared satellites include better vehicle capacity utilization, reduction of congestion and of the number of vehicles entering a city (van Heeswijk et al., 2019). Since these logistics initiatives are essentially business driven, a

* Corresponding author at: Department of Mechanical, Energy and Management Engineering, University of Calabria, Rende, 87036, Cosenza, Italy.
E-mail address: mariaelena.bruni@unical.it (M.E. Bruni).

https://doi.org/10.1016/j.trc.2024.104495
Received 18 March 2023; Received in revised form 2 January 2024; Accepted 23 January 2024
Available online 8 February 2024
0968-090X/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

business model should consider financial feasibility for both the owners/initiators (often city administrators) and the operators of the system. From the point of view of the initiator, to guarantee the continuation of the initiative, proper tariffs should be set (Janjevic and Ndiaye, 2017b,a) to make the satellite-based system competitive against non-consolidated express urban distribution services, reducing, as a byproduct, the congestion in the urban areas. The definition of tariffs for satellite infrastructure usage should align with the growing trend of implementing time and vehicle-dependent tariffs to regulate city access and congestion (Marciani and Cossu, 2014). The use of dynamic tariff definitions in particular has long been recognized as a crucial tool for promoting more sustainable behaviors among various stakeholders. The fundamental claim behind this is that, although the corporate social responsibility strategy of the carrier company also plays a significant role (Crotti and Maggi, 2022), tariffs ultimately influence carriers' decisions regarding satellite usage, since decisions are primarily based on pricing considerations (Kin et al., 2016; Isa et al., 2021). Current investigations into the potential of dynamic tariff definition for the successful implementation of satellite-based logistic solutions have largely overlooked the importance of collaboration and coordination among different stakeholders (Björklund and Johansson, 2018; Dreischerf and Buijs, 2022; Crainic et al., 2018). Considering different stakeholders poses a challenge since they have often different interests, goals, and needs (de Carvalho et al., 2019; Ballantyne et al., 2013).

This paper addresses this complex problem by:

- Presenting an integrated approach for setting the tariffs in a shared satellite-based last-mile delivery system that explicitly considers the hierarchical and complex relationship between two of the main stakeholders, the satellite-based infrastructure manager (SM) and the satellite operator (SO).
- Contributing to the literature presenting a bi-level problem, namely the Bi-level Last-Mile Delivery Problem with Multiple Satellites, to appropriately address the dynamism of pricing, and costing schemes, as well as operational issues of the system, considering the interests of both the stakeholders involved (Kaspi et al., 2022).
- Presenting two mathematical models and an exact solution approach, properly customized for the problem at hand. To the best of our knowledge, this is one of the very few contributions to bi-level last-mile delivery.
- Conducting an extensive set of computational experiments, using benchmark data sets, to qualify the model and to assess the computational tractability of the proposed model and the performance of the exact solution approach.
- Applying the model to a real case study of a medium-sized metropolitan area[1], and deriving managerial insights with respect to the structure of the tariffs and the fleet mix, given various urban distribution characteristics.

The remainder of the paper is organized as follows.

Section 2 provides a review of the relevant literature. Section 3 describes the problem and presents two mathematical formulations. Section 4 is devoted to the description of the exact solution approach. Section 5 discusses the numerical experiments conducted on a set of instances taken from the benchmark test set, appropriately modified to account for the characteristic of the problem. In addition, interesting policy-making and managerial insights are derived from the real case study on parcel delivery in the metropolitan area of Turin, Italy in Section 6. Finally, Section 7 summarizes the paper and presents some directions for future research.

2. Literature review

We analyze the literature along two axes. First, the literature on satellite-based multi-tier urban delivery systems is discussed. Second, the relevant literature on bi-level optimization in last-mile and urban delivery is reviewed.

From the point of view of *satellite-based multi-tier urban delivery systems*, the literature mainly focused on the family of problems known as two-echelon vehicle routing problems (Perboli et al., 2011; Crainic et al., 2009; Qiu et al., 2021). They have been extended in various forms but always consider the costs and the tariffs as given. Recently, a new direction in the literature used bin packing problems to tackle the operational and tactical issues in the management of a single satellite depot. In Perboli et al. (2021a), the joint problem of satellite management and urban delivery optimization has been addressed. The problem has been modeled as a variant of bin packing, which takes into account some specific features of the on-demand economy and e-commerce as, for instance, the time-dependent structure of the costs and the effects of customers' preferences. From a transportation perspective, the more considered issue is tactical capacity planning, arising in many contexts characterizing the new generation of multi-stakeholder systems, e.g., synchromodal (Qu et al., 2019; Perboli et al., 2017; Giusti et al., 2018) and physical-internet-based (Ballot et al., 2014; Saglietto, 2013; Skender et al., 2017), and city logistics (Crainic and Montreuil, 2016; Crainic et al., 2021). Nonetheless, there is a need to integrate the vision of the business models into optimization, in particular by providing tools for planning the use of satellite repositories and defining appropriate pricing (Crainic et al., 2018, 2021). These systems must respond to the needs of public stakeholders towards the use of vehicles with low environmental impact and vehicle congestion reduction, without hampering the correct execution of operations by delivery workers (Perboli et al., 2021a). In this context, the correct definition of appropriate dynamic tariff schemes over time is essential (Perboli et al., 2017; Marciani and Cossu, 2014; Musso et al., 2022).

From the point of view of *bi-level optimization in last-mile and urban delivery*, the literature is quite limited and mainly focused on bi-level location-routing models (Zhu and Ursavas, 2018). Following this stream, Xu et al. (2018) presented a bi-level programming model for a location-routing problem considering time window, vehicle capacity, and vehicle backhaul cost and proposed a genetic

[1] The problem setting comes from collaborations of the authors on urban distribution in the metropolitan area of Turin, Italy, as part of the development of the new Logistics and Mobility Plan of the Regional Government of the Piedmont region to be activated in 2025 (Perboli et al., 2021a, 2022)

algorithm solution approach. In another paper, Yang et al. (2020) presented a bi-level model to handle the location and demand distribution decisions arising in a parcel locker management problem where a delivery company as the upper level stakeholder minimizes the total cost associated with the locker construction, operation, transportation and parcel delivery to the lockers and the customers as the lower level decision makers minimize the pick-up cost. The model is also solved by a genetic algorithm. Santos et al. (2021) investigated an integrated inbound and outbound transportation planning problem in the realm of a bi-level vehicle routing problem with selective backhauls. The shipper is the leader: it plans the routes and proposes incentives to the carrier for performing integrated routes. The carrier is the follower: it can accept or refuse the offer and optimizes the routes visiting backhaul customers. Of course, the pricing notion in the latter research is on incentives offered by the shipper to the carrier for visiting backhaul customers, which is completely different from the satellite pricing plans addressed in the present paper. In addition, the aforementioned contribution is more focused on operational routing plans, and the tactical issues are not addressed. In Arrieta-Prieto et al. (2022), the authors model the interplay between the policy planners and the carriers in a bi-level framework. The public sector as the leader controls the location and demand coverage decisions for a set of uncapacitated urban micro-consolidation centers, promoting the use of sustainable delivery modes such as bikes and electric vehicles, and aims to minimize the social costs corresponding to the total emission which is, in turn, expressed in terms of the total distance traveled. On the other side, the private carriers control the transportation and fleet assignment decisions to minimize their operational delivery cost which is expressed as a linear function of the carrier tour length. Since the upper level and the lower level objective functions differ only on a multiplicative constant, the problem is reduced to a single-level problem that is solved by a greedy heuristic. This research is different from the present paper in many aspects. First of all, the pricing decisions are missing there; secondly, the time-dependent nature of the costs, which is a typical aspect of last-mile operations is not considered. For the sake of completeness, we also mention the work of Ji et al. (2017) studying a problem with an urban consolidation center operator: the urban consolidation center operator is the leader that sets the delivery time windows, whilst a third-party logistics follower delivers the orders to a set of retail stores by considering the time windows set in the upper level. The uncertainty on the supply and demand sides is tackled by adopting a risk-averse approach.

Concerning the specific problem of the service tariff definition, the literature mainly considered intermodal transportation, with a focus on long and medium-range transportation (Tawfik and Limbourg, 2018; Brotcorne et al., 2008; Tawfik and Limbourg, 2019; Hu et al., 2022). Bi-level optimization plays a central role in those pricing schemes, but their focus on long and medium-range transportation makes these models not usable in urban logistics. First, from a model point of view, they make use of simplified versions of service network design models. This is due to the rather small size of the instances in long and medium-distance transportation if compared to last-mile and urban logistics (Crainic et al., 2021). Second, due to the different problem settings of long and medium-distance transportation, they do not incorporate the specific constraints needed in last-mile and urban delivery. Thus, there is a general lack in terms of the integration of pricing plans into bi-level optimization models for last-mile and urban delivery.

It is evident how the literature lacks in terms of models and methods to guide the creation of time and vehicle-dependent tariffs in multi-tier urban systems. We provide the first answer to this need, by proposing a bi-level approach that considers both the multi-tier tariff fixing and the optimization of the operation at the satellites. An optimized tariff definition, in such complex systems, could help to achieve both high service quality, economic efficiency, and competitive advantages in the long run.

3. Problem description and model formulation

In two-tier systems, the delivery is based on the use of multiple transportation tiers, using specialized fleets and infrastructure. The first tier generally includes a set of terminals, known as consolidation distribution centers, used as the entry (exit) points to and from the city, where inbound (outbound) freight is consolidated. Lower tiers are connected to the first tier at transshipment facilities, called satellites, through urban trucks. Freight is then transshipped to city freighters, which deliver freight within the city. In systems with more than two tiers, the deliveries are performed through smaller facilities (e.g., mini hub and lockers) by drones, robots or bicycles. Fig. 1 shows the Social Business Network for a satellite-based multi-tier urban delivery system. The Social Business Network represents a complex system in a standard visual manner and is part of the GUEST methodology (Perboli, 2016; The GUEST Initiative, 2017). Different actors are involved in the system, including institutional authorities and city managers who establish rules and boundaries, customers and citizens who demand fast, affordable, and environmentally friendly services, carriers responsible for delivering the products, and facility and physical infrastructure managers who ensure efficient utilization of the infrastructure (Isa et al., 2021; Quak and Tavasszy, 2011; Crainic et al., 2018). The system exhibits complexity due to the interconnectedness, interactions, and interdependencies among these actors, as well as the frequent presence of conflicting goals (Perboli et al., 2021a, 2014).

Satellites are at the core of the system (Facility and Infrastructure Management). Satellites are transshipment facilities with no or low warehousing capabilities (Perboli et al., 2011; Crainic et al., 2009; Allen et al., 2012), operated by private or public/private partnerships (ULaaDS, 2020) that offer value-added logistics activities, such as packing, unpacking, and freight consolidation into environmentally friendly vehicles (Crainic et al., 2004). Concerning satellite management, most part of the literature focuses on the operational and routing part, disregarding the infrastructure managers' view.

In this paper, we particularly focus on the two main types of infrastructure managers in a two-tier urban delivery system: the SM and SO. The SM is in charge of operating the system, ensuring the economic sustainability of the service. The actual implementation of the deliveries is left to the SOs, often private logistic operators, that have to cope with the operational issues, including the integration of different delivery methods, as electric vehicles and cargo bikes (Crainic et al., 2021; Perboli et al., 2018) and whose aim is cost minimization (Pahwa and Jaller, 2022). The SM can be a private company or a public–private partnership in which

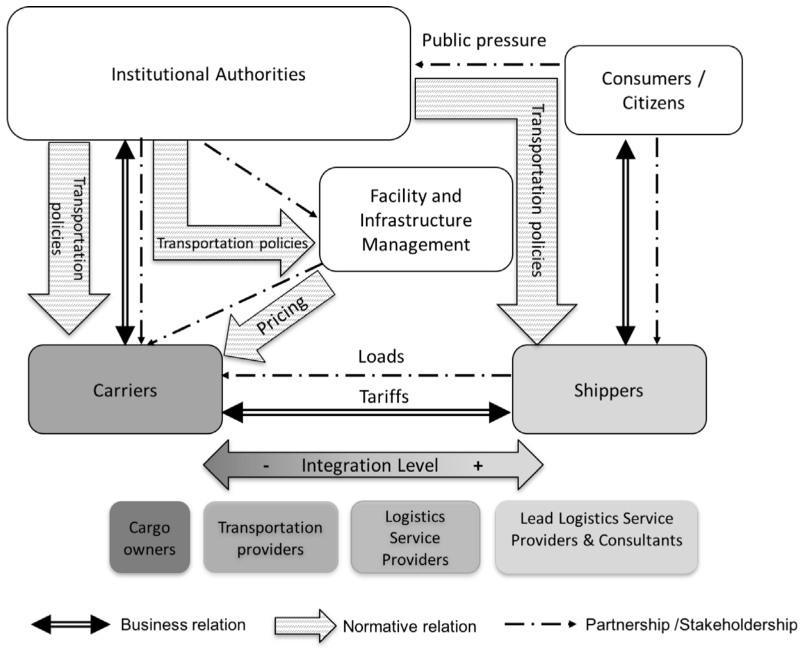

Fig. 1. Relationships among the main actors in freight transportation systems (Crainic et al., 2018).

local authorities and private stakeholders involved in delivery cooperate. This partnership/stakeholdership is often useful to enable better use of the transport infrastructure (local authorities might concede logistics spaces in strategic locations to set up satellites) or to financially support satellite-based initiatives. As a matter of fact, most projects that received public subsidies to operate, failed once subsidies terminated, for the lack of self-financing ability. The most important and frequent reason of non-viability is the presence of too high operating costs in comparison with the revenues coming from the SOs for the use of satellites. This problem is exacerbated when small numbers of users and, in turn, insufficient flows of goods are attracted. Hence, setting proper tariffs for using the satellites is an important task for the SM, crucial for the survivability of the system, since most carriers will not be willing to participate if the tariffs are high, preferring to outsource the deliveries by using external services (Alcaraz et al., 2019; Janjevic and Ndiaye, 2017a).

To fill the gap in the literature regarding the existence of models that explicitly consider the hierarchical and complex relationship between SM and SO, in the next Section we present a model framework, in the realm of bi-level optimization. Our model ensures the definition of a proper tariff for the SM, to cover the costs and reach the break-even point, ensuring at the same time a sufficient participation of SOs. Bi-level optimization approach is a suitable framework to address this problem with a hierarchical structure, since it accounts for the interplay and the interactions of different agents at two different levels.

3.1. A bi-level mathematical formulation

The origin of bi-level optimization problems dates back to the seminal works of von Stackelberg and Peacock (1952) on game theory and, in particular, leader–follower games. The leader, that has information about the follower's objectives, takes decisions first and communicates them to the follower, which reacts to the decision of the leader optimizing its own objective (Colson et al., 2007). As a result, the leader's optimization problem is a nested problem, where the feasible set is partly determined by a second optimization problem (Dempe et al., 2019). This approach has been successfully applied to tackle decision making problems with a hierarchical nature in various fields, such as supply chain, energy sector, transportation network design, revenue management, to mention a few (Kleinert et al., 2021).

Our bi-level formulation allows to explicitly model the hierarchy and the interaction between the SM and SO. The SM is the upper level decision maker interested to define time-dependent volume-based tariffs for satellites, with the aim of maximizing the total revenue to ensure the economic viability of the system. Clearly, the final gain depends on the response/reaction of the lower level system user, the SO, to the announced tariffs. The SO is responsible for the efficient delivery of a set of customers' demands, $d_i, i \in I$ within a planning horizon $\mathcal{H}$ using a limited and heterogeneous fleet of vehicles $\mathcal{K}$ dispatched from a set of existing satellite depots $\mathcal{J}$, each characterized by a time-dependent capacity $D_j^h, j \in \mathcal{J}, h \in \mathcal{H}$. Also, each vehicle type $p \in \mathcal{P}$ has a limited and time-dependent capacity V_p^h and if deployed, a vehicle usage cost is imposed. The vehicle costs δ_p^h associated to the use of the fleet depend on the vehicle type p an on the timeslot h. The total cost of the SO can be divided into three separate elements: the cost of the satellite infrastructure (usually volume-based), the delivery cost and the cost of the fleet. The delivery cost ϕ^h_{ipj} [2] depends on the relative distance between the satellite j and the order i, on the type of vehicle p used and on the timeslot h. Outsourcing the deliveries to external carriers is considered as an option available to the SO. In general, the cost of this service (E) is much higher than the ordinary delivery cost. We notice that this always ensures the feasibility of the SO's problem.

The SO should take the following decisions: *(i)* the selection of a subset of satellite depots from which the deliveries are performed, *(ii)* the allocation of vehicles to selected depots, and *(iii)* the delivery schedule. We assume that the SM and the SO cooperate, that is for any given decision of the SM, the SO is expected to choose the optimal solution that leads to the best objective function value for the leader. This is referred to as the "optimistic approach".

Before presenting the mathematical formulations, we first provide a brief discussion on tariff setting and we address some modeling issues. Let variable T^h_j represent the tariff assigned to satellite $j \in \mathcal{J}$ at timeslot $h \in \mathcal{H}$. Tariffs are time-varying for a number of reasons. Cheaper rates can be charged at certain times of day or night, with the aim of smoothing out the utilization of satellites. They can be also influenced by specific access restrictions set by the local authorities to limit traffic, noise and pollution in given areas of the cities during specific hours of the day. Hence, the tariffs are required to satisfy some "regulation constraints" expressed as lower and upper bounds ($\underline{T}^h_j \leq T^h_j \leq \overline{T}^h_j,\ \forall j \in \mathcal{J},\ h \in \mathcal{H}$). Moreover, the average tariff for each satellite over all timeslots should be below a pre-specified threshold ($\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}} T^h_j \leq \Delta_j,\ \forall j \in \mathcal{J}$).

In practice, the tariffs take discrete values that belong to a set of known and finite realizations as $\{\underline{T}^h_j,\ \underline{T}^h_j + s,\ \underline{T}^h_j + 2s,\ \ldots,\ \underline{T}^h_j + \left\lfloor \frac{\overline{T}^h_j - \underline{T}^h_j}{s} \right\rfloor s\}$ where s is the incremental rate. Therefore, we may express tariff T^h_j in terms of binary variables γ^{lh}_j as $T^h_j = \underline{T}^h_j + s\sum_{l\in\Lambda^h_j} l\,\gamma^{lh}_j$ where $\Lambda^h_j = \{0, 1, 2, \ldots, \left\lfloor \frac{\overline{T}^h_j - \underline{T}^h_j}{s} \right\rfloor\}$ and $\sum_{l\in\Lambda^h_j}\gamma^{lh}_j \leq 1\ \ \forall j \in \mathcal{J},\ h \in \mathcal{H}$.

It is clear that by construction, the regulatory constraints on the upper and lower bounds are always satisfied. We notice that the variables T^h_j play the role of auxiliary variables, being the choice of the tariff governed by the values assumed by the binary variables γ^{lh}_j.

3.2. First mathematical formulation

Let define three set of binary variables: x^h_{ikj} to represent if order i is delivered by vehicle k from satellite j at timeslot h; X_i to denote if customer i is served using the express delivery service and finally, y^h_{kj} to decide if vehicle k is used to deliver any orders from satellite depot j at timeslot h. By using the notation reported in Table 1, the Bi-level Last-Mile Delivery Problem with Multiple Satellites can be formulated as follows:

$$\text{(M1)}\quad \max_{\gamma, T, x, y, X} \sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} (d_i T^h_j) x^h_{ikj} \tag{1}$$

$$T^h_j = \underline{T}^h_j + s\sum_{l\in\Lambda^h_j} l\,\gamma^{lh}_j,\quad \forall j \in \mathcal{J},\ h \in \mathcal{H} \tag{2}$$

$$\frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}} T^h_j \leq \Delta_j,\ \forall j \in \mathcal{J} \tag{3}$$

$$\sum_{l\in\Lambda^h_j} \gamma^{lh}_j \leq 1,\quad \forall j \in \mathcal{J},\ h \in \mathcal{H} \tag{4}$$

$$\gamma^{lh}_j \in \{0,1\},\quad \forall j \in \mathcal{J},\ h \in \mathcal{H},\ l \in \Lambda^h_j \tag{5}$$

$$\begin{aligned}\min_{x, y, X} &\sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} (d_i T^h_j) x^h_{ikj} + \sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \phi^h_{ipj} x^h_{ikj}\\ &+ \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \delta^h_p\, y^h_{kj} + E\sum_{i\in I} X_i\end{aligned} \tag{6}$$

$$\sum_{i\in I}\sum_{k\in\mathcal{K}} d_i\, x^h_{ikj} \leq D^h_j,\ \forall j \in \mathcal{J},\ h \in \mathcal{H} \tag{7}$$

$$\sum_{i\in I} d_i x^h_{ikj} \leq V^h_p y^h_{kj},\ \forall j \in \mathcal{J},\ p \in \mathcal{P},\ k \in \mathcal{K}_p,\quad h \in \mathcal{H} \tag{8}$$

$$\sum_{j\in\mathcal{J}} y^h_{kj} \leq 1,\ \forall k \in \mathcal{K},\ h \in \mathcal{H} \tag{9}$$

$$\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} x^h_{ikj} + X_i = 1,\ \forall i \in I \tag{10}$$

[2] The SO delivery cost may also include environmental costs, congestion, etc. as generalized costs (Baldi et al., 2012).

Table 1
Notation for the mathematical model.

Sets	
$\mathcal{I}$	Set of orders indexed by i
$\mathcal{J}$	Set of satellite depots indexed by j
$\mathcal{H}$	Set of timeslots indexed by h
$\mathcal{K}$	Set of vehicles indexed by k
$\mathcal{P}$	Set of vehicle types indexed by p
$\mathcal{K}_p \subseteq \mathcal{K}$	Set of vehicles of type p where $\cup_{p\in\mathcal{P}}\mathcal{K}_p = \mathcal{K}$
$\Lambda_j^h = \{0, 1, \dots, \left\lfloor \frac{\overline{T}_j^h - \underline{T}_j^h}{s} \right\rfloor\}$	Set of tariff steps of satellite j at time slot h (indexed by l)
Parameters	
$\underline{T}_j^h$	Minimum tariff for satellite depot j at time slot h
$\overline{T}_j^h$	Maximum tariff for satellite depot j at timeslot h
Δ_j	Average tariff for satellite depot j within the day
s	Incremental rate for tariff
d_i	Demand associated to order i
ϕ_{ipj}^h	Cost-per-stop for vehicle of type p at timeslot h allocated to satellite depot j to deliver order i
δ_p^h	Usage cost for vehicle type p at timeslot h
E	Cost for the express delivery
D_j^h	Capacity of satellite j at timeslot h
V_p^h	Capacity of vehicle type p at timeslot h
Decision variables	
γ_j^{lh}	Binary variable which takes value 1 if the lth discrete tariff is set for satellite j at timeslot h and 0 otherwise; (upper level variable)
T_j^h	Auxiliary variable representing the tariff assigned to satellite j at timeslot h; (upper level variable)
x_{ikj}^h	Binary variable which takes value 1 if order i is delivered at timeslot h by vehicle k from satellite j and 0 otherwise; (lower level variable)
y_{kj}^h	Binary variable which takes value 1 if vehicle k is used to deliver the orders from satellite j at timeslot h and 0 otherwise (lower level variable)
X_i	Binary variable which takes value 1 if order i is delivered with express delivery option and 0 otherwise (lower level variable)

$$x_{ikj}^h \in \{0,1\}, \ \forall i \in I, \ j \in \mathcal{J}, \ k \in \mathcal{K}, \ h \in \mathcal{H} \tag{11}$$

$$y_{kj}^h \in \{0,1\}, \ \forall j \in \mathcal{J}, \ k \in \mathcal{K}, \ h \in \mathcal{H} \tag{12}$$

$$X_i \in \{0,1\}, \ \forall i \in I \tag{13}$$

The upper level objective function (1) expresses the SM's total revenue. Constraints in (2) set the tariff for each satellite at each timeslot. Constraints (3) are the "regulation constraints" that require the average of tariffs set for each satellite over all timeslots should be below a defined threshold. Constraints (4) are logical constraints ensuring that for each satellite j at each timeslot h, only one tariff is set. Finally, constraints (5) define the domain of the upper level variables. It is important to note that the SM can control only the variables related to tariffs while the delivery plan decisions are handled by the SO. The lower level problem corresponding to the SO is described by the objective function (6) and the set of constraints (7)–(13). The lower level objective function (6) represents the total cost in terms of satellite usage, the routing cost, the vehicle usage cost, and the cost for the express deliveries. Constraints (7) and (8) ensure that at each timeslot, the satellite depot capacity and the vehicle capacity are not exceeded, respectively. Constraints (9) are logical constraints ensuring that each vehicle at each timeslot can be sited in at most one satellite. Constraints (10) ensure that each order is delivered whether as an ordinary or express delivery. Finally, the set of constraints (11)–(13) show the nature of variables.

3.3. Second mathematical formulation

The lower level problem (6)–(13) is a multi-depot extension of the model presented in Perboli et al. (2021a) and, therefore, is NP-hard. The time-dependent structure of the objective function greatly increases the inherent complexity of the bin packing problem, which represents the core of the model. Moreover, the size of the model drastically grows with the increase in the number of orders and brings up serious computational tractability challenges.

To overcome this issue, we derive an aggregated formulation with a lower number of constraints and variables. With this aim, we assume that customers sharing similar characteristics, for example in terms of position location, are grouped into clusters (whose set is denoted by $\mathcal{C}$). We should note that the level of granularity of clusters can be increased to have as many clusters as the number of customers, thus leading back to the first model, and therefore, the aggregated formulation (M2) is a generalization of M1. In the computational results, we will show that even using coarse-grained clustering, the aggregated model is a good approximation of the non-aggregated one.

We introduce three sets of variables (as reported in Table 2): the aggregated amount of delivered demands in each cluster (from a given satellite, using a given vehicle type and in a given timeslot), the number of vehicles of each type dispatched from each satellite during each timeslot, and the total demands in each cluster delivered by the express delivery.

Table 2
Notation for the aggregated model.

Sets	
C	Set of clusters indexed by c
Parameters	
α_c	Average of demands in cluster c ($\alpha_c = \frac{\sum_{i\in c} d_i}{\lvert c\rvert}$)
f_{pj}^{ch}	Routing cost-per-stop for vehicle type p at timeslot h between satellite j and the customers in cluster c
β^c	Total amount of orders in cluster c ($\beta^c = \sum_{i\in c} d_i$)
Decision variables	
$q_{pj}^{ch} = \sum_{i\in c}\sum_{k\in\mathcal{K}_p} d_i x_{ikj}^h$	Aggregated amount of demands in cluster c delivered by vehicles type p from satellite j at timeslot h
$Y_{pj}^h = \sum_{k\in\mathcal{K}_p} y_{kj}^h$	Number of vehicles of type p dispatched from satellite j at timeslot h
$Q^c = \sum_{i\in c} d_i X_i$	Demand of cluster c delivered by the express delivery

Following the notation in Table 2, the aggregated model M2 can be written as follows:

$$\text{(M2)} \quad \max_{\gamma, T, q, Y, Q} \sum_{c\in C}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} T_j^h\, q_{pj}^{ch} \tag{14}$$

(2)–(5)

$$\min_{q,Y,Q} \sum_{c\in C}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} T_j^h\, q_{pj}^{ch} + \sum_{c\in C}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} f_{pj}^{ch} \frac{q_{pj}^{ch}}{\alpha_c} + \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \delta_p^h Y_{pj}^h + E\sum_{c\in C}\frac{Q^c}{\alpha_c} \tag{15}$$

$$\sum_{c\in C}\sum_{p\in\mathcal{P}} q_{pj}^{ch} \le D_j^h, \ \forall j\in\mathcal{J},\ h\in\mathcal{H} \tag{16}$$

$$\sum_{c\in C} q_{pj}^{ch} \le V_p^h Y_{pj}^h, \ \forall j\in\mathcal{J},\ p\in\mathcal{P},\ h\in\mathcal{H} \tag{17}$$

$$\sum_{j\in\mathcal{J}} Y_{pj}^h \le \lvert\mathcal{K}_p\rvert, \ \forall p\in\mathcal{P},\ h\in\mathcal{H} \tag{18}$$

$$\sum_{h\in\mathcal{H}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}} q_{pj}^{ch} + Q^c = \beta^c, \ \forall c\in C \tag{19}$$

$$q_{pj}^{ch} \ge 0, \ \forall c\in C,\ j\in\mathcal{J},\ p\in\mathcal{P},\ h\in\mathcal{H} \tag{20}$$

$$Q^c \ge 0, \ \forall c\in C \tag{21}$$

$$Y_{pj}^h \in \mathbb{Z}^+, \ \forall j\in\mathcal{J},\ p\in\mathcal{P},\ h\in\mathcal{H} \tag{22}$$

The upper level objective function (14) displays the total SM's revenue in terms of the aggregated deliveries. It is easy to see that the upper level objective functions (1) and (14) are equal and therefore the upper level problems in M1 and M2 are equivalent. The aggregated lower level problem corresponding to the SO is reformulated as follows: The lower level objective function (15) represents the aggregated total costs of the SO. While the satellite usage and the vehicle usage costs are equivalent in M1 and M2 (by the definition of the aggregated variables q_{pj}^{ch} and Y_{pj}^h), the routing and the express delivery costs in (15) are an approximation to their counterparts in M1. In fact, $\frac{q_{pj}^{ch}}{\alpha_c}$ approximates the number of deliveries in cluster c and $\frac{Q^c}{\alpha_c}$ is an estimation of the number of express deliveries in cluster c. Constraints (16) and (17) ensure that the restrictions on the capacity of satellite depots and vehicles are respected, respectively. Constraints (18) require that the total number of vehicles of each type to be deployed at each timeslot should be below the total number of existing vehicles of that type. Constraints (19) guarantee that the demand of each cluster should be delivered either with ordinary or express service. Finally, constraints in (20)–(22) show the nature of variables.

We assume without loss of generality that d_i are integer valued for all $i\in I$. Hence, variables q_{pj}^{ch} and Q^c are integer as well. The mathematical formulation M2 has $\lvert\mathcal{H}\rvert(\lvert\mathcal{K}\rvert-\lvert\mathcal{P}\rvert)(\lvert\mathcal{J}\rvert+1)+(\lvert I\rvert-\lvert C\rvert)$ less binary variables and $\lvert\mathcal{J}\rvert\,\lvert\mathcal{H}\rvert\left[\lvert\mathcal{K}\rvert(\lvert I\rvert+1)-\lvert\mathcal{P}\rvert(\lvert C\rvert+1)\right]+(\lvert I\rvert-\lvert C\rvert)$ less constraints, compared to M1. We further notice that leader's objective function contains bilinear terms resulting from the product of upper level and lower level variables. The same is true for the objective of the lower level problem, too, which, however, becomes linear once the upper-level decision variable values are set to a given value.

4. Solution approach

Despite the large amount of applications fitting the bi-level programming framework, real-life implementations are quite limited for the lack of efficient algorithms able to handle both the problem complexity and the large size of models. In general, bi-level problems are neither convex nor differentiable and even the simplest bi-level problem with linear upper and lower level models is strongly NP-hard (Hansen et al., 1992; Jeroslow, 1985). Therefore, most of the contributions to the solution methods for the bi-level models are based on the exploitation of the problem structure.

For instance, in case the lower level problem is convex and satisfies a suitable constraint qualification (which, in the convex case, usually is Slater's constraint qualification), a single-level reformulation can easily be derived using the KKT conditions or the strong duality theorem that replaces the lower level problem by a system of equations or inequalities (Dempe et al., 2015; Leyffer et al., 2006). The presence of integer variables in the follower's problem prevents the application of standard single-level reformulation techniques such as KKT. Hence, exact methods for mixed integer bi-level problems are based on the high-point relaxation model and bi-level infeasible solutions are discarded by branching, adding cutting planes, approximating the value function, or by a combination of the approaches above mentioned (Kleinert et al., 2021).

To exactly solve our bi-level models we present an exact solution approach adapted from the value-function-based approach developed by Lozano and Smith (2017) and tailored for the problems considered in this paper. To ease the description of the solution approach, we present a compact formulation for models M1 and M2. Since models M1 and M2 belong to the class of bi-level problems with mixed integer variables, the same is true for the compact formulation of the bi-level problem (BLP, for short)

$$\max_{(\mathbf{x}^l,\mathbf{x}^f)} \{Z^l(\mathbf{x}^l,\mathbf{x}^f) | \mathbf{x}^l \in \Omega^l, \mathbf{x}^f \in \Psi(\mathbf{x}^l)\} \tag{23}$$

where

$$\Psi(\mathbf{x}^l) = \underset{\mathbf{x}^f}{arg\,min}\{Z^f(\mathbf{x}^l,\mathbf{x}^f) | \mathbf{x}^f \in \Omega^f\} \tag{24}$$

$Z^l(\mathbf{x}^l,\mathbf{x}^f)$ denotes the leader's objective function and $Z^f(\mathbf{x}^l,\mathbf{x}^f)$ denotes the follower's objective function. $\Psi(\mathbf{x}^l)$ is the follower's *rational reaction set* corresponding to leader's solution $\mathbf{x}^l$. We also refer to (24) as the lower level problem.

Definition 1. A solution $(\mathbf{x}^l,\mathbf{x}^f)$ is called bi-level feasible if $\mathbf{x}^l \in \Omega^l$ and $\mathbf{x}^f \in \Psi(\mathbf{x}^l)$.

Lemma 1. *A solution $(\mathbf{x}^l,\mathbf{x}^f) \in \Omega$ is bi-level feasible iff $Z^f(\mathbf{x}^l,\mathbf{x}^f) \le Z^f(\mathbf{x}^l,\bar{\mathbf{x}}^f)$ for every $\bar{\mathbf{x}}^f \in \Omega^f$.*

Proof. The proof is straightforward by the definition of $\Psi(\mathbf{x}^l)$. ■

It is easy to recognize that M1 can be represented by setting $\mathbf{x}^f = (\{x^h_{ikj}\}, \{y^h_{kj}\}, \{X_i\})$, $\Omega = \{(\mathbf{x}^l,\mathbf{x}^f)|(2)–(5), (7)–(13)\}$, and $\Omega^f = \{\mathbf{x}^f|(7)–(13)\}$ where $Z^l(\mathbf{x}^l,\mathbf{x}^f)$ and $Z^f(\mathbf{x}^l,\mathbf{x}^f)$, respectively, denote the leader's and follower's objective functions (1) and (6).

For model M2, it is enough to set: $\Omega = \{(\mathbf{x}^l,\mathbf{x}^f)|(2)–(5), (16)–(22)\}$, where $\mathbf{x}^l = (\{\gamma^{lh}_j\})$ and $\mathbf{x}^f = (\{q^{ch}_{pj}\}, \{Y^h_{pj}\}, \{Q^c\})$ respectively, encode the vector of leader's and follower's decision variables. Moreover, we set $\Omega^l = \{\mathbf{x}^l|(2)–(5)\}$, and $\Omega^f = \{\mathbf{x}^f|(16)–(22)\}$ as the leader's and follower's decision spaces and $Z^l(\mathbf{x}^l,\mathbf{x}^f)$ and $Z^f(\mathbf{x}^l,\mathbf{x}^f)$, respectively, denote the leader's and follower's objective functions (14) and (15).

A key characteristic of this model is that the upper level variables do not influence the feasible set of the lower level problem, appearing instead as coefficients of the lower level objective. This implies that for any possible upper level solution, the remaining lower level problem has a finite optimal value, and the relatively complete response property holds for the BLP.

The bi-level problem in (23) can be expressed as a single-level optimization problem that is amenable to solution via a cutting-plane algorithm as follows.

$$\max_{(\mathbf{x}^l,\mathbf{x}^f)} \; Z^l(\mathbf{x}^l,\mathbf{x}^f) \tag{25}$$

$$\text{s.t.} \quad Z^f(\mathbf{x}^l,\mathbf{x}^f) \le Z^f(\mathbf{x}^l,\bar{\mathbf{x}}^f), \; \forall \bar{\mathbf{x}}^f \in \Omega^f \tag{26}$$

$$(\mathbf{x}^l,\mathbf{x}^f) \in \Omega \tag{27}$$

Problem (25)–(27) is called the extended high-point problem (EHPP).

Theorem 1. *The EHPP is equivalent to the BLP.*

Proof. Since EHPP and BLP share the same objective function, it is enough to illustrate that $S_{EHPP} = S_{BLP}$, where $S_{\cdot}$ denotes the feasible region.

To show $S_{EHPP} \subseteq S_{BLP}$, let $(\mathbf{x}^l,\mathbf{x}^f)$ be an arbitrary solution in S_{EHPP}. Therefore, $(\mathbf{x}^l,\mathbf{x}^f) \in \Omega$ (i.e., $\mathbf{x}^l \in \Omega^l$, $\mathbf{x}^f \in \Omega^f$). Since the constraints $Z^f(\mathbf{x}^l,\mathbf{x}^f) \le Z^f(\mathbf{x}^l,\bar{\mathbf{x}}^f)$, $\forall \bar{\mathbf{x}}^f \in \Omega^f$ enforce the bi-level feasibility, they are equivalent to $\mathbf{x}^f \in \Psi(\mathbf{x}^l)$ where $\mathbf{x}^l \in \Omega^l$. This means that $(\mathbf{x}^l,\mathbf{x}^f) \in S_{BLP}$.

To show that $S_{BLP} \subseteq S_{EHPP}$, let $(\mathbf{x}^l,\mathbf{x}^f)$ be an arbitrary solution in S_{BLP}, by definition we have, $\mathbf{x}^l \in \Omega^l$ and $\mathbf{x}^f \in \Psi(\mathbf{x}^l)$. The latter is equivalent to $Z^f(\mathbf{x}^l,\mathbf{x}^f) \le Z^f(\mathbf{x}^l,\bar{\mathbf{x}}^f)$ $\forall \bar{\mathbf{x}}^f \in \Omega^f$ where $(\mathbf{x}^l,\mathbf{x}^f) \in \Omega$. Hence, $(\mathbf{x}^l,\mathbf{x}^f) \in S_{EHPP}$.

From $S_{BLP} \subseteq S_{EHPP}$ and $S_{EHPP} \subseteq S_{BLP}$, we conclude $S_{EHPP} = S_{BLP}$ and the proof is complete. ■

Remark 1. The EHPP includes bilinear terms both in the objective function and in the constraints (26). To tackle this issue, we use McCormick's reformulation (McCormick, 1976), which in our case is exact, since the bilinear terms include binary variables γ^{lh}_j (Costa et al., 2017). The EHPP can be formulated as an integer model with linear objective function and constraints as presented in Appendix A.1. We remark that this reformulation is applied to expedite the solution process and neither changes the structure of the follower's rational reaction set nor invalidates Theorem 1. Clearly, a black-box solver for non-linear programs can also be applied to solve the non-linearized EHPP.

Theorem 1 implies that the BLP (23) can be solved to optimality by solving the EHPP. This, in turn, requires the enumeration of all feasible follower's responses $\bar{\mathbf{x}}^f \in \Omega^f$ that can be an overwhelming task in case Ω^f is an infinite set or its size is exponentially large. To overcome this drawback, we solve a relaxation of EHPP, named (REHPP), where Ω^f in (26) is replaced by $\hat{\Omega}^f \subseteq \Omega^f$, which includes a subset of sampled solutions.

Let $\hat{\Omega}^f \subseteq \Omega^f$ be a finite set which is specified by its elements $\bar{\mathbf{x}}_1^f, \bar{\mathbf{x}}_2^f, \dots, \bar{\mathbf{x}}_\kappa^f, \dots \bar{\mathbf{x}}_K^f$. The REHPP corresponding to $\hat{\Omega}^f$, denoted by REHPP($\hat{\Omega}^f$), is defined as

$$\max_{(\mathbf{x}^l, \mathbf{x}^f)} \; Z^l(\mathbf{x}^l, \mathbf{x}^f) \tag{28}$$

$$\text{s.t.} \;\; Z^f(\mathbf{x}^l, \mathbf{x}^f) \le Z^f(\mathbf{x}^l, \bar{\mathbf{x}}_\kappa^f), \;\; \forall \bar{\mathbf{x}}_\kappa^f \in \hat{\Omega}^f \tag{29}$$

$$(\mathbf{x}^l, \mathbf{x}^f) \in \Omega \tag{30}$$

Since $\hat{\Omega}^f \subseteq \Omega^f$, we have $S_{EHPP} \subseteq S_{REHPP}$, that implies $Z^{l*}_{EHPP} \le Z^{l*}_{REHPP}$. From the equivalency of EHPP and BLP, we conclude $Z^{l*}_{EHPP} = Z^{l*}_{BLP}$ which means that the optimal solution of REHPP provides a valid upper bound to BLP. Obviously, constraints (29) do not eliminate any bi-level feasible solution. The upper bound provided by the REHPP can be tightened by adding more cuts (29), progressively enlarging the set $\hat{\Omega}^f$.

This idea forms the core of an exact solution approach which iteratively solves the REHPP. By enlarging the current sample set and adding the corresponding cuts, tighter upper bounds are obtained. on the other hand, the response of the follower, corresponding to the optimal upper level variables in REHPP, is a bi-level feasible solution providing a lower bound.

The pseudocode of the exact approach is reported in Algorithm 1, where LB and UB refer to the BLP lower and upper bounds, respectively.

Algorithm 1: Pseudocode of the exact approach

1 **Initialization** ε, $\kappa \leftarrow 0$, $LB \leftarrow -\infty$, $UB \leftarrow \infty$, $\hat{\Omega}^f \leftarrow \emptyset$
2 Solve REHPP and obtain an optimal solution $(\bar{\mathbf{x}}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$
3 Obtain an optimal follower's response $\bar{\mathbf{x}}_\kappa^f \in \Psi(\bar{\mathbf{x}}_\kappa^l)$
4 $UB \leftarrow Z^l(\bar{\mathbf{x}}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$
5 $\hat{\Omega}^f \leftarrow \hat{\Omega}^f \cup \{\bar{\mathbf{x}}_\kappa^f\}$
6 **while** $((UB - LB) \ge \varepsilon)$ **do**
7     $\kappa \leftarrow \kappa + 1$
8   Solve REHPP($\hat{\Omega}^f$) and obtain an optimal solution $(\bar{\mathbf{x}}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$
9     $UB \leftarrow Z^l(\bar{\mathbf{x}}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$
10   Obtain an optimal follower's response $\bar{\mathbf{x}}_\kappa^f \in \Psi(\bar{\mathbf{x}}_\kappa^l)$
11     $\hat{\Omega}^f \leftarrow \hat{\Omega}^f \cup \{\bar{\mathbf{x}}_\kappa^f\}$
12   **if** $Z^f(\bar{\mathbf{x}}_\kappa^l, \bar{\mathbf{x}}_\kappa^f) = Z^f(\bar{\mathbf{x}}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$ **then**
13         $(\mathbf{x}^{*l}, \mathbf{x}^{*f}) \leftarrow (\mathbf{x}_\kappa^l, \hat{\mathbf{x}}_\kappa^f)$
14         $UB = LB$
15   **end**
16   **else if** $Z^l(\bar{\mathbf{x}}_\kappa^l, \bar{\mathbf{x}}_\kappa^f) > LB$ **then**
17         $LB \leftarrow Z^l(\bar{\mathbf{x}}_\kappa^l, \bar{\mathbf{x}}_\kappa^f)$
18         $(\mathbf{x}^{*l}, \mathbf{x}^{*f}) \leftarrow (\bar{\mathbf{x}}_\kappa^l, \bar{\mathbf{x}}_\kappa^f)$
19   **end**
20 **end**
21 **return** $(\mathbf{x}^{*l}, \mathbf{x}^{*f})$

The algorithm starts in Line 1 by initializing the counter κ, the sample set $\hat{\Omega}^f$, and the initial lower and upper bounds LB and UB. In Line 2, the REHPP is solved and the optimal values $\bar{\mathbf{x}}_\kappa^l$ of the upper level problem are stored. Fixing the upper level variables $\mathbf{x}^l$ to $\bar{\mathbf{x}}_\kappa^l$, in Line 3, the follower's problem is solved and the optimal follower's response $\bar{\mathbf{x}}_\kappa^f$ is obtained. The leader can either (i) select a feasible solution such that $Z^f(\mathbf{x}^l, \mathbf{x}^f) \le Z^f(\mathbf{x}^l, \bar{\mathbf{x}}_\kappa^f)$ or (ii) block the follower solution $\bar{\mathbf{x}}_\kappa^f$, by selecting an $\mathbf{x}^l$ such that $\bar{\mathbf{x}}_\kappa^f \notin \Omega^f$. In our case, for the relatively complete response strategy, for any upper level feasible solution, the follower's reaction $\bar{\mathbf{x}}_\kappa^f$ remains feasible. Since case (ii) is not possible the leader can only rely on option (i) which is nothing but the constraint (29).[3]

In Line 4 the upper bound UB is updated appropriately. The set of bi-level feasible solutions $\hat{\Omega}^f$ is updated in Line 5. In Lines 7–20, a loop is repeated until the upper and lower bounds match each other. First, the iteration counter is updated (Line 7). Then, the REHPP is solved considering the current bi-level feasible solutions in $\hat{\Omega}^f$ and the optimal values corresponding to upper level variables $\mathbf{x}^l$ are saved (Line 8). In Line 9, the upper bound is updated and, next, in Line 10 the follower's problem is solved for $\mathbf{x}^l$ fixed to $\bar{\mathbf{x}}_\kappa^l$. In Line 11, set $\hat{\Omega}^f$ is updated. In Lines 13–14, the incumbent and the upper bound are updated, if necessary. In Lines

[3] Note that Lozano and Smith (2017) introduce binary variables $w_{\bar{x}^f j} = 1$ if constraint j blocks solution $\bar{x}^f$ and since in our proposed models no constraint can block solution $\bar{x}^f$, the value of such variables is always set to zero. In this case, constraints (10b) of Lozano and Smith (2017) reduce to $g_j^2(x) \ge -M_j^1, \forall j = 1, \dots, m_2$ that are always valid because in our case $g_j^2(x) = 0$ and $M_j^1 \ge 0$.

17–18, the lower bound, if necessary, is updated together set $\hat{\Omega}^f$. Finally, the algorithm ends in Line 21 and the optimal solution is returned.

Remark 2. The leader's feasible region Ω^l corresponding to (BLP) (23) is a finite set and $|\Omega^l| \le |\mathcal{J}| \times |\mathcal{H}|$.

It is easy to see that all possible combination of values for binary variables $\{\gamma_j^{lh}\}$ satisfying constraints (4) is less than $|\Omega^l| \le |\mathcal{J}| \times |\mathcal{H}|$. Clearly, the cardinality of such feasible combinations satisfying simultaneously constraints (3) is also less than $|\mathcal{J}| \times |\mathcal{H}|$.

Theorem 2. *Algorithm 1 provides the optimal solution and converges in a finite number of iterations.*

The existence of the optimal solution for the BLP is guaranteed since the leader's and the follower's objective functions are continuous and the feasible regions Ω^l, Ω^f are compact sets for both M1 and M2. Now, it is enough to show that in at most $|\Omega^l|+1$ iterations, the condition in Line 12 holds, meaning that we have found an upper bound to the problem that is also a bi-level feasible solution (lower bound). Obviously, such solution is the optimal one.

Proof. Assume that Algorithm 1 finishes in a number of iterations less than or equal to $|\Omega^l|+1$. Since the number of leader's solutions is finite (see Remark 2), two indexes exist, let say κ_1 and κ_2, $1 \le \kappa_1 < \kappa_2 \le |\Omega^l|+1$, such that the corresponding upper level variables $\bar{\mathbf{x}}^l_{\kappa_1}$ and $\bar{\mathbf{x}}^l_{\kappa_2}$ are equal ($\bar{\mathbf{x}}^l_{\kappa_1} = \bar{\mathbf{x}}^l_{\kappa_2}$). At iteration κ_1, $\bar{\mathbf{x}}^f_{\kappa_1}$ is added to the set $\hat{\Omega}^f$. Hence, at iteration $\kappa_2 > \kappa_1$, $\hat{\Omega}^f$ already includes $\bar{\mathbf{x}}^f_{\kappa_1}$. At iteration κ_2 we solve REHPP($\hat{\Omega}^f$), obtaining the optimal solution $(\bar{\mathbf{x}}^l_{\kappa_2}, \hat{\mathbf{x}}^f_{\kappa_2})$, which is an upper bound for BLP. This solution satisfies constraints (29), hence $Z^f(\bar{\mathbf{x}}^l_{\kappa_2}, \hat{\mathbf{x}}^f_{\kappa_2}) \le Z^f(\bar{\mathbf{x}}^l_{\kappa_2}, \bar{\mathbf{x}}^f_{\kappa}) \forall \kappa = 1, \dots, \kappa_1, \dots, \kappa_2$. Since $\bar{\mathbf{x}}^l_{\kappa_1} = \bar{\mathbf{x}}^l_{\kappa_2}$, we can rewrite $Z^f(\bar{\mathbf{x}}^l_{\kappa_2}, \hat{\mathbf{x}}^f_{\kappa_2}) \le Z^f(\bar{\mathbf{x}}^l_{\kappa_2}, \bar{\mathbf{x}}^f_{\kappa_1})$ as $Z^f(\bar{\mathbf{x}}^l_{\kappa_1}, \hat{\mathbf{x}}^f_{\kappa_2}) \le Z^f(\bar{\mathbf{x}}^l_{\kappa_1}, \bar{\mathbf{x}}^f_{\kappa_1})$. We notice that $\bar{\mathbf{x}}^f_{\kappa_1}$ is the optimal reaction of the follower to the leader's solution $\bar{\mathbf{x}}^l_{\kappa_1}$. Then, $Z^f(\bar{\mathbf{x}}^l_{\kappa_1}, \hat{\mathbf{x}}^f_{\kappa_2})$ cannot be strictly less than $Z^f(\bar{\mathbf{x}}^l_{\kappa_1}, \bar{\mathbf{x}}^f_{\kappa_1})$. Therefore, the equality holds and $\hat{\mathbf{x}}^f_{\kappa_2}$ is also the optimal follower's reaction to the solution $\bar{\mathbf{x}}^l_{\kappa_2}$ and, as such, $(\bar{\mathbf{x}}^l_{\kappa_2}, \hat{\mathbf{x}}^f_{\kappa_2})$ is a bi-level feasible solution and hence, a lower bound for the BLP, which completes the proof. ■

4.1. An efficient procedure to obtain the follower's rational reaction

Finding an optimal follower's response in Lines 3 and 11 in Algorithm 1 requires solving the follower's problem (24) parameterized by the values of leader's decisions $\bar{\mathbf{x}}^l$. In case $\Psi(\bar{\mathbf{x}}^l)$ is not a singleton, following the "optimistic approach", the follower always responds in favor of the leader and selects a $\mathbf{x}^f \in \Psi(\bar{\mathbf{x}}^l)$ that maximizes the leader's revenue Z^l. To implement this approach, the following auxiliary problem should be solved:

$$\max_{\mathbf{x}^f} \; Z^l(\bar{\mathbf{x}}^l, \mathbf{x}^f) \tag{31}$$

$$\text{s.t.} \quad Z^f(\bar{\mathbf{x}}^l, \mathbf{x}^f) \le \Psi(\bar{\mathbf{x}}^l) \tag{32}$$

$$\mathbf{x}^f \in \Omega^f. \tag{33}$$

Solving problem (31)–(33) imposes an additional computational burden that may be efficiently handled to explore all the multiple alternative solutions of the follower's problem. This procedure is described as follows.

First, the follower's problem (24) corresponding to $\mathbf{x}^l = \bar{\mathbf{x}}^l$ is solved to optimality. Then, a no-good cut is added to the follower's problem (24) and the model, amended with the constraint which excludes the current optimal solution $\mathbf{x}^{*f} = (x_1^{*f}, x_2^{*f}, \dots, x_n^{*f})$ from solution space of the follower's problem, is solved again. The no-good cut

$$\sum_{i=1}^{n} |x_i^f - x_i^{*f}| \ge 1$$

can be rewritten as

$$z_i \le |x_i^f - x_i^{*f}|, \; \sum_{i=1}^{n} z_i \ge 1, \; \forall i = 1, \dots, n$$

and further reformulated as

$$z_i \le x_i^f - x_i^{*f} + M_i \, \delta_i, \; \forall i = 1, \dots, n \tag{34}$$

$$z_i \le -(x_i^f - x_i^{*f}) + M_i \, (1 - \delta_i), \; \forall i = 1, \dots, n \tag{35}$$

$$\sum_{i=1}^{n} z_i \ge 1 \tag{36}$$

$$\delta_i \in \{0, 1\}, \; \forall i = 1, \dots, n \tag{37}$$

$$z_i \in \mathbb{Z}^+, \; \forall i = 1, \dots, n \tag{38}$$

where M_i is a big-M value set as $M_i = x_i^{''f} - x_i^{'f}$, $i = 1, \dots, n$ and $x_i^{'f}, x_i^{''f}$ are such that $x_i^{'f} \le x_i^f \le x_i^{''f}$.

This provides another optimal solution, if any; otherwise, the optimal objective value deteriorates and the search ends. This process iteratively adds the no-good cuts one at a time until all the optimal solutions are found. Clearly, among the set of solutions found, the one that benefits the leader the most is chosen as the follower's rational response.

The customized procedure is described in Algorithm 2.

Algorithm 2: Pseudocode of the customized procedure

Obtain an optimal follower's response $\mathbf{x}^{*f} \in \Psi(\bar{\mathbf{x}}^l)$
$\bar{\mathbf{x}}^f \leftarrow \mathbf{x}^{*f}$
$\overline{Z}^l \leftarrow Z^l(\bar{\mathbf{x}}^l, \bar{\mathbf{x}}^f)$, $\overline{Z}^f \leftarrow Z^f(\bar{\mathbf{x}}^l, \bar{\mathbf{x}}^f)$
repeat
    Add cuts (34)-(38) to $\min_{\mathbf{x}^f \in \Omega^f} Z^f(\bar{\mathbf{x}}^l, \mathbf{x}^f)$ and get the optimal solution $\dot{\mathbf{x}}^f$
    if $Z^f(\bar{\mathbf{x}}^l, \dot{\mathbf{x}}^f) > \overline{Z}^f$ **then**
        break
    else
        $\overline{Z}^l \leftarrow Z^l(\bar{\mathbf{x}}^l, \dot{\mathbf{x}}^f)$
        $\bar{\mathbf{x}}^f \leftarrow \dot{\mathbf{x}}^f$
    end
return $(\bar{\mathbf{x}}^l, \bar{\mathbf{x}}^f)$

5. Computational experiments

In this section, we begin by providing comprehensive experimental findings conducted on a collection of benchmark instances in order to examine the effectiveness of the proposed models and solution approach. Furthermore, we present a practical case study pertaining to last-mile parcel delivery in the metropolitan area of Turin. This case study is derived from recent industrial and institutional collaborations between the authors, aimed at the development of the new Piedmont Logistics and Mobility Plan, which is scheduled to be implemented in 2025 (Perboli et al., 2021a, 2022).

5.1. Testing environment

The instances used in this paper are a multi-satellite extension of the instances presented in Perboli et al. (2021a) available in a BitBucked repository[4].

The number of orders to be delivered in one day (divided into five or three timeslots) varies in the set {200, 500, 1000}. Order volumes consider realistic settings in parcel delivery (De Marco et al., 2017; Perboli and Rosano, 2019) since small (with demand $d_i \in \{1, \dots, 15\}$) and medium orders (with demand $d_i \in \{16, \dots, 20\}$) are mixed in different percentages.

The fleet is composed by cargo bikes (with capacity of 100 kg in all the timeslots), electric vans (with a capacity of 150 kg in all the timeslots), and fossil-fueled vehicles (with a capacity of 200 kg in all the timeslots). Also, the case with a homogeneous fleet has been considered. The vehicle usage cost δ_p^h (Brotcorne et al., 2019) and the time-dependent cost-per-stop ϕ_{ipj}^h (Crainic et al., 2011) are reported in the dataset. The cost is made time-dependent through a vector [1.0, 0.3, 0.7] for three timeslots and [1.0, 0.1, 0.3, 0.5, 0.7] for five timeslots. The number of satellites for instances with 200, 500, and 1000 orders has been set to two, three, and four, respectively. The capacity of each satellite is equal to the capacity of a single depot case, as reported in the benchmark, divided by the number of satellites.

The lower and upper bounds for the satellite tariff are set to [30, 3, 9, 15, 21] and [45, 18, 24, 30, 36], respectively when five timeslots are considered and for the case of three timeslots, the bounds are set as [30, 9, 21] and [45, 24, 36].

All the experiments have been performed on a laptop with CPU Intel Core i7 with 2.60 GHz CPU and 16 GB RAM. The exact method has been coded in AIMMS 4.79.2.5, with Cplex 20.1.0 used as MIP solver. A time limit of 1800 s has been imposed on all the instances.

5.2. Models and exact solution approach performance analysis

In the first set of experiments, we perform a comparison between M1 and M2. The question that we want to address here is how much we lose in terms of the quality of the solution by solving M2, and if this is worth it from a computational viewpoint. Clearly, M2 has fewer variables and constraints compared to M1, on one hand. On the other hand, M2 provides an approximation of the optimal delivery plans of M1. To compare the performance of the proposed models in terms of solution quality and computational time, we fix the tariffs in the upper level model to the values reported in the benchmark instances. This provides a fair setting to compare the lower level problems in M1 and M2. The comparison is made on the set of instances with 200 orders and two satellite depots. For instances with more than 200 orders, the non-aggregated model M1 cannot be solved within a reasonable solution time while the size of aggregated model M2 is independent of the orders and can be efficiently solved. For the sake of brevity, Table 3 reports the summary of results (detailed results are reported in Appendix A.2) in terms of the average of CPU time ($CPU_{(.)}$) for M1 and M2, the average of relative percentage solution gap (Gap_M), where $Gap_M = \frac{|Z^{*f}_{M2} - Z^{f*}_{M1}|}{|Z^{f*}_{M1}|} 100$.

[4] https://bitbucket.org/orogroup/vcsbpp-td/src/master/.

Table 3
Summary of results: Comparison of M1 and M2.

Instance	$\|\mathcal{H}\|$	CPU_{M1} (s)	CPU_{M2} (s)	Gap_M (%)
200	3	71.23	0.03	0.86
200	5	131.83	0.03	0.86

Table 4
Optimal tariffs in the bi-level model versus fixed tariffs ($|\mathcal{J}| = 1$).

Instances	Gap_{Tariff}(%)			Gap_{CPU}(%)
	$h = 1$	$h = 2$	$h = 3$	
200	1.07	0.00	41.66	15.86
500	1.74	0.87	41.67	15.20
1000	1.91	0.00	40.63	9.46

Table 5
Summary of results: exact solution approach.

Instances	$\|\mathcal{H}\| = 3$				$\|\mathcal{H}\| = 5$			
	CPU_{REHPP} (s)	CPU_{LL} (s)	CPU (s)	Opt (%)	CPU_{REHPP} (s)	CPU_{LL} (s)	CPU (s)	Opt (%)
200	0.04	0.03	0.39	0	0.05	0.03	0.61	0
500	0.09	0.05	0.83	0	0.13	0.15	2.33	0
1000	0.31	0.62	3.29	0	7.31	0.04	205.12	11.43

The solution time of model M2 is quite stable while M1 is sensitive to the increase in timeslots as its CPU time increases by about 46%.

The values of Gap_M are quite small confirming the validity of model M2.

To investigate the quality of the optimal tariffs provided by the bi-level model (M2), we compared them with the suggested tariffs as reported in the benchmark (Municipality of Turin, 2019; Perboli et al., 2018). In fact, this gives us insights on the performance of the bi-level versus the single-level approach. Table 4 shows the summary of the results for instances up to 1000 orders and one satellite where Gap_{Tariff} refers to the average relative gap between the optimal tariff and the tariffs in the data set for each timeslot, Gap_{CPU} denotes the average relative gap in the solution time of bi-level and single-level models. As we can see, the optimal tariffs are 41.67% higher than the real tariffs for distribution of the e-commerce parcels in urban area (Perboli and Rosano, 2019). However, this superiority comes at the price of an increase in the solution time which is always below 15.86%. Of course, such an increase in CPU time is quite affordable since a bi-level model is clearly more complicated.

Table 5 reports the summary of the computational results gathered by applying Algorithm 1 on the aggregated model M2 (the detailed results are reported in Appendix A.2). In particular, we report the average values corresponding to the CPU time in the REHPP and the LL problem, the total CPU time, and the relative gap of the best upper and lower bounds, denoted by Opt ($Opt = \frac{|UB-LB|}{|LB|} 100$). The solution approach provides the optimal solution for all the instances but four instances with 1000 orders and five timeslots for which the average relative gap is around 11.43%. All the instances with three timeslots are solved to optimality and the average CPU time is below four seconds. For the instances with 1000 orders and five timeslots, the average CPU time is less than four minutes.

To investigate the effect of the numbers of clusters, we ran a set of experiments for instances with 1000 orders, four satellite depots, three timeslots, and a number of clusters ranging in the set $\{1, 2, 5, 10, 20\}$. Fig. 2 shows the average CPU time for such test cases. As expected, the solution time increases considerably with a higher number of clusters: for example, going from one to 20 clusters, the average solution time increases 18 times but it is still affordable and around 60 s. It is noticeable that the optimal objective function value remains the same.

6. Case study and managerial insights

Overall, dynamic tariff definition is considered attractive for the SM. One perceived positive implication of this framework is the control it offers over the utilization of satellites in terms of time and space. However, the impact of dynamic tariffs on the overall system is influenced not only by the economic attractiveness of the tariffs themselves but also by the decision-making processes of the SO and the preferences of the customers. In particular, the decisions made by the SO are influenced by its knowledge and understanding of the demand and the tariffs being offered, as well as the available fleet of satellites that the SO has at its disposal. These factors play a significant role in shaping the outcomes of the dynamic tariff system.

Based on these observations, several research questions arise. Firstly, do dynamic tariffs ultimately have a positive impact on the efficiency of the distribution system? Moreover, do dynamic tariffs contribute to the reduction of non-consolidated distribution services, such as those provided by express couriers? Secondly, what role does the composition of the satellite fleet play in this regard? Finally, what is the monetary value of the collaboration facilitated by dynamic tariffs?

In this paper, we aim to answer these research questions by presenting a real case study focused on the metropolitan area of Turin, Italy. In order to utilize the data from e-commerce and parcel delivery companies, it was necessary to anonymize and normalize

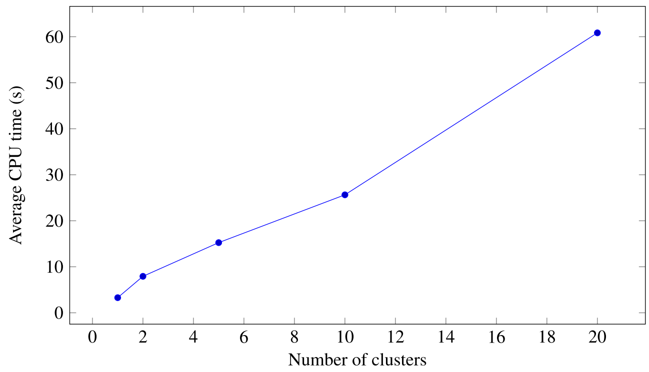

Fig. 2. Solution time versus the number of clusters.

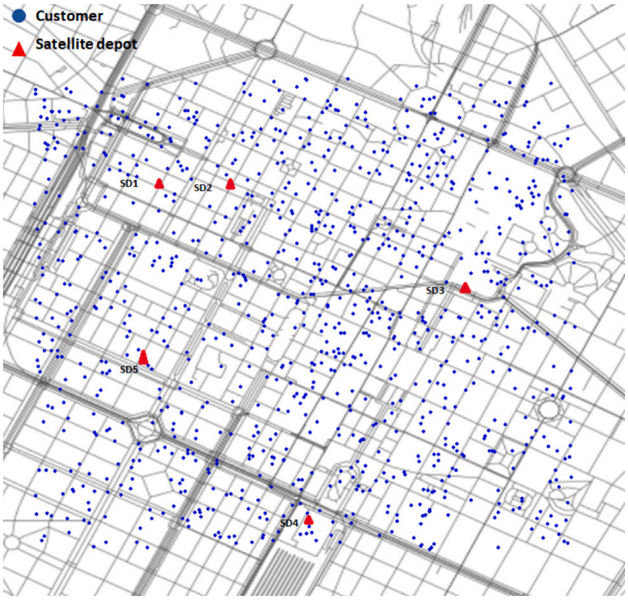

Fig. 3. Service area in the case study.

the data and process several data sources (e.g., city maps, travel times, and user preferences). This process was carried out using the data-fusion tool presented in the work of Perboli et al. (2018) (Perboli et al., 2018). The case study concerns a real last-mile delivery application in the Turin city center area (2.805×2.447 km^2). Fig. 3 displays the service area, the location of customers (blue circles) and the satellite depots (red triangles). More details on the case study and input parameters are reported in Appendix A.3.

Customers' orders (1000, in the case study) should be serviced within a day (from 9 AM to 12 PM), divided into three-hour timeslots. A heterogeneous fleet composed of three types of vehicles – cargo-bikes (CBs), vans (VANs), and electric vehicles (EVs) – is available for deliveries. Each customer order should be delivered within a time window, specified by the timeslots in which the customer prefers to receive the order. We have considered three different operational scenarios. As a baseline, the first scenario considers the case of a hard single-period time window. In the second scenario, customers place orders to be delivered in a few consecutive timeslots. This scenario is referred to as the flexible configuration. Finally, in the last scenario (free delivery time scenario), customers' preferences are not considered. Customers are divided into clusters on the basis of spatial and temporal characteristics. We notice that clustering the customers also on the basis of the time windows enables us to implicitly account for the time window constraints in an efficient way. In fact, it is sufficient to prevent the delivery of orders outside the time window. Therefore, we impose the following restrictions:

$\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}} q_{pj}^{ch} = 0,\ \forall c \in \mathcal{C},\ h \in \mathcal{H} - \{h^c\}$ where $\{h^c\}$ denotes the subset of timeslots in the time window associated to orders in cluster c. For each cluster, the cost-per-stop f_{pj}^{ch} has been evaluated through the Daganzo's formula (Daganzo, 1984) widely used to estimate routing costs in real applications (Franceschetti et al., 2017).

To evaluate the potential impact of dynamic tariffs on the efficiency of the distribution system, we first analyze the solution provided by model M2 under the baseline scenario. Table 6 reports the optimal solution in terms of number of deployed vehicles of

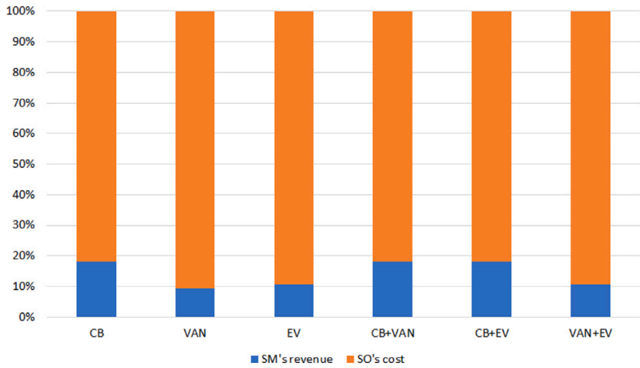

Fig. 4. SM's revenue and SO's cost under different fleet configurations.

Table 6
Results: Baseline scenario.

	# of deployed vehicles					Average vehicle fill ratio					Delivery ratio				
	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5
CB	8	7	7	8	15	0.95	0.93	0.94	0.95	1.00	1.00	1.00	1.00	1.00	0.75
EV	–	–	–	–	5	–	–	–	–	0.99	–	–	–	–	0.25

Table 7
Utilization of satellites in different timeslots.

Satellite	Satellite fill ratio				
Id	h = 1	h = 2	h = 3	h = 4	h = 5
SD1	0.10	–	–	–	–
SD2	–	–	–	–	0.52
SD3	–	–	–	0.10	0.94
SD4	–	0.68	–	–	0.63
SD5	0.66	–	0.69	0.66	–

each type, average vehicle fill ratio, and delivery ratio evaluated as the ratio between the demand serviced at that specific timeslot and the total demand[5].

We observe that the vehicle fill ratios are always above 93% of the vehicle capacity and in the last timeslot, in which the demands are the highest (about 42% of the total demands), the vehicle fill ratio increases to 100% and 99% for CBs and EVs, respectively. This efficient use of the vehicle capacity leads to a reduction of the number of empty trips, enhancing the liveability of the city and mitigating the negative impact (in terms of costs, congestion, noise, etc.) associated with the logistic activities. Only CBs and EVs are used. In particular, all the orders within the first four timeslots are handled by CBs and only in the fifth timeslot, the EVs are deployed delivering 25% of the orders. This is reasonable since the cost-per-stop of CBs and EVs are lower than the cost of VANs (see Table 16 in the Appendix).

We notice also that the utilization of satellites throughout the day is quite high, and that in particular, SD5 is heavily congested (see Table 7) during the rush hours. The operations nearby this satellite increases traffic congestion, producing unwanted externalities for the neighboring area. To address this issue, the public authority can implement policies to make the satellite less attractive to the delivery companies, especially in specific timeslots, in order to alleviate the traffic load around the area. For instance, the SM can raise the minimum tariff rates in the first three timeslots by 30%. In the new solution, the choice of satellite depots is different since the deliveries are shifted from SD5 to SD4 (see Table 8). More restrictive policies can directly impose some traffic rules restricting the drivers' access to some highly congested areas.

Fig. 4 shows the impact of different fleet configurations on the SM's revenue and SO's cost. If we only consider fleets composed of EVs and VANs, we observe an increase in the delivery cost of 6.67% and a considerable decrease in the SM's revenue (around 74.56%). The use of homogeneous fleets, with only CBs, VANs, or EVs, increases the delivery cost up to 1.44%, 11.20%, 8.36%, respectively. This situation is also inconvenient for the SM since its revenue is 92.58%, 70.57% lower in case of VANs and EVs, respectively (except for CBs where the SM's revenue increases by 1.32%).

As a result, we can conclude that it is beneficial to integrate CBs into the logistic system and use them alongside EVs to distribute last-mile deliveries. An efficient integration of CBs into city logistics (Perboli and Rosano, 2019) should become a priority for urban planners and city managers, since undoubtedly, it is a smart eco-friendly strategy, contributing to a more sustainable last-mile

[5] Tables 17–18 in Appendix A.3 report the results for other scenarios.

Table 8
Utilization of satellites in different timeslots under the congestion scenario.

Satellite	Satellite fill ratio				
Id	h = 1	h = 2	h = 3	h = 4	h = 5
SD1	0.73	–	–	–	0.52
SD2	0.03	–	–	0.76	0.94
SD3	–	–	–	–	0.63
SD4	–	0.68	0.69	–	–

Table 9
Results: express delivery and tariff.

Express delivery cost (per parcel)	Ordinary delivery cost (per parcel)	% of express deliveries (per parcel)	Ratio% (per parcel)
3.50	1.52	0	2.30
3.15	1.52	0	2.07
2.80	1.52	0	1.84
2.45	1.52	0	1.61
2.10	1.52	0	1.38
1.75*	1.48	2.60	1.18
1.40	1.47	4.20	0.95
1.05	1.36	23.20	0.77
0.70	1.30	41.60	0.54

delivery system. Nevertheless, still there is a lack of regulation for CBs. Excessive use of CBs could even decrease the sustainability of the city from the viewpoint of citizens, cyclists, and pedestrians. Clearly, any successful policy measure should be adapted to the specific characteristics and needs of the local context.

In the following, we consider the effect of the customers' preferences, by analyzing the two other scenarios already mentioned.

First we comment on the results for the second scenario with flexible time windows (see Table 17 in Appendix A.3). We observe that increasing the flexibility is advantageous for the SO, because it may allow consolidations that were previously impossible and can result in reduced delivery costs (by 19.36%). However, the leader revenue decreases by about 16.69%, given that all the deliveries are scheduled in the cheapest timeslots. Even though this solution can reduce the congestion in rush periods, shifting the last-mile deliveries in the early morning and in the late evening, it would not necessarily accepted by the SM, unfairly penalized. A third actor must be found who sees the potential of running such a business model. In particular, this motivates the involvement of public authorities in the management of satellites. This has been already observed in Crainic et al. (2004), where the authors argued that public–private collaboration in freight activities can increase the efficiency of the system.

Under the third scenario with free delivery time (see Table 18 in Appendix A.3), when customers do not express preferences, the SO can arrange all the deliveries on the fourth and the fifth timeslots in which both the cost-per-stop and the vehicle usage costs are the lowest, leading to a total delivery cost decrease (with respect to the baseline scenario) by 39.06%. VANs are heavily used in the free delivery time scenario, due to their larger capacity and the increased efficiency in this case. We remark that this solution is only viable in traditional offer-driven logistics, where customer's preferences and the quality of service are not primary goals. In general, the more flexibility is given to the customers, the more EVs and CBs become efficient options (Simoni et al., 2020; Perboli et al., 2021a).

It is beneficial to remark that a proper definition of tariffs for the usage of the satellite infrastructure guarantees the continuation of the initiative, safeguarding competitiveness against non-consolidated express urban distribution services. The ultimate effect is the reduction of the congestion in the urban areas, thanks to the consolidation effect achieved through the use of shared satellites. To support this claim, we investigate the sensitivity of the solution with respect to the variations in the express delivery cost (Brotcorne et al., 2019). Column 1 in Table 9 displays the express delivery cost (per parcel); column 2 the ordinary delivery cost (per parcel); column 3 the percentage of orders delivered using the express service; finally, column 4 the ratio between express and ordinary delivery costs (in percentage). We observe that for $E > 1.75$ €, the express delivery option is not selected at all since it is at least 1.38 times more costly than the ordinary delivery (see ratios in column 4, varying in the range of $[2.30, 1.38]$). Hence, the SO performs all the deliveries by the ordinary delivery service. With the decrease of express cost to $E = 1.75$ €, the two delivery options become quite comparable and therefore, 2.60% of orders are delivered using the express service. When the express delivery cost decreases to $E = 0.70$ €, thanks to the collaborative setting implemented by the bi-level model, the SM reduces satellite tariffs by about 30%, to make again competitive the deliveries made by the SO, compared to the express delivery option. In this case, the SO delivers about 58.40% of orders. From a policy maker's viewpoint, the external service will work properly and the war of price is prohibited. Even in this situation, the payback of investments in the satellite depot is also assured. In fact, considering for each satellite depot a setup cost of 2,500,000 € and a staff of two workers, with an average salary of 2000 € per person, the costs will be regained in 1250 days (3–4 years).

6.1. The benefits of collaboration

In this section, we assess the benefits of the collaboration. Collaboration can be vertical or horizontal. Vertical collaboration (Santos et al., 2021) is intended as the alliance between decision-makers at different levels, whilst horizontal collaboration addresses

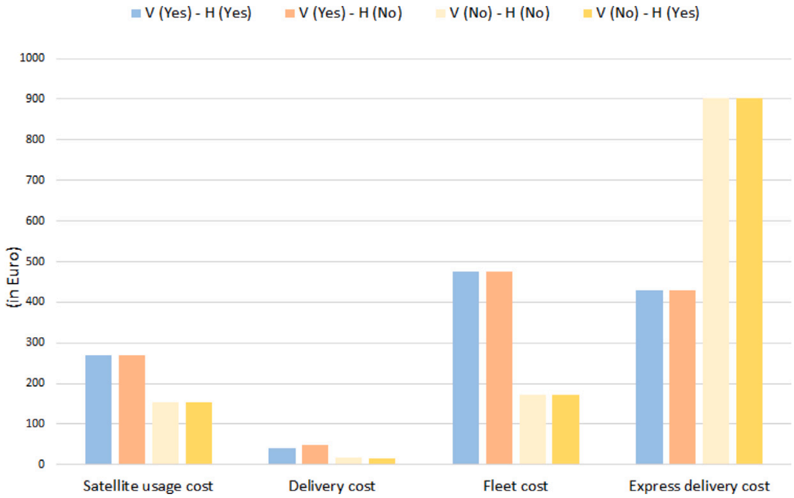

Fig. 5. Vertical and horizontal collaboration: comparative results.

the cooperation between agents at the same level. The proposed model already implements what is sometimes called lateral collaboration, that is, both the horizontal collaboration – since the set of constraints (17) allow to share each satellite among all vehicles – as well as the vertical collaboration-through the bi-level framework.

The vertical non-collaborative scenario corresponds to the case where the upper level agent totally ignores the reaction of the lower level player and sets the tariffs just considering its own benefits. Obviously, the SM will set the tariffs to the highest values. In the horizontal non-collaborative scenario, each satellite is exclusively occupied by the fleet of that specific satellite. This can be reflected into the model adding the set of constraints

$$Y^h_{pj} \leq V^h_{pj}, \ \forall p \in \mathcal{P}, \ j \in \mathcal{J}, \ h \in \mathcal{H} \tag{39}$$

ensuring that the number of vehicles at satellite depot j is below the fleet size denoted by V^h_{pj} where $\sum_{j\in\mathcal{J}} V^h_{pj} = V^h_p$.

In the following, we analyze four possible scenarios depending on whether vertical (V, for short) and/or horizontal (H, for short) collaboration are considered or not. Each scenario is evaluated in terms of SM's gain and SO's cost reduction. Fig. 5 illustrates the results. Comparing the vertical and non-vertical cases, we notice that the vertical collaboration (scenarios V (Yes) - H (Yes) and V (Yes) - H (No)) benefits the SM with an increase of revenue of about 43%, compared to the non-vertical scenarios (V (No) - H (Yes) and V (No) - H (No)). In fact, in the non-collaborative setting, the SM sets the tariffs to the maximum value, but the lower level reaction is to not use satellites since deemed too expensive. In a vertical collaboration setting, the tariffs are appropriately set, taking into account the lower level response and promoting the use of satellite depots against the express delivery option. This in turn leads to considerably higher revenue for the SM. Interestingly, the vertical collaboration is also in favor of the lower level decreasing the total cost by 2%. To explain this phenomenon, we should note that under the vertical collaboration, the tariffs are competitive to the express delivery cost, and therefore, the SOs prefer to deliver the majority of the orders as ordinary shipments operated from satellites; as a matter of fact, the total amount of express deliveries is 47% lower than the total amount of the non-vertical scenarios. This highlights the importance of establishing a collaboration between stakeholders, bringing benefits for both sides.

If horizontal collaboration is implemented, the total lower level cost decreases by 7% (comparing two cases of V (Yes) - H (Yes) and V (Yes) - H (No)) and 3% (V (No) - H (Yes) and V (No) - H (No)). This also confirms the utility of horizontal collaboration, motivating the agents to work cooperatively. The SM's revenue is indifferent with respect to the horizontal collaboration, as expected, since the collaboration in the lower level attempts to decrease the SO's cost.

For the sake of completeness, we also report the Synergy Value (SV) (Santos et al., 2021) for the percentage gain derived from the collaboration, evaluated as $SV = \frac{NC_{\text{Non-Col}} - NC_{\text{Col}}}{NC_{\text{Non-Col}}} 100$ where $NC = \text{SO}_{\text{cost}} - \text{SM}_{\text{Revenue}}$ and labels Col and Non-Col refer to the collaborative and non-collaborative cases, respectively. We should note that we evaluated all the SVs compared to scenario V (Yes) - H (Yes). The SVs are as follows: $SV_{\text{V(No)-H(Yes)}} = 13.24$, $SV_{\text{V(Yes)-H(No)}} = 0.87$, $SV_{\text{V(No)-H(No)}} = 13.52$. The SVs give informative insights on the impact of collaborative operations at different levels. Considering these values, we may draw a few conclusions. First, only the vertical collaboration increases the total gain to 13.24%. Second, the share of horizontal collaboration is limited to 0.87%, confirming our previous claim on the preference of vertical cooperation over horizontal collaboration. Third, the combination of both vertical and horizontal cooperation increases the total gain to 13.52%. Based on these results, we argue that the bi-level model proposed can guarantee a more sustainable solution with a higher synergy value.

7. Conclusions and future research directions

In this paper, we have studied a last-mile logistics problem with multiple satellites in the realm of bi-level optimization. The proposed mathematical model is a generalization of the single-level single-depot version proposed in Perboli et al. (2021a) with a great potential for application in practical settings. The model captures the interplay between the SM and the SO, the two most important agents of the system, interacting in a hierarchical fashion. Notably, in addition to the traditionally considered costs at the lower level, the model explicitly accounts for the leader's tariff-setting problem. An exact solution approach is also discussed and applied in large-sized benchmark instances and in a real-world case study that corresponds to actual transport practices in Turin, Italy. To increase the realism of our study, we have also incorporated customers' preferences. The results show that a broader consideration of service quality brings, as expected, a higher cost on the SO. Another important cost-generating factor for the SO is fleet composition. When the vehicle fleet is heterogeneous, in terms of typical use and capacity, it is preferred to service customers with small and eco-friendly vehicles like EVs and CBs. A homogeneous fleet of only CBs or EVs causes dramatic increases in delivery costs. As already observed (Perboli et al., 2021a), an appropriate mix of different vehicles gives the flexibility to allocate the capacity according to the customer's varying demand, in a more cost-effective way. The use of traditional VANs is not beneficial for any of the actors involved. The modal shift goal to be attained by 2030 is hence necessary for the sustainability of satellite-based two-tier systems and should be addressed by stronger political measures.

As part of future research, this study holds promising opportunities for further development and expansion. One potential avenue worth exploring is the consideration of competition among several SOs. Investigating how multiple SOs interact and compete within the same market context can provide valuable insights into the dynamics of the problem. This could involve analyzing factors such as pricing strategies, service differentiation, and resource allocation among competing entities. By incorporating the competitive aspect, a more comprehensive understanding of the overall system can be achieved.

Another interesting direction for future research lies in the development of a multi-level model that captures the decision-making processes of various stakeholders involved in the supply chain, namely the SM, SO, and customers. This multi-level framework would enable a more nuanced representation of the interactions and dependencies between these stakeholders. It would allow for the exploration of decision trade-offs, coordination mechanisms, and the impact of each stakeholder's choices on the overall system performance. Understanding the different perspectives and objectives of each stakeholder can lead to more effective decision-making and improved supply chain management strategies.

Furthermore, future research could focus on developing stochastic or robust variants of the model to address the inherent uncertainty present in the SO's problem. By incorporating stochastic elements, such as uncertain demand patterns, supply disruptions, or fluctuating market conditions, the model can better reflect the real-world complexities faced by the SO. This would provide more robust and adaptable decision-making capabilities, allowing for more resilient and responsive supply chain operations.

In summary, future research can build upon this work by exploring the dynamics of competition among multiple SOs, developing a multi-level modeling framework, and incorporating stochastic or robust variations to handle uncertainty. These avenues have the potential to enhance the understanding and management of supply chain systems, leading to more effective strategies and improved operational outcomes.

CRediT authorship contribution statement

Maria Elena Bruni: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – original draft. **Sara Khodaparasti:** Methodology, Software, Validation, Writing – original draft. **Guido Perboli:** Conceptualization, Resources, Supervision, Writing – review & editing.

Appendix

A.1. The linearized EHPP

The linearized EHPP corresponding to model M1 is cast as

$$\max \sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} d_i \underline{T}_j^h\, x_{ikj}^h + s\sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l\,\zeta_{ikj}^{lh} \tag{40}$$

$$(2)\text{–}(5),\ (7)\text{–}(13) \tag{41}$$

$$\zeta_{ikj}^{lh} \le x_{ikj}^h,\ \forall i\in I,\ k\in\mathcal{K},\ j\in\mathcal{J},\ h\in\mathcal{H},\ l\in\Lambda_j^h \tag{42}$$

$$\zeta_{ikj}^{lh} \le \gamma_j^{lh},\ \forall i\in I,\ k\in\mathcal{K},\ j\in\mathcal{J},\ h\in\mathcal{H},\ l\in\Lambda_j^h \tag{43}$$

$$\zeta_{ikj}^{lh} \ge x_{ikj}^h + \gamma_j^{lh} - 1,\ \forall i\in I,\ k\in\mathcal{K},\ j\in\mathcal{J},\ h\in\mathcal{H},\ l\in\Lambda_j^h \tag{44}$$

$$\begin{aligned}&\sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} d_i \underline{T}_j^h\, x_{ikj}^h + s\sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l\,\zeta_{ikj}^{lh} + \sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \phi_{ipj}^h x_{ikj}^h\\ &+ \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \delta_p^h\, y_{kj}^h + E\sum_{i\in I} X_i \le \sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}} d_i \underline{T}_j^h\, \bar{x}_{ikj}^h +\\ &s\sum_{i\in I}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l\,\bar{x}_{ikj}^h\gamma_j^{lh} + \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \phi_{ipj}^h \sum_{i\in I}\bar{x}_{ikj}^h\\ &+ \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{k\in\mathcal{K}_p}\sum_{h\in\mathcal{H}} \delta_p^h\, \bar{y}_{kj}^h + E\sum_{i\in I}\bar{X}_i,\ \forall(\bar{x}_{ikj}^h, \bar{y}_{kj}^h, \bar{X}_i)\in\Omega^f\end{aligned} \tag{45}$$

$$\zeta_{ikj}^{lh} \in \{0,1\}, \ \forall i \in \mathcal{I}, \ k \in \mathcal{K}, \ j \in \mathcal{J}, \ h \in \mathcal{H}, \ l \in \Lambda_j^h \tag{46}$$

Also, the EHPP corresponding to model M2 is cast as the mixed integer problem (48)–(53) with linear objective (47).

$$\max \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \underline{T}_j^h \, q_{pj}^{ch} + s \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l \, \mu_{pj}^{clh} \tag{47}$$

$$(2)\text{–}(5),\ (16)\text{–}(22) \tag{48}$$

$$\mu_{pj}^{clh} \le q_{pj}^{ch}, \ \forall c \in \mathcal{C}, \ p \in \mathcal{P}, \ j \in \mathcal{J}, \ h \in \mathcal{H}, \ l \in \Lambda_j^h \tag{49}$$

$$\mu_{pj}^{clh} \le U_{pj}^{ch} \gamma_j^{lh}, \ \forall c \in \mathcal{C}, \ p \in \mathcal{P}, \ j \in \mathcal{J}, \ h \in \mathcal{H}, \ l \in \Lambda_j^h \tag{50}$$

$$\mu_{pj}^{clh} \ge U_{pj}^{ch}(\gamma_j^{lh} - 1) + q_{pj}^{ch}, \ \forall c \in \mathcal{C}, \ p \in \mathcal{P}, \ j \in \mathcal{J}, \ h \in \mathcal{H}, \ l \in \Lambda_j^h \tag{51}$$

$$\begin{aligned}
&\sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \underline{T}_j^h \, q_{pj}^{ch} + s \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l \, \mu_{pj}^{clh} + \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \frac{f_{pj}^h}{\alpha_c} q_{pj}^{ch} \\
&+ \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \delta_p^h \, Y_{pj}^h + E \sum_{c\in\mathcal{C}} \frac{Q^c}{\alpha_c} \le \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \underline{T}_j^h \, \bar{q}_{pj}^{ch} + \\
&s \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}}\sum_{l\in\Lambda_j^h} l \, \gamma_j^{lh} \, \bar{q}_{pj}^{ch} + \sum_{c\in\mathcal{C}}\sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \frac{f_{pj}^h}{\alpha_c} \bar{q}_{pj}^{ch} + \sum_{j\in\mathcal{J}}\sum_{p\in\mathcal{P}}\sum_{h\in\mathcal{H}} \delta_p^h \, \bar{Y}_{pj}^h \\
&+ E \sum_{c\in\mathcal{C}} \frac{\bar{Q}^c}{\alpha_c}, \ \forall (\bar{q}_{pj}^{ch}, \bar{Y}_{pj}^h, \bar{Q}^c) \in \Omega^f
\end{aligned} \tag{52}$$

$$\mu_{pj}^{clh} \ge 0, \ \forall c \in \mathcal{C}, \ p \in \mathcal{P}, \ j \in \mathcal{J}, \ h \in \mathcal{H}, \ l \in \Lambda_j^h \tag{53}$$

where U_{pj}^{ch} is the upper bound on variables q_{pj}^{ch}.

A.2. Detailed computational results

Here we first report the details corresponding to the summarized results of Table 3 in Tables 10 and 11. The results include the CPU time for models M1 and M2 (denoted by CPU_{M1} and CPU_{M2}), the relative percentage solution gap (Gap_M), and the speed up

Table 10
Comparison of models M1 and M2 for $|\mathcal{H}| = 3$.

Instance	CPU_{M1} (s)	CPU_{M2} (s)	Gap_M (%)	Δ(%)
1-200-32-2-3	49.36	0.02	0.90	0.04
5-200-42-2-3	82.89	0.03	0.79	0.04
6-200-42-2-3	38.52	0.03	0.79	0.08
7-200-42-2-3	189.92	0.02	0.78	0.01
8-200-42-2-3	25.89	0.03	0.79	0.12
9-200-42-2-3	69.56	0.03	0.79	0.04
10-200-42-2-3	24.69	0.02	0.78	0.08
2-200-32-2-3	31.69	0.02	0.89	0.06
3-200-32-2-3	16.84	0.03	0.90	0.18
1-200-126-2-3	54.52	0.03	0.81	0.06
2-200-126-2-3	59.50	0.03	0.81	0.05
1-200-132-2-3	64.00	0.03	0.80	0.05
3-200-126-2-3	46.64	0.03	0.82	0.06
4-200-126-2-3	62.94	0.03	0.81	0.05
5-200-126-2-3	110.13	0.03	0.82	0.03
6-200-126-2-3	53.42	0.03	0.81	0.06
1-200-96-2-3	48.09	0.05	0.92	0.10
3-200-96-2-3	54.36	0.03	0.92	0.06
4-200-96-2-3	38.86	0.03	0.92	0.08
5-200-96-2-3	34.28	0.03	0.91	0.09
6-200-96-2-3	70.34	0.03	0.92	0.04
7-200-96-2-3	107.08	0.03	0.92	0.03
8-200-96-2-3	35.11	0.03	0.93	0.09
9-200-96-2-3	44.06	0.03	0.91	0.07
7-200-126-2-3	76.06	0.03	0.82	0.04
8-200-126-2-3	143.80	0.05	0.96	0.03
1-200-120-2-3	27.08	0.03	0.83	0.11
1-200-102-2-3	38.27	0.03	0.90	0.08
4-200-32-2-3	38.09	0.02	0.89	0.05
5-200-32-2-3	21.23	0.03	0.90	0.14

(continued on next page)

Table 10 (*continued*).

Instance	CPU_{M1} (s)	CPU_{M2} (s)	Gap_M (%)	Δ(%)
1-200-34-2-3	65.70	0.03	0.88	0.05
2-200-34-2-3	105.61	0.02	0.87	0.02
3-200-34-2-3	433.63	0.02	0.87	0.00
4-200-34-2-3	59.00	0.03	0.88	0.05
5-200-34-2-3	61.67	0.02	0.88	0.03
1-200-42-2-3	81.30	0.02	0.79	0.02
Avg.	71.23	0.03	0.86	0.06

Table 11
Comparison of models M1 and M2 for $|\mathcal{H}| = 5$.

Instance	CPU_{M1} (s)	CPU_{M2} (s)	Gap_M (%)	Δ(%)
1-200-32-2-5	22.73	0.02	0.90	0.09
5-200-42-2-5	179.42	0.02	0.79	0.01
6-200-42-2-5	148.27	0.02	0.79	0.01
7-200-42-2-5	67.58	0.03	0.78	0.04
8-200-42-2-5	244.39	0.03	0.79	0.01
9-200-42-2-5	46.03	0.03	0.79	0.07
10-200-42-2-5	134.33	0.02	0.78	0.01
2-200-32-2-5	27.39	0.02	0.89	0.07
3-200-32-2-5	87.20	0.03	0.90	0.03
1-200-126-2-5	107.33	0.05	0.81	0.05
2-200-126-2-5	92.81	0.03	0.81	0.03
1-200-132-2-5	197.52	0.02	0.80	0.01
3-200-126-2-5	113.17	0.03	0.82	0.03
4-200-126-2-5	83.52	0.03	0.81	0.04
5-200-126-2-5	174.38	0.05	0.82	0.03
6-200-126-2-5	77.17	0.02	0.81	0.03
1-200-96-2-5	92.41	0.05	0.92	0.05
3-200-96-2-5	100.28	0.05	0.92	0.05
4-200-96-2-5	73.66	0.03	0.92	0.04
5-200-96-2-5	73.41	0.03	0.91	0.04
6-200-96-2-5	76.83	0.03	0.92	0.04
7-200-96-2-5	97.55	0.03	0.92	0.03
8-200-96-2-5	34.53	0.02	0.93	0.06
9-200-96-2-5	57.86	0.03	0.91	0.05
7-200-126-2-5	155.66	0.05	0.82	0.03
8-200-126-2-5	157.08	0.03	0.96	0.02
1-200-120-2-5	51.89	0.03	0.83	0.06
1-200-102-2-5	57.48	0.03	0.90	0.05
4-200-32-2-5	17.55	0.02	0.89	0.11
5-200-32-2-5	22.86	0.02	0.90	0.09
1-200-34-2-5	48.47	0.02	0.88	0.04
2-200-34-2-5	38.28	0.03	0.87	0.08
3-200-34-2-5	32.31	0.02	0.87	0.06
4-200-34-2-5	498.73	0.03	0.88	0.01
5-200-34-2-5	37.52	0.02	0.88	0.05
1-200-42-2-5	1218.30	0.02	0.79	0.00
Avg.	131.83	0.03	0.86	0.04

rate (Δ) calculated as $\Delta = \frac{CPU_{M2}}{CPU_{M1}}100$. The column with heading "Instance" displays the instance name, for example, 5-200-42-2-3 refers to the fifth test case in the class of instances with 200 orders, 42 vehicles, two satellites, and three timeslots.

The detailed results corresponding to Table 5 are shown in Tables 12–14 that report the best iteration of the algorithm (denoted by BI), the total CPU time (CPU) and the average CPU time for the REHPP (CPU_{REHPP}) and LL problem (CPU_{LL}) over all iterations. For cases in which the time limit is reached, we have also reported the relative gap of the best upper and lower bounds (OPT).

A.3. Case study: detailed computational results

In this subsection, we report the information on the case study in detail.

As for the fleet size, we considered a total number of 45 vehicles, shared equally between three vehicle types. The fleet size has been computed in such a way that all the orders can be delivered with a single type of vehicle. The capacity of each vehicle type is set to 45 kg. We consider five satellite depots with capacity of 429 kg, and the minimum and the maximum tariffs (€ per kg) are the same for each satellite and vary based on the timeslot in the following sets [0.07, 0.10, 0.13, 0.05, 0.03] and [0.10, 0.13, 0.16, 0.08, 0.06]. The express delivery cost is set to 3.5 €. The order volumes (in kg) follow a uniform distribution $d_i \sim U(0,3]$ (Perboli et al., 2018).

Tables 15 and 16 report the vehicle usage cost and cost-per-stop parameters used in the case study.

Table 12
Exact solution approach: 200 orders.

Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)	Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)
5-200-42-2-3	2	0.25	0.02	0.02	5-200-42-2-5	5	0.75	0.03	0.02
1-200-42-2-3	3	0.29	0.02	0.02	1-200-42-2-5	5	0.64	0.02	0.02
6-200-42-2-3	2	0.28	0.03	0.02	6-200-42-2-5	5	0.77	0.03	0.03
7-200-42-2-3	2	0.24	0.02	0.02	7-200-42-2-5	4	0.55	0.03	0.02
8-200-42-2-3	2	0.24	0.02	0.02	8-200-42-2-5	4	0.62	0.04	0.03
9-200-42-2-3	2	0.48	0.06	0.09	9-200-42-2-5	5	0.70	0.02	0.02
10-200-42-2-3	2	0.21	0.02	0.02	10-200-42-2-5	4	0.80	0.04	0.04
1-200-132-2-3	2	0.40	0.05	0.03	1-200-132-2-5	2	0.43	0.06	0.03
1-200-126-2-3	2	0.43	0.07	0.03	1-200-126-2-5	5	1.17	0.06	0.04
2-200-126-2-3	2	0.38	0.05	0.02	2-200-126-2-5	4	0.86	0.06	0.04
3-200-126-2-3	3	0.46	0.05	0.03	3-200-126-2-5	5	1.14	0.05	0.03
4-200-126-2-3	2	0.35	0.03	0.02	4-200-126-2-5	5	0.93	0.05	0.02
5-200-126-2-3	3	0.67	0.09	0.03	5-200-126-2-5	2	0.65	0.11	0.05
6-200-126-2-3	2	0.34	0.03	0.02	6-200-126-2-5	4	0.88	0.05	0.03
7-200-126-2-3	3	0.61	0.05	0.04	7-200-126-2-5	4	1.16	0.07	0.05
8-200-126-2-3	3	0.96	0.09	0.08	8-200-126-2-5	2	0.66	0.11	0.04
1-200-96-2-3	2	0.37	0.05	0.03	1-200-96-2-5	2	0.56	0.09	0.04
3-200-96-2-3	2	0.40	0.06	0.03	3-200-96-2-5	2	0.58	0.10	0.03
4-200-96-2-3	2	0.40	0.05	0.03	4-200-96-2-5	2	0.64	0.08	0.03
5-200-96-2-3	3	0.55	0.05	0.04	5-200-96-2-5	2	0.51	0.09	0.03
6-200-96-2-3	2	0.45	0.05	0.05	6-200-96-2-5	2	0.52	0.09	0.05
7-200-96-2-3	3	0.57	0.06	0.03	7-200-96-2-5	2	0.68	0.13	0.04
8-200-96-2-3	3	0.67	0.11	0.03	8-200-96-2-5	3	0.44	0.05	0.02
9-200-96-2-3	3	0.49	0.06	0.03	9-200-96-2-5	2	0.50	0.09	0.03
2-200-96-2-3	2	0.41	0.04	0.02	2-200-96-2-5	3	0.45	0.05	0.02
1-200-120-2-3	3	0.51	0.04	0.01	1-200-120-2-5	3	0.48	0.04	0.03
1-200-102-2-3	2	0.98	0.11	0.22	1-200-102-2-5	3	0.59	0.06	0.04
5-200-32-2-3	2	0.24	0.02	0.02	5-200-32-2-5	2	0.29	0.04	0.02
4-200-32-2-3	2	0.22	0.02	0.02	4-200-32-2-5	2	0.29	0.02	0.02
3-200-32-2-3	3	0.30	0.02	0.02	3-200-32-2-5	3	0.37	0.03	0.02
2-200-32-2-3	3	0.29	0.01	0.02	2-200-32-2-5	3	0.38	0.03	0.02
1-200-32-2-3	2	0.22	0.02	0.02	1-200-32-2-5	2	0.52	0.08	0.05
1-200-34-2-3	2	0.22	0.02	0.02	1-200-34-2-5	2	0.36	0.05	0.03
2-200-34-2-3	3	0.29	0.02	0.02	2-200-34-2-5	3	0.37	0.02	0.03
3-200-34-2-3	2	0.21	0.02	0.01	3-200-34-2-5	3	0.37	0.02	0.03
4-200-34-2-3	3	0.29	0.02	0.02	4-200-34-2-5	3	0.39	0.02	0.03
5-200-34-2-3	2	0.25	0.02	0.02	5-200-34-2-5	2	0.31	0.03	0.02
2-200-42-2-3	3	0.30	0.02	0.02	2-200-42-2-5	4	0.61	0.03	0.02
3-200-42-2-3	2	0.24	0.02	0.02	3-200-42-2-5	4	0.63	0.04	0.02
4-200-42-2-3	2	0.25	0.02	0.02	4-200-42-2-5	5	0.74	0.03	0.03
Avg.		0.39	0.04	0.03			0.61	0.05	0.03

Table 13
Exact solution approach: 500 orders.

Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)	Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)
1-500-246-3-3	3	0.95	0.11	0.06	1-500-246-3-5	3	1.18	0.15	0.06
2-500-246-3-3	2	1.00	0.16	0.08	2-500-246-3-5	2	1.25	0.24	0.09
3-500-246-3-3	3	1.01	0.14	0.06	3-500-246-3-5	3	1.46	0.21	0.10
4-500-246-3-3	4	0.92	0.06	0.04	4-500-246-3-5	3	0.99	0.12	0.05
1-500-258-3-3	4	1.04	0.09	0.05	1-500-258-3-5	4	1.24	0.12	0.09
1-500-318-3-3	3	0.58	0.04	0.04	1-500-318-3-5	7	3.28	0.06	0.23
2-500-318-3-3	4	0.88	0.08	0.03	2-500-318-3-5	6	3.19	0.09	0.20
3-500-318-3-3	3	0.66	0.08	0.03	3-500-318-3-5	7	1.98	0.08	0.05
4-500-318-3-3	3	0.74	0.08	0.05	4-500-318-3-5	7	2.40	0.08	0.10
5-500-318-3-3	2	0.87	0.15	0.09	5-500-318-3-5	7	3.62	0.14	0.18
6-500-318-3-3	3	1.45	0.24	0.09	6-500-318-3-5	7	9.04	0.75	0.28
1-500-258-3-3	4	0.89	0.08	0.03	1-500-258-3-5	3	1.16	0.16	0.05
1-500-108-3-3	4	0.97	0.11	0.03	1-500-108-3-5	5	6.64	0.19	0.85
2-500-108-3-3	3	1.14	0.17	0.07	2-500-108-3-5	5	6.21	0.21	0.43
3-500-108-3-3	2	0.73	0.14	0.09	3-500-108-3-5	5	8.16	0.16	0.88
4-500-108-3-3	2	0.61	0.06	0.06	4-500-108-3-5	6	3.36	0.15	0.21
1-500-252-3-3	4	0.86	0.08	0.03	1-500-252-3-5	2	0.80	0.16	0.07
2-500-252-3-3	3	0.92	0.10	0.06	2-500-252-3-5	2	0.90	0.20	0.09
3-500-252-3-3	3	1.15	0.15	0.09	3-500-252-3-5	2	1.38	0.29	0.09
4-500-252-3-3	2	0.86	0.15	0.10	4-500-252-3-5	3	1.14	0.16	0.07
1-500-82-3-3	4	0.56	0.02	0.03	1-500-82-3-5	3	0.56	0.05	0.04

(continued on next page)

Table 13 (*continued*).

Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)	Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)
2-500-82-3-3	4	0.78	0.04	0.02	2-500-82-3-5	4	0.78	0.05	0.04
3-500-82-3-3	4	0.65	0.04	0.04	3-500-82-3-5	4	0.82	0.07	0.05
4-500-82-3-3	3	0.49	0.05	0.02	4-500-82-3-5	4	0.77	0.06	0.04
5-500-82-3-3	3	0.47	0.03	0.03	5-500-82-3-5	4	0.63	0.04	0.03
1-500-312-3-3	3	1.12	0.16	0.08	1-500-312-3-5	6	5.77	0.17	0.43
1-500-324-3-3	3	1.18	0.14	0.12	1-500-324-3-5	6	4.87	0.16	0.43
2-500-324-3-3	2	0.94	0.18	0.10	2-500-324-3-5	6	3.34	0.16	0.15
3-500-324-3-3	2	0.99	0.19	0.10	3-500-324-3-5	6	2.74	0.17	0.09
1-500-84-3-3	4	0.61	0.05	0.02	1-500-84-3-5	4	0.85	0.06	0.04
2-500-84-3-3	4	0.79	0.04	0.05	2-500-84-3-5	3	0.72	0.05	0.07
3-500-84-3-3	4	1.06	0.06	0.10	3-500-84-3-5	4	1.30	0.07	0.13
1-500-80-3-3	4	0.60	0.04	0.02	1-500-80-3-5	4	0.74	0.05	0.03
1-500-86-3-3	3	0.53	0.06	0.03	1-500-86-3-5	4	1.16	0.06	0.10
1-500-106-3-3	4	0.74	0.05	0.03	1-500-106-3-5	4	0.93	0.07	0.05
2-500-106-3-3	3	0.65	0.05	0.04	2-500-106-3-5	7	1.64	0.05	0.05
3-500-106-3-3	3	0.52	0.03	0.04	3-500-106-3-5	7	1.74	0.03	0.07
4-500-106-3-3	4	0.72	0.05	0.03	4-500-106-3-5	7	2.02	0.05	0.08
5-500-106-3-3	4	0.72	0.03	0.03	5-500-106-3-5	7	1.39	0.03	0.04
6-500-106-3-3	4	0.67	0.03	0.04	6-500-106-3-5	7	1.46	0.03	0.04
7-500-106-3-3	4	0.81	0.05	0.04	7-500-106-3-5	7	2.02	0.05	0.07
Avg.		0.83	0.09	0.05			2.33	0.13	0.15

Table 14
Exact solution approach: 1000 orders.

Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)	Instance	BI	CPU (s)	CPU_{REHPP} (s)	CPU_{LL} (s)	OPT (%)
1-1000-170-4-3	3	3.70	0.19	0.90	1-1000-168-4-5	3	0.71	0.14	0.03	0.00
2-1000-164-4-3	3	1.91	0.11	0.41	2-1000-168-4-5	3	2.33	0.53	0.06	0.00
3-1000-160-4-3	3	3.06	0.10	0.83	3-1000-168-4-5	3	0.79	0.14	0.02	0.00
4-1000-166-4-3	3	1.30	0.08	0.27	1-1000-218-4-5	5	3.40	0.47	0.06	0.00
5-1000-166-4-3	3	1.35	0.03	0.32	2-1000-216-4-5	7	3.30	0.29	0.05	0.00
6-1000-166-4-3	3	2.20	0.04	0.58	3-1000-214-4-5	7	3.04	0.25	0.04	0.00
7-1000-166-4-3	3	2.88	0.57	0.31	4-1000-214-4-5	7	9.16	1.00	0.05	0.00
1-1000-168-4-3	3	2.19	0.22	0.31	5-1000-214-4-5	8	3.89	0.32	0.03	0.00
2-1000-168-4-3	3	2.76	0.51	0.32	6-1000-214-4-5	6	5.92	0.82	0.03	0.00
3-1000-168-4-3	3	1.75	0.18	0.31	7-1000-214-4-5	8	2.40	0.14	0.02	0.00
1-1000-498-4-3	2	3.47	0.21	1.00	8-1000-214-4-5	6	6.03	0.83	0.04	0.00
2-1000-498-4-3	3	2.03	0.20	0.32	9-1000-212-4-5	8	2.91	0.17	0.03	0.00
3-1000-498-4-3	3	1.28	0.27	0.04	10-1000-212-4-5	6	2.27	0.22	0.03	0.00
4-1000-498-4-3	3	8.46	0.59	2.00	1-1000-170-4-5	4	1.88	0.32	0.02	0.00
5-1000-498-4-3	3	2.28	0.48	0.15	2-1000-164-4-5	3	1.43	0.23	0.07	0.00
6-1000-498-4-3	3	1.69	0.21	0.17	3-1000-160-4-5	3	2.50	0.69	0.03	0.00
7-1000-498-4-3	3	5.10	0.58	0.97	4-1000-166-4-5	3	1.24	0.23	0.03	0.00
8-1000-498-4-3	2	6.97	1.00	2.00	5-1000-166-4-5	3	2.03	0.49	0.03	0.00
9-1000-498-4-3	3	4.26	0.73	0.44	6-1000-166-4-5	3	1.14	0.19	0.04	0.00
10-1000-504-4-3	3	3.32	0.57	0.32	7-1000-166-4-5	3	1.07	0.20	0.03	0.00
1-1000-642-4-3	4	7.91	0.33	1.00	1-1000-498-4-5	2	12.69	6.00	0.05	0.00
2-1000-642-4-3	2	3.70	0.23	1.00	2-1000-498-4-5	3	42.17	10.00	0.03	0.00
3-1000-642-4-3	2	5.58	0.75	2.00	3-1000-498-4-5	2	1.81	0.65	0.02	0.00
4-1000-642-4-3	3	2.04	0.14	0.33	4-1000-498-4-5	2	103.36	50.00	0.07	0.00
5-1000-642-4-3	3	3.37	0.53	0.40	5-1000-498-4-5	1	1800	0.63	0.05	114.29
6-1000-642-4-3	3	2.70	0.33	0.27	6-1000-498-4-5	2	7.31	2.00	0.02	0.00
7-1000-636-4-3	2	7.22	0.21	3.00	7-1000-498-4-5	2	11.84	6.00	0.02	0.00
8-1000-654-4-3	4	9.43	0.65	1.00	8-1000-498-4-5	2	1.44	0.52	0.02	0.00
9-1000-648-4-3	4	8.60	0.64	1.00	9-1000-498-4-5	1	1800	0.50	0.02	114.29
10-1000-630-4-3	2	0.58	0.15	0.02	10-1000-504-4-5	3	5.50	0.41	0.01	0.00
1-1000-218-4-3	3	0.89	0.10	0.08	1-1000-642-4-5	4	718	200	0.13	0.00
2-1000-216-4-3	3	1.74	0.03	0.45	2-1000-642-4-5	6	5.21	0.70	0.02	0.00
3-1000-214-4-3	3	3.72	0.88	0.25	3-1000-642-4-5	6	3.14	0.30	0.02	0.00
4-1000-214-4-3	3	1.95	0.11	0.43	4-1000-642-4-5	4	5.87	1.00	0.12	0.00
5-1000-214-4-3	3	1.59	0.18	0.25	5-1000-642-4-5	4	16.23	4.00	0.06	0.00
6-1000-214-4-3	3	1.91	0.08	0.45	6-1000-642-4-5	4	2.55	0.30	0.22	0.00
7-1000-214-4-3	4	1.13	0.07	0.09	7-1000-636-4-5	6	3.60	0.40	0.02	0.00
8-1000-214-4-3	4	1.86	0.09	0.25	8-1000-654-4-5	1	1800	0.52	0.03	114.29
9-1000-212-4-3	4	1.77	0.07	0.25	9-1000-648-4-5	1	1800	0.59	0.05	114.29
10-1000-212-4-3	3	1.92	0.14	0.39	10-1000-630-4-3	9	6.70	0.20	0.02	0.00
Avg.		3.29	0.31	0.62			205.12	7.31	0.04	11.43

Table 15
Input parameters: Vehicle usage cost.

	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
CB	20.16	28.80	37.44	14.40	8.64
VAN	30.24	43.20	56.16	21.60	12.96
EV	24.64	35.20	45.76	17.60	10.56

Table 16
Input parameters: Cost-per-stop.

Satellite	Cluster	CB					VAN					EV				
		$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
SD1	1	0.09	0.12	0.15	0.06	0.03	0.25	0.36	0.47	0.18	0.11	0.21	0.29	0.38	0.15	0.09
	2	0.09	0.12	0.15	0.06	0.03	0.25	0.36	0.47	0.18	0.11	0.21	0.29	0.38	0.15	0.09
	3	0.16	0.22	0.28	0.11	0.06	0.45	0.65	0.84	0.32	0.19	0.37	0.53	0.68	0.26	0.16
	4	0.16	0.22	0.28	0.11	0.06	0.46	0.66	0.85	0.33	0.19	0.38	0.53	0.69	0.27	0.16
	5	0.13	0.18	0.23	0.09	0.05	0.37	0.53	0.69	0.27	0.16	0.31	0.43	0.56	0.22	0.13
SD2	1	0.10	0.14	0.17	0.07	0.04	0.29	0.41	0.53	0.20	0.12	0.23	0.33	0.43	0.17	0.10
	2	0.10	0.14	0.17	0.07	0.04	0.29	0.41	0.53	0.20	0.12	0.23	0.33	0.43	0.17	0.10
	3	0.16	0.22	0.28	0.11	0.06	0.46	0.66	0.85	0.33	0.20	0.38	0.54	0.70	0.27	0.16
	4	0.16	0.22	0.29	0.11	0.06	0.47	0.67	0.87	0.33	0.20	0.38	0.55	0.71	0.27	0.16
	5	0.12	0.17	0.22	0.08	0.05	0.36	0.51	0.66	0.25	0.15	0.29	0.41	0.53	0.21	0.12
SD3	1	0.11	0.16	0.20	0.08	0.04	0.34	0.48	0.62	0.24	0.14	0.27	0.39	0.50	0.19	0.11
	2	0.11	0.16	0.20	0.08	0.04	0.34	0.48	0.62	0.24	0.14	0.27	0.39	0.50	0.19	0.11
	3	0.11	0.15	0.19	0.07	0.04	0.31	0.44	0.57	0.22	0.13	0.25	0.36	0.46	0.18	0.11
	4	0.11	0.15	0.19	0.07	0.04	0.31	0.45	0.58	0.22	0.13	0.26	0.36	0.47	0.18	0.11
	5	0.09	0.12	0.15	0.06	0.03	0.25	0.36	0.47	0.18	0.11	0.21	0.29	0.38	0.15	0.09
SD4	1	0.07	0.09	0.12	0.05	0.03	0.20	0.28	0.36	0.14	0.08	0.16	0.23	0.29	0.11	0.07
	2	0.07	0.09	0.12	0.05	0.03	0.20	0.28	0.36	0.14	0.08	0.16	0.23	0.29	0.11	0.07
	3	0.09	0.13	0.16	0.06	0.04	0.27	0.39	0.50	0.19	0.11	0.22	0.31	0.41	0.16	0.09
	4	0.10	0.13	0.17	0.07	0.04	0.28	0.40	0.51	0.20	0.12	0.23	0.32	0.42	0.16	0.10
	5	0.14	0.19	0.25	0.10	0.05	0.41	0.58	0.75	0.29	0.17	0.33	0.47	0.61	0.24	0.14
SD5	1	0.06	0.08	0.10	0.04	0.02	0.16	0.23	0.30	0.11	0.07	0.13	0.19	0.24	0.09	0.06
	2	0.06	0.08	0.10	0.04	0.02	0.16	0.23	0.30	0.11	0.07	0.13	0.19	0.24	0.09	0.06
	3	0.14	0.20	0.26	0.10	0.06	0.42	0.6	0.78	0.30	0.18	0.34	0.49	0.63	0.24	0.14
	4	0.15	0.20	0.26	0.10	0.06	0.43	0.61	0.79	0.31	0.18	0.35	0.50	0.64	0.25	0.15
	5	0.15	0.21	0.27	0.10	0.06	0.44	0.63	0.81	0.31	0.19	0.36	0.51	0.66	0.26	0.15

Table 17
Results: Flexible delivery time scenario.

	# of deployed vehicles					Average vehicle fill ratio					Delivery ratio				
	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5
CB	15	–	–	–	15	1.00	–	–	–	1.00	0.74	–	–	–	0.55
EV	6	–	–	–	13	0.90	–	–	–	0.97	0.26	–	–	–	0.45

Table 18
Results: Free delivery time scenario.

	# of deployed vehicles					Average vehicle fill ratio					Delivery ratio				
	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5	h = 1	h = 2	h = 3	h = 4	h = 5
CB	–	–	–	3	15	–	–	–	1.00	1.00	–	–	–	0.15	0.23
VAN	–	–	–	–	15	–	–	–		0.97	–	–	–	–	0.31
EV	–	–	–	–	15	–	–	–	–	1.00	–	–	–	–	0.31

Tables 17–18 report the optimal solution in terms of the number of deployed vehicles of each type, average vehicle fill ratio, and delivery ratio evaluated as the ratio between the demand serviced at that specific timeslot and the total demand.

References

Alcaraz, J.J., Caballero-Arnaldos, L., Vales-Alonso, J., 2019. Rich vehicle routing problem with last-mile outsourcing decisions. Transp. Res. E (ISSN: 1366-5545) 129, 263–286. http://dx.doi.org/10.1016/j.tre.2019.08.004.

Allen, J., Browne, M., Woodburn, A., Léonardi, J., 2012. The role of urban consolidation centres in sustainable freight transport. Transp. Rev. 32, 473–490.

Arrieta-Prieto, M., Ismael, A., Rivera-Gonzalez, C., Mitchell, J.E., 2022. Location of urban micro-consolidation centers to reduce the social cost of last-mile deliveries of cargo: A heuristic approach. Networks 79 (3), 292–313.

Baldi, M.M., Crainic, T.G., Perboli, G., Tadei, R., 2012. The generalized bin packing problem. Transp. Res. E 48 (6), 1205–1220.

Ballantyne, E.E., Lindholm, M., Whiteing, A., 2013. A comparative study of urban freight transport planning: addressing stakeholder needs. J. Transp. Geogr. (ISSN: 0966-6923) 32, 93–101. http://dx.doi.org/10.1016/j.jtrangeo.2013.08.013.

Ballot, E., Montreuil, B., Meller, R.D., 2014. The Physical Internet. La Documentation Française.

Björklund, M., Johansson, H., 2018. Urban consolidation centre – a literature review, categorisation, and a future research agenda. Int. J. Phys. Distrib. Logist. Manage. 48 (8), 745–764.

Brotcorne, L., Labbé, M., Marcotte, P., Savard, G., 2008. Joint design and pricing on a network. Oper. Res. 56 (5), 1104–1115. http://dx.doi.org/10.1287/opre.1080.0617.

Brotcorne, L., Perboli, G., Rosano, M., Wei, Q., 2019. A managerial analysis of urban parcel delivery: a lean business approach. Sustainability 11, 34–39.

de Carvalho, P.P.S., de Araújo Kalid, R., Rodríguez, J.L.M., Santiago, S.B., 2019. Interactions among stakeholders in the processes of city logistics: a systematic review of the literature. Scientometrics 120 (2), 567–607.

Colson, B., Marcotte, P., Savard, G., 2007. An overview of bilevel optimization. Ann. Oper. Res. (ISSN: 1572-9338) 153 (1), 235–256. http://dx.doi.org/10.1007/s10479-007-0176-2.

Costa, A., Ng, T.S., Foo, L.X., 2017. Complete mixed integer linear programming formulations for modularity density based clustering. Discrete Optim. 25, 141–158.

Crainic, T.G., Gendreau, M., Jemai, L., 2020. Planning hyperconnected, urban logistics systems. Transp. Res. Procedia 47, 35–42. http://dx.doi.org/10.1016/j.trpro.2020.03.070.

Crainic, T.G., Montreuil, B., 2016. Physical internet enabled hyperconnected city logistics. In: Taniguchi, E., Thompson, R.G. (Eds.), 9th International Conference on City Logistics, June 2015, Tenerife, Spain. In: Transportation Research Procedia, vol. 12, Elsevier, pp. 383–398.

Crainic, T.G., Perboli, G., Rei, W., Tadei, R., 2011. Efficient lower bounds and heuristics for the variable cost and size bin packing problem. Comput. Oper. Res. 38 (11), 1474–1482. http://dx.doi.org/10.1016/j.cor.2011.01.001.

Crainic, T.G., Perboli, G., Ricciardi, N., 2021. City logistics. In: Crainic, T.G., Gendreau, M., Gendron, B. (Eds.), Network Design with Applications in Transportation and Logistics. Springer, Boston, pp. 507–537, (Chapter 16).

Crainic, T.G., Perboli, G., Rosano, M., 2018. Simulation of intermodal freight transportation systems: a taxonomy. European J. Oper. Res. 270, 401–418. http://dx.doi.org/10.1016/j.ejor.2017.11.061.

Crainic, T.G., Ricciardi, N., Storchi, G., 2004. Advanced freight transportation systems for congested urban areas. Transp. Res. C (ISSN: 0968-090X) 12 (2), 119–137. http://dx.doi.org/10.1016/j.trc.2004.07.002.

Crainic, T.G., Ricciardi, N., Storchi., G., Storchi, G., 2009. Models for evaluating and planning city logistics systems. Transp. Sci. 43, 432–454. http://dx.doi.org/10.1287/trsc.1090.0279.

Crotti, D., Maggi, E., 2022. Social responsibility and urban consolidation centres in sustainable freight transport markets. Ital. Econ. J. 9, 829–850.

Daganzo, C.F., 1984. The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. Transp. Sci. 18 (4), 331–350.

De Marco, A., Mangano, G., Zenezini, G., Cagliano, A.C., Perboli, G., Rosano, M., Musso, S., 2017. Business modeling of a city logistics ICT platform. In: Proceedings - International Computer Software and Applications Conference. Vol. 2, pp. 783–789.

Dempe, S., Dinh, N., Dutta, J., Pandit, T., 2019. Simple bilevel programming and extensions, part-1: Theory. http://dx.doi.org/10.48550/ARXIV.1912.06376, URL https://arxiv.org/abs/1912.06376.

Dempe, S., Kalashnikov, V., Pérez-Valdés, G.A., Kalashnykova, N., 2015. Bilevel programming problems. Energy Syst. Springer, Berlin.

Dreischerf, A.J., Buijs, P., 2022. How urban consolidation centres affect distribution networks: An empirical investigation from the perspective of suppliers. Case Stud. Transp. Policy (ISSN: 2213-624X) 10 (1), 518–528. http://dx.doi.org/10.1016/j.cstp.2022.01.012.

Franceschetti, A., Jabali, O., Laporte, G., 2017. Continuous approximation models in freight distribution management. Top 25, 413–433.

Giusti, R., Manerba, D., Perboli, G., Tadei, R., Yuan, S., 2018. A new open-source system for strategic freight logistics planning: The SYNCHRO-NET optimization tools. In: Transp. Res. Procedia. 30, pp. 245–254.

Hansen, P., Jaumard, B., Savard, G., 1992. New branch-and-bound rules for linear bilevel programming. SIAM J. Sci. Statist. Comput. 13 (5), 1194–1217.

van Heeswijk, W., Larsen, R., Larsen, A., 2019. An urban consolidation center in the city of copenhagen: A simulation study. Int. J. Sustain. Transp. 13 (9), 675–691. http://dx.doi.org/10.1080/15568318.2018.1503380.

Hu, Q., Gu, W., Wang, S., 2022. Optimal subsidy scheme design for promoting intermodal freight transport. Transp. Res. E 157, http://dx.doi.org/10.1016/j.tre.2021.102561.

Isa, S.S., Lima, O.F., Vidal Vieira, J.G., 2021. Urban consolidation centers: Impact analysis by stakeholder. Res. Transp. Econom. (ISSN: 0739-8859) 90, 101045. http://dx.doi.org/10.1016/j.retrec.2021.101045.

Janjevic, M., Ndiaye, A., 2017a. Investigating the financial viability of urban consolidation centre projects. Res. Transp. Bus. Manage. 24, 101–113.

Janjevic, M., Ndiaye, A., 2017b. Investigating the theoretical cost-relationships of urban consolidation centres for their users. Transp. Res. A 102 (C), 98–118.

Jeroslow, R.G., 1985. The polynomial hierarchy and a simple model for competitive analysis. Math. Program. 32 (2), 146–164.

Ji, Y., Qu, S., Yu, Z., 2017. Bi-level multi-objective optimization model for last mile delivery using a discrete approach. J. Difference Equ. Appl. 23 (1–2), 179–190.

Kane Logistics, 2011. Hershey and ferrero embrace collaborative distribution. https://www.kanelogistics.com/blog/hershey-and-ferrero-embrace-collaborative-distribution. Last access: 11/07/2023.

Kaspi, M., Raviv, T., Ulmer, M.W., 2022. Directions for future research on urban mobility and city logistics. Networks 79 (3), 253–263. http://dx.doi.org/10.1002/net.22092.

Kin, B., Verlinde, S., van Lier, T., Macharis, C., 2016. Is there life after subsidy for an urban consolidation centre? An investigation of the total costs and benefits of a privately-initiated concept. Transp. Res. Procedia (ISSN: 2352-1465) 12, 357–369. http://dx.doi.org/10.1016/j.trpro.2016.02.072, Tenth International Conference on City Logistics 17-19 June 2015, Tenerife, Spain.

Kleinert, T., Labbé, M., Ljubić, I., Schmidt, M., 2021. A survey on mixed-integer programming techniques in bilevel optimization. EURO J. Comput. Optim. 9, 100007.

Leyffer, S., López-Calva, G., Nocedal, J., 2006. Interior methods for mathematical programs with complementarity constraints. SIAM J. Optim. 17 (1), 52–77. http://dx.doi.org/10.1137/040621065.

Lozano, L., Smith, J.C., 2017. A value-function-based exact approach for the bilevel mixed-integer programming problem. Oper. Res. 65 (3), 768–786.

Marciani, M., Cossu, P., 2014. How the {URBeLOG} project will enable a new governance model for city logistics in Italian metropolitan areas. Procedia Soc. Behav. Sci. 151, 230–243.

McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. Math. Program. 10 (1), 147–175.

Municipality of Turin, 2019. Torino living lab. https://www.torinocitylab.it/en/component/tags/tag/torino-living-lab. Last access: 11/07/2023.

Musso, S., Perboli, G., Apruzzese, M., Renzi, G., Selini, N., 2022. Innovative business models in ports' logistics. In: Proceedings - 2022 IEEE 46th Annual Computers, Software, and Applications Conference. COMPSAC 2022, pp. 1702–1707. http://dx.doi.org/10.1109/COMPSAC54236.2022.00271.

Pahwa, A., Jaller, M., 2022. A cost-based comparative analysis of different last-mile strategies for e-commerce delivery. Transp. Res. E (ISSN: 1366-5545) 164, 102783. http://dx.doi.org/10.1016/j.tre.2022.102783.

Perboli, G., 2016. GUEST-OR. Linking lean business and OR. In: 28th European Conference on Operation Research. p. Poznan, Poland.

Perboli, G., Brotcorne, L., Bruni, M.E., Rosano, M., 2021a. A new model for last-mile delivery and satellite depots management: The impact of the on-demand economy. Transp. Res. E 145, 102184.

Perboli, G., Musso, S., Rosano, M., Tadei, R., Godel, M., 2017. Synchro-modality and slow steaming: New business perspectives in freight transportation. Sustainability (Switzerland) 9 (10), 1843.

Perboli, G., Rosano, M., 2019. Parcel delivery in urban areas: Opportunities and threats for the mix of traditional and green business models. Transp. Res. C 99, 19–36. http://dx.doi.org/10.1016/j.trc.2019.01.006.

Perboli, G., Rosano, M., Saint-Guillain, M., Rizzo, P., 2018. Simulation–optimisation framework for city logistics: an application on multimodal last-mile delivery. IET Intell. Transp. Syst. 12 (4), 262–269. http://dx.doi.org/10.1049/iet-its.2017.0357.

Perboli, G., Rosano, M., Wei, Q., 2022. A simulation-optimization approach for the management of the on-demand parcel delivery in sharing economy. IEEE Trans. Intell. Transp. Syst. 23 (8), 10570–10582.

Perboli, G., Tadei, R., Gobbato, L., 2014. The multi-handler knapsack problem under uncertainty. European J. Oper. Res. 236 (3), 1000–1007.

Perboli, G., Tadei, R., Vigo, D., 2011. The two-echelon capacitated vehicle routing problem: Models and math-based heuristics. Transp. Sci. 45, 364–380. http://dx.doi.org/10.1287/trsc.1110.0368.

Qiu, Y., Zhou, D., Du, Y., Liu, J., Pardalos, P.M., Qiao, J., 2021. The two-echelon production routing problem with cross-docking satellites. Transp. Res. E (ISSN: 1366-5545) 147, 102210. http://dx.doi.org/10.1016/j.tre.2020.102210.

Qu, W., Rezaei, J., Maknoon, Y., Tavasszy, L., 2019. Hinterland freight transportation replanning model under the framework of synchromodality. Transp. Res. E 131, 308–328. http://dx.doi.org/10.1016/j.tre.2019.09.014.

Quak, H., Tavasszy, L., 2011. Customized solutions for sustainable city logistics: The viability of urban freight consolidation centres. In: van Nunen, J.A., Huijbregts, P., Rietveld, P. (Eds.), Transitions Towards Sustainable Mobility: New Solutions and Approaches for Sustainable Transport Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 213–233.

Saglietto, I., 2013. Towards a classification of fourth party logistics (4PL). Univ. J. Ind. Bus. Manage. 1 (3), 104–116.

Santos, M.J., Curcio, E., Amorim, P., Carvalho, M., Marques, A., 2021. A bilevel approach for the collaborative transportation planning problem. Int. J. Prod. Econ. 233, 108004.

Simoni, M.D., Kutanoglu, E., Claudel, C.G., 2020. Optimization and analysis of a robot-assisted last mile delivery system. Transp. Res. E (ISSN: 1366-5545) 142, 102049. http://dx.doi.org/10.1016/j.tre.2020.102049.

Skender, H.P., Mircović, P.A., Prudky, I., 2017. The role of the 4PL model in a contemporary supply chain. Sci. J. Marit. Res. 31, 96–101.

von Stackelberg, H., Peacock, A., 1952. The Theory of the Market Economy. William Hodge.

Tawfik, C., Limbourg, S., 2018. Pricing problems in intermodal freight transport: Research overview and prospects. Sustainability (Switzerland) 10 (9), 33–41. http://dx.doi.org/10.3390/su10093341.

Tawfik, C., Limbourg, S., 2019. A bilevel model for network design and pricing based on a level-of-service assessment. Transp. Sci. 53 (6), 1609–1626. http://dx.doi.org/10.1287/trsc.2019.0906.

The GUEST Initiative, 2017. http://www.theguestmethod.com. Last Access: 11/03/2023.

ULaaDS, 2020. Urban logistics as an on-demand service web site. http://hulaads.eu/. Last access: 11/07/2023.

World Economic Forum, 2020. The Future of the Last-Mile Ecosystem. Technical Report.

Xu, X., Zheng, Y., Yu, L., 2018. A bi-level optimization model of LRP in collaborative logistics network considered backhaul no-load cost. Soft Comput. 22 (16), 5385–5393.

Yang, G., Huang, Y., Fu, Y., Huang, B., Sheng, S., Mao, L., Huang, S., Xu, Y., Le, J., Ouyang, Y., et al., 2020. Parcel locker location based on a bilevel programming model. Math. Probl. Eng. 2020, 1–12.

Zhu, S.X., Ursavas, E., 2018. Design and analysis of a satellite network with direct delivery in the pharmaceutical industry. Transp. Res. E (ISSN: 1366-5545) 116, 190–207. http://dx.doi.org/10.1016/j.tre.2018.06.005.