EURO Journal on Transportation and Logistics 12 (2023) 100114

Contents lists available at ScienceDirect

EURO Journal on Transportation and Logistics

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

Clustering and routing in waste management: A two-stage optimisation approach

Massimiliano Caramia [a], Diego Maria Pinto [b], Emanuele Pizzari [a,*], Giuseppe Stecca [b]

[a] Dipartimento di Ingegneria dell'Impresa, Università di Roma "Tor Vergata", Viale del Politecnico 1, 00133 Rome, Italy
[b] Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", Consiglio Nazionale delle Ricerche, Via dei Taurini 19, 00185 Rome, Italy

A R T I C L E I N F O

Keywords:
Waste management
Clustering
Routing
Two-stage model
Facility location
Fairness

A B S T R A C T

This paper proposes a two-stage model to tackle a problem arising in Waste Management. The decision-maker (a regional authority) is interested in locating sorting facilities in a regional area and defining the corresponding capacities. The decision-maker is aware that waste will be collected and brought to the installed facilities by independent private companies. Therefore, the authority wants to foresee the behaviour of these companies in order to avoid shortsighted decisions. In the first stage, the regional authority divides the clients into clusters, further assigning facilities to these clusters. In the second stage, an effective route is defined to serve client pickup demand. The main idea behind the model is that the authority aims to find the best location–allocation solution by clustering clients and assigning facilities to these clusters without generating overlaps. In doing so, the authority tries to (i) assign the demand of clients to the facilities by considering a safety stock within their capacities to avoid shortages during the operational phase, (ii) minimise Greenhouse Gases emissions, (iii) be as compliant as possible with the solution found by the second stage problem, the latter aiming at optimising vehicle tour lengths. After properly modelling the problem, we propose a matheuristic solution algorithm and conduct extensive computational analysis on a real-case scenario of an Italian region. Validation of the approach is achieved with promising results.

1. Introduction

In recent years, sustainability has become the centre of public attention. This global environmental concern results in new recovery and recycling targets imposed by national and international waste directives. To achieve these goals, specific policies must be defined and employed (Šomplák et al., 2019). The need to achieve these goals has caused an increasing interest in the recovery of materials from the waste streams (Pinto and Stecca, 2020) along with the rising prices of raw materials. Recovery of materials is a common concept of the Circular Economy. Every object that reaches an end-of-life state is put back into the stream to create additional value. Therefore, it is vital to have technologically advanced facilities able to manage waste, although with the possible drawback of increasing costs (Swart and Groot, 2015). Since the recycling industry is already characterised by low margins and high operations and logistics costs, optimising the supply chain processes becomes critical to turn it into a feasible and profitable market.

The increased interest in sustainability also influenced the evolution of the supply chain, with the rise of the term Green Supply Chain (GSC) (Beamon, 1999). GSC is a complex field of research with several tasks involved. Some of these are Green Design, Green Manufacturing, Network Design and Waste Management (WM) (Srivastava, 2007). The abundance of decision problems in GSC led to the rise of several optimisation models and approaches in the literature (Tseng et al., 2019). Each decision problem has a specific time horizon. Three main categories have been individuated, such as long-term, medium-term, and short-term. Problems falling into each of these categories are called, respectively, strategic, tactical, and operational. An example of a strategic task is facility location, transportation is considered a tactical task, and an example of an operational task is routing. Each decision problem may be organised by different decision-makers, who could even be in contrast with each other.

Indeed, considerable attention has been directed towards the optimisation of aspects related to WM, considered one of the main tasks of GSC (Srivastava, 2007). The Waste Framework Directive 2008/98/EC of the European Parliament (Directive, 2008) defines WM as a group of several tasks, ranging from waste collection to disposal to after-care of the sites designated for disposal. WM is a vast field of optimisation,

* Corresponding author.
E-mail addresses: caramia@dii.uniroma2.it (M. Caramia), diegomaria.pinto@iasi.cnr.it (D.M. Pinto), emanuele.pizzari@uniroma2.it (E. Pizzari), giuseppe.stecca@iasi.cnr.it (G. Stecca).

https://doi.org/10.1016/j.ejtl.2023.100114
Received 2 August 2022; Received in revised form 6 June 2023; Accepted 10 July 2023
Available online 18 July 2023

comprising several strategic, tactical, and operational tasks, such as the location of collection areas (Bautista and Pereira, 2006), waste collection (Li et al., 2008), planning of waste sorting operations (Pinto and Stecca, 2020; Pinto et al., 2020; Gentile et al., 2022).

When dealing with a WM network, there is a concurrence of critical actions, such as collection, sorting, disposal, waste-to-energy practices, and recycling. Moreover, the impact of these tasks must be analysed from an economic and environmental perspective. Therefore, the simultaneity of strategic, operational and environmental problems is highlighted. The findings in Vigo et al. (2007) state that integrated planning of different decision problems may lead to much larger savings, a result confirmed in the work of Hemmelmayr et al. (2014).

Focusing on the potential imbalance among regions in charge of WM is essential. Indeed, the rates at which countries manage solid waste using landfill, waste-to-energy plants, or composting facilities differ considerably at the global level (Kaza et al., 2018). Di Foggia and Beccarello (2021) try to analyse these differences, further noting that some regions struggle to achieve self-sufficiency. One of the main reasons is the level of industrialisation, most strikingly characterised by the critical under-capacity. The authors underline how some regions' overcapacity counterbalances others' under-capacity. This situation generates negative environmental and economic externalities due to waste exports (Di Foggia and Beccarello, 2021), further highlighting the importance of proper facility location and capacity allocation.

We want to focus on the design of a WM regional network. Sorting facilities must be located and waste should be collected from various municipalities and brought to appropriate facilities. The objective of being sustainable is translated by focusing on emissions due to facilities and transportation, as well as considering avoiding wastefulness of allocated capacity. The location of facilities and the collection of waste are not conducted by the same decision-maker. Indeed, facility location is handled by a regional authority, while waste collection is handled by private companies.

The remainder of this paper is organised as follows: Section 2 is dedicated to the literature review; at the end of the section, we highlight the findings of the literature review that we included in our paper, as well as the main contributions of the paper; Section 3 is devoted to the problem formulation, while Section 4 details the solving algorithm; results and managerial implications are reported in Section 5; Section 6 gives conclusions and future perspectives.

2. Literature review

Facility location is a staple of OR literature. Melo et al. (2009) provide a survey for facility location models specifically for supply chains. Farahani et al. (2010) focus on multi-objective and multi-attribute models for facility location, while Farahani et al. (2014) surveyed hierarchical models. While facility location is a well-known and researched optimisation problem, sustainable facility location is not. Interest in this topic has increased in recent years (Terouhid et al., 2012). Sustainable facility location takes into account the three pillars of sustainability, i.e. economic, environmental, and social. Cost is a common proxy for the economic pillar, emission and usage of land is a proxy for the environmental pillar, while job creation and impact on people's health is a proxy for the social pillar (Barbosa-Póvoa et al., 2018). Olapiriyakul et al. (2019) propose a multiobjective optimisation model to design a cost-effective waste management supply chain, including all three aspects of sustainability. They study the effect of the presence of a facility on the land and on people's health. They also study the effect on people's health of emissions due to transportation. The impact of the presence of a facility on people's health is also studied by Tirkolaee et al. (2021), who focused on medical waste treatment during the COVID-19 outbreak. They specifically focused on the infection and environmental risk to the population living close to the disposal centres. The distance between facilities and residential areas is also investigated in the works of Kumar et al. (2020) and Sherif et al. (2022) via means of multi-criteria decision-making for the selection of proper locations. They put an emphasis on recycling centres for electronic equipment and batteries, respectively. Saadatlu et al. (2023) design a sustainable waste management system. The effect on the environment is considered by studying the effect of leachates, while the social pillar is included by studying how the presence of a facility affects the creation of jobs. Tirkolaee et al. (2020) design an urban waste management system using robust optimisation focusing on pollution, linking the cost of pollution to the distance between collection centres and demand nodes. A greater distance yields a lower cost of pollution because it will impact fewer people. They also put a cost on non-collected waste to account for the damage to the environment and the population. Concerning transportation, one of the most common approaches is the Vehicle Routing Problem (VRP), which deals with a set of customers that need to be visited. The VRP generalises the Travelling Salesman Problem (TSP). This class of problems was first introduced in 1959 by Dantzig and Ramser (1959), who also presented a heuristic for the solution. Since the TSP is a particular case of the VRP, the VRP is NP-hard (Garey and Johnson, 1978). There are two main classes of VRPs, i.e. arc routing and node routing. Arc routing problems deal with models that place the service required on the arc of the networks, whereas node routing problems represent clients by nodes of the network. The literature has historically mainly focused on node routing (Corberán et al., 2021). The base VRP model has been enriched by scholars with the addition of several constraints, such as time-window constraints (i.e. certain nodes/arcs may be visited only in a specific time frame) (Dumas et al., 1991), maximum limit of working hours, emissions, maximum limit of functioning hours of electric vehicles (Kucukoglu et al., 2021), and others. The Green VRP is a version of VRP that includes not only economic objectives but also takes into account the usage of fuels, the adoption of alternative fuels, and emissions. Erdoğan and Miller-Hooks (2012) theorised this model, introducing fuel stations in the network. Vehicles need to visit these stations to replenish their tanks and continue operating. The literature on Green VRP also includes other environmental objectives, such as environmental impact, and social objectives, such as level of service and journey risk (Asghari et al., 2021). Bruglieri et al. (2019) introduce the Green Vehicle Routing Problem with capacitated fuel stations, a more realistic variant of the Green Vehicle Routing Problem.

As mentioned above, integrated approaches lead to better results. The idea of combining location and routing dates back fifty years (Prodhon and Prins, 2014). Older models mainly dealt with uncapacitated facilities or vehicles, but after the survey of Nagy and Salhi (2007), scholars started dedicating their resources to models with capacitated facilities and vehicles. In the field of Green VRP, Tricoire and Parragh (2017) investigate the trade-off between location and routing by devising a multi-objective model that takes care of location costs and emissions from transportation. Toro et al. (2017) also develop a multi-objective model for a location-routing problem, focusing on environmental impact. Zhang et al. (2018) focus on emergency facility location, including uncertainty in the model. Li et al. (2019) develop a genetic algorithm combining tabu search and local search to solve a biomass feedstock delivery problem with carbon emissions constraints.

Models employing an integrated approach are quite common for the optimisation of WM tasks. Hemmelmayr et al. (2014) develop a model for the integrated planning of bin allocation and routing for waste collection. They first solve the two problems in sequence, and then in an integrated model, further proving that the integrated approach delivers better results. The aforementioned Olapiriyakul et al. (2019) develop a multi-objective mixed-integer problem for locating sorting centres, landfills and incinerators and for the transportation of waste. Kůdela et al. (2019) formulate a multi-objective two-stage mixed-integer stochastic programming problem for locating waste transfer stations and allocating the right capacity. Viktorin et al. (2023) apply hierarchical clustering to split the complex problem of waste bin location–allocation into several more solvable subproblems. In addition to multi-objective optimisation, integrated approaches can also be implemented through multilevel optimisation (He et al., 2011; Sharif et al., 2018; Caramia and Pizzari, 2022b) and fractional optimisation (Caramia and Pizzari, 2022a). As is clear from the references here reported, the waste collection problem has been tackled either as a transportation issue or a routing issue. Routing allows delving into details at the cost of increased complexity.

The VRP is an NP-hard problem; therefore, researchers have focused extensively on solution approaches, may they be exact or metaheuristic (Elshaer and Awad, 2020). Among the various methods in the literature, cluster first-route second logic is a fairly standard approach. The large set of customers is firstly divided into smaller groups of customers (clusters) not to exceed the overall capacity of a vehicle. Several attributes can be taken into account when defining the clusters. Afterwards, the vehicles are assigned to specific clusters and routing is executed. Beltrami and Bodin (1974) first introduced this method, using it to solve a problem concerning municipal waste collection. Barreto et al. (2007) implemented several hierarchical and non-hierarchical clustering techniques in a sequential heuristic algorithm. Hintsch and Irnich (2018) design a large multiple neighbourhood search which makes use of multiple clusters destroy and repair operators and a variable-neighbourhood descent for post-optimisation. Hintsch and Irnich (2020) design and analyse different branch-and-price algorithms for the exact solution of the soft-clustering VRP, i.e. a VRP with clusters of clients that can be visited only by one vehicle. Battarra et al. (2014) define two exact methods to solve the clustered VRP, i.e. a branch and cut and a branch and cut and price. Vidal et al. (2015) devise three metaheuristics, two of which are based on Iterated Local Search, while the third is a Hybrid Genetic Algorithm with a cluster-based solution representation. Defryn and Sörensen (2017) integrate a cluster level and a customer level local search phase in their metaheuristic, effectively exploiting the specific clustered structure to reduce complexity.

2.1. Main contributions

While OR scholars focused heavily on optimising WM problems, there are only a few contributions for multi-level and multi-stage models (Ghiani et al., 2014; Van Engeland et al., 2020). Some papers dealing with integrated approaches put on the same level decision problems belonging to different decision categories. Hemmelmayr et al. (2014) models both facility and routing, the former being a strategic task and the latter being an operational task. Kůdela et al. (2019) and Olapiriyakul et al. (2019) follow a similar logic by putting location and transportation, i.e. a strategic and tactical task, respectively, on the same level. Different categories of decision problems, perhaps even referring to different decision-makers in the real world, should be treated with multi-level or multi-stage models. These methods still allow pursuing integrated approaches without sacrificing adherence to reality.

Therefore, our purpose is to propose a WM network design two-stage model to achieve the recycling targets of the circular economy paradigm. Following the findings in the Literature, we included the computation of emissions due to a facility's presence and due to transportation to take care of the environmental pillar of sustainability. We adopted an integrated approach; we concurrently optimise location and allocation, considering information arising from the optimisation of the transportation problem.

This model aims to provide an optimal distribution of treatment capacity that limits waste movements while optimising the overall economic and environmental impact. In terms of theoretical implications, the main contributions of this paper to the existing literature are as follows:

- Integrated approach, i.e. two-stage model solved via a matheuristic that allows the exploration of several different starting points.
- Dynamic clusters. Most articles dealing with cluster first-route second (and its counterpart route first-cluster second) define clusters that cannot be changed. Moreover, these clusters are done without taking into account information arising from the routing. In our model, the clusters are updated via this information.
- The use of cluster first-route second logic in a two-stage model. Information from the routing (second stage) will be used as a new starting point for the clustering phase (first stage).
- Fairness of capacity allocation between facilities. The unbalanced capacity and the consequent economic and environmental impact has already been addressed in the Introduction.
- Penalty for capacity saturation. This feature has been included to consider uncertainties in waste streams and to better address the issue of unbalanced capacity.

The model is tested in a real-world case scenario, namely, the Lazio region, Italy. Although the real-case scenario application deals with an Italian region's situation, the model can provide several useful insights for other regions and countries. Moreover, the model mainly deals with Municipal Waste, but with the proper modifications, it is applicable to other scenarios.

3. Model definition and mathematical formulation

In this section, we introduce the studied WM problem, and next, we introduce its mathematical formulation.

3.1. Model definition

A regional authority, denoted for short as RA, is involved in solving a strategical problem, hereafter denoted as SP. The SP consists of opening a set of *collection facilities* in proper geographical points (drawn from a set F of prescribed locations). Waste must be routed after being picked up from *collection nodes*, i.e., *client nodes*, where waste and the resulting collection demand are generated. Let us denote this client set with C and demands with d_i, with $i \in C$.

Collection facilities act as hubs for the (spokes) client nodes. They may also be referred to as *sorting facilities* in case waste needs to be sorted in order to be finally routed to recycling plants, incinerators or landfills for their disposal. However, in this paper, we will not focus our attention on this downstream part of the supply chain, which requires studying another strategic problem.

Facilities to be opened by the RA have different sizes and capacities; let H be the set of sizes associated with a facility, e.g., H={small, medium, large} and let cap^f_{jh} be the capacity of a facility located in candidate node $j \in F$ with size $h \in H$. Depending on both location and size, facilities have different opening costs c_{jh}, with $j \in F$ and $h \in H$, and a different environmental impact in terms of CO_2 emissions em^f_{jh}, with $j \in F$ and $h \in H$.

The clustering phase in the strategic problem. In an attempt to decide which facilities have to be opened and which sizes have to be assigned to each of them, RA searches for a partition of set C into clusters, each with its own facilities. In doing so, RA aims at implementing a proximity logistic model where capacities are fairly allocated to opened facilities and, consequently, to clusters based on the following assumptions:

- h_1: demands d_i, with $i \in C$, are i.i.d. stochastic variables;
- h_2: the number of clients in each cluster is large enough to let the overall demand be served in each cluster to follow the Central Limit Theorem.

Consider a generic cluster; let $C' \subseteq C$ be the subset of clients belonging to this cluster. Moreover, let $d(C') = \sum_{i \in C'} d_i$ be the overall demand of the cluster. By hypotheses (h_1) and (h_2), $d(C')$ is a Gaussian variable with expected value $\mu(d(C'))$ equal to the sum of expected values of variables d_i, with $i \in C'$, and variance $\sigma^2(d(C'))$ equal to the sum of variances of the same variables.

With this setting, RA can define the probability (i.e., the service level) with which an assignment of clients to a cluster will be obeying the cluster capacity, the sum of the capacities of the facilities therein assigned.

For instance, further considering the example of a cluster with a set C' of assigned clients, in order to guarantee a service level of 0.99, RA may impose 2.33 standard deviations of $d(C')$ from the expected demand of the cluster $\mu(d(C'))$ to determine the minimum capacity to be assigned to that cluster.

The hypothesis in h_1 related to the independence of client demands appears to be realistic, since clients produce waste independently one to each other; hypothesis in h_1 for which client demands are identically distributed appears to be realistic in those scenarios in which clients are represented by, e.g., comparable sets of families or commercial activities. In the case in which clients are substantially heterogeneous and, seemingly, it would be quite unrealistic to assume identical distributions, we can properly relax this assumption as follows (we will denote this option as option (b), as opposed to the current option (a)). Client demands d_i, with $i \in C$, are modelled as (deterministic) parameters such that $d_i = \mu d_i$, i.e., each demand is set equal to its expected value, and we consider a safety capacity sc on the capacity of a cluster, say, for instance, $\bar{cc}$, to reserve and penalise in case of being used. By means of sc, we model the action of the previously defined standard deviations without making use of any assumption on the client demand stochastic variables. Formally, we have

$$\sum_{i \in C'} d_i + (1-\eta) \cdot sc \leq \bar{cc}, \tag{1}$$

where η is a continuous variable which should be minimised to let it be as close as possible to 0 to let, in turn, the safety capacity to be untouched. In the worst case $\eta = 1$, we will be using all the safety capacity and, therefore, the cluster will be assigned a capacity only on the basis of the expected values of the demands, which would correspond to a service level of 0.5 in the case of identical distribution of the client demands.

The operational problem nested inside the strategic problem. After opening the facilities, RA cannot play any role in determining how clients will be served and which facilities will be used. Moreover, it appears not to be realistic to impose clusters at an operational routing level since carriers are free to choose routes based on their minimum cost criterion. Therefore, to minimise the risk of failure in the clustering task, RA tries to infer which will be the best route to serve clients from the carrier's point of view, given the set of opened facilities; this task requires that RA nests a new optimisation problem in their problem formulation. This new problem will be denoted as OP. The optimal solution of OP allows the calculation of miss-clustered clients to be penalised in the SP full objective function, i.e., a penalty arises in the case a client allocated in a cluster and to be served by the associated facilities in that cluster, is instead served by a different facility in the routing solution offered by OP. The different nature of both SP and OP leads to the definition of a two-stage optimisation problem (denoted as TSOP) that represents well the interactions between the two problems. In order to fully define TSOP, we first have to consider OP in more detail; to this end, we need to introduce some more notation.

Let $G = (N, A)$ be a graph modelling the geographical area of the problem, where N is the set of nodes that encompasses three different subsets, i.e., $N = C \cup F \cup D$, being C the subset of node clients, F the subset of sites where facilities can be opened, and D the subset of depot nodes. Let t_{ab} be the travelling time between nodes $a \in N$ and $b \in N$. There is a set V of vehicles in charge of collecting wastes from clients in C. Each vehicle $l \in V$ has a capacity cv_l. The demand d_i of each client $i \in C$ has to be served by vehicles $l \in V$. Each depot, in turn, is associated with a set of vehicles. Matrix $A \in \mathbb{Z}^{|D| \times |V|}$ stores this allocation of vehicles to depots in such a way that its generic element a_{kl} is equal to 1 if vehicle l starts its tour from depot k, and holds 0 otherwise. Each node must be served by a fleet of vehicles associated with a single depot. In OP we have to find tours of vehicles in such a way that a vehicle $l \in V$ starts its collection tour from a depot $k \in D$, serves a subset of the clients in C obeying its capacity cv_l and then goes to an opened facility $j \in F$ of size $h \in H$. Finally, it routes back to the same depot $k \in D$. A service operated by a vehicle $l \in V$ should respect a maximum servicing time per tour, denoted as T_l.

The objective function of OP is to minimise the total travel time of all vehicles.

As mentioned above, the regional authority first decides which facilities to open by minimising emissions from the installation of facilities and minimising predicted capacity saturation. Afterwards, the authority uses this information to infer the optimal routing in the second stage, which gives feedback on the eventual miss-clustering and on the actual capacity saturation. Therefore, it makes sense to have two different objective functions for SP, namely, a reduced objective function computed in the first stage and a full objective function computed after the second stage.

The reduced objective function of SP is to minimise the conic combination of two functions to be minimised:

- (i) emissions due to installation of facilities
- (ii) penalty associated with the usage of the safety stock capacity of a facility

The full objective function of SP is to minimise the conic combination of three functions to be minimised:

- (i) emissions due to installation of facilities and transportation
- (ii) penalty associated with the usage of the safety stock capacity of a facility
- (iii) penalty associated with a possible miss-match between the clustering solution, represented by the solution of SP, and the assignments clients/routes, offered by the solution of OP.

Fig. 1 briefly displays the interaction between the two models. In the figure, *OF* shortens *objective function*.

3.2. The mathematical formulation

After the definition of the problem, we now give the mathematical formulation of TSOP. In order to get to the overall formulation, we will separately present the constraints of SP and OP.

The strategical problem constraints. Let S be the set of clusters the municipal firm may define, i.e., $|S|$ is an upper bound on the number of non-empty clusters defined by the problem solution. Let sc_{jh} be the safety capacity associated with facility $j \in F$ of size $h \in H$. The decision variables of SP are:

$$x_{is} = \begin{cases} 1 \text{ if client } i \in C \text{ is assigned to cluster } s \in S \\ 0 \text{ otherwise} \end{cases}$$

$$r_{jhs} = \begin{cases} 1 \text{ if facility } j \in F \text{ of size } h \in H \text{ is assigned to cluster } s \in S \\ 0 \text{ otherwise} \end{cases}$$

$$y_j = \begin{cases} 1 \text{ if facility } j \in F \text{ has been opened} \\ 0 \text{ otherwise} \end{cases}$$

$\eta_{jhs} \in [0,1]$ = the fraction of used safety capacity sc_{jh} associated with facility $j \in F$ of size $h \in H$ assigned to cluster $s \in S$

The constraints of SP are described in the following; note that, in defining the capacity of each cluster, they encompass option (b) (Eq. (1)).

$$\sum_{s \in S} x_{is} = 1, \quad \forall i \in C, \tag{2}$$

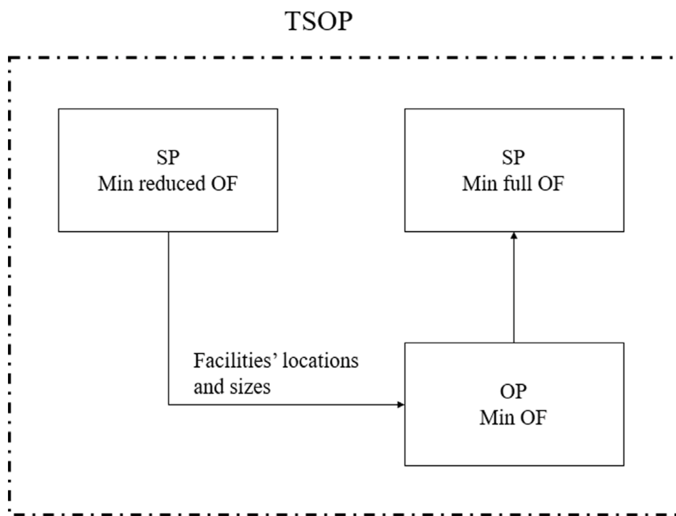

Fig. 1. The two-stage optimisation model.

$$\sum_{s\in S}\sum_{h\in H} r_{jhs} = y_j, \quad \forall j \in F, \tag{3}$$

$$\sum_{i\in C} d_i \cdot x_{is} \le \sum_{j\in F}\sum_{h\in H} (r_{jhs}\cdot cap^f_{jh} - (r_{jhs}-\eta_{jhs})\cdot sc_{jh}), \quad \forall s \in S, \tag{4}$$

$$x_{is} \le \sum_{j\in F}\sum_{h\in H} r_{jhs}, \quad \forall i \in C,\ \forall s \in S, \tag{5}$$

$$\eta_{jhs} \le r_{jhs}, \quad \forall j \in F, h \in H, s \in S, \tag{6}$$

$$\sum_{j\in F}\sum_{h\in H}\sum_{s\in S} c_{jh}\cdot r_{jhs} \le B, \tag{7}$$

$$\sum_{j\in F}\sum_{h\in H} r_{jhs} \le mc, \quad \forall s \in S, \tag{8}$$

$$\eta_{jhs} \in [0,1] \quad \forall j \in F, h \in H, s \in S \tag{9}$$

$$x_{is}, r_{jhs}, y_j \in \{0,1\} \quad \forall i \in C,\ \forall j \in F,\ \forall h \in H,\ \forall s \in S. \tag{10}$$

Constraints (2) say that every client $i \in C$ must be assigned to one cluster in S. Constraints (3) impose that a facility $j \in F$ can have at most one size $h \in H$ and can be allocated to at most one cluster $s \in S$; moreover, in the case that facility $j \in F$ of size $h \in H$ is assigned to cluster $s \in S$, facility j is opened, i.e., $y_j = 1$. Constraints (4) guarantee that the overall demand of clients assigned to a cluster $s \in S$ (considering demands as given parameters, as detailed in Section 3) must not exceed the sum of the capacities cap^f_{jh} of all facilities $j \in F$ (each with its own size h) assigned to cluster s minus the fractions η_{jhs} of the safety capacities sc_{jh} of the same facilities. Constraints (5) assign at least one facility to a cluster $s \in S$ if at least one client $i \in C$ is assigned to that cluster s. Constraints (6) impose that if no facility $j \in F$ is opened with size $h \in H$ and assigned to cluster $s \in S$, then no safety stock can be used. Constraint (7) imposes that the cost of locating facilities cannot exceed a given budget B. Constraints (8) place a limit on the maximum amount mc of facilities that can be assigned to a cluster. Constraints (9) and (10) define the range of feasible values for the decision variables.

The operational problem constraints. Let us now define the constraints of OP. Let

$$h_l^a = \begin{cases} 1 \text{ if vehicle } l \in V \text{ visits node } a \in N \text{ during its tour} \\ 0 \text{ otherwise} \end{cases}$$

$$z_l^{ab} = \begin{cases} 1 \text{ if vehicle } l \in V \text{ travels from node } a \in N \text{ to node } b \in N \\ 0 \text{ otherwise} \end{cases}$$

$\theta_l^i \ge 0$; the amount of waste picked up by vehicle $l \in V$ from node $i \in C$

$p_l^a \ge 0$; the arrival time of vehicle $l \in V$ in node $a \in N$

$v_l^j \ge 0$; the total amount of waste transported by vehicle $l \in V$ to facility $j \in F$

Constraints are as follows:

$$\sum_{j\in F} h_l^j = \sum_{k\in D}\sum_{i\in C} z_l^{ki}, \quad \forall l \in V, \tag{11}$$

$$\sum_{k\in D} h_l^k = 1, \quad \forall l \in V, \tag{12}$$

$$\sum_{j\in F} h_l^j = 1, \quad \forall l \in V, \tag{13}$$

$$\sum_{l\in V} \theta_l^i = d_i, \quad \forall i \in C, \tag{14}$$

$$h_l^i + h_r^i \le a_{kl} + a_{kr}, \quad \forall i \in C, k \in D l \in V, r \in V : l \ne r \ \&\ h_l^k = 1, \tag{15}$$

$$z_l^{ij} = 0, \quad \forall l \in V, i \in N, j \in N : i = j \tag{16}$$

$$\sum_{i\in C} z_l^{ki} \le a_{kl}, \quad \forall k \in D,\ \forall l \in V, \tag{17}$$

$$\sum_{j\in F} z_l^{jk} \le a_{kl}, \quad \forall k \in D,\ \forall l \in V, \tag{18}$$

$$\sum_{i\in C} z_l^{ji} = 0, \quad \forall j \in F, l \in V, \tag{19}$$

$$\sum_{k\in D} z_l^{jk} = \sum_{i\in C} z_l^{ij}, \quad \forall j \in F, l \in V, \tag{20}$$

$$\sum_{j\in C+F} z_l^{ij} = h_l^i, \quad \forall i \in C, l \in V, \tag{21}$$

$$\sum_{j\in D+C} z_l^{ji} = h_l^i, \quad \forall a \in C, l \in V, \tag{22}$$

$$\sum_{i\in C} \theta_l^i \le cv_l, \quad \forall l \in V, \tag{23}$$

$$\theta_l^i \le M \cdot h_l^i, \quad \forall l \in V, i \in C, \tag{24}$$

$$\theta_l^i \ge \sigma \cdot d_i \cdot h_l^i, \quad \forall l \in V, i \in C, \tag{25}$$

$$p_l^b = p_l^a + t_{ab}, \quad \forall l \in V, a \in D \cup C, b \in C \cup F : z_l^{ab} = 1, \tag{26}$$

$$\sum_{l\in V}\sum_{k\in D} p_l^k = 0, \tag{27}$$

$$p_l^a = 0, \quad \forall l \in V, a \in N : h_l^a = 0, \tag{28}$$

$$\sum_{j\in F} p_l^j \le T_l, \quad \forall l \in V, \tag{29}$$

$$\sum_{i\in C} \theta_l^i = v_l^j, \quad \forall j \in F, l \in V : h_l^j = 1, \tag{30}$$

$$\sum_{l\in V} v_l^j \le \sum_{h\in H}\sum_{s\in S} r_{jhs}\cdot cap_{jh}^f, \quad \forall j \in F, \tag{31}$$

$$v_l^j = 0, \quad \forall j \in F, l \in V : h_l^j = 0, \tag{32}$$

$$h_l^i \in \{0,1\}, \quad \forall i \in N, l \in V, \tag{33}$$

$$z_l^{ij} \in \{0,1\}, \quad \forall i,j \in N, l \in V, \tag{34}$$

$$p_l^a \ge 0, \quad \forall a \in N, l \in V \tag{35}$$

$$v_l^j \ge 0, \quad \forall j \in F, l \in V. \tag{36}$$

Constraints (11) say that if vehicle $l \in V$ travels from depot $k \in D$, i.e., $\sum_{k\in D}\sum_{i\in C} z_l^{ki} = 1$, then there must exist exactly one facility $j \in F$ visited by vehicle l on its tour. Constraints (12) impose that each vehicle $l \in V$ must have exactly one depot $k \in D$ on its route (that is, the one for which $a_{lk} = 1$), while constraints (13) impose that each vehicle must have exactly one facility on its route. Constraints (14) impose the demand of each node to be satisfied. Constraints (15) impose that multiple vehicles can visit the same node if they belong to the same deposit. Constraints (16) do not permit loops on the same node. Constraints (17) define that a vehicle $l \in V$ in a depot $k \in D$, i.e., one for which $a_{kl} = 1$, can exit from depot k to visit as immediate successors only (client) nodes $i \in C$ (i.e., facility or depot nodes are not allowed). Constraints (18) define that a vehicle leaving a facility $j \in F$ can visit only one deposit $k \in D$ afterwards, namely the one for which $a_{kl} = 1$. Constraints (19) guarantee that a vehicle $l \in V$ cannot visit a client node $i \in C$ after visiting a facility node $j \in F$. Constraints (20) define that if a vehicle $l \in V$ has visited a facility $j \in F$, the next node visited by vehicle l must be its depot. Constraints (21) guarantee that after visiting a client, we can go to another client or to a facility, while constraints (22) state that we can enter a client node only from a deposit or another client. Constraints (23) impose to respect the maximum capacity for each vehicle. Constraints (24) and (25) state that a vehicle can collect waste from a node if it visits the node and picks up a minimum amount σ, respectively. Constraints (26)–(29) regulate the behaviour of variable p_l^a, i.e., the arrival time of vehicle l in node a. In detail, Constraints (26) compute the arrival time at node b from node a as the arrival time at node a plus the time needed to traverse arc (a,b) only if arc (a,b) is on the route of vehicle l. Constraint (27) initialises the arrival time for each vehicle to be zero at each deposit. Constraints (28) set the arrival time to zero for each node not visited. Finally, Constraints (29) put a limit to the time length of a route ending in a facility. Constraints (30) compute the overall load of a vehicle $l \in V$, linking this information with the facility $j \in F$ visited during the route. Constraints (31) are the capacity constraints for the facilities; they ensure that the overall load of the vehicles does not exceed the installed capacity. Constraints (32) put the load towards the facilities not visited equal to zero. Constraints (33)–(36) state the feasible domain of the decision variables.

The operational problem objective function. In OP, we aim to minimise the sum of the vehicles' travel times. This means that the objective function associated with OP is:

$$of_{\mathtt{OP}} : \min \sum_{a\in D+C}\sum_{b\in C+F}\sum_{l\in V} z_l^{ab}\cdot t_{ab}. \tag{37}$$

The strategic problem objective function. We try to give a more in-depth understanding of the term (iii) mentioned at the end of Section 3; to this end, in Figs. 2 and 3, we illustrate a gadget example of an instance of our problem.

In the first figure (upper part), we report a gadget graph $G = (N, A)$ with $C = \{1,2,3,4,5,6,7\}$, F={Facility 1, Facility 2}, and D={Depot 1, Depot 2}. In the lower part of Fig. 2, we report a clustering solution consisting of two clusters, i.e., Cluster 1, including clients 1, 2, and 3, and Cluster 2, including clients 4, 5, 6, and 7.

Fig. 3 shows (upper part) an example of routing operated by the company in charge of collecting the waste demands of the network, i.e., we have Routing 1, which starts from Depot 1, visits clients 2, 5, 4, and 7, and, finally, reaches Facility 2; and Route 2, which starts from Depot 2 and visits clients 1, 3, and 6, and, finally, reaches Facility 1. Both routes, after visiting their respective facilities, end up at the initial depot.

Looking at the lower part of Fig. 3, we see that there is a misalignment between the cluster composition and the clients in each route. In particular, clients 1, 2, and 3, belong to the same cluster but are not served on the same route. A similar situation occurs for clients 4, 5, 6, and 7, which belong to cluster 2 but are not served on the same route. Therefore, the unmatched assignments, e.g., those related to clients 2 and 5 (same route, different clusters), or those related to clients 2 and 6 (again, same route, different clusters), will activate a penalty term in the objective function of SP. The full objective function of SP is, therefore, the following:

$$\begin{aligned} of_{\mathtt{SP}} : \min \quad & \gamma_1 \cdot \left(\sum_{i\in N}\sum_{j\in N}\sum_{l\in V} z_l^{ij}\cdot em_{ij}^t + \sum_{j\in J}\sum_{h\in H}\sum_{s\in S} em_{jh}^f \cdot r_{jhs}\right) + \\ & \gamma_2 \cdot \sum_{j\in F}\sum_{h\in H}\sum_{s\in S} P_j \cdot \eta_{jhs} + \\ & \gamma_3 \cdot \sum_{i\in C}\sum_{i'\in C}\sum_{l\in V}\sum_{s\in S} \max\{0, z_l^{ii'} - x_{is}\cdot x_{i's}\}, \end{aligned} \tag{38}$$

where γ_1, γ_2, and γ_3 are real scalars. The first piece of the function defines the overall emissions, given by the sum of the emissions caused

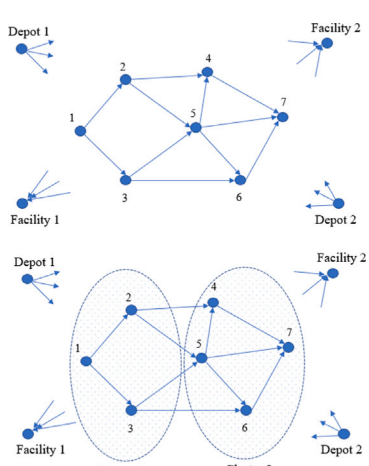

Fig. 2. An example of a network and a clustering.

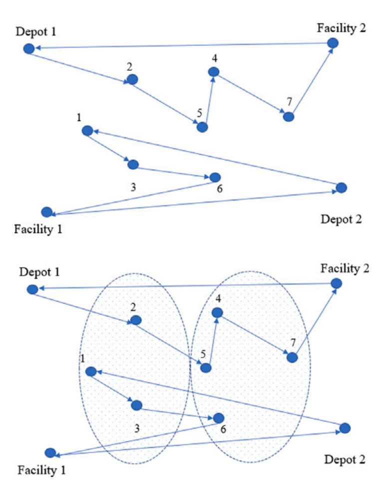

Fig. 3. An example of routing and the associated penalty occurrence.

by transportation (the first term in parentheses, where em^t_{ab} is the emission for each arc (a,b)) and the sum of the emissions caused by the facility locations and their assigned size (second term in parentheses). The second piece of the function models the amount of penalty $\eta_j \cdot P_j$ to be paid in case a fraction $0 \leq \eta_j \leq 1$ of safety capacity sc_j is used in a facility j, $\forall j \in F$, P_j being a real scalar that defines the maximum penalty in a facility $j \in F$. The third term of of_{SP} refers to the penalty defined at the beginning of this paragraph, i.e., the penalty occurring when two clients are assigned to the same route in OP but to different clusters in SP. It can be linearised by introducing the following:

- a real non negative variable $g_{ii'ls}$, with $i \in C$, $i' \in C$, $l \in V$, $s \in S$,
- a binary variable $u_{ii's}$, with $i \in C$, $i' \in C$, $s \in S$,

and by adding the following constraints in SP:

$$g_{ii'ls} \geq 0, \quad \forall i \in C,\ \forall i' \in C,\ \forall l \in V,\ \forall s \in S, \tag{39}$$

$$g_{ii'ls} \geq z_l^{ii'} - u_{ii's}, \quad \forall i \in C,\ \forall i' \in C,\ \forall l \in V,\ \forall s \in S, \tag{40}$$

$$u_{ii's} \geq x_{is} + x_{i's} - 1, \quad \forall i \in C,\ \forall i' \in C,\ \forall s \in S, \tag{41}$$

$$u_{ii's} \leq \frac{x_{is} + x_{js}}{2}, \quad \forall i \in C,\ \forall i' \in C,\ \forall s \in S, \tag{42}$$

$$u_{ii's} \in \{0,1\}, \quad \forall i \in C,\ \forall i' \in C,\ \forall s \in S, \tag{43}$$

and further rewriting the objective function as follows:

$$of_{\mathrm{SP}} : \min \quad \gamma_1 \cdot \left(\sum_{i\in N} \sum_{j\in N} \sum_{l\in V} z_l^{ij} \cdot em^t_{ij} + \sum_{j\in J} \sum_{h\in H} \sum_{s\in S} em^f_{jh} \cdot r_{jhs} \right) + \gamma_2 \cdot \sum_{j\in F} \sum_{h\in H} \sum_{s\in S} P_j \cdot \eta_{jhs} + \gamma_3 \cdot \sum_{i\in C} \sum_{i'\in C} \sum_{l\in V} \sum_{s\in S} g_{ii'ls}, \tag{44}$$

For clarity, we present a more compact version of the objective function, where we name the three different parts of the full objective function.

$$of_{\mathrm{SP}} : \min \quad \gamma_1 \cdot of^1_{\mathrm{SP}} + \gamma_2 \cdot of^2_{\mathrm{SP}} + \gamma_3 \cdot of^3_{\mathrm{SP}} \tag{45}$$

As mentioned earlier, the first-stage model also employs a reduced objective function, which is the following:

$$of'_{\mathrm{SP}} : \min \quad \gamma_1 \cdot \sum_{j\in J} \sum_{h\in H} \sum_{s\in S} em^f_{jh} \cdot r_{jhs} + \gamma_2 \cdot \sum_{j\in F} P_j \cdot \eta_j \tag{46}$$

4. A matheuristic solution approach

Decisions taken in the first stage of the model affect decisions taken in the second stage. Furthermore, the second-stage decisions affect the full objective function of the regional authority. If we employ a simple two-stage scheme, i.e. the first stage is optimally solved, then the second stage is optimally solved, and finally, we compute the overall results, we may select several shortsighted decisions. Therefore, we developed a matheuristic in order to provide the RA with better solutions. The proposed matheuristic combines a local search and a tabu search logic. The overall framework is displayed in Fig. 4, where the reader can easily recognise three main sectors, i.e. sectors A, B, and C.

At the beginning of sector A, the RA solves SP on the reduced objective function. Information regarding opened facilities and their size is then passed to the second stage, which is then solved. After OP is solved, RA has the information it needs to calculate the full objective function. At the end of sector A, the RA checks if there are facilities whose safety stock is being used, i.e. if there are facilities that are overly used by the OP.

Sector B comprises all the possible actions that could be taken. The authority selects the facility that struggles the most, i.e. has the most amount of safety stock used. The authority first tries to enlarge the under-sized facility. If it is not possible, the authority checks if there are facilities whose capacity is used under a certain threshold, eventually closing them and replacing them with other facilities. The reason behind this is that facilities under-used may not be attractive to second-stage routing. Therefore, the RA tries to open other facilities, which may be more attractive. If there are no under-used facilities, the RA opens an additional facility in order to try to assist the one in need.

In sector C, the action undertaken is then passed on to the second stage, which is again solved. Finally, the authority is able to infer if the new combination of facilities and sizes yields better results for the full objective function. If it is so, the authority continues the matheuristic. If the solution is worse than the one already obtained, the change in facility location/size is reverted and added to a tabu list to avoid making the same decision. When there is no more room for improvement, i.e. all facilities are acceptably used, or if there are no available moves, i.e. the tabu list contains all the possible moves from the current solution, then the matheuristic stops.

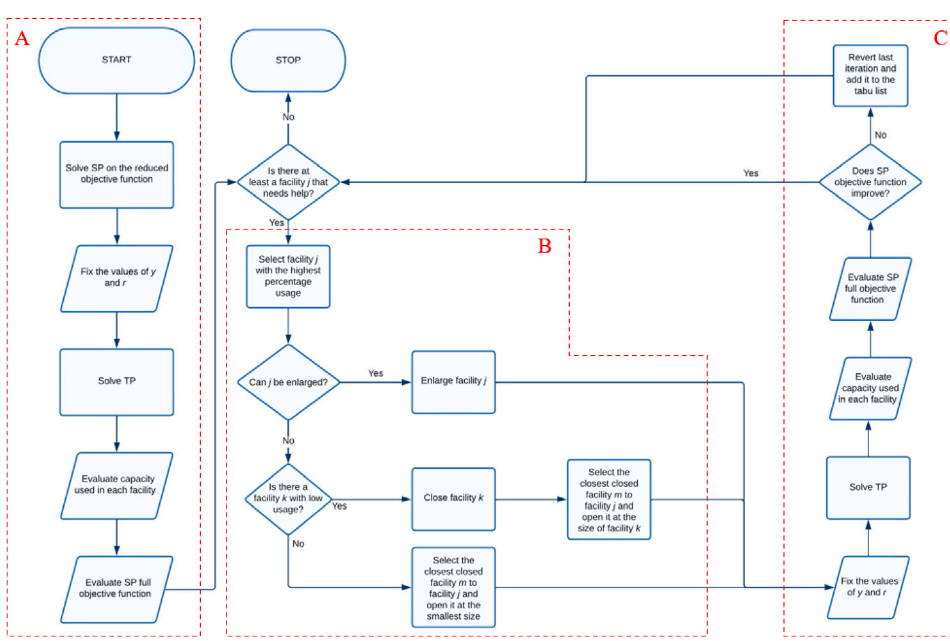

Fig. 4. Framework of the matheuristic.

Table 1
Random instances - details.

Instance	Facilities	Clients	Deposits	Vehicles	SP Time lim (s)	OP Time lim (s)
Small	5	15	2	10	25	200
Medium	10	25	5	20	35	400
Large	15	40	8	30	50	800

Table 2
SP results for random instances.

Size	Instance	First value	Best value	Improvement
Small	1	290.73	290.73	0.00%
	2	728.55	728.55	0.00%
	3	375.66	375.66	0.00%
	4	323.94	323.94	0.00%
	5	286.68	286.68	0.00%
Medium	6	445.42	420.37	5.62%
	7	665.59	566.28	14.92%
	8	380.30	374.98	1.40%
	9	585.25	421.04	28.06%
	10	811.74	764.48	5.82%
Large	11	561.64	561.64	0.00%
	12	706.74	667.12	5.61%
	13	1,002.92	701.79	30.03%
	14	970.54	876.72	9.67%
	15	658.74	599.46	9.00%

5. Computational analysis and case study

Both the model and the matheuristic are coded in Python3 programming language. All SP and OP problems instances are solved via branch-and-cut using the Gurobi 9.5.2 solver on a PC running an AMD Ryzen 7 4800H Processor with 16 GB of RAM. In order to test the behaviour of the two-stage model and of the matheuristic before applying them to a real-world case study, we created a set of random instances of different sizes. Table 1 displays the number of facilities, clients, deposits, and vehicles for each size of the instances, as well as the solver time limit for each stage of the model (expressed in seconds). For each size, we developed 5 different instances. For what concerns the network nodes of each instance, these are the results of different random samples of real locations of waste disposal facilities, industrial waste clients and depots of waste companies' trucks fleets. These types of nodes are all located within the Lazio region of centre-Italy. Dealing with data from a real-case scenario, the expected waste demand of served clients is computed according to 5 years (2017–2022) of waste pick-up instances. Graph edges attributes, such as travel distance and duration, are generated by Open Source Routing Machine (OSRM), a high-performance routing engine written in C++14 designed to run on OpenStreetMap data (Luxen and Vetter, 2011). For the first-stage model, the values of the weights are $\gamma_1 = 0.3$, $\gamma_2 = 0.3$, and $\gamma_3 = 0.4$.

Table 2 displays the results for each random instance. Specifically, it shows the value of the full SP objective function computed in the first iteration and the best value of the objective function computed according to multiple starting points provided by the matheuristic. The improvement column shows that the proposed approach improved the results with respect to the initial solution for most instances. For the few instances not improved, the matheuristic did not start, meaning that the allocation of capacity done at the SP stage was proved acceptable by the OP. Table 3 shows how the SP objective function values evolve across the iterations. The underlined values highlight the instance with the best value. The results certify that the matheuristic is able to improve upon the original solution. A limit of 10 maximum iterations bounds the matheuristic test. Instance 14 is the only instance for which the solver was not able to explore all the possible starting points in less than 10 iterations.

Table 4 displays the values of the second stage model for each iteration. Underlined results do not represent the best OP values, they instead represent the iteration the RA settles on (i.e. the iteration for which the value of SP is the lowest). The SP problem was always solved at the optimum, while the OP is a more complex problem and its optimal solution is not always found in a reasonable time (Table 5). All nonzero gap values of Table 5 refer to the solver time limits previously reported in Table 1.

5.1. Case study

The case study focuses on the Lazio region, located in the centre of Italy. Lazio region has an extension of $17,242\ \text{km}^2 (6,657\ \text{mi}^2)$ and a population of $5,864,321$. Considering such a large area is indeed a challenging network design problem. Therefore, only some municipalities are selected for each district. The Lazio region has 5 districts: Rome, Viterbo, Rieti, Latina, and Frosinone. For each district, the provincial county seat and the most populated municipalities are considered in the use case application. Rome is excluded from this case study due to its size as a metropolitan city, which implies specific management approaches. Several facilities are selected for each province, and all data regarding waste production, population, and processing capacity of each possible facility have been taken from ISPRA's *Catasto dei rifiuti* (Waste cadastre) (ISP). We decided to focus on data concerning paper waste generation. Although, we remind the reader that the model is also applicable to other types of municipal waste. We selected a total of 40 municipal clients and 10 possible facilities. The 40 municipalities are inhabited by 1.4 million citizens, circa, and generate a daily demand of 180 tonnes of paper waste. They can be served by a fleet of 60 vehicles distributed across 10 depots. Fig. 5 displays the selected municipal clients (green), the facilities (orange) and the truck depots (blue). Once again, the values of the weights of the first-stage objective function are $\gamma_1 = 0.3$, $\gamma_2 = 0.3$, and $\gamma_3 = 0.4$.

The optimal use-case solution that results from employing the matheuristic is described in the following. In particular, Table 6 shows the number of opened facilities, their size, and the used capacity.

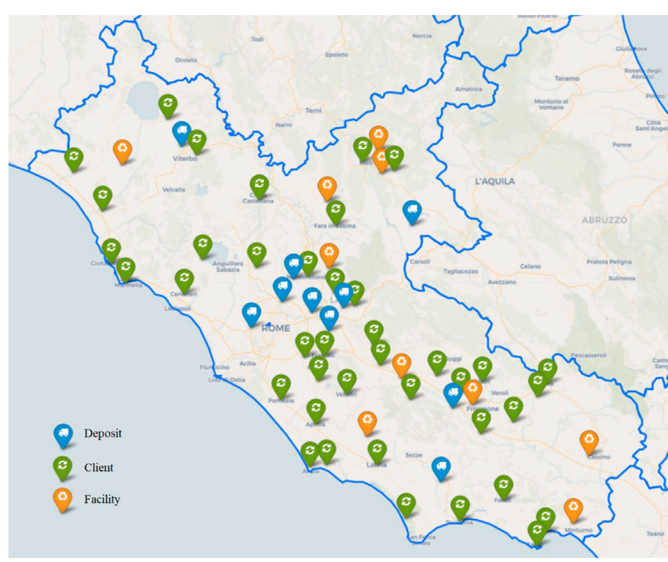

Fig. 5. Real case instance with vehicle depots, facilities, and demand points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Evolution of SP objective function.

Instance	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8	Iter 9	Iter 10
1										
2										
3										
4										
5										
6	420.37									
7	566.28									
8	374.98									
9	541.35	421.04	428.35							
10	764.48									
11										
12	700.12	702.24	667.12							
13	964.37	736.78	722.81	701.79						
14	967.56	898.16	943.42	959.75	950.09	944.73	876.72	940.24	951.08	928.98
15	677.86	635.46	674.71	635.55	641.96	599.46	599.46			

Table 4
OP results for random instances.

Inst.	First Value	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8	Iter 9	Iter 10
1	494.40										
2	1099.30										
3	753.54										
4	562.85										
5	644.45										
6	744.44	744.21									
7	1175.73	1051.23									
8	686.77	688.06									
9	1,169.15	1,204.92	946.70	946.70							
10	1,295.61	1,267.07									
11	877.42										
12	1,262.58	1,289.84	1,224.43	1,219.35							
13	1,445.40	1,312.24	1,015.54	1,015.54	1,026.31						
14	1,368.77	1,318.06	1,351.42	1,410.06	1,405.73	1,491.80	1,421.46	1,345.72	1,406.63	1,396.67	1,462.77
15	1,322.35	1,340.50	1,271.44	1,336.25	1,295.04	1,304.51	1,262.00	1,262.00			

Table 5
Percentage gap of OP for each iteration.

Instance	First Value	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8	Iter 9	Iter 10
1	0.00%										
2	0.44%										
3	0.80%										
4	0.00%										
5	6.04%										
6	0.30%	0.27%									
7	23.33%	14.25%									
8	5.57%	3.89%									
9	0.86%	3.10%	4.73%	4.73%							
10	0.44%	2.01%									
11	8.38%										
12	8.90%	10.25%	8.88%	7.83%							
13	30.98%	23.97%	4.02%	4.03%	6.26%						
14	6.82%	3.42%	2.05%	1.88%	1.57%	7.34%	2.38%	2.53%	1.64%	0.94%	5.50%
15	8.65%	9.88%	9.72%	9.73%	9.25%	9.91%	6.59%	6.59%			

Table 7 shows the values of the SP objective function, as well as the evolution of the matheuristic's iterations. The time limit for solving the SP is 50 s, while the time limit for the OP is 10 000 s. The limit of 10 maximum iterations does not bind the matheuristic implementation. Table 8 shows the solution results w.r.t. the OP; the first row is the result, while the second row is the optimality gap. The number in bold corresponds to the iteration in which the regional authority settles, i.e., the eighth iteration, which is the one with the best result for RA, as shown in Table 7. Although the value of the OP in the eighth iteration improves over the first solution, the second stage model could settle for an even better result in the sixth iteration.

Table 6
Open facilities and capacity allocation in use case application.

Facility	Size	Capacity used (%)	Safety stock used (%)
1	Large	74.00%	13.33%
5	Medium	49.44%	0
6	Small	50.19%	0
8	Large	44.76%	0
9	Medium	60.12%	0

Table 7
Results of SP in the case study.

First	Best	Impr.	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8
1,977.94	1,782.79	9.87%	1,953.89	1,835.74	1,855.54	1,957.97	1,980.45	1,788.89	1,986.36	**1,782.79**

Table 8
Results of OP in the case study.

First value of OP	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8
4,000.36	4,011.35	4,088.7	3,856.5	3,924.43	3,944.35	3,800.89	4,010.88	**3,822.37**
4.16%	4.00%	5.12%	4.55%	6.12%	5.09%	5.12%	9.42%	5.95%

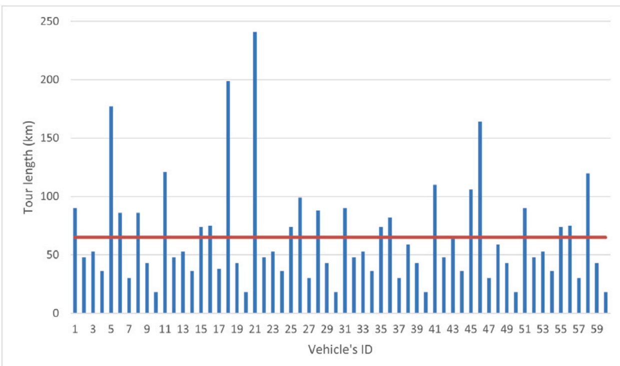

Fig. 6. Tour lengths of the vehicles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 6 showcases the tour lengths of the vehicles in km for each of the sixty considered vehicles. The red line shows the average tour length.

The first solution found opens four facilities, mainly located in the central area of the region. The model ultimately settles on changing a subset of these facilities and on opening a fifth additional facility to better deal with demand distribution and allocated capacity. The average tour length of the vehicles is relatively low, with some vehicles being barely used. This may be caused by the homogeneous capacity of the fleet's vehicles. Since each vehicle has the same capacity, those visiting higher-demand clients are unlikely able to serve additional nodes, thus decreasing the overall length of their routes.

The relatively high computational time is in line with the objective of network optimisation, i.e., a strategic task concerning facility location and capacity allocation. These long-term decisions consider the effects they have over those regarding the operational level, from which they are able to gain insights to further improve upon the original shortsighted network configuration.

5.2. Sensitivity analyses

To further test the model, we conducted two sensitivity analyses concerning the effects of small perturbations on the capacities of the facilities and the weights assigned to the strategic problem objective function, respectively.

Hence, as per the first sensitivity analysis, we varied the capacities $c\bar{a}p^{f}_{jh}$ of each facility j of size h in the range $[0.8 \cdot c\bar{a}p^{f}_{jh}, 1.2 \cdot c\bar{a}p^{f}_{jh}]$, with a step equal to 0.05, thus producing nine different scenarios (see Table 9). Clearly, the fifth scenario is the same as the one discussed in the Case Study subsection. We highlight which facilities are open and their size $h \in \{S, M, L\}$. We also present two percentage values, i.e., the percentage of capacity used and the percentage of safety stock used, respectively.

At a glance, facilities 0 and 4 are never opened. On the contrary, facilities 5, 8, and 9 are always open. Therefore, these potential facilities are critical for the WM network. The decision-maker should focus on the efficiencies produced by the most used facilities while studying the reasons why facilities 0 and 4 are never considered.

For the second sensitivity analysis, we changed the values of the weights of each objective function's components, being γ_1, γ_2, and γ_3. We first analysed the case where only one of the weights is equal to 1 (and the others are equal to zero), thus finding the ideal point $\mathtt{id} = (5,698.29, 40.75, 132)$, i.e., the point with coordinates equal to the lowest possible values for each component of of_{SP}.

Table 10 reports different configurations of weights and the resulting values for each component of of_{SP}, being of^1_{SP}, of^2_{SP}, and of^3_{SP}. The first row of Table 10 displays the case-study configuration of weights and the following rows report the cases where these change between 0 and 1 with a step equal to 0.25, and $\gamma_1 + \gamma_2 + \gamma_3 = 1$. We also compute the overall value of of_{SP} and the distance $\mathtt{dist}$ (the Euclidean norm) from the ideal point.

5.3. Managerial implications

The proposed two-stage model and matheuristic effectively handle the two different decision-makers. The matheuristic is capable of improving on the original first solution for RA, also providing solutions with a lower total travel time for OP, although certain network configurations (matheuristic iterations) allow even better results for the latter. One of the inherent objectives of the matheuristic can be appreciated in terms of capacity allocation. Accordingly, the values reported in Table 6 highlight that only one facility uses safety stocks, albeit in a minimal percentage. The matheuristic also allows scouting potential facilities location, as displayed in the "Sensitivity Analyses" sub-section. Being able to locate critical infrastructure is a great asset. Unsurprisingly, the opened facilities are closer to the centre of the region, which is the most densely inhabited part of Lazio and of our instance. The homogeneous capacity of the fleet's vehicles influences the low average length. The vehicles visiting higher-demand clients are unlikely able to serve additional nodes, thus decreasing the overall length of their routes. Considering a tailored heterogeneous fleet may assess this issue at the cost of increased complexity.

One of the main understandings is the concurrent optimisation of the three objective functions. Moreover, we believe they should hold relatively similar importance. Concurrent optimisation of tasks is vital when managing supply chains, especially with a focus on sustainability. Sustainability is, in fact, a multi-faceted concept comprising economic, environmental, and social aspects that must be dealt with together. Failing to do so results in sub-optimal decisions.

The cooperation between the two stages also allows for better results. Although complexity is an issue for the second stage problem, the first stage is relatively easy to solve. In order to gain information from the second stage, the RA should also solve the other problem, which has a higher level of complexity. Nonetheless, the RA faces a strategic problem, so a higher computational time is not a hindrance. Indeed, strategic problems concern long-term decisions, such that a decision-maker can surely afford the high computational time.

A limitation of this model is the deterministic nature of data, which does not reflect reality. Although, the implementation of safety stock in the facilities eases this issue. Indeed, the main benefit of safety stock is

Table 9
Results concerning the sensitivity analysis on facilities' capacity.

	Scenarios								
Facility	S1	S2	S3	S4	S5	S6	S7	S8	S9
0									
1	M 50% 0%	M 50% 0%	L 67% 0%	L 90% 67%	L 74% 13%	S 50% 0%		L 75% 25%	M 78% 33%
2							S 45% 0%		
3	L 68% 0%	L 64% 0%				L 52% 0%		M 62% 0%	
4									
5	L 47% 0%	L 70% 0%	L 94% 81%	L 71% 6%	M 49% 0%	L 69% 0%	L 97% 90%	L 59% 0%	L 75% 18%
6	S 100% 100%				S 50% 0%			S 68% 0%	
7	S 64% 0%	S 60% 0%	S 56% 0%						
8	L 75% 21%	L 68% 0%	M 87% 58%	L 92% 75%	L 45% 0%	L 57% 0%	M 96% 87%	M 55% 0%	L 76% 24%
9	L 91% 71%	L 84% 50%	L 89% 63%	L 97% 90%	M 60% 0%	L 83% 44%	L 52% 0%	M 82% 39%	L 60% 0%

Table 10
Results for different combinations of the weights γ_1, γ_2, and γ_3.

γ_1	γ_2	γ_3	of^1_{SP}	of^2_{SP}	of^3_{SP}	of_{SP}	dist
0.3	0.3	0.4	5,718.58	44.05	135	1,782.79	20.77
0.25	0.25	0.5	5,718.58	44.05	134	1,507.66	20.65
0.25	0.5	0.25	5,698.29	99.07	138	1,508.61	58.63
0.5	0.25	0.25	5,718.58	44.05	135	2,904.05	20.77

to dampen the fluctuation of demand. Moreover, the objective function of the second-stage concerns only the length of the routes. It is a fair assumption, considering that a RA usually outsources waste collection tasks to smaller local companies. However, the RA could also push the smaller companies to pursue sustainable objectives, perhaps by using incentives. This could benefit the RA's sustainable objective.

6. Conclusions

In this paper, we developed a novel two-stage model to assist a regional authority in designing a WM network. The authority wanted to cluster clients, locate facilities for these clusters, and define the correct amount of capacity to install. The regional authority is aware that the network will then be used by another decision-maker, i.e. the carriers in charge of picking up waste and delivering it to the opened facilities. In order to foresee the behaviour of the second decision-maker and in order to avoid possible miss-clustering or misallocation of capacity, the regional authority relies on the second stage of the model and a matheuristic. The matheuristic provides several starting points to the authority, which is then able to predict the best outcome.

Several random instances are used to test the two-stage optimisation problem TSOP within the proposed matheuristic. Given the promising results, TSOP and the matheuristic are then applied to a real-world case study. Once again, the model provides promising results, and the matheuristic is able to improve upon the first shortsighted solution.

Given that some tour lengths are uneven, future research may address tour length balance. Moreover, uncertainty in waste generation and time length could be considered. Most importantly, a few instances of the operational problem were optimally solved. For these instances, the whole TSOP model is effectively a Bilevel optimisation model, which can be used with multiple decision-makers with an inherent hierarchy and a possible degree of cooperation. For these instances, the first stage could be perceived as the leader, whereas the second stage could be perceived as the follower. Although, we were not always able to locate the optimal solution in the second stage. For all instances where the second stage was not optimally solved, we cannot refer to TSOP as a bilevel optimisation problem. Indeed, the optimality of the follower's problem solution is a strict requirement in order to have a feasible bilevel solution. Therefore, a follow-up could be to improve on solving the second stage VRP in order to achieve the optimality needed for a bilevel optimisation approach.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Diego Maria Pinto reports fnancial support was provided by Lazio Region.

Acknowledgements

This work has been partially supported by EU POR FESR program of LAZIO Region on "Research Groups 2020" through the project PIPER - Piattaforma intelligente per l'ottimizzazione di operazioni di riciclo. [Grant No. A0375 - 2020 - 36611, CUP B85F21001480002]

References

https://www.catasto-rifiuti.isprambiente.it/index.php?pg=&width=1536&height=864&advice=si.

Asghari, M., Al-e, S.M.J.M., et al., 2021. Green vehicle routing problem: A state-of-the-art review. Int. J. Prod. Econ. 231, 107899.
Barbosa-Póvoa, A.P., da Silva, C., Carvalho, A., 2018. Opportunities and challenges in sustainable supply chain: An operations research perspective. European J. Oper. Res. 268 (2), 399–431.
Barreto, S., Ferreira, C., Paixao, J., Santos, B.S., 2007. Using clustering analysis in a capacitated location-routing problem. European J. Oper. Res. 179 (3), 968–977.
Battarra, M., Erdoğan, G., Vigo, D., 2014. Exact algorithms for the clustered vehicle routing problem. Oper. Res. 62 (1), 58–71.
Bautista, J., Pereira, J., 2006. Modeling the problem of locating collection areas for urban waste management. An application to the metropolitan area of Barcelona. Omega 34 (6), 617–629.
Beamon, B.M., 1999. Designing the green supply chain. Logist. Inf. Manage.
Beltrami, E.J., Bodin, L.D., 1974. Networks and vehicle routing for municipal waste collection. Networks 4 (1), 65–94.
Bruglieri, M., Mancini, S., Pisacane, O., 2019. The green vehicle routing problem with capacitated alternative fuel stations. Comput. Oper. Res. 112, 104759.
Caramia, M., Pizzari, E., 2022a. Clustering, location, and allocation in two stage supply chain for waste management: A fractional programming approach. Comput. Ind. Eng. 108297.
Caramia, M., Pizzari, E., 2022b. Novel bilevel formulations for waste management. Discrete Appl. Math.
Corberán, Á., Eglese, R., Hasle, G., Plana, I., Sanchis, J.M., 2021. Arc routing problems: A review of the past, present, and future. Networks 77 (1), 88–115.
Dantzig, G.B., Ramser, J.H., 1959. The truck dispatching problem. Manage. Sci. 6 (1), 80–91.
Defryn, C., Sörensen, K., 2017. A fast two-level variable neighborhood search for the clustered vehicle routing problem. Comput. Oper. Res. 83, 78–94.
Di Foggia, G., Beccarello, M., 2021. Designing waste management systems to meet circular economy goals: The Italian case. Sustain. Prod. Consum. 26, 1074–1083.
Directive, E., 2008. Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives. Off. J. Eur. Union L 312 (3).
Dumas, Y., Desrosiers, J., Soumis, F., 1991. The pickup and delivery problem with time windows. European J. Oper. Res. 54 (1), 7–22.
Elshaer, R., Awad, H., 2020. A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. Comput. Ind. Eng. 140, 106242.
Erdoğan, S., Miller-Hooks, E., 2012. A green vehicle routing problem. Transp. Res. E 48 (1), 100–114.
Farahani, R.Z., Hekmatfar, M., Fahimnia, B., Kazemzadeh, N., 2014. Hierarchical facility location problem: Models, classifications, techniques, and applications. Comput. Ind. Eng. 68, 104–117.
Farahani, R.Z., SteadieSeifi, M., Asgari, N., 2010. Multiple criteria facility location problems: A survey. Appl. Math. Model. 34 (7), 1689–1709.
Garey, M.R., Johnson, D.S., 1978. A guide to the theory of NP-completeness.
Gentile, C., Pinto, D.M., Stecca, G., 2022. Price of robustness optimization through demand forecasting with an application to waste management. Soft Comput. 1–12.
Ghiani, G., Laganà, D., Manni, E., Musmanno, R., Vigo, D., 2014. Operations research in solid waste management: A survey of strategic and tactical issues. Comput. Oper. Res. 44, 22–32.
He, L., Huang, G., Lu, H., 2011. Greenhouse gas emissions control in integrated municipal solid waste management through mixed integer bilevel decision-making. J. Hazard. Mater. 193, 112–119.
Hemmelmayr, V.C., Doerner, K.F., Hartl, R.F., Vigo, D., 2014. Models and algorithms for the integrated planning of bin allocation and vehicle routing in solid waste management. Transp. Sci. 48 (1), 103–120.
Hintsch, T., Irnich, S., 2018. Large multiple neighborhood search for the clustered vehicle-routing problem. European J. Oper. Res. 270 (1), 118–131.
Hintsch, T., Irnich, S., 2020. Exact solution of the soft-clustered vehicle-routing problem. European J. Oper. Res. 280 (1), 164–178.
Kaza, S., Yao, L., Bhada-Tata, P., Van Woerden, F., 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. World Bank Publications.
Kucukoglu, I., Dewil, R., Cattrysse, D., 2021. The electric vehicle routing problem and its variations: A literature review. Comput. Ind. Eng. 161, 107650.
Kůdela, J., Šomplák, R., Nevrlỳ, V., Lipovskỳ, T., Smejkalová, V., Dobrovskỳ, L., 2019. Multi-objective strategic waste transfer station planning. J. Clean. Prod. 230, 1294–1304.
Kumar, A., Wasan, P., Luthra, S., Dixit, G., 2020. Development of a framework for selecting a sustainable location of waste electrical and electronic equipment recycling plant in emerging economies. J. Clean. Prod. 277, 122645.
Li, J.-Q., Borenstein, D., Mirchandani, P.B., 2008. Truck scheduling for solid waste collection in the City of Porto Alegre, Brazil. Omega 36 (6), 1133–1149.
Li, S., Wang, Z., Wang, X., Zhang, D., Liu, Y., 2019. Integrated optimization model of a biomass feedstock delivery problem with carbon emissions constraints and split loads. Comput. Ind. Eng. 137, 106013.
Luxen, D., Vetter, C., 2011. Real-time routing with OpenStreetMap data. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. GIS '11, ACM, New York, NY, USA, ISBN: 978-1-4503-1031-4, pp. 513–516. http://dx.doi.org/10.1145/2093973.2094062, URL http://doi.acm.org/10.1145/2093973.2094062.
Melo, M.T., Nickel, S., Saldanha-Da-Gama, F., 2009. Facility location and supply chain management–A review. European J. Oper. Res. 196 (2), 401–412.
Nagy, G., Salhi, S., 2007. Location-routing: Issues, models and methods. European J. Oper. Res. 177 (2), 649–672.
Olapiriyakul, S., Pannakkong, W., Kachapanya, W., Starita, S., 2019. Multiobjective optimization model for sustainable waste management network design. J. Adv. Transp. 2019.
Pinto, D.M., Stecca, G., 2020. Optimal planning of waste sorting operations through mixed integer linear programming. In: Proceedings of 18th Cologne-Twente Workshop on Graphs and Combinatorial Optimization, September 14-16, Online. In: AIRO Springer Series, vol. 5.
Pinto, D.M., Stecca, G., Gentile, C., 2020. Robust optimal planning of waste sorting operations. In: Proceedings of International Conference on Optimization and Decision Science ODS2020, 2020 November 19, Online. In: AIRO Springer Series ODS2020, vol. 6.
Prodhon, C., Prins, C., 2014. A survey of recent research on location-routing problems. European J. Oper. Res. 238 (1), 1–17.
Saadatlu, E.A., Barzinpour, F., Yaghoubi, S., 2023. A sustainable municipal solid waste system under leachate treatment impact along with leakage control and source separation. Process Saf. Environ. Prot. 169, 982–998.
Sharif, N.S., Pishvaee, M.S., Aliahmadi, A., Jabbarzadeh, A., 2018. A bi-level programming approach to joint network design and pricing problem in the municipal solid waste management system: A case study. Resour. Conserv. Recy. 131, 17–40.
Sherif, S.U., Asokan, P., Sasikumar, P., Mathiyazhagan, K., Jerald, J., 2022. An integrated decision making approach for the selection of battery recycling plant location under sustainable environment. J. Clean. Prod. 330, 129784.
Šomplák, R., Kůdela, J., Smejkalová, V., Nevrlỳ, V., Pavlas, M., Hrabec, D., 2019. Pricing and advertising strategies in conceptual waste management planning. J. Clean. Prod. 239, 118068.
Srivastava, S.K., 2007. Green supply-chain management: a state-of-the-art literature review. Int. J. Manag. Rev. 9 (1), 53–80.
Swart, J., Groot, L., 2015. Waste management alternatives:(Dis) economies of scale in recovery and decoupling. Resour. Conserv. Recy. 94, 43–55.
Terouhid, S.A., Ries, R., Fard, M.M., 2012. Towards sustainable facility location–a literature review. J. Sustain. Dev. 5 (7), 18–34.
Tirkolaee, E.B., Abbasian, P., Weber, G.-W., 2021. Sustainable fuzzy multi-trip location-routing problem for medical waste management during the COVID-19 outbreak. Sci. Total Environ. 756, 143607.
Tirkolaee, E.B., Mahdavi, I., Esfahani, M.M.S., Weber, G.-W., 2020. A robust green location-allocation-inventory problem to design an urban waste management system under uncertainty. Waste Manage. 102, 340–350.
Toro, E.M., Franco, J.F., Echeverri, M.G., Guimarães, F.G., 2017. A multi-objective model for the green capacitated location-routing problem considering environmental impact. Comput. Ind. Eng. 110, 114–125.
Tricoire, F., Parragh, S.N., 2017. Investing in logistics facilities today to reduce routing emissions tomorrow. Transp. Res. B 103, 56–67.
Tseng, M.-L., Islam, M.S., Karia, N., Fauzi, F.A., Afrin, S., 2019. A literature review on green supply chain management: Trends and future challenges. Resour. Conserv. Recy. 141, 145–162.
Van Engeland, J., Beliën, J., De Boeck, L., De Jaeger, S., 2020. Literature review: Strategic network optimization models in waste reverse supply chains. Omega 91, 102012.
Vidal, T., Battarra, M., Subramanian, A., Erdoğan, G., 2015. Hybrid metaheuristics for the clustered vehicle routing problem. Comput. Oper. Res. 58, 87–99.
Vigo, D., Bonoli, A., Gricinella, A., Zarri, G., 2007. Key Elements for Optimal Integrated USW-Management-International Experiences. Esculapio, Bologna, Italy.
Viktorin, A., Hrabec, D., Nevrlỳ, V., Šomplák, R., Šenkeřík, R., 2023. Hierarchical clustering-based algorithms for optimal waste collection point locations in large-scale problems: A framework development and case study. Comput. Ind. Eng. 109142.
Zhang, B., Li, H., Li, S., Peng, J., 2018. Sustainable multi-depot emergency facilities location-routing problem with uncertain information. Appl. Math. Comput. 333, 506–520.