Master Operations Research and Combinatorial Optimization
Master Informatique / Master Mathématiques & Applications

Waste Collection Optimization: A Reverse Inventory Planning Problem

Louis Gourinchas

August 30th, 2024

Research project performed at laboratory G-SCOP

Under the supervision of:
Van-Dat CUNG, Grenoble INP-UGA

Defended before a jury composed of:
Prof. Nadia BRAUNER-VETTIER
Associate Prof. Siao-Leu PHOURATSAMAY

August 2024

Abstract

This work explores the applications of Reverse Inventory Routing Problems in the creation of a waste collection schedule. Our main objective is to find a way to reduce the distance travelled by collection vehicles, which would be beneficial both monetarily and environmentally. With this in mind, we will study whether or not the reverse inventory routing problem (RIRP) is the right framework of consideration for this situation.

For our work, we will research different techniques to try and replicate the solving of a RIRP formulation of our situation. For this, we implemented techniques of Routing, Clustering and Date picking, recreating a complete collection schedule. The results of these methods are then compared to the efficiency of the current system in place, as well as, where applicable, the previous work that was done on this subject. This teaches us that, while our experiment weren't as successful as we hoped, it is possible to recreate the solving of a RIRP through decomposition.

Overall, this work shows that applying inventory routing techniques even partially can have beneficial effects on the waste collection process. We have also identified that there are multiple possible options to investigate further, including the use of heuristics to improve tours, or different date picking techniques.

Résumé

Ce travail explore les applications des problèmes de routage avec gestion de stocks inversés dans la création d'un calendrier de collecte des déchets. Notre objectif principal est de trouver un moyen de réduire la distance parcourue par les véhicules de collecte, ce qui serait bénéfique à la fois sur le plan financier et environnemental. Dans cette optique, nous étudierons si le problème de routage avec gestion d'inventaire inversé est ou non le cadre de considération approprié pour cette situation.

Pour notre travail, nous allons étudier différentes techniques pour essayer de reproduire la résolution d'une formulation de notre situation selon le probleme de routage avec gestion d'inventaire inversé (RGSI). Pour ce faire, nous avons mis en œuvre des techniques de routage, de regroupement et de sélection des dates, recréant ainsi un programme de collecte complet. Les résultats de ces méthodes sont ensuite comparés à l'efficacité du système actuellement en place, ainsi que, si applicable, aux travaux antérieurs réalisés sur ce sujet. Cela nous apprend que, même si notre expérience n'a pas été aussi fructueuse que nous l'espérions, il est possible de recréer la résolution d'un probleme de RGSI par le biais de la décomposition.

Dans l'ensemble, ce travail montre que l'application, même partielle, des techniques de routage avec gestion de stocks inversé peut avoir des effets bénéfiques sur le processus de collecte des déchets. Nous avons également constaté qu'il existe de nombreuses options possibles qu'il serait possible d'approfondir, notamment l'utilisation de techniques heuristiques pour améliorer les tournée ou de différentes methodes de sélection des dates.

Contents

— 1 —
Introduction

For as long has there as been a human civilization, there has been a waste production, and as populations grew throughout history, so did the quantity of waste produced, and so appeared the need for waste management solutions. Aggravated especially by the growing urban population, sprung up by industrialisation, waste management became an unignorable problem during the late 18th and early 19th century, where the buildup of waste in cities caused the quality of life to rapidly plummet. Since that time, ever-innovating waste management solution have been researched and put in effect up until today, where new technology keeps being applied to make these systems more efficient, like the use of RFID tags and level censors to better monitor every bin's activity.

The subject of this internship was brought forward by the *Communauté de Commune de Serre-Ponçon* (referred to as CCSP going forward), and is about optimizing waste collection in the region. This is a follow-up of previous work by Antonio Lopez-Ropero in [14], which focused on applying Machine Learning techniques in order to predict the waste production in the same territory. In this work, our approach will be more focused on how to improve the waste collection planning of the CCSP using optimization techniques.

With this objective in mind, section 1.1 goes further into detail about the waste management process and what parts of it are relevant to our research. Section 1.2 will explore our problem in more details, examining the real life system we are working with, and potential problems or limits we might face. We will then move on to reviewing the original work that was previously done on this project in section 1.3. Finally, we will go over the methodology to be used in this work, and finally present our research objective in section 1.4.

1.1 Waste Management Overview

Waste collection is the core part of the work on this project, but to better understand the objectives and implications of our work, we need to fully grasp the whole waste management process. After a quick study of the subject, we are able to details this process in 5 different phases:

1. Waste Generation: Simply a byproduct of daily life, for both households and businesses, waste production happens every day throughout every part of the city. However, multiple types of waste exist, like household waste or recyclables (like paper, plastics or glass), and are all produced at different proportion by different actors.

2. On-Site Management: This phase corresponds to how the waste is stored and separated (or not) to await it's collection. Multiple systems exist here, but are mostly simply a question of scale: Is each waste collection point meant to receive the waste of a street block, or an entire neighbourhood? But this is also where a first round of sorting happens, where household waste, recyclables or hazardous waste are all separated.

3. Collection: In this phase, either the municipality (like in our project) or a private company will have the task of collecting the waste from the containers and transporting it to the relevant facilities. In our case, each type of waste is collected separately, but it is also possible for some of them to be collected simultaneously using specific trucks. However, we will deal only with the collection of household waste in this work.

4. Treatment: For some type of waste, further treatment is necessary before it is able to be processed, even if it was already partially done in the on-site management phase. Here, each type of waste is separated depending on which way it is going to be processed, be it recycling, incineration, or going into various waste-from-energy systems.

5. Disposal: For any waste that cannot be recycled or reused in any way, this is the final stage of the process. The waste that reaches this phase is most often sent to landfills.

These phases give a somewhat simplified look on the waste management process, not detailing the multiple ways some types of waste might be treated once collected. But as we can see, waste collection is still a complex issue, involving multiple phases and actors in coordination. For this work, our focus is mainly going to be on phase 3, where we will try to develop efficient strategies for the collection and transport of the household waste.

1.2 Current Situation of the CCSP and the Research Question

As mentioned before, this section will first examine the existing physical situation, before giving the research question that we will try to answer in our work.

As we have previously said, this waste collection takes place in the *Communauté de Communes de Serre-Ponçon*, which is a diverse area. The main population of just under 17,000 inhabitants lives mostly in the twenty towns that form the CCSP. The region is highly touristic, with Lake Serre-Ponçon attracting visitors in summer, and the ski resorts of Les Orres in winter. The area is also mainly mountainous, resulting in the waste collection truck having a very high fuel consumption during some routing trips, up to 100 liters per 100 kilometers!

On the physical system of the region, we have a set of 329 waste collection points, that are all over the territory. Each collection point could have one to several waste bins, and people closed to the bins are invited to bring their waste to the bins. Waste is no longer collected in front of each house or each building. There are here two interesting problems related to the location of the collection points and the sizing of the bins, but in this work, we assume that the decisions to these two problems have already been made.

Besides, no distinction is made between these bins, whether they belong to different towns or have very different waste production rates. Clearly, there is no common waste generation behavior for all bins, as they may be in very different locations, for example one being useful to a relatively small group of households and the other in the middle of a ski resort.

Each of these waste collection points has it's own characteristics. Firstly, each point does not necessarily have a bin ready to collect waste of every type (meaning either household waste, mixed waste, or recyclables), and they also might have more than one bin used to collect a specific type of waste. Each bin also has it's own capacity. The size of a bin can be 3000, 4000, or 5000 liters, the latter being by far the most common. For the sake of simplicity and without loss of generality, we will assume that there is only one bin at every collection point, and we will refer to them simply as points without distinction for the remaining of the report.

The only details left of the system is the fleet which, according to UNICO France, consists of 7 trucks, with a capacity of either 10 or 12 tons. UNICO France is the company who handles the waste management system for CCSP, and also the company who provides us the waste collection data.

Finally, all of this is completed by a special point called the *depot*, which is the point from where trucks depart their waste collection tours, and the point to where they must bring back the waste.

The cycle of tourist seasons is of course reflected in the CCSP's waste collection management, because an increase in population inevitably means an increase in waste production. To better handle this seasonal behavior, four general periods of time representing different seasons have been created: Spring-April 15th to June 15th; Summer-June 16th to September 14th; Autumn-September 15th to December 15th; and Winter-December 16th to April 14th. Autumn and Spring are the two *off-seasons* that we will be focusing on in this work, because the CCSP staff thinks that there is more savings that can be generated in these two periods. During the two peak-seasons, Summer and Winter, waste generation is so high that most of the points have to be collected daily and savings are almost impossible to achieve.

For an idea of the scale of the system, during the 2023 spring season, the staff collected 518 tons of waste in 72 tours, travelling over 4300 kilometers to collect waste from points 2174 times, resulting in an average of 118kg of waste collected per kilometer traveled.

Furthermore, another decision made by CCSP is to associate a *frequency* with every collection point, corresponding to how often they will collect it. If a point has an associated frequency of 1, it will be collected once a week, and if it has a frequency of 0.5, it will be every two weeks, etc.

Now that we know more about the context of this project and the physical system, we can state our objective in the form of the following research question: **Can we reduce the travelled distance of the fleet by applying an optimized collection planning at the tactical level?**

In other words, can we formulate our problem as a Reverse Inventory Routing Problem in order to improve the efficiency of the waste collection planning over an entire period of 2-4 months? Readers are invited to refer to the end of section 2.1 for an explanation on the way we consider the waste collection problem as a Reverse Inventory Routing Problem.

1.3 Existing Data and Preliminary Work

As mentioned previously, our work on this project comes directly after some recent work in [14], and one of our first task was to examine and understand this work, to see which methods used or conclusion drawn might be useful to us. The main focus of this previous work was on the use of Machine Learning techniques, with the objectives of creating a waste generation forecast. Then, it was to develop strategies using clustering and routing techniques related to a reverse inventory routing problem to determine the collection dates of each points. For the first part of this work, the data provided by UNICO France was consolidated, merging different parts of the provided data to create a consolidated dataset to be used in the upcoming work. This part is not exactly applicable to the work that we will do on our own data, as ours is already way more complete than those used then. It is still interesting to study, as we have our own consolidation and selection step applied to our data, and we might be required to apply some of the same methods to obtain a clean data set.

The major part of this work is focused on using three different Machine Learning methods to predict waste generation of points for an the upcoming season of spring 2023 using historical data: Linear and polynomial regressions, and time series. This is completed with an analysis of the resulting prediction, determining which method produces the best results for each points, while pointing out the possible limitations of each of them whenever the results appear to be unreliable. While this analysis is interesting on it's own, it is not relevant at all for the purpose of this work in it's nature, as we will not be using Machine Learning techniques to predict waste generation when we eventually need it. It will be interesting to look at the results and see if we can approach them using the recorded data provided by UNICO France for the same period. We note that these recorded data were not available at the time of the previous worK.

After this, a 2-phases clustering technique is used: An agglomerative clustering algorithm that creates 22 clusters based on proximity, and a balancing algorithm that results in 5 clusters with the most equal volume of waste produced possible within reason, while respecting adjacency. This method of clustering seems logical enough, but we see as a result that one cluster ends up being about 30% of the produced waste, while the other 4 are more or less balanced around 17% of the total.

1.4 Methodology and Research Objective

Our main goal with this project is to analyse the possible application of Reverse Inventory Routing Problem solving techniques to the waste collection process of the CCSP, and more precisely the elaboration of a waste collection planning. To achieve this general objective, we will adopt the methodology described below:

Since the waste collection problem has been identified as a Reverse Inventory Routing Problem which is basically the same as an Inventory Routing Problem (IRP), we want to investigate the techniques used to solve the IRP and look for their adaptation to our problem.

Indeed, instead of considering maintaining product stocks above a certain minimum level in order to avoid stock shortages as in the IRP by deciding when and which product stocks should be replenished to avoid product disruption and how to route trucks, in the waste collection problem we need to decide when and which points to should be emptied to prevent them from overflowing.

1. In chapter 2, reviewing the current state-of-the-art in Inventory Routing Problems applied to waste collection, and more specifically some of the sub-problems making up IRPs, such as clustering and vehicle routing.

2. Find a method of solving simple routing problems that can be applied within our context and improves the distance travelled in respect to already recorded tours, detailed in section 2.3.

3. Determine whether or not we can come up with and carry out a beneficial clustering steps that makes waste collection more efficient, which we theorize in section 3.3 and detail in section 4.1.

4. Develop a reliable strategy to determine waste collection dates at each point, discussed in section 3.2 and section 3.4, and then experimented on in section 4.2 and section 4.3. The goal of this strategy is to be used in addition with the two other methods we developed above in order to solve the reverse inventory routing problem of our project.

5. For our experimentations, chapter 4 will study the efficiency of our methods in different scenarios and with different parameters, to compare with the current planning strategy of CCSP.

6. Finally, look back on our work and identify potential limitations or blind spots, but also possible improvements for future iterations, and open the research to new possible angles.

— 2 —
State of the art

In this section, we will be exploring the current literature to better understand our problem. We will first explore articles treating Inventory Routing Problems, with the hopes of finding possible applications for our situation considering the reverse inventory management and/or handling the periodic planning due to the collection frequency of the collection points. Then, we will discuss different aspects of the decomposition of an RPIRP, more precisely we will seek to better understand clustering methods used in this kind of problems. Finally, we will conclude by reviewing the different ways a routing problem can be considered in a waste collection situation, as well as how to solve it in an efficient way.

2.1 Reverse, Periodic, Inventory Routing Problems

The Inventory Routing Problem (IRP) presents a complex optimization challenge integrating inventory management and vehicle routing. The objective, given a network of roads and a list of customer demands, is to determine an optimal delivery planning (including the routes travelled), that meets every customer demand while also minimizing the overall cost of both travel and inventory holding. The IRP has been studied widely for various industries, and we refer to [6] for a review of the literature surrounding the problem, which goes over multiple variants of the problem, formulations, and solving algorithms.

A variant of the IRP that is of interest to us in this context is the Periodic Inventory Routing Problem (PIRP), where deliveries are made at a fixed, regular interval. This is relevant because it approaches the way waste collection is currently handled in the CCSP: There is a set of preexisting tours, that are each performed at their own regular interval over a time period. Following this, one of our goals for this project will be trying to determine the set of points to be collected together, and solve a PIRP concerning these specific points, resulting in multiple plannings covering all points that can be merged in a complete planning.

Another study we can turn to in order to better understand the IRP is [15], which details the different important points of interest of an IRP that can create variations of the problem depending on their nature. This is for example the client's expectations, be it in time of delivery, size of inventory or unpredictable demands, but it also includes the vehicle fleet aspect, in number or capacity of vehicle, as well as constraints on time of availability. This PhD thesis then goes further into the methods used in the literature to solve uncertain IRP problems. It is from this thesis that we obtain a formulation of a generic IRP that we adapt to our problem, detailed below in section 2.1.1.

A specificity of our problem, and a difference from the usual IRP, is that we need to collect products from clients, and not deliver them. We can, however, quickly come back to a normal IRP for our situation. The simple solution to this is to consider that the products we deliver to our clients is empty space. Thinking like this, the collection points would be consuming stock of their product of empty space by filling up with waste, until their stock of empty space is low, meaning that they are almost full, and we come to deliver more empty space by collecting waste. If we consider the problem like this, the filling rate of the points can simply be seen as the empty space consumed daily by each point. Thus, by considering this interpretation of the waste collection problem regarding the IRP, we adapted the data we have so that they can be used with a generic formulation of the IRP.

From [15], we obtain the following formulation of a Periodic Reverse IRP:

2.1.1 Formulation of the IRP

Definition Sets:

- $G(V,E)$ is a graph where vextex $0 \in V$ is the depot, $V \setminus \{0\}$ are the collection points, and the set of edges E represents the roads connecting points to each other.
- The set $H = \{0,1,2,|H|\}$ is the time horizon of the period we consider, where $p \in H$ represents the index of the period. $p = 0$ represents the initial state .

Client Data:

- D_i^t is the demand of client $i \in V \setminus \{O\}$ at period $p \in H$.
- I_i^0 represents the initial inventory (at period $p = 0$) of client $i \in V \setminus \{O\}$.
- I_i^{max} is the maximum inventory level for client $i \in V \setminus \{O\}$.

Supplier data:

- R^p is the quantity of products available or produced at supplier $0 \in V$ for period $p \in H$.
- I_0^0 represents the initial inventory of the supplier.
- C represents the capacity of the vehicle.

Costs:

- h_i is the holding cost paid for each product in the inventory of the client/supplier $i \in V$ at the end of period $p \in H$.
- $f(i,j)$ is the travelling distance of edge $(i,j) \in E$.
- c is the cost of going through one distance unit.

Mathematical variables:

- $x_{i,j}^p = 1$ if edge $(i,j) \in E$ is travelled by vehicle et period $p \in H$, 0 otherwise.
- $y_i^p = 1$ if client $i \in V \setminus \{0\}$ is visited at paeriod $p \in H$, 0 otherwise.

- $I_i^p \in \mathbb{R}$ is the inventory level of client $i \in V \setminus \{0\}$ at the end of period $p \in H$.
- $q_i^p \in \mathbb{R}$ is the quantity sent to client $i \in V \setminus \{0\}$ at period $p \in H$.

$$min \quad c * \sum_{i \in V} \sum_{j \in V, j<i} \sum_{p \in H \setminus \{0\}} x_{i,j}^p * f(i,j) + \sum_{i \in V} \sum_{p \in H} I_i^p * h_i$$

$$I_0^p = I_0^{p-1} - \sum_{i \in V \setminus \{0\}} q_i^p + R^p \qquad \forall p \in H \setminus \{0\} \tag{2.1}$$
$$I_i^p = I_i^{p-1} + q_i^p - D_i^p \qquad \forall i \in V \setminus \{0\}, \forall p \in H \setminus \{0\} \tag{2.2}$$
$$I_i^p \leq I_i^{max} \qquad \forall i \in V \setminus \{0\}, \forall p \in H \setminus \{0\} \tag{2.3}$$
$$q_i^p + I_i^{p-1} \leq I_i^{max} \qquad \forall i \in V \setminus \{0\}, \forall p \in H \setminus \{0\} \tag{2.4}$$
$$q_i^p \leq y_i^p * I_i^{max} \qquad \forall i \in V \setminus \{0\}, \forall p \in H \setminus \{0\} \tag{2.5}$$
$$q_0^p \leq y_0^p * C \qquad \forall p \in H \setminus \{0\} \tag{2.6}$$
$$\sum_{j \in V \setminus \{0\}} x_{i,j}^p + \sum_{j \in V \setminus \{0\}} x_{i,j}^p = 2 * y_i^p \qquad \forall i \in V, \forall p \in H \setminus \{0\} \tag{2.7}$$
$$\sum_{i \in T} \sum_{j \in T, i<j} x_{i,j}^p \leq |T| - 1 \qquad \forall T \subseteq V \setminus \{0\}, p \in H \setminus \{0\} \tag{2.8}$$
$$x_{i,j}^p \in \{0,1\} \qquad \forall i,j \in V, \forall p \in H \setminus \{0\} \tag{2.9}$$
$$y_i^p \in \{0,1\} \qquad \forall i \in V, \forall p \in H \setminus \{0\} \tag{2.10}$$
$$q_i^p \geq 0 \qquad \forall i \in V \setminus \{0\}, \forall p \in H \setminus \{0\} \tag{2.11}$$
$$I_i^p \geq 0 \qquad \forall i \in V, \forall p \in H \setminus \{0\} \tag{2.12}$$

The objective function minimizes the holding cost and the total travel cost over the entire period H. Constraint (2.1) is a flow constraint that computes the inventory level of the supplier at each period $p \in H \setminus \{0\}$ from it's previous inventory level, the quantity produced and the quantities sent to the clients at period p. Constraint (2.2) is the flow conservation for the clients, calculating the inventory level of each client $i \in V \setminus \{0\}$ for each period $p \in H \setminus \{0\}$ from it's previous inventory level, the quantity received from the supplier and it's demand for period p. Constraint (2.3) states that the inventory level of client $i \in V \setminus \{0\}$ at any period $p \in H$ must be lower than I_i^{max}, and constraint 2.4 states that a replenishment of this client at period $p \in H \setminus \{0\}$ cannot exceed it's maximum inventory level. Constraint (2.5) links variables y_i^p and q_i^p, stating that a client $i \in V \setminus \{0\}$ that receives at period $p \in H \setminus \{0\}$ is necessarily visited at period p. Constraint (2.6) works similarly for the supplier, stating that the quantity leaving the supplier at period $p \in H \setminus \{0\}$ is limited by vehicle capacity C. Constraint (2.7) ensures that if a location is visited, it is entered and left once, and constraint (2.8) eliminates subtours. Finally, constraints (2.9) through (2.12) enforce integrality and non-negativity conditions on the variables.

There are, however, multiple problems when directly applying that formulation to our situation as it is. The first is that we will not exactly know the filling rate, and so the demand, of each point of collection during the season. This could be worked on by pursuing further in the trail of Antonio Lopez-Ropero's work, with the goal to reliably predict waste production, and it is a subject studied in papers like [3] or [11], which both explore different ways to predict waste generation. It is also something we will come back to in section 3, with much more approximate methods. The second is that even 'only' an inventory routing problem is already NP-Hard, as seen in [1], and our particular instance has over 300 clients to satisfy, over a period of around 60 days.

It is for these two reasons that we decided to try to approach the problem from a new angle and tried to break down the PIRP into smaller, more manageable problems. As a result of this process, we end up with with three different parts that we will explore and combine, corresponding to the three important components of a PIRP: Clustering, Routing and Planning.

As the creation of a planning will make both clustering and routing techniques, we will now follow up with a literature exploration of these two components making up a regular IRP: Clustering and Routing. The planning aspect will be handled separately, when deciding of a collection frequency for each point and utilizing both other components.

2.2 Clustering

A crucial step in solving the reverse inventory routing problem is clustering, or cluster analysis. This is the act of sorting a group of objects into multiple different, smaller groups, or clusters. In this part, we will learn about the current methods used in aiding in the waste collection process. The first article we refer ourselves to is [2], which performs an analysis of clustering methods employed in papers centered around waste management. One of the interesting results presented in this paper is the part focused on waste collection problems, as this is exactly what our problem is. What transpires from this section is that the most used category of techniques used in waste collection problems is heuristics, which encapsulates a vast number of different things like k-means, ant colony systems or capacitated savings algorithms. However, this paper does no go into further details about what exactly these techniques are and how they are implemented and used, and so we look to explain some of these in greater details:

1. K-means Clustering: One of the most widely used clustering methods, K-means partitions the dataset into K distinct clusters, minimizing the variance within each cluster. In waste collection, K-means can be used to group similar waste generation sites, allowing for the design of efficient collection routes that minimize travel time and costs. We can turn to [9], for an example of K-means clustering separating cities to prepare for the installation of different waste collection strategies.

2. Hierarchical Clustering: Unlike K-means, which requires the number of clusters to be defined beforehand, hierarchical clustering builds a tree-like structure (dendrogram) to represent nested groupings of data points. This method is particularly useful in waste collection for identifying clusters at different scales, such as neighborhood-level or city-wide groupings. In [16], this method is applied to solve large waste collection network design problems in cites.

3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): DBSCAN is a density-based clustering method that groups points based on their proximity and density, making it effective in identifying clusters of various shapes and sizes. This technique is advantageous in waste collection for handling irregularly distributed waste generation

sites or for areas where data points are sparse. Research showed that DBSCAN could effectively segment urban areas with varying population densities, enabling more flexible and adaptive waste collection strategies.

2.3 Routing Problem

The other component of our problem is a Vehicle Routing Problem (VRP) which a generalisation of the Traveling Salesman Problem [10]. This is a widely known combinatorial optimization problem in logistics and transportation management, the main goal being determining the most efficient routes for a vehicle, or fleets of vehicles, to deliver goods to a set of customers. The objective is typically to minimize the total distance traveled or the overall cost of the routing, and it might be subject to several constraints in addition to fleet size, such as the vehicle load limits, specific delivery time windows, customer demand, or integrating delivery and pickup at the same time.

The variant of this problem that interests us is the Capacitated VRP, where the vehicles we use have a limited capacity. Other specificities of this problem include the fact that our routes start and end at the same location (henceforth called the depot), each point is visited only once, and we do not exceed the capacity of the truck. This problem also has it's own variations, that we can learn more about in [13], but that are not relevant to our situation.

This is once again an easy match for our problem, by simply reversing how we think: While in practice our vehicles start empty, and pick up waste as they progress, we can consider that they start full of empty space, and distribute it to our clients: the collection points. The first formulation of this problem was propose by Dantzig and Ramserc in 1959 [7], and Clarke and Wright proposed a good savings heuristic to solve this NP-hard problem in 1964 [5].

We can consider the flow-based formulation of the CVRP given by Borcinova in 2017 [4]:

We consider a complete directed graph $G(V,E)$, with $V = \{0,1,...,n\}$ is the set of nodes and $H = \{(i,j) : i,j \in V, i \neq j\}$ the set of arcs, where node 0 represents the depot. We also have a fleet of p vehicles with the same capacity Q, and the remaining n nodes are the customers. Each customer $i \in V \setminus \{0\}$ has a certain demand $d_i \leq Q$. We associate the distance traveled $c_{i,j}$ to each arc $(i,j) \in H$. The cost matrix is symmetric, i.e. $c_{i,j} = c_{j,i}$ for all $i,j \in V, i \neq j$ and satisfies the triangular inequality $c_{i,j} + c_{j,k} \geq c_{i,k}$ for all $i,j,k \in V$. The minimum number of vehicles needed to serve all customers is $\lceil \frac{\sum_{i=1}^{n} d_i}{Q} \rceil$.

We also add the binary variable $x_{r,i,j}$ to indicate if vehicle $r, r \in \{1,2,...,n\}$ travels the arc (i,j) in the optimal solution.

$$minimize \sum_{r=1}^{p} \sum_{i=0}^{n} \sum_{j=0, i\neq j}^{n} c_{i,j} x_{r,i,j}$$

Subject to

$$\sum_{r=1}^{p}\sum_{i=0,i\neq j}^{n} x_{r,i,j} = 1 \qquad \forall j \in \{1,...,n\} \tag{2.13}$$

$$\sum_{j=1}^{n} x_{r,0,j} = 1 \qquad \forall r \in \{1,...,p\} \tag{2.14}$$

$$\sum_{i=0,i\neq j}^{n} x_{r,i,j} = \sum_{i=0}^{n} x_{r,j,i} \qquad \forall j \in \{0,...,n\}, r \in \{1,...,p\} \tag{2.15}$$

$$\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n} d_j x_{r,i,j} \leq Q \qquad \forall r \in \{1,...,p\} \tag{2.16}$$

$$\sum_{r=1}^{p}\sum_{i\in S}\sum_{j\in S,i\neq j} x_{r,i,j} \leq |S|-1 \qquad \forall S \subseteq \{1,...,n\} \tag{2.17}$$

$$x_{r,i,j} \in \{0,1\} \qquad \forall r \in \{1,...,p\}, i,j \in \{0,...,n\}, i \neq j \tag{2.18}$$

Here, the main objective is the minimization of the cost of the roads travelled. Constraint (2.13) ensures that every customer is visited by exactly one customer. The flow constraints (2.14) and (2.15) guarantee that each vehicle can leave the depot only once, and the number of vehicles arriving at every customer and entering the depot is equal to the number of vehicles leaving it. Constraint (2.16) is the capacity constraint, making sure that the sum of demands of the customers visited in a route is less than or equal to the capacity of the vehicle performing the service. The sub-tour elimination constraint (2.17) ensures that the solution contains no cycle disconnected from the depot. The remaining constraint (2.18) specifies the definition domain of the variables. This model is known as a three-index flow formulation. The number of inequalities of the sub-tour elimination constraints grows exponentially with the number of nodes (from [4]).

Clarke and Wright's method to this problem uses a savings approach, where we start by considering the length of travel induced by delivering each customer one by one, returning to the depot each time. This results in a total cost of $2 * \sum_{j\in V\setminus\{0\}} c_{0,j}$.
We then consider delivering two points i and j on the same tour, which saves us the following amount:

$$\begin{aligned} s(i,j) &= 2c_{0,i} + 2c_{0,j} - [c_{0,i} + c_{i,j} + c_{0,j}] \\ &= c_{0,i} + c_{0,j} - c_{i,j} \end{aligned} \tag{2.19}$$

This quantity $s(i,j)$ is the "savings" gained from combining i and j. The larger $s(i,j)$ is, the better the option to combine them is. However, we still need to check if combining i and j results in a tour that doesn't violate the constraints of the VRP. The complete algorithm follows the steps below:

- Step 1: Calculate the value of s(i,j) for every pair of points (i,j) excluding the depot.
- Step 2: Rank the savings obtained in step 1 in descending order.

- Step 3: Consider the pair (i,j) with the largest s(i,j) that has not been treated yet, and treat it as follows:

    a. If neither i nor j are already part of a route, create a new route including both of them.

    b. If exactly *one* of i or j is already included in a route and is adjacent to the depot, then the link (i,j) is added to that already existing route

    c. If *both* i and j are already part of a route, these routes are merged.

  Note: As mentioned before, all of these modifications only occur if they do not break any constraints of the VRP

- Step 4: If there are still unprocessed savings, go back to step 3. Otherwise, stop. The solution of the VRP is the ensemble of routes created in step 3. Note that any points that have not been added to routes are to be collected alone.

As we have said before, if we consider only the pick up of waste from a small group of points, our situation translates easily to a single vehicle CVRP. Because we chose to try and cluster small amount of points together to create tours, we can make the assumption that there is no time constraint (because we are not collecting enough points to go over the duration of one shift) and there is only one vehicle. This is also in line with how the collection tours are done in reality, where a truck will be responsible for a tour, and might go back to the depot at some point in the tour if the collected waste is too much. Because of this, Clarke and Wright's solving approach will be used to determine the precise routes of our tours. While this algorithm is a single-vehicle problem and ours is not, we make the hypothesis that because we want to determine tours for limited clusters, a single truck will be able to meet this demand alone. This is also in line with the current real situation, where trucks that find themselves too full at some point of their tours are able t oreturn to the depot early and continue the tour after.

— 3 —
Solving Methods

As mentioned before, our goal in this work is narrowed down to trying to find methods for the creation of better collection routes, and a better overall collection planning. While we will explore methods such as clustering or routing options further down the road, one of our first concerns was starting the creation of a collection planning from scratch. As a starter, our first focus is trying to choose, for every waste collection point, on which days it should be collected.

3.1 Checking Data

The previous work on this problem by Antonio Lopez-Ropero resulted in the prediction of waste production data for the 2023 Spring period (which was not available at the time). Our first step is to check the results we can obtained using Antonio Lopez-Ropero's solving method on the now available real-life data provided by UNICO France with respect to Antonio Lopez-Ropero's prediction and the data he used in his report. For more clarity, going forward we will refer to the period of September first to December 16th as 'Fall', and the period of April 15th to June 15th as 'Spring'.

From Antonio's report, we get a part of real data, for Fall 2022, and a part of predicted data, for spring 2023, which is one of the results of his work. We compare those to our own data, which is the real provided data for both periods.

		Previous Results	Provided Data
Fall 2022	Volume	8,132,409 L	8,439,670 L
	Mileage	7,467 Km	9,330 Km
	Efficiency	1,089 L/km	904,6 L/km
Spring 2023	Volume	10,725,094 L	6,688,318 L
	Mileage	7,814 Km	4,379 Km
	Efficiency	1,372 L/km	1,527 L/km

Table 3.1: Comparison of results obtained on data in the previous work vs. provided data

What we can see in table 3.1 is that the data for Fall 2022 looks similar, enough that the differences can be considered filtering errors on our part of refining of the data by UNICO France in the time between our two works. We observe a larger discrepancy between the numbers for Spring 2023, which may be explained by multiple factors. Most importantly, as mentioned

before, the data presented by Antonio Lopez-Ropero for this period is the result of a prediction made using data from spring 2021 and 2022, and ours are simply recorded collection data. Also, the mileage noted here for spring 2023, 7,814 Km, is the result of an execution of Clarke and Wright's algorithm to determine the tours.
While the first factor might help us understand how the total collected volume might be inaccurate, the second one indicates that if anything, the mileage *should* be lower than ours, and so we will investigate further as to why that might be. Not presented in the table, we can also consider the data we were provided for the period of Spring 2022, which indicates that 6,736,700L were collected over 3,862Km, for an efficiency of approximately 1,744 L/km, overall way closer to the recorded 2023 data than to the prediction.

3.1.1 Using Mass instead of Volume

Because of the anomalies that we already noticed, a more in-depth exploration of the previous work was done. The most important thing that we discovered was that for the routing step, the Clark and Wright algorithm is executed using collection volumes in liters, and with a truck capacity of 18,000L. This is not an accurate capacity for the trucks, and is even not the one that was specified in Antonio Lopez-Ropero's report. This, of course, results in extremely high mileage for truck. As an indication, using this criteria for the trucks being full resulted in a return to depot over twice per tour on average (with some tours having upwards of 12 returns), which is clearly unrealistic. After a meeting with UNICO France's CEO, Clément MARTY, we learned that the metric used by truck drivers to determine if and when they need to go empty the truck were tons, and not liters. The two models of truck having an effective capacity of either around 10 and 12 tons, respectively, the drivers will choose to go empty the truck if they reach around 8/9 tons or 10/11 tons collected, and they know the remainder of their tour is going to end up overflowing the truck. After learning this, we re-explored our previous results using that new metric:

Distance calculated on real data	Without going back to the depot	With going back every 18,000L	With going back every 10t
Fall 2022	9,329 Km	26,438 Km	11,273 Km
Spring 2023	4,378 Km	13,029 Km	5,311 Km

Table 3.2: Comparison of travelled distance obtained using liters or tons on real data.

In tables 3.2 and 3.3, we can instantly see that while still adding some mileage, this method of choosing when to empty the truck results in way more acceptable total kilometers driven. However, this does mean that the results obtained by Antonio Lopez-Ropero might be slightly inaccurate, and it will require us being careful when using them in the future.

As mentioned earlier, the previous work makes use of a clustering technique to group points in 5 clusters, and succeeds in reducing the overall distance travelled using this step alone. However, we were not able to re-obtain the clustering results of the previous work from the

Distance calculated using Clarke&Wright	Without going back to the depot	With going back every 18,000L	With going back every 10t
Fall 2022	5,046 Km	9,657 Km	5,369 Km
Spring 2023	3,949 Km	9,411 Km	4,142 Km

Table 3.3: Comparison of distance travelled obtained using liters or tons and an execution of the Clarke & Wright algorithm.

material left over, and will have to re-establish their efficiency during our own experiments with clustering.

3.2 Picking Dates

Using what we have learned in our literature review of Inventory Routing Problems, we decided to try and pick a collection date for a point as soon as it is filled to a certain threshold. The problem with this approach, is that in practice, it is not realistic to check each point everyday to determine if it is filled enough to warrant a collection. However, creating a collection planning this way could give us a good baseline of what to expect from further methods and, if it is good enough, a base to iterate upon using further historical data.

Another difficulty for this method is that the data we have is not continuous: we know each collection point's volume only once at each collection, and never in between. This is where we can take inspiration from the previous work that was done on this subject, and try to fill in the missing part of our data. We picked two methods for this purpose, simply out of convenience: (1) Treating the production of the points as linear over the whole period, distributing the total of waste produced evenly over all days, and (2) treating it as linear over every period, where we distribute the waste recorded during one collection over the period of time passed since the last one. As an example, here is the resulting production for two collection points:

Once we have this data, the rest of the method is fairly straightforward. Every day, we looked at every point and their accumulated waste, taking into account any potential waste that we would have already collected, and decided to collect it or not depending on whether that waste is more than a certain threshold or not. At first, our threshold was placed at exactly 90% of the point's capacity, but this is one of the parameters that we can modify later, when doing our sensitive analysis.

3.3 Clustering

As discussed previously, another parameter that might help us to make a better collection planning is adding a step of clustering. In practice, this is to find groups of points that are close enough geographically that collecting them together would always be worth it. This is close to how the CCSP planners actually create their planning: There are sets of points, each collected at a regular frequency, and all points to be collected on one specific day will make up the tour, or multiple tours, for this day.

Following this idea, we wanted to determine whether or not we could apply clustering techniques to create separate clusters that would be interesting to collect together. As we have already mentioned in our literature review, multiple types of clustering techniques can be used efficiently to create more optimized groups of collection points. And so to try and recreate the CCSP's notions of tours in an efficient manner, we will try different techniques to create group, and then, assuming a similar planning of collection for all techniques, we will compare them. For this, we will look at indicators such as the raw distance travelled over the period, the number of collections, the number of tours, or the efficiency. From this, we can choose which of these is the most interesting in our situation.

3.4 Determining Collection Frequency

In addition to grouping their points in different tours, the CCSP's staff also assignsa unique collection frequency to each of them, and each day, the final tours are made up of the points whose collection day is today. This results in regular tours comprised of the collection points that fill up the fastest, occasionally supplemented with less frequently visited points. This idea of adjustable tours is one that is incredibly interesting, and if we can find a method to automatically create such tours from recorded collection data, the optimization of the planning will be far easier.

Our objective here is to use the collection point's filling rate to try and determine a collection frequency, that we can then couple with a step a clustering to imitate the CCSP's system. The challenge will be to try and find a correct formula such that the frequency chosen is regular enough to permit the creation of a planning, while remaining faithfully to the point's real collection needs.

Once we have explored our methods of creating a collection frequency completely enough, our final goal is to use it in association with what we will have explored as a result of section 3.2 and section 3.3, to hopefully succeed in creating a collection planning, as we initially intended when we defined our problem.

— 4 —

Numerical Experiments

Before detailing what experiments we conducted and what we learn out of them, let us recall what data we will be utilising. We are considering a region containing 329 collection points and a single depot, and we will focus on studying the Spring 2023 period, from the 15th of April to the 15th of June. We were provided with UNICO France's collection records for these points in 2023, up to August. Of these records, we will only consider the data tat is (1) relevant to our period, and (2) relevant to our type of waste, organic matter.

This data included two different documents: The first one a record of every point collected, totaling 2451 entries detailing the date of collection, the tour of which the collection was a part of, the point collected, how full the point was, and how many liters of waste were collected. The second one was a record of the 84 tours recorded during the period, including once again the date, but also the name of the tour, the truck and the truck driver, and the mileage recorded for this tour. Because we established earlier that calculating the return to the depot using tons was more accurate than using liters, we had to complete the first document, calculating how many kilograms of waste were collected.

4.1 Clustering

As mentioned in section 3.3, here the goal is to find out if using clustering to separate points to collect would be beneficial to our situation. At first, we wanted to try and get an idea of the overall efficiency of the method, and so we decided to start with a test that would go as follows: Separate our points into clusters geographically first, and then balance out the amount of waste produced by each cluster (while maintaining a geographical proximity), and compare the results of a planning created using those clusters to one created using randomly created clusters. Because we have 329 points, we settled on creating 6 clusters, which would map well to the 6 days of the week if necessary.

For our random algorithm, we simply go over each point and randomly assign him a non-full cluster (we know a cluster is full if it contains 55 points, and the last one will contain 54). For our non-random algorithm, we first regroup our points into 25 clusters using scikit's agglomerative algorithm in [12], which will merge pairs of clusters based on the distance between them until there are only 25 left. For this, the initial distance used is a matrix distance of every point created using the OSRM API [8], and the number of clusters was chosen arbitrarily. Then, we start by making a complete distance matrix of clusters, using the average coordinates of the points as a cluster's coordinates. Then, we explore different permutations of groupings these 25 clusters into 6 final clusters, while maintaining geographical closeness. We then select the permutation that has the 6 clusters that are the most balanced in terms of waste produced over the period.

Our next step is creating a collection plan over these clusters, and compare the execution of the plan. For this, we assumed that every collection point has a similar, regular production using the total produced quantity of waste produced, resulting in a final 115,479L/day overall, and 350L per day per point. While we came to the conclusion that using kilograms was more efficient earlier, the total quantity of waste collected in liters is more accurate than the one we estimated in kgs, and we already have efficiencies calculated in liters to compare the results to. Please note that while the production is regular for every point to calculate our results, the clusters of our second algorithm were balanced using the real quantities of waste produced by the points. Using this metric, is it possible to plan a collection every two weeks and still have no point overflow (resulting in a total of 4,900L per point every two weeks), but because so little collections is unrealistic, we will also try with one collection a week. The results of this experiment are presented in table 4.1, with the total distance obtained as well as the efficiency that results. Not detailed in the table is the number of tours that each method results in, this is because it is highly regular: There are 9 weeks in our period, and so our weekly collection will result in 9 total tours, while the fortnightly collection will only have 4.5.

Random Clustering			
Weekly Collection	With returns	18,089 Km	402.2 L/Km
	Without returns	11,116 Km	654.5 L/Km
Fortnightly Collection	With returns	9,044 Km	804.4 L/Km
	Without returns	5,558 Km	1,308.9 L/Km
Balanced Geographic Clustering			
Weekly Collection	With returns	9,782 Km	743.7 L/Km
	Without returns	5,026 Km	1,447.5 L/Km
Fortnightly Collection	With returns	4,891 Km	1,460.5 L/Km
	Without returns	2,513 Km	2,895 L/Km

Table 4.1: Comparison of distance travelled and efficiency for random clusters vs. balanced geographical clusters.

From this, we learn that this method of clustering is beneficial, at least when every point is at equal production and on a regular collection planning. We also note that while efficient compared to a random clustering, our results are still not as good as what we had obtained when applying Clark and Wright's algorithm on the real data. With this, we end up travelling 5,026 kilometers over the period, on a weekly collection basis and without going back to the depot, the corresponding result for the real data we obtained in section 3.1.1, table 3.3 is 3,949 kilometers (or 4,142 kilometers if we consider return using tons). This means that we still have a lot of differences with the complete system used by the CCSP, but is also expected at this stage. Our next goal will be to try and apply it to a more realistic scenario.

4.2 Picking Dates for Collection Planning

Following up on section 3.2 and the previous section, our goal in this part if to apply both the calculated filling rate and our clustering to a more realistic collection planning. The technique we decided to try is as follows: Every day, we check which points are over a certain collection threshold, and create the tours of the day using these points. For a more complete overview, we start by applying this method to every point without any clustering, and present the results in tables 4.2 and 4.3. We also separate the results according to which filling rate was used.

Production linear over the entire period		
Collection Threshold	Total Distance	Efficiency
95%	5,864 Km	1,288 L/km
90%	6,022 Km	1,254.2 L/km
85%	6,591 Km	1,146 L/km
80%	6,709 Km	1,125.8 L/km

Table 4.2: Detail of the travelled distance and efficiency of date picking with no clusters, using production linear over the entire period.

Production linear between each collection		
Collection Threshold	Total Distance	Efficiency
95%	5,821 Km	1,297.5 L/km
90%	5,992 Km	1,260.5 L/km
85%	6,218 Km	1,214.7 L/km
80%	6,495 Km	1,162.8 L/km

Table 4.3: Detail of the travelled distance and efficiency of date picking with no clusters, using production linear between each collection.

From this, we can first see that the production linear by collection seems to yield slightly better results, although not by a significant margin. This might be because of the way the production data was distributed, which results in some points not having any production at toward the end of the period, and so is not a relevant result. We also notice that planning to collect waste later seems beneficial, which seems logical since it results in planning less tours overall. This is counterbalanced however, by the fact that the real recorded collection on this period is still better than this result, once again with a total distance of 4,142 kilometers. We can also add that for the real collections, the points are 60% full on average, far lower than what we used, and still way more efficient. Next, we separated out points into clusters following the technique from section 4.1, and collected the results in table 4.4.

Overall, this results in 8,141 kilometers traveled over the period, in a total of 305 tours, for an efficiency of 927.8L/km, which is much worse than any of our previous results. As indicated by the total number of tours, and as found out after looking into our results, this is because we plan an abundance of very small tours every day, collecting waste from one or two points per cluster. The conclusion is that these two techniques are not immediately compatible, and require us to either find alternatives, or change the way we apply them. One of the alternatives that we explored is the creation of a frequency indicator for collection, and is detailed in section 4.3.

Cluster number	1	2	3	4	5	6
Total Distance	414 Km	1,861 Km	926 Km	2,498 Km	1,004 Km	1,438 Km
Number of tours	53	49	56	45	53	49

Table 4.4: Travelled distance and number of tours obtained using a combination of date picking and clustering.

4.3 Determining Collection Frequency

Inspired by the CCSP's system, and an alternative to picking collection dates from a recorded historical threshold, is the creation of a frequency indicator per point, which would help plan tours more easily. The way we decided to calculate this indicator was, like date picking, based on the point's filling rate that we calculates in section 3.2. For every point, we looked at how long it would usually take to fill (following the production linear overall, and considering a point to be full at 90% of it's capacity), and reverse it into a frequency. Because of the very diverse filling rate of points, we also choose to have "breakpoints" at which we would round the obtained frequency. The basics of these breakpoints is presented in table 4.5, we also note that when rounding the frequency indicator, we always did so in a way that would result in the collection point being over-collected, or never collected in the opposite.

Frequency indicator	.25	.5	.75	1
Collection Frequency	4x/week	2x/week	3x/2 weeks	1x/week
Effective collection days	Mon, Wed, Thu, Sat	Mon, Thu	Mon, Sat during week 1, Wed during week 2	Mon

Table 4.5: Details of the frequency collection breakpoints

Presented in table 4.5 is another decision we had to make: Given each point's collection frequency, which days do we choose to collect them on? The decision here is arbitrary, and it might be interesting to explore different planning options. While not reflective of the real collection planning, we made the decision to no spread collection over every day of the week, which may have resulted in something similar to the planning obtain in the previous section, with a vast majority of very small tours. Once we had this for every point, we checked every day for which points were marked for collection, and created tours using them. The results are visible in table 4.6

Cluster number	1	2	3	4	5	6
Total Distance	1,493 Km	6,167 Km	1,157 Km	2,130 Km	2,066 Km	8,567 Km
Number of Tours	13	39	33	33	52	52

Table 4.6: Travelled distance and number of tours obtained using a combination of collection frequencies and clustering.

As we can see, this is even worse than what we obtained using our previous method of picking date, and far worse than the real travelled distance: we travel 17,080 kilometers over 222 tours, for a final efficiency of 442.2L/km. Multiple factors can be pointed at to find out where we went wrong with this approach: A better choice of collection days could have been more efficient, or a different threshold would've resulted in a more accurate indicator. While this is possible, another look at the calculation of the indicator calculation reveals that for the vast majority of these points, their raw indicator is well over 1, and so they would require far less collections than what they were given. Changing our calculation to construct a planning over the entire period of 9 weeks (instead of the 2 weeks we used) would've allowed us to better account for these "slow" collection points. Another problem, that was already visible in section 4.2, is both our methods of creating collection plans do not make great use of the clustering. With our methods, we frequently collect only one or two points of each cluster every day, creating a lot of small, un-optimized tours. Our original goal with clustering was to tours that collect most of the cluster's points together, creating larger tours, but with points that are closer geographically.

— 5 —
Conclusions and Future Works

5.1 Conclusions

To conclude this work, we can return to our original premise, as detailed in section 1.4. From the study of the state of the art, we established that while the Inventory Routing Problem is very well studied in the literature, it is also an extremely complex problem that does not instantly translate well to our situation.

We successfully confirmed that decomposing the problem was possible, and that the subproblems we identified were also already studied in the context of waste collection, and so we were able to propose a strategy to solve the problem in several steps: (1) a clustering algorithm, creating groups of points that would be more efficient to collect together, (2) a picking date technique, resulting in the creating of a collection plan over our chosen period, and finally (3) compute the vehicle routing problem using Clarke and Wright's algorithm inside of each planned collection.

However, we were less successful on the execution of the techniques we identified to be relevant in section 2. Section 4 proved that even with techniques that we know are relevant to our problem, we could not approach the efficiency of the solution operated by CCSP.

5.2 Future Works and Perspectives

While we could not obtain a result that improves on the current state of the problem, we still managed to produce multiple tracks onto which continuing this work is possible. With all the tools we now have, it is possible to imagine creating a solution to the problem that integrates them better together, instead of trying to measure their efficiency with less-than-adequate tests. It also transpired that while a lot of papers exist on using various vehicular routing problems, a lot of them do not treat the planning aspect in the way we introduced in this paper, and this could be explored further.

There are also unexplored solution that could help to come up with a better solution. An idea that was mentioned during our research, but not explored further, is establishing heuristics to improve a collection planning once is it created. The idea is to examine tours one by one, and possibly add the points that do not make the distance traveled significantly worse. The difficulty in this is first, creating a base of tours that are good enough, and second, knowing which metrics to look at to determine if including a new point of collection is worth it. It was also mentioned that in a real tour, the driver can go over the recommended capacity of a truck if they know that the tour is nearly over, which is another thing that could be integrated in a future planning method.

Overall, we can firmly say that it is definitely interesting to continue researching solution to this problem, as efficient waste collection planning is beneficial all over the world, and we have already uncovered multiple possible tracks onto which this work can be continued.

Bibliography

[1] Claudia Archetti, Luca Bertazzi, Gilbert Laporte, and Maria Grazia Speranza. A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation science*, 41(3):382–391, 2007.

[2] Fernanda Medeiros Assef, Maria Teresinha Arns Steiner, and Edson Pinheiro de Lima. A review of clustering techniques for waste management. *Heliyon*, 8(1), 2022.

[3] Sara Ojeda Benítez, Gabriela Lozano-Olvera, Raúl Adalberto Morelos, and Carolina Armijo de Vega. Mathematical modeling to predict residential solid waste generation. *Waste management (Elmsford)*, 28(suppl.1), 2008.

[4] Zuzana Borcinova. Two models of the capacitated vehicle routing problem. *Croatian Operational Research Review*, 8:463–469, 12 2017.

[5] G Clarke and J. W Wright. Scheduling of vehicles from a central depot to a number of delivery points. *Operations research*, 12(4):568–581, 1964.

[6] Leandro C. Coelho, Jean-François Cordeau, and Gilbert Laporte. Thirty years of inventory routing. *Transportation science*, 48(1):1–19, 2014.

[7] G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.

[8] OSRM developers. Osrm api documentation. *https://project-osrm.org/docs/v5.24.0/api/*, Last accessed 06-29-2024.

[9] Xingyu Du, Dongjie Niu, Yu Chen, Xin Wang, and Zhujie Bi. City classification for municipal solid waste prediction in mainland china based on k-means clustering. *Waste Management*, 144:445–453, 2022.

[10] Gulcin Ermis Francesco Alesiani and Konstantinos Gkiotsalitis. Constrained clustering for the capacitated vehicle routing problem (cc-cvrp). *Applied Artificial Intelligence*, 36(1), 2022.

[11] Nicholas E. Johnson, Olga Ianiuk, Daniel Cazap, Linglan Liu, Daniel Starobin, Gregory Dobler, and Masoud Ghandehari. Patterns of waste generation: A gradient boosting model for short-term waste prediction in new york city. *Waste management (Elmsford)*, 62:3–11, 2017.

[12] Scikit learn developers. Agglomerative clustering - scikit-learn 1.5.1 documentation, Last accessed 07-23-2024.

[13] V. Praveen, Dr.P. Keerthika, G. Sivapriya, A. Sarankumar, and Boddu Bhasker. Vehicle routing optimization problem: A study on capacitated vehicle routing problem. *Materials Today: Proceedings*, 64:670–674, 2022. International Conference on Advanced Materials for Innovation and Sustainability.

[14] Antonio López Ropero. Machine learning in waste demand forecast for collection optimization. Master's thesis, Université Grenoble Alpes, 2023.

[15] Fayçal Atik Touzout. Inventory routing problem : Managing demand and travelling time uncertainties, 2021.

[16] Adam Viktorin, Dušan Hrabec, Vlastimír Nevrlý, Radovan Šomplák, and Roman Šenkeřík. Hierarchical clustering-based algorithms for optimal waste collection point locations in large-scale problems: A framework development and case study. *Computers Industrial Engineering*, 178:109142, 2023.