AARHUS UNIVERSITY

Cover sheet

Publication metadata

Solution of the Maximal Covering Tour Problem for locating recycling drop-off stations

Maximiliano Cubillos* and Sanne Wøhlk*

*Department of Economics and Business Economics,
Aarhus University, Denmark

Abstract

Tactical decisions on the location of recycling drop-off stations and the associated collection system are essential in order to increase recycling amounts while keeping operational costs at a minimum. The conflicting nature of the objectives of the problem can be modelled as a bi-objective location-routing problem. In this paper, we address the location-routing problem of recycling drop-off stations by solving the Maximal Covering Tour Problem. To this aim, we propose a heuristic inspired by a variable neighbourhood search. The heuristic is tested on a set of benchmark instances from the TSPLIB and applied to a set of real-life instances from both urban and rural areas in Denmark. Based on the results of the real-life cases, we provide insights on the trade-off between recycling rates and transportation costs.

Keywords: Location; Recycling; Maximal Covering Tour; Heuristics

6.1   Introduction

In this paper, we study a location problem motivated by the situation in which the locations of recycling drop-off stations must be selected in order to increase recycling rates while maintaining low collection costs. The problem is formulated as a Maximal Covering Tour Problem (MCTP), first proposed by Current and Schilling (1994). With this approach, we maximize the covering level of p drop-off stations while minimizing the collection costs estimated as the length of the tour that visits all the stations. We propose a heuristic for solving this problem for large instances, and present results based on real-life data from Denmark.

The collection of materials for recycling is usually performed either via on-site collection or through the use of bring systems (Beullens et al., 2004). The collection policy

depends on the area of service, and it is usually decided by local governments at a municipality level. In Denmark, local governments independently decide on the collection policy in terms of sorting (e.g., separating glass, paper, or plastic) and collection system (on-site, bring in, or mixed). On-site collection systems, where the material is collected directly from the households, are typically modelled as arc routing problems and are described in Ghiani et al. (2015), Mourão and Pinto (2017) and Kiilerich and Wøhlk (2018). In bring systems, the residents bring their waste to containers at predefined drop-off stations, from where it is collected. In such bring systems, there are two main decisions to take. At the tactical level, the number and locations of the drop-off stations must be determined, and at the operational level, routes for the collection vehicles must be planned.

In bring systems, the amount of materials to be collected for recycling is largely affected by public policies and recycling programs. Recycling programs that increase the availability and accessibility of drop-off stations have been reported to significantly increase recycling rates (Kannangara et al., 2018). Hence, to maximize recycling amounts, many drop-off stations are preferred over few. However, increasing the number of drop-off stations involves equipment investment and an increase in operational costs of transportation. Therefore, it is important that the drop-off stations are located carefully in order to balance the quantity of recyclable materials collected and the operational costs.

The balance between the amount of recyclable materials and the related collection costs can differ immensely between rural and urban planning areas. In rural areas, distances tend to be long and households are concentrated in small villages, making collection costs more sensitive to location decisions. In Fig. 6.1, we illustrate the location of drop-off stations for two, rural and urban, postal districts in Denmark. The red triangles indicate the potential locations, the blue triangles show the locations selected in the solutions, the red circles indicate the covered households, and the blue lines show an approximated collection tour between them. The figure shows that with the same number of drop-off stations, the number of covered households and the distances between the stations change significantly from one area to the other. One of the aims of this paper is to provide insights into these differences, measuring the effect of different location decisions.

There are two main factors that should be taken into consideration when deciding on the locations of a given number of recycling drop-off stations. Firstly, the distance between households and the drop-off stations is one of the most important factors determining the willingness of citizens to adopt recycling behaviours (Lange et al., 2014). Hence, the drop-off stations should cover as many households as possible by being located close enough to them. Secondly, the operational cost incurred in relation to collection should be kept at a minimum. Balancing these two objectives results in a location-routing

87

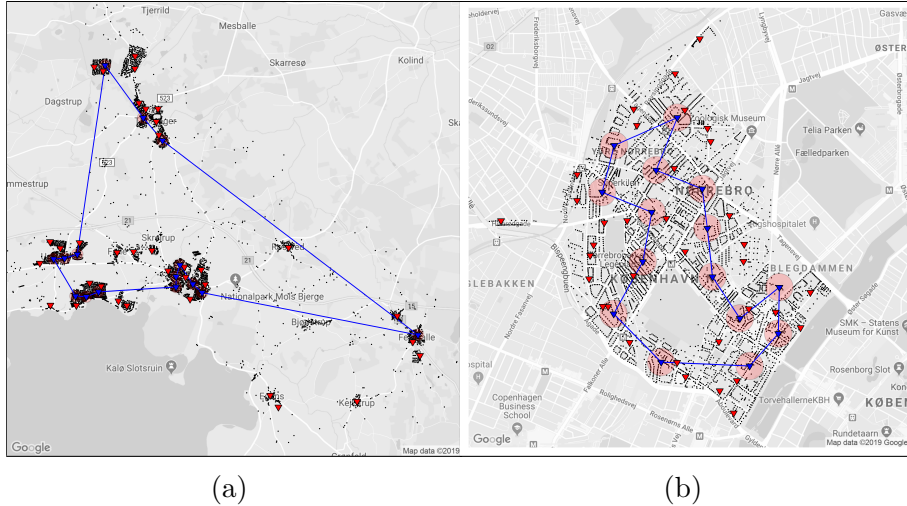|	
(a)	(b)

Figure 6.1: Example of a selection of 15 stations from 50 potential locations in (a) rural and (b) urban areas in Denmark.

problem.

In location-routing problems considering recycling drop-off stations, the proximity of the drop-off stations to the citizens is usually modelled using a cut-off distance as the maximal distance within which a citizen is willing to use the station. If a household is within this cut-off distance, then it is covered, and the material brought to the station is processed. Considering that the covering distances used in the literature vary between 150 and 350 meters (Gautam and Kumar, 2005; Lin and Chen, 2009; Rahim and Sepil, 2014), we adopt 200 meters as the covering distance of a drop-off station in our real-life application.

In terms of operational costs, the drop-off stations should be located in a way that leads to efficient collection routes once the system starts operating. From conversations with various municipalities in Denmark, we note that operational decisions vary significantly among them. In some municipalities, sensors in the recycling containers provide information on current fill levels, and an automated system plans daily routes based on this information. Other municipalities use their gut feeling about the filling at each location and manually plan a daily route. Yet other municipalities use periodic routes. Considering the different collection schemes and considering the location problem as a tactical decision, we use an approximation of the collection costs to study the impact of different location decisions.

Several approaches can be considered to approximate the collection cost. In the case of highly fluctuating filling rates at each drop-off station, Elbek and Wøhlk (2016) and Bogh et al. (2014) argue that the collection of recyclable material is best modelled as an inventory-routing problem. Several algorithms and meta-heuristics have been developed

to solve inventory-routing problems, within a rich area of applications (Coelho et al., 2014). However, at a tactical level of planning, determining the involved costs as a sub-problem of an inventory-routing problem is computationally too expensive. On the other hand, if filling rates at the stations have low variability, the collection can be modelled as a Periodic Routing Problem, and if filling rates do not vary much from one station to another, the collection can be modelled as a Vehicle Routing Problem (VRP) (Toth et al., 2014). To use such approximations of costs, assumptions regarding recycling amounts and vehicle capacities have to be decided upon such that the tactical location problem is not the focal point. For this reason, and in order to keep our contribution as generic as possible, we approximate the collection costs by the Travelling Salesman Problem (TSP). Although it does not provide exact collection costs, this approach enables us to assess the cost effect of selecting near or distant locations, without entailing a need to solve the complex inventory-routing problem as a sub-problem.

Since we are interested in solving real-life sized problems with thousands of households and several potential locations, it is not tractable to solve the TSP sub-problem of the MCTP to optimality. An alternative is to consider the continuous approximation of the TSP tour cost proposed by Daganzo (2005). However, this approximation assumes that the stations are uniformly distributed and, therefore, that the costs only depend on the number of locations and the size of the area considered. By using that approach, the collection cost would be fixed and independent of the actual locations for a given number of stations, failing to consider the impact of selecting different locations on the total cost. We therefore choose to approximate the collection cost by heuristically solving the TSP of the tour that visits all the stations to be located.

To summarize, we study an application of the MCTP where we need to determine the location of a set of p recycling drop-off stations in such a way that the number of citizens covered is maximized, and the total distance of the TSP tour connecting the selected locations is minimized. As this is a bi-objective problem, we use a weighted objective function. The contribution of this study is twofold. Firstly, we propose a heuristic for solving the MCTP for real-life sized problems and validate its performance by comparing the heuristic solutions to optimal solutions obtained by CPLEX for a total of 324 different problem instances based on benchmark instances from the TSPLIB (Reinelt, 1991). Secondly, we use our heuristic to study the problem of locating recycling drop-off stations using real-life data from Denmark. For the real-life case, we propose a procedure to select potential locations, study the trade-off between recycling and collection costs, and compare results by considering urban and rural areas.

The rest of this paper is organized as follows. We first review the related literature in Section 6.2. Section 6.3 presents the model formulation of the MCTP, and in Section 6.4,

we propose our solution method. In Section 6.5, we present the numerical results using benchmark data to assess the performance of our heuristic. In Section 6.6, we use real-life data from different areas in Denmark, and we discuss the implications of our results on the recycling problem. Finally, Section 6.7 concludes.

6.2 Literature review

The Maximal Covering Tour Problem (MCTP) is a bi-objective location-routing problem and it was first proposed by Current and Schilling (1994). In the problem, a tour must visit p nodes out of n potential locations in a network. The two objectives are to minimize the total length of the tour and to maximize the covered demand. In order to address the bi-objective nature of the model, Current and Schilling (1994) propose a heuristic to generate an approximate set of efficient optimal solutions. In spite of the large application potential of the MCTP, related works have focused on variants of the problem rather than on solving the original problem.

Gendreau et al. (1997) propose a single-objective variant of the MCTP referred to as the Covering Tour Problem (CTP). In the CTP, the covering objective of the MCTP is replaced with a constraint that requires complete covering of a given subset of the nodes. The constraint of having exactly p locations to be placed is replaced by the construction of two subsets of nodes: one subset of nodes that must be visited in the tour, and another for which the visit is optional. The authors propose an exact branch-and-cut algorithm and a heuristic to solve the problem. A limitation of the CTP is that it is restricted to problems in which several assumptions about the nodes have to be made, namely, the mandatory demand that must be covered, and the nodes in the tour that must be visited. In our problem, this is not the case since there are no predetermined households that must be covered, and there is an ample range of potential locations where stations can be placed.

Berman et al. (2003) propose a variant of the MCTP that considers a gradual decay covering function. The authors define two critical distances, a lower distance in which a location is fully covered and a larger distance where locations are not covered. In between the two distances, a gradual coverage decreasing from full coverage at the lower distance to no coverage is assumed. Jozefowiez et al. (2007) propose a bi-objective variation of the MCTP where the covering part of the objective function is replaced by the largest distance between a node of some given set and the nearest visited node. The authors propose a two-phase cooperative strategy that combines a multi-objective evolutionary algorithm with a branch-and-cut algorithm to find sets of efficient optimal solutions. Tricoire et al. (2012) formulate a variant of the MCTP considering stochastic demand where the two

objectives are given by cost (opening and routing costs) and expected uncovered demand. They propose a branch-and-cut algorithm within an epsilon-constraint algorithm to find optimal sets. For a general overview of location-covering problems see Laporte et al. (2015) and Church and Murray (2018).

The field of applications of the MCTP and its variants has mainly focused on emergency supply and disaster relief problems. Hodgson et al. (1998) propose a CTP model for planning mobile health care facilities. The model minimizes a mobile facility's travel while serving all population centers within range of a feasible stop depending on weather conditions. Doerner et al. (2007) propose a three-objective covering tour model for mobile health care units in a developing country and solve it by using Genetic Algorithms and Ant Colony optimization. Nolz et al. (2010) formulated the problem of the delivery of drinking water to the affected population in a post-disaster situation as a multi-objective covering tour problem. Naji-Azimi et al. (2012) propose a generalization of the Covering Tour Problem by considering split delivery for the location of satellite distribution centers to supply humanitarian aid and propose a local search heuristic to solve large sized instances efficiently. Finally, Abounacer et al. (2014) propose an exact solution approach to generate the set of efficient solutions of a three-objective covering tour model for disaster response.

In the last few years, there has been an increase in the application of location models in multi-criteria optimization models for waste management problems (Coelho et al., 2017). Those models mainly consider the problem of locating processing plants and waste deposits (Khan, 1987; Antunes et al., 2008), but little attention has been paid to the location of recycling drop-off stations (Purkayastha et al., 2015). Studies that address the location problems in a recycling context have mainly used multi-objective approaches combined with Geographical Information Systems (GIS), and only a few applications of covering problems in the area of recycling have been published (see, for example, Valeo et al. (1998); Gautam and Kumar (2005); Lin et al. (2010)). Regarding applications for recycling drop-off stations, Chang and Wei (1999) propose a multi-objective evaluation of the trade-off between the number and size of drop-off recycling stations, the citizens covered in the service network, the average walking distance to the drop-off stations for the citizens, and the distance travelled by collection vehicles.

Within facility location models used in relation to waste management, the concept of covering has been applied in different problems. Bautista and Pereira (2006) formulate the problem of locating waste collection sites using set covering formulations. They propose a genetic algorithm with four variations of the set covering formulation and show results for real-life instances from Barcelona. In a different approach, Farhan and Murray (2006) consider the location of both desirable and undesirable facilities simultaneously,

where recycling stations are considered as undesirable facilities. They propose the Maximal/Minimal Covering-Distance Decay Problem to locate $p = 38$ park-and-ride facilities and recycling facilities in Columbus, Ohio. The location of recycling facilities can be seen from two different approaches depending on whether existing facilities are considered or not. Ye et al. (2011) study the problem to reduce the number of existing recycling centers from 79 to 2 centers using a 2-stage location set covering–p-median problem. First, they locate the recycling centers using set covering and then assign the collection depots using a p-median model. They propose a greedy algorithm and present results for a case study in Taiwan. One of the few studies that apply the location-routing problem using maximal covering is the study by Rahim and Sepil (2014), which formulates a combined maximal covering location problem in the presence of partial coverage and a selective TSP to determine the location of bottle banks. The objective of the problem is to maximize the profit of a glass recycling company, and a nested heuristic based on a variable neighbourhood search is proposed as a solution method. The authors present results using $p = 1$ to $p = 20$ for TSPLIB instances, and a case study using $p = 12$ locations. In Erfani et al. (2017), the location-allocation problem and capacitated vehicle routing problem are solved using GIS. The ESRI ArcGIS network analysis extension was used for the analysis on maximum walking distances and covered demand. Erfani et al. (2018) propose a two-step model that first minimizes the number of facilities using set covering, and then maximizes the demand covering using maximal capacitated covering with different values of facility locations. They perform statistical analysis of the results analysing total service covering, total attendance derived by maximize attendance analysis, and surplus devoted capacity for 26 locations.

6.3   Model formulation

In order to formally define the MCTP, we consider a directed complete graph $G = (N, A)$, where, in our application, each node $i \in N$ represents a household that produces a certain amount of recycling material a_i. Each arc $(i, j) \in A$ is associated with a distance d_{ij}, where $d_{ii} = +\infty$.

A subset of the nodes, $V \subseteq N$, are potential nodes for the location of drop-off stations. The objective is to find a subset $V' \subseteq V$ of p nodes that minimizes the tour length and simultaneously maximizes the covering of the nodes in N. A node $i \in N$ is said to be covered if the distance between i and any node $j \in V'$ is less than a predefined maximal covering distance S. For each node $j \in V$, we define the set $N_j = \{i \in N \mid d_{ij} \leq S\}$, that contains all nodes i that can be covered by j.

To model the problem, we define the following two types of variables. For all pairs

$i, j \in V$, we define x_{ij} as a binary variable with a value equal to 1 if the arc from i to j is in the selected tour, and 0 otherwise. For all nodes $i \in N$, we define a binary variable y_i, which is 1 if the node i is not covered by any node $j \in V'$ in the selected tour, and 0 otherwise. Using this notation, we can formulate the MCTP as follows:

$$\text{Min} \quad Z = (Z_L, -Z_C) \tag{6.1}$$

Subject to

$$\sum_{i \in V} x_{ij} - \sum_{k \in V} x_{jk} = 0 \quad \text{for all} \quad j \in V \tag{6.2}$$

$$\sum_{i \in Q} \sum_{j \in Q} x_{ij} \leq |Q| - 1 \quad \text{for all} \quad Q \subset V \text{ such that} \quad 2 \leq |Q| < p \tag{6.3}$$

$$\sum_{i \in V} \sum_{j \in V} x_{ij} = p \tag{6.4}$$

$$\sum_{l \in V} \sum_{j \in N_i} x_{lj} + y_i \geq 1 \quad \text{for all} \quad i \in N \tag{6.5}$$

$$x_{ij} \in \{0, 1\} \text{ for all } (i, j) \in A \tag{6.6}$$

$$y_i \in \{0, 1\} \quad \text{for all } i \in N \tag{6.7}$$

where,

$$Z_L = \sum_{i \in V} \sum_{j \in V} d_{ij} x_{ij}$$

$$Z_C = \sum_{i \in N} a_i (1 - y_i)$$

The two objectives, Z_L and Z_C, minimize the tour length and maximize the covered demand, respectively. Constraint set (6.2) ensures that the tour leaves a node if it enters it. Constraint set (6.3) eliminates subtours in the solution. These constraints are the usual TSP subtour elimination constraints with the added limitation that the number of nodes in Q must be less than p. Constraint (6.4) ensures that there are exactly p nodes on the tour, and constraint sets (6.5) ensure that $y_i = 1$, for all nodes i which are not covered by the tour. Finally, constraints (6.6)–(6.7) are the domain constraints.

Since the MCTP has two conflicting objectives, there is no single optimal solution. Instead, a set of efficient solutions representing the trade-off between the two objectives can be found. An efficient solution is one for which an improvement in one objective requires a degradation of the second one. Since the number of efficient solutions can grow

exponentially with the number of nodes in the network (Current and Schilling, 1994), approximation methods for finding the trade-off curve must be applied. A well-known approximation method for multi-objective problems that does not alter the structure of constraints (6.2)–(6.7) is the weighting method (Zadeh, 1963). With this approach, the two conflicting objectives are combined into a single objective by adding a positive parameter α, with $0 \leq \alpha \leq 1$, that defines a convex combination of the two objectives. The model is then solved using different sets of relative weights to generate an approximate trade-off curve between the two objectives. In this study, we focus on presenting a heuristic to solve the problem for fixed values of α rather than on finding the complete set of efficient solutions. Using this approach, the objective function we minimize for the MCTP is:

$$\text{Min} \quad Z' = \alpha Z_L - (1 - \alpha)Z_C \tag{6.8}$$

6.4 Solution approach

In our problem, the number of nodes in the network corresponds to the number of households for a recycling planning area, which may be a significantly large number. In our experiments, we found that the model (6.2)–(6.8) can be solved optimally for small values of the total number of nodes in the network and the number of nodes in the tour.

Considering the limitations of solving the MCTP to optimality for real-life size problems, we propose a heuristic solution approach inspired by a Variable Neighbourhood Search (VNS). The key idea of the heuristic is to systematically diversify an incumbent solution using a shaking step followed by a local search. These two steps are embedded into a main step, which is repeated until a maximal number of iterations is performed. The application of VNS approaches for location problems (Hansen and Mladenović, 1997; Mladenović et al., 2003) and location-routing problems (Pérez et al., 2003; Rahim and Sepil, 2014) has yielded good results previously. The pseudo-code of our heuristic is presented in Algorithm 6.2, and it is explained in the following.

We represent a feasible solution as a subset $V' \subseteq V$ containing p nodes, and an associated vector $\mathbf{x}$, of length p, dictating the order in which they are visited in the tour. The heuristic is initialized with a solution $\mathbf{x}$ obtained by a greedy construction heuristic. In the shaking step (line 5), k with $0 < k \leq p$ nodes from the incumbent solution $\mathbf{x}$ are randomly dropped and replaced with k nodes randomly selected from the set of potential locations. This change of neighbourhood is intended to perturb the incumbent solution to escape from local minima and provide a starting point for the local search. In the local search (lines 7–16), we obtain the best objective value from all the neighbour solutions reachable from the incumbent solution by performing a one-swap move. The local search

continues as long as improvements are found.

The main step (lines 4–22) involves a systematic shake that changes k nodes in the incumbent solution. This systematic change can be briefly described as follows. The main step starts with $k = k_{min}$, the minimum size of the shaking. If a better solution is found, we increase $k \leftarrow k + 1$ and repeat the shaking step. Otherwise, we set $k \leftarrow k_{min}$. The process continues until $k = k_{max}$. Based on experiments for different parameters for the main step, we selected $k_{min} = 2/3 \; p$ and $k_{max} = p$. The latter means that the shake can get as large as the total number of locations in the solution.

Each time a set of nodes V' has been identified, both the tour length and tour coverage have to be computed. To compute the covering of the nodes in the tour is straightforward, but determining the length of the tour is equivalent to solving a TSP, which is NP-hard. After each shaking step where we drop k nodes from the solution, we solve the TSP to optimality using CPLEX (line 6). This serves as a base for the local search (lines 7-16). In order to reduce the number of times the heuristic has to solve a TSP to optimality, we approximate the tour length in the local search step during our search for the best neighbour, by using the nearest insertion rule for each potential node to be included (line 9). After identifying the best neighbour, we re-optimize the TSP tour using CPLEX in order to proceed with the optimal tour in the next iteration of the local search (line 10).

6.5 Computational results for benchmark data

In this section, we present the computational results of the proposed heuristic and evaluate the performance of the heuristic using benchmark instances. We compare the results to the optimal solutions obtained by solving the model presented in Section 6.3 using CPLEX. To the best of our knowledge, no benchmark instances exist specifically for the MCTP. We therefore use a subset of instances from the TSPLIB as the basis for our comparison. This subset consists of four instances containing 100 nodes (kroA100, kroB100, kroC100, and kroD100) and two instances with 200 nodes (kroA200 and kroB200). For each of them, the subset V of potential locations was randomly selected using $|V| = 25$ and $|V| = 50$ nodes, resulting in 12 instances. Our heuristic was coded in Java 1.8.0_201 and run on a 3 GHz Intel X5450 processor, 24 GB RAM. Each result of our heuristic is an average of 5 runs, and the exact solutions were obtained using CPLEX V12.8.0 with a maximum run time of 2 hours.

The problem parameters used for the comparison were selected according to three criteria. First, S was selected such that each node in V covers at least two nodes in N. This means that each potential location can cover at least two demand nodes in the network. For the value of p, we selected values for which solutions can cover between

Algorithm 6.2 Heuristic

1: $\mathbf{x} \leftarrow$ Get greedy initial solution
2: **for** iter $= 1 :$ MaxIter **do**
3:      $k \leftarrow k_{min}$
4:      **while** $k < k_{max}$ **do**
5:          $\mathbf{x}' \leftarrow$ ShakeRandom($\mathbf{x}$, k)
6:          $\mathbf{x}' \leftarrow$ TSP exact solution of $\mathbf{x}'$
7:          improvement $\leftarrow$ true
8:          **while** improvement **do**
9:              $\mathbf{x}'' \leftarrow$ Get best approximate solution in the neighbourhood of $\mathbf{x}'$
10:              $\mathbf{x}'' \leftarrow$ TSP exact solution of $\mathbf{x}''$
11:              **if** $f(\mathbf{x}'') < f(\mathbf{x}')$ **then**
12:                  $\mathbf{x}' \leftarrow \mathbf{x}''$
13:              **else**
14:                  improvement $\leftarrow$ false
15:              **end if**
16:          **end while**
17:          **if** $f(\mathbf{x}') < f(\mathbf{x})$ **then**
18:              $\mathbf{x} \leftarrow \mathbf{x}'$
19:              $k \leftarrow k_{min}$
20:          **else**
21:              $k \leftarrow k + 1$
22:          **end if**
23:      **end while**
24: **end for**

$N/2$ and N nodes. This criteria is intended to obtain non-trivial solutions. The weighting parameter α was selected such that we can solve three scenarios: covering-oriented, collection-oriented, and balanced. Three values for each parameter were selected in order to analyse the impact of each of them in the heuristic performance. The selected values are $S \in \{600, 700, 800\}$, $p \in \{4, 6, 8\}$, and $\alpha \in \{0.001, 0.01, 0.1\}$. This selection and the rest of the results consider euclidean distances between nodes in the network. Thereby, we tested 27 problems for each of the 12 instances, resulting in a total of 324 problem instances. Regarding the demands of the nodes, we considered them to be $a_i = 1$ for all nodes $i \in N$.

In Table 6.1, we show detailed results for a selected instance for each value of s, p and α. The table shows the average objective value and average and maximal run times in seconds for the heuristic over 5 runs, and in addition it shows the results obtained by CPLEX as well as the gap % calculated as the percentage difference for our heuristic and the optimal solutions, respectively. In the cases where the maximal run time of CPLEX was reached, the upper and lower bounds of the solutions are reported, and no optimality gap is reported. Table 6.2 presents a comparison of the aggregated results of our heuristic and the optimal solutions obtained by CPLEX. Each line corresponds to the average results of 27 variations of each instances as shown in Table 6.1. The gap % is computed as the average gap of the instances for which CPLEX solved the problem to optimality.

As can be seen in Tables 6.1 and 6.2, our heuristic can effectively find good quality solutions. In addition, our heuristic shows lower average and lower variability in CPU times. As shown in Table 6.1, the CPLEX results show a high variability in times depending on the weighting parameter. In particular, increased run times were found in the balanced scenario ($\alpha = 0.01$) for all test problems, which reached the maximal run time in most cases. These increases in run times indicate that the problem resulting from the need to balance covering and travel cost is generally harder to solve than the problem that focuses solely on either of the two objectives. In 6 of the 8 instances in which CPLEX could not solve the problem to optimality within the time limit, the solution of our heuristic was better than the best solution from CPLEX. When considering the total 324 test instances, in 50 instances CPLEX could not solve the problem to optimality, and from those instances, our heuristic performed better in 36 cases. These results show that our heuristic yields high quality results for the MCTP in short computation times.

S	p	α	Heuristic			CPLEX			Gap(%)
			Avg. Obj. Val.*	Avg. Time*	Max. Time	Obj. Val.	Lower Bound†	Time	
600	4	0.001	45.1	3.4	4.1	45.1	-	1.7	0.0
		0.01	83.6	6.4	10.3	83.6	-	5786.3	0.0
		0.1	134.0	5.6	7.5	134.0	-	8.7	0.0
	6	0.001	27.1	6.9	9.4	27.1	-	9.3	0.0
		0.01	84.0	8.3	9.7	86.9	61.8	7204.8	-
		0.1	169.2	12.0	15.1	169.2	-	8.8	0.0
	8	0.001	15.6	9.0	13.7	15.6	-	8.6	0.0
		0.01	84.0	10.2	12.8	90.3	56.2	7215.8	-
		0.1	250.2	16.1	17.3	250.2	-	97.7	0.0
700	4	0.001	35.7	3.6	4.3	35.7	-	8.0	0.0
		0.01	78.1	4.7	5.5	78.1	63.4	7212.1	-
		0.1	131.3	5.5	6.3	131.3	-	21.0	0.0
	6	0.001	17.2	6.8	8.1	17.2	-	12.7	0.4
		0.01	76.9	9.7	11.7	81.2	47.7	7209.6	-
		0.1	166.5	10.6	12.0	166.5	-	25.1	0.0
	8	0.001	10.1	7.4	8.3	10.0	-	35.3	0.6
		0.01	77.0	10.1	11.6	78.6	44.4	7209.3	-
		0.1	248.4	15.4	18.5	248.4	-	139.9	0.0
800	4	0.001	25.7	3.6	4.6	25.7	-	7.3	0.0
		0.01	72.1	4.5	5.3	72.0	52.1	7205.9	-
		0.1	128.6	7.7	9.9	128.6	-	10.5	0.0
	6	0.001	11.7	6.3	8.2	11.5	-	21.5	2.1
		0.01	70.9	9.2	11.7	75.0	40.3	7208.4	-
		0.1	165.6	10.7	12.5	165.6	-	20.2	0.0
	8	0.001	8.7	8.2	8.9	8.7	-	4977.2	0.9
		0.01	71.6	9.9	11.0	72.7	36.7	7208.7	-
		0.1	247.5	16.3	17.9	247.5	-	117.5	0.0
Global average			93.9	8.4	10.2	94.7	-	2555.3	0.2

Table 6.1: Detailed computational results for the instance kroA100, $|V| = 50$. *Average results over 5 runs of the instance. †A '-' indicates an optimal solution is obtained.

Name	$|V|$	Heuristic		CPLEX			Gap(%)
		Avg. Time	Max. Time	Avg. Time	Max. Time	# opt.*	
kroA100	25	7.6	16.1	79.8	669.3	27	0.07
kroA100	50	8.4	18.5	2,555.3	7,215.8	19	0.15
kroA200	25	7.9	15.0	494.8	4,923.2	27	0.15
kroA200	50	8.4	19.0	3,044.1	7,210.1	16	0.04
kroB100	25	7.8	14.0	310.6	2,857.0	27	0.07
kroB100	50	8.7	19.0	2,677.9	7,216.1	19	0.19
kroB200	25	7.3	14.8	1,439.9	7,200.3	27	0.19
kroB200	50	8.5	20.1	2,760.3	7,216.3	20	0.04
kroC100	25	8.0	18.4	1,012.8	7,206.9	27	0.04
kroC100	50	9.4	23.7	2,486.9	7,208.0	19	0.30
kroD100	25	8.0	14.5	112.0	923.5	27	0.00
kroD100	50	8.5	17.7	2,523.4	7,211.5	19	0.70
Global average		8.2	17.6	1,624.8	5,588.2	22.8	0.2

Table 6.2: Aggregated computational results for the benchmark instances. *Number of instances solved to optimality by CPLEX out of 27 instances.

6.6 Results for real-life data

In this section, we apply our heuristic to large-scale real-life data. We first present the data selected from five areas in Denmark in Section 6.6.1. In Sections 6.6.2 and 6.6.3, we explain the procedure we use to select potential locations for the drop-off stations and the scale used to compare urban and rural areas. Finally, Sections 6.6.4–6.6.6 present results regarding the trade-off between covering and distance, sensitivity analysis on our procedure to select potential locations, and managerial insights for urban and rural areas.

6.6.1 Data from Denmark
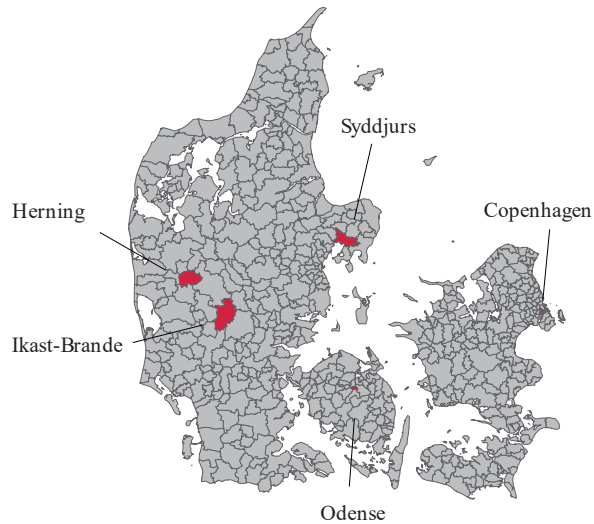
Figure 6.2: Selected postal code areas from Denmark.

We selected one postal code area from each of five municipalities in Denmark: Copenhagen, Odense, Herning, Ikast-Brande, and Syddjurs. In Fig. 6.2, we show the map of Denmark highlighting the selected postal code areas used in this study. The selected areas allow us to study the effect of different geographical layouts, namely urban and rural, with a comparable number of households. In urban areas (Copenhagen and Odense), household density is larger and distances between households are shorter. In the rural cases (Herning, Ikast-Brande and Syddjurs), households tend to concentrate in small villages with larger distances between them. Our data consists of the coordinates of the households for each of the selected areas. In Table 6.3, we provide the postal codes, type of area (Urban/Rural), number of households (N), and the average and maximal distances, in meters, between any pair of households in the data set.

Name	Postal code	Type	$|N|$	Avg. dist. (m)	Max. dist. (m)
Copenhagen	2200	Urban	5,186	1,015	3,066
Odense	5230	Urban	6,032	1,296	3,841
Herning	7480	Urban/Rural	3,395	3,352	14,994
Ikast-Brande	7330	Rural	4,816	4,183	21,321
Syddjurs	8410	Rural	4,853	4,683	17,610

Table 6.3: Description of the data sets from the five selected zones in Denmark.

6.6.2 Selecting potential locations
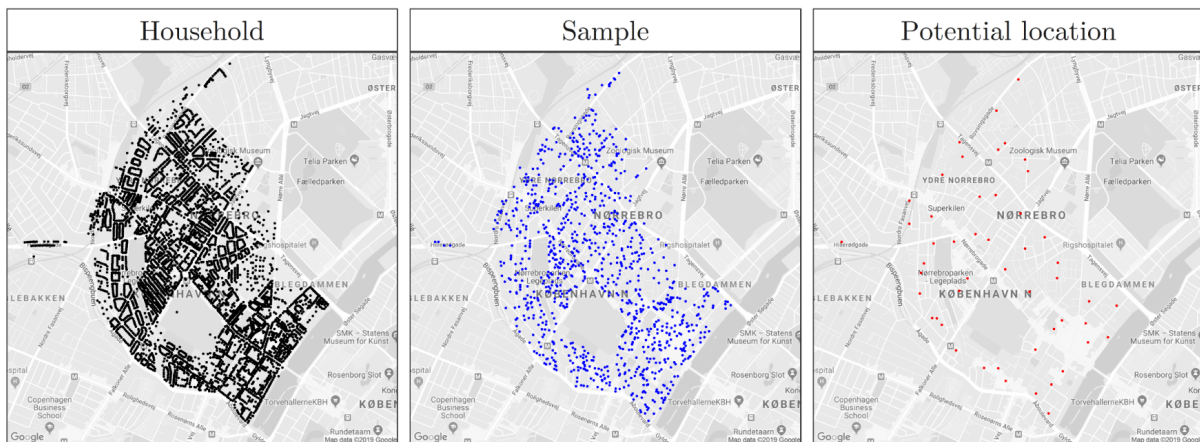
Figure 6.3: The process for selecting potential locations from the household data from the selected area in Copenhagen.

Before solving the MCTP for these large instances, we need to determine the set of potential locations (V) for drop-off stations. We use a 2-step procedure to select a set of well-dispersed potential locations chosen so as to maintain problem tractability. First, we randomly sample a set of $M \subseteq N$, $|M| = 1000$ nodes from the total number of nodes. Second, based on the sampling M, we determine the set $V \subseteq M$ using an aggregation demand method similar to Francis et al. (1999). In this step, we solve the Maximal Covering Problem (MCP) (Church and ReVelle, 1974) for a large number of locations ($p = 50$) using the set M as the set of nodes. The MCP is a maximization problem that seeks to select a subset of p locations from $|M|$ potential stations that maximizes the total covering. That is, the MCP corresponds to the MCTP, but does not consider the tour length.

We formulate the MCP as follows. For all $i \in M$ we define the binary variable y_i with value equal to 1 if the node i is covered by any selected location. For all $j \in M$ the variable x_j is equal to 1 if the node j is selected as a location. The MCP is then

formulated as:

$$\text{Max} \quad \sum_{i=1}^{|M|} y_i \tag{6.9}$$

Subject to

$$\sum_{j \in N_i} x_j \geq y_i \quad \forall \, i \in M \tag{6.10}$$

$$\sum_{j \in M} x_j = p \tag{6.11}$$

$$x_i, y_i \in \{0, 1\} \quad \forall \, i \in M \tag{6.12}$$

The objective (6.9) maximizes the total demand covered by the selected p locations. Constraints (6.10) define if a specific demand node i is covered or not. Constraint (6.11) ensures that exactly p locations are selected, and constraints (6.12) are the binary domain constraints. In our problem, we define $p = 50$ as the number of potential locations for drop-off stations, a covering radius of $S = 200$, and the nodes with $x_i = 1$ constitute the set V of potential locations. In Fig. 6.3, we show the households ($|N| = 5,186$), the sampled locations ($|M| = 1,000$), and the resulting potential locations ($|V| = 50$) for the area of Copenhagen.

6.6.3 Scaling urban and rural areas

The order of magnitude of the two objectives, distance and covering, varies depending on the area. In rural areas, where distances between households are large, the distance values in the objective function are increased compared to the urban areas. In the same fashion, since households are more uniformly distributed in urban areas, the covering values are greater in urban areas for the same covering radius. This means that in order to compare results between urban and rural areas, we have to scale the weights in the objective function accordingly with the balance between distance and covering. To scale the two objectives, we determine an upper bound for the distance and the covering for each area by solving the MCTP problem with $\alpha = 0$, i.e. only maximizing covering. We use the results of the upper bound case as an approximation of the order of magnitude of the two objectives. In Table 6.4, we provide the upper bound results for distance and covering for $p = 15$, and the proportion in the objective value of the covering (k), for each instance, respectively. We can see that the relative proportion of the covering from the total sum changes significantly between rural and urban instances. Thus, using the same α to test similar scenarios can be misleading. Instead, we use a parameter β that

Upper bound results ($\alpha = 0$)					$\beta = 0.5$			
Name	Cov.*	Dist.	Sum	k (Cov./Sum)	β	$\alpha = 2k\beta$	$1 - \alpha$	α Dist.	$(1 - \alpha)$ Cov.
Copenhagen	3,393	6,427	9,820	0.34		0.35	0.65	2,221	2,221
Odense	2,722	8,231	10,953	0.25		0.25	0.75	2,045	2,045
Ikast-Brande	1,713	7,444	9,157	0.19	0.5	0.19	0.81	1,392	1,392
Herning	1,748	18,175	19,923	0.09		0.09	0.91	1,594	1,594
Syddjurs	1,775	31,455	33,230	0.05		0.05	0.95	1,680	1,680

Table 6.4: Upper bound results ($\alpha = 0$, $p = 15$) are used as an approximation of the order of magnitude to scale α for the different instances. *Number of covered households.

accounts for the approximation of the proportion between distance and covering. We compute the values of the weighting parameter α as:

$$\alpha = 2k\beta \tag{6.13}$$

where k is the ratio of the covering results over the sum of covering and distance. To illustrate this, Table 6.4 shows the corresponding values of α when considering the case in which we want the distance and the covering to have the same value ($\beta = 0.5$) and the effect on the two parts of the objective value. For the rest of the analysis, we obtain results using $\beta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

6.6.4 Covering and distance trade-off

We use our heuristic to obtain results on covering, tour distance, and the total run time in seconds, for different weighting parameters, for each of our five instances. The covering was computed as the number of households covered as a percentage of the total number of households of the instance. In Table 6.5, we illustrate the trade-off between covering and distance for the selected area in Copenhagen using $p = 15$ drop-off stations. In addition, we present the percentage decrease from the upper bound solution ($\beta = 0$) to compare the relative change of the solution when increasing the weighting parameter on the distance.

From a managerial point of view, the number of recycling drop-off stations (p) reflects the size and installation cost of the recycling network. At the same time, the tour distance approximates the transportation costs, and the covering represents the recycling level. Fig. 6.4 allows us to directly compare different configurations regarding installation costs, operational costs, and recycling levels. For each value of $p \in \{10, 15, 20, 25\}$, we obtain the trade-off between covering and tour distance. In all cases, our results show a non-linear relationship between covering and distance with decreasing increments. This means that an increase in covering for a fixed number of stations has a higher impact on distance

		Heuristic results			% decrease from u.b.*	
β	α	Dist.	Cov. (%)	Time (sec)	Dist. (%)	Cov. (%)
0	0	6,427	65.43	45.5	-	-
0.1	0.06	6,427	65.43	98.5	0.00	0.00
0.2	0.13	5,884	64.58	106.2	8.45	1.30
0.3	0.19	6,047	64.85	108.3	5.92	0.88
0.4	0.25	5,361	61.24	112.0	16.60	6.40
0.5	0.31	4,675	56.48	114.7	27.26	13.68
0.6	0.38	4,154	51.08	116.6	35.38	21.93
0.7	0.44	3,539	43.27	168.3	44.94	33.86
0.8	0.50	3,453	41.79	200.5	46.28	36.13
0.9	0.56	3,453	41.79	173.7	46.28	36.13

Table 6.5: Results for Copenhagen using $p = 15$ drop-off stations for different values of β. *Computed as the percentage difference compared with the upper bound case ($\alpha = 0$).

when aiming at higher values of covering.

Although all results show similar trade-off trends between covering and distance, there are differences when comparing urban and rural areas. In Fig. 6.5, we show the percentage decrease from the upper bound case ($\beta = 0$), in both distance and covering, for each instance. Each of the lines in Fig. 6.5 corresponds to the results of the last two columns in the example given in Table 6.5. One main difference between the two types of areas is that the range in which we obtain results for covering is smaller for rural areas. This means that, independently on the weighting parameters and the number of drop-off stations, different solutions have less impact on the total covering of the selected locations. We illustrate this difference in Fig. 6.6, where we compare the solution for the upper bound case ($\beta = 0$) and the balanced case ($\beta = 0.5$) for each area. Given that rural areas have small villages in which most of the households are concentrated, a solution adding a new village has a large impact on the distance. For example, in Herning the upper bound case includes three villages, while increasing the weight of the distance reduces the tour into a single village.

6.6.5 Stability of solution sets

From a managerial point of view, it is important to understand how the selected locations change when considering different balances in the objective function and different areas. From our experiments, we see that for rural areas, locations tend to change less when we increase the weight on the distance ($\beta > 0.3$) compared to urban areas. To quantify this difference, we used the Hamming distance to count the number of locations that differ from the upper bound solution. In Fig. 6.7, we show the Hamming distance for each instance using $p = 15$. The distance increases when increasing the weighting parameter
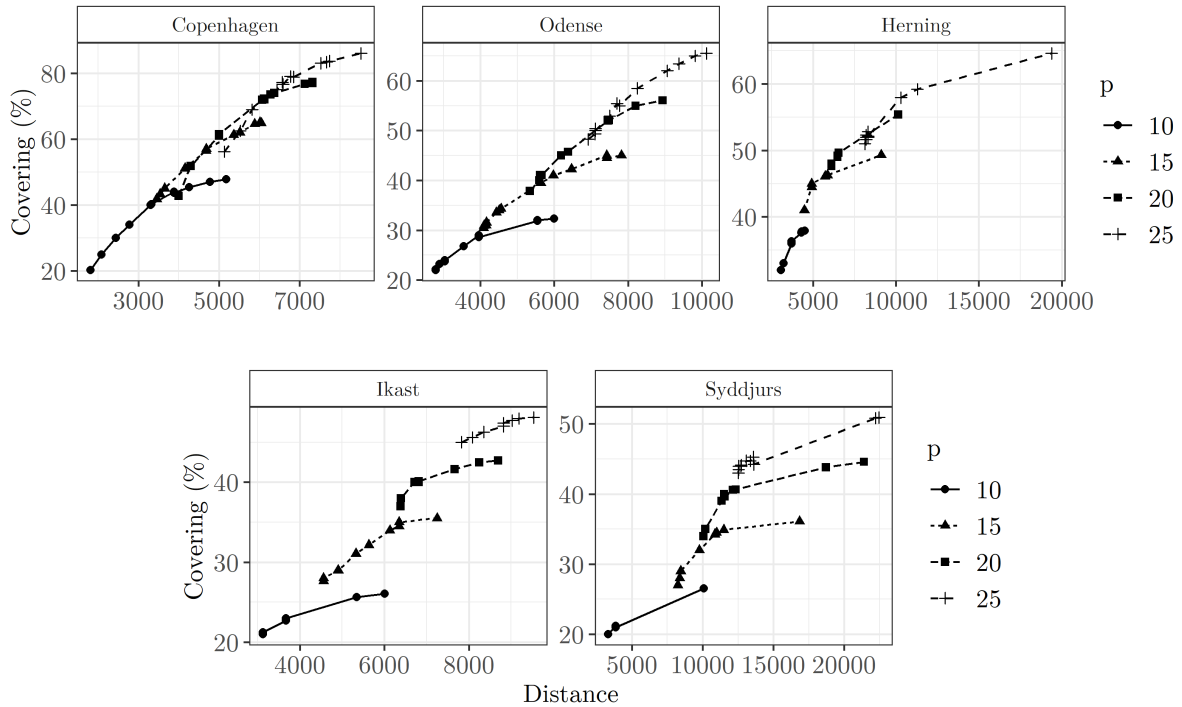
Figure 6.4: Trade-off between distance and covering for different numbers of drop-off stations ($p = 10, 15, 20, 25$) for each instance.

β, but the increase is greater and has more variability for urban areas. To illustrate this location by location, we show two set of solutions, for Copenhagen (urban) and for Syddjurs (rural) in Fig 6.8. In the figure, each row represents the selection of 15 locations (in grey) over the 50 potential locations. As can be seen, the selection of locations is more sensitive to changes in β for the urban case. In urban cases, potential locations cover similar number of households making potential locations equally good in the solution. In the rural case, where rather few locations concentrate most of the covering in small villages, the impact of changing β is lower. Finally, this visualization allows us to identify the locations that are chosen independently of the choice of β, which can serve as a criteria for location selection.

6.6.6 Sensitivity analysis on sampling

In order to validate the robustness of our results, we perform sensitivity analysis on the parameters for the process for selecting the potential locations. We analyse the effect of three different parameters: random sampling, number of random samples, and the number of potential locations. First, we analyse the impact of the sampling process by comparing the results for each instance. We replicate our results for 10 different random
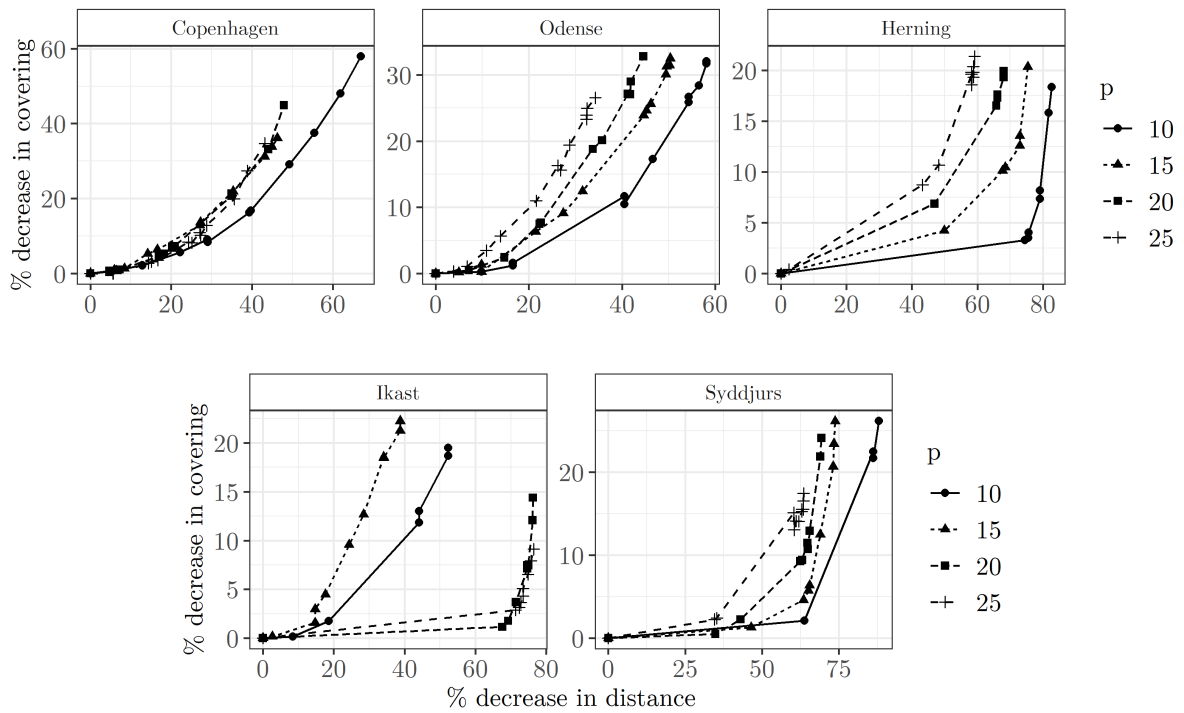
Figure 6.5: Trade-off between percentage decrease in covering and percentage decrease in distance, computed as the decrease from the upper bound case ($\beta = 0$), for $p = 10, 15, 20$, and 25.
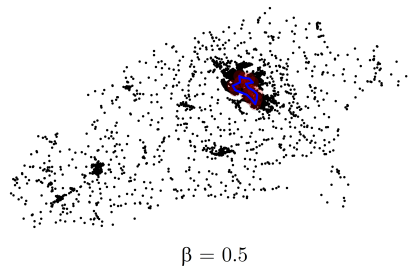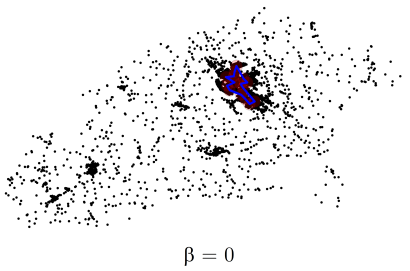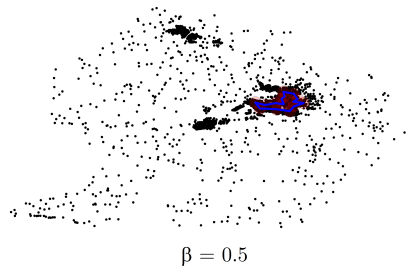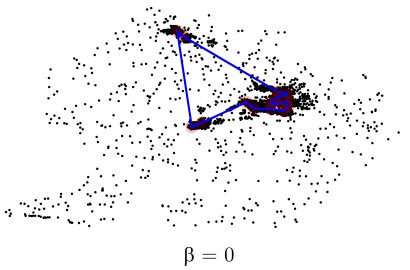
Copenhagen

β = 0   β = 0.5

Odense

β = 0   β = 0.5

Ikast-Brande

β = 0   β = 0.5

Herning

β = 0   β = 0.5

Syddjurs

β = 0   β = 0.5
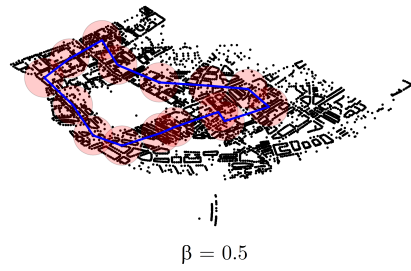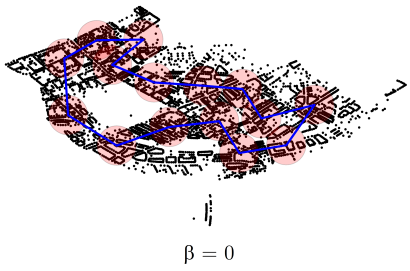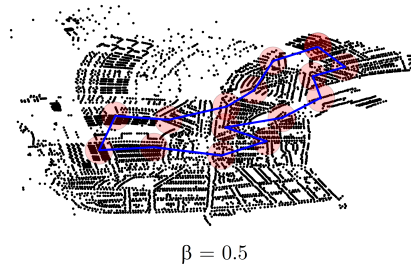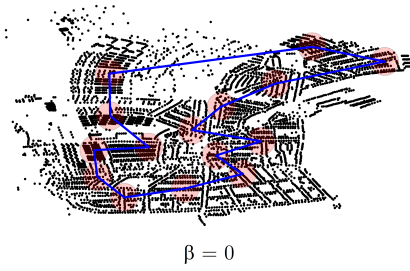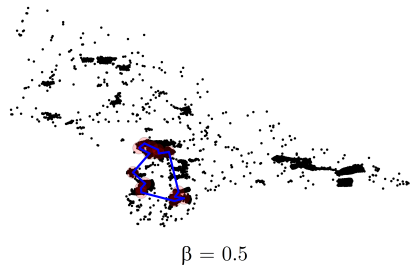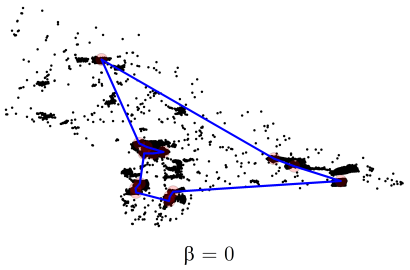
Figure 6.6: Comparison of the solution of the upper bound case ($\beta = 0$) and the balanced case ($\beta = 0.5$), for $p = 15$ recycling drop-off stations.
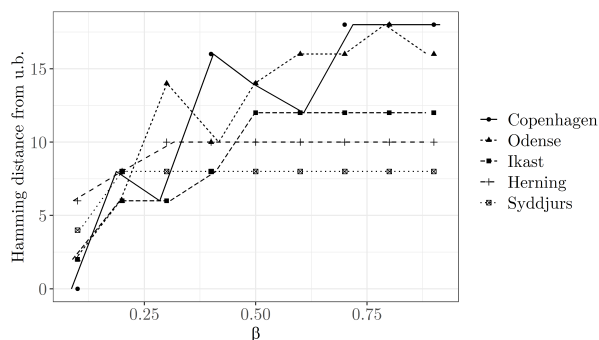
Figure 6.7: Hamming distance from the solution set of the upper bound case ($\beta = 0$) to the solution set depending on β, for $p = 15$. Solution sets for rural areas differ less from the upper bound solution when increasing the weight on distance.
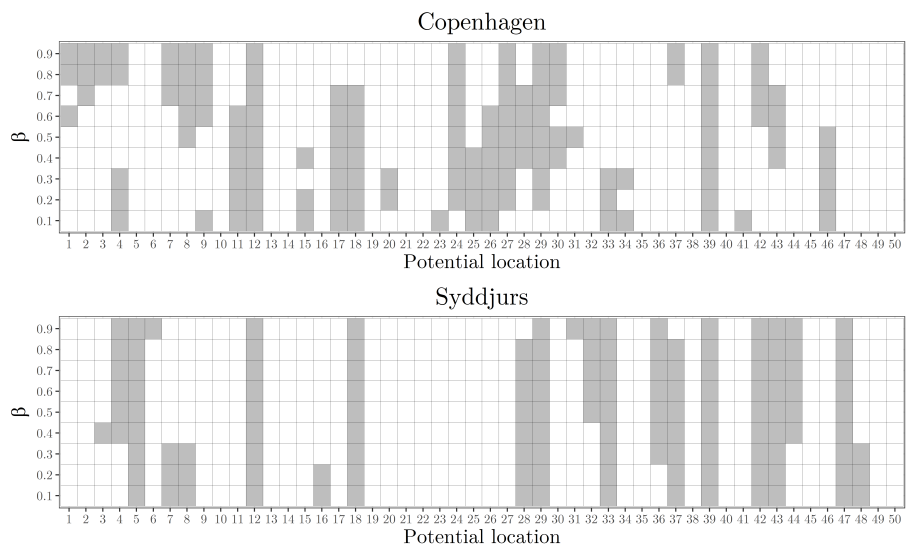

Figure 6.8: Representation of the selection of $p = 15$ drop-off stations from the total of $|M| = 50$ potential locations, for different values of β for Copenhagen (urban) and Syddjurs (rural).

samples, for each area. Table 6.6 shows the mean, standard deviation, and coefficient of variation (C.V.), computed as the percentage of the standard deviation on the average, for both covering and distance. The results show a low variability for all cases, but increased variability can be found for specific values of β. In Fig 6.9, we disaggregate the results to show the coefficient of variation of the absolute value of the objective value. The results show that the variability of the solution remains within 5% in the majority of the instances, which is in accordance with Table 6.6. However, an increased variability can be seen in urban areas in more balanced problems, a fact that can be attributed to the increase in the complexity of the solution rather than to the sampling selection.

	Covering			Distance		
Instance	Mean	St. dev.	C.V. (%)*	Mean	St. dev.	C.V.(%)*
Copenhagen	53.5	3.5	9.3	4983	353.4	9.5
Syddjurs	34.7	1.1	3.3	10734	663.9	7.5
Odense	42.9	1.1	2.8	6338	243.8	4.0
Ikast	36.3	0.9	2.6	6616	261.5	4.4
Herning	46.8	0.9	2.0	6579	300.4	4.8

Table 6.6: Summary statistics of the average results ($p = 10, 15, 20, 25$, $\beta = 0.1$ to 0.9), using 10 different random samples in the selecting potential locations step. The sampling phase has a low impact on the variability of both distance and covering results. *Average coefficient of variation (%).

To have a measure of the impact of the sample size, we use the aggregation error definition similar to Francis et al. (1999). We define the distance error as the euclidean distance between the household location and the location of the closest potential location. With this measure we compare the level of sparsity of our potential locations. In Fig. 6.10, we present the average distance error when varying (a) sample size, (b) number of potential locations, and (c) the covering radius, using $p = 15$ and $\beta = 0.5$. Fig. 6.10a shows the average distance error obtained by varying the size of the random sample $|M|$, from $|M| = 500$ to $|M| = 2000$ points. For this range, we observe that on average, the sample size has no significant impact on the average distance error. The different levels for each area correspond to the scale of distances between households, which are shorter for urban zones. These results confirm that the selection of a sample size of $|M| = 1000$ is an adequate value.

An important parameter in the decision of the location of drop-off stations is the size of the set of potential locations. We use the distance error to measure the impact of the number of potential locations on their spatial distribution. In Fig. 6.10b, we present the average distance error when varying the number of potential locations from $N = 10$ to $N = 100$, with steps of 10 locations. In general, the error decreases quickly when increasing the number of locations from 10 to 50, and then the error stabilizes.
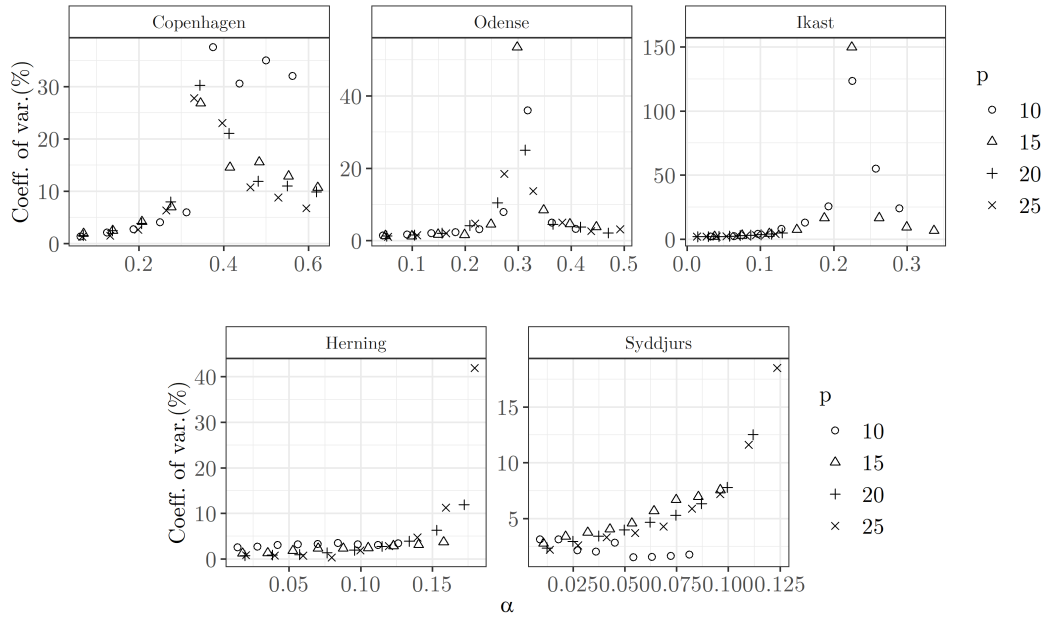
Figure 6.9: Coefficient of variation (%) of the objective value ($p = 10, 15, 20, 25$, $\beta = 0.1$ to 0.9, and its corresponding α), using 10 different random samples in the selecting potential locations step. Variability is increased in balanced scenarios of the objective function for urban areas.

Given these results, we conclude that a selection of $N = 50$ potential locations is an adequate value that balances computational complexity and satisfactory distribution of the locations on the map. Finally, in Fig. 6.10c, we show the effect of the covering radius parameter (S). We see that S has little effect on the distance error for urban areas, whereas for rural areas the error stabilizes when we consider values greater than $S = 200$, which is the selected value for the results in Section 6.

6.7 Conclusion

In this paper we presented a heuristic approach to solve the problem of deciding on the locations of recycling drop-off stations. The problem is motivated by the situation in which recycling rates must be increased by strategically locating drop-off stations while maintaining low installation and collection costs. We formulated the problem as an MCTP and presented a heuristic approach inspired by a VNS. Computational results were presented for a set of benchmark instances and compared to the optimal solutions. Results show that the proposed heuristic can effectively find good quality solutions for the MCTP. Finally, we used our heuristic to solve a set of real-life instances to provide insights into the trade-off between covering and collection costs for both urban and rural
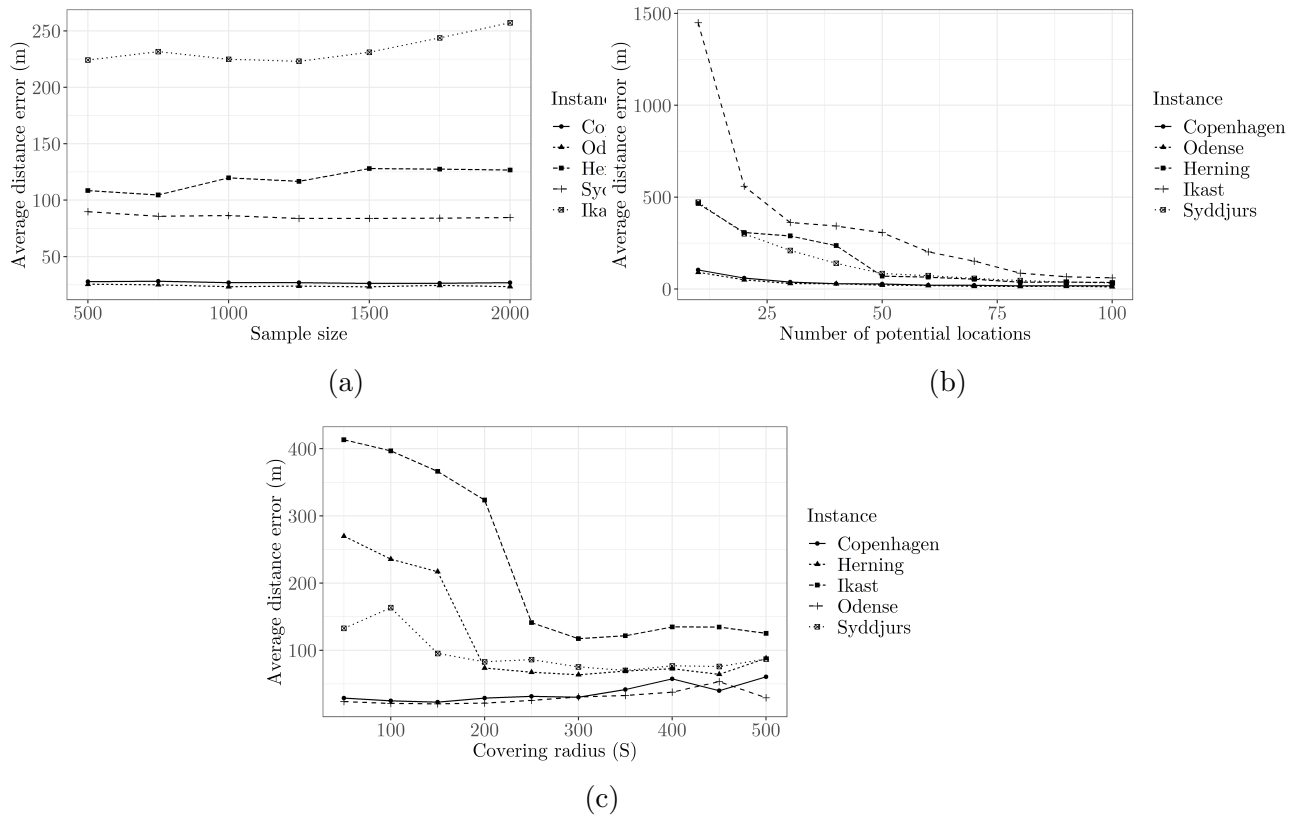
Figure 6.10: Sensitivity analysis on parameters for selecting potential locations. Average distance error for (a) Sample size, (b) Number of potential locations, and (c) Covering radius, using $p = 15$ locations and $\beta = 0.5$.

scenarios.

Acknowledgments