Logistics Research (2019) 12:5
DOI_10.23773/2019_5

Integrating Customer Choice in Differentiated Slotting for Last-Mile Logistics

Jochen Mackert[1], Claudius Steinhardt[2], Robert Klein[1]

Received: 28 February 2018 / Accepted: 08 April 2019 / Published online: 24 April 2019

ABSTRACT

In this paper, we study an e-grocer's problem of differentiated slotting in last-mile logistics for attended home delivery. The purpose of differentiated slotting is to statically determine which time slots should be on offer throughout the selling horizon in each area of a delivery region on each future service day such that cost-effective last-mile delivery schedules on an operational level can be expected. For this purpose, in addition to an appropriate approximation of potential delivery tours and costs, it is crucial to adequately consider the customer choice behavior between slots. In this paper, we propose a model-based, profit-oriented slotting approach that accounts for such choice behavior in a realistic, sophisticated fashion using a finite-mixture multinomial logit choice model. We formulate the resulting optimization problem as a non-linear mixed-integer program and show how it can be linearized. Further, we conduct a computational study to examine the benefit of the new approach. In particular, we demonstrate the superiority of our approach to existing approaches in the academic literature, which neglect or simplistically approximate customer choice behavior and which are purely cost-based.

Jochen Mackert[1]
jochen.mackert@wiwi.uni-augsburg.de

Claudius Steinhardt[2]
claudius.steinhardt@unibw.de

Robert Klein[1]
robert.klein@wiwi.uni-augsburg.de

[1] Chair of Analytics & Optimization,
University of Augsburg,
Universitätsstraße 16,
D-86159 Augsburg, Germany

[2] Chair of Business Analytics and Management Science,
Bundeswehr University Munich (UniBw),
Werner-Heisenberg-Weg 39,
D-85577 Neubiberg, Germany

KEYWORDS: last-mile logistics · attended home delivery · e-grocery · delivery time slot · customer choice behavior · mixed-integer programming

1 INTRODUCTION

Given the progressive digitization and the raising delivery service demand in urban areas as well as increasing customer service expectations, many researchers and practitioners emphasize the organization of urban logistics as a major challenge for the next years to come (BVL International 2017a, 2017b). In the area of urban logistics, we focus on deliveries to the front door of the customer (referred to as "last-mile") within a service time slot the service provider and customers have agreed on in advance. An example of the application of this service model, known as attended home delivery (AHD), is the e-grocery sector.

In an e-grocery, the entire order process takes place online via the e-grocer's website. Usually, at some point within the order process (before or after selecting the groceries), the customer is requested to login and to reveal some information to the e-grocer's system (e.g., her/his address). Based on the information regarding her/his location, the customer can choose from a set of specific delivery time slots (or decide to leave the website without ordering). At a certain time in advance of a specific delivery day, e.g., on its eve, the e-grocer stops offering service for this day and starts planning the operational delivery schedule based on the collected orders, known as the "last-mile" problem.

Besides an expensive delivery service and thin profit margins entailed in the e-grocery sector, new market entrants, such as Amazon Fresh, increase the competition for market share. Since a positive service experience is one of the main drivers for customers shopping for groceries online, an e-grocer's success highly depends on maintaining cost efficient operations while simultaneously ensuring the fulfilment of customers' expectations regarding service quality and reliability (BVL International 2017a). As customers' delivery time slot choices highly impact the e-grocer's operations (especially the delivery tours in last-mile logistics), the active management of the customers' demand, already in the selling horizon, is one of the key factors of success.

In this paper, we focus on differentiated static slotting, which is one of four demand management concepts prevalent in the academic literature on last-mile logistics in AHD as well as in practice. In static slotting, the e-grocer divides the delivery region into a number of delivery areas and strives, for each delivery day in the near future and each delivery area, to define a selection of time slots that will be on offer for all customers originating from that area. This selection is static and binding in the sense that it is fixed over the entire selling horizon regarding delivery on the specific day. The resulting problem is referred to as time slot management problem (TSMP) (Agatz et al. 2011), whereas the underlying concept stems from the literature on dynamic vehicle routing (Kunze 2004, 2005).

The main motivation of the research presented in this paper is the hypothesis that modeling demand in a realistic fashion is crucial for an efficient time slot management. In particular, we contribute to the existing literature as follows:

First, to the best of our knowledge, our approach is the first to incorporate detailed choice behavior into the TSMP and, thus, demand that arises depending on the e-grocer's offer decisions, meaning that different customer preferences regarding the time slots are realistically taken into account. This is in contrast to current static slotting approaches that assume demand to be independent of the e-grocer's offer decisions and, in most cases, to be spread evenly across the offered time slots. To reflect choice behavior in a realistic and sophisticated fashion, we incorporate a generalization of the multinomial logit model (MNL), known as the finite-mixture MNL in a model-based approach and combine it with a route approximation drawing on Agatz et al.'s (2011) mixed-integer program for the TSMP. With the finite-mixture MNL, multiple customer segments (e.g., students, family, and professionals) can be distinguished, and decision making can be further improved by exploiting segment specific information as, for instance, a segment's expected size in a delivery area, its value in terms of profit (before order fulfilment) and individual choice behavior.

Second, we aim at profit maximization in comparison to current approaches that aim at minimizing the cost of the expected delivery schedule. The latter technically requires the ex-ante specification of service frequencies that specify the number of time slots on offer in the different delivery areas, as otherwise the cost-based solution will always result in not to offer anything at all, or an additional artificial objective, as the minimization of unserved expected demand which is assumed to occur independently of the time slot offer decisions. Our profit maximizing approach is much more natural and enables the endogenous determination of the number of time slots on offer in the different delivery areas, without requiring the ex-ante specification of service frequencies.

The resulting model, which we refer to as TSMP with customer choice (TSMPCC), is non-linear. Therefore, we also show how it can be linearized in a lossless fashion to obtain a mixed-integer linear programming (MILP) formulation, by adapting techniques that have been proposed in the context of location planning. Further, we design and conduct a computational study to evaluate the benefit of our new approach at various capacity levels in comparison to existing approaches from the academic literature. In a series of computational experiments, we address the impact of adequately modeling time slot preferences, confirming our initial hypothesis, as well as the benefit of focusing on profit maximization instead of cost minimization. We also examine an approach based on the decomposition of the problem to find promising solutions for large problem sizes.

The remainder of the paper is organized as follows. In §2, an overview of the related literature and the different demand management concepts in AHD is provided. In §3, we introduce the non-linear TSMPCC and derive the corresponding MILP formulation. In §4, the computational study and results are presented. In §5, we discuss our results and provide managerial implications.

2 LITERATURE REVIEW

As depicted in Table 1 and mentioned earlier, the academic community in the e-fulfilment area considers four concepts for managing customer demand. While time slot allocation approaches ("slotting") seek to make a decision about which delivery time slots to offer, time slot pricing approaches ("pricing") deal with the determination of delivery fees for the different time slot alternatives. Dynamic approaches allow for making decisions for every individual customer who wants to place an order. In contrast, static approaches aim to make binding decisions prior to the first customer arrival (i.e., prior to the selling horizon), which are then valid for all customers.

By using pricing, the e-grocer is much more flexible in decision making than with slotting, which is restricted to offer decisions. Further, customers are not restricted in the choice of time slots, since always all available time slots are offered, which might be perceived as good service. However, at the same time, if the delivery of a customer in a specific time slot is costly, the delivery price might be inappropriately high, which might be negatively perceived by the customer. To circumvent this negative perception and possible inefficient delivery commitments, slotting offers the possibility to close a time slot for a customer request, if its delivery in this time slot is particularly costly. Regarding delivery price changes in the context of AHD, they are not as established as, for instance, in the airline industry and, thus, customers might perceive pricing as opaque and unfair (Xia et al. 2004), whereas slotting decisions are not perceived as discriminating as significantly differing prices (Agatz et al. 2013).

Concerning dynamic and static approaches, dynamic approaches allow to use the most current information prevalent at the time of decisions making. This can further improve the decisions' quality with respect to the objective (e.g. profit maximization). However, an individual decision for every customer request might again be perceived as opaque and unfair from a customer's perspective. Since decisions based on static approaches are valid and binding for the entire booking horizon, this can be circumvented. However, because decisions are not adjusted in the curse of the booking horizon, mistakes due to, for instance, errors in the forecast, cannot be corrected. For application purposes, the technical implementation of dynamic approaches is more challenging than that of static approaches, since decisions have to be made in real-time for every customer request.

All of the aforementioned concepts have in common that customer choice behavior and vehicle routing have to be incorporated in some way. Regarding the latter, the operational solution of a vehicle routing problem with time windows (VRPTW) must be anticipated.

Campbell and Savelsbergh (2005) and Ehmke and Campbell (2014) solve problems of *dynamic slotting.* Campbell and Savelsbergh (2005) assume a profit-maximizing e-grocer who decides to accept or reject a customer request and which time slots to offer in case of acceptance. They assume that customers have the same choice probability for each delivery time slots. Due to the NP-hardness of the VRPTW, vehicle routing decisions are made heuristically by an insertion heuristic drawing on Solomon (1987). Ehmke and Campbell (2014) base their work on the ideas of Campbell and Savelsbergh (2005). They propose customer acceptance mechanisms aiming at maximizing profits by accepting as many customer requests as possible. The impact of a customer request on the delivery tour is assessed drawing on Campbell and Savelsbergh (2005), but is extended by time-dependent and stochastic travel time information. The proposed mechanisms differ in the extent of travel-time information that is considered for the acceptance decision. Customer choice is considered by mainly assuming uniform demand for all offered delivery time slots. In their computational study, they briefly evaluate the impact of different popular delivery time slots among customers by simply exogenously predefining different choice probabilities for different delivery time slots.

Campbell and Savelsbergh (2006), Asdemir et al. (2009), and Yang et al. (2016) propose *dynamic pricing* solution approaches. Campbell and Savelsbergh (2006) build on Campbell and Savelsbergh (2005) and examine the potential of giving incentives to certain customers to influence their delivery time slot choice. The customers' choice probabilities for delivery times slots are assumed to be known, and can be influenced by giving discounts. Thereby, the probability of choosing a delivery time slot increases proportionally to the given discount. Asdemir et al. (2009) propose a dynamic programming (DP) formulation for the dynamic time slot pricing problem decomposed by delivery area. The delivery capacity for each area and time slot is assumed to be given in advance. Hence, they do not deal with the solution of a VRPTW, but the delivery cost is fixed. Asdemir et al. (2009) build on discrete choice modelling for customer behavior and use the MNL model. Yang et al. (2016) also use the MNL and approximate the delivery cost for every customer request with an insertion heuristic approach loosely based on Campbell and Savelsbergh (2006). However, for making pricing decisions, the underlying DP would have to be solved to determine a customer request's opportunity cost, which is computationally intractable for real-world instances. Therefore, Yang et al. (2016) propose a simple approach based on each customer request's insertion cost as an approximation for the request's opportunity cost. Yang and Strauss (2017) and Klein et al. (2018) propose more sophisticated approaches to approximate this opportunity cost and hence to improve pricing decisions.

To our best knowledge, the study by Klein et al. (2019) is the only one dealing with *differentiated pricing.* They propose a mixed-integer linear program to determine delivery prices for the different delivery slots. These prices are equal for all customers originating from the same delivery area. Routing approximations for real-world instances are based on a seed-based scheme drawing on Fisher and Jaikumar (1981). Customer choice behavior is modeled via a general non-parametric approach in which customers' preference lists are used to obtain the choice probabilities for the different delivery time slot alternatives.

Table 1: Classification of demand management concepts (Agatz et al. 2013, Yang et al. 2016)

	Time slot allocation	Time slot pricing
Dynamic (real-time)	Dynamic slotting	Dynamic pricing
Static (off-line)	**Differentiated slotting**	Differentiated pricing

Agatz et al. (2011) are among the first to consider *differentiated slotting*. For the solution of the TSMP, they propose a continuous approximation approach and an integer programming alternative. Their two approaches aim to find the time slots on offer in each delivery area based on exogenously given service frequencies, with the objective of minimizing the expected delivery cost. In their continuous approximation approach, the expected delivery cost is approximated according to the concepts of Daganzo (1987). The integer programming approach draws on a seed-based scheme from Fisher and Jaikumar (1981). Customer choice behavior is not explicitly considered, as they assume constant total demand in every area, independent of the time slots on offer and evenly spreading among the offered time slots. Cleophas and Ehmke (2014) tackle the problem of static slotting by determining booking limits for every delivery time slot dependent on the customers' order values. Therefore, they make use of historical booking data and adapt the expected marginal seat revenue heuristic for airline revenue management (Belobaba 1987). For this purpose, they draw on Ehmke and Campbell (2014) to determine the reserved transport capacity in an area and time slot. Concerning customer choice, the adapted EMSR based approach relies on the first choice principle, and a discrete demand probability distribution per customer segment for the different time slots is assumed. Hernandez et al. (2017) and Bruck et al. (2018) are closely related to Agatz et al. (2011), since they also propose a solution approach for the TSMP. In contrast to Agatz et al. (2011), both consider a planning horizon of one week. Analogous to the service levels necessary in the formulation of Agatz et al. (2011), Hernandez et al. (2017) use service frequencies for every delivery zone. They propose a MILP that aims at minimizing cost, where routing decisions are modelled based on graph theory by choosing the arc a vehicle must travel, which represents the distance between two delivery zones. Since the MILP is computationally intractable for problems of reasonable size, they suggest two heuristics for its solution. The demand assumptions are made in line with Agatz et al. (2011). Bruck et al. (2018) propose an approach to create time slot tables that minimize expected unserved demand by means of a large neighborhood search. Customer choice is considered based on four different scenarios. While the first and second are in line with assumption of Agatz et al. (2011), the third and fourth are based on historical data. The actual routing is optimized relying on integer linear programming.

In the line of that literature, in this paper, we propose a new differentiated slotting approach that relaxes the simplified and restrictive assumptions existing so far, especially regarding the demand model as well as the cost-orientation requiring given service frequencies of Agatz et al. (2011) and Hernandez et al. (2017). Our model formulation aims at maximizing the expected total profit, and we model customer choice by means of a finite-mixture MNL model that also enables the incorporation of multiple customer segments The expected routing decisions are made based on the integer programming formulation of Agatz et al. (2011).

Besides the literature on AHD, the approximation of total delivery cost dependent on the offered time slots as a crucial part of the TSMPCC also links our work to the vast body of literature on vehicle routing. We draw on Agatz et al. (2011)'s seed-based routing cost approximation. This idea was first presented by Fisher and Jaikumar (1981) in the context of a cluster-first, route-second heuristic for the VRP. Afterwards, this approximation method was adapted and extended by many researchers. For the classical VRP and its generalization the VRPTW, Koskosidis et al. (1992), Bramel and Simchi-Levi (1995), Russell (1995), Bramel and Simchi-Levi (1996) and Baker and Sheasby (1999) apply or extend Fisher and Jaikumar (1981)'s heuristic within their proposed solution approaches. In an extended version of the VRP with partially accessibility constraints, Semet (1995) base their two-phase algorithm on this approximation idea. Further modifications are proposed by Chao (2002) and Derigs et al. (2013) in the context of the Truck and Trailer Routing problem. For school bus routing, Pacheco and Martí (2006) apply the seed-based approximation to construct feasible solutions. One of the most current works is Holzapfel et al. (2016), which apply this approximation in an approach for a retailer's transportation planning. For an extensive review besides the seed-based approach, we refer to Bräysy and Gendreau (2005a, 2005b), Laporte (2009) and Baldacci et al. (2012).

Beyond, the TSMPCC particularly shares aspects with dynamic vehicle routing (DVRP), in which new information is revealed during the execution of an initial delivery plan that is adapted as soon as new information becomes apparent. An extensive review on this class of problems is given by Pillac et al. (2013) and Ulmer et al. (2018). With specific regard to our work, Kunze (2004, 2005) and Bent and van Hentenryck (2004) need to be explicitly mentioned. The former is relevant because they propose a dynamic vehicle routing approach in the context of AHD, drawing on the concept to make delivery time suggestions to customer requests based on their area of origin. These suggestions rely on the actual route plan and initial "Tour-Templates" generated based on expectations. The latter is mentioned because they are among the first to take stochastic information and expectations about future customer requests into account when dynamically deciding about a current customer request.

3 THE TIME SLOT MANAGEMENT PROBLEM WITH CUSTOMER CHOICE

In this section, we introduce the TSMPCC. The objective is to maximize the expected total profit by determining the time slots on offer for each delivery area. Since customer choices are influenced by the offered time slots, customer behavior as well as an approximation of the resulting delivery cost need to be incorporated. In §3.1 and §3.2, we explain the underlying ideas and introduce the notation for vehicle routing and customer choice, respectively. In §3.3, we formulate the problem as a non-linear mathematical optimization model and derive a linearization.

3.1 Delivery cost approximation in the TSMPCC

We consider an e-grocer who divides the delivery region into $i \in \mathcal{I} = \{1, \dots, I\}$ delivery areas $a_i \in \mathcal{A} = \{a_i: i \in \mathcal{I}\}$ and that offers $n \in \mathcal{N} = \{1, \dots, N\}$ non-overlapping time slots $s_n \in \mathcal{S} = \{s_n: n \in \mathcal{N}\}$ with length l^n in each delivery area for a fixed delivery day in the near future. Thereby, s_{n+1} is the direct successor time slot of s_n. The e-grocer's depot is located in a_0. Further, the e-grocer operates a homogenous delivery vehicle fleet with $m \in \mathcal{M} = \{1, \dots, M\}$ vehicles $v_m \in \mathcal{V} = \{v_m: m \in \mathcal{M}\}$. Each vehicle $v_m \in \mathcal{V}$ has the capacity Q to serve the customer demand.

The approximation of the expected delivery cost builds on a well-established seed-based scheme that draws on Fisher and Jaikumar (1981) and that has been introduced in the AHD literature by Agatz et al. (2011). This type of modelling allows high-quality solutions to be found, as shown by Klein et al. (2019). We conduct a time slot-based seed assignment. For this purpose, in each delivery area $a_i \in \mathcal{A}$, one seed $\dot{a}_i \in \dot{\mathcal{A}} = \{\dot{a}_i: i \in \mathcal{I}\}$ is predefined (e.g. the center of the area). Customers are served – figuratively speaking – from these seeds by the respective vehicles. The depot is denoted by $\dot{a}_0$. Now, in each time slot $s_n \in \mathcal{S}$, exactly one seed $\dot{a}_i \in \dot{\mathcal{A}}$ is assigned to each vehicle $v_m \in \mathcal{V}$ in use. For each vehicle, these assigned seeds can be thought of as forming a "rough" tour. If vehicle v_m is in use, the binary decision variable z^m equals 1, and 0 otherwise. The seed assignment is denoted by the decision variable σ_i^{mn}, which equals 1 if seed $\dot{a}_i$ is assigned to vehicle v_m in time slot s_n, and 0 otherwise. We approximate the travel distance of a vehicle by taking the following distances into account: First, d_{ij} denotes the *seed-to-seed distance* and is equal to the Euclidean distance between two seeds $\dot{a}_i$, $\dot{a}_j \in \dot{\mathcal{A}}$. If the seed assignment does not change in two consecutive time slots, i.e., $i = j$, we set d_{ii} to 0. Since the delivery tour starts and ends at the e-grocer's depot, d_{0i} (d_{i0}) denotes the distance from the depot to the assigned seed in the first time slot (from the assigned seed in the last time slot back to the depot) and is referred to as the *depot-to-seed distance*. Second, we incorporate the same distance d_{ij} (with $i \neq j$) for approximating the travel distance to each area $a_i \in \mathcal{A}$ in which customers are served in time slot $s_n \in \mathcal{S}$ from the assigned seed $\dot{a}_j$, referred to as the *seed-to-delivery-area distance*. The difference here is that the delivery area's seed is not assigned to the vehicle in the respective time slot, and, thus, it is not incorporated in the "rough" tour. Instead, the corresponding delivery area is thought of as being connected to the assigned seed in the sense of a "hub". Third, we assume a *to-customer distance* as the average expected travel distance $\tilde{d}_i$ to serve a customer within delivery area $a_i \in \mathcal{A}$. Corresponding to the distances, three further decision variables are introduced. Decision variables $D_i^{mn} \geq 0$ are defined to track a vehicle's *seed-to-delivery-area distance* and the decision variables $\widetilde{D}_i^{mn} \geq 0$ are defined to track the *to-customer distance* of vehicle v_m in time slot s_n. The vehicle v_m's *seed-to-seed distance* between its assigned seed in time slot s_{n-1} and in the consecutive time slot s_n with $n \in \mathcal{N} \backslash \{1\}$ are tracked by the decision variables $\widehat{D}_{ij}^{mn} \geq 0$. The sum of all the aforementioned distances approximates the distance of the operational vehicle routing tour. Multiplying any approximated travel distance with a cost factor c (cost per unit of travel distance) yields the corresponding approximated travel cost. Note, the approximations of *seed-to-seed-area distances* and *seed-to-delivery-area distances* are transferred from Agatz et al. (2011) who apply this way of route approximation on geographically aggregated customers. Since we consider arising demand in dependence of the offered time slots, we model the route approximation in dependence of this demand and, thus, additionally introduce the *to-customer distance*.

Figure 1 illustrates a fictional delivery region consisting of four areas a_i and the corresponding seeds $\dot{a}_i$ with $i \in \{1,2,3,4\}$. We assume an e-grocer who decides which time slots to offer in each area a_i from the set $\mathcal{S} = \{s_1, s_2\}$.

For clarity, in the illustration, we only consider a single vehicle v_1 (i.e., $M = 1$). The black (grey) triangles represent the expected demand occurring in time slot s_1 (s_2) as they result from the e-grocer's offer decisions which are considered as given here for the ease of explanation. Note that the exact location of expected customers is not known and is not of importance for our approximation approach but is only depicted for illustration purposes. All distances that need to be (virtually) traveled before (d_{01}) or in s_1 are depicted in black, and after (d_{30}) or in s_2 in grey. The *depot-to-seed* (d_{01} and d_{30}) and *seed-to-seed* (d_{13}) distances are indicated by the arrows between the depot and the seed $\dot{a}_1$ (assigned seed in s_1) and $\dot{a}_3$ (assigned seed in s_2). The arrows form a directed cycle that can be thought of as defining a "rough" routing of the vehicle. In the example, the *seed-to-seed distance* d_{13} is assumed to be equally divided between the time slots (which is later endogenously determined by the model). Within time slot s_1, for instance, expected demand in areas a_1 and a_2 must be served. On the one hand, the solid line from the assigned seed $\dot{a}_1$ to the seed $\dot{a}_2$ depict the approximated distances that must be travelled to reach the delivery-area a_2 in time slot s_1 (*seed-to-delivery-area distance* d_{12}), which can be thought of as forming a "hub connection" with regard to the rough routing. On the other hand, the radius of the broken circles inside of the areas illustrate the

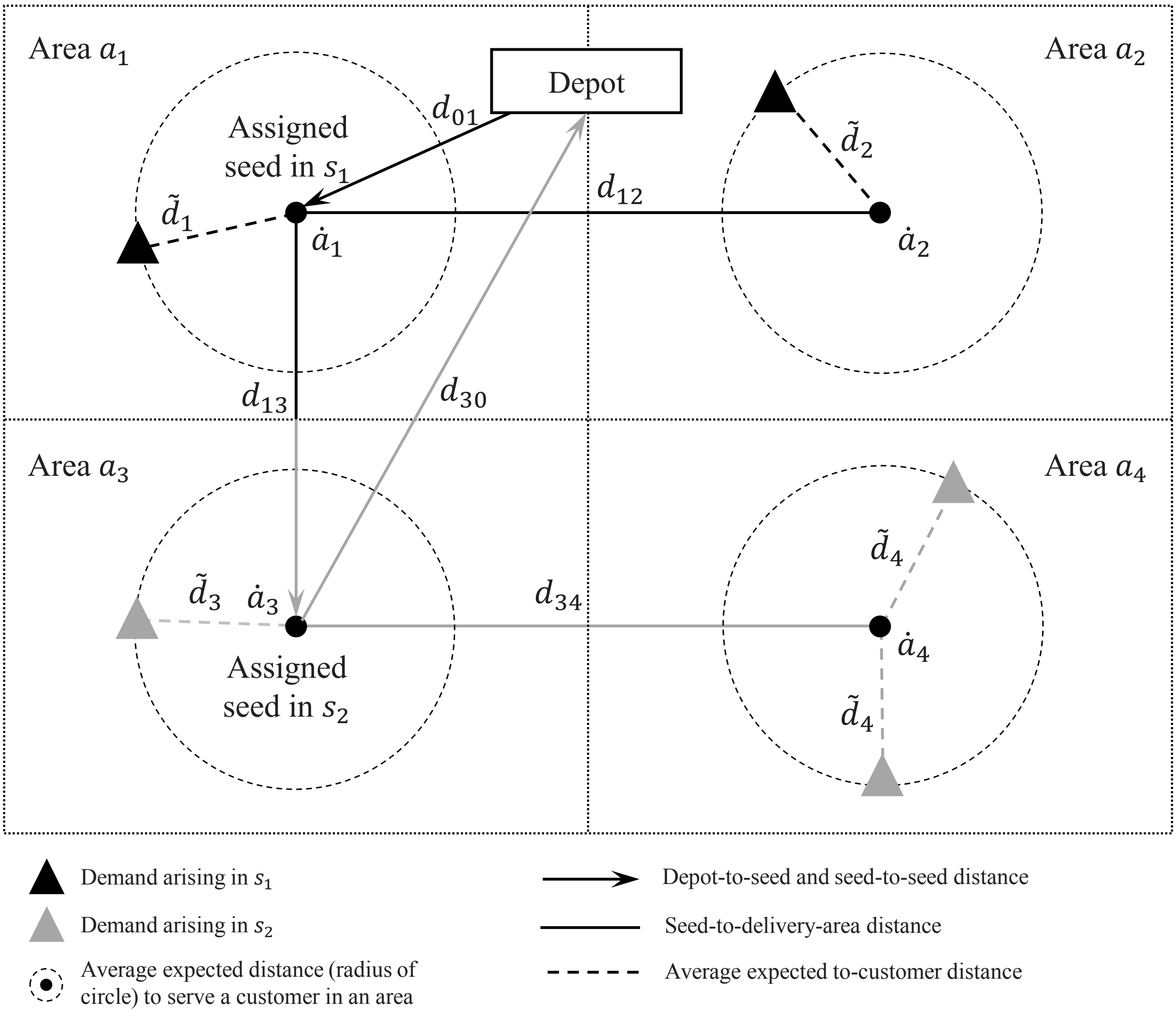

Figure 1: Approximation of the delivery cost by means of a seed-based scheme

average distance that must be (virtually) travelled to serve a customer in a specific area (*to-customer distance*), defining the "spokes" of the hub. Thus, the approximated delivery route distance in time slot s_1 is equal to $\tilde{d}_1 + \tilde{d}_2 + d_{12} + \alpha_{12} \cdot d_{13}$, with $\alpha_{12} \in [0,1]$ and $(1 - \alpha_{12})$ being the portion of d_{13} which needs to be (virtually) travelled in time slot s_1 and time slot s_2 (in the example, $\alpha_{12} = 0.5$) and which is endogenously determined by the model. Analogously, for time slot s_2, the approximated delivery route distance is equal to $(1 - \alpha_{12}) \cdot d_{13} + \tilde{d}_3 + d_{34} + 2 \cdot \tilde{d}_4$. The approximation of the overall delivery tour distance is equal to the total length of all the solid and broken straight lines (without the circles) and arrows in Figure 1. Note, following the common idea of seed-based approximations (e.g. Agatz et al. 2011), we do not additionally incorporate any "return" distances to the seed, e.g., in the illustration, $\tilde{d}_1$, $\tilde{d}_2$ and d_{12} are only taken into account once in time slot s_1. This is because on an operational level, in a specific time slot, a vehicle would first serve all customers from one area and would then directly drive to the next area serving all customers living there. If, however, on an operational level, a vehicle needs to travel back and forth between two delivery areas (e.g. a_1 and a_2), owing to the fact that in both of two consecutive time slots (e.g. s_1 and s_2) there are customers originating from these areas who require service, our approximation approach takes the distance d_{12} into account twice; once for the travel distance between the areas in time slot s_1 and once for the travel distance between the areas in time slot s_2.

3.2 Incorporation of customer choice in the TSMPCC

We model the customer choice behavior in line with discrete choice theory by means of a finite-mixture MNL model (cf. e.g. McFadden and Train 2000, Train 2009). More precisely, we consider a market that consists of different customer segments $k_o \in \mathcal{K} = \{k_o : o \in \mathcal{O} = \{1, \dots O\}\}$. All customers belonging to a segment $k_o \in \mathcal{K}$ homogenously perceive a specific utility U_o^n for every time slot $s_n \in \mathcal{S}$ and are assumed to rationally choose the time slot with the highest utility. Rooted in random utility theory, this utility is defined as $U_o^n := u_o^n + \epsilon_o^n$, with u_o^n being the deterministic part and ϵ_o^n being the random part. The deterministic parts are assumed to be segment-specific linear functions $u_o^n = \boldsymbol{\beta}_o^T \hat{\mathbf{a}}^n$ of observable time slot attributes $\hat{\mathbf{a}}^n$ weighted by utility parameters $\boldsymbol{\beta}_o$. The random parts are assumed to be independent and identically distributed random variables following a Gumbel distribution with zero mean, which implies that each customer segment is assumed to follow a specific MNL. Then, according to the MNL properties, the choice probability of an offered time slot of a customer belonging to segment k_o is equal to its attraction $\hat{A}_o^n = e^{u_o^n}$ relative to the overall attraction of the offered alternatives (cf. e.g. Train 2009 for details). To formalize this probability, we define the vector $\boldsymbol{\gamma}_i = [\gamma_i^n]_{a_i \in \mathcal{A}, s_n \in \mathcal{S}}$, containing the decision variable γ_i^n, which equals 1 if time slot $s_n \in \mathcal{S}$ is offered in area a_i, and 0 otherwise. Given the selection of offered time slots $\boldsymbol{\gamma}_i$ in a delivery area a_i, the choice probability $P_{io}^n(\boldsymbol{\gamma}_i)$ of a time slot s_n of a customer belonging to segment k_o and originating from area a_i is then given by

$$P_{io}^n(\boldsymbol{\gamma}_i) = \frac{\hat{A}_o^n \gamma_i^n}{C_o + \sum_{p \in \mathcal{N}} \hat{A}_o^p \gamma_i^p}, \quad (1)$$

with $C_o = e^{u_o^0}$ representing the segment-specific attraction value of all outside alternatives aggregated over the e-grocer's competitors. Additionally, it includes the customers' attraction toward the alternative of not choosing a time slot at all (which is always available). Hence, the no-purchase probability of a customer belonging to segment k_o facing the selection of time slots $\boldsymbol{\gamma}_i$ is given by

$$P_{io}^0(\boldsymbol{\gamma}_i) = \frac{C_o}{C_o + \sum_{n \in \mathcal{N}} \hat{A}_o^n \gamma_i^n}. \quad (2)$$

Following from (1) and (2) and in line with discrete choice theory, the choice probabilities and the no-purchase probability of a customer belonging to segment k_o and originating from area a_i sum to 1, independent of the offered selection $\boldsymbol{\gamma}_i$:

$$\sum_{n \in \mathcal{N}} P_{io}^n(\boldsymbol{\gamma}_i) + P_{io}^0(\boldsymbol{\gamma}_i) = 1. \quad (3)$$

Given the expected number of customers from each segment k_o originating from area a_i by parameter ω_{io}, we obtain the overall finite mixture MNL by aggregation over the segment-specific probabilities. More precisely, the aggregated expected number of customers in area a_i over all segments k_o ordering and choosing a time slot s_n when offering the selection $\boldsymbol{\gamma}_i$ is given by

$$\sum_{o \in \mathcal{O}} \omega_{io} P_{io}^n(\boldsymbol{\gamma}_i) = \sum_{o \in \mathcal{O}} \omega_{io} \frac{\hat{A}_o^n \gamma_i^n}{C_o + \sum_{p \in \mathcal{N}} \hat{A}_o^p \gamma_i^p}. \quad (4)$$

Table 2: Notation summary

Indices			Input parameters	
$\mathcal{I}$	$= \{1, \dots, I\}$	Indices of areas and seeds	$\hat{A}_o^n$	Attraction of slot $s_n \in \mathcal{S}$ of segment $k_o \in \mathcal{K}$
$\mathcal{M}$	$= \{1, \dots, M\}$	Indices of vehicles	C_o	Attraction of outside alternative of segment $k_o \in \mathcal{K}$
$\mathcal{N}$	$= \{1, \dots, N\}$	Indices of time slots	c	Cost per unit of travel distance
$\mathcal{O}$	$= \{1, \dots, O\}$	Indices of segments	d_{ij}	Distance between seeds/the depot $\dot{a}_i, \dot{a}_j \in \dot{\mathcal{A}} \cup \{\dot{a}_0\}$
Sets			$\tilde{d}_i$	Average distance to serve a customer in area $a_i \in \mathcal{A}$
$\mathcal{A}$	$= \{a_i : i \in \mathcal{I}\}$	Set of areas	f	Fixed cost per vehicle in use
$\dot{\mathcal{A}}$	$= \{\dot{a}_i : i \in \mathcal{I}\}$	Set of seeds	l^n	Length of slot $s_n \in \mathcal{S}$
$\mathcal{K}$	$= \{k_o : o \in \mathcal{O}\}$	Set of customer segments	M_i^n	Large number for area $a_i \in \mathcal{A}$ and slot $s_n \in \mathcal{S}$
$\mathcal{S}$	$= \{s_n : n \in \mathcal{N}\}$	Set of time slots	ω_{io}	Size of segment $k_o \in \mathcal{K}$ belonging to area $a_i \in \mathcal{A}$
$\mathcal{V}$	$= \{v_m : m \in \mathcal{M}\}$	Set of delivery vehicles	π^n	Time needed for one unit of distance between areas in slot $s_n \in \mathcal{S}$
			$\hat{\pi}^n$	Time needed for one unit of distance between areas between slots $s_{n-1}, s_n \in \mathcal{S}$
			$\bar{\pi}_i^n$	Travel time to serve a customer in area $a_i \in \mathcal{A}$ and slot $s_n \in \mathcal{S}$
			Q	Vehicle capacity
			$\bar{r}_o$	Average profit generated by segment $k_o \in \mathcal{K}$
			τ	Service time needed to serve a customer

Decision variables

$D_i^{mn} \geq 0$	Travel distance of vehicle $v_m \in \mathcal{V}$ in slot $s_n \in \mathcal{S}$ that emanates from seed $\dot{a}_i \in \dot{\mathcal{A}}$
$\widehat{D}_{ij}^{mn} \geq 0$	Travel distance of vehicle $v_m \in \mathcal{V}$ between the assigned seed $\dot{a}_i \in \dot{\mathcal{A}}$ in slot s_{n-1} and the assigned seed $\dot{a}_j \in \dot{\mathcal{A}}$ in slot s_n with $n \in \mathcal{N} \setminus \{1\}$
$\widetilde{D}_i^{mn} \geq 0$	Travel distance of vehicle $v_m \in \mathcal{V}$ in slot $s_n \in \mathcal{S}$ to serve demand in area $a_i \in \mathcal{A}$
$e_i^{mn} \geq 0$	Demand occurring in area $a_i \in \mathcal{A}$ in slot $s_n \in \mathcal{S}$ served by vehicle $v_m \in \mathcal{V}$
$\gamma_i^n \in \{0,1\}$	$= 1$, if slot $s_n \in \mathcal{S}$ is offered in area $a_i \in \mathcal{A}$, 0 otherwise
$\sigma_i^{mn} \in \{0,1\}$	$= 1$, if seed $\dot{a}_i \in \dot{\mathcal{A}}$ is assigned to vehicle $v_m \in \mathcal{V}$ in slot $s_n \in \mathcal{S}$, 0 otherwise
$x_{io}^n \geq 0$	Number of customers of segment $k_o \in \mathcal{K}$ from area $a_i \in \mathcal{A}$ choosing slot $s_n \in \mathcal{S}$
$z^m \in \{0,1\}$	$= 1$, if vehicle $v_m \in \mathcal{V}$ is used, 0 otherwise
$\tilde{z}_i^{mn} \in \{0,1\}$	$= 1$, if vehicle $v_m \in \mathcal{V}$ visits area $a_i \in \mathcal{A} \cup \{a_0\}$ or the depot in slot $s_n \in \mathcal{S}$, 0 otherwise

3.3 Optimization model

In addition to the notation explained in the preceding subsections, we first need to introduce some further decision variables and parameters, as follows: The decision variable $\tilde{z}_i^{mn}$ equals 1 if vehicle v_m visits area a_i in time slot s_n, and 0 otherwise. The decision variable e_i^{mn} gives the demand that has to be delivered by vehicle v_m in area a_i and time slot s_n. In line with Agatz et al. (2011), Klein et al. (2019) and since the time slot length rather than physical vehicle capacity is usually the constraining factor in AHD (Campbell and Savelsbergh 2005, Ehmke and Campbell 2014), we assume unit demand for all customers. To enhance readability, we introduce the auxiliary decision variable x_{io}^n to track the disaggregated results of (4) for all areas a_i, segments k_o, and time slots s_n, that is, the number of expected customers from segment k_o originating from area a_i which choose time slot s_n.

Regarding the model's parameters, π^n ($\hat{\pi}^n$) represents the time needed to travel one unit of distance between areas in time slot $s_n \in \mathcal{S}$ (between time slot s_{n-1} and s_n). The average travel time to serve a customer within a delivery area $a_i \in \mathcal{A}$ and slot $s_n \in \mathcal{S}$ is denoted by $\bar{\pi}_i^n$. The parameter τ represents the service time. The use of a vehicle causes the fixed cost of f. Every customer belonging to segment k_o generates an average profit of $\bar{r}_o$. A comprehensive notation summary is presented in Table 2.

3.3.1 *Mathematical formulation for the TSMPCC*

Based on the introduced notation, the TSMPCC can be formulated by the following fractional mixed-integer program (MIP):

$$\text{Maximize} \quad \sum_{i\in\mathcal{I}}\sum_{o\in\mathcal{O}}\sum_{n\in\mathcal{N}} x_{io}^n \bar{r}_o - \sum_{m\in\mathcal{M}} f z^m - \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\sum_{n\in\mathcal{N}:n>1}\sum_{m\in\mathcal{M}} c\widehat{D}_{ij}^{mn} - \sum_{i\in\mathcal{I}}\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} cD_i^{mn} - \sum_{i\in\mathcal{I}}\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}} c\widetilde{D}_i^{mn} - \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} c\left(\sigma_i^{m1} d_{0i} + \sigma_i^{mN} d_{i0}\right) \tag{5}$$

subject to the following:

$$\tilde{z}_0^{m1} = z^m \qquad \forall m \in \mathcal{M} \tag{6}$$

$$\tilde{z}_0^{mN} = z^m \qquad \forall m \in \mathcal{M} \tag{7}$$

$$\sum_{i\in\mathcal{I}} \sigma_i^{mn} = z^m \qquad \forall m \in \mathcal{M}, n \in \mathcal{N} \tag{8}$$

$$D_i^{mn} \geq d_{ij}\left(\tilde{z}_i^{mn} + \sigma_j^{mn} - 1\right) \qquad \forall i,j \in \mathcal{I}, m \in \mathcal{M}, n \in \mathcal{N} \tag{9}$$

$$\widetilde{D}_i^{mn} \geq \tilde{d}_i e_i^{mn} \qquad \forall i \in \mathcal{I}, m \in \mathcal{M}, n \in \mathcal{N} \tag{10}$$

$$\hat{D}_{ij}^{mn} \geq d_{ij}\left(\sigma_i^{mn-1} + \sigma_j^{mn} - 1\right) \qquad \forall i,j \in \mathcal{I}, m \in \mathcal{M}, n \in \mathcal{N}: n > 1 \tag{11}$$

$$x_{io}^n = \omega_{io} \frac{\hat{A}_o^n \gamma_i^n}{C_o + \sum_{p\in\mathcal{N}} \hat{A}_o^p \gamma_i^p} \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \tag{12}$$

$$\sum_{m\in\mathcal{M}} e_i^{mn} \geq \sum_{o\in\mathcal{O}} x_{io}^n \qquad \forall i \in \mathcal{I}, n \in \mathcal{N} \tag{13}$$

$$e_i^{mn} \leq \tilde{z}_i^{mn} M_i^n \qquad \forall i \in \mathcal{I}, m \in \mathcal{M}, n \in \mathcal{N} \tag{14}$$

$$Qz^m \geq \sum_{i\in\mathcal{I}} \sum_{n\in\mathcal{N}} e_i^{mn} \qquad \forall m \in \mathcal{M} \tag{15}$$

$$\sum_{i\in\mathcal{I}} \sum_{n=p}^{q} \pi^n D_i^{mn} + \sum_{i\in\mathcal{I}} \sum_{n=p}^{q} (\tau + \bar{\pi}_i^n) e_i^{mn} + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} \sum_{n=p+1}^{q} \hat{\pi}^n \hat{D}_{ij}^{mn} \leq \sum_{n=p}^{q} l^n \qquad \forall m \in \mathcal{M}, p, q \in \mathcal{N}: q \geq p \tag{16}$$

The objective function of the TSMPCC (5) aims at maximizing the expected total profit by summing up the expected number of customers from each segment multiplied by their average profit and subtracting the total expected cost, consisting of the fixed cost for the vehicles in use and the travel cost. Remember, the total delivery cost are approximated as explained in §3.1, and thus, no operational delivery tour is generated. The approximated travel cost for a vehicle v_m arises from the distances between the seeds in consecutive time slots, the distances to the areas and to the expected customers in the offered time slots, and the distances from (to) the depot in the first (last) time slot. Constraints (6) – (11) model the routing requirements. Constraints (6) and (7) ensure that every vehicle in use ($z^m = 1$) starts (cf. (6)) and ends (cf. (7)) its tour at the e-grocer's depot. Constraints (8) assign exactly one seed per time slot s_n to every vehicle in use ($z^m = 1$). Constraints (9) – (11) define the total expected travel distance, except for the distance from and to the depot. Constraints (9) – (10) model the travel distance of a vehicle v_m within a time slot s_n and constraints (11) model the travel distance between consecutive time slots. If a vehicle v_m visits area a_i in time slot s_n ($\tilde{z}_i^{mn} = 1$) and $\dot{a}_j$ is assigned as the seed for vehicle v_m in time slot s_n ($\sigma_j^{mn} = 1$), then constraints (9) force variable D_i^{mn} to be equal to the distance between seed $\dot{a}_i$ and seed $\dot{a}_j$. Constraints (10) models the distance necessary to serve the expected customers in area a_i and time slot s_n. Constraints (11) work analogously to constraints (9), but is not defined for the first time slot since the distance between the depot and the seed in the first time slot is separately considered in the objective function (5). Constraints (12) – (14) model the customer demand and its fulfilment. The fractional constraints (12) assign the disaggregated results of (4) to the auxiliary variable x_{io}^n that is used within the objective function to enhance readability. Constraints (13) ensure that all expected demand is served by the vehicles in use. To prevent vehicles in use from serving the demand of area a_i in time slot s_n even though they are not visiting area a_i in that time slot, constraints (14) are defined. To obtain a tight bound, M_i^n for area a_i and time slot s_n has to be equal to $\min\left\{\sum_{o\in\mathcal{O}} \omega_{io} \frac{\hat{A}_o^n}{C_o + \hat{A}_o^n}; \frac{l^n}{\bar{\pi}_i^n + \tau}\right\}$.

Note that this represents the minimum of either the maximum demand for area a_i in time slot s_n, which can possibly occur due to the properties of the assumed customer choice model (cf. §3.2), or the maximum demand, which can possibly be served within time slot s_n. Constraints (15) – (16) model the vehicle capacity and time constraints. On the one hand, constraints (15) prevent every vehicle's load from exceeding its capacity Q, and on the other hand, that its capacity is only available if it is in use ($z^m = 1$). Constraints (16) ensure that customer service is provided within the time slots. In other words, the travel time and service time of vehicle v_m in time slot s_n are not allowed to exceed the time slot's length l^n. For each single slot $s_n \in \mathcal{S}$, i.e. $q = p$, the first and second term on the left-hand side represent the time necessary for customer service and respective travel time in slot s_n which is not allowed to exceed the time slot's length l^n. For consecutive time slots $q = p + 1$, the third term additionally accounts for the travel times between the seeds and it is ensured, that the overall service and travel times for both time slots do not exceed their aggregated length. By defining constraints (16) analogously for every combination of time slots with $q \geq p$, it is ensured that no single time slot's length is exceeded. Note, since we draw on the route approximation of Agatz et al. (2011), constraints (6) – (9) and (11) are directly transferred from their formulation.

3.3.2 Transformation into a mixed-integer linear program

In this section, we show how the model (5) – (16) can be transformed into a MILP by replacing the fractional demand constraints (12) by linear constraints. For this purpose, we adapt the linearization techniques proposed by Aros-Vera et al. (2013) and Haase and Müller (2014) from the location planning stream of research. The advantage of this reformulation is that any standard MILP software package can be used to solve the TSMPCC.

For the linearization, we define the additional decision variables p_{io}^{n} and p_{io}^{0}. They represent the choice probabilities for slot s_n and the no-purchase probability for customer segment k_o from area a_i (cf. §3.2). Constraints (12) are replaced by the following constraints (17):

$$x_{io}^{n} = \omega_{io} p_{io}^{n} \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \qquad (17)$$

To assign the correct values to the decision variable p_{io}^{n}, which is dependent on the decisions for the offered time slots $\boldsymbol{\gamma}_i$ in area a_i, we define constraints (18) – (22), as follows:

$$p_{io}^{0} + \sum\nolimits_{n \in \mathcal{N}} p_{io}^{n} = 1 \qquad \forall i \in \mathcal{I}, o \in \mathcal{O} \qquad (18)$$

$$p_{io}^{n} \le \frac{\hat{A}_o^n}{C_o + \hat{A}_o^n} \gamma_i^n \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \qquad (19)$$

$$\frac{\hat{A}_o^n}{C_o + \sum_{n \in \mathcal{N}} \hat{A}_o^n} \gamma_i^n \le p_{io}^{n} \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \qquad (20)$$

$$p_{io}^{n} \le \frac{\hat{A}_o^n}{C_o} p_{io}^{0} \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \qquad (21)$$

$$p_{io}^{0} \le \frac{C_o}{\hat{A}_o^n} p_{io}^{n} + (1 - \gamma_i^n) \qquad \forall i \in \mathcal{I}, o \in \mathcal{O}, n \in \mathcal{N} \qquad (22)$$

In line with the properties of the underlying choice model (cf. §3.2), all the choice probabilities of customer segment k_o belonging to area a_i must sum up to 1 (constraints (18)). Constraints (21) ensure that the choice probability of time slot s_n of segment k_o belonging to area a_i is only greater than 0 if s_n is offered in area a_i. The fractions on the right-hand side of constraints (19) are tighter upper bounds for p_{io}^{n} than those obtainable from only taking $\gamma_i^n = 1$ in case time slot s_n is offered in area a_i. Constraints (20) are not necessary for the correctness of the model, but provide the tightest lower bounds for the choice probabilities p_{io}^{n} in case time slot s_n is offered in area a_i. Constraints (21) – (22) ensure the correct ratio between the choice probabilities dependent on the offered time slots $\boldsymbol{\gamma}_i$ resulting from the choice model as given in (1) and (2). The complete MILP formulation for the TSMPCC is then given by (5) – (11) and (13) – (22).

By adding the symmetry-breaking constraints (23) – (24), the efficiency of a standard solver can be increased. Similar to constraints (20), they are not mandatory for the model correctness.

$$\sum\nolimits_{i \in \mathcal{I}} \sum\nolimits_{n \in \mathcal{N}} \tilde{z}_i^{mn} \le \sum\nolimits_{i \in \mathcal{I}} \sum\nolimits_{n \in \mathcal{N}} \tilde{z}_i^{(m-1)n} \qquad \forall m \in \mathcal{M}: m > 1 \qquad (23)$$

$$z^{m} \le z^{m-1} \qquad \forall m \in \mathcal{M}: m > 1 \qquad (24)$$

4 COMPUTATIONAL STUDY

We will now evaluate the proposed MILP for the TSMPCC. In particular, we examine it in comparison to neglecting or simplistically approximating customer choice behavior, as assumed in current approaches of the TSMP in the academic literature (cf. e.g. Agatz et al. 2011). In §4.1, we describe the data used for the computational experiments. In §4.2, we evaluate the incorporation of appropriate choice behavior. For this purpose, we present a benchmark approach, which is based on approaches in which the choice behavior can only be approximated, and propose an evaluation framework and statistics for comparison (§4.2.1). In §4.2.2, we present the results of exploiting an adequate modeling of time slot preferences with a customer choice model as realized by our new approach. In §4.2.3, we present the results of profit orientation of the TSMPCC that is possible with our new approach, in comparison to cost orientation of existing approaches. Besides, we also investigate a decomposition approach to find promising solutions for the TSMPCC for large problem sizes in §4.3.

4.1 Description of data

For our computational experiments, we define two problem classes. In the first problem class, we exactly evaluate the proposed MILP for the TSMPCC. For this purpose, we assume a delivery region consisting of 5 delivery areas served by one depot. In the second problem class, we evaluate how promising solutions can be found for larger problem sizes. For this purpose, the delivery region is assumed to consist of 24 areas, and a heuristic decomposition approach will be used (cf. §4.3). For both problem classes and in line with Klein et al. (2019), we use publicly available statistical data based on zip code areas in the Munich region for the distances d_{ij} between areas to realistically model the routing cost. These distance vary between 1 and 6 kilometers, and we assume the distance $\tilde{d}_i$ that needs to be traveled to serve a customer within an area equal to $\tilde{d}_i = 0.4$ kilometers $\forall i \in \mathcal{I}$. For the delivery day under consideration, we assume 6 equally long and non-overlapping time slots of 2 hours (Campbell and Savelsbergh 2005, 2006, Klein et al. 2019). The generation of customer data for evaluation purposes is described in Appendix A.

In line with the publicly available statistical data and Klein et al. (2019), we assume an average speed of 30 km/h for travel distances between areas (i.e., $\pi^n = \hat{\pi}^n = 2\ \forall n \in \mathcal{N}$). The travel time to serve a customer within an area is also assumed to be $\bar{\pi}_i^n = 2\ \forall n \in \mathcal{N}, i \in \mathcal{I}$. The size Q of a delivery van is assumed to be non-binding (cf. Ehmke and Campbell 2014), and the fixed cost for its usage (i.e. $z^m = 1$) is $f = 150$. The variable cost factor per unit of travel distance is equal to $c = 0.25$ (Yang and Strauss 2017). The fleet size is chosen to be restrictive, which means that there is not enough available (timely) capacity to offer all slots in all areas. In the first (second) problem class, the number of vehicles which can be used at most is assumed to be restricted to 2 (12). For the first problem class, we artificially change the available capacity by varying the service time τ between 8 (least scarce capacity level), 10, 12, and 14 (scarcest capacity level).

We perform all computations on an Intel(R) Xeon(R) processor with 16 cores, 3.4 GHz and 192 GB RAM. We implemented the models in Python 3.5 and used Gurobi 7.0.2 as the solver.

4.2 Evaluation of the new approach

To evaluate our approach, we investigate the profit potential of incorporating customer choice in the TSMP by benchmarking it to only simplistically and approximately taking choice behavior into account which is in line with current approaches of the academic literature (e.g. Agatz et al. 2011). For this purpose, we assume customer demand that arises in dependence of the time slots on offer $\boldsymbol{\gamma}_i$ according to the MMNL, as the true choice behavior of customers (referred to as TRUE). In comparison, current approaches in the academic literature approximate this behavior by assuming equally popular time slots and, thus, evenly distributed demand across the offered time slots, independent of the e-grocer's offer decisions (referred to as APPROX). Moreover, our new approach pursues profit maximization instead of cost minimization. This profit orientation does not require the predefinition of service frequencies (i.e., the number of time slots which need to be offered in a delivery area), which is mandatory in cost oriented approaches as mostly proposed in the current academic literature.

4.2.1 Benchmark, evaluation framework, and statistics

For the evaluation, we define two different configurations of input parameters for the TSMPCC: For the first configuration TRUE, reflecting the accurate modeling of choice behavior and resulting demand, customer attraction values are assumed to be different for the different time slot alternatives. The TSMPCC allows for different attraction values, which means that there can be more and less popular time slots. Therefore, demand can be different in and dependent on the different time slots on offer $\boldsymbol{\gamma}_i$. For the benchmark configuration APPROX, reflecting the approximation of choice behavior and resulting demand, we assume equal attraction values, meaning that customers are indifferent between the different time slots, and that are consistently derived from the data used for the configuration TRUE (cf. Appendix A.2). Hence, in APPROX, for a given number of offered time slots, this leads to equal demand in every time slot on offer, independent of the offered time slots $\boldsymbol{\gamma}_i$. This is in line with the assumptions currently made in the academic literature for static slotting as, for instance, by Agatz et al. (2011) or Hernandez et al. (2017). Thus, we implement APPROX in line with this literature to reflect the "status quo" of static slotting approaches. The attraction values used for TRUE and APPROX can be found in Appendix A.

To compare ourselves towards the benchmark approach that approximates choice behavior according to the aforementioned explanations, the underlying question for the evaluation is as follows: If an e-grocer derives time slot offer decisions relying on APPROX (e.g. by applying the approach of Agatz et al. 2011), but in reality, customer demand arises according to TRUE, what is the e-grocer's loss in profit if time slots are offered according to the derived decisions by means of APPROX?

Technically, we implement this idea by first solving the TSMPCC with the configurations TRUE and APPROX. This yields the optimal time slot offer decisions relying on the assumed choice behavior. To mimic the time slot offers $\boldsymbol{\gamma}_i^{\text{APPROX}}$ $\forall i \in \mathcal{I}$ resulting from APPROX to customers that, in reality, behave according to TRUE, we plug the decisions $\boldsymbol{\gamma}_i^{\text{APPROX}}$ $\forall i \in \mathcal{I}$ in the TSMPCC formulation by means of constraints (25) and evaluate it with configuration TRUE.

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i^{\text{APPROX}} \qquad \forall i \in \mathcal{I} \qquad (25)$$

I.e., we aim at finding the true objective function value when fixing time slots on offer in every delivery area as obtained from APPROX. The resulting expected maximum profit is the expected profit that can be obtained when managing customer demand based on the decisions made by approximately assuming customer choice behavior.

Note that due to the different configurations, the solution $\boldsymbol{\gamma}_i^{\text{APPROX}}$ might yield more demand with configuration TRUE than capacity allows (overdemand), leading to an infeasible solution. Therefore, to compare the results, the MILP is slightly adjusted for the evaluation of $\boldsymbol{\gamma}_i^{\text{APPROX}}$ with TRUE. An additional decision variable is introduced that allows to determine the demand in an area and time slot that needs to be considered as overdemand at the given capacity level, if demand arises according to TRUE but $\boldsymbol{\gamma}_i^{\text{APPROX}}$ are applied. The demand captured by this variable is then technically not included in the calculation of profit and capacity consumption. This decision variable is weighted with a large negative value in the objective function to only consider demand as overdemand which indeed cannot be physically served but not in favor of a better profit performance.

Then, the results are compared. The comparison is based on solving 100 instances for every examined capacity level. Note, the solution of one instance always comprises solving the instance with TRUE and APPROX and then, evaluate APPROX's solution with TRUE (cf. §4.2.1). The comparison of results is based on the average over the instances' solutions. The following statistics will be provided:

- REV refers to the average revenue (profits before distribution cost) obtained from serving the arising customer demand.
- COST refers to the average expected travel cost for serving the arising customer demand and the fixed cost for the vehicles in use.
- PROF refers to the average expected profit obtained after the distribution cost.
- TIME refers to the average solution time of the model in seconds.
- OVD refers to the share of expected demand that needs to be considered as overdemand at the given capacity level due to the decisions for the time slots on offer $\boldsymbol{\gamma}_i^{\text{APPROX}}$.
- #SLO refers to the average number of time slots on offer in a delivery area.
- REV REL refers to the average relative difference in revenue to the revenue obtained with configuration TRUE.
- DEM REL refers to the average relative difference in arising demand to the demand arising with configuration TRUE.
- SER ABS refers to the absolute demand served.
- SER REL refers to the average relative difference in served demand to the demand served with configuration TRUE.

4.2.2 *Benefit of customer choice modeling*

In this section, we compare the results of TRUE and APPROX, focusing on the impact of modeling the choice preferences for the time slot alternatives in a detailed fashion as accomplished by our new approach. More precisely, there are more and less popular time slots and, thus, demand arises in dependence of the offered time slots. We consider the first problem class (cf.§4.1). To exclusively investigate the impact of

Table 3: Results for adequately considering time slot preferences

		Solution						Relative Performance [%]			
Configuration	Service time τ	REV	COST	PROF	TIME [s]	OVD [%]	# SLO	REV REL	DEM REL	SER REL	PROF REL
TRUE	14	953.3	317.6	635.7	413.6	–	2.43	0.0	0.0	0.0	0.0
APPROX		946.9	318.1	628,8	370.6	3.76	3.21	-0.67	3.12	-0.75	-1.09
TRUE	12	1101.3	319.2	782.1	1198.7	–	2.66	0.0	0.0	0.0	0.0
APPROX		1082.6	319.5	763.1	1330.4	3.55	3.48	-1.70	2.59	-1.05	-2.43
TRUE	10	1267.1	321.9	945.2	1628.4	–	3.22	0.0	0.0	0.0	0.0
APPROX		1238.2	321.5	916.7	1463.8	3.97	4.19	-2.28	1.53	-2.50	-3.02
TRUE	8	1413.6	325.2	1088.4	439.9	–	5.37	0.0	0.0	0.0	0.0
APPROX		1393.0	326.2	1066.8	167.6	2.53	5.87	-1.46	1.14	-1.14	-1.99

Table 4: Overdemand rate and resulting demand when managing demand by means of APPROX

Service time τ		s_1	s_2	s_3	s_4	s_5	s_6
14	OVD [%]	1.77	0	0	0	0.57	1.42
	SER ABS	14.4	12.2	13.2	12.2	14.0	13.8
12	OVD [%]	1.68	0	0	0	0.55	1.32
	SER ABS	16.5	14.0	15.1	14.1	16.4	16.6
10	OVD [%]	1.67	0	0	0	0.66	1.64
	SER ABS	19.3	16.4	17.7	16.4	19.3	19.3
8	OVD [%]	1.07	0	0	0	0.35	1.10
	SER ABS	22.7	18.4	19.9	18.4	22.2	22.7

the adequate consideration of time slot preferences, we assume one customer segment only and use the respective data as described in Appendix A1.

Table 3 reports the results. The first column states the configuration. The second column reports the scarcity of the available capacity, which we artificially enlarge by varying the service time τ between 14 and 8 minutes. The following columns report the statistics explained in §4.2.1 which are additionally summarized in the overview given in Appendix C.

The configuration TRUE clearly outperforms the configuration APPROX. This is intuitive, as the latter assumes that demand in a delivery area spreads evenly over all offered time slots. Hence, it neglects the fact that there are more and less popular time slots in reality, leading to an overall loss in profit of up to 3.02%. For APPROX, the (true) expected demand resulting from the approximation's slotting decisions exceeds the capacity availability in more popular time slots, while there remains idle capacity in less popular time slots, leading to expected demand, which needs to be considered as overdemand. This is further underlined by the results shown in Table 4, which gives the percentage of overdemand and the resulting absolute demand that needs to be served across the different time slots with decisions resulting from APPROX.

While there is no overdemand in the less popular time slots 2, 3, 4 (cf. attraction values in Appendix A.1), overdemand occurs in the more popular time slots 1, 5 and 6. A further consequence of APPROX's assumption is that the average number of time slots on offer is higher than that of TRUE. Table 5 reports the average offer frequency of the single time slots. The term "offer frequency" refers to the average number of delivery areas in which a specific time slot is offered (e.g., if s_1 is offered in two areas, the offer frequency of s_1 is equal to two). For all capacity levels, all time slots are offered less frequently with the configuration TRUE than with APPROX.

Table 5: Average offer frequency of single time slots

		Average offer frequency of single time slots					
Configuration	Service time τ	s_1	s_2	s_3	s_4	s_5	s_6
TRUE	14	2.02	2.07	1.91	2.08	2.02	2.03
APPROX		2.68	2.70	2.74	2.73	2.66	2.59
TRUE	12	2.13	2.28	2.22	2.33	2.13	2.20
APPROX		2.86	2.89	2.89	2.92	2.86	2.77
TRUE	10	2.61	2.89	2.63	2.87	2.54	2.63
APPROX		3.43	3.47	3.55	3.53	3.48	3.45
TRUE	8	4.18	4.85	4.63	4.85	4.18	4.18
APPROX		4.88	4.88	4.90	4.90	4.90	4.90

For the more popular time slots 1, 5, and 6, this is because the time slot restrictions prevent to offer them more often, since otherwise, capacity would be exceeded, which is further supported by the rate of overdemand in these time slots (cf. Table 4). For the less popular time slots, this may have the following reasons: First, it may not be worth to additionally offer an unpopular time slot in an area, since the generated additional revenue would either not overcompensate the delivery cost and the spilled demand of the more popular time slot. Second, it may be that a more popular time slot might not be able to be on offer to not exceed the overall (timely) capacity level, leading to less demand and, thus, less profit in general. If demand is (wrongly) assumed to spread evenly over all offered time slots (APPROX), these effects level off, and more time slots are offered. Further, the offer frequencies are on average almost equal for all time slots owing to the fact that demand arises equally across the time slots and independent of the offered time slots which also leads to an even utilization of the (timely) capacity across the time slots.

For an increased but still scarce capacity level represented by a service time of 10, the loss in profit of APPROX (i.e. PROF REL) has its peak with 3.02%. This is because due to an increased capacity level (compared to a service time of 14 or 12), with APPROX, more popular time slots are offered in which the capacity level is exceeded, and more unpopular time slots are offered in which idle capacity is left than in a case with a tighter capacity level. Hence, the overdemand is higher, and thus, the difference of demand which can possibly be served is greater. If capacity is further increased so that there is almost enough capacity to offer all time slots in most cases (service time 8), the differences in the profit performance as well as the overdemand reduce again. This is due to the fact that, in many cases, all time slots can be offered without timely restrictions and, thus, the solutions of TRUE and APPROX do not deviate much in most cases.

The computational time (cf. Table 3 Column 6) for the different scaling factors and, thus, capacity levels varies strongly. For the least tight and tightest capacity level, it is much smaller than for the medium capacity levels. For a service time factor of 8, there is almost no capacity restriction at all, what significantly decreases the problem's complexity. For the tightest capacity level ($\tau = 14$), a reason might be the smaller solution space than for the medium capacity levels.

Note, instead of merging the different customer segments, their different characteristics (i.e., their different choice behavior) can also be taken into account, since this is explicitly featured by our modeling approach incorporating the MMNL. Following this approach, we did some additional experiments (not shown here), in which the profit potential of the TSMPCC for the scarcest capacity level with a service time of 14 could be further increased compared to APPROX (on average further 1.4%; cf. Appendix A for the data used).

4.2.3 Benefit of profit orientation

In addition to the possibility of an adequate consideration of customer preferences in the TSMPCC, a further consequence arises from its objective function, which maximizes the expected total profit. In comparison, existing approaches based on the TSMP aim to minimize expected total cost (e.g. Agatz et al. 2011, Hernandez et al. 2017). This necessitates the definition of service frequencies (i.e., the number of time slots that needs to be offered in an area) in the different delivery areas, since otherwise, no time slots would be offered at all leading to total cost of zero. Due to the profit orientation of the TSMPCC, these service frequencies are no longer required to be exogenously given input parameters. Instead, these service frequencies are endogenously resulting from the solution of the TSMPCC (equal to $\sum_{n \in \mathcal{N}} \gamma_i^n$ for area a_i). However, if an e-grocer's decision support is based on approximated choice behavior and the TSMP and, thus, needs to specify service frequencies (represented by $sf_i \in \mathbb{N}_0$) the following additional constraints (26) need to be applied:

$$\sum_{n \in \mathcal{N}} \gamma_i^n = sf_i \qquad \forall i \in \mathcal{I} \qquad (26)$$

Since these service frequencies influence arising demand, delivery cost and, hence, expected total profit, we evaluate the benefit of profit orientation and hence, the endogenous determination of service frequencies in comparison to cost orientation with exogenously predefined service frequencies as input parameters. For this purpose and to compare ourselves to decisions made based on assumptions in the academic literature, we assume an e-grocer who derives decisions based on APPROX but with the specification of service frequencies (thus, referred to as $\text{APPROX}_{\text{SF}}$). To simulate realistic service frequencies, which only slightly deviate from the optimal ones, we use the following evaluation procedure: First, we derive decisions based on the configuration APPROX. Second, we alter the resulting service frequencies, i.e., the number of offered time slots, in an area (this is $\sum_{n \in \mathcal{N}} \gamma_i^n$) by adding a randomly drawn number from the set $\{-1, 0, 1\}$. Third, we derive final offer decisions by solving $\text{APPROX}_{\text{SF}}$, imposing the service frequencies obtained from the second step by means of constraints (26). Fourth, we follow our evaluation idea described in §4.2.1 to calculate the "true" profit from the final offer decisions obtained from the third step by plugging them into configuration TRUE according to constraints (25).

Table 6: Results for exogenously and endogenously given service frequencies

		Solution						Relative Performance [%]			
Configuration	Service time τ	REV	COST	PROF	TIME [s]	OVD [%]	# SLO	REV REL	DEM REL	SER REL	PROF REL
TRUE	14	953.3	317.6	635.7	413.6	–	2.43	0.0	0.0	0.0	0.0
$APPROX_{SF}$		938.4	318.3	620.1	352.9	15.83	3.23	-1.56	18.14	-0.56	-2.45
TRUE	12	1101.3	319.2	782.1	1198.7	–	2.66	0.0	0.0	0.0	0.0
$APPROX_{SF}$		1058.2	319.6	738.6	728.5	12.99	3.50	-4.07	13.03	-1.66	-5.56
TRUE	10	1267.1	321.9	945.2	1628.4	–	3.22	0.0	0.0	0.0	0.0
$APPROX_{SF}$		1205.7	321.5	884.2	1433.3	7.30	4.11	-4.85	3.95	-3.64	-6.45
TRUE	8	1413.6	325.2	1088.4	439.9	–	5.37	0.0	0.0	0.0	0.0
$APPROX_{SF}$		1380.7	325.7	1055.0	520.1	3.35	5.52	-2.33	1.68	-1.73	-3.07

Table 6 reports the results. For all capacity levels, the solution of $APPROX_{SF}$ leads to a worse profit performance than TRUE and, in addition, also to a worse profit performance than applying the solution of APPROX with TRUE (cf. §4.2.2 Table 3). The exogenous specification of the service frequencies in $APPROX_{SF}$ leads to the following effects: First, the fulfilment of exogenous service frequencies may lead to a time slot offer or several time slot offers in an area rising demand that is not worth the travel and service time since only a small portion or 'less valuable' portion of demand lives there, or it leads to less time slot offers in areas with valuable demand which negatively impacts the profit performance. Second, much more demand is raised with the decisions of $APPROX_{SF}$ (around 18% at the tightest capacity level) which leads to much higher rates in overdemand of up to 15.83% in case of the tightest capacity level. This is underlined by the results shown in Table 7. Even in the less popular time slots 2, 3, and 4, overdemand occurs. This also results from more inefficient delivery tours due to suboptimal exogenous service frequencies that can force a vehicle to travel to areas in time slots which may not be worth the travel time and cost and, thus, also leads to less served demand compared to TRUE.

Table 7: Overdemand rates and resulting demand when managing demand by means of APPROXSF

Service time τ		s_1	s_2	s_3	s_4	s_5	s_6
14	OVD [%]	4.33	1.34	1.57	1.61	2.09	4.89
	SER ABS	12.9	12.3	12.8	11.8	13.2	12.7
12	OVD [%]	4.26	0.63	1.41	0.37	2.06	4.26
	SER ABS	15.6	14.1	14.5	14.4	15.6	14.9
10	OVD [%]	2.16	0.36	0.44	0.65	0.99	2.70
	SER ABS	19.4	16.7	17.7	16.9	18.1	18.3
8	OVD [%]	1.23	0	0	0	0.36	1.16
	SER ABS	22.2	18.2	20.1	18.8	21.4	21.5

By enlarging the delivery capacity, first, the loss in profit increases while the overdemand rate decreases. Hence, the tighter the capacity level is, the more demand arises due to decisions of $APPROX_{SF}$ that exceeds the capacity restrictions. Since customers considered as overdemand is decided endogenously within the optimization calculus, there is much more flexibility in deciding which demand is served and which not the tighter the capacity is (because much more demand exceeds the capacity restriction) and, thus, the loss in profit is smaller. The less restrictive capacity level ($\tau = 8$) often leads to an optimal solution comprising all 6 time slots. Since no more than 6 time slots can be offered, the randomly altered service frequencies sometimes lead to less time slots on offer than the optimal solution but in a lot of cases also to the same number than the optimal solution. Consequently, the loss in profit decreases again if there is almost no capacity restriction at all. Since the number of offered time slots are given as input parameters for $APPROX_{SF}$, its solution time is lower than without specifying them (cf. APPROX in Table 3), except in case of a service time of 8 minutes, since not simply all slots can be offered than in most cases without given service frequencies.

Even though we simulated only slight deviations from the optimal number, the difference in profit is substantial. This underlines the importance of an integrated optimization of service frequencies in the delivery areas.

4.3 Enlargement of the delivery region

The previous investigations are based on the first problem class and can be solved to optimality in a reasonable amount of computational time. In this section, we investigate how promising solutions for the TSMPCC can be found for an enlarged delivery region represented by the second problem class. For the corresponding problem size (24 areas and 12 vehicles), it is not possible to solve the proposed MILP formulation to optimality within an imposed 12 hours computational time limit due to its computational complexity. Therefore, we use a straightforward heuristic decomposition approach for its solution.

4.3.1 Decomposition approach and evaluation framework

The idea of the decomposition approach is as follows: We geographically partition the set of delivery areas $\mathcal{A}$ in disjoint subsets, assign a certain amount of the given overall capacity (number of vehicles) to each subset dependent upon its size, and then solve separate instances of the MILP, one for each subset. Finally, we aggregate the obtained results. This leads to a reduction of computational effort and is likely to improve the solution obtained within the given time limit by the full model. In more detail, we perform the following steps:

1. Partitioning: Decomposition of the set of delivery areas $\mathcal{A}$ in ρ disjoint subsets $\mathcal{A}_q$ with $q \in \{1, \dots, \rho\}$ and $\bigcup_{q \in \{1,\dots,\rho\}} \mathcal{A}_q = \mathcal{A}$ (cf. Appendix B for details).
2. Proportional assignment of the available vehicles to the ρ subsets. If this is not possible, the remaining vehicles can be assigned, for instance, to the subsets which include the areas that house the most potential customers or the most valuable ones.
3. Instantiation and solving of ρ subproblems with generated subsets $\mathcal{A}_q$, the corresponding distances, the customer segment sizes, the assigned vehicles, and the general input parameters (cf. Figure 2).
4. Aggregation of the results obtained by the subproblems' best solutions found within the imposed time limit (cf. Figure 2). The subproblems' results are aggregated as follows:
   a. Profit: $\sum_{q=1}^{\rho} F_q$, where F_q is the best objective value (profit) found by the solution of the q^{th} subproblem within the imposed time limit;
   b. Cost: $\sum_{q=1}^{\rho} C_q$, where C_q is the total given cost by the solution of the q^{th} subproblem within the imposed time limit;
   c. Integrality gap: $\frac{1}{\rho}\sum_{q=1}^{\rho} GAP_q$, where GAP_q is the integrality gap of the q^{th} subproblem given by the solver after the imposed time limit; and

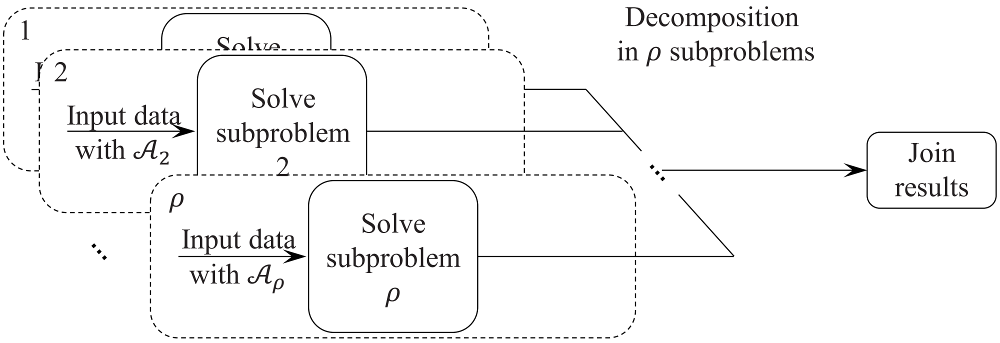

Figure 2: Solution methodology for the decomposition approximation

d. Upper bound (UB): $\frac{1}{\rho}\sum_{q=1}^{\rho} F_q \cdot (GAP_q + 1)$, gives the tightest upper bound of the q^{th} subproblem after the imposed time limit.

For the evaluation, we decompose the original problem into 2, 4, 6, and 12 subproblems. As the configuration, we choose TRUE from §4.2.1 with a service time of 12 minutes, representing a medium capacity level. The overall imposed 12-hour solution time limit is evenly assigned to the subproblems. After solving the MILP instance of the original problem as well as the instances of the subproblems for every ρ with the imposed time limit, we compare the aggregated results.

4.3.2 Results

In Table 8, we report the results for the different granular decompositions. The structure of Table 8 is similar to those presented in the subsections before. The first column states the number of subproblems into which the original one is decomposed, and the second column contains the number of areas $|\mathcal{A}_q|$ considered in each subproblem. Note that $\rho = 1$ represents the original problem, simultaneously considering all 24 areas. Columns 3, 4, 5, and 6 state the averages over all instances of the aggregated profit, the aggregated cost, the aggregated gap, and the aggregated upper bound, respectively. The following columns are in line with the statistics provided in previous tables (cf. Table A.3 in Appendix C). The reported UB in column 6 is an overestimation of the optimal solution's objective value. The true optimal objective value deviates less than the difference between PROF and UB. However, comparing PROF to the highest UB found for the original problem ($\rho = 1$) gives a conservative UB for the maximal loss in profit resulting from not solving the original problem with 24 areas to optimality. This relative maximum loss in profit is given in column 9.

Decomposing the original problem into $\rho = 2$ subproblems (with 12 areas and 6 vehicles each) results in the highest feasible and expected profit, which is 2.44% higher than the profit resulting from the best feasible solution found for the original problem. Even though all subproblems are solved to optimality for $\rho = 6$, the feasible solution found for $\rho = 2$ yields a higher expected profit since the loss in flexibility when only jointly considering 4 areas with 2 vehicles in the optimization is too high. For the same reason, the decomposition with $\rho = 12$ is clearly outperformed by the other decompositions. The loss in flexibility if the decomposition granularity is too high is further emphasized by the loss in demand of 9.36% for the decomposition with $\rho = 12$ in comparison to the original problem.

Comparing the highest PROF (for $\rho = 2$) to the highest UB of the original problem, there is a deviation of 8.25%, which means that the best feasible solution found deviates less than 8.25% from the real optimal solution for the original problem with 24 areas. This might seem like a lot at first glance, but one must keep in mind that it is a very conservative UB for the loss in profit that, in reality, is most likely far less and that the solution of the original problem is far worse.

5 MANAGERIAL IMPLICATIONS AND DISCUSSION

In this paper, we propose a model-based approach for solving an e-grocer's problem of differentiated (static) slotting, i.e., of determining which delivery time slots should be on offer in each delivery area to maximize the expected total profit. To our best knowledge, we are the first to integrate sophisticated customer choice behavior into a model of this problem, which we incorporate by means of a finite-mixture MNL. We show that the resulting non-linear optimization model, that we term TSMPCC, can (exactly) be linearized, obtaining a MILP formulation solvable by linear standard solvers.

Additionally, we systematically set up a number of evaluation experiments to intensively investigate the novel MILP formulation and its profit potential in comparison to the current literature, which

Table 8: Results for different granularities of the decomposition

Parameters		Solution						Relative Performance [%]		
ρ	$\lvert\mathcal{A}_q\rvert$	PROF	COST	GAP [%]	UB	TIME [min]	# SLO	UB LOSS REL	SER REL	PROF REL
1	24	3712.7	1902.3	11.66	4144.8	720	4.08	-10.43	0.00	0.00
2	12	3802.8	1904.9	8.30	4118.0	360	4.04	-8.25	2.41	2.44
4	6	3798.6	1898.7	2.34	3884.3	180	3.88	-8.36	1.70	2.31
6	4	3685.0	1898.6	0.00	-	0.95	3.82	-11.11	-0.63	-0.74
12	2	3155.7	1898.1	0.00	-	0.01	3.81	-23.89	-9.36	-15.01

simplistically approximates customer choice. The managerial implications of our computational study are as follows: First, neglecting or simplistically approximating customer choice behavior leads to a loss in profit of 3.3% on average over all investigated experiments, compared to adequately incorporating customer choice. Second, if cost orientation for decision making is favored over profit orientation, service frequencies in every delivery area need to be specified as input parameters. However, this specification requires a high level of diligence since only slight deviations from the optimal solution lead to an average loss in profit of 4.4% in our study. Third, large problem sizes can heuristically be solved by the decomposition approach we propose, leading to a profit gain of 2.4% over solving the original problem within an imposed 12-hour time limit in our experiments. However, a trade-off between the solution time and the degree of decomposition needs to be found.

Since these implications are derived from computational results that are mainly based on simulated data, we critically discuss the results in the light of the assumptions that are made: First, the results strongly depend on the attraction values for the different time slot alternatives. The stronger these attraction values differ, the higher the profit potential of the TSMPCC in comparison to approaches in the academic literature which approximate customer choice behavior and assume that demand spreads equally across the offered time slots, independent of the e-grocer's offer decisions. For practical application, these attraction values need to be reliably estimated from real world data to gain promising results. If they are poorly estimated due to, for instance, an insufficient data basis and, thus, do not adequately reflect true choice behavior, the profit potential of including customer choice when managing demand by time slot offer decisions will be most likely offset.

Second, for our experiments, we assume a constant travel time for all time slots and a constant service time for all customers. In reality, this requires drivers who are well oriented within their delivery areas to, for instance, circumvent traffic related circumstances (e.g. traffic jams) by taking alternative routes and to ensure smooth operations without delays due to, for instance, orientation purposes (e.g., to find a customer address) (Kunze 2006). Further, construction sites and resulting detours are not modelled in our approximation approach. However, our TSMPCC is formulated to take time slot dependent travel times into account and can be easily extended by customer segment specific service times. On the same lines, in an evaluation of our routing approximation based on real-world customer data, it would be interesting to evaluate whether even better results could be achieved by using several service distances per delivery area, especially in cases with customers being non-homogenously spread among the delivery area. Therefore, an exhaustive evaluation of the delivery cost approximation approach would be insightful.

Third, since our approach is static, and, thus, decisions are not adjusted as soon as information is revealed, the decision quality strongly depends on a reliable data basis and good forecasts. If this is not available (e.g., because of daily changing routes), the decisions made supported by our approach can strongly impact the service quality as well as the profit performance. However, in such cases, the solutions found can serve as a good starting point for subsequent real-time adjustments as soon as new information is revealed, for instance, in the context of dynamic vehicle routing, as far as such dynamic adjustments are acceptable in the specific business context.

Fourth, in our approach, we assume that profit is only influenced by the time slot offer decisions of the e-grocer. However, in practice, non-measurable factors might also impact the long-term profit. For instance, customer satisfaction or the customer lifetime value can be such factors, whose consideration in decision-making might positively impact the long-term profit of an e-grocer (cf. Buhl et al. 2011). Further criteria, which are already examined in a multi-criteria dynamic time window allocation approach in the AHD literature, are customers' social influence and the visibility of deliveries (Lang et al. 2017). Since hardly all such factors can be considered, the big challenge will always be to select the appropriate ones and to quantify them for a promising incorporation in mathematical decisions support.

Beyond, there remain several interesting research directions: First, the results of the decomposition approach might be improved by decomposing the delivery region and assigning the available vehicles to the decomposed parts in a more advanced way, and other heuristics could be developed to solve large problem sizes. Second, since the MILP formulation is computationally expensive, instead of investigating significant efforts into a model-based approach, it could be promising to combine customer choice behavior with descriptive routing approximations (e.g. Daganzo 1987). However, especially with demand that is dependent on the time slots on offer in a delivery area and that mainly influences routing considerations, this might be challenging. Third, with respect to a potentially uncertain and volatile environment, there are limitations of our study that could serve as starting point for further research. Especially an investigation of our approach's robustness towards stochastic influences and systematic forecast errors and how such influences impact the profit performance in comparison to other approaches would be insightful. Fourth, even though not common in the existing literature, considering a planning horizon comprising more than one day might further improve the profit performance due to choice dependencies prevailing beyond the time slots of one delivery day.

REFERENCES

Agatz, N., A. M. Campbell, M. Fleischmann, M. Savelsbergh. 2011. Time slot management in attended home delivery. *Transp Sci* **45**(3) 435–449.

Agatz, N., A. M. Campbell, M. Fleischmann, J. van Nunen, M. Savelsbergh. 2013. Revenue management opportunities for internet retailers. *J Revenue Pricing Manag* **12**(2) 128–138.

Aros-Vera, F., V. Marianov, J. E. Mitchell. 2013. p-Hub approach for the optimal park-and-ride facility location problem. *Eur J Oper Res* **226**(2) 277–285.

Asdemir, K., V. S. Jacob, R. Krishnan. 2009. Dynamic pricing of multiple home delivery options. *Eur J Oper Res* **196**(1) 246–257.

Baker, B. M., J. Sheasby. 1999. Extensions to the generalised assignment heuristic for vehicle routing. *Eur J Oper Res* **119**(1) 147–157.

Baldacci, R., A. Mingozzi, R. Roberti. 2012. Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints. *Eur J Oper Res* **218**(1) 1–6.

Belobaba, P. 1987. Air travel demand and airline seat inventory management, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Bent, R. W., P. van Hentenryck. 2004. Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Oper Res* **52**(6) 977–987.

Bramel, J., D. Simchi-Levi. 1995. A location based heuristic for general routing problems. *Oper Res* **43**(4) 649–660.

Bramel, J., D. Simchi-Levi. 1996. Probabilistic analyses and practical algorithms for the vehicle routing problem with time windows. *Oper Res* **44**(3) 501–509.

Bräysy, O., M. Gendreau. 2005a. Vehicle routing problem with time windows, part I: Route construction and local search algorithms. *Transp Sci* **39**(1) 104–118.

Bräysy, O., M. Gendreau. 2005b. Vehicle routing problem with time windows, part II: Metaheuristics. *Transp Sci* **39**(1) 119–139.

Bruck, B. P., J.-F. Cordeau, M. Iori. 2018. A practical time slot management and routing problem for attended home services. Omega 81 208-219

Buhl, H. U., R. Klein, J. Kolb, A. Landherr. 2011. CR2M – an approach for capacity control considering long-term effects on the value of a customer for the company. *J Manag Control* **22**(2) 187–204.

BVL International. 2017a. *BVL Magazin Drei*.

BVL International. 2017b. *Digitalisierung in der Logistik*: *Antworten auf Fragen aus der Unternehmenspraxis*.

Campbell, A. M., M. Savelsbergh. 2005. Decision support for consumer direct grocery initiatives. *Transp Sci* **39**(3) 313–327.

Campbell, A. M., M. Savelsbergh. 2006. Incentive schemes for attended home delivery services. *Transp Sci* **40**(3) 327–341.

Chao, I.-M. 2002. A tabu search method for the truck and trailer routing problem. *Comp Oper Res* **29**(1) 33–51.

Cleophas, C., J. F. Ehmke. 2014. When are deliveries profitable? *Bus Inf Syst Eng* **6**(3) 153–163.

Daganzo, C. F. 1987. Modeling distribution problems with time windows: Part I. *Transp Sci* **21**(3) 171–179.

Derigs, U., M. Pullmann, U. Vogel. 2013. Truck and trailer routing – Problems, heuristics and computational experience. *Comp Oper Res* **40**(2) 536–546.

Ehmke, J. F., A. M. Campbell. 2014. Customer acceptance mechanisms for home deliveries in metropolitan areas. *Eur J Oper Res* **233**(1) 193–207.

Fisher, M. L., R. Jaikumar. 1981. A generalized assignment heuristic for vehicle routing. *Networks* **11**(2) 109–124.

Haase, K., S. Müller. 2014. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *Eur J Oper Res* **232**(3) 689–691.

Hernandez, F., M. Gendreau, J.-Y. Potvin. 2017. Heuristics for tactical time slot management: A periodic vehicle routing problem view. *Intl Trans Oper Res* **24**(16) 1233–1252.

Holzapfel, A., A. Hübner, H. Kuhn, M. G. Sternbeck. 2016. Delivery pattern and transportation planning in grocery retailing. *Eur J Oper Res* **252**(1) 54–68.

Klein, R., J. Mackert, M. Neugebauer, C. Steinhardt. 2018. A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum* **40**(4) 969–996.

Klein, R., M. Neugebauer, D. Ratkovitch, C. Steinhardt. 2019. Differentiated time slot pricing under routing considerations in attended home delivery. *Transp Sci* **53**(1) 236–255.

Koskosidis, Y. A., W. B. Powell, M. M. Solomon. 1992. An optimization-based heuristic for vehicle routing and scheduling with soft time window constraints. *Transp Sci* **26**(2) 69–85.

Kunze, O. 2004. A new interactive approach on route planning with tight delivery time windows. E. Taniguchi, R. G. Thompson, eds. *Logistics systems for sustainable cities*. Emerald Group Publishing Limited.

Kunze, O. 2005. Ein praxistauglicher Ansatz zur Lösung eines spezifischen D-VRSP-TW-UC. H. Fleuren, D. den Hertog, P. Kort, eds. *Oper Res Proc 2004*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kunze, O. 2006. *Tourenplanung für den eCommerce-Lebensmittel-Heimlieferservice*. Dissertation. University Press Karlsruhe, Karlsruhe.

Lang, M., J. F. Ehmke, C. Cleophas. 2017. Multi-criteria time window allocation for attended home deliveries. Working Paper. RWTH Aachen University.

Laporte, G. 2009. Fifty years of vehicle routing. *Transp Sci* **43**(4) 408–416.

McFadden, D., K. Train. 2000. Mixed MNL models for discrete response. *J Appl Econ* **15**(5) 447–470.

Pacheco, J., R. Martí. 2006. Tabu search for a multi-objective routing problem. *J Oper Res Soc* **57**(1) 29–37.

Pillac, V., M. Gendreau, C. Guéret, A. L. Medaglia. 2013. A review of dynamic vehicle routing problems. *Eur J Oper Res* **225**(1) 1–11.

Russell, R. A. 1995. Hybrid heuristics for the vehicle routing problem with time windows. *Transp Sci* **29**(2) 156–166.

Semet, F. 1995. A two-phase algorithm for the partial accessibility constrained vehicle routing problem. *Ann Oper Res* **61**(1) 45–65.

Solomon, M. M. 1987. Algorithms for the vehicle routing and scheduling problems with time window constraints. *Oper Res* **35**(2) 254–265.

Train, K. 2009. *Discrete choice methods with simulation, 2nd Edition*. Cambride University Press, Cambridge, MA.

Ulmer, M. W., D. C. Mattfeld, F. Köster. 2018. Budgeting time for dynamic vehicle routing with stochastic customer requests. *Transp Sci* **52**(1) 20–37.

Xia, L., K. B. Monroe, J. L. Cox. 2004. The price is unfair!: A Conceptual Framework of Price Fairness Perceptions. *J Mark* **68**(4) 1–15.

Yang, X., A. K. Strauss. 2017. An approximate dynamic programming approach to attended home delivery management. *Eur J Oper Res* **263**(3) 935–945.

Yang, X., A. K. Strauss, C. S. M. Currie, R. Eglese. 2016. Choice-based demand management and vehicle routing in e-fulfillment. *Transp Sci* **50**(2) 473–488.

APPENDIX A CUSTOMER DATA GENERATION

For customer data generation purposes, we generally assume the market to be split into three customer segments (students, family, and professionals). The customer segments' sizes ω_{io} in the delivery areas are drawn randomly and equally distributed from the interval [5, 15]. Even though this distribution is artificial, it reflects the fact that different areas house different proportions of the customer segments (Cleophas and Ehmke 2014). The average revenue of every customer segment is drawn equally from the intervals $[€20, €40]_{\text{Student}}$, $[€40, €80]_{\text{Family}}$, and $[€35, €55]_{\text{Professionals}}$ (Klein et al. 2019). By multiplying the drawn average revenue with a commonly used gross profit margin for e-grocers of 25% (Klein et al. 2019, Yang et al. 2016), the customer segments' average profit $\bar{r}_o$ (before order fulfilment) is obtained. The attraction values are loosely derived from the practical inspired setting of Klein et al. (2019) which assume that different customer segments perceive different attractions for the time slot alternatives with a clear preference for a certain time of the day. The attraction values are given in Table A.1.

Table A.1: Customer segments' attraction values

Segments	Attraction Values						
	None	1	2	3	4	5	6
Student	1.0	0.3	0.9	2.1	2.0	2.2	1.5
Family	1.0	2.6	1.2	0.5	0.5	1.6	2.6
Professional	1.0	2.2	1.8	1.6	1.4	1.0	1.0

A.1 Merging customer segments for evaluation purposes

To conduct evaluations which are based on one customer segment, we merge the data generated for the different customer segments (cf. Appendix A) to obtain data for only one customer segment, i.e. $\mathcal{O} = \{1\}$. Note, this merge in segments is only done for evaluation purposes to exclusively focus on the investigated dimensions. In the following, we omit the segment index o for the ease of explanation and define the merged segment's size in a delivery area $i \in \mathcal{I}$ as $\omega_i = \sum_{o\in\mathcal{O}} \omega_{io}$. The merged segments' average profit in a delivery area $i \in \mathcal{I}$ is obtained by $\bar{r}_i = \frac{\sum_{o\in\mathcal{O}} \bar{r}_o \omega_{io}}{\sum_{o\in\mathcal{O}} \omega_{io}}$.

Note, these values are calculated based on the drawn segments' sizes and revenues for every instance and are used in the objective function (5) instead of the segment specific profit $\bar{r}_o$.

The attraction values for the merged customer segment are obtained from the three defined segments (cf. Table A.1) by adjusting the attraction values, so that equal expected demand results for every time slot (cf. Table A.2). Technically, that means $\frac{\sum_{o\in\mathcal{O}} \bar{\omega}_o \hat{A}_o^n}{\sum_{o\in\mathcal{O}} \bar{\omega}_o}$ $\forall n \in \mathcal{N}$, representing the average over the segments' attraction values for slot $n \in \mathcal{N}$ weighted by the segments' sizes mean $\bar{\omega}_o$, must be equal for the merged segment and the joint consideration of the three segments. Since we assume an equal distribution and, thus, an equal mean segment size $\bar{\omega}_o$ for all segments $o \in \mathcal{O}$, the calculation reduces to $\frac{\sum_{o\in\mathcal{O}} \hat{A}_o^n}{|\mathcal{K}|}$. For time slot 1, for instance, we obtain $\frac{0.3+2.6+2.2}{3} = 1.7$. Note, the merged customer segment's attraction values remain constant throughout the entire computational experiments.

Table A.2: Merged customer segment's attraction values

Segment	Attraction Values						
	None	1	2	3	4	5	6
Merged	1.0	1.7	1.3	1.4	1.3	1.6	1.7

A.2 Customer's attractions for approximated choice behavior

Current approaches in the academic literature often assume equal popular time slots and, thus, demand which equally spreads among the time slots and which arises independently of the e-grocer's time slot offer decisions. To ensure a fair comparison between adequate time slot preferences (cf. Table A.1 and Table A.2) and approximated (i.e. equal) ones, the attraction values in the approximated case are adjusted, so that the same amount of total expected demand results, if all time slots are offered. Technically, that means that the average over all attraction values $\frac{\sum_{n\in\mathcal{N}} \sum_{o\in\mathcal{O}} \hat{A}_o^n}{|\mathcal{S}|\cdot|\mathcal{K}|}$ must be equal for the true and the approximated choice behavior. For the approximated case and with respect to the chosen attraction values in Table A.1, the attraction value of the no choice probability remains 1. Thus, for all time slot alternatives, the attraction values become $\frac{(0.3+0.9+2.1+2+2.2+1.5)+(2.6+1.2+0.5+0.5+1.6+2.6)+(2.2+1.8+1.6+1.4+1+1)}{6\cdot 3} = 1.5$.

APPENDIX B DECOMPOSITION APPROACH: PARTITIONING

To find a partition of set $\mathcal{A}$, we define ρ, which gives the number of subproblems into which the original problem considering all $|\mathcal{A}|$ delivery areas simultaneously is divided (degree/granularity). Now, we consider the delivery areas "less distant" to each other in one subproblem. For this purpose, we step by step add the closest delivery area to the set of areas $\mathcal{A}_q$ (and the corresponding index to the set $\mathcal{I}_q$) already considered in subproblem $q \in \{1, \ldots, \rho\}$; starting with $\mathcal{A}_q = \{a_0\}$ for every subproblem q. Algorithm 1 describes the procedure. Note that, with regard to line 8, we additionally assume the minimum to be unique, which can be ensured by the input data.

Algorithm 1: Greedy heuristic to partition the delivery region

1 0. Initialization:

2 $\quad \mathcal{I} := \{1, \ldots, I\}$

3 $\quad \mathcal{A} := \{a_i : i \in \mathcal{I}\}$

4 $\quad \rho \in \{1, \ldots, |\mathcal{A}|\}$

5 1. Decomposing the delivery region:

6 $\quad$ **for** $q = 1$ to ρ **do**

7 $\quad\quad \mathcal{A}_q := \{a_0\}$

8 $\quad\quad$ **while** $|\mathcal{A}_q| < \left\lceil \frac{|\mathcal{A}|}{\rho - q + 1} \right\rceil + 1$ **do**

9 $\quad\quad\quad \mathcal{A}_q := \mathcal{A}_q \cup \left\{a_{i^*} | d_{i^*j^*} = min\left\{d_{ij} | i \in \mathcal{I}\backslash\mathcal{I}_q, j \in \mathcal{I}_q\right\}\right\}$

10 $\quad\quad$ **end while**

11 $\quad\quad \mathcal{A} = \mathcal{A}\backslash\mathcal{A}_q, \mathcal{I} = \mathcal{I}\backslash\mathcal{I}_q$

12 $\quad$ **end for**

APPENDIX C ABBREVIATION SUMMARY OF SECTION 4

Table A.3: Abbreviation summary regarding §4

Sets/Parameters/Indices	
$\mathcal{A}_q = \{a_i : i \in \mathcal{I}_q\}$	q^{th} subset of $\mathcal{A}$
$\mathcal{I}_q = \{1, \dots, I_q\}$	q^{th} subset of $\mathcal{J}$
$\rho \in \{1, \dots, \lvert\mathcal{A}\rvert\}$	Number of subproblems
$i \in \{1, \dots, \rho\}$	Index of subproblem

Statistics	
APPROX	Attraction value configuration reflecting the approximated customer choice behavior
$\text{APPROX}_{\text{SF}}$	Attraction value configuration APPROX with exogenously given service frequencies
COST	Average expected travel cost for serving customer demand and fixed cost of vehicles in use
C_q	COST of solution of q^{th} subproblem
DEM REL	Average relative difference in arising demand to demand that arises with configuration TRUE
F_q	Best objective value found for the q^{th} subproblem within the imposed time limit
GAP_q	Integrality gap of the q^{th} subproblem after the imposed time limit
PROF	Average expected profit obtained after the distribution cost (COST)
OVD	Share of expected demand which needs to be considered as overdemand at the given capacity
REV	Average revenue (profits before COST) obtained from serving the arising customer demand
REV REL	Average relative difference in revenue to the revenue obtained with configuration TRUE
SER ABS	Absolut demand which is served
SER REL	Average relative difference in demand served to the demand with configuration TRUE
TIME	Average solution time of the model in seconds
TRUE	Attraction value configuration reflecting the true customer choice behavior
UB	Tightest upper bound of the q^{th} subproblem after the imposed time limit
UB LOSS REL	Relative difference between UB and PROF