Territory-Based Vehicle Routing in the Presence of Time Window Constraints

Michael Schneider* Andreas Stenger† Fabian Schwahn* Daniele Vigo‡

December 13, 2013

Abstract

Territory-based routing approaches (TBRAs) are commonly used to achieve high service consistency, e.g., in the small package shipping industry, but their drawback is a decline in routing flexibility. Consequently, a high percentage of time-definite deliveries, as common in the small package shipping sector, should have a significant negative effect on the solution quality of TBRAs. To the best of our knowledge, no study exists on the magnitude of this effect and the factors that influence it. Therefore, we develop a two-phase TBRA and use it i) to investigate the design requirements of a TBRA for successfully handling time windows, and ii) to study the influence of time window constraints on the performance of such an approach. We find that the consideration of geographical aspects in the districting is paramount for generating high-quality territories, while explicitly incorporating time window characteristics and historical demand data does not lead to a perceptible improvement of the solution quality. Moreover, the efficiency and feasibility forfeits of our TBRA in comparison to daily route reoptimization (RR) are larger if time windows are present. However, significantly higher consistency improvements compared to RR are achieved for time-constrained problems. This is due to the fact that RR solutions to time-definite problems exhibit lower consistency and thus a higher potential for improvement by using a TBRA, which constitutes an important insight for practitioners.

Keywords: routing consistency, service territories, small package shipping, vehicle routing problem, time windows, tabu search.

1 Introduction

Due to high competition on saturated markets, small package shipping (SPS) companies are forced to shift more and more attention to customer service and efficient workforce management in order to stay competitive. By having the same driver visit the same set of customers regularly, the driver becomes acquainted with the region and the customer locations therein. Hence, he is able to serve the customers more efficiently while enhancing service consistency. Experienced drivers use shortcuts, know about traffic light intervals, anticipate road or traffic problems and

*{schneider@bwl.tu-darmstadt.de, fabian.schwahn@gmail.com}, Logistics Planning and Information Systems, TU Darmstadt, Germany
†andistenger@gmail.com, Lufthansa, Hamburg, Germany
‡daniele.vigo@unibo.it, Dipartimenti di Ingegneria dell’Elettrica e dell’Informazione, University of Bologna, Italy
find parking space more easily, which leads to reduced travel and service times. The regularity of service creates a bond between customer and driver, which yields improved customer service and a competitive advantage resulting from higher customer loyalty and improved reputation (Wong 2008, Smilowitz et al. 2013).

To achieve such consistency benefits, the routing operations of SPS companies are commonly based on the division of the depot area into service territories (STs), each visited by a single driver (Malandraki et al. 2001, Wong 2008). Due to the pre-assignment of customers to drivers, such territory-based routing approaches (TBRAs) implicitly achieve service consistency, decrease routing complexity and reduce administrative costs (Wong and Beasley 1984). The drawback of TBRAs is the decline in routing flexibility, which yields route configurations that are suboptimal concerning route efficiency (e.g., measured in total traveled distance) when faced with varying demand and/or numerous and possibly tight time window requirements.

In order to find a tradeoff between achieving service consistency and maintaining routing flexibility, some recent scientific approaches present optimization models for SPS routing problems that explicitly integrate consistency requirements and are solved without utilizing STs (see Groër et al. 2009, Sungur et al. 2010, Smilowitz et al. 2013). On the other hand, TBRAs are adapted by allowing to adjust them based on the daily demand, i.e., by excluding a percentage of customers from being assigned to territories as done at UPS and described in Zhong et al. (2007). However, to the best of our knowledge, none of the published works on routing consistency explicitly considers the existence of time windows with the partial exception of Sungur et al. (2010), who consider soft time windows. This strongly conflicts with recent practical developments. Our industry contacts state that up to 60% of their orders are time-definite, which is consistent with the statistics given in Campbell and Thomas (2009).

If time windows are considered, routing flexibility is not only needed to achieve distance-efficient route configurations but also to fulfill the time requirements of customer deliveries. Thus, the value of routing flexibility should increase, which is likely to have a negative effect on the solution quality of TBRAs. However, no study exists on the magnitude of this effect and the factors that influence it. Nevertheless, several private communications with employees of German SPS companies and the literature on SPS routing indicate that, despite the high percentage of time-definite deliveries, TBRAs are widely utilized.

The central questions studied in this work are:

- How well do TBRAs perform in the presence of time window requirements? More precisely, we study i) to what extent the goals of feasible and distance-efficient routes can be achieved with a TBRA under time window constraints, and ii) how well the implicit realization of consistency benefits (without directly optimizing consistency in any way but only due to the restriction to STs) is achieved.
- How strong is the influence of factors such as the size of the STs, the customer distribution, the tightness, frequency and spatial distribution of time windows or the variance of the daily demand on this performance?

To address these questions, we develop a TBRA to solve a series of Vehicle Routing Problems with Time Windows (VRPTW) that are linked by a common customer base set. On every day of a given time horizon, a random subset of these customers requires service. This corresponds
to the practical problem faced by many SPS companies. The proposed TBRA is a two-phase method called Modular Territory Routing (MTR) that first divides the delivery area into STs based on historical, spatial and time window related information. In the second phase, it defines the daily routing based on these STs. The aim of our approach is to provide a performance that is good enough to achieve meaningful results in the numerical studies addressing the above questions while maintaining (relative) simplicity.

The paper is organized as follows. In Section 2, a review of the related literature is presented. In Section 3, an overview of MTR is provided and its two phases are detailed in Sections 4 and 5. The numerical studies are presented in Section 6. Apart from studying the above described research questions, we provide a comparison of MTR with approaches presented in the literature to show the competitiveness of our approach. Section 7 presents the conclusions and an outlook on possible future research questions. A list of notations and abbreviations is included in Appendix E of the provided online appendix. Note that in the following text, all references to the appendix relate to the online appendix.

2 Literature Review

This work is closely related to the literature on VRPTW because the daily problems faced by SPS companies generally belong to this problem class. The VRPTW considers time intervals in which service at each customer has to take place and it can thus be used to accurately model many real-world distribution problems, such as, e.g., small package deliveries. The solution of a VRPTW calls for the determination of a cost-minimal set of routes carried out by a set of homogeneous vehicles located at a single depot. Each route starts and ends at the depot within a given scheduling horizon and the cumulative demand of the customers visited on a route does not exceed vehicle capacity. Each customer is served by exactly one vehicle within the associated time window.

The VRPTW is NP-hard and only small to medium-sized instances can be solved by exact methods even within large amounts of computing time (see, e.g., Kallehauge 2008, Baldacci et al. 2012). Consequently, intensive research effort has been given to metaheuristic optimization approaches for VRPTW, which resulted in a large number of successful methods that are able to produce high-quality solutions for more realistically sized instances in reasonable time. The best-performing algorithms are evolution strategies (Mester and Bräysy 2005), memetic algorithms (Nagata et al. 2010, Vidal et al. 2013), large neighborhood search (Pisinger and Ropke 2007, Prescott-Gagnon et al. 2009), iterated local search (Ibaraki et al. 2008), and multi-start local search (Lim and Zhang 2007). Recent surveys of metaheuristic solution methods for VRPTW have been presented in Bräysy and Gendreau (2005a,b), Gendreau et al. (2008) and Gendreau and Tarantilis (2010).

Moreover, approaches addressing driver knowledge and workforce management in vehicle routing are of relevance for this paper. An overview of the literature concerning these topics is presented in the following.

Wong and Beasley (1984) divide the complete delivery area into a number of fixed STs to be served by a single vehicle. The STs are determined based on historical demand data. The main idea of their approach is to aggregate those customers into the same ST that often appear
together on the same route during an initialization phase of several sample days. On each of these sample days, a classical VRP is solved independently without considering any consistency requirements. Their approach is quite simple and provides strong consistency benefits because the STs are completely fixed. However, the inflexibility of the STs leads to significant problems when demand varies strongly as routes are often infeasible due to capacity constraints. Other early works addressing consistency requirements by means of fixed routes or territories are Christofides (1971) and Beasley (1984). The latter work addresses the design of fixed routes that serve the same set of customers each day, where the requested demand of the customers differs from day to day.

Zhong et al. (2007) present a two-stage method for solving large-scale vehicle dispatching problems with uncertain customer locations and demand. In the “strategic phase”, customers are first aggregated in “cells” and a percentage of cells are assigned to “core areas” that serve as STs. Core areas are always visited by the same driver. Cells located between core areas and those included in the “flex zone” around the depot remain unassigned. In the “ operational phase”, routes within the core areas are constructed and the unassigned cells are inserted at the lowest cost in order to generate the daily routes. The insertion cost incorporates a driver learning model that explicitly considers the familiarity of a driver with a certain cell. Numerical studies show that the approach is well suited to balance the tradeoff between driver familiarity and routing flexibility.

Haugland et al. (2007) study the problem of designing fixed STs for a stochastic VRP. For this two-stage (partitioning and routing) problem, they present a tabu search (TS) and a multistart heuristic and perform numerical studies on classical VRP benchmarks instances that they adapted to their problem. Further works addressing districting problems include Daganzo and Erera (1999), Erera (2000), Ouyang (2007), Carlsson (2012) and Salazar-Aguilar et al. (2012).

Haughton (2008) studies the effect of exclusive territory assignment on routing efficiency against the benchmark of daily route reoptimization (RR), where drivers are flexibly assigned based on the day-to-day demand situation. The possibility of territory sharing is considered, i.e., a team of drivers visits a territory in order to reduce the negative impact of exclusive assignment on routing efficiency by means of the pooling effect. Simulations are used to investigate how the variance of customer demand, the vehicle capacity and the size of the team of drivers assigned to each territory influence the routing efficiency.

Groër et al. (2009) introduce the Consistent VRP (ConVRP), a multi-period routing problem that considers service consistency requirements. The problem requires that “the same driver visits the same customers at roughly the same time on each day that these customers need service”, in addition to the traditional constraints on capacity and route length. They develop a two-phase algorithm based on record-to-record (RTR) travel, which first constructs template routes and then uses them to generate the daily routes by removing non-occurring customers and inserting new ones. The template routes are based on a simple precedence principle. The principle states that if two customers \(a\) and \(b\) are served by the same vehicle on a specific day, then the vehicle that serves them and the order in which they are served must be the same on all days on which they both require service. Numerical studies on a number of generated and one
real-world data set show that their approach is able to produce routes that fulfill the consistency requirement to a high degree. However, compared to the results of a standard heuristic VRP solution, travel distance and the number of vehicles increase slightly.

Sungur et al. (2010) present the courier delivery problem, modelled as a multi-day VRP with soft time windows, using robust optimization and scenario-based stochastic programming to represent uncertainty in service time and probabilistic customers. An insertion heuristic enhanced by a TS improvement method is presented and tested in numerical studies. With slight adaptations, they can provide solutions that adhere to the precedence principle of Groër et al. (2009) and where each customer is served by the same vehicle whenever it requires service. In this way, they are able to compare their approach to that of Groër et al. (2009) on the ConVRP benchmark and are able to improve the results.

Smilowitz et al. (2013) present measures to quantify the effect of visiting a customer repeatedly with the same driver. Thus, they are able to balance the resulting consistency benefits against additional routing costs such as, e.g., increased traveled distance. They propose three different periodic vehicle routing problems (PVRPs) that include consistency measures as a part of the objective function and use a base PVRP with the objective of minimizing traveled distance for comparison. To solve the PVRPs, a TS is developed and studies are carried out on the ConVRP benchmark proposed by Groër et al. (2009). They demonstrate that by including the consistency measures in the objective function, an appropriate balance between route efficiency and consistency can be obtained. On the other hand, solving the base PVRP and applying post-processing steps to increase consistency does not lead to convincing results.

Another recent approach to address consistency requirements in VRPs is presented by Coelho et al. (2012), who propose an inventory routing model that integrates consistency measures, including driver consistency, and develop a matheuristic to solve the model.

As mentioned above, apart from Sungur et al. (2010), none of the above presented works considers the existence of time windows. Moreover, Sungur et al. (2010) only account for soft time windows and we are not aware of any TBRA dealing with time requirements.

### 3 Overview of Modular Territory Routing Approach

Our general goal is to provide a simple and flexible TBRA with a solution quality that is good enough to provide meaningful results in our numerical studies, which analyze the influence of time window restrictions on the design requirements and the performance of a TBRA. The proposed method, called Modular Territory Routing (MTR), is divided into two phases. The districting phase generates the STs that are permanently assigned to drivers, i.e., each customer within an ST is later served by the same driver whenever requiring service. The routing phase produces the daily delivery routes based on the STs created in the districting phase. MTR is designed to solve multi-day series of VRPTW achieving efficient routes (in terms of both the number of employed vehicles and the traveled distance) while providing consistency in the delivery to customers. The latter is accomplished through the restrictions imposed by the STs on the daily service. Our approach does not consider overtime expenditures as overtime is strongly limited in practice due to legal restrictions on working hours.

Thus, the goal of the districting phase is to create an assignment of customers to STs which is
of high quality with respect to both consistency and flexibility requirements. Routing flexibility is important to obtain vehicle- and distance-efficient routes. The concepts for incorporating flexibility in MTR, namely excluding a percentage of customers from being assigned to STs and the utilization of an exclusion zone around the depot as shown in Figure 1, are inspired by Zhong et al. (2007).

![Figure 1: The delivery area: Service territories, exclusion zone and seed customers](image)

We design MTR in a modular fashion in order to be able to study the design requirements of a TBRA dealing with time window constraints and in order to increase its potential practical usability. The two phases of MTR are independent of one another and can be implemented by different modules. The same applies for the steps of the districting phase. To analyze the design requirements, we present several approaches for realizing the modules, whose quality is later investigated in the numerical studies. Generally, for a step of the districting phase, we present both a basic procedure and an advanced one that incorporates customer time window requirements in a more sophisticated manner. The steps of the districting phase are:

**Definition of the Exclusion Zone** The districting phase starts with the definition of an exclusion zone around the depot. The customers within this zone are always assigned to drivers on a daily basis and are therefore never assigned to one of the STs (see Figure 1).

**Solution of Sample Day Problems** Inspired by real-world situations, where SPS companies store and analyze detailed delivery and customer data to support their route planning operations, we generate a series of sample days (scenarios) representing historical demand as proposed by Wong and Beasley (1984). More precisely, the historical data is used to generate a series of VRPTW instances each representing one day of customer requirements. The instances are solved by means of a so called sample-day VRPTW method and the information on the structure of the obtained solutions are then used both to determine the desired number of STs to be generated, and to assist in the creation of reasonable STs. Note that each of the sample days is solved separately without utilizing STs or imposing any restrictions concerning consistency. Thus, any VRPTW solution method can be used to solve the sample days, while the potential size of the
daily problems of an SPS company recommends the utilization of metaheuristics instead of exact methods. In the computational testing, we investigate the potential impact of different qualities of sample-day VRPTW methods on the overall MTR performance.

**Selection of Seed Customers** Next, a set of seed customers (see Figure 1) is chosen based on spatial and time window related characteristics, such as, e.g., the number of neighboring customers and the compatibility of the selected customer’s time window with those of the neighbors.

**Assignment of Customers to Territories** Finally, a predefined percentage of customers are added to the seed customers in order to generate the STs as illustrated in Figure 1. Spatial characteristics are considered to generate “well-shaped” STs and time window characteristics are also incorporated. In addition, the selection of the customers to be assigned to the STs is based on the above described routing solutions obtained on the VRPTW sample days (cp. Wong and Beasley 1984). To describe the basic idea, we say that if two customers are often served on the same route during the sample days, then they are potentially good candidates to belong to the same ST. On the contrary, customers that are never served on the same route seem to be incompatible concerning spatial and/or time characteristics and should probably not be included in the same ST.

In the *routing phase*, daily routes are designed based on the STs developed in the districting phase. The daily problems are solved completely independent of one another, i.e., we are not optimizing consistency goals such as, e.g., delivery consistency over the considered time period. Our optimization goal is to minimize the number of employed vehicles and the traveled distance on each day while adhering to the restriction posed by the STs, the latter promoting delivery consistency. This is achieved by means of a so called *territory-based VRPTW method* that is capable of solving VRPTW problems with the additional feature that part of the customers are preassigned to STs and are therefore served by specific drivers.

An overview of our MTR approach is given in Figure 2. The districting phase is detailed in Section 4, the routing phase is described in Section 5.

**Figure 2:** Overview of the districting and routing phase of MTR
Similar to our approach, the methods presented by Groër et al. (2009) and Sungur et al. (2010) are designed to address multi-day VRP instances pursuing service consistency goals. Both works forego the use of STs and instead utilize some sort of master schedule that is adapted for the daily routing requirements, thus promoting consistency. Like MTR, both approaches exploit historical demand data (scenarios) to generate their template schedule. However, the main constraints or optimization goals are quite different from MTR. Groër et al. (2009) minimize traveled distance while guaranteeing that the same driver always visits a customer requiring service and service should take place at approximately the same time on each service day. Sungur et al. (2010) optimize a weighted function of customer coverage, total travel time, lateness and earliness penalties and route similarity.

In contrast, MTR aims at generating efficient and feasible vehicle routes given hard time window constraints. Consistency is realized by adhering to the customer-driver assignment restrictions imposed by the STs, which closely relates our approach to the territory-based method of Zhong et al. (2007). They proposed the general ideas for enhancing flexibility in a TBRA that are implemented in MTR. However, as described above they do not consider time window requirements, although these play a major role in the routing operations of SPS. Our numerical studies show that MTR obtains quite competitive results on the multi-day VRP instances without time windows of Groër et al. (2009).

4 The Districting Phase: Designing the Service Territories

This section details the districting phase, which consists of the following steps: i) definition of the exclusion zone, ii) solution of sample-day VRPTW problems, iii) selection of seed customers for the STs, and iv) iterative assignment of customers to the STs.

4.1 Definition of Exclusion Zone

As already noted by O’Brien (1975) and Beasley (1984), all vehicle routes pass through the area around the depot and the customers in this area can be reached from every route without making long detours. Thus, such customers can be used to efficiently balance loads between routes in daily routing. As previously mentioned, this insight was also used in the creation of a flex zone in Zhong et al. (2007) and inspired the implementation of the exclusion zone in our MTR.

We present two methods for generating the exclusion zone. The first, basic method, called $\text{EZ}_{\text{basic}}$, does not take into account the time window characteristics of customers but solely considers geographical distance from the depot. Thus, $\text{EZ}_{\text{basic}}$ assigns all customers that are located within a radius $r_{\text{EZ}}$ around the depot to an exclusion zone $\text{EZ}$ and prohibits the assignment of those customers to an ST. Based on the historical demand data, the radius is chosen to include a given percentage $\omega$ of $|\text{V}_{\text{total}}|$, where $\text{V}_{\text{total}}$ denotes the set of all customers in the customer base.

The second, advanced method, called $\text{EZ}_{\text{tw}}$, incorporates time window characteristics of customers. As mentioned above, customers within $\text{EZ}$ are preferably visited on leaving or returning to the depot. Therefore, the idea is to include those customers into $\text{EZ}$ that have a time window
close to the beginning or the end of the depot’s scheduling horizon. Again, a percentage \( \omega \) of the total customers are added to the exclusion zone. However, the selection of these customers is based on a weighted function of the spatial distance to the depot and the temporal distance of the time window to the beginning or end of the scheduling horizon. More precisely, let \( 0 \) denote the depot vertex and \( d_{vw}, v, w \in V_{total} \cup \{ 0 \} \) the spatial distance between two vertices. Further, let \( e_v (l_v) \) denote the earliest (latest) time for the start of service at customer \( v \) and \( e_0 (l_0) \) the start (end) of the scheduling horizon. The temporal distance of a customer \( v \) to the scheduling horizon is defined as follows: \( d_{\text{temp}}^0 v = \min (e_v - e_0, l_0 - l_v). \)

Now, \( EZ \) is generated iteratively by selecting the customer \( v \) with

\[
v = \arg \min_{w \in V_{total} \setminus EZ} \left( \varepsilon \cdot d_{0v} + (1 - \varepsilon) \cdot d_{\text{temp}}^0 v \cdot \sum_{w \in V_{total}} \frac{d_{0w}}{d_{\text{temp}}^0 w} \right).
\]

The parameter \( \varepsilon \) is used to weight the influence of the spatial and temporal distance, while the term \( \sum_{w \in V_{total}} \frac{d_{0w}}{d_{\text{temp}}^0 w} \) relates the amounts of the average spatial and temporal distance. After the selection of a customer, we determine the convex hull of all currently selected customers. We include all previously unassigned customers that are now located within the convex hull. This is done to ensure that there are no enclaves of customers within \( EZ \). The process stops as soon as the number of customers in \( EZ \) has reached the desired number.

We have investigated the similarity of the exclusion zones generated by \( EZ_{\text{basic}} \) and \( EZ_{tw} \) are. To this end, we conducted numerical studies on 60 series of 100 VRPTW days described in Section 6.4.1. We found that the percentage of customers that the exclusion zones have in common is 66.7% on average. Considering that both approaches select customers that are geographically close to the depot, this indicates that the two methods are able to generate quite different results.

### 4.2 Solution of Sample Day Problems

From the historical demand data, we create a series of \( \tau_1 \) sample days, where on each sample day \( t = 1, \ldots, \tau_1 \), a subset \( V_t \) of the set of all customers \( V_{total} \) require service, i.e., the days of the series are linked by a common customer base. As mentioned above, neither consistency requirements nor any other dynamic effects are considered for generating the sample day solutions. Thus, the problem to be solved here is a VRPTW. The sample-day VRPTW method follows a hierarchical objective and minimizes the number of employed vehicles before minimizing the traveled distance. This is motivated by the fact that later the number of generated STs (the number of employed vehicles) is set to achieve a certain service level over the sample days. Therefore, we are interested in the minimal number of vehicles with which one can achieve complete service of the customers on each day.

As sample-day VRPTW method, we use a tabu search (TS). TS is a powerful metaheuristic, which guides local search heuristics in order to escape from local optima and provide near-optimal solutions (Glover 1986). TS has successfully been applied to various combinatorial optimization problems including VRPs (Gendreau and Potvin 2010). The proposed TS, called high-quality TS (TS_{hq}), is similar to the TS methods in Schneider et al. (2012, 2013) but differs regarding
the approach for generating the initial solution and the set of diversification methods. TS\textsubscript{hq} is designed in a simple fashion and incorporates well-known concepts from the literature (see, e.g., Cordeau et al. 2001, Toth and Vigo 2003). A detailed description of TS\textsubscript{hq} is given in Appendix A.

On the well-known VRPTW test instances of Solomon (1987), TS\textsubscript{hq} achieves results of a quality comparable to the best approaches in the literature in fast run-time (see Appendix A.6 for more details). In order to investigate the influence of the utilized sample-day VRPTW method on the overall performance of MTR in numerical studies, a second, simpler TS method was also developed. This method, denoted as medium-quality TS (TS\textsubscript{mq}), has a significantly lower solution quality than TS\textsubscript{hq} and is detailed in Appendix B.

### 4.3 Selection of Seed Customers

The number \( m_{st} \) of STs to be created depends on the solutions obtained for the sample VRPTW problems. More precisely, \( m_{st} \) corresponds to the number of vehicles required by the sample-day VRPTW method to find a feasible solution for a given percentage \( \theta \) of the VRPTW sample days. This procedure is again inspired by industry practice, as it is not realistic to have a number of vehicles/drivers that guarantee a 100\% percent service level (see also Beasley 1984).

The task to be performed is to partition a set of customer locations into districts in order to define the STs. For solving such a districting problem, one needs a well-defined notion of adjacency. By only allowing customers to be added to an ST if they are adjacent to it, we guarantee the creation of spatially well-defined STs: We avoid generating disconnected STs or enclaves (i.e., STs within STs), which are very unlikely to be of high quality. For our districting approach, we use the concept of adjacency introduced by Haugland et al. (2007), which was successfully applied in the context of a districting problem. Note that other adjacency concepts, like, e.g., Delaunay triangulation, would also be valid choices.

In the concept introduced by Haugland et al. (2007), two customers \( v \) and \( w \) are called adjacent if the edge between these customers is not intersected by any shorter edge between two other arbitrary customers \( u \) and \( y \). The adjacency graph \( \mathfrak{A} \) consists of all edges connecting adjacent nodes. If a node has several neighbors in two districts then that node is likely to be located close to the border of the two districts (Haugland et al. 2007). To reduce computing times, we remove all edges whose length exceeds \( d_{\text{max}} \) from the adjacency graph regardless of whether they are intersected by any shorter edge or not. A figure illustrating the adjacency concept can be found in Appendix C.1.

Using this definition of adjacency, each of the \( m_{st} \) STs is built around a seed customer as depicted in Figure 1. For selecting the seed customers, we again propose a basic method, called Seed\textsubscript{basic}, and an advanced method incorporating time window aspects, called Seed\textsubscript{tw}.

In Seed\textsubscript{basic}, the selection of seed customers bases on the following criteria: i) a large distance to the depot and all other seed customers, ii) a large number of neighboring customers and iii) a low average spatial distance to neighboring customers. Customers within the exclusion zone are obviously not eligible as seed customers. Let \( I \) denote the set of selected seed customers, vertex 0 the depot, \( \mathfrak{A}(v) \) the direct neighbors of customer \( v \) in \( \mathfrak{A} \), \( \mathfrak{A}_\pi(v) \) the set of neighbors of vertex \( v \) within path length \( \pi \) and \( \bar{d}(\mathfrak{A}_\pi(v)) \) the average distance of \( v \) to these neighbors. The length \( \pi \) is chosen such that the number of customers that are reachable from \( v \) on a path of maximal
length $\pi$ is at least as high as the desired number of customers in the ST. Starting from $I = \emptyset$, we iteratively select $m_{st}$ seed customers according to the following formula and add them to set $I$:

$$
    v_{add} = \arg \max_{v \in V_{total} \setminus (I \cup EZ)} \left( \frac{|A(v)| \cdot \min (d_{vw} | w \in (I \cup 0))}{d(A(v))} \right),
$$

(1)

To incorporate time window aspects in the selection of seed customers, we introduce a measure that evaluates the compatibility of the time window of the seed customer with the time windows of his neighbors. Thus, the advanced method $Seed_{tw}$ selects seed customers based on the following formula:

$$
    v_{add} = \arg \max_{v \in V_{total} \setminus (I \cup EZ)} \left( \frac{|A(v)| \cdot \min (d_{vw} | w \in (I \cup 0))}{d(A(v))} \cdot (1 + \frac{|A_{infeas}(v)|}{|A_{v}(v)|}) \right),
$$

(2)

where $A_{infeas}(v)$ is the set of time window incompatible neighbors of $v$ that can be reached on a path of at most length $\pi$. The time windows of two customers $a$ and $b$ are incompatible if it is not possible to reach customer $b$ within its time window starting from customer $a$ and vice versa.

In the described methods to select the seed customers, only the locations of previously chosen seed customers are taken into account. This results in a non-uniform distribution of seed customers over the delivery area as the distance between adjacent seed customers decreases with every additional seed customer. To provide a more uniform distribution of seed customers over the delivery area and thus a better starting point for well-shaped STs, we use the relocation procedure given in pseudocode in Figure 3.

```
repeat
    for each customer $v \in I$ do
        Determine closest vertex $w_1 \in I \cup 0 \setminus \{v\}$ with $d_{vw_1} = \min d_{vw}$
        Determine second closest vertex $w_2 \in I \cup 0 \setminus \{v, w_1\}$ with $d_{vw_2} = \min d_{vw}$
        $d(v) = d_{vw_2} - d_{vw_1}$
    end for
    Select customer $v_{remove} = \arg \max_{v \in I} d(v)$
    Remove $v_{remove}$ from $I$
    Select customer $v_{add}$ according to Equation (1)/(2) and add it to $I$
until $v_{remove} = v_{add}$ or maximal iteration number reached
```

Figure 3: Pseudocode for the relocation of seed customers

Note that we have chosen to select the seed customers based on heuristic measures trading off several different desirable characteristics of seed customers. Another possibility would have been to solve a p-median problem, i.e., to minimize the distance of all customers to the seed points. However, this can result in spatially close seed customers or seed customers being located between customer clusters, which is both not desirable for our approach.

### 4.4 Customer Assignment to Territories

After the set of seed customers $I$ has been selected, unassigned customers are iteratively added to the STs. We generate a set of $m_{st}$ territories that approximately include a given percentage $\rho$ of the total customer demand and are roughly of equal size, measured by the sum of expected
demands of the customers assigned to the ST. Let $E(Q)$ denote the average demand of all customers in the delivery area during the $\tau_1$ sample days. The capacity constraint of each ST is calculated by $C_{ST} = \lceil \rho \cdot \frac{1}{m_{st}} \cdot E(Q) \rceil$.

Starting from the STs initialized with the seed customers, we use an iterative approach to decide which of the unassigned customers to add to which ST. The basic procedure for this step, called $Assign_{basic}$, is given in pseudocode in Figure 4 and bases the assignment decision on:

- the average insertion cost with the neighbors in the ST as introduced by Wong and Beasley (1984). Let $a_{vw}$ denote the number of times customers $v$ and $w$ are served by the same driver and $h_v$ the total number of times that customer $v$ is served during the sample days. Then, the cost $b_{vw}$ of having customers $v$ and $w$ together in the same ST is $b_{vw} = \min(h_v, h_w) - a_{vw}; v, w \in V_{total}$. Consequently, an insertion cost of $b_{vw} = 0$ means that customers $v$ and $w$ were visited by the same driver whenever both of them required service on the same day. Note that in this way temporal aspects of a good customer allocation are also implicitly considered since, e.g., customer sequences that are infeasible due to time window violations are penalized by this measure.

- the number of direct neighbors $|A(v, k)|$ that each customer $v$ has in service territory $k$.

- the distance $d(v, g(v))$ to the center of gravity $g(v)$ of the nearest ST that customer $v$ is not adjacent to or the distance to the depot $d(v, 0)$. The underlying idea is to prefer customers that are as far away as possible from all other STs and the depot. Customers close to the depot can easily be added to different routes on a flexible basis. A high distance between STs enables a more uniform distribution of STs and enough of a flexible buffer zone between them. We use a distance exponent $\psi$ as weighting factor to determine the influence of the distance to the other STs and the depot.

To incorporate time windows, we favor customers with wide time windows for assignment to a territory. Consequently, the advanced method $Assign_{tw}$ uses the following formula in Line 10
of the pseudocode in Figure 4:

\[
v_{\text{add}} = \arg \max_{v \in \text{Candidates}} \left( |\mathcal{A}(v, k)|^2 \cdot \min \left( d(v, g(v)), d(v, 0) \right)^{\nu} \right)
\]

\[
\max \left( \sum_{w \in \mathcal{A}(v, k)} b_{vw} \cdot (1 + \frac{\sum_{i \in V_{\text{total}} (l_i - e_i)}}{|V|} \cdot 0.5) \right)
\]

The factor \( \frac{\sum_{i \in V_{\text{total}} (l_i - e_i)}}{|V|} \) decreases if the time window of the customer under inspection is wide in comparison to the average time window width of the instance.

5 The Routing Phase: Route Design Respecting the Preassignment of Customers to Service Territories

In the second phase of MTR, the operational routing of vehicles based on the STs generated in the districting phase is carried out for \( \tau_2 \) evaluation days. As territory-based VRPTW method, we use a modified version of TS\(_{hq}\), called TS\(_{terr}\). Inspired by practice, the number of employed vehicles corresponds to the number of STs, i.e., one vehicle is assigned to each of the STs and serves all customers within the ST that require service on the given day. Opening additional routes is not allowed since we aim at evaluating the suitability of TBRAs for handling problems with time windows. This suitability depends on the feasibility of the obtained solutions, for which we introduce adequate measures in the numerical studies (see Section 6.1). If we allowed to open additional routes to always obtain a feasible solution, we would only be able to indirectly assess the feasibility of the solutions provided by the TBR.

For the initialization of routes in TS\(_{terr}\), we use the modified Solomon I\(_1\) insertion heuristic presented in Appendix A.3. In a first step, the initialization is individually applied to each ST and all customers assigned to an ST are iteratively inserted into the corresponding vehicle route. Subsequently, the remaining customers are added to the initialized routes one at a time at the best place of insertion. The initialized routes are then iteratively improved by TS\(_{terr}\). Those customers that are assigned to STs must stay in the route of the respective ST and must not be considered for relocation. They are simply neglected by inter-route moves.

Since the demand situation in the routing phase is not known during the districting phase, on some days it might not be possible to generate a valid solution that satisfies all capacity and time window constraints and respects the pre-assignment of customers based on STs. This is especially true if demand peaks occur, STs are large and time windows are tight. To mitigate the problem, we increase the routing flexibility by using semi-fixed STs, i.e., we allow to expel customers from their STs and assign them to different drivers. This is incorporated into TS\(_{terr}\) by allowing solutions during the search that are infeasible with respect to the preassignment of customers. The number of customers that can maximally be expelled is gradually increased as long as no feasible solution is found. More precisely, after every \( \eta_{\text{feas}} / m_{\text{st}} \) iterations without finding a feasible solution, where \( \eta_{\text{feas}} \) denotes the maximal number of iterations TS\(_{terr}\) is run to find a feasible solution (see also Appendix A.6), the number of customers that can be expelled is increased by one. In this way, the total number of expelled customers is restricted to the number of STs, i.e., we allow to expel on average one customer per ST.
The violation of the preassignment constraint is handled by means of a penalty mechanism (cp. Appendix A.2). Each customer that is expelled beyond the currently allowable number is penalized. The penalty factor is initialized to $\zeta_0 = 100$ and is restricted to the interval $[1, 6400]$, which is chosen equal to the intervals for the capacity and time window penalty factors in $\text{TS}_{hq}$ (see Appendix A.6). It is modified after every $\eta_{\text{penalty}} = 2$ iterations by multiplying or dividing by $\delta = 1.2$ depending on whether the number of currently expelled customers exceeds or falls below the maximal allowable amount. Note that it is nevertheless possible to end up with an infeasible solution. This problem cannot be avoided if the STs are designed to trade off feasibility and efficiency considerations.

In case $\text{TS}_{terr}$ finds a feasible solution with the given number of vehicles, it tries to reduce the number of vehicles as long as a feasible solution can be found. From a practical perspective, this can be a significant factor for cost reduction if a flexible workforce and fleet management allows to reduce the number of employed drivers and vehicles on a daily basis. To find a feasible solution with a reduced vehicle number, $\text{TS}_{terr}$ first sorts the STs according to their coverage on the considered day. In other words, we determine for each ST the number of customers within this ST that have to be served on the considered day and we order the STs in increasing order of this measure. The first ST is emptied and the route serving this ST is removed from the solution. The customers that were previously assigned to that route are relocated into the other routes at the best possible insertion position. Then, $\text{TS}_{terr}$ is run to find a feasible solution with the reduced number of vehicles. If the route removal is successful, the next ST in the list is emptied until no feasible solution can be found anymore. In case $\text{TS}_{terr}$ is not able to determine a feasible solution, we repeat the emptying step iteratively using the next three STs in the list. The vehicle minimization is stopped as soon as the procedures failed for all three STs.

An illustrative figure that shows the outcome of applying MTR to a with 1000 customers can be found in Appendix C.2.

6 Computational Studies

We performed extensive numerical studies to investigate the performance of the proposed MTR and analyze the influence of several method and problem parameters, such as, e.g., the size of the generated STs or the density and tightness of time windows on the performance of our approach. In this way, we want to gain insights concerning the suitability of TBRAs for different application scenarios. In other words, we study to what extent the goals of feasible and distance-efficient routes can be achieved with a TGRA under time window constraints and how well the implicit realization of consistency benefits (without directly optimizing consistency in the routing but only due to the restriction to STs) is achieved. Moreover, we want to gain insights concerning the design requirements of a TGRA to be successfully utilized for time-definite vehicle routing.

First, we explain how to assess the solution quality of MTR in terms of routing efficiency, consistency and feasibility of the produced results (Section 6.1). Subsequently, the parameter setting of MTR is described in Section 6.2. To prove the quality of the developed TGRA, we compare its performance to existing approaches on available multi-day VRP benchmarks without time windows (Section 6.3). In Section 6.4, we generate a comprehensive set of multi-day benchmarks with time windows and use them to investigate the influence of the following method parameters.
on the performance of MTR: i) the number of sample days considered in the districting phase, ii) the specific time window treatment in the advanced method for determining the exclusion zone, seed customers and customers to assign to STs, iii) the solution quality of the sample-day VRPTW method and iv) the size of the STs. Moreover, we study the influence of the following problem parameters: i) geographical distribution of customers, ii) existence, density, width and spatial distribution of time windows, and iii) variability of the number of customers to be served each day.

6.1 Assessment of the Solution Quality of MTR

MTR is designed to address multi-day series of routing problems as otherwise no statements about delivery consistency are possible. Another requirement for consistency evaluation is that the customers for each day stem from the same base set. This mimics the practical situation, where each day a (more or less) random subset of the SPS company’s customer base requires service. Thus, adequate benchmark sets consist of $\tau_1 + \tau_2$ single-day VRP problems, where the set of customers $V_t$ that requires service on day $t = 1, \ldots, \tau_1 + \tau_2$ is a random subset of the customers of a larger base problem with customer set $V_{\text{total}}$. Throughout the experimental analysis, we separate the generated single-day problems in two groups. The data of the first group consists of the first $\tau_1$ sample days and is used as input to the districting phase. The remaining $\tau_2$ evaluation days are used to assess the performance of the STs generated in the districting phase.

The performance of MTR is assessed with respect to efficiency, feasibility and consistency measures. For the solutions obtained by MTR during the $\tau_2$ evaluation days, the following efficiency and feasibility measures are calculated i) the average number of employed vehicles ($NV$), ii) the average traveled distance ($TD$), iii) the total number of days with an invalid solution before outsourcing customers ($\text{Inv}$), distinguished in time window infeasible ($\text{Inv}_{\text{tw}}$) and capacity infeasible ($\text{Inv}_{\text{cap}}$), and iv) the total number of customers that have to be outsourced to achieve a feasible solution on all $\tau_2$ days ($\text{OC}$). The last measure is inspired by practice, where unprofitable or difficult-to-serve customers are subcontracted (Stenger et al. 2013). We use a straightforward approach to determine the customers to outsource if a solution is infeasible. In each iteration, we find the customer that causes the largest time window and capacity violation and remove it until we end up with a feasible solution.

In order to quantify route consistency, we use the following consistency measures, which are similar to the measures introduced in Smilowitz et al. (2013).

- **Customer familiarity** ($\text{CF}$) for a certain customer is defined as the percentage of deliveries to that customer during the $\tau_2$ evaluation days which are carried out by the one driver who visits the customer most frequently. By averaging over customers, we get the CF measure of the solution of a problem instance.

- **Driver diversity** ($\text{DD}$) is the number of different drivers serving a customer during the $\tau_2$ evaluation days. The average over all customers represents the DD measure of a solution. It should be noted that these two measures have an inverse behavior, i.e., a high customer familiarity value and respectively a low driver diversity value indicate route consistency.

As comparison method for MTR, we use a strategy of daily route reoptimization (RR)
carried out with TS$_{hq}$, where vehicle routes for the $\tau_2$ evaluation days are determined without considering any fixed assignment of customers to drivers. To ensure a fair comparison of the consistency results obtained by MTR and RR, we apply a post-processing step to the RR solution as commonly done in the literature (cp. Zhong et al. 2007, Groër et al. 2009, Sungur et al. 2010, Smilowitz et al. 2013). Based on the driver-route assignment of the first $t-1$ evaluation days, we determine the assignment for day $t$ that maximizes the sum of consistency rewards of day $t$. The consistency reward of a driver-customer assignment is defined by the number of previous deliveries that the customer received from the assigned driver. The consistency reward of a driver-route assignment is given by the sum of the driver-customer rewards of all customers contained in the route. Thus, we model this step as an assignment problem which aims at finding the optimal driver-route assignment for day $t$ in terms of maximizing consistency rewards and solve the problem using CPLEX 12.1.

6.2 Parameter Setting and Testing Environment

The parameter setting of the sample-day VRPTW methods, TS$_{hq}$ and TS$_{mq}$, are described in Appendix A and B. The parameter values of TS$_{hq}$ are also used for the territory-based VRPTW method TS$_{terr}$.

During our extensive testing, we found the following base parameter setting for the districting phase. The number of vehicles $m_{st}$ and thus the number of STs to be created corresponds to the number of vehicles that the sample-day VRPTW method requires to find a feasible solution for $\theta = 95\%$ of the $\tau_1$ sample days, i.e., the number of vehicles is set to achieve the given service level over the sample days when RR is used. The STs are created with $\rho = 0.6$, so as to include 60% of the customers in set $V_{total}$, which we denote as medium-sized STs. We found that the ratio $\omega$ of the total customers $V_{total}$ to be included in the exclusion zone has to be chosen in dependence of the ST size in order to guarantee that enough “free” customers exist in between STs. For medium-sized STs, we use $\omega = 0.05$. The maximum distance $d_{max}$ for the adjacency graph is set to the average distance between customers in the respective base instance divided by four. Furthermore, we use a distance exponent equal to $\psi = 2$.

MTR is implemented as single-thread code in Java. Due to the considerable computing power needed for our computational studies, we conducted all tests on the high-performance cluster Elwetritsch of the University of Kaiserslautern.

6.3 Performance of MTR on Related Multi-Day VRP Instances without Time Windows

To prove the competitiveness of MTR compared to other approaches presented in the literature, we conduct tests on the multi-day VRP instances without time windows suggested by Groër et al. (2009). The instances are based on the 12 CVRP instances of Christofides and Eilon (1969), which contain between 50 to 199 customers.

To address the test instances with MTR, we generate five additional sample days for each instance using the approach described in Groër et al. (2009) and use them as input for the districting phase. Based on the generated STs, the routing phase is conducted on the five days of the original benchmark instances and the obtained routing solutions are evaluated. Problems
6–10 are distance-constrained and this is handled in MTR by setting appropriate time windows for the depot. The results of MTR and our RR approach are compared to the methods of Groër et al. (2009) (ConRTR as consistent routing approach and RTR as RR approach) and Sungur et al. (2010) (MADS as consistent routing approach). As all solutions have to be feasible in order to allow for a fair comparison, we adopt TS$_{terr}$ to open further routes until a feasible solution is found.

For the solution approaches described above, we report the averages of i) the total travel time TTT, and ii) the number of vehicles NV (not available for MADS) over the 12 test instances in Table 1. Detailed results on an instance basis are provided in Appendix D.1. A direct comparison is only possible for the RR approaches due to the different objectives of the consistent routing approaches ConRTR, MADS and MTR. ConRTR aims at minimizing the difference in the daily arrival times and produces solutions where each customer is visited by the same driver whenever service is required, i.e., the driver diversity of the solutions is one and the customer familiarity is 100%. While the approach of Sungur et al. (2010) is adapted to provide results with the same characteristics, this is not possible for MTR. MTR is neither designed to provide 100% CF nor to provide solutions with small deviations of the arrival times at the customers. Such objectives are better pursued by approaches based on master schedules which are adapted on a daily basis. Therefore, the efficiency measures NV and TTT are not directly comparable as MTR does not guarantee a 100% CF. However, to be able to assess the tradeoff between efficiency and consistency achieved by MTR solutions in comparison to ConRTR and MADS, we report the CF of MTR.

<table>
<thead>
<tr>
<th></th>
<th>ConRTR</th>
<th>MADS</th>
<th>MTR</th>
<th>RTR</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>NV</td>
<td>10.50</td>
<td>6.090.44</td>
<td>7.90</td>
<td>7.77</td>
<td>7.65</td>
</tr>
<tr>
<td>TTT</td>
<td>6,115.42</td>
<td>5,687.28</td>
<td>5,372.66</td>
<td>5,393.54</td>
<td></td>
</tr>
<tr>
<td>TTT</td>
<td>6,090.44</td>
<td>5,687.28</td>
<td>5,372.66</td>
<td>5,393.54</td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>85.63%</td>
<td>85.63%</td>
<td>86.83%</td>
<td>86.83%</td>
<td>86.83%</td>
</tr>
<tr>
<td>NV</td>
<td>7.65</td>
<td>7.65</td>
<td>7.65</td>
<td>7.65</td>
<td>7.65</td>
</tr>
<tr>
<td>TTT</td>
<td>5,393.54</td>
<td>5,393.54</td>
<td>5,393.54</td>
<td>5,393.54</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** Comparison of MTR and RR to the methods of Groër et al. (2009) (RTR as route reoptimization and ConRTR as consistent routing approach) and Sungur et al. (2010) (MADS as consistent routing approach) on the multi-day VRP instances of Groër et al. (2009).

Comparing the results of RTR and our RR, we find that our RR achieves solutions with a slightly lower NV while the TTT increases marginally. This can be explained by the hierarchical objective function of VRPTW that is used in our RR method. The solution quality of our RR can thus be said to be on par with that of RTR, which is a very strong result considering that our RR was specifically developed and parameter tuned for VRPTW.

Although MTR results are not directly comparable to ConRTR and MADS results as described above, the MTR solutions are clearly superior concerning the efficiency measures, reducing the NV value of ConRTR by 24.8% and the TTT value of ConRTR and MADS by 7.00% and 6.6% respectively, while still achieving an average CF of nearly 86%. Thus, MTR is capable of strongly reducing the NV and TD values while achieving high route consistency. These results indicate a very good solution quality of MTR and its ability to produce meaningful results in the subsequent studies.
6.4 Performance of MTR on Multi-day Instances with Time Windows

This section analyzes the application of MTR to multi-day instances with time windows. As no adequate benchmark set is available from the literature, we follow the approach of Groër et al. (2009) and use well-known VRPTW instances as a basis to generate a multi-day series of problems (Section 6.4.1). On these instances, we study the influence of method and problem parameters on the performance of MTR in Sections 6.4.2 and 6.4.3.

6.4.1 Generation of Test Problems

To study the performance of MTR on time-definite routing problems, we require multi-day series of VRPTW single-day problems in order to be able to investigate delivery consistency. Thus, we generate new multi-day benchmark problems based on the well-known VRPTW benchmark problems proposed by Gehring and Homberger (1999). More precisely, we use the Gehring and Homberger 1000-customer instances as base problems for generating multi-day series of VRPTW problems.

The Gehring and Homberger benchmarks are generated in a manner similar to the well-known VRPTW instances of Solomon (1987). They consist of five sets with customer numbers of 200, 400, 600, 800 and 1000 respectively. Each set contains 60 instances. Within each set six problem groups exist involving instances with a clustered (C), random (R) and random-clustered (RC) customer distribution. The instance groups differ with regard to the scheduling horizon of the depot \( l_0 - e_0 \) and the capacities of vehicles. Groups C1, R1 and RC1 have relatively short scheduling horizons and low-capacity vehicles, generally resulting in solutions with a high number of short routes containing only few customers. Groups C2, R2, RC2 have considerably longer scheduling horizons and high-capacity vehicles, resulting in solutions with a comparatively low vehicle number performing longer routes containing a higher number of customers.

The instances within a group differ in terms of time window density (TWD), i.e., the percentage of customers with a time window (25%, 50%, 75%, 100%), and time window width (TWW), defined as the average width of \((l_v - e_v)\) across all customers in the problem instance. The detailed TWDs and TWWs of the Gehring and Homberger 1000-customer instances plus additional information on the benchmark are provided in Appendix C.3.

Two aspects should be noted in the choice of the base problem instances. First, as the Gehring and Homberger benchmark contains instances with 75% and 100% TWD, we clearly go beyond the TWD characteristics of real-world SPS companies, which face up to 60% of time-definite deliveries (Campbell and Thomas 2009). Second, the instances are very hard to solve and the very restrictive time windows play a major role in this respect. This is witnessed by the regular update of the best-known solutions to these instances (see, e.g. Nagata et al. 2010, Blocho and Czech 2012, Vidal et al. 2013).

By drawing customers from the base instances, we generate for each of the 60 base instances a series consisting of \( \tau_1 + \tau_2 \) days of VRPTW problems, where \( \tau_1 \) sample days are used as input for the districting phase and daily routes are generated for \( \tau_2 \) evaluation days. To allow for a certain degree of demand variation, not only the choice of customers is randomized, but the number of selected customers \(|V_t|\) for day \( t \) is also a stochastic variable. We assume \(|V_t|\) to be normally distributed with an expected value of \( \mu = 120 \) customers and a standard deviation
Table 2: Influence of the number of sample days on the performance of MTR\textsubscript{basic}.

<table>
<thead>
<tr>
<th>$\tau_1$</th>
<th>NV</th>
<th>TD</th>
<th>DD</th>
<th>CF</th>
<th>Inv</th>
<th>OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11.18</td>
<td>10,574.35</td>
<td>2.11</td>
<td>78.33%</td>
<td>5.25</td>
<td>20.27</td>
</tr>
<tr>
<td>50</td>
<td>11.37</td>
<td>10,490.04</td>
<td>2.11</td>
<td>78.06%</td>
<td>2.62</td>
<td>8.78</td>
</tr>
<tr>
<td>100</td>
<td>11.32</td>
<td>10,494.06</td>
<td>2.13</td>
<td>77.73%</td>
<td>2.50</td>
<td>8.72</td>
</tr>
</tbody>
</table>

of $\sigma = 10$. Moreover, we make the assumption that each customer has the same probability of requiring service on a given day. This is the worst case scenario for a TBRA and we did not want to have a distribution that prejudices the generated instances towards problems which are more easily tractable by TBRA\textsc{s}. Note however that such distributions, with several core customers that have high order volumes and frequencies, may well occur in practice.

6.4.2 Influence of Method Parameters and Components

We use the above described multi-day instances with time windows to perform the following tests on the sensitivity of MTR concerning modifications in algorithmic parameters and components. The base parameter setting of the distance phase described in Section 6.2 is used for all tests that do not state otherwise. Moreover, the reported results are averages over all 60 benchmark instances as long as not stated differently.

6.4.2.1 Influence of the Number of Considered Sample Days

In this test, we study the influence of the number of sample days considered in the districting phase on the performance of MTR. To this end, we generate $\tau_1 = 100$ sample days and use the first 10, 50, and 100 days as input for the districting phase and compare the results using the same $\tau_2 = 50$ evaluation days for the routing phase. We report the results for MTR\textsubscript{basic}, the variant of MTR that is composed of the basic methods for determining the exclusion zone $EZ_{\text{basic}}$, the seed customers $Seed_{\text{basic}}$, and for assigning the customers $Assign_{\text{basic}}$. This is motivated by the fact, that MTR\textsubscript{basic} depends strongly on the solutions of the sample days as this is the only way in which time window requirements are considered. In Table 2, results concerning the measures described in Section 6.1 are given.

The table shows that using the results of 10 sample days in the districting phase is not sufficient to create adequate STs. Using few sample days results in an insufficient number of employed vehicles (and thus generated STs), which leads to a high number of days with invalid solutions and thus a high number of customers to outsource (OC) in order to obtain feasible solutions. By contrast, using 50 sample days allows MTR to create well formed STs resulting in an efficient daily routing. Significantly fewer invalid solutions are obtained and the number of customers that have to be outsourced to obtain a feasible solution is also less than half of the value obtained for 10 sample days. Meanwhile, the consistency values are approximately on the same level for both numbers of sample days. Moreover, we find that increasing the number of sample days to 100 does not significantly improve the solution quality of MTR\textsubscript{basic}. Consequently, for all of the following tests, we use $\tau_1 = 50$ days as input for the districting phase and $\tau_2 = 50$ days for the routing phase.
6.4.2.2 Influence of Time-Window Handling Components

In this section, we study the impact of incorporating time window aspects in the steps of the districting phase. To this end, we compare MTR\textsubscript{basic} to MTR\textsubscript{tw}, which utilizes \textit{EZ\textsubscript{tw}}, \textit{Seed\textsubscript{tw}} and \textit{Assign\textsubscript{tw}}. As further comparison methods, we use i) MTR\textsubscript{geo}, which uses only spatial information in the districting phase, i.e., the assignment of customers to STs is carried out without considering the history-based insertion costs, ii) MTR\textsubscript{history}, in which the customer assignment is purely based on the insertion cost $b_{vw}$ and no additional geographical or time window characteristics are considered, and iii) MTR\textsubscript{twOnly}, that bases the customer assignment only on time window characteristics and disregards any historical or geographical information. Table 3 shows the results of the five methods for the measures described in Section 6.1.

<table>
<thead>
<tr>
<th>Method</th>
<th>NV</th>
<th>TD</th>
<th>CF</th>
<th>DD</th>
<th>Inv</th>
<th>Inv\textsubscript{cap}</th>
<th>Inv\textsubscript{tw}</th>
<th>OC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTR\textsubscript{basic}</td>
<td>11.37</td>
<td>10490.04</td>
<td>78.06%</td>
<td>2.11</td>
<td>2.62</td>
<td>0.32</td>
<td>2.42</td>
<td>8.78</td>
</tr>
<tr>
<td>MTR\textsubscript{tw}</td>
<td>11.36</td>
<td>10562.43</td>
<td>77.90%</td>
<td>2.12</td>
<td>2.55</td>
<td>0.32</td>
<td>2.33</td>
<td>8.27</td>
</tr>
<tr>
<td>MTR\textsubscript{geo}</td>
<td>11.36</td>
<td>10511.38</td>
<td>77.62%</td>
<td>2.13</td>
<td>2.45</td>
<td>0.25</td>
<td>2.25</td>
<td>8.05</td>
</tr>
<tr>
<td>MTR\textsubscript{history}</td>
<td>11.72</td>
<td>15975.87</td>
<td>72.89%</td>
<td>2.53</td>
<td>36.27</td>
<td>4.05</td>
<td>36.15</td>
<td>506.27</td>
</tr>
<tr>
<td>MTR\textsubscript{twOnly}</td>
<td>11.69</td>
<td>16167.21</td>
<td>66.33%</td>
<td>2.90</td>
<td>33.70</td>
<td>3.42</td>
<td>33.57</td>
<td>512.88</td>
</tr>
</tbody>
</table>

Table 3: Influence of time window handling components

Comparing the results obtained with MTR\textsubscript{basic}, MTR\textsubscript{tw} and MTR\textsubscript{geo}, we gain a major insight in the design of STs when dealing with time definite customer deliveries. The results of the three methods are approximately of the same quality with respect to efficiency, feasibility and consistency. To be more precise, even when time windows are present, a customer assignment only based on geographical aspects leads to similar results as MTR\textsubscript{basic}, which additionally integrates historical information, or as MTR\textsubscript{tw}, which additionally uses historical information and time window characteristics. This indicates that the consideration of geographical data is paramount for the design of high-quality STs even in the presence of (restrictive) time window requirements.

At first glance, one might speculate that this result indicates inadequate benchmark instances which are rather straightforward to solve in the routing phase independent of the quality of the STs generated in the districting phase. We have already observed in Section 6.4.1 that this is very unlikely for instances based on the notoriously difficult Gehring and Homberger benchmark. However, to further prove our point we have conducted the tests with MTR\textsubscript{history} and MTR\textsubscript{twOnly}, which base only on historical and time window characteristics and neglect geographical information. The quality of the results obtained with these methods are much worse, characterized by very strong increases in traveled distance and infeasible solutions, while also achieving significantly lower consistency.

To summarize, our results indicate that a TBRA for time-definite routing has to rely strongly on geographical characteristics to achieve good solution quality. While historical data is clearly important to determine a reasonable number of STs to generate, its influence for generating high-quality STs seems less significant. This also seems true for the time window characteristics of the customers included in the STs. However, as shown in Section 6.4.2.1, if historical data is considered in the districting phase, an adequate amount of information has to be used. Otherwise the process is misguided and the solution quality can even suffer from the inclusion of this
additional information.

Finally, note the number of time window invalid solutions in the study is an order of magnitude higher than the number of capacity infeasible solutions. This confirms that the instances are restrictive concerning time requirements.

6.4.2.3 Influence of Solution Quality of Sample-Day VRPTW Method and Territory-Based VRPTW Method

To study the influence of the solution quality of the solution methods used as sample-day VRPTW method and as territory-based VRPTW method, we use the two TS methods described in Appendix A and B. As described above, TS$_{hq}$ is quite competitive, whereas TS$_{mq}$ achieves only moderate solution quality. We use MTR$_{basic}$ to study the following combinations of the two TS as sample-day VRPTW method in the districting phase and territory-based VRPTW method in the routing phase: TS$_{hq}$/TS$_{hq}$, TS$_{mq}$/TS$_{mq}$, TS$_{mq}$/TS$_{hq}$.

When using TS$_{mq}$ as territory-based VRPTW method, we employ a different mechanism for freeing STs, namely, an ST can only be emptied if no customer of that ST requires service on the given day. To make the results comparable, we adapted our TS$_{hq}$ in the same fashion for this test. This is the reason why the results obtained with the combination TS$_{hq}$/TS$_{hq}$ are slightly different from those reported for MTR$_{basic}$ in Table 3.

The results are reported in Table 4 and show that the overall solution quality of MTR seems to depend very strongly on the quality of the solution method used in the routing phase. If STs are generated based on the sample day solution evaluated by TS$_{mq}$, using TS$_{hq}$ as routing methods achieves approximately the same solution quality as using TS$_{hq}$ for both phases. Contrary to this, using TS$_{mq}$ in both phases leads to a significantly inferior solution quality concerning efficiency, feasibility and consistency measures. These results also indicate that the quality of the sample-day VRPTW method has only moderate impact on the overall quality of MTR; a result that is perfectly compatible with the findings of Section 6.4.2.2.

The results concerning efficiency and consistency values are displayed in Figure 5(a) and the
results concerning feasibility measures in Figure 5(b). To better highlight the tradeoff between consistency, efficiency and feasibility, the figures are normalized by using the value achieved with medium-sized STs as 100% level. A table listing the complete results for MTR$_{tw}$, MTR$_{basic}$ and MTR$_{geo}$ can be found in Appendix D.2.

Comparing the results of RR and MTR$_{tw}$, we find that vehicle routes designed with MTR$_{tw}$ are generally characterized by a deterioration of efficiency measures, i.e., higher number of vehicles, traveled distances, number of infeasible days and number of customers to outsource in order to achieve feasible solutions over the evaluation days. On the other hand, MTR$_{tw}$ achieves considerably higher consistency, namely an increase in customer familiarity and a decrease in driver diversity.

Increasing the size of the STs leads to a decrease in the flexibility of MTR$_{tw}$ and thus to a deterioration of efficiency and feasibility measures together with an improvement of the consistency measures. An important point can be noted: The changes in the efficiency and consistency measures behave approximately proportionally to the increase of the customer percentage permanently assigned to drivers. This suggests that the size of the STs can be utilized to control the achieved tradeoff between service consistency and route efficiency.

Moreover, we can see that from a percentage perspective the consistency improvement achieved by MTR$_{tw}$ seems to outweigh the deterioration of efficiency measures. Clearly, the optimal tradeoff can only be determined if a cost value can be put on each of the considered measures. However, taking into account that daily customer selection is based on a uniform distribution and the instances have partially very high time window densities, the results indicate only a moderate deterioration in efficiency measures of MTR$_{tw}$. On the other hand, the implicit consideration of driver familiarity enables MTR$_{tw}$ to improve consistency results significantly.

Furthermore, we note that while the increase in the number of days with invalid solutions seems rather moderate for small and medium-sized STs, large STs show a steeper increase. However, even with large STs, less than 4 out of the $\tau_2 = 50$ evaluation days are not feasibly solvable and outsourcing less than 13 customers in total is sufficient to restore feasibility (see Appendix D.2).

Thus, with the real-world possibility of subcontracting, which reduces the risk of having unsatisfied customers, the advantages of a TBRA might outweigh the efficiency forfeits even in the presence of time window constraints and an unfavorable distribution of customer occurrence.
probabilities. To further assess the performance of MTR in comparison to an RR approach, we conducted a worst case analysis, which is reported in Appendix D.3.

6.4.3 Influence of Problem Parameters

The following tests investigate the influence of changes in the problem parameters on MTR. All tests are carried out with medium-sized STs and $\tau_1 = 50$ sample days are used for the districting phase. We conducted all tests with MTR$_{tw}$, MTR$_{basic}$ and MTR$_{geo}$, but since the methods showed a very similar behavior in all studies, we restrict ourselves to the presentation of MTR$_{tw}$. Again, results are presented as averages over instance sets. Tables listing the complete results for all three methods can be found in Appendix D.

6.4.3.1 Influence of Geographical Customer Distribution

To analyze the influence of the geographical distribution of customers, we present the results of MTR$_{tw}$ for the test instances described above, categorized by the customer distribution classes R, C, RC. Since an interpretation of absolute values is not meaningful due to the strong structural differences of the instances of the distribution classes, we report the gaps to the results obtained with our RR approach for each class.

The results concerning efficiency and consistency values are displayed in Figure 6(a), the feasibility measures in Figure 6(b).

![Figure 6: Influence of geographical customer distribution (R, C and RC) on (a) efficiency and consistency measures and (b) feasibility measures of MTR$_{tw}$ as gaps to the RR approach. For presentation purposes we reverse the sign of the gaps of the driver diversity values.](image)

The test reveals that MTR works best on instances with a random customer distribution, achieving the highest consistency improvements and the lowest efficiency and feasibility forfeits compared to the RR solution. The clustered instances show the highest losses in efficiency and the number of invalid solutions increases significantly. This is a very interesting insight as one would generally expect that clustered customer instances are in favor of a TBRA. However, the existence of time window requirements and/or a mismatch in the size of the geographical clusters and the STs might explain this results. The results for the RC instances lie between the R and C results for the efficiency and feasibility measures, while the consistency improvements on RC are the weakest. This might however be explained by the already high consistency level of the RR solutions on RC as shown in Appendix D.4.
6.4.3.2 Influence of Time Windows

We analyze the effect of time window existence on the performance of MTR by comparing the results obtained on the test instances introduced above with the results of MTR on the same instances after removing all customer time windows. We report for both test sets the gaps of MTR to the results obtained with our RR approach. For the instances with time windows MTR\textsubscript{tw} is used, for those without we used MTR\textsubscript{basic}. The results concerning efficiency and consistency values are displayed in Figure 7(a), the feasibility measures in Figure 7(b). The complete results can be found in Appendix D.5.

![Figure 7: Influence of time window existence on (a) efficiency and consistency and (b) feasibility measured of MTR\textsubscript{tw} as gaps to the RR result](image)

The results show that the gaps of the efficiency measures between MTR and RR clearly increase when time windows are present. The same is true for the feasibility measures. For the instances without time windows, MTR is basically able to hold the same feasibility level as the RR approach, while this is not possible for MTR if time windows are present. On the other hand, MTR achieves a drastically higher improvement of consistency measures if time windows are present. This can be explained by the fact that RR solutions to problems without time windows offer higher consistency than solution for time window instances (see Appendix D.5). Nevertheless, the fact that RR solutions to time-definite problems show lower consistency and thus high potential for improvement by using a TBRA is an important insight for practitioners, e.g., in the SPS industry.

To get a more differentiated picture of the problem characteristics producing the above results, we further study the influence of time window density (TWD) and time window width (TWW) on the performance of MTR. To evaluate the influence of different TWDs, we generate test instances as follows. Since we want to eliminate any influences caused by different customer distribution, we use the Gehring and Homberger instances from classes R1 and R2 with 100% time window densities, namely R1\textsubscript{10}1, R1\textsubscript{10}5, R1\textsubscript{10}9, R1\textsubscript{10}10, R2\textsubscript{10}1, R2\textsubscript{10}5, R2\textsubscript{10}9 and R2\textsubscript{10}10. From each of these 8 instances, we generate 5 instances with a TWD of 25% (50%, 75% respectively) by randomly removing 75% (50%, 25% respectively) of the time windows. Moreover, we generate instances with a TWD of 0% by removing all time windows. Based on the existence of time windows in these newly generated base instances, we remove the time windows from the original multi-day series associated with the respective base instance in order to obtain the final test instances.

As the number of vehicles employed in a solution strongly influences the traveled distance, feasibility and consistency values, we decided to set the number of STs to generate for all
instances to the values of the solutions obtained on the instances with 100% TWD. In this way, we provide a better comparability of the results.

The results concerning efficiency and consistency values are displayed in Figure 8(a), where NV is not reported due to the fixing of the vehicle number described above. Feasibility results are shown Figure 8(b). Again, the figures are normalized using the value achieved for the 100%-TWD instances as 100% level.

![Figure 8](a) Influence of time window density (a) efficiency and consistency measures and (b) feasibility measures of MTRtw

The efficiency and consistency measures behave as could be expected: an increase of TWD leads to a deterioration of the measures due to the growing hardness caused by the time windows that prohibits the design of distance-efficient and consistent routes. Concerning the feasibility measures, the number of invalid solutions and the number of customers to be outsourced remain relatively stable up to a TWD of 75% and then increase significantly for a TWD of 100%. This is caused by a rise in time-window infeasible solutions for a TWD of 100% while capacity infeasible solutions remain on a relatively constant level for all TWDs.

To analyze the effect of the tightness of the time windows, we use the four Gehring and Homberger R1 instances with a 100% TWD as initial instances. Again, we want to erase influences caused by different customer distributions and we restrict ourselves to the instances of class R1 as these have tight time windows in comparison to instances in class R2. From this set of instances, we generate four new sets of base instances by multiplying the TWW with a factor of 2, 3, 5 and 10. Based on the TWWs in these newly generated base instances, we adjust the time windows of the original multi-day series associated with the respective base instance in order to obtain the final test instances. The number of STs to generate is set to the value of the solutions of the original instances. Figure 9(a) shows the efficiency and consistency measures in dependency of the TWW, Figure 9(b) the feasibility measures. The complete results of this study can be found in Appendix D.5.

Similarly to a decrease in TWD, an increase in TWW leads to an improvement of efficiency and consistency measures as less restrictive time windows enhance the flexibility to create efficient and consistent routes. An increase of the TWW by factor two leads to a significant decrease of invalid solutions as it is already enough to render time-window infeasible solutions close to zero. Interestingly, although it is possibly to always generate time-window feasible solutions already with a TWW factor of 3, further enhancing the factor leads to additional improvements in efficiency and consistency.

Summarizing, one can see that an increase of TWD and a decrease of TWW both play a
significant part in the deterioration of efficiency, consistency and feasibility measures caused by the existence of time windows.

6.4.3.3 Influence of the Spatial Distribution of Time Windows

We now test the influence of the spatial distribution of time windows: Customers with time-definite requirements either occur in clusters or they are uniformly distributed. In the Gehring and Homberger instances, the customers with time windows are uniformly distributed and thus, we generate new instances to cover clustered time windows aspects as follows. As base instance, we use $R1_{10,1}$, which is the most restrictive R instance concerning time windows (100% TWD and TWW of 10.00). From this instance, we generate a set of 30 instances, 20 with a TWD of 50% and 10 with a TWD of 75%.

To generate an instance, we use a sweep approach starting with a ray leaving the depot at an arbitrary angle and separate the delivery area into four sectors that contain the same number of customers. The second and fourth sector keep their time windows, whereas the time windows in the other sectors are removed. By rotating the ray in steps of 9 degrees, we are able to generate an additional 9 instances. The same procedure is applied to generate another 10 instances, where the delivery area is separated into 8 equal sectors, where every second one keeps its time window. Another ten instances with a TWD of 75% are generated by randomly drawing 6 out of the 8 parts and keeping the time windows. In the comparison set, time windows are randomly distributed, i.e., we generate 20 instances by randomly removing 50% of the time windows in the base instance and 10 instances by removing 25%. In both sets, we use the 30 instances as base for generating a series of 100 single-day problems, which are the final test instances for MTR$_{tw}$ and RR.

Solutions of MTR always trade-off efficiency, consistency and feasibility considerations, which makes a comparison of the results for the different instance sets difficult. Therefore, we decided to keep solutions mostly feasible for this test by using the maximal number of STs required for a test instance as initial number of STs for all other instances. This initial number can clearly be reduced by freeing STs, however, this only happens if a feasible solution is found beforehand. In this way, only a marginal percentage of daily solutions is infeasible, which allows to omit feasibility measures from the analysis and provide a fair comparison of the efficiency and consistency measures of both instance sets.

Results are averaged over all 30 instances. Figure 10 depicts the efficiency and consistency
measures as gaps with respect to the RR results.

![Figure 10: Influence of the spatial distribution of time windows on the efficiency and consistency measures of MTR\textsubscript{tw}](image)

The results contradict the expectation that instances with clustered time windows are harder to solve using a TBRA (especially if very restrictive time windows are present as it is the case here). By contrast, the solution quality of MTR\textsubscript{tw} is higher for the clustered instances concerning all efficiency and consistency measures. This might be explained by the fact that not all customers within a time window cluster require service simultaneously. Moreover, STs located in a way that they include clusters with time windows and clusters without time windows might gain the necessary flexibility to fulfill the time window requirements by also serving a sufficient percentage of easy-to-schedule customers.

### 6.4.3.4 Influence of Variability of the Number of Customers Requiring Service

Last, we study the influence of the variability of the number of customers requiring service on each day on the performance of MTR in comparison to an RR approach. In addition to our 60 base instances, which are generated with a standard deviation of $\sigma = 10$ for the number of customers to be served on each day, we generate two more sets using standard deviations of 20 and 30. In order to achieve comparable results, we fixed the number of STs to be generated to the value used for the instances with $\sigma = 10$. However, given the possibility of MTR to free STs, this does not mean that the number of vehicles employed in the solution has to be equal. Figures 11(a) and 11(b) depict the gap with respect to the RR solutions for the efficiency, consistency and feasibility measures.

In principle, a higher variability in the number of customers requiring service is expected to have a negative influence on the performance of a TBRA. However, increasing the standard deviation seems to have no negative effect on MTR\textsubscript{tw}. The gaps for the efficiency and consistency are almost identical for the different standard deviations. The same holds for the absolute values of the measures as Appendix D.6 shows. For the feasibility measures, we notice that while the absolute values of Inv and OC clearly increase for higher standard deviations, the gap to the RR solution is maximal for $\sigma = 10$ and decreases with higher standard deviations. These results show that variability in the number of customers to be served does not have the detrimental effect on the performance of a TBRA that one could expect. Instead, MTR\textsubscript{tw} seems to find a very good tradeoff between routing efficiency and consistency in such settings.
7 Conclusion and Outlook

We develop Modular Territory Routing (MTR) to study the performance of territory-based routing approaches (TBRAs) in the presence of time window requirements. In the districting phase, service territories (STs) are constructed based on spatial, temporal and historical information by means of a modular approach, selecting a set of seed customers and iteratively adding further customers to the seeds until STs of the desired size are created. In the subsequent routing phase, the daily routing is conducted based on the generated STs. MTR is able to obtain very competitive results compared to approaches presented in the literature on multi-day VRP test instances without time windows. Studies on an extensive set of newly generated multi-day VRPTW benchmark instances are used to investigate the design requirements of a TBRA suitable for handling time windows and the influence of the time requirements on the performance of such an approach.

The main insights of our studies are:

- The consideration of geographical aspects in the districting is paramount for generating high-quality STs, while explicitly incorporating time window characteristics and historical demand data in ST design does not lead to a perceptible improvement of the solution quality.
- Increasing the size of STs leads to approximately proportional changes in the efficiency and consistency measures, which suggests that these parameter can be utilized to control the achieved tradeoff between driver consistency and route efficiency.
- The efficiency and feasibility forfeits of MTR in comparison to daily route reoptimization (RR) are higher if (restrictive) time windows are present. On the other hand, significantly stronger consistency improvements are achieved. This is due to the fact that RR solutions to time-definite problems show lower consistency and thus higher potential for improvement by using a TBRA, which constitutes an important insight for practitioners.
- Increasing the variance of the number of customers requiring service on each day does not have the detrimental effect that one could expect. Thus, even with high variations of the number of customers requiring service, an unfavorable uniform distribution of customers and high time window densities, MTR shows only moderate efficiency and feasibility forfeits while achieving significantly higher routing consistency compared to an RR strategy.
Such routing consistency is highly valued by SPS companies.

Several interesting topics for future research arise with regard to MTR. First, we assume that each customer has the same probability of requiring service on a given day in order to generate the worst case scenario for a TBRA. However, the performance of MTR shall also be investigated on benchmark instances with more realistic customer distributions, where a set of core customers have higher order volumes and occurrence frequencies than standard customers. Second, the application of MTR on real-life data of an SPS company and comparison to their route planning method would allow valuable conclusions. Third, investigating the impact of further practical routing constraints such as, e.g., precedence constraints in pickup and delivery operations on the performance of a TBRA is another interesting topic for future research.

Acknowledgements The authors would like to thank Henning Lagemann for his help in coding an earlier version of the algorithm. Moreover, the authors gratefully thank three anonymous referees and the associate editor for their valuable comments.

References


