The vehicle routing problem with zone tariffs: a matheuristic framework and insights into detour limit design

Niklas Tuma, Manuel Ostermeier, Alexander Hübner

Angaben zur Veröffentlichung / Publication details:

Tuma, Niklas, Manuel Ostermeier, and Alexander Hübner. 2026. "The vehicle routing problem with zone tariffs: a matheuristic framework and insights into detour limit design." *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2026.07.012.

ARTICLE IN PRESS

European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

The vehicle routing problem with zone tariffs: A matheuristic framework and insights into detour limit design

Niklas Tuma [a], Manuel Ostermeier [b,*], Alexander Hübner [a]

[a] Technical University of Munich, Supply and Value Chain Management, 94315 Straubing, Germany
[b] University of Augsburg, Resilient Operations, 86159 Augsburg, Germany

ARTICLE INFO

Keywords:
Logistic services
Detour limit
Retail operations
Adaptive large neighborhood search
Real-world application

ABSTRACT

Companies commonly outsource transportation logistics to logistics service providers (LSPs) to reduce complexity and leverage economies of scale. The coordination between shippers and LSPs for regional distribution is typically organized through zone-based tariffs, in which shippers propose joint customer deliveries to an LSP. This approach decouples customer allocation and pricing from the delivery sequence, simplifying planning. To minimize costs, shippers must form customer clusters that meet capacity and specific distance constraints, known as detour limits. These detour limits ensure reasonable driving distances within a customer cluster. Despite their widespread application in practice, zone tariffs have received little research attention to date. Early related research either addresses dedicated case applications with a specific single detour limit or relies on exact methods suited only for small instances. In contrast, general real-world applications involve hundreds of customers and heterogeneous detour limit options.

We address this gap by tailoring an Adaptive Large Neighborhood Search–based matheuristic that features problem-specific operators and two adaptive mechanisms. Combining heuristic tour generation with a set partitioning model, our approach efficiently handles multiple detour limits. Numerical studies, including a case study and benchmark instances with up to 1000 customers, demonstrate the approach's scalability and effectiveness. This study provides the first comprehensive analysis of how different detour limitations relate and affect routing efficiency as well as zone tariff design, revealing key trade-offs between cost and travel distance. We identify an efficiency threshold with moderate detour values, offering actionable guidance for the design of real-world zone-based tariffs.

1. Introduction

In today's highly connected and specialized economy, outsourcing transportation has become a strategic necessity across many producing and retailing industries (Business Research Insights, 2025). The increasing specialization of logistics firms has made the logistics service provider (LSP) market a cornerstone of global trade (Future Market Insights, 2025). Its rapid expansion is fueled by rising yet volatile transportation demand, driven by seasonality, supply chain disruptions, and the surge in middle- and last-mile deliveries (Alicke et al., 2016). Globally, LSP revenues exceeded USD 1 trillion in 2020 and are projected to surpass USD 1.4 trillion by 2026 (Statista, 2024). This growth reflects the focus of producing and retailing companies on their core competencies while leveraging the scale, efficiency, expertise, and technology offered by specialized LSPs (Gosling et al., 2023; NTT Data, 2024; United Parcel Service, 2023). Effective coordination between these companies, collectively referred to as *shippers*, and LSPs for transportation optimization has thus become a critical driver of competitive advantage, operational resilience, and efficiency.

Our work addresses this planning problem from the shipper's perspective, relying on an LSP to handle distribution within a regional network. The corresponding process for the shipper's distribution requests typically follows two steps: the shipper issues a request to an LSP based on a predefined tariff structure, and the LSP subsequently executes the delivery tours. To facilitate efficient day-to-day coordination, LSPs commonly employ tariff systems, which enable shippers to plan more easily by offering predictable and straightforward pricing for distribution requests. While the actual tariff system and its underlying pricing logic may differ between LSPs, zone-based tariffs are the standard for freight forwarding applications, including retail distribution, parcel, and courier services (see, e.g., DHL Express, 2025;

* Corresponding author.
E-mail address: manuel.ostermeier@uni-a.de (M. Ostermeier).

https://doi.org/10.1016/j.ejor.2026.07.012
Received 11 November 2025; Accepted 7 July 2026
Available online 14 July 2026

Please cite this article as: Niklas Tuma et al., *European Journal of Operational Research,* https://doi.org/10.1016/j.ejor.2026.07.012

Masson International, 2025; SEKO Logistics, 2023; United Parcel Service, 2024). Zone-based tariffs divide the delivery area into geographic areas (zones), similar to fare zones in public transport. According to this definition of zones, the tariff does not explicitly account for a tour's total distance but approximates distance-based costs by charging based on the farthest zone visited during a delivery tour. Along with the zones and their distance-based costs, the distribution request's total load determines the total costs. The load costs per unit decrease with increasing load sizes due to volume discounts granted by the LSP. These discounts incentivize shippers to consolidate single deliveries into tours. Shippers therefore bundle customer orders from the same area into joint distribution requests (tours) to reduce costs. The total transportation costs for the resulting tours are then determined by the total load (including volume discounts) and the farthest destination zone. While zone-based tariff schemes are relatively easy to implement and communicate, owing to their aggregation of destinations into predefined zones, this simplification often comes at the expense of cost accuracy. Specifically, coarse spatial groupings do not reflect the actual driving distances and tours to be executed.

The shipper's resulting transportation problem that covers all distribution requests of a planning period is the Capacitated Vehicle Routing Problem with a Zone Tariff (VRP-ZT), which minimizes costs across all tours while accounting for the requirements imposed by LSPs and their tariff systems. The distribution requests forming tours must satisfy vehicle capacity constraints and comply with specific routing constraints to ensure operational feasibility and cost efficiency for the LSP. As driving distances are not directly covered by the zone tariffs, the LSP applies various measures to curtail them. One approach is to restrict the *total tour distance*, another is to limit the deviation imposed by bundling customers, referred to as the *detour* limit. A detour is the additional distance a vehicle must travel when visiting an extra customer during a tour, rather than approaching the farthest away customer currently on the tour directly. Detours are the key mechanism in zone-based tariff systems, as they primarily restrict the creation of tour requests by shippers. In detail, these constraints determine which customers can be clustered and, consequently, the extent to which consolidation is feasible and economically viable. To correctly account for detours, a routing solution is needed that measures the additional distances incurred when an additional customer is added to the tour request (see, e.g., Khodabandeh et al., 2021; Tuma et al., 2024). The permitted detour may be specified uniformly across all tours (e.g., an absolute threshold such as 50 kilometers) or relative to the tour length (e.g., 20%). In practice, however, different combinations of tour and detour limit formulations and implementations are used, reflecting different operational contexts and managerial preferences. We denote these different limits under the general term *(de)tour limit*.

Solving the VRP-ZT considering these limits and their combinations is highly relevant in practice. Its application arises in many real-world scenarios and across a wide range of sectors, including grocery retail (e.g., Aldi, Carrefour), consumer electronics (e.g., MediaMarkt, Best Buy), and home improvement, furniture, and department store chains (e.g., Home Depot, IKEA, Macy's). These companies engage LSPs to manage their regional distribution networks for store replenishment and home delivery services (Hübner et al., 2022; Rooderkerk et al., 2023). This involves the shipment of deliveries to several hundred customers (see, e.g., Kuhn & Sternbeck, 2013; United Parcel Service, 2023). Instead of running their own fleets, partnering with specialized LSPs such as C.H. Robinson, DHL Supply Chain, Kuehne+Nagel, or XPO Logistics enables these firms to achieve flexibility, scalability, and cost efficiency in their transportation operations. Many shippers encounter similar challenges in their day-to-day logistics operations, with the general problem setting often being transferable across industries. This makes the VRP-ZT a problem of substantial practical importance, rendering its solution fundamental for enhancing transportation efficiency. Despite these facts, research on solving the VRP-ZT is limited. A small body of related research focuses on specific applications with case-dependent structures (see e.g., Khodabandeh et al., 2021) and one particular type of a (de)tour limit. General problem settings for large-scale applications involving hundreds of customers, as well as the effects of different (de)tour limit formulations on solution quality, remain largely unexplored. The literature consequently lacks a generally applicable solution approach for the VRP-ZT that efficiently handles different (de)tour limits, analyzes their differences, and assesses their impacts on vehicle utilization, routing efficiency, and overall operational costs. Handling different (de)tour limits is an essential tool for evaluating routing solutions for versatile LSPs that apply these limits. Addressing these issues is thus vital to strengthening the connection between theoretical research and practical applications.

We approach this gap by introducing a generalized VRP-ZT that integrates different (de)tour limits. We develop an Adaptive Large Neighborhood Search (ALNS)-based matheuristic framework to efficiently solve the new variant of the problem. The algorithm applies novel problem-specific operators, exact components, and multiple adaptive mechanisms to efficiently solve large, practically relevant instances. We benchmark the matheuristic against an exact approach and validate its practicality by solving a case study with a major European retailer. We obtain optimal solutions for the benchmark and show runtime efficiency even for large problems with up to 1000 customers. Finally, we analyze differing (de)tour limits and their implications on costs and tariff design.

The remainder is organized as follows: Section 2 provides an overview of the theoretical and practical background. Section 3 formulates the VRP-ZT with different (de)tour limits, and Section 4 details the solution framework. We conduct computational studies to examine the efficiency and managerial implications of detour limits in Section 5. Section 6 concludes and discusses further areas for research for this novel application.

2. VRP with zone-based tariffs: Problem description and related literature

This section further details the VRP-ZT components – including zone structures, tariff structures, and (de)tour limits (Section 2.1) – and discusses the relevant literature (Section 2.2). We then integrate practice-based problem framing with the academic state of the art to identify the research gap and delineate the contribution of this work (Section 2.3).

2.1. Problem description of VRPs with zone-based tariffs and (de)tour limits

Zone-based tariffs as considered in our work are commonly used for transportation services across various industries (see, e.g., Elicit Technology, 2025; FleetX, 2025; FreightAmigo, 2025; ShipBob, 2025; Waredock Logistics Glossary, 2025). Applications include public transport, postal services, and long-haul transportation (see, e.g., DHL Express, 2025; Masson International, 2025; SEKO Logistics, 2023). Motivated by real-world applications in retail (see, e.g., Euro-CASE, 2001; Khodabandeh et al., 2021; Tuma et al., 2024), this paper focuses on a setting in which a shipper engages an LSP to supply up to several hundred stores (customers) daily via tours from a distribution center (DC). The shipper cannot propose all customer orders at once; instead, it must group distribution requests suitable for dedicated delivery tours operable by the LSP. We will refer to these grouped distribution requests as (customer) clusters going forward. The LSP executes the requested tour for each cluster and bills according to a zone-based tariff. This tariff involves three main components:

(1) *Zone structure*: Divides the delivery area and its customers into geographical tariff zones;
(2) *Two-dimensional tariff with volume discounts*: Prices transportation jobs according to the zone-load combination of the respective orders and grants volume discounts;
(3) *Vehicle capacity and (de)tour restrictions*: Maintains (economic) feasibility by enforcing vehicle capacity and routing constraints.

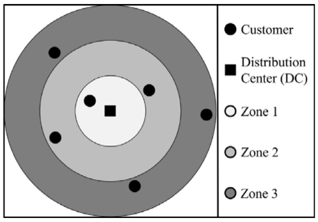


Fig. 1. Zone structure (example).

Load and zone-dependent tour costs			
	Farthest Zone: 1	**…**	**Farthest Zone: 10**
Total Load: 5	235 €	…	335 €
…	…	…	…
Total Load: 34	595 €	…	695 €

Fig. 2. Two-dimensional tariff with volume discounts (example) (see also Tuma et al., 2026).

(1) Zone structure. The key element of the zone tariff is the segmentation of the delivery area into a limited number of zones. Each customer is assigned to a distinct tariff zone, allowing planners to work with aggregated, easy-to-handle distance approximations rather than detailed route calculations. Instead of relying on exact travel distances to each customer, the system deliberately simplifies reality by approximating travel distances through zone establishment and the corresponding customer assignments. In this way, the exact distance resulting from a customer cluster is not assessed; instead, the visited zones serve as a practical proxy. Fig. 1 illustrates a zone structure consisting of three circular zones centered around the shipper's DC. Zone layouts can take different forms or reflect specific geographical features, such as postal areas or administrative districts (see Ceselli et al., 2009).[1]

(2) Two-dimensional tariff with volume discounts. The zone-based tariff combines the defined zones with the shipper's load for the customer cluster and applies the actual pricing (see Fig. 2). The distance dimension considers the distance from the DC to the zones visited by the tour, and the load summarizes the total number of load units (e.g., pallets or tons) across all customers for a cluster. To further encourage the consolidation of shipments from multiple customers, the LSP offers volume discounts applying to the total tour load. The cost per load unit, therefore, decreases as volume increases. Total transportation costs are then determined by the farthest zone of the delivery tour (i.e., the most distant customer on the tour) and the total load (i.e., the sum of all ordered units), subject to the volume discount. This information can then be easily extracted from the tariff table (see Fig. 2) by the shipper to instantly evaluate the price of a planned customer cluster. The zone tariff requires shippers to form clusters with high capacity utilization and coherent zones to ensure cost-efficient deliveries under the zone tariff.

(3) Vehicle capacity and (de)tour restrictions. An LSP accepts customer clusters from a shipper only if they comply with predefined constraints imposed to ensure that the resulting tours are both cost-efficient and operationally feasible. First, the vehicle capacity must not be exceeded. Second, the customer cluster for delivery must yield routes with economically reasonable driving distances that comply with technical constraints (e.g., the maximum truck driving distance) and legal requirements (e.g., driver working hours). While the capacity limit is a straightforward restriction on the truck's load (e.g., weight, volume, number of transportation units), limiting driving distances is more intricate. LSPs typically use out-of-route distance limits, also referred to as detour limits, to implement these restrictions. The detour is defined as the deviation from the direct distance between the DC and the farthest customer on the tour. In industry applications, the detour is usually calculated as an absolute value or as a relative fraction. Fig. 3 illustrates these detour options. An absolute limit is indicated by a fixed detour distance (e.g., 50 km) (see, e.g. Tuma et al., 2026), a relative limit by a fraction (e.g., 1.2) (see Khodabandeh et al., 2021; Lindsey et al., 2013). This implies that the allowed detour in distance units is consistent for all tours in the absolute case but changes for the relative limit. A relative limit thus allows longer detours when reaching farther customers, and is more complex to handle. To accurately account for the detour limits requires implicitly solving the routing for each requested tour to verify adherence to the detour limit. Taking a more granular look, this requires solving an open Traveling Salesperson Problem (TSP) (i.e., open tours), as LSPs price tours that start from the shipper and visit customer zones. Tours end at an LSP site, with the return to this site implicitly included in the tariff. As an alternative to detour limits, the LSPs might also apply a simple limit on the total tour distance. However, strict distance limits are rather myopic and do not reflect the LSP's intentions accurately. Our work jointly considers the resulting (de)tour limit variants and combinations.

2.2. Related literature with zone-based tariffs

Even though distribution problems with zone-based tariffs are widely encountered in practice, the academic literature that jointly considers such tariffs and (de)tour limitations remains limited. In the first set of related papers with a zone-based tariff, distance constraints either concern a total distance limit or are not explicitly integrated. Ceselli et al. (2009) present a first VRP-ZT. Their tariff system considers multiple cost criteria (e.g., distance, value, load, and stops), volume discounts, and a total distance limit (based on driving time). They solve a rich VRP with a heterogeneous fleet, multiple LSPs, multiple depots, time windows, open tours, split delivery, two delivery modes, and incompatibility constraints. The authors apply a Branch&Price with three sequential phases to increase the degree of order splitting. They solve a case study with data obtained from a software company. Similarly, Ceschia et al. (2011) consider various tariff structures and address a VRP with a heterogeneous fleet, time windows, two delivery modes, and a multi-period planning horizon. They consider multiple different tariff cost functions. Tours are limited by a maximum total driving time, and a tabu search is applied to solve their case studies from an industrial partner. Tordecilla et al. (2021) propose a specialized heuristic for an agri-food rich VRP with a heterogeneous fleet, incompatibility constraints, stop priorities, and multiple compartments. The tour costs depend on the most expensive zone visited, the number of stops, and the total load without volume discounts. A further rich application in this context is Soleilhac et al. (2026). Drawing on a case from the retail industry, they study a VRP with a heterogeneous fleet, multiple LSPs, time windows, and both less-than-truckload and full truckload shipping modes. The authors motivate a tariff scheme for tour costs across multiple LSPs and determine the assignment of orders to less-than-truckload or full truckload tours, accounting for the limited fleets and capacities of LSPs. Tour length and detour constraints are motivated but not explicitly reflected in their approach. A common strength of these papers is their strong grounding in practice. All consider rich VRP variants with various real-world features and embed zone tariff structures. They incorporate pricing elements reflecting the complexity of contractual agreements between shippers and LSPs. While the problems considered are rich, this richness often obscures structural insights, making it difficult to generalize the findings or

[1] Note that there is a separate stream of literature (see e.g., Carlsson et al., 2025) where zoning is referred to as the partitioning of the decision problem into smaller sub-problems, which is different from our application.

Absolute Detour Example	Relative Detour Example
$Tour - Direct\ Connection =$ $566\ km - 525\ km = 41\ km$	$\frac{Tour}{Direct\ Connection} = \frac{566\ km}{525\ km} = 1.08$

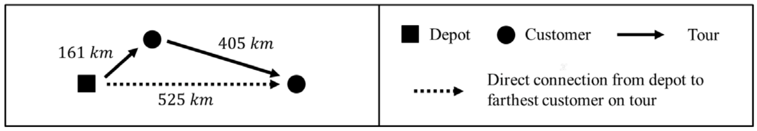


Fig. 3. Detour limit options (example) (see also Tuma et al., 2024).

derive broader managerial implications. From a conceptual standpoint, these early studies treat zone-based tariffs inconsistently and largely implicitly. Although some kind of tariffs are incorporated, they are embedded in specific cost functions rather than explicitly modeled and examined as a central structural element. There is no unified framework that isolates and systematically studies zone-based tariffs as the core of the models. More strikingly, they do not include actual detour constraints. As a result, they fail to capture an essential operational mechanism used by LSPs to control inefficiencies arising from indirect routing.

The second line of work explicitly incorporates detour limits alongside zone-based tariffs, thereby aligning the approach more closely with operational practice. Tuma et al. (2024) study a variant with a zone-based tariff and a heterogeneous fleet, multiple LSPs, two delivery modes, open tours, and incompatibility constraints. The zone tariff is based on the total load and the distance to the farthest zone visited. The costs are subject to volume discounts, and a relative detour limit is applied. A tree-based set partitioning algorithm is tailored to the application case. The constrained problem setting enables the computation of the full feasible set of tours and the determination of optimal solutions. Tuma et al. (2026) are the first to formalize the VRP-ZT with absolute detour limits. The zone tariff reflects the farthest zone and the total load. The authors decompose the problem and apply a Branch&Check (BAC) approach with multiple acceleration techniques. Their exact approach can only handle smaller problem sizes, with fewer than 100 customers, and an absolute detour limit.

These papers address routing inefficiencies by imposing detour restrictions while maintaining tariff-based pricing schemes. Despite these advances, the treatment of detour limits remains partial: each paper considers either absolute or relative detour constraints, without a unified, general framework that integrates flexible zone-based tariffs with different or combined types of (de)tour constraints.

2.3. Synthesizing industry practice and research in VRPs with zone-based tariffs

The VRP-ZT has not yet been systematically investigated in the academic literature, with most existing insights originating from industry-driven applications. While prior research demonstrates the clear practical relevance of zone-based tariffs, it largely embeds them within highly specific and rich VRP settings, often with tailored cost functions and isolated modeling choices. As a result, the literature remains fragmented, lacking a unified perspective. In particular, there is no general modeling framework that abstracts from these application-specific formulations to consistently capture zone-based tariffs and their interaction with different types of (de)tour constraints. Table 1 summarizes the existing contributions and highlights these structural gaps, thereby emphasizing the need for a more systematic and generalized approach.

The overview reveals that prior studies differ considerably in the tariff structures they consider. While all contributions incorporate zones in some form, the definition of zones ranges from geographic and concentric structures to more general or unspecified designs. In addition to zone-based pricing, papers include various tariff components such as volume discounts or additional criteria (e.g., load, stops, or multiple attributes). Our model and solution approach are generally suited to general zone design and multiple zone criteria, but focus on the load component and concentric zones motivated by our case application. More importantly, the integration of actual detour limitations remains fragmented. Only two works address such constraints, and they consider only one variant in isolation. There is no approach to assess existing limits simultaneously or to compare tariffs that rely on diverging limit applications. This means the impact of different (de)tour limit options and their differences remains largely unexplored. Limits influence tour construction, thereby affecting vehicle utilization, route efficiency, and operational costs. The specific magnitude of each (de)tour limit is expected to produce a distinct impact on routes. Insights from our collaborating industry partner reveal similar knowledge gaps regarding the impact of (de)tour limits. Limits are often set ad hoc to prevent excessively long tours, a simplification that may arise from a lack of advanced decision support. The actual effects on shipper costs, tour efficiency, and LSP revenues are therefore unclear. Bridging theory and practice requires an in-depth study that analyzes the effects of different (de)tour limits and provides decision support in choosing a (de)tour option.

Furthermore, various solution approaches exist, including exact methods, metaheuristics, and hybrid approaches. However, the scalability of these methods is limited, with most studies focusing on relatively small to medium-sized instances or case-specific settings. Many industry applications require advanced routing methods that can serve several hundred customers while accommodating different (de)tour limit options, as not all collaborating LSPs may apply the same restrictions.

This paper addresses this gap and provides a more comprehensive treatment by considering different types of (de)tour limitations simultaneously, while also demonstrating scalability to significantly larger instances of up to 1000 customers. We provide the first approach to handle large-scale, general problem settings applicable to common (de)tour limitation rules. Current contributions primarily address highly specific industry applications, focusing on tailored solutions for specific settings (e.g., heterogeneous fleets) and reporting case-based results. We adopt a more general perspective, aiming to generate broader methodological insights, particularly regarding the implications of different detour limits, without integrating case-individual routing constraints.

3. Mathematical model

This section depicts the VRP-ZT, originally introduced by Tuma et al. (2026), and extends the model to cope with absolute detour, relative detour, and total distance limits simultaneously. The model minimizes the shipper's cost by determining cost-efficient tours under a zone tariff while respecting the LSP's constraints. Table 2 summarizes the notation.

The VRP-ZT is defined on a directed graph $G(N, A)$. N defines the set of nodes and includes all customer nodes $i \in I$ and the node 0 for the DC ($N = I \cup \{0\}$). The arc set $A = \{(i, j) : i, j \in N, i \neq j\}$ defines the corresponding connections, and each arc is weighted by the

Table 1
Overview of related literature with zone-based tariffs.

Publication	Applied tariff structure						Solution approach	Instance size[b]
	Criteria[a]	Zones	Vol. Disc.	(De)tour limit				
				Abs.	Rel.	Dist.		
Ceselli et al. (2009)	multiple	geographic	✓	–	–	✓	Branch&Price	47
Ceschia et al. (2011)	load	not specified	✓	–	–	✓	Tabu Search	56
Tordecilla et al. (2021)	stop	not specified	–	–	–	–	Set Part. Heuristic	214
Soleilhac et al. (2026)	stop	concentric	–	–	–	–	LNS	100
Tuma et al. (2024)	load	concentric	✓	–	✓	–	Set Part. Exact	143
Tuma et al. (2026)	load	concentric	✓	✓	–	–	BAC	60
This paper	load	concentric	✓	✓	✓	✓	ALNS	1000

[a] Cost components additional to zone-based criteria; Multiple includes distance, number of pallets, weight, volume, value & stops.
[b] Maximum instance size (number of customers) solved by the proposed algorithm in the paper.

Table 2
Notation.

Index sets	
I	Set of customers, with $i = 1, \ldots, I$
L	Set of possible vehicle loads, with $l = 1, \ldots, Q$
T	Set of vehicle tours, with $t = 1, \ldots, T$
Z	Set of tariff zones, with $z = 1, \ldots, Z$
Parameters	
$c_{z,l}$	Tariff costs for zone z with total load l [in currency units]
$d_{i,j}$	Distance from location i to location j, $i, j \in I \cup \{0\}$ [in distance units]
Δ^a	Absolute detour limit [factor]
Δ^r	Relative detour limit [factor]
Δ^k	Total distance limit [factor]
q_i	Order size of customer i [in load units]
Q	Vehicle capacity [in load units]
f_i	Tariff zone of customer i
Variables	
$n_{i,j}$	Binary: 1, if customer i is visited before customer j on the same tour; 0, otherwise
u_i	Continuous: tour distance to reach customer i
v_t	Continuous: total distance of open tour t
$w_{i,t}$	Binary: 1, if customer i is selected for the calculation of the detour t; 0, otherwise
$x_{i,t}$	Binary: 1, if customer i is included in tour t; 0, otherwise
$y_{t,z,l}$	Binary: 1, if the tariff for tour t is defined by zone z and load l; 0, otherwise

distance $d_{i,j}$. All distances adhere to the triangle inequality. The shipper clusters customers $i \in I$ to open tours $t \in T$ that originate at node 0 but do not return to it (open tours). Each customer has a non-negative demand q_i and belongs to a tariff zone f_i. The tours are executed by an LSP operating a sufficiently large and homogeneous fleet of vehicles with a capacity of Q. The LSP may restrict the tour lengths by a total distance limit Δ^k, an absolute detour limit Δ^a, a relative detour limit Δ^r, or a combination of them. The simultaneous consideration of relative detour and total distance limit, in addition to an absolute limit, is the essential model extension compared to Tuma et al. (2026). The costs of a vehicle tour in the zone tariff $c_{z,l}$ depend on the carried load units $l \in L$, with $L = \{1, \ldots, Q\}$ and the tariff zone $z \in Z$.

We introduce the following continuous variables: u_i tracks the distance traveled to reach customer i, and v_t represents the total distance of a tour t. Further, we introduce four binary variables. $n_{i,j}$ indicates if two nodes i and j are visited consecutively, $w_{i,t}$ selects a customer i for the calculation of the detour for tour t, $x_{i,t}$ assigns customer i to tour t, and $y_{t,z,l}$ determines the tariff with zone z and load l for tour t. The complete model is formalized with Eqs. (1) to (22):

$$Minimize\ TC = \sum_{t \in T} \sum_{z \in Z} \sum_{l \in L} c_{z,l} \cdot y_{t,z,l} \tag{1}$$

subject to

$$\sum_{t \in T} x_{i,t} = 1 \qquad \forall i \in I \tag{2}$$

$$\sum_{z \in Z} \sum_{l \in L} y_{t,z,l} \le 1 \qquad \forall t \in T \tag{3}$$

$$\sum_{i \in I} q_i \cdot x_{i,t} \le \sum_{z \in Z} \sum_{l \in L} l \cdot y_{t,z,l} \qquad \forall t \in T \tag{4}$$

$$f_i \cdot x_{i,t} \le \sum_{z \in Z} \sum_{l \in L} z \cdot y_{t,z,l} \qquad \forall t \in T,\ i \in I \tag{5}$$

$$n_{i,j} + n_{j,i} \ge x_{i,t} + x_{j,t} - 1 \qquad \forall t \in T,\ i, j \in I,\ i < j \tag{6}$$

$$u_i \ge d_{0,i} \qquad \forall i \in I \tag{7}$$

$$u_j \ge u_i + d_{i,j} - M(1 - n_{i,j}) \qquad \forall i, j \in I,\ i \ne j \tag{8}$$

$$v_t \ge u_i - M(1 - x_{i,t}) \qquad \forall t \in T,\ i \in I \tag{9}$$

$$v_t - \sum_{i \in I} d_{0,i} \cdot w_{i,t} \le \Delta^a \cdot max_{i \in I} d_{0,i} \qquad \forall t \in T \tag{10}$$

$$v_t - \sum_{i \in I} d_{0,i} \cdot w_{i,t} \le \Delta^r \cdot \sum_{i \in I} d_{0,i} \cdot w_{i,t} \qquad \forall t \in T \tag{11}$$

$$v_t \le \Delta^k \cdot max_{i \in I} d_{0,i} \qquad \forall t \in T \tag{12}$$

$$w_{i,t} \le x_{i,t} \qquad \forall t \in T,\ i \in I \tag{13}$$

$$\sum_{i \in I} w_{i,t} \le 1 \qquad \forall t \in T \tag{14}$$

$$x_{j,t+1} \le \sum_{i=1}^{m} x_{i,t} \qquad \forall j, m \in I,\ j \le m,\ t \in \{1, \ldots, |T| - 1\} \tag{15}$$

$$\sum_{z \in Z} \sum_{l \in L} y_{t,z,l} \le \sum_{i \in I} x_{i,t} \qquad \forall t \in T \tag{16}$$

$$\sum_{i \in I} q_i \cdot x_{i,t} \le Q \qquad \forall t \in T \tag{17}$$

$$u_i \in \mathbb{R}_0^+ \qquad \forall\ i \in I \tag{18}$$

$$v_t \in \mathbb{R}_0^+ \qquad \forall t \in T \tag{19}$$

$$n_{i,j} \in \{0, 1\} \qquad \forall\ i, j \in I,\ i \ne j \tag{20}$$

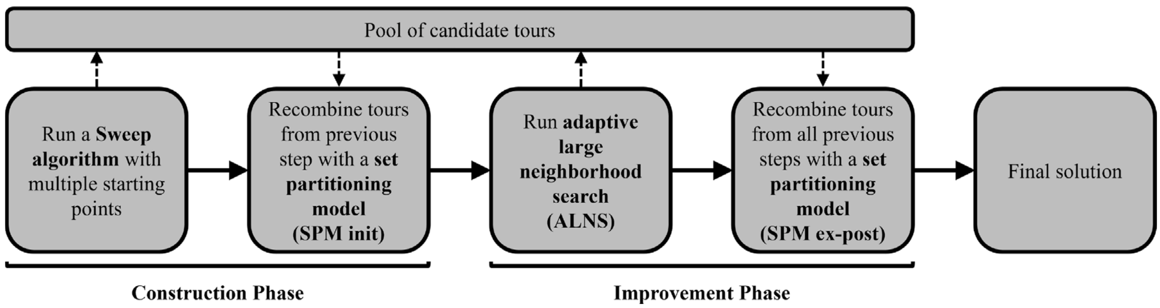


Fig. 4. Overview of the Adaptive Learning Partitioning Heuristic Algorithm (ALPHA).

$$w_{i,t},\ x_{i,t} \in \{0,1\} \qquad \forall\ i \in I, t \in T \tag{21}$$

$$y_{t,z,l} \in \{0,1\} \qquad \forall t \in T,\ z \in Z,\ l \in L \tag{22}$$

The Objective Function (1) minimizes the shipper's total costs (TC) based on the zone tariff for each tour t. Constraints (2) ensure that all customers are visited by exactly one tour (no split delivery). Inequalities (3)–(5) concern the tariff specification. Constraints (3) make sure that a tariff is selected for each non-empty tour, while none is selected for empty tours (no pricing of empty tours). Constraints (4) determine the load dimension of the tariff according to the tour load. Analogously, Constraints (5) determine the tariff zone depending on the customers visited on the tour. Constraints (4) and (5) ensure that the highest zone and the total load of all customers on the tour are selected. Constraints (6) define the customer sequence of tour t. Using this sequence, Constraints (7)–(9) calculate the tour distance, where M is a sufficiently large number. Constraints (10)–(12) model the different (de)tour options: the absolute detour limit (10), the relative detour limit (11), and the total distance-based limit (12). Constraints (10) and (11) ensure that detour limits are maintained by determining the detour based on the tour distance and the direct distance from the DC to one selected customer of the tour. This further requires Constraints (13) and (14) to determine w_{it}, which selects one customer for each tour and specifies the direct distance from the DC for the detour calculation. The farthest customer need not be chosen to meet the detour limit, but selecting it ensures the full detour allowance is used. Constraints (12) limit the maximum tour distance. Valid inequalities (15)–(17) strengthen the MIP (Thorsteinsson, 2001; Tuma et al., 2026). Inequalities (15) prevent parallel solutions by restricting the tour selection, (16) ensure that a tariff is only selected for an active tour, and (17) defines the capacity limit Q. Constraints (18)–(22) define the variable domains.

4. Matheuristic solution framework

The VRP-ZT is an NP-hard Generalized Assignment Problem (GAP) to multiple resources, in which customers are assigned to clusters (tours) subject to capacity and distance restrictions. While the routing sequence is not relevant to cluster pricing, determining (de)tour adherence requires solving an open TSP for each tour (see Section 2.1). The arising complexity renders exact methods quickly impractical as problem sizes grow. Current exact methods (see Section 2.2) exhibit limited scalability and adaptability and are thus not applicable to industry problems involving hundreds of customers per day and the existing (de)tour limits. We address these challenges and propose a matheuristic, the Adaptive Learning Partitioning Heuristic Algorithm (ALPHA), tailored to the problem at hand. It exploits problem-specific features and reflects requirements observed in real-world operations. This enables us to efficiently solve large problems, accommodating all relevant constraints and limits. ALPHA operates with a *construction phase* and an *improvement phase*, combining three different components: A *Sweep algorithm*, two applications of a *Set Partitioning Model (SPM)*, and an *ALNS*. Fig. 4 illustrates the complete matheuristic framework.

We construct initial solutions using a Sweep algorithm with multiple start nodes. The tours of these solutions are saved as a pool of candidate tours. The SPM then recombines the generated tours to improve the initial solution. The improvement phase applies a tailored ALNS algorithm. All newly created tours during the ALNS are added to the candidate tour pool. After the ALNS terminates, we reapply the SPM as a post-optimization step. Tours of the VRP-ZT are represented by an unordered cluster of customers, as the actual sequence is only relevant for (de)tour adherence.

4.1. Sweep algorithm

The well-known Sweep algorithm (Gillett & Miller, 1974; Wren & Holliday, 1972) clusters customers based on their polar angles relative to the DC location. The algorithm provides fast solutions well-suited to our problem by combining geographically related customers to construct tours with low detours and high vehicle utilization. Due to these properties, the algorithm performs well with the zone tariff structure and provides fast, diverse initial solutions. The Sweep algorithm controls for capacity and (de)tour constraints when adding a new customer. We implement the dynamic programming algorithm of Held and Karp (1962) to solve an open TSP and verify adherence to the (de)tour limit. We apply the Sweep algorithm once for each customer (as a starting point), yielding $|I|$ initial VRP-ZT solutions, whose tours populate the candidate pool G.

4.2. Set Partitioning Model (SPM)

Our solution approach is based on creating a diversified pool of tour candidates from which we can select the best combination to obtain a final solution for the VRP-ZT. Therefore, we pool all tours generated (and discarded) during the construction and improvement phases and consider these tours for evaluation. We employ an SPM at the end of the construction and improvement phase for this purpose. As the SPM is NP-hard and its solving consumes substantial runtime, we deliberately limit the number of applications to a minimum. However, especially for large problems, this post-optimization step can improve the final solution, as promising tours may be discarded during ALNS runs. The model specifies for each tour t in the pool of generated tours G whether it is part of the optimized VRP-ZT solution using the binary variable s_t (1 if tour t is selected; 0, otherwise). The SPM then minimizes the total costs across all tours (see (23)), where h_t represents a tour's costs. Eqs. (24) ensure that each customer is visited exactly once by a tour. The binary parameter $k_{i,t}$ indicates if a customer is part of a tour (1 if tour t serves customer i; 0, otherwise; see (25)).

$$Minimize \sum_{t\in G} h_t \cdot s_t \tag{23}$$

subject to

$$\sum_{t\in G} k_{i,t} \cdot s_t = 1 \qquad \forall i \in I \tag{24}$$

$$s_t \in \{0,1\} \qquad \forall t \in G \tag{25}$$

4.3. Adaptive Large Neighborhood Search (ALNS)

The ALNS was introduced by Ropke and Pisinger (2006a, 2006b) and is successfully applied for versatile VRP variants (see, e.g., Fontaine, 2022; François et al., 2019; Lehmann & Winkenbach, 2024; Mara et al., 2022). The ALNS excels through its flexibility and adaptability. Its suitability for the VRP-ZT stems in particular from its high flexibility in handling complex and non-standard cost structures. It combines diverse removal and insertion operators within an adaptive framework that dynamically guides the search towards promising regions of the solution space.

We build on the tariff features and design problem-specific removal operators (see Section 4.3.1) and insertion operators (see Section 4.3.2) that explicitly account for tariff zone characteristics and their impact on routing costs. This allows the search to better exploit cost-saving opportunities arising from zone configurations. We apply an adaptive mechanism to determine the extent of destruction for each removal operator (see Section 4.3.3), enabling a more balanced exploration–exploitation trade-off in the presence of the intricate cost relationship induced by zone tariffs and their routing requirements. Simulated Annealing (SA) is employed as an acceptance criterion and terminates the search after a maximum number of iterations (see Section 4.3.4).

Our ALNS formulation further differs from standard ALNSs in its solution representation. We use a random order to represent the customers assigned to a tour, not the actual visit sequence. When assigning customers to a tour, we compute the shortest possible distance of the new tour to check whether it meets the detour limit. Solving an open TSP ensures that only feasible solutions are explored during the search. We solve the open TSPs efficiently using the dynamic programming algorithm of Held and Karp (1962), as for the Sweep algorithm. All ALNS solutions' tours are then added to our tour pool as candidates. Our approach improves the insertion step by selecting the optimal sequence and minimizing costly attempts to insert at different positions. This approach works especially well for the limited cluster sizes typical for VRP-ZT applications, as we observe a substantial reduction in computation time using the random order.

4.3.1. Removal operators

We propose three removal operators $\rho \in P^r$, which combine VRP-ZT-specific operators with standard operators that perform well in related problem settings. The removal operators remove b customers from the current solution and are defined as follows.

- **Random Removal:** Fast diversification operator that randomly removes b customers.
- **Similar Zone Removal:** Problem-specific operator that leverages the tariff structure to minimize zone variation. It selects a random customer, calculates the zone difference for all remaining customers and removes the chosen customer, as well as the $b-1$ customers with the lowest zone difference.
- **Concatenated Random Removal:** Diversification operator that concatenates all tours into a single giant tour, chooses a random customer, and removes the customer along with the $b-1$ next customers from the concatenated tour. Owing to our solution representation with random order, the operator differs from traditional VRP applications (see Voigt, 2024). The random customer order is maintained for the created sub-tours of the concatenated tour, enabling greater diversification in the selection of removal customers than traditional operators.

4.3.2. Insertion operators & tour feasibility

Once customers have been removed from the current solution, they must be reinserted while ensuring tour feasibility. As detailed above, the feasibility of each tour concerning the detour limit is ensured by solving an open TSP when inserting a customer to a tour. Insertions resulting in an infeasible tour are discarded. We propose six insertion operators $\rho \in P^i$ to reinsert customers into existing tours or construct new tours. We again employ a combination of novel problem-specific and generally proven operators.

- **Greedy Insertion:** Inserts removed customers in a random order into the position with the smallest increase in cost. A new tour is created if no tour can take the additional customer.
- **Greedy Demand Insertion:** Sorts the removed customers by demand and inserts the customers with the highest demand first into the best positions.
- **Greedy Distance Insertion:** The detour limits generally make it more favorable to insert a customer close to the DC than farther away. This operator therefore sorts the removed customers by their distance to the DC in ascending order and inserts the customers with the lowest distance first at the best position.
- **Random Insertion:** Inserts removed customers in random order in a random tour. A new tour is created if no tour can incorporate the additional customer.
- **Random Zone Insertion:** Inserting a customer that changes the highest zone of a tour leads to increased costs in a VRP-ZT. The higher the difference in the maximum zone, the higher the costs. This operator therefore inserts removed customers in a random order so that the insertion results in the smallest possible increase in the highest zone.
- **Zone Demand Insertion:** Higher vehicle utilization leads to lower unit costs. A change in the highest zone results in higher costs for the entire tour. To exploit these tariff characteristics, this operator inserts each removed customer into a tour, maintaining the highest zone while maximizing capacity utilization.

4.3.3. Adaptive mechanisms

A key feature of the ALNS is its ability to learn during the search. ALPHA applies two adaptation methods – *(1) adaptive operator selection* and *(2) adaptive removal value* – which adjust probabilities based on past operator performance.

(1) Adaptive operator selection. We use a roulette wheel with adaptive weights to select one removal and one insertion operator at each iteration. Every operator $\rho \in P^r \cup P^i$ has a selection probability ϕ_ρ, which depends on the relative weight of the operator compared to all other operators of the same type (i.e., P^r or P^i). The selection probability is defined by $\phi_\rho = \frac{\omega_\rho}{\sum_{\rho \in P^r/P^i} \omega_\rho}$. The initial weight of each operator ρ is set at $\omega_\rho = 1$. After each iteration, we update the weights of the selected operators based on their performance, using Eq. (26).

$$\omega_\rho = \alpha \cdot \omega_\rho + (1-\alpha) \cdot \psi_\rho^{score} \tag{26}$$

The update is based on the current weight ω_ρ, the operator score ψ_ρ^{score}, and a parameter α. The operator score depends on the operator performance and increases depending on the quality of the produced solution: (1) new global best solution, (2) improved incumbent solution, (3) accepted (worse) solution, and (4) rejected solution.

(2) Adaptive removal value. We adapt the number of customers for removal b in each iteration γ using a binomial distribution $B(n,p)$ (see Voigt et al., 2022) and assigning an individual probability p_ρ^{binom} to each removal operator. In each ALNS iteration γ, we first select an operator ρ and then draw the respective b_γ using a binomial distribution, i.e., $b_\gamma \sim B(|I|, p_\rho^{binom})$. The probability p_ρ^{binom} is updated after each iteration: $p_\rho^{binom} = \beta \cdot p_\rho^{binom} + (1-\beta) \cdot p_\rho^{binom\_avg}$, with $p_\rho^{binom\_avg} = \frac{\sum_{q=1}^{\gamma} b_q/|I| \cdot \psi_{\rho,q}^{score}}{\sum_{q=1}^{\gamma} \psi_{\rho,q}^{score}}$ and β as weighting parameter. The value $p_\rho^{binom\_avg}$ relates the actual number of removed customers and the operators' performance ($\psi_{\rho,q}^{score}$) of each iteration.

Table 3
Overview of numerical analyses.

Sec.	Rationale	Data (No. of instances)	Results
5.1	Application to case company		
	• Show real-world applicability	Case company (10)	Table 4
	• Show impact of different detour limit magnitudes	Case company (10)	Table 5
5.3	Computational efficiency of ALPHA		
	• Compare ALPHA to BAC on small instances	Solomon-based (90)	Table 6
	• Show scaling of ALPHA on large instances	Uchoa-based (100)	Fig. 5
	• Show performance of different framework components	Uchoa-based (35)	Table 7
	• Show performance of different ALNS operators	Uchoa-based (35)	Table 8
5.4	Managerial insights on (de)tour limits		
	• Match different (de)tour limits	Uchoa-based (35)	Table 9
	• Show implications of different (de)tour limits	Uchoa-based (35)	Table 10
	• Show functioning of absolute and relative limits	Uchoa-based (1)	Fig. 6
	• Show functioning of hybrid limits	Uchoa-based (1)	Fig. 7
	• Compare absolute and relative limits	Uchoa-based (35)	Table 11
	• Show trade-off when choosing detour limit value	Uchoa-based (1)	Fig. 8
	• Show impact of spatial customer distribution on KPIs	Uchoa-based (35)	Table 12
	• Show impact of demand distribution on KPIs	Uchoa-based (35)	Table 13
	• Show impact of order sizes on detour utilization	Uchoa-based (1)	Fig. 9

4.3.4. Acceptance criterion

We use Simulated Annealing (SA) to evaluate the acceptance of each new ALNS solution S' for further search (Kirkpatrick et al., 1983). We set the initial SA temperature $\tau = \frac{TC(S^{init})\cdot\overline{\sigma}}{log(1/\sigma^{init})}$ to achieve a specific initial acceptance percentage σ^{init} of solutions that are at most $\overline{\sigma}$ worse compared to the initial total costs $TC(S^{init})$. The temperature τ then decreases following the cooling rate $\tau = \epsilon \cdot \tau$. We choose the step size ϵ in the interval $[0,1]$ such that the temperature converges to one after γ^{max} iterations: $\epsilon = (1/\tau)^{(1/\gamma^{max})}$. Solution improvements are always accepted. Worse solutions are further evaluated for their difference to the incumbent solution S and accepted with probability $e^{(TC(S)-TC(S'))/\tau}$ (Roozbeh et al., 2018; Ropke & Pisinger, 2006a). Note that new tours are added to the candidate tour pool independent of the solution acceptance.

Algorithm 1 summarizes the complete ALPHA framework.

5. Numerical analysis

We begin our numerical analysis by solving a case from a European retailer in Section 5.1. This establishes the practical relevance of the approach by applying it to real-world data, systematically comparing different detour limit settings, and demonstrating its direct applicability and impact. Building on these insights, we derive a set of generic problem instances that abstract from the case-specific setting while preserving its key structural characteristics. These instances, described in Section 5.2, form a robust foundation for generating broadly applicable insights across different operational contexts and are used in all subsequent analyses. Section 5.3 then provides a comprehensive evaluation of the algorithmic performance. We benchmark the proposed approach against exact methods on small instances, analyze its scalability on large-scale problems, and isolate the contribution of individual algorithmic components. This enables a rigorous assessment of both solution quality and computational efficiency. Finally, Section 5.4 derives managerial insights by examining the effects of (de)tour limit types and magnitudes, as well as the influence of spatial and demand characteristics on key performance indicators. This analysis reveals the operational trade-offs and decision-relevant implications of different modeling choices. Table 3 highlights a coherent and multi-layered analysis that integrates practical validation, methodological rigor, and actionable managerial insights.

Test bed and parameter setting. We implement ALPHA and the exact benchmark approach in Python (version 3.12) using Gurobi (version 12.0.1) as the solver. ALPHA's source code can be shared upon request to the authors. An AMD 5950X processor (16 cores, 3.4 GHz, disabled turbo boost, individual Passmark single-thread score: 2540) with 64 GB of random-access memory is used for all computations. We limit the processor to a single thread and the runtime of the SPM in our framework to $|I|/10$ seconds. We report, unless stated otherwise, the average result of ten runs per instance.

We follow the systematic parameter calibration of Ropke and Pisinger (2006a) and apply a two-phase approach. In the first phase, we establish an initial configuration through experiments conducted during the heuristic's development. In the second phase, we refine this configuration by systematically varying one parameter at a time while holding all others constant. For each candidate configuration, we run ALPHA 10 times on 10 Uchoa-based instances with uniformly distributed sizes ranging from 100 to 1000 customers. We select the configuration that yields the best average performance. We then advance the process to the next parameter, carrying forward the best values identified so far and retaining the initial values for any parameters not yet evaluated. We repeat this cycle until all parameters are optimized. The tuning rationale results in the following parameter values: Removal and roulette wheel weight parameters $\alpha = \beta = 0.7$; Initial operator weights $\omega_\rho = 1$; Operator scores (best/better/accepted/rejected) (ψ_ρ^{score}): 10/10/0/0; ALNS iterations $\gamma^{max} = 3000$; SA parameters: $\overline{\sigma} = 0.25$, $\sigma^{init} = 0.6$; Initial binomial probability $p_\rho^{binom} = 0.15$;

5.1. Application to case company

The real-world case of a European do-it-yourself retailer serves as the starting point of our numerical analysis, illustrating the practical relevance of the VRP-ZT and demonstrating the applicability of ALPHA to the retailer's decision problem. The case data comprises the supply of about 300 stores from a central warehouse in Germany. Stores are supplied by an LSP multiple times a week. The retailer operates with fixed delivery schedules that define the delivery days per store. The LSP bills according to a zone tariff with volume discounts across five concentric zones based on postal areas. The LSP employs homogeneous trucks with a capacity of 34 euro pallets and allows a relative detour limit of $\Delta^r = 0.2$. Due to a non-disclosure agreement, further details of the instances and tariff data can only be shared upon reasonable request to the authors.

Table 4 summarizes the case study results over two operational weeks. The data shows notable day-to-day variations in order volumes and tour requirements, reflecting actual dynamics in retail distribution. ALPHA solves each instance in well under one minute, demonstrating its efficiency and applicability in large-scale, real-world settings.

Algorithm 1 ALPHA framework

Initialize parameters and weights
Initialize set of all generated tours: $G \leftarrow \{\}$
for $i \in I$ **do** ▷ Construction phase (Sweep algorithm)
  Solve Sweep algorithm with customer i as starting point and save solution S
  $G \leftarrow G \cup \text{tours}(S)$
end for
$S^{init} \leftarrow SPM(G)$
$S \leftarrow S^{init}$
for $\gamma \in \{1, ..., \gamma^{max}\}$ **do** ▷ Improvement phase (ALNS)
  Select one $\rho \in P^r$, one $\rho \in P^i$ and determine the number of removals b_γ
  Remove b_γ customers of the current solution using the selected removal operator
  Insert removed customers using the selected insertion operator while ensuring feasibility
  Save new solution as S'
  $G \leftarrow G \cup \text{tours}(S')$
  Update ω_ρ and p_ρ^{binom} for all operators
  Check acceptance of S' based on SA
  if S' is accepted **then**
    $S \leftarrow S'$
  end if
end for
$S \leftarrow SPM(G)$ ▷ Post-optimization step (SPM):
return S

Moreover, we achieve substantial savings of 4.9% compared to the company's current solution (rule-based approach).[2] Extrapolating these weekly savings on an annual level results in yearly savings in routing costs of approx. € 350k.

Week 1 shows an average order volume of 918 pallets, corresponding to approximately 75 orders per day, and resulting in 28 tours and a total distance of around 10,800 km per day. The average detour utilization was 7.7%, meaning that on average the total distance was 7.7% higher than the direct distance from the depot to the farthest customer. This detour utilization is well below the 20% detour limit, suggesting that routes are operated efficiently within the permitted flexibility. The limited use of the detour can be attributed to large order sizes, which result in shorter tours, making vehicle capacity the primary binding constraint. Week 2 shows a comparable number of tours but an increase in travel distances (+82%) due to a different customer structure. Detour utilization further decreases – only 5.2% of the allowed 20% – due to higher volumes per order (+26%) and fewer orders (stops) per tour (−21%). The increasing order sizes in Week 2 result in more single-stop tours. Single-stop tours, by definition, do not involve any detours.

Altogether, the results show that ALPHA achieves efficient tours for both weeks; however, operational conditions – such as spatial demand distribution and daily order variability – strongly influence tour structure and detour utilization. Lower realized detours in the second week are due to the more geographically dispersed demand pattern, highlighting the method's adaptability to varying logistics conditions and parameter constellations.

We further use the case to conduct a sensitivity analysis of Δ^r, aiming to gain insights into the impact of the detour limit. Table 5 shows the corresponding results. Using the same cost tariff, a stricter detour limit results in less efficient tours for the shipper (e.g., fewer orders/stops per tour and lower vehicle utilization) and higher costs. Without tariff adaptations, the costs increase by 1.3 to 1.7%, resulting in approx. € 100k in additional annual expenses for the retailer. Intuitively, one might expect that lower detour limits reduce total travel distances, benefiting the LSP. However, stricter limits can prevent combining certain customers, sometimes increasing total travel distances. The analyses show that both low and high limits increase distances. Lower limits do not allow for efficient truck tours, while higher limits do not significantly restrict the distance. The analysis shows that the contracted detour $\Delta^r = 0.2$ is balanced, offering a good trade-off between distance and vehicle utilization. Furthermore, a detour limit of 0.4 already provides substantial flexibility for customer bundling. Increasing this flexibility further (e.g., to 1.0) has a relatively small impact on the solutions, as vehicle capacity constraints become binding and additional cost reductions, e.g., achieving further volume discounts within a zone, are limited. Comparing the two weeks shows that the total distance and the detour are less sensitive to looser detour limits for the second week. This lower sensitivity is due to larger order sizes and the more dispersed customer locations in week 2. Table A.14 in the Appendix further provides more detailed insights of the different relative detour limits.

5.2. General data instances adapted from literature

We supplement our case study insights with analyses based on data generated from the literature and informed by our case study. This allows us to generalize the findings beyond the specific case study context and to systematically assess the computational performance of the solution approach.

Overview of customer data. We utilize two sets of VRP-ZT instances for our numerical analysis. The first set, introduced by Tuma et al. (2026), comprises 90 small- to medium-sized instances with 30, 45, 60, and 100 customers. These instances are adapted from the well-known benchmark data of Solomon (1987). The second set consists of 100 large-scale instances, each containing between 100 and 1000 customers, generated based on Uchoa et al. (2017) and adapted using the logic by Tuma et al. (2026) to fit the VRP-ZT requirements. In both cases, we uniformly draw the order size (demand) for each customer in the interval [1;34], matching the vehicle capacity of 34 load units. As shown in Table 3, we use a representative subset of the Uchoa-based instances for most tests. The subset consists of 35 instances, representing all instances from three size brackets (number of customers): [[100;199],[500;599],[900;1000]]. This allows us to provide specific insights for small-, medium-, and large-scale problems.

[2] The retailer's approach first sorts customers by postal code to obtain a list of geographically related locations. In a further step, customers with similar postal codes are assigned to a vehicle to achieve a high capacity utilization. Finally, the vehicle's detour is checked, and customers may be reassigned to a different tour. This procedure is repeated until every customer is assigned to a vehicle.

Table 4
Case study results.

Week 1								
Data			Run[b]	No. of	Zones	Distance	Detour[c]	Savings[d]
Day	Orders	Order volume[a]	time	tours	/ tour	in km	in %	in %
Mo	97	1008	46.7	30	1.31	13,504	9.0	5.9
Tue	84	1007	23.2	30	1.38	11,666	8.2	12.3
Wed	96	1161	35.9	35	1.28	12,858	8.6	2.1
Thu	29	472	6.1	15	1.13	4179	5.4	0.9
Fri	71	944	20.8	29	1.21	11,697	7.2	7.1
Average	75.4	918	26.5	27.8	1.26	10,781	7.7	5.7
Week 2								
Mo	81	1162	19.1	36	1.19	18,291	6.2	3.4
Tue	62	1028	13.6	31	1.49	20,539	3.1	6.5
Wed	66	1117	14.7	34	1.26	21,839	4.0	4.7
Thu	25	352	5.9	11	1.36	9058	6.7	1.4
Fri	78	1111	19.0	33	1.22	28,606	6.0	4.1
Average	62.4	954	14.5	29.0	1.31	19,667	5.2	4.0

Data: 10 planning days from case partner.
[a] Total number of euro pallets ordered.
[b] Runtime in seconds.
[c] Driven distance over direct distance in % (20% allowed).
[d] Cost savings obtained in comparison to the company's solution.

Table 5
Case study: Comparison of different relative detour limits, average over weekdays.

Week 1						
Detour Δ^r	Objective Total Costs	No. of tours	Capacity utilization[a]	Avg. stops per tour	Change tot. distance[b]	Detour[c]
0.10	+1.3%	29.0	93.1%	2.53	−0.4%	4.0%
0.15	+0.6%	28.5	94.6%	2.57	+0.3%	5.4%
0.20	baseline	27.9	96.4%	2.62	baseline	7.7%
0.30	−0.7%	27.7	97.0%	2.64	+5.4%	12.5%
0.40	−1.0%	27.6	97.2%	2.64	+10.8%	16.5%
1.00	−1.6%	27.6	97.2%	2.64	+32.4%	37.5%
Week 2						
0.10	+1.7%	30.1	92.4%	2.07	+0.9%	3.9%
0.15	+0.6%	29.5	94.4%	2.12	+1.5%	5.2%
0.20	baseline	29.0	96.4%	2.17	baseline	5.2%
0.30	−0.4%	29.0	96.3%	2.16	+0.2%	6.7%
0.40	−0.8%	28.8	97.0%	2.18	+0.3%	8.2%
1.00	−1.5%	28.9	96.7%	2.17	+11.0%	19.1%

Data: 10 planning days from case partner; Values: All values are averages over five different planning days and ten runs of ALPHA per day.
[a] Average change in utilization of vehicle capacity calculated using accumulated load and the number of tours for each day.
[b] Average change in total travel distance calculated using accumulated travel distance of all tours for each day.
[c] Average driven detour in % of direct distance calculated using the accumulated travel distances and the accumulated direct distances of all tours for each day.

Tariff structures, vehicles and financials. We derive the zone tariff-specific parameters from our case study. Specifically, we apply a 5% volume discount, with five concentric zones of constant arc-radius differences around a DC. We specify the absolute detour limit with $\Delta^a = 0.1$ for the following experiments as a base value. This absolute limit is used in related benchmarks. We set all other tour limits relative to this value. All data is available at https://github.com/NikTuma/CVRPZT.git.

5.3. Computational efficiency of ALPHA

We assess the runtime and solution quality of our matheuristic in comparison to the BAC algorithm proposed by Tuma et al. (2026), as it is the only available benchmark that solves general VRP-ZT instances with volume discounts and an absolute detour limit. Using the BAC as a benchmark enables comparisons with optimal solutions. We compare the two approaches for small and medium-sized instances. In a subsequent analysis we demonstrate ALPHA's scalability using the large-scale (Uchoa et al., 2017)-based instances.

Comparison of ALPHA with BAC. Tuma et al. (2026) proposed a total of 81 benchmark instances for the VRP-ZT. We extend these instances with the 100-customer versions by Solomon (1987) for a total of 90 instances. The instances include three cluster types, four sizes, three different customer location settings, and three demand scenarios ($d \in [1; 34]$). We run all instances on the identical computational setup. Table 6 summarizes the performance. ALPHA achieves the same results as the BAC for up to 60 customers and thus solves these instances to optimality. Furthermore, when solving the larger instances, the matheuristic scales almost linearly while the runtimes of the BAC rise exponentially. In the 100-customer setting, the BAC is able to find feasible solutions for eight of the nine instances within the time limit, achieving an average optimality gap of 2.8%, though without proven optimality. ALPHA improves each BAC solution (found after 24 h) by 0.2% on average, again showing a high solution quality with near-optimal, potentially optimal solutions.

Analysis of large instances. To evaluate the computational effort on larger instances, we adapt the instances proposed by Uchoa et al. (2017) to the VRP-ZT. We specifically accommodate the cost function,

Table 6
Computational efficiency for small and medium instances.

Customers	Runtime (seconds)[a]		Object. gap (%)		Improvement (%)[b]
	ALPHA[c]	BAC[d]	ALPHA	BAC	ALPHA over BAC
30	8	8	0.0	0.0	0.0
45	11	127	0.0	0.0	0.0
60	14	11,134	0.0	0.0	0.0
100	25	>86,400	–[e]	>2.8	0.2

Data: 90 Solomon-based instances from Tuma et al. (2026) or self-adapted (100 customers).
[a] Time limit 24h (86,400 s).
[b] Improvement in objective function value (in %).
[c] Avg. total runtime, 10 runs per instance.
[d] Avg. total runtime, incl. master- and subproblem.
[e] No proven optimal solutions for comparison.

Table 7
Impact of individual framework components.

Customers	Average objective difference (in %)				Maximum objective difference (in %)			
	Sweep	& SPM (init)	& ALNS	& SPM (ex-post)	Sweep	& SPM (init)	& ALNS	& SPM (ex-post)
100–199	baseline	−0.2	−11.8	−0.5	baseline	−0.6	−14.4	−0.9
500–599	baseline	−0.4	−12.8	−1.3	baseline	−0.6	−14.2	−1.8
900–1000	baseline	−0.5	−13.3	−1.7	baseline	−0.6	−14.7	−2.2

Data: 35 Uchoa et al. (2017)-based instances; Values: All values are stated as relative improvements (in %) of a setting including the specified component and all previous components compared to a setting including only the previous components.

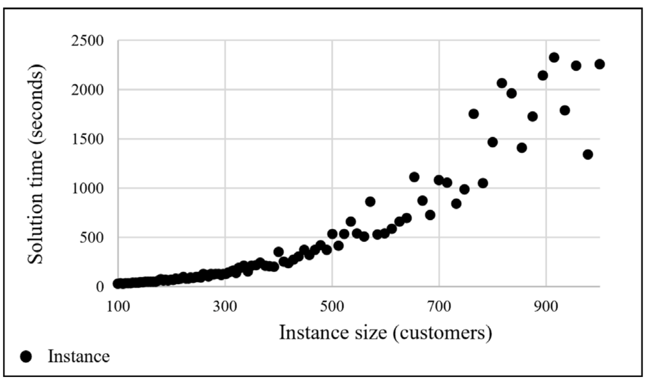


Fig. 5. Runtime with large instances.
Data: Uchoa et al. (2017)-based instances.

the demand data, and the vehicle capacity according to the instances from Tuma et al. (2026). Fig. 5 highlights the runtimes of our framework for the large instances and depicts the scaling of the runtimes with an increasing number of customers. The runtimes range from 25 s (100 customers) to 2256 s (1000 customers). This demonstrates that the matheuristic is computationally efficient, highly scalable, and capable of solving large instances within practical timeframes, highlighting its strong real-world applicability.

Algorithmic components. We evaluate the performance of the (i) matheuristic's components, and the (ii) adaptive removal value and various removal and insertion operators, using 35 representative large instances across three size buckets.

(i) Components and solution quality. Table 7 summarizes the contribution of each major algorithmic component. We begin with the baseline solution, which involves the Sweep as the first component, and then add the subsequent components sequentially, as shown in Fig. 4. For each added component, we report the relative improvement in objective value compared to the case where only the previous components are applied. Applying the initial SPM, denoted as *SPM (init)*, to the Sweep solution yields an average improvement of 0.4%. The ALNS component achieves the most substantial reduction in objective value, with an average improvement of 12.7%. Finally, the *SPM (ex-post)* step further refines the solution, yielding an additional average improvement of 1.3%. The improvements are more pronounced for larger instance sizes. The performance of both SPMs depends more on the instance size than the ALNS results. Our findings highlight the contribution of individual solution components to the overall framework and, in particular, demonstrate the added value of the post-optimization step. Table B.15 in the Appendix provides further insights into the runtime contributions of the components.

(ii) Operator statistics and adaptive removal value. We developed and tested 12 removal and insertion operators, evaluating their effectiveness by successively removing each from the full set and observing the resulting changes in objective and runtime. This enabled us to identify three operators without a positive impact. The first discarded destroy operator removes customers from tours with low demand first. The second discarded destroy operator removes customers from tours with high (de)tour first. The last discarded destroy operator removes customers from tours with a high difference between the highest and lowest zones. After excluding these destroy operators, the nine operators presented in the final algorithm remain (see Section 4). Table 8 reports the operators' performance. We run the subset of 35 Uchoa et al. (2017)-based instances ten times each and report the percentages of new best, improving, non-improving but accepted and rejected solutions. We sort Table 8 by the percentage of new best solutions in descending order. The best-performing removal operators are Concatenated Random Removal and Random Removal. Both find similarly often new best solutions (1.5% to 1.6%), but the Concatenated Random Removal operator results in significantly more improved solutions (32.3% vs. 24.3%). The best-performing insertion operators are those tailored for the VRP-ZT: Greedy Distance Insertion (1.7% of new best solutions) and Zone Demand Insertion (1.3% of new best solutions). Random Insertion shows a relatively high share of new best solutions (1.3%) compared to a low share of improving solutions (4.1%).

In the destruction phase, we apply an adaptive removal value to determine how many customers to remove from the current solution. This value is updated after each iteration based on the observed performance (see Section 4.3.3). Although the adaptive mechanism has only a minor effect on the objective value, it reduces runtime by an average of 1.7%, with a maximum improvement of 18.0%.

Table 8
Analysis of operator performance, representative large instances.

Removal operators	% Best	% Better	% Accept	% Reject
• Concatenated Random Removal	1.6	32.3	28.8	37.2
• Random Removal	1.5	24.3	32.6	41.7
• Similar Zone Removal	0.8	23.7	35.4	40.0
Insertion operators				
• Greedy Distance Insertion	1.7	32.9	31.7	33.7
• Zone Demand Insertion	1.3	27.9	33.7	37.1
• Random Insertion	1.3	4.1	22.5	72.2
• Greedy Insertion	1.2	29.7	33.1	36.1
• Greedy Demand Insertion	0.8	15.5	32.1	51.7
• Random Zone Insertion	0.8	6.7	24.8	67.7

Data: 35 Uchoa et al. (2017)-based instances.

Table 9
Relation analysis for (de)tour limits, impact on total costs.

Abs. detour	Rel. detour		Total distance	
Δ^a	Δ^r	TC Diff.[a]	Δ^k	TC Diff.[a]
0.05	0.05	0.78%	**1.00**	**1.47%**
	0.06	0.46%	1.01	1.64%
	0.07	0.24%	1.02	1.66%
	0.08	**0.13%**	1.03	1.68%
	0.09	0.20%	1.04	1.71%
	0.10	0.31%	1.05	1.72%
0.10	0.10	0.58%	**1.00**	**0.75%**
	0.13	0.23%	1.01	0.87%
	0.14	0.15%	1.02	0.89%
	0.15	**0.12%**	1.03	0.91%
	0.16	0.16%	1.04	0.94%
	0.17	0.23%	1.10	1.03%
0.20	0.20	0.37%	**1.00**	**0.28%**
	0.29	0.09%	1.01	0.33%
	0.30	0.07%	1.02	0.33%
	0.31	**0.06%**	1.03	0.33%
	0.32	0.07%	1.04	0.33%
	0.33	0.08%	1.20	0.52%

Data: 35 Uchoa et al. (2017)-based instances.
[a] Deviation in objective value to absolute limit Δ^a in %.

5.4. Managerial insights on (de)tour limits

(De)tour limits govern the customer clustering to tours and are thus a fundamental design element of every zone-based tariff. Despite this central role, the relationship between the different (de)tour variants and their impact on vehicle utilization, route efficiency, and operational costs has not been analyzed. ALPHA provides the first approach that can simultaneously consider total distance, absolute detour, relative detour, and any feasible combination of these limits. This enables us to (a) relate the limits and their impact with one another, (b) investigate the impact of limit magnitudes on the routing, (c) identify which limits are most effective under various settings, and (d) ultimately assess the utilization of detour limits.

(a) Relationship of (de)tour limit types and their impact. We first study the relationships between the different limits. The advanced, modular design of ALPHA and our VRP-ZT formulation enables the flexible application of absolute, relative, and distance-based (de)tour limits. While all limits can be found in practice, a relative limit theoretically allows for higher detours with increasing tour lengths, whereas an absolute limit always allows the same detour, regardless of tour length. Limiting the detour instead of the total distance is more flexible, as the minimum distance limit must at least equal the distance to the farthest customer from the DC. These differences raise the question of whether limits can be calibrated to enable comparability across tariffs – even when LSPs apply different limit rules. In other words, we can compare different tariffs independent of potentially heterogeneous limit types. We therefore first relate the limits to one another and analyze which limit values have a comparable impact on total costs and tours. This allows us to tune the different limits and determine whether they are interchangeable across applications to still achieve comparable objectives. The absolute limit ($\Delta^a = \{0.05; 0.1; 0.2\}$) serves as a reference for this purpose, and we vary the other two limits to derive values for Δ^r and Δ^k.

Table 9 shows changes in total costs compared to the baseline obtained with Δ^a for different values of Δ^r and Δ^k. The analysis reveals a linear relationship between absolute and relative detour limits. To achieve equivalent objectives, we require relative limits with about 50% higher values (0.05 vs. 0.08; 0.1 vs. 0.15 and 0.20 vs. 0.31). Simply applying the same value to all detours increases solution deviations, as it neglects the different characteristics of the two detour limits. However, the differences in objective value remain relatively small.

Comparing the absolute detour with the total distance limit yields higher deviations, even for the smallest feasible value of $\Delta^k = 1.0$. The differences in the results between the two limits decrease as the absolute detour value increases. However, distance limits alone do not provide adequate constraints for VRP-ZT applications, as they imply that distant customers must always be visited by a dedicated vehicle.

Our results demonstrate that Δ^a and Δ^r can be calibrated to yield almost identical total costs. To further understand these effects, we examine the resulting solution structures in more detail. Table 10 details the effects on costs, tours, and distances for related settings, using $\Delta^a = 0.1$ as reference. We include a baseline without any (de)tour limitation for comparison. All three (de)tour limit types – relative, absolute, and distance-based – lead to a clear reduction in total distance driven (−15% to −23%) compared to the unrestricted case, confirming their effectiveness in constraining excess travel. This comes at a modest

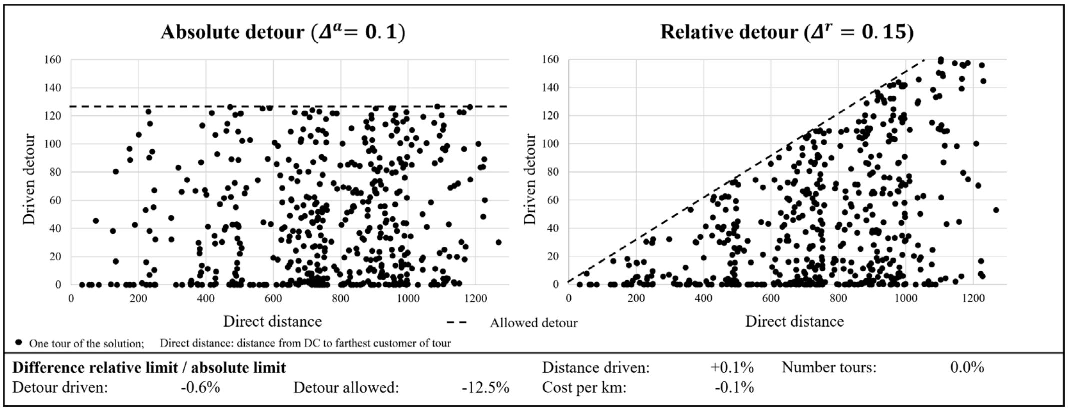


Fig. 6. Comparison of absolute and relative detour limits, in relation to direct distance. Data: Largest Uchoa et al. (2017)-based instance.

Table 10
Implications of different (de)tour limits on routing.

Detour	Runtime (in sec.)	Change (in %), compared to no (de)tour limit				
		Detour used[a]	Total costs	No. of tours	Total distance[b]	Detour driven
$\Delta^r = 0.15$	507	34.3%	+0.9	+0.9	−22.5	−86.5
$\Delta^a = 0.10$	496	32.5%	+0.9	+0.8	−22.6	−86.1
$\Delta^k = 1.00$	398	25.7%	+0.4	+0.4	−15.5	−58.1

Data: 35 Uchoa et al. (2017)-based instances.
[a] Detour driven divided by detour allowed.
[b] Total distance of all tours from DC to last customer.

Table 11
Impact of different detour limit levels.

Absolute detour[a] Δ^a					Relative detour[a] Δ^r					Δ^r/Δ^a			
Limit	TC	Dist.	Det.	Tours	Limit	TC	Dist.	Det.	Tours	TC	Dist.	Det.	Tours
0.050	+0.6	−1.3	−51.1	+0.6	0.08	+0.5	−1.3	−47.5	+0.7	+0.1	+0.2	+4.1	+0.1
0.075	+0.3	−0.7	−26.1	+0.3	0.12	+0.2	−0.7	−21.0	+0.3	0.0	+0.2	+3.7	0.0
0.100	base	base	base	base	0.15	base	base	base	base	+0.1	+0.2	−3.0	0.0
0.150	−0.3	+1.8	+47.7	−0.3	0.25	−0.4	+1.9	+58.0	−0.4	0.0	+0.3	+3.8	−0.1
0.200	−0.4	+3.6	+91.8	−0.5	0.31	−0.5	+3.2	+90.7	−0.5	0.0	−0.1	−3.6	0.0

Data: 35 Uchoa et al. (2017)-based instances.
[a] Difference to baseline in %; TC: change in total costs; Dist.: change in total travel distance; Det: change in detour distance; Tours: change in number of tours.

cost increase of less than 1%, driven by a slightly higher number of tours (+0.4 to +0.9%). The (de)tour limits lead to cost increases for the shipper, which in turn reduces travel distance for the LSP. The relative ($\Delta^r = 0.15$) and absolute ($\Delta^a = 0.10$) limits yield nearly identical numbers of tours, confirming our earlier observation of their near-linear relationship. Both limits produce similar runtimes, costs, and detour usage levels, suggesting that, if tuned correctly for the specified application and data, they can be used interchangeably. The distance-based limit ($\Delta^k = 1.0$) is less restrictive overall, resulting in smaller reductions. These findings demonstrate that absolute and relative detour limits effectively reduce travel distances with minimal trade-offs in terms of cost. The choice between related absolute and relative values can be made flexibly, as their operational impact is nearly equivalent on average.

Fig. 6 further illustrates a more detailed detour analysis for both detour limits that yields comparable total costs (using an exemplary instance). Both plots show the driven detours in relation to the distance from the DC to the farthest customer on a tour. For the relative limit, the maximum detour driven increases proportionally with the direct distance, while for the absolute limit, the maximum driven detour is independent of the direct distance. Absolute detour limits allow larger detours for customers near the DC, whereas relative detour limits allow larger detours for customers farther away. Despite this structural difference, when choosing equivalent detour limit values, both limits yield nearly identical tours and distance driven (see Table 10).

We further analyze ALPHA's ability to combine different limits and examine how combined limits affect solutions. Fig. 7 illustrates the impact of the hybrid configurations. Applying two limits simultaneously reduces the solution space, while having only a minor effect on total distance and costs per distance. The combination of relative and absolute limits behaves like a relative limit up to the maximum detour permitted by the absolute limit – represented by the inflection point – beyond which it follows the behavior of the absolute limit. The combination of a relative and a distance limit behaves like a relative limit until the allowed detour equals the gap between the distance limit and the farthest customer distance; beyond this point, the permitted detour decreases rapidly and reaches zero when the direct distance meets the distance limit.

(b) Impact of (de)tour limit magnitudes. In the following, we examine how varying limit levels influence solutions, ultimately supporting the design of the tariff. Table 11 shows that both detour types exhibit similar trends: tighter limits result in slightly higher costs and number of tours, but decrease distance and detours to a greater extent. Conversely, larger limits lower costs and the number of tours, yet increase the distance and detours driven. Vehicle utilization inversely follows the number of tours, meaning tighter limits lead to lower vehicle utilization. The right columns highlight a near-linear correspondence between absolute and relative limits, confirming that either formulation can be applied interchangeably when calibrated appropriately.

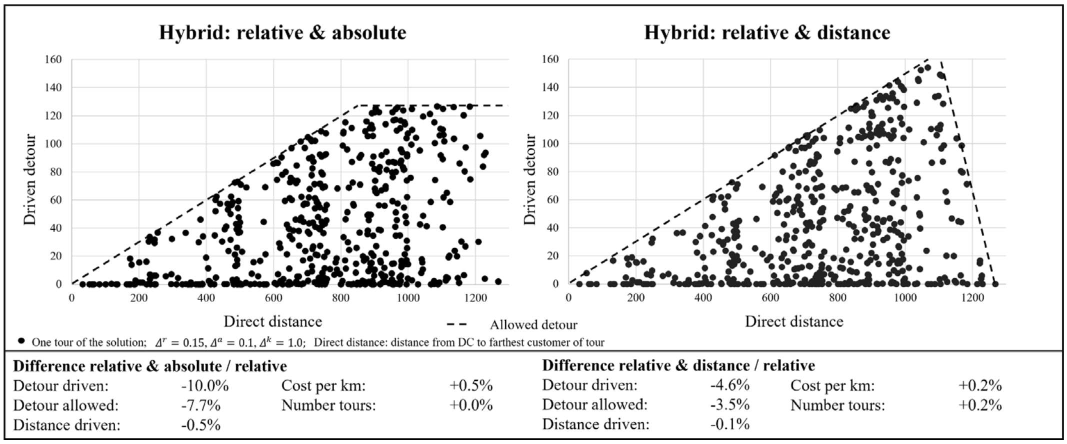
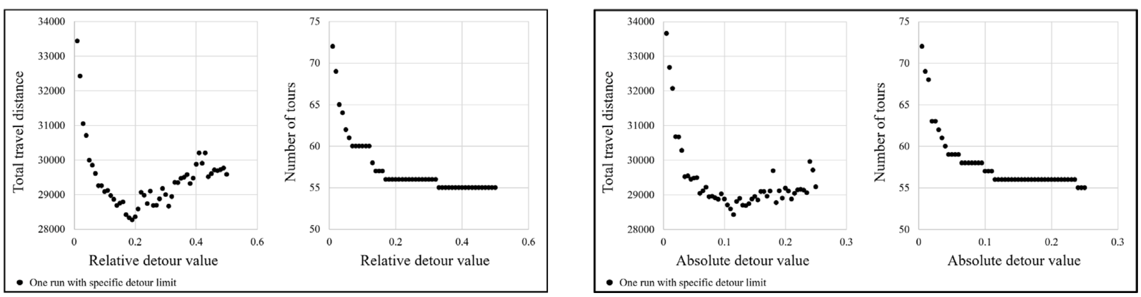


Fig. 7. Comparison of hybrid relative (de)tour limits. Data: Largest Uchoa et al. (2017)-based instance.



(a) **Relative detour limit** (b) **Absolute detour limit**

Fig. 8. Impact of detour values on total travel distance and number of tours. Data: Smallest Uchoa et al. (2017)-based instance.

The previous analysis indicates a trade-off between vehicle utilization and distance. We further explore this aspect by testing both detour limits with 50 values ($\Delta^r \in [0.01;0.5]$; $\Delta^a \in [0.005;0.25]$). Figs. 8(a) and (b) illustrate the relationship between detour values and total travel distance as well as the number of tours. Strict detour limits result in sharply increasing total travel distances for both limit types. There are more tours with lower vehicle utilization, as fewer customers can be bundled into one tour, resulting in inefficient routing due to overly restrictive limits. As the detour limits increase, the number of tours rapidly decreases, stabilizing around $\Delta^a = 0.1$ and $\Delta^r = 0.2$. The total travel distances are more pronounced when considering a relative detour. The distances follow a U-shaped pattern: they initially decrease as routing flexibility improves, but increase again once the limit allows excessive detours to outweigh consolidation benefits. This perfectly reflects the LSP's concern of extensive driving tours as soon as limits allow the shipper to exploit this and justifies the use of these limits. The total travel distances for the absolute limit do not exhibit the same effect but remain more stable above $\Delta^a = 0.2$, with some outliers showing increased distances. This suggests that a relative detour limit with its detour allowance proportional to the tour length should be chosen more carefully, particularly in customer settings allowing extensive driving distances. Overall, the results reveal a clear efficiency threshold: moderate detour allowances ($\Delta^r \simeq 0.2$; $\Delta^a \simeq 0.1$) yield the lowest total distance while maintaining a compact number of tours. Beyond these limits, additional detour flexibility offers diminishing or even adverse returns.

(c) Limit effects for changing settings. Spatial customer distributions and various demand scenarios typically have a significant impact on routing decisions. Understanding these effects in the context of zone tariffs is crucial, as real-world logistics networks often exhibit varying distributions. Table 12 presents the results for selected customer subsets with specific spatial distributions (C, R, RC) compared to the setting that includes all customers. As total costs, travel distances, and detour utilization depend on the spatial distribution, we compare them in relative terms.

The spatial distribution of customers also has a substantial impact on routing efficiency for VRP-ZT applications. Counterintuitively, randomly distributed customers achieve the lowest costs (−2.5 to −2.7%), shortest travel distances (−20%), and highest detour utilization (+3.7 to +5.9%), indicating that spatial diversity enhances tour consolidation. This finding contradicts the usual findings in the VRP literature, which attest to a higher potential for customer consolidation in clustered settings. A random (i.e., spatially diverse) customer distribution results in the VRP-ZT to lower costs and shorter distances per order unit. The increased detour utilization at the absolute limit further indicates that diversity enables more effective customer combinations than clustered distributions, in which customers are concentrated in a single area. In the randomly distributed cases, the zone tariff offers greater flexibility in assigning customers to tours as long as detour limits are met, meaning that customers within the same zone do not necessarily need to be geographically close to one another. As the absolute detour does not adapt to customer distances (clustered and short distances vs. random and longer distances) but relies on a fixed value across all spatial distributions, the detour utilization is higher than that of the adaptive relative detour. This also shows in the slightly stronger decrease in utilization of the relative limit for the clustered distribution. The analysis confirms that customer spatial distribution has a substantial impact on economics, whereas the choice between absolute and relative detour limits – when set to values that yield comparable total costs – has only a minor influence.

Clustering options also depend on order sizes. Table 13 compares the effect of smaller order sizes. We consider four demand scenarios

Table 12
Impact of spatial distribution of customers, relative to baseline with all customers included.

	Absolute detour, $\Delta^a = 0.1$			Relative detour, $\Delta^r = 0.15$		
Customer locations	Costs / ord. unit	Distance / ord. unit	Detour utilization	Costs / ord. unit	Distance / ord. unit	Detour utilization
Clustered (C)	+1.9%	+20.9%	−6.8%	+1.8%	+21.1%	−7.6%
Random (R)	−2.7%	−20.0%	+5.9%	−2.5%	−20.1%	+3.7%
Random Clustered (RC)	+0.1%	−10.5%	+3.3%	+0.1%	−10.7%	+3.3%

Data: 35 Uchoa et al. (2017)-based instances.

Table 13
Impact of varying demand distribution, relative (in %) to baseline with scaling factor 1.0.

Order size	Ordered units	Tours	Costs / ord. unit	Distance / ord. unit	Detour / ord. unit	Detour utilization	Capacity utilization
Absolute detour $\Delta^a = 0.1$							
0.25	−71.4	−70.8	+2.8	+24.9	+164.0	+160.4	−2.0
0.5	−48.6	−49.4	−0.9	+8.8	+110.8	+114.1	+1.7
0.75	−22.9	−24.3	−1.2	+2.8	+46.2	+48.7	+1.8
Relative detour $\Delta^r = 0.15$							
0.25	−71.4	−71.0	+3.1	+28.9	+193.6	+143.9	−2.0
0.5	−48.6	−49.4	−0.7	+9.9	+112.4	+103.1	+1.6
0.75	−22.9	−24.3	−1.1	+2.7	+42.7	+41.7	+1.8

Data: 35 Uchoa et al. (2017)-based instances.

using a uniform distribution over the intervals [1,34], [1,26], [1,17], and [1,9], corresponding to scaling factors of 1.00, 0.75, 0.50, and 0.25, respectively. As order sizes decrease, the number of tours declines markedly, indicating higher customer consolidation on tours and improved vehicle utilization. However, distance per order unit increases, and detours rise substantially, reflecting longer routes required to consolidate smaller loads. Smaller order sizes and the same number of customer locations to be visited result in higher efforts for the LSP. Despite these structural shifts, shippers' costs per order unit decrease slightly when order sizes decline by 50%, suggesting that the volume discount embedded in the zone tariff effectively rewards a higher vehicle utilization in these scenarios. When dealing with very small orders (a factor of 0.25), more tours are required, as the detour limit is stricter than the vehicle capacity. This results in a significantly higher detour utilization and increasing costs per order unit.

(d) Utilization of detour allowance. Our previous findings indicate that the driven detour in practical settings depends on order sizes. Order sizes determine the consolidation possibilities for the routing, which is constrained by vehicle capacity and detour limits. We analyze this relation in more detail by examining the development of detour utilization as order sizes increase. We generate scenarios in which customer order sizes are uniformly distributed over the intervals $[1, 5]$ and $[1, 34]$. Fig. 9 illustrates the development. The results show a clear decreasing trend: as order sizes increase, detour utilization declines. This indicates that larger customer orders limit the potential to combine multiple stops within a single tour, causing vehicle capacity to become binding earlier. Consequently, tours have fewer customers and are less flexible, resulting in lower detour utilization. A further analysis with a relative limit shows similar results.

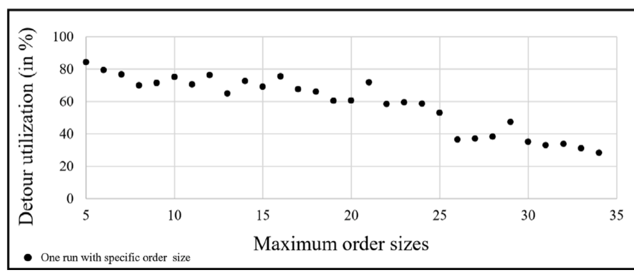


Fig. 9. Detour limit utilization (absolute limit) across varying customer order sizes.
Data: Largest Uchoa et al. (2017)-based instance.

6. Conclusion and future areas of research

Conclusion. The VRP-ZT addresses a class of routing problems of high practical relevance in today's logistics applications. Applying zone tariffs creates complex trade-offs among cost efficiency, service quality, and fairness between shippers and LSPs. Despite its practical importance, this problem type has received little attention in the literature. Our work addresses this gap and introduces the first scalable approach for the VRP-ZT, enabling the efficient solution of large-scale instances while accommodating various (de)tour limit configurations flexibly. We propose the matheuristic framework ALPHA, integrating a tailored ALNS with customized operators and enhancing it through hybridization with exact optimization steps. We demonstrate the practical applicability of our model and solution approach in a real-world case study, and show that ALPHA optimally solves related benchmark instances.

Beyond methodological advances, this study deepens understanding of how different (de)tour limits, as a defining element of zone-based tariffs, relate to and influence routing efficiency when zone tariffs are applied. We analyze the prevailing (de)tour limit types and provide valuable insights for both researchers and practitioners seeking to design cost-efficient, operationally feasible routing strategies for such logistics systems. We show that absolute and relative detour limits can be calibrated to produce comparable results, thereby enabling a direct and meaningful comparison of LSP tariffs regardless of which limit is applied. Absolute and relative detour formulations perform similarly, resulting in distance reductions exceeding 20% compared to unrestricted routing. However, setting (de)tour limits too strictly can have adverse effects, as they prevent efficient customer combinations and increase overall travel distances. Hence, careful calibration of the detour value is crucial for achieving optimal performance. In this sense, we identify an efficiency threshold with moderate (de)tour limit values (absolute detour of 0.1; relative detour of 0.2) that provides actionable guidance for designing compact, cost-effective tours and offers practical decision support for configuring zone tariff systems.

Future areas of research. Our adaptable and scalable matheuristic establishes a foundation for future research. The flexibility of ALPHA allows the integration of additional costing criteria (e.g., the number of stops), enabling its adaptation to additional applications. Cross-domain applications of zone-based tariffs, such as urban logistics and reverse logistics, provide promising opportunities to test the robustness and scalability of ALPHA. Further research could also incorporate additional problem features that reflect practical routing complexities, including time windows, multiple depots, split deliveries, heterogeneous vehicle fleets, and multi-tier or collaborative distribution systems involving several LSPs. The design and analysis of zone-based tariff structures in VRPs is another promising area of research. Future studies could systematically examine how different spatial zoning schemes, pricing functions, and discount mechanisms influence routing efficiency, cost distribution, and market fairness. This can provide the basis for empirically grounded design principles and managerial guidelines for equitable and economically sustainable tariff systems. The proposed modeling framework also opens opportunities to study collaborative and competitive coordination mechanisms between logistics actors. Incorporating game-theoretic or auction-based formulations could help identify stable and efficient allocation schemes under zone-based pricing.

From a methodological perspective, promising avenues for further enhancing computational efficiency and adaptability emerge. One option is to develop a hybrid detour feasibility check that dynamically selects between exact and heuristic evaluation methods. Such a dynamic mechanism could substantially reduce computational effort for tours with many stops, addressing scalability challenges posed by open TSPs that must be solved for each customer cluster. The ALNS component could be adapted to allow for infeasible interim solutions. Allowing infeasible solutions during the ALNS broadens the search and could improve solutions, but also requires tailored mechanisms to control the degree of infeasibility and ensure a feasible final solution. Furthermore, adapting exact approaches such as the BAC into a heuristic framework by relaxing optimality guarantees could offer a flexible trade-off between precision and computational speed. As for exact solution methods, column generation could be a promising approach for solving larger instances to proven optimality, allowing further assessment of ALPHA's performance.

CRediT authorship contribution statement

Niklas Tuma: Writing – original draft, Visualization, Methodology. **Manuel Ostermeier:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Alexander Hübner:** Writing – review & editing, Supervision, Formal analysis, Data curation, Conceptualization.

Acknowledgments

We would like to acknowledge our industry partner for providing real-world data, sharing insights on the planning problem, and actively collaborating throughout the development of this paper. We thank the Editor and the three anonymous referees for their valuable suggestions and constructive feedback, which greatly improved the quality of this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Case study: sensitivity analysis

See Table A.14.

Table A.14
Case study: Comparison of different relative detour limits, minimum/median/maximum over weekdays.

Week 1

Detour	Objective Total Costs			No. of tours			Capacity utilization[a]			Avg. stops per tour			Change tot. distance[b]			Detour[c]		
Δ^r	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max
0.10	+0.3%	+1.3%	+1.3%	15.0	31.4	36.5	92.5%	92.6%	94.3%	1.93	2.63	3.03	−0.3%	+0.4%	+2.0%	3.3%	4.1%	4.9%
0.15	+0.3%	+0.5%	+0.5%	15.0	31.0	35.3	92.5%	95.5%	96.7%	1.93	2.71	3.13	−1.9%	+2.0%	−1.0%	3.7%	6.1%	6.7%
0.20	baseline			15.0	30.0	35.0	92.5%	97.5%	98.7%	1.93	2.74	3.19	baseline			5.4%	8.2%	9.0%
0.30	−0.0%	−0.7%	−0.9%	15.0	30.0	35.0	92.5%	97.6%	99.2%	1.93	2.74	3.18	+0.7%	+7.5%	+6.3%	5.6%	13.7%	15.2%
0.40	−0.0%	−0.7%	−1.5%	15.0	30.0	35.0	92.5%	98.2%	99.2%	1.93	2.74	3.21	+0.8%	+10.7%	+12.9%	5.9%	18.5%	21.6%
1.00	−1.0%	−1.2%	−2.2%	15.0	30.0	35.0	92.5%	98.2%	99.2%	1.93	2.74	3.21	+7.7%	+32.8%	+39.9%	13.3%	42.8%	48.9%

Week 2

Detour	Objective Total Costs			No. of tours			Capacity utilization[a]			Avg. stops per tour			Change tot. distance[b]			Detour[c]		
Δ^r	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max
0.10	+2.4%	+2.0%	+1.1%	12.0	34.0	36.0	86.3%	92.0%	96.6%	1.88	2.08	2.25	+2.9%	+2.1%	+0.2%	3.3%	3.5%	5.2%
0.15	+2.4%	+0.8%	+0.2%	11.5	34.0	36.0	90.0%	94.9%	96.6%	1.94	2.17	2.29	+4.2%	+2.5%	+1.6%	2.7%	5.2%	8.2%
0.20	baseline			11.0	33.0	36.0	94.1%	96.6%	99.0%	1.94	2.25	2.36	baseline			3.1%	6.0%	6.7%
0.30	−0.0%	−0.7%	−0.4%	11.0	33.2	35.0	94.1%	96.6%	98.4%	1.94	2.27	2.35	+2.8%	+2.7%	−2.5%	4.4%	6.2%	9.0%
0.40	−0.0%	−1.1%	−1.0%	11.0	33.0	35.0	94.1%	97.5%	99.0%	1.94	2.27	2.36	+3.8%	+1.6%	−4.2%	5.8%	8.3%	10.7%
1.00	−1.0%	−1.3%	−2.1%	11.0	33.5	35.0	94.1%	97.5%	97.6%	1.94	2.27	2.33	+9.9%	+14.3%	+3.0%	14.0%	15.7%	33.3%

Data: 10 planning days from case partner; Values: Minimum/Median/Maximum over five different planning days. Each value is also an average over ten runs of ALPHA per day.
[a] Average change in utilization of vehicle capacity calculated using accumulated load and the number of tours.
[b] Average change in total travel distance calculated using accumulated travel distance of all tours.
[c] Average driven detour in % of direct distance calculated using the accumulated travel distances and the accumulated direct distances of all tours.

Table B.15
Runtime contribution of individual framework components.

	Average runtime contribution (in %)			
Customers	Sweep	SPM (init)	ALNS	SPM (ex-post)
100–199	0.8	0.1	95.2	3.9
500–599	1.0	0.1	93.5	5.4
900–1000	0.9	0.1	91.5	7.5

Data: 35 Uchoa et al. (2017)-based instances.

Appendix B. Algorithmic components

In addition to comparing the contribution of each major algorithmic component to solution quality (see Table 7), we compare the runtime contributions of these components. Table B.15 summarizes the results.

References

Alicke, K., Rexhausen, D., & Seyfert, A. (2016). Supply chain 4.0 in consumer goods. McKinsey & Company, URL: https://www.mckinsey.de/~/media/McKinsey/Industries/Consumer%20Packaged%20Goods/Our%20Insights/Supply%20Chain%204%200%20in%20consumer%20goods/Supply-Chain-4-0-in-consumer-goods-vf.pdf. (Accessed: 09 October 2025).

Business Research Insights (2025). *Logistics Outsourcing Market to Hit USD 103.55 Billion By 2035*: *Technical Report ID: 108890*, Business Research Insights, URL: https://www.businessresearchinsights.com/market-reports/logistics-outsourcing-market-108890. (Accessed: 09 October 2025).

Carlsson, J. G., Lim, S. F. W. T., Liu, S., Yu, H., Arntong, W., & Tan, E. H. (2025). Redesigning zoning systems for equitable and efficient last-mile delivery at ninja van. *INFORMS Journal on Applied Analytics*, *55*(5), 412–423.

Ceschia, S., Di Gaspero, L., & Schaerf, A. (2011). Tabu search techniques for the heterogeneous vehicle routing problem with time windows and carrier-dependent costs. *Journal of Scheduling*, *14*(6), 601–615.

Ceselli, A., Righini, G., & Salani, M. (2009). A column generation algorithm for a rich vehicle-routing problem. *Transportation Science*, *43*(1), 56–69.

DHL Express (2025). Service & rate guide 2025: Zones and rates. https://mydhl.express.dhl/content/dam/downloads/us/en/rate-guide/service_and_rate_guide_us_en_2025.pdf.coredownload.pdf. (Accessed: 06 November 2025).

Elicit Technology (2025). Zone-based freight tariffs. https://elicittechnology.com/glossary/zone-based-freight-tariffs/. (Accessed: 06 November 2025).

Euro-CASE (2001). Freight logistics and transport systems in europe: Trends in the location of European industry and its interaction with logistics and transport. https://www.euro-case.org/wp-content/uploads/2019/07/Eurocase/PDF/Freight_Final-Report.pdf. (Accessed: 06 November 2025).

FleetX (2025). Zone pricing in logistics. https://blog.fleetx.io/zone-pricing-in-logistics/. (Accessed: 06 November 2025).

Fontaine, P. (2022). The vehicle routing problem with load-dependent travel times for cargo bicycles. *European Journal of Operational Research*, *300*(3), 1005–1016.

François, V., Arda, Y., & Crama, Y. (2019). Adaptive large neighborhood search for multitrip vehicle routing with time windows. *Transportation Science*, *53*(6), 1706–1730.

FreightAmigo (2025). Understanding zone rates: A key concept in freight pricing. https://www.freightamigo.com/all-promotion-and-news/understanding-zone-rates-a-key-concept-in-freight-pricing/. (Accessed: 06 October 2025).

Future Market Insights (2025). Logistics outsourcing market analysis - growth & forecast 2025 to 2035. Future Market Insights, URL: https://www.futuremarketinsights.com/reports/global-logistics-outsourcing-market. (Accessed: 10 November 2025).

Gillett, B. E., & Miller, L. R. (1974). A heuristic algorithm for the vehicle-dispatch problem. *Operations Research*, *22*(2), 340–349.

Gosling, S., Li, J. D., Martinez, A., Miguel, M., & Perez, F. (2023). Digital logistics: Technology race gathers momentum. URL: https://www.mckinsey.com/capabilities/operations/our-insights/digital-logistics-technology-race-gathers-momentum. (Accessed: 09 October 2025).

Held, M., & Karp, R. M. (1962). A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, *10*(1), 196–210.

Hübner, A., Hense, J., & Dethlefs, C. (2022). The revival of retail stores via omnichannel operations: A literature review and research framework. *European Journal of Operational Research*, *302*(3), 799–818.

Khodabandeh, E., Snyder, L. V., Dennis, J., Hammond, J., & Wanless, C. (2021). C. H. Robinson uses heuristics to solve rich vehicle routing problems. *INFORMS Journal on Applied Analytics*, *52*(2), 173–188.

Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

Kuhn, H., & Sternbeck, M. (2013). Integrative retail logistics: An exploratory study. *Operations Management Research*, *6*, 2–18.

Lehmann, J., & Winkenbach, M. (2024). A matheuristic for the two-echelon multi-trip vehicle routing problem with mixed pickup and delivery demand and time windows. *Transportation Research Part C: Emerging Technologies*, *160*, Article 104522.

Lindsey, K. A., Erera, A. L., & Savelsbergh, M. W. P. (2013). A pickup and delivery problem using crossdocks and truckload lane rates. *EURO Journal on Transportation and Logistics*, *2*(1–2), 5–27.

Mara, S. T. W., Norcahyo, R., Jodiawan, P., Lusiantoro, L., & Rifai, A. P. (2022). A survey of adaptive large neighborhood search algorithms and applications. *Computers & Operations Research*, *146*, Article 105903.

Masson International (2025). Understanding 3PL costs in europe | real costs explained. https://www.massoninternational.com/blog/third-party-logistics-costs-europe. (Accessed: 06 November 2025).

NTT Data (2024). *Third-Party Logistics Study - The Power of Partnership: Exploring Collaboration Through Data and Innovation - Results and Findings of the 28th Annual Study*: *Technical Report*, NTT DATA, URL: https://us.nttdata.com/en/-/media/content-hub/ch-gated/3PL-Study-2024.pdf. (Accessed: 09 November 2025).

Rooderkerk, R., de Leeuw, S., & Hübner, A. (2023). Advancing the marketing-operations interface in omnichannel retail. *Journal of Operations Management*, *69*(2), 188–196.

Roozbeh, I., Ozlen, M., & Hearne, J. W. (2018). An adaptive large neighbourhood search for asset protection during escaped wildfires. *Computers & Operations Research*, *97*, 125–134.

Ropke, S., & Pisinger, D. (2006a). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, *40*(4), 455–472.

Ropke, S., & Pisinger, D. (2006b). A unified heuristic for a large class of vehicle routing problems with backhauls. *European Journal of Operational Research*, *171*(3), 750–775.

SEKO Logistics (2023). How 3PLs calculate fees & rates. https://www.sekologistics.com/en/resource-hub/knowledge-hub/how-3pls-calculate-fees-and-rates-factors-that-impact-your-shipping-costs/. (Accessed: 06 November 2025).

ShipBob (2025). Shipping zones: What are they and how to use them. https://www.shipbob.com/ecommerce-shipping/shipping-zones/. (Accessed: 06 November 2025).

Soleilhac, G., Lehuédé, F., Medina, J., & Péton, O. (2026). Vehicle routing problem with multiple fleets and carrier limit. *EURO Journal on Transportation and Logistics*, *15*, Article 100177.

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

Statista (2024). *Revenue of the global third-party logistics (3PL) industry from 2010 to 2022, with forecasts until 2026*: *Technical Report*, Statista GmbH, URL: https://www.statista.com/statistics/418206/3pl-logistics-industry-revenue-worldwide/. (Accessed: 09 October 2025).

Thorsteinsson, E. S. (2001). Branch-and-check: A hybrid framework integrating mixed integer programming and constraint logic programming. In *Principles and practice of constraint programming—cP 2001: 7th international conference, CP 2001 paphos, cyprus, November 26–December 1, 2001 proceedings* (pp. 16–30). Springer.

Tordecilla, R. D., Copado-Méndez, P. J., Panadero, J., Martins, L. d. C., & Juan, A. A. (2021). An agile and reactive biased-randomized heuristic for an agri-food rich vehicle routing problem. *Transportation Research Procedia*, *58*, 385–392.

Tuma, N., Ostermeier, M., & Hübner, A. (2024). Optimal transportation planning for a do-it-yourself retailer with a zone-based tariff. *INFORMS Journal on Applied Analytics*, *54*(4), 312–328.

Tuma, N., Ostermeier, M., & Hübner, A. (2026). Capacitated vehicle routing problem with a zone tariff. *Networks*, *87*(2), 210–227.

Uchoa, E., Pecin, D., Pessoa, A., Poggi, M., Vidal, T., & Subramanian, A. (2017). New benchmark instances for the capacitated vehicle routing problem. *European Journal of Operational Research*, *257*(3), 845–858.

United Parcel Service (2023). *The Reality of Logistics Outsourcing*: *Technical Report*, United Parcel Service of America, Inc., URL: https://www.ups.com/assets/resources/supplychain/media/reality-of-logistics-outsourcing.pdf. (Accessed: 09 October 2025).

United Parcel Service (2024). *2024 UPS Rates Daily*: *Technical Report*, United Parcel Service of America, Inc., URL: https://www.ups.com/assets/resources/webcontent/en_US/Daily_US48_Final_CORRECTED.pdf. (Accessed: 09 October 2025).

Voigt, S. (2024). A review and ranking of operators in adaptive large neighborhood search for vehicle routing problems. *European Journal of Operational Research*, *318*(2), 393–410.

Voigt, S., Frank, M., Fontaine, P., & Kuhn, H. (2022). Hybrid adaptive large neighborhood search for vehicle routing problems with depot location decisions. *Computers & Operations Research*, *146*, Article 105856.

Waredock Logistics Glossary (2025). Zone pricing. https://www.waredock.com/glossary/zone-pricing/. (Accessed: 06 November 2025).

Wren, A., & Holliday, A. (1972). Computer scheduling of vehicles from one or more depots to a number of delivery points. *Journal of the Operational Research Society*, *23*(3), 333–344.