Transportation Research Part D 156 (2026) 105349

Contents lists available at ScienceDirect

Transportation Research Part D

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

From willingness to wait to sustainable impact: Smoothing e-commerce delivery volume peaks

Marith J. Zeelenberg, Onur A. Kilic, Paul Buijs *

University of Groningen, Faculty of Economics and Business, Department of Operations, Nettelbosje 2, 9747 AE Groningen, The Netherlands

A R T I C L E I N F O

Keywords:
E-commerce
Sustainability
Last-mile delivery
Consumer behavior
Life-cycle emissions
Demand peaks

A B S T R A C T

Willingness to wait is often promoted as a path towards more sustainable e-commerce deliveries, yet its impact depends on how firms use the operational flexibility it creates. We analyze two approaches: concentrating deliveries to minimize route distance versus spreading deliveries to smooth peaks and minimize fleet size. We adopt a strategic modeling approach that captures the flexibility created by customers' willingness to wait and uses continuous route-length approximations to estimate total route distance and fleet requirements. Using e-commerce delivery data from the Netherlands, we find that concentrating deliveries yields substantial distance savings only at very high willingness-to-wait levels and requires much larger fleets, whereas spreading deliveries can reduce fleet size with only a modest increase in distance. Once life-cycle emissions of the delivery fleet are considered, spreading deliveries can therefore lower CO_2 emissions, whereas concentrating deliveries can increase them.

1. Introduction

In the search for more sustainable e-commerce logistics, postponed delivery has gained traction as a promising solution (Zhang et al., 2023; Voigt et al., 2025). A widely shared belief is that if customers are willing to wait longer for their delivery, environmental benefits will naturally follow. Empirical studies suggest a broad willingness to wait, particularly when customers are informed of the resulting benefits (Buldeo Rai et al., 2019, 2021; Cauwelier et al., 2025; Dietl et al., 2024; Ignat and Chankov, 2020; Kokkinou et al., 2024; Nguyen et al., 2019). In this context, an implicit assumption—shared by practitioners and scholars—is that willingness to wait directly translates into more efficient vehicle use, fewer kilometers driven, and lower emissions. This promise has intuitive appeal. If deliveries are no longer rushed to meet the common next-day deadlines, there is more room to plan routes, consolidate deliveries, and reduce pressure on the parcel delivery system. Yet, this simple line of reasoning overlooks how firms actually use the flexibility created by customers' willingness to wait.

In practice, firms can use delivery flexibility in different ways. For instance, they could avoid visiting certain regions on specific days to reduce route distances and emissions (Muñoz-Villamizar et al., 2024). They could also concentrate deliveries on particular days and increase delivery density to improve routing efficiency by reducing the distance between deliveries (Muñoz-Villamizar et al., 2022; Voigt et al., 2025). Alternatively, they may spread deliveries across days or shifts to reduce peak delivery volumes and the required fleet size. Whether these approaches actually reduce environmental impact is uncertain, for two main reasons. First, the feasibility and benefits of any approach depend heavily on context, including spatial factors such as delivery density and temporal demand patterns. Moreover, willingness to wait is multidimensional: it varies in terms of how many customers are willing to wait

* Corresponding author.
E-mail address: p.buijs@rug.nl (P. Buijs).

https://doi.org/10.1016/j.trd.2026.105349
Received 31 December 2025; Received in revised form 22 March 2026; Accepted 25 March 2026
Available online 16 April 2026
1361-9209/

and for how long. Together, these factors determine whether an approach is even feasible. For instance, avoiding certain regions on specific days may improve route efficiency but would require very low delivery density or a high willingness to wait to divert all regional demand on those days. Second, different approaches interact with one another and involve trade-offs. For instance, smoothing delivery volume peaks can reduce stress for drivers, but may also decrease route efficiency. Conversely, concentrating deliveries on particular days can improve route efficiency, but may come at the expense of demand peaks that require additional resources such as vehicles and staff. These trade-offs obfuscate how increased flexibility driven by customers' willingness to wait actually yields sustainability benefits.

In this study, we assess postponed delivery at a strategic level and evaluate if and how willingness to wait can translate into reduced environmental impact. Prior research suggests that, without financial incentives, 51% to 68% of customers are willing to wait for their delivery (Buldeo Rai et al., 2021; Cauwelier et al., 2024; Ignat and Chankov, 2020), while 56% to 68% report being willing to wait when financial incentives are in place (Cauwelier et al., 2025). Here, financial incentives refer to, for example, reduced delivery fees or loyalty rewards. Non-financial incentives rely on behavioral and informational nudges (e.g., sustainability awareness or social influence) (Agatz et al., 2021; Buldeo Rai et al., 2021; Ignat and Chankov, 2020). All these estimates come from stated-preference surveys rather than observed customer behavior, and prior work highlights a broader intention-behavior gap: although many customers express environmental concerns, they often default to faster or cheaper options (Cauwelier et al., 2024). Given this uncertainty, we explore the full range of willingness to wait by varying (i) the share of deliveries that can be postponed from 0% to 100% and (ii) the postponement window, defined as the maximum number of days a customer order may be postponed, from one to five days.

To this end, we develop a mathematical model that captures how willingness to wait can be leveraged in e-commerce delivery operations, using continuous route-length approximation to estimate the resulting impacts over a delivery horizon. Our analyses show that postponed delivery can improve or worsen the environmental impact of the delivery operations, depending on how firms use the resulting flexibility. When customers' willingness to wait is used to concentrate deliveries, the total route distance decreases, resulting in lower operational emissions. However, it also concentrates more deliveries on the busiest day and thereby increases the number of delivery vehicles needed on that day. In our analysis, this offsets the operational emission reductions once the total life-cycle CO_2 emissions of the delivery fleet are considered. By contrast, when willingness to wait is used to spread deliveries across the delivery horizon, fewer delivery vehicles are needed on the busiest day, with only a modest increase in total route distance. In our analysis, this reduces the life-cycle CO_2 emissions of the delivery fleet.

2. Theoretical background

This section positions our study within the last-mile delivery literature and specifies the gap in the literature that our study intends to address.

2.1. Decision levels in last-mile delivery research

Last-mile delivery research looks at decisions from several angles, from short-term operational choices to longer-term planning, with different levels of detail and different aims. At the operational level, studies optimize decisions such as routing and dispatching, typically using order- and vehicle-level detail to improve metrics such as distance, time, fleet requirements, and cost (Arslan et al., 2019; Muriel et al., 2022; Perboli et al., 2021; Yang and Hyland, 2024). At the tactical level, research informs service design and demand-management choices that shape operations over weeks or months, such as the design of delivery systems and time-slot offerings, differentiated pricing or incentives, and service-region districting (Agatz et al., 2011; Bender et al., 2020; Klein et al., 2019; Lin et al., 2018; Stroh et al., 2022; Waßmuth et al., 2023). The output is typically policy guidelines or parameter settings rather than daily operational plans. At the strategic level, research evaluates system-level design choices and long-term implications, such as last-mile network, infrastructure and delivery fleet configurations, with an emphasis on robustness across spatial contexts, trade-offs across objectives, and the assessment of sustainability-oriented transitions (Arevalo-Ascanio et al., 2025; de Mello Bandeira et al., 2019; Janjevic et al., 2021; Stokkink and Geroliminis, 2025).

2.2. Postponement in delivery operations

Within this broader last-mile literature, research that explicitly models postponed delivery is scarce and predominantly operational in scope. Existing studies mainly embed customers' willingness to wait in short-horizon dispatch and vehicle-routing problems, asking how temporal flexibility can be leveraged to better use vehicle capacity and reduce the number and distance of routes. Two complementary approaches have been studied.

First, split-order consolidation targets multiple parcels for the same customer. Rather than dispatching each parcel as it arrives from different retailers (and their warehouses), shipments are temporarily held at the delivery station and combined into a single delivery to the customer (Zhang et al., 2019). In this setting, green labels have been shown to increase customers' willingness to wait, generating additional consolidation opportunities (Zhang et al., 2023).

The second approach targets orders of multiple customers and aims to temporally concentrate deliveries by treating the dispatch day as a decision variable. While this approach aligns with the general focus of our study, prior work has largely studied it as an operational decision problem, emphasizing dispatch and routing performance. There are three closely related studies in this domain. Muñoz-Villamizar et al. (2022) formulate a multi-period vehicle routing problem in which orders can be assigned to feasible dispatch days within short postponement windows. Relative to an earliest-dispatch baseline, they report large reductions in route-level outcomes, with four-day postponement windows reducing distance by 57%, total costs by 61%, and CO_2 emissions by 56%. Muñoz-Villamizar et al. (2024) extend this to a larger-scale setting by using rolling-horizon decision-making with region-level postponement penalties and a continuous route-length approximation. They find that postponement can reduce the number of routes by 57%, distance by 46%, transportation time by 43%, and total cost by 29%, while more than doubling the number of deliveries per route. Finally, Voigt et al. (2025) study a vehicle routing problem with a some-day option, where postponement is guided by dynamically updated order-level opportunity costs. Their results similarly indicate substantial operational gains: a three-day postponement window can reduce costs by up to 60% relative to an earliest-policy baseline with a mean delay below 1.5 days.

2.3. Synthesis and literature gap

The three prior studies considering the implications of postponed delivery all suggest a clear positive link between customers' willingness to wait and operational benefits. Each adopts an optimization approach to support day-by-day order dispatch and vehicle routing decisions. In addition, they consider a single e-commerce retailer with its own fleet, and settings with relatively modest delivery volumes and low baseline delivery densities. In such settings, postponement can readily increase delivery density and thereby improve vehicle utilization, yielding large relative savings in route distance and delivery costs.

Existing literature provides limited strategic guidance on how to use the flexibility provided by customers' willingness to wait and what different approaches could realistically deliver across different contexts. In particular, prior work has not explicitly compared the approach of spreading deliveries to smooth demand peaks with the approach of concentrating deliveries to minimize route distance. Context sensitivity, including differences between urban and rural demand volumes, service-area characteristics, and fleet constraints, also remains underexplored. Finally, prior studies typically prioritize cost minimization through operational efficiency, with more limited attention to the levels of customer willingness to wait required to obtain environmental benefits.

Together, these limitations define the main gap in the literature that our study aims to address. Instead of an optimization-based approach focused on supporting daily vehicle routing and order dispatch decisions, we adopt a strategic modeling approach, using real-world e-commerce delivery volumes across various spatial contexts, considering both fleet and service-area characteristics, and explicitly examining how operational and environmental outcomes vary with different levels of customers' willingness to wait.

3. Modeling approach

3.1. Problem description

We consider a parcel delivery system over a finite delivery horizon of N days. Let p_i denote the number of orders initially scheduled for delivery on day i. A fraction of customers may be willing to postpone the delivery of their order to a later day. Accordingly, p_{ij} denotes the number of orders initially scheduled for delivery on day i that can be postponed to day j, with $i \leq j$. For instance, if a customer is willing to postpone an order for delivery on day $i = 1$ for up to three days, the order can be delivered on any of the days $j = 1, 2, 3, 4$. If a customer is not willing to wait, the order must be delivered on day $i = j = 1$. We take a deterministic perspective, assuming full knowledge of delivery demand and customers' willingness to wait over the entire delivery horizon.

Fig. 1 presents a network flow diagram of the postponement mechanism, illustrating how orders initially scheduled for delivery on a given day can either be fulfilled on that day or deferred to one of the subsequent days within the postponement window. In this illustrative five-day delivery horizon, the postponement window is three days, so orders initially scheduled for delivery on day i can be delivered no later than day $j = i + 3$. That is, $p_{ij} = 0$ for all $j > i + 3$. Similarly, orders on later days (e.g., day 5) may also be postponed within the same three-day limit, though this is omitted from the figure for clarity. In the network, the flows x_{ij} represent the number of orders initially scheduled for delivery on day i that are delivered on day j, subject to the restriction $x_{ij} \leq p_{ij}$. The inflows p_i denote the number of orders initially scheduled for delivery on day i, while the outflows n_j denote the number of orders actually delivered on day j after postponement decisions have been made. The flow balance ensures that the total number of orders originating from day i equals p_i, and that the total number of orders delivered on day j equals n_j.

Consistent with the strategic scope of our study, we analyze route characteristics using regional-level approximations rather than optimizing specific delivery routes. In particular, we use the well-established route length approximation of Daganzo (1984), which estimates the total distance based on the number of deliveries and the size of the region served. The size of the service region is denoted by A (in km^2), and the number of delivery locations on day j is denoted by n_j. In this framework, the service region should be interpreted as an analytical aggregation unit over which deliveries are represented in average terms, rather than as an exact operational route area. In line with (Daganzo, 1984), we assume that the delivery locations are uniformly distributed within A, which allows the approximation to capture the expected intra-region travel pattern without modeling exact delivery locations. A routing constant β adjusts the expected intra-region travel distance and $\bar{d}$ denotes the average line-haul distance, defined as the one-way distance from the depot to the centroid of the service region. The line-haul term approximates the average distance between the depot and the region. Thus, each route incurs an expected round-trip line-haul distance of $2\bar{d}$. Daganzo-style VRP approximations have been empirically validated as a good predictor of average route length and generally perform better than approximation methods that ignore depot line-haul distances (Robusté et al., 1990, 2004; Merchán and Winkenbach, 2019; Chien, 1992; Kwon et al., 1995). Let m_j denote the number of vehicles on day j. Assuming each vehicle completes one route per day, the total distance on day j is

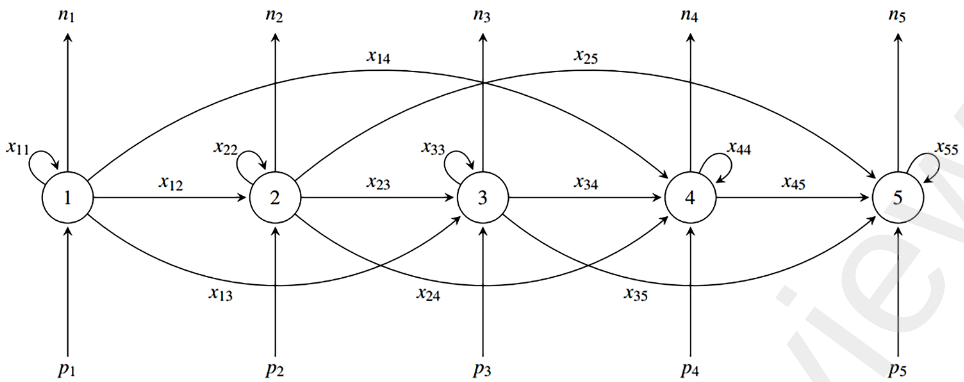

Fig. 1. Illustration of the postponement mechanism across a five-day delivery horizon.

$$2\bar{d}\, m_j \; + \; \beta\sqrt{n_j A}\,. \tag{1}$$

The approximate number of vehicles needed on day j is determined in connection with volume and time limitations. First, assuming a fixed and known average order volume per delivery, and a given vehicle capacity C, in volume per vehicle in the fleet, we can determine the maximum number of deliveries that can be made in a single route. Second, we consider a maximum allowable route duration h in hours. To convert distance to time, two average driving speeds are defined: s_1 (km/h) for line-haul distance, and s_2 (km/h) for travel within the service region. Each delivery incurs a fixed stop time λ (in hours) for parking and parcel handover. Accordingly, the average time per route on day j is

$$\frac{\frac{2\bar{d}m_j}{s_1} + \frac{\beta\sqrt{n_j A}}{s_2} + n_j\lambda}{m_j}. \tag{2}$$

3.2. Model

We formulate a continuous non-linear programming model to determine the optimal distribution of deliveries over the planning horizon. Table 1 provides an overview of the nomenclature used in this model.

The variable x_{ij} denotes how many orders with the earliest delivery day i are carried out on day j. From these postponement choices, the model derives the daily number of deliveries n_j, which determines the number of vehicles m_j required on that day. The overall fleet size F is defined as the maximum number of vehicles used on any day j over the delivery horizon. The decision variables x_{ij}, n_j, m_j, and F are in $\mathbb{R}_+$. Observe that we treat x_{ij}, n_j, m_j, and F as nonnegative reals, even though deliveries and vehicles are discrete in reality. Given our strategic, long-term perspective, approximating variables as real numbers enables us to better capture trends and trade-offs.

In the following sections, we present two alternative objective functions to evaluate the trade-off between spreading deliveries to minimize fleet size and concentrating them to minimize total route distance.

3.2.1. Minimization of total route distance

We include distance minimization as an objective because fewer vehicle kilometers imply less energy use for driving and therefore lower CO_2 emissions in delivery operations. The objective function (3) minimizes the approximate distance traveled across all delivery routes:

$$\min \quad \sum_{j=1}^{N} \left\{ 2\bar{d}m_j + \beta\sqrt{n_j A} \right\} \tag{3}$$

where the term $2\bar{d}m_j$ is the estimated line-haul distance incurred by dispatching m_j vehicles on day j, with each vehicle traveling an average round-trip distance of $2\bar{d}$, and the term $\beta\sqrt{n_j A}$ captures the approximate total distance traveled to serve n_j delivery locations uniformly distributed across the service region A. By minimizing the sum of these terms over the delivery horizon, we determine the optimal distribution of deliveries across days within the allowed postponement windows to reduce total route distance.

3.2.2. Minimization of fleet size

We consider fleet size minimization as an objective because a smaller fleet implies fewer vehicles to be manufactured and therefore lower CO_2 emissions from production. The objective function (4) minimizes the fleet size:

$$\min \quad F + \epsilon \sum_{j=1}^{N} m_j \tag{4}$$

Table 1
Nomenclature of the continuous non-linear model.

Parameters	
p_{ij}	Number of orders on day i that can be postponed to day j, $i = 1, \dots, N$; $j = 1, \dots, N$
p_i	Total number of orders with the earliest delivery on day i
A	Size of the service region (km^2)
$\bar{d}$	Average distance between the depot and the service region (km)
β	Constant
s_1	Driving speed on line-haul (km/h)
s_2	Driving speed inside the service region (km/h)
λ	Stop time per delivery (hours)
h	Maximum allowable route duration for each vehicle (hours)
C	Maximum volume that can be delivered per vehicle (number of deliveries)
ϵ	Small positive constant
Decision variables	
x_{ij}	Number of deliveries on day i that are postponed to day j, $i = 1, \dots, N, j = 1, \dots, N$
n_j	The number of deliveries that are performed on day j
m_j	The number of vehicles needed on day j
F	Fleet size

where the term F denotes the minimum fleet size sufficient to satisfy all deliveries across the delivery horizon, which we will establish as $F = \max\{m_j\}_{j=1,\dots,N}$ by means of the constraints of the model. Since m_j is determined using the continuous route-length approximation, F should likewise be interpreted as an estimated fleet requirement rather than an exact operational fleet size. The term $\epsilon \sum_{j=1}^{N} m_j$, with the small positive constant ϵ, is a tiebreaker that discourages excess daily vehicle use on non-peak days.

3.2.3. Constraints

Both objectives are subject to the following constraints:

$$\text{s.t.} \quad n_j = \sum_{i=1}^{N} x_{ij} \qquad \forall j \tag{5}$$

$$x_{ij} \le p_{ij} \qquad \forall i \neq j \tag{6}$$

$$x_{ii} \ge p_{ii} \qquad \forall i \tag{7}$$

$$\sum_{j=i}^{N} x_{ij} = p_i, \qquad \forall i \tag{8}$$

$$\frac{\frac{2\bar{d}m_j}{s_1} + \frac{\beta\sqrt{n_j A}}{s_2} + n_j\lambda}{m_j} \le h \qquad \forall j \tag{9}$$

$$\frac{n_j}{m_j} \le C \qquad \forall j \tag{10}$$

$$m_j \le F \qquad \forall j \tag{11}$$

$$n_j, m_j \in \mathbb{R}_+ \qquad \forall j \tag{12}$$

$$x_{ij} \in \mathbb{R}_+ \qquad \forall i, j \tag{13}$$

$$F \in \mathbb{R}_+ \tag{14}$$

Constraint (5) defines the realized deliveries on each day j. It states that the number of deliveries performed on day j must equal the total number of deliveries assigned to day j after the postponement decisions are made. Constraint (6) ensures that the number of deliveries postponed from day i to a later day j does not exceed the corresponding upper bound p_{ij}. Constraint (7) makes sure that the deliveries that cannot be postponed have to be delivered on day $i = j$. Constraint (8) enforces delivery balance. It states that all orders are delivered within the respective postponement windows for each day i. Constraint (9) ensures that deliveries remain feasible in terms of time, and Constraint (10) imposes the vehicle volume capacity limit. Each constraint therefore yields a lower bound on the number of vehicles required on day j. Because both constraints must be satisfied, m_j is effectively determined by the maximum of these two bounds. Constraint (11) ensures that the fleet size F is at least the approximate daily vehicle requirement m_j for all days j. Finally, Constraints (12)–(14) define the domains of the decision variables.

4. Analysis

Our modeling approach can be applied to any delivery region, characterized by its service area, the line-haul distance, and the number of deliveries to be served in that region each day over a delivery horizon. In our analysis, we model customers' willingness

Table 2
Parameter values for the representative urban and rural regions.

Category	Parameter (unit)	Urban	Rural
Service region characteristics	Population	160,800	46,400
	Area (km^2)	32.2	517.77
	Average deliveries per day	12,902	4,419
	Delivery density (deliveries/day/km^2)	400.7	8.5
	$\bar{d}$ (km)	10	23
Model parameters[a]	s_1 (km/h)	80	80
	s_2 (km/h)	20	50
	λ (min)	4	2
	β (constant)	0.57	0.57
	h (hours)	7.5	7.5
	C (deliveries)	120	120

[a] Stop time is guided by Allen et al. (2018), Dalla Chiara and Goodchild (2025). We set $\beta = 0.57$ following (Daganzo, 1984). We set the vehicle volume capacity C to 120 deliveries per route. This value is derived from an effective vehicle volume capacity of approximately 5 m^3 and an average delivery volume of 41.6 L (representing 1.6 parcels of 26 L).

to wait as the share of deliveries that can be postponed for a fixed number of days, which we refer to as the postponement window. We adopt this simplified representation—where a variable share of customers accepts a uniform postponement window—to keep the analysis tractable and to ensure results are easily interpretable and comparable across different settings.

Our analysis draws on detailed e-commerce delivery data from a Dutch logistics technology provider. The provider coordinates deliveries for around 300 e-commerce retailers in the Netherlands, while delivery firms operate the routes. The dataset covers all business-to-consumer (B2C) deliveries processed via the platform in 2024[1]. The data exhibits a clear weekly seasonal pattern in delivery volumes. In line with this, we analyze the impact of postponed delivery over a weekly delivery horizon that repeats itself perpetually with average daily delivery volumes derived from the data. To that end, we impose periodic boundary conditions on the day index of our finite-horizon model. We index the days of the week as $1, \dots, 7$ (Monday–Sunday). Then, for an order scheduled on day i, we express the set of possible delivery days by the index set $j = ((i + t - 1) \bmod 7) + 1$ for $t = 0, \dots, K$, where K is the length of the postponement window. This allows us to model postponements from day i of a week to day j of the next week through the postponement feasibility parameter p_{ij} and decision variable x_{ij}, in the context of the optimization model and constraints in Section 3.2. Following the practice in the Netherlands, we assume Sunday is a non-service day and set $p_{i7} = 0$. Consequently, with a three-day postponement window, an order initially scheduled for delivery on Saturday can be postponed at most to the following Monday or Tuesday.

In our analysis, we focus on two representative yet contrasting service regions in the Netherlands: the city of Groningen as an urban region and a nearby rural region in the province of Drenthe. These regions differ markedly in service area size, daily delivery volumes, and resulting delivery density. Table 2 reports their key characteristics, along with the rest of the model parameters used in the analysis.

4.1. *The impact of willingness to wait*

We solve our model with the alternative objectives of minimizing total route distance and minimizing fleet size, and compare the results with the baseline without postponement. We present the results graphically to indicate how objective values change as a function of willingness to wait, as we vary both the share of deliveries that can be postponed and the length of the postponement window. We report the results separately for the urban and rural region to compare the benefits of postponement across different spatial contexts.

We begin by analyzing the results for distance minimization. Fig. 2a and 2b show how the weekly route distance varies with the share of deliveries that can be postponed and the length of the postponement window for the urban and rural regions, respectively. In both regions, the curves indicate that increasing either the share of deliveries that can be postponed or the length of the postponement window leads to a reduction in weekly route distance. This is because postponement allows deliveries to be concentrated on fewer days, thereby increasing local stop density and reducing distance driven per delivery. It is important to highlight that a substantial reduction in distance can be achieved only when a large share of deliveries can be postponed.

Comparing Fig. 2a and 2b, we observe how delivery density affects the potential distance savings from postponement. In the dense urban region, even in the extreme case where all deliveries can be postponed for five days, postponed delivery reduces weekly distance by only 8.8%. In absolute terms, this corresponds to a reduction of 19 meters per delivery from an average distance of 197 meters per delivery. In the more dispersed rural region, the maximum possible reduction is 19.8%, which corresponds to a reduction

[1] The available e-commerce delivery data only covers deliveries mediated by the logistics technology provider. To approximate total regional deliveries, we assume that the provider's market share remains constant across weekdays and consistent across geographic regions and, accordingly, scale the data using national delivery volumes reported by the Dutch Authority for Consumers and Markets.

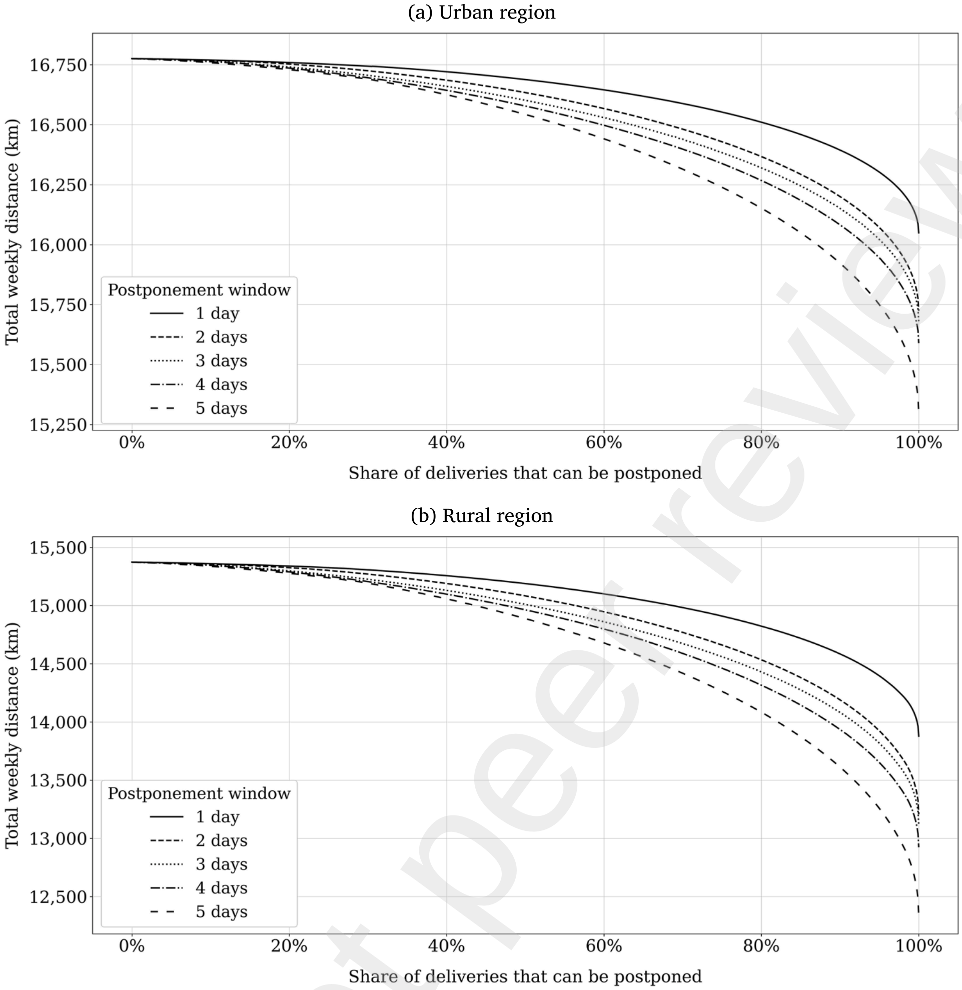

Fig. 2. Impact of delivery postponement on weekly distance for (a) urban and (b) rural regions.

of 115 meters per delivery from an average distance of 465 meters per delivery. In both regions, the reduction in distance quickly diminishes as the share of deliveries that can be postponed decreases or the postponement window becomes shorter.

Fig. 3 presents the distribution of delivery volume over the week when the objective is to minimize distance for postponement windows from one to five days, assuming that all deliveries can be postponed. This illustrates the mechanism behind the distance reductions observed in Fig. 2a and 2b. When all deliveries can be postponed by one day, deliveries take place only every other day. This is intuitive, since route distance is mainly driven by delivery density. This becomes even more pronounced with longer postponement windows, and in the most extreme case all deliveries are concentrated on a single day. We note that this pattern persists even when not all deliveries can be postponed, as the model still concentrates as many deliveries as possible on a few days, leaving only the non-postponable deliveries on the remaining days. By concentrating deliveries into a few peak days, weekly route distance decreases, but peak-day resource requirements rise as substantially more vehicles and delivery staff are needed on those days, with underutilized capacity on the remaining days.

Next, we analyze the results for fleet size minimization. Fig. 4a and 4b illustrate how the minimum fleet size varies with the share of deliveries that can be postponed and the length of the postponement window, for the urban and rural regions, respectively. The minimum fleet size is driven by the day with the highest delivery volume, as this day requires the most vehicles. Hence, the effectiveness of delivery postponement in minimizing fleet size lies in its ability to smooth variation in delivery volumes across days. In both urban and rural regions, increasing either the share of deliveries that can be postponed or the length of the postponement window reduces the required fleet size. This is because postponement allows deliveries to be shifted away from peak days, thereby smoothing delivery volumes across the week and reducing the number of vehicles needed on the busiest day. The figure shows postponement windows up to three days, as extending the postponement window beyond three days does not yield a further reduction in fleet size. For any postponement window, the minimum fleet size is achieved even when only a modest share of deliveries can be postponed. This suggests that postponed deliveries can effectively smooth delivery volume peaks, even when customers' willingness to wait is limited or when operational constraints prevent the full exploitation of the resulting flexibility.

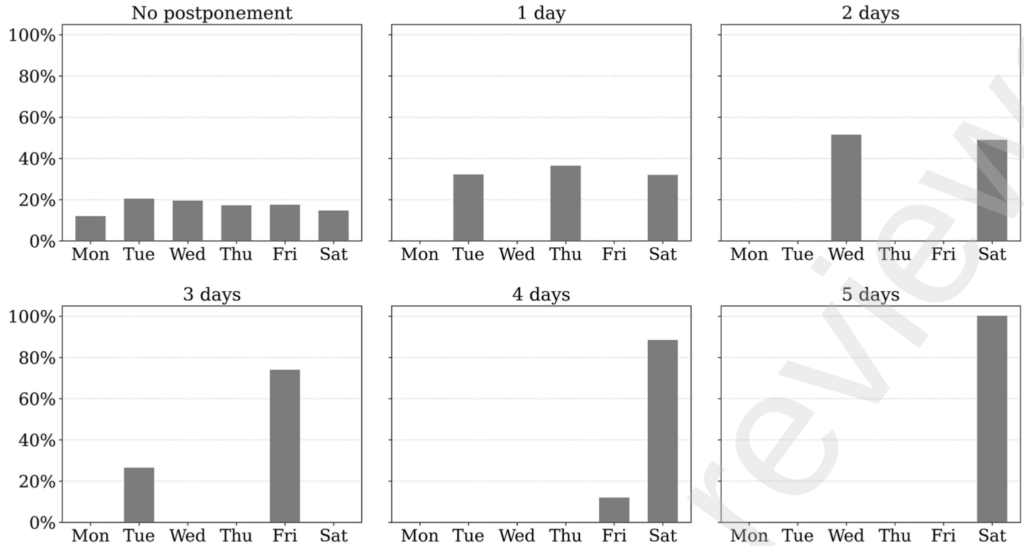

Fig. 3. The (percentage) distribution of weekly delivery volume that minimizes total route distance for different postponement windows.

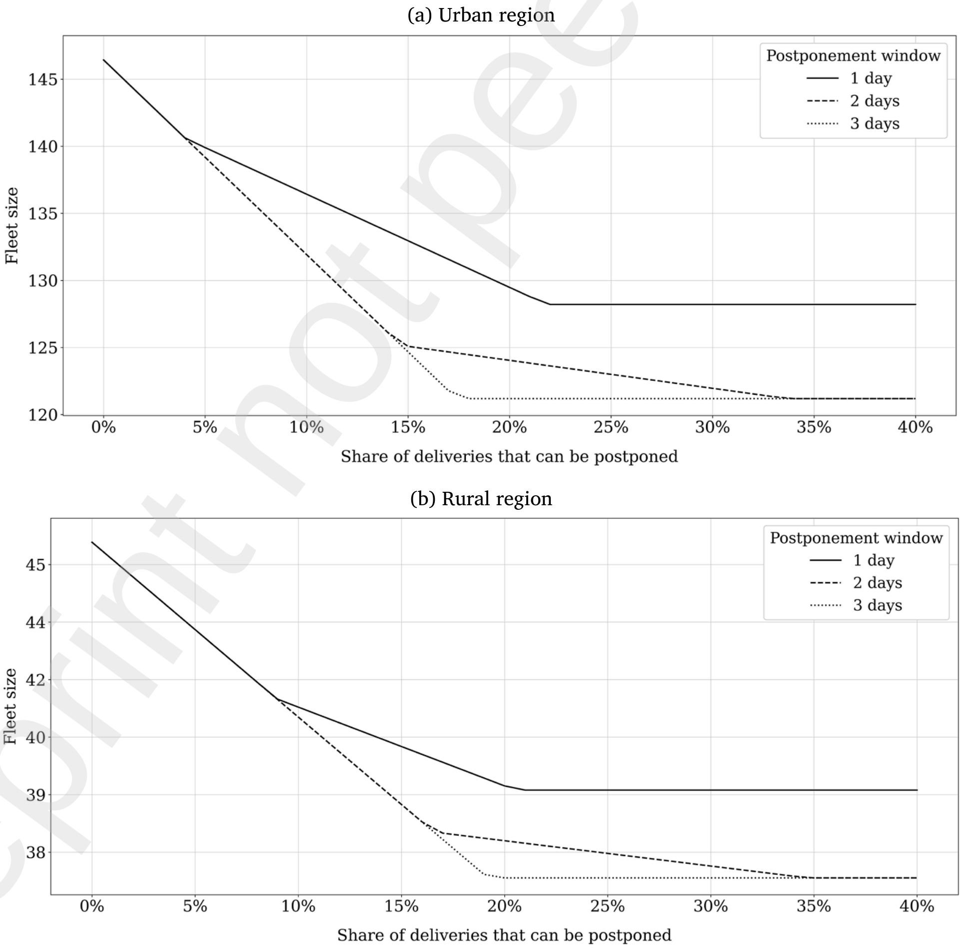

Fig. 4. Impact of delivery postponement on fleet size for (a) urban and (b) rural regions.

Comparing Fig. 4a and 4b we observe how delivery density affects the minimum fleet size. For any given postponement window, the minimum fleet size can be achieved with a smaller share of deliveries that can be postponed in the urban region as compared to the rural region. Consider a postponement window of three days. In the urban region, the minimum possible fleet size is achieved when 17.5% of deliveries can be postponed. This is a reduction of 17.2%, which corresponds to 25.25 vehicles. In the rural region, the minimum possible fleet size is achieved when 19.3% of deliveries can be postponed, which is a reduction of 19.2% or 8.75 vehicles. The difference between these figures can also be attributed to the relatively larger variation in daily delivery volumes in the rural region.

Finally, we analyze the extent to which the flexibility arising from customers' willingness to wait is utilized. Across both objectives, both regions, and all postponement windows, the mean realized delay is consistently below the maximum allowable postponement window—typically around half the window length—and scales approximately linearly with the postponement window length. This pattern is nearly identical under distance minimization and fleet minimization, implying that average customer delays are largely independent of the firm's strategic priority.

4.2. *Trade-off between distance and fleet size*

We now analyze the trade-off between distance and fleet size. To characterize this trade-off, we compute Pareto frontiers in the (D, F) plane, where D denotes the approximate weekly route distance and F the estimated required fleet size. Because D and F differ in units and scale, we normalize both objectives using bounds derived from the two single-objective optima. Specifically, minimizing distance yields $D_{\min}$ with associated fleet size $F_{@D_{\min}}$, while minimizing fleet size yields $F_{\min}$ with associated distance $D_{@F_{\min}}$. In this setting, maximizing either objective is not meaningful as both distance and fleet size can grow without bound. We therefore approximate the nadir point for each objective as the worst value observed when the other objective is minimized, so that the upper bound for distance is $D_{@F_{\min}}$ and the upper bound for fleet size is $F_{@D_{\min}}$. By means of these bounds, we define

$$d_{\text{norm}} = \frac{D - D_{\min}}{D_{@F_{\min}} - D_{\min}} \quad \text{and} \quad f_{\text{norm}} = \frac{F - F_{\min}}{F_{@D_{\min}} - F_{\min}},$$

and trace the frontier by solving the weighted-sum problem

$$\min \; \theta \, f_{\text{norm}} + (1 - \theta) \, d_{\text{norm}}$$

for different values of $\theta \in [0, 1]$, thereby combining the two objectives on the same scale.

Fig. 5a and 5b illustrate the Pareto frontier between fleet size and weekly route distance for urban and rural regions, respectively. The results are based on a three-day postponement window and the share of deliveries that can be postponed is 100%. The horizontal axis reports weekly distance, where the minimum distance $D_{\min}$ is 15,680 km in the urban region and 13,118 km in the rural region. The vertical axis reports fleet size, where the minimum fleet size $F_{\min}$ is 121.2 vehicles in the urban region and 36.8 vehicles in the rural region. The weekly distance when operating at the minimum fleet size $D_{@F_{\min}}$ is 16,785 km in the urban region and 15,394 km in the rural region. Conversely, the fleet size required to attain the minimum-distance solution $F_{@D_{\min}}$ is 530.2 vehicles in the urban region and 162.6 vehicles in the rural region.

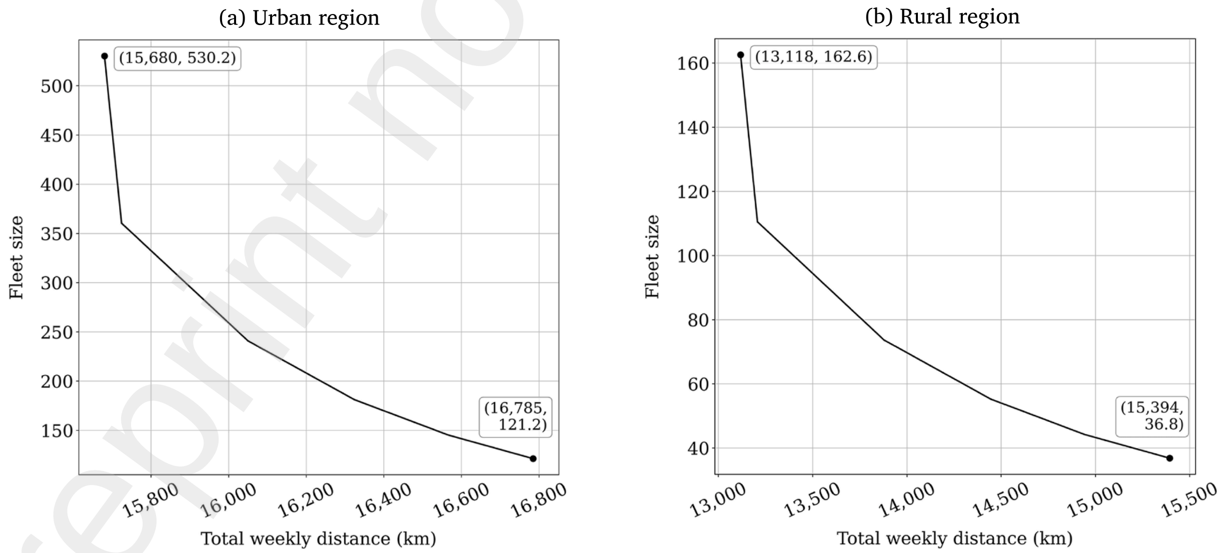

Fig. 5. Pareto frontier between fleet size and weekly route distance for (a) urban and (b) rural regions.

We observe a clear asymmetry in the endpoints of the Pareto frontier. In the urban region, moving from the fleet-optimal to the distance-optimal endpoint requires a 337.46% increase in fleet size, corresponding to 409 additional vehicles. In contrast, moving from the distance-optimal to the fleet-optimal endpoint increases weekly distance by only 7.05% or 1,105 km. In the rural region, the corresponding changes are an increase in fleet size of 341.85% or 125.8 additional vehicles, and an increase in weekly distance of 17.35% equivalent to 2,276 km. The larger relative increase in weekly distance in the rural region can be attributed to longer distances per delivery resulting from more spatially dispersed delivery locations. Overall, these results indicate that substantial fleet reductions can be achieved at the cost of only a relatively small increase in weekly distance.

4.3. Sensitivity analysis

Fig. 6a and 6b illustrate how the trade-off between distance and fleet size is affected by time and volume capacity constraints. We examine this effect by perturbing three representative parameters—one at a time by $\pm 20\%$ relative to the baseline—as shown by the dashed and dotted curves. Specifically, we vary the routing constant β to explore how robust the trade-off is to differences in spatial distribution and road network topology. We vary the stop time λ because it affects the time constraint, which reflects broader operational factors related to the time required and available to complete deliveries, such as service time per stop, handling time, and route duration limits. We also vary the vehicle volume capacity C because it affects the volume constraint, which reflects broader operational factors related to the vehicle's carrying capacity and the capacity required to serve all deliveries. Accordingly, both time-related constraints and vehicle volume capacity are key determinants of the driven distance and the required fleet size. The three parameters influence the trade-off in distinct yet intuitive ways. The routing constant β scales route lengths and therefore shifts the Pareto frontier along the distance axis, while leaving fleet requirements essentially unchanged. By contrast, the stop time λ and vehicle volume capacity C primarily affect the fleet dimension of the trade-off by limiting the number of deliveries per vehicle, thereby dictating the required fleet size. While we do not model all operational constraints explicitly, these two high-level constraints capture the primary drivers of vehicle requirements in our strategic setting.

Comparing Fig. 6a and 6b, we observe pronounced differences in sensitivity to stop time and vehicle volume capacity. Fleet size is driven by whichever constraint (i.e., time or volume) is binding. Consequently, changes to the non-binding constraint have negligible effect. In the urban region, the time constraint binds and increasing vehicle volume capacity has no effect, whereas reducing it shifts the Pareto frontier only if the volume constraint becomes binding. In the rural region, the volume constraint binds and varying stop time by $\pm 20\%$ does not alter this status and leaves the frontier unchanged.

The trade-off remains clearly asymmetric across all parameter variations in both regions. Reaching the distance-optimal endpoint consistently requires a large increase in fleet size, whereas the fleet-optimal endpoint incurs only a modest distance penalty. This confirms that the shape of the Pareto frontier is robust.

5. Broader environmental impact

To assess the broader environmental impact of postponement, we extend our analysis in two steps. First, we scale our findings to the national level to examine how market share affects the strategic value of postponed delivery. Second, we combine these results with life-cycle assessment to quantify the net CO_2 emissions, explicitly trading off the emissions from vehicle manufacturing against the emissions from delivery operations.

5.1. Impact at national scale

To estimate the impact of postponed deliveries at the national level, we consider two B2C delivery firms operating in the Dutch e-commerce sector: the market leader with a 48% market share and a smaller competitor with a 5% market share, and examine how each would benefit from postponed deliveries.

To model these national networks, we require the depot delivery regions and delivery volumes for each firm. Since proprietary network data are unavailable, we construct approximate delivery regions using a transparent partitioning procedure based on depot locations and delivery volumes derived from our e-commerce dataset. Our procedure is as follows: First, we create depot delivery regions by assigning each postcode to the geographically nearest depot, excluding islands which represent operational edge cases. Second, we categorize each postcode as urban or rural using a population density threshold of 500 inhabitants per square kilometer. Third, we subdivide each depot delivery region into distinct service regions by grouping contiguous postcodes and classifying them as urban or rural, yielding analytically tractable regions for the strategic approximation. Finally, for each service region, we approximate the number of deliveries by scaling the e-commerce delivery data according to the firm's market share and compute the line-haul distance from the assigned depot to the centroid of the service region, which serves as an approximation of the average depot-to-region distance. The resulting network, illustrated in Fig. 7, provides a consistent spatial framework for our national-level analysis.

We examine the impact of delivery postponement at the national scale by solving the model for each service region and aggregating the results to obtain the total weekly route distance and fleet size for the market leader and the smaller competitor to serve all their deliveries across the Netherlands. In our analysis, we focus on the impact of the share of deliveries that can be postponed, and keep the postponement window fixed at three days. Furthermore, we keep the number of vehicles required per service region as a continuous variable to avoid systematic overestimation when these requirements are aggregated to the depot level.

For the market leader delivery firm, Table 3 reports the nationwide results under the distance minimization and fleet size minimization objectives. When the objective is to minimize the weekly distance, increasing the share of deliveries that can be postponed from 0% to 100% reduces weekly distance from 1,109,089.2 km to 1,071,604.1 km, a reduction of about 3.4% or 37,500 km. However, this comes at the cost of a much larger fleet. The fleet size increases from 8,508.6 vehicles to 29,928.5 vehicles, a more than threefold

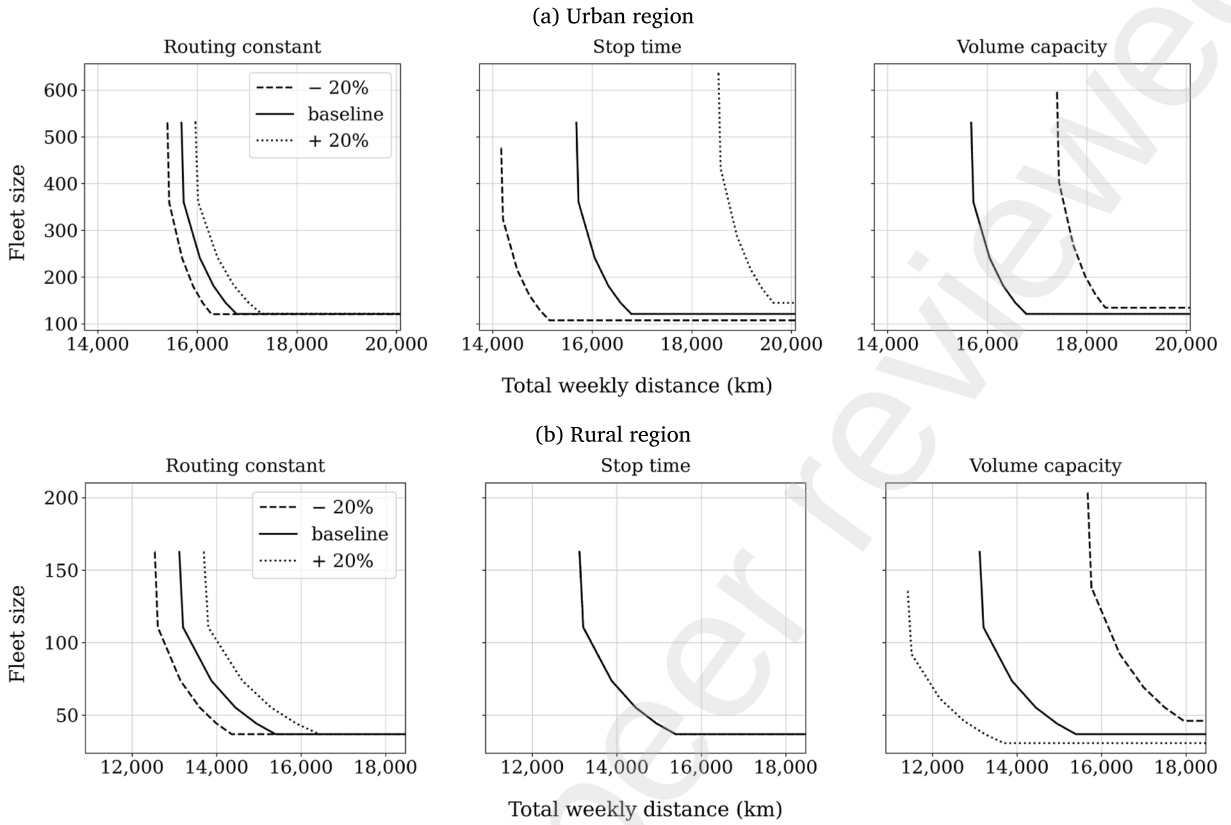

Fig. 6. Sensitivity of the Pareto frontier between weekly route distance and fleet size for (a) urban and (b) rural regions. Each panel shows the distance–fleet trade-off under a ± 20% change in the indicated parameter, relative to the baseline.

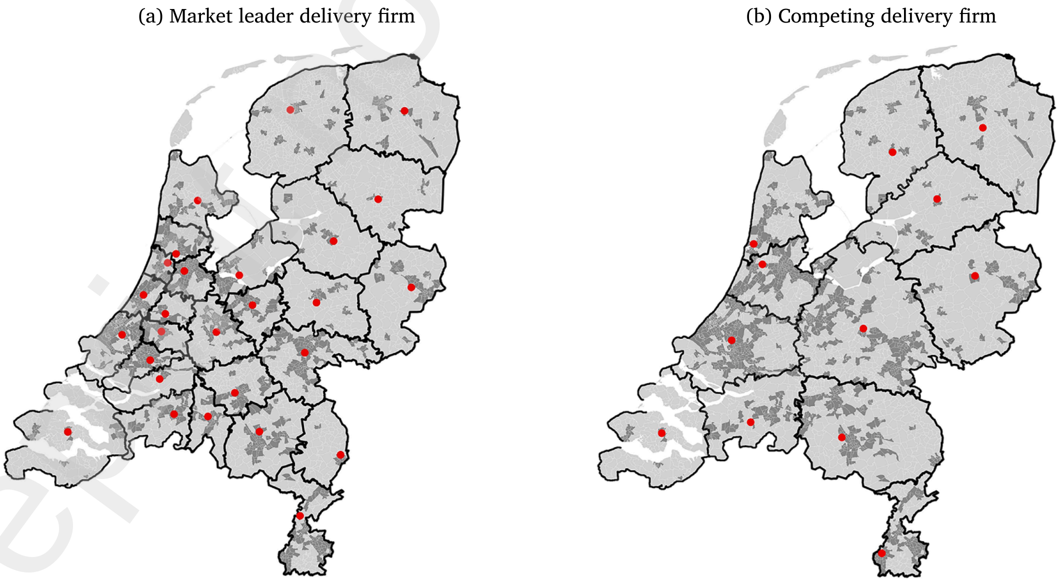

Fig. 7. Depot locations and approximate depot delivery regions for (a) the market leader and (b) a smaller competitor. Red dots represent depot locations, black lines delineate depot delivery regions, dark gray areas indicate urban zones, and light gray areas represent rural zones.

Table 3
The impact of postponement on the market leader's weekly route distance and fleet size, with varying shares of deliveries that can be postponed over a three-day postponement window.

Distance minimization	**0%**	**20%**	**40%**	**60%**	**80%**	**100%**
Fleet size	8508.6	11677.9	16273.5	20847.0	25315.7	29928.5
Weekly distance (km)	1109089.2	1107832.0	1105015.6	1100521.1	1093368.8	1071604.1

Fleet size minimization	**0%**	**5%**	**10%**	**15%**	**20%**	**25%**
Fleet size	8508.6	8084.6	7660.6	7236.6	6829.0	6766.3
Weekly distance (km)	1109089.2	1109086.1	1109119.4	1109292.0	1109512.8	1109525.7

Table 4
The impact of postponement on the smaller competitor's weekly route distance and fleet size, with varying shares of deliveries that can be postponed over a three-day postponement window.

Distance minimization	0%	20%	40%	60%	80%	100%
Fleet size	917.6	1257.4	1749.8	2241.8	2722.4	3213.6
Weekly distance (km)	209908.3	209571.9	208816.8	207609.2	205685.8	199844.9

Fleet size minimization	**0%**	**5%**	**10%**	**15%**	**20%**	**25%**
Fleet size	917.6	872.2	826.7	781.2	736.9	730.7
Weekly distance (km)	209908.3	209907.0	209912.5	209965.3	210021.2	210024.5

increase. Even at lower shares of deliveries that can be postponed, the fleet needs to expand substantially to achieve relatively modest distance savings. When the objective is to minimize the fleet size, increasing the share of deliveries that can be postponed from 0% to 25% reduces the required fleet from 8,508.6 to 6,766.3 vehicles, a reduction of about 20% or 1,742.3 vehicles. This comes at almost no cost in terms of the weekly distance, which increases by less than 0.05%, or a mere 436.5 km. Further increasing the share of deliveries that can be postponed does not yield an additional benefit, as the minimum fleet size is already achieved at 25%.

For the smaller competitor delivery firm, Table 4 reports the nationwide results under the distance minimization and fleet size minimization objectives. When the objective is to minimize the weekly distance, increasing the share of deliveries that can be postponed from 0% to 100% reduces weekly distance from 209,908 km to 199,845 km, a saving of about 4.8% or 10,000 km. These savings again require a substantial increase in fleet size. The fleet size increases from 917.6 to 3,213.6 vehicles, which is more than a threefold increase. When the objective is to minimize the fleet size, increasing the share of deliveries that can be postponed from 0% to 25% reduces the fleet size from 917.6 to 730.7 vehicles, a reduction of roughly 20% or 187 vehicles. This is achieved with negligible impact on weekly distance, which increases by less than 0.1% or 116.2 km. The minimum fleet size is once again achieved when the share of deliveries that can be postponed reaches 25%.

Fleet capacity is naturally allocated at the depot level, allowing vehicles to be shared across the service regions assigned to a depot. To compute depot fleet sizes, we therefore aggregate the continuous vehicle requirements from all assigned service regions. In our national-scale setting, depot fleets are large in every scenario. Across both objectives and all postponement shares, the smallest depot fleet is about 89.1 vehicles for the market leader and 18.1 vehicles for the smaller competitor. At this scale, treating fleet size as a continuous variable is appropriate, as rounding to integers at the service-region level would systematically overestimate total fleet requirements.

Taken together, the results at the national scale show that postponed delivery offers very limited potential to reduce route distance for either the market leader or the competitor. Even when all deliveries can be postponed, only marginal distance savings can be achieved, and these come at the cost of a substantially larger fleet. This may appear surprising for the competitor, as its smaller market share and lower delivery density would provide greater potential to reduce route distance due to postponed deliveries. However, because the competitor operates with fewer depots, its average line-haul distance is substantially longer and accounts for a larger share of total distance, thereby offsetting any distance savings achieved at the regional level.

In sharp contrast, postponed delivery provides both the market leader and the smaller competitor considerable scope to reduce their fleet size at the expense of a very limited increase in their total route distance. In addition, sizable reductions in fleet size can be achieved even when only a small share of deliveries can be postponed. To achieve the minimum possible fleet size, only 25% of deliveries must be postponed. This is slightly higher than what we reported in our analysis of individual regions, which can be attributed to the fact that some regions in the Netherlands exhibit stronger day-to-day variation in delivery volumes over the week.

5.2. Life-cycle emission impact

We now evaluate the net environmental impact of postponed deliveries at the national level. Given the trade-off between fleet size and route distance, we model total life-cycle CO_2 emissions as the sum of emissions from vehicle manufacturing and delivery operations. We focus this analysis on the market leader with a three-day postponement window. To establish the lower and upper

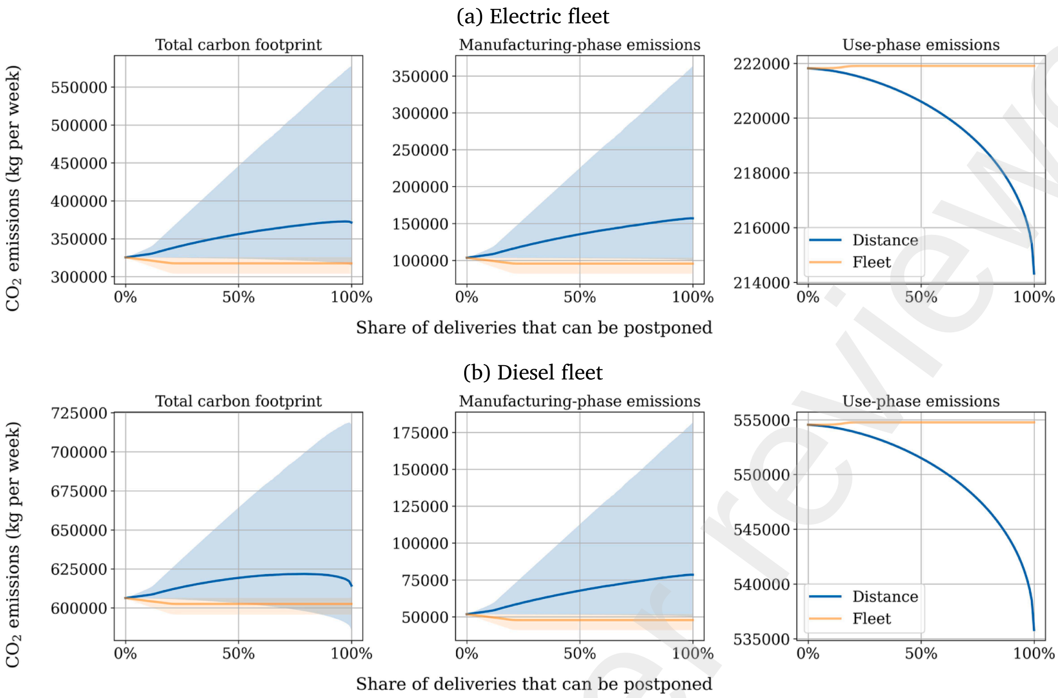

Fig. 8. Weekly CO_2 emissions as a function of the share of deliveries that can be postponed with an (a) electric fleet and (b) diesel fleet, decomposed into manufacturing-phase and use-phase emissions.

bounds of these environmental implications, we analyze two extreme fleet compositions: a fully diesel and a fully electric fleet. In practice, delivery firms often operate mixed fleets that fall between these two extremes.

Following (Giordano et al., 2018), we set manufacturing emissions to $E^{man}_{electric} = 7{,}600$ kg CO_2 for an electric van and $E^{man}_{diesel} = 3{,}800$ kg CO_2 for a diesel van. Manufacturing emissions cover the full vehicle. The main difference is that electric vans include battery and electric drivetrain production (including one battery replacement), whereas diesel vans include the internal combustion engine and related systems. Use-phase emissions per kilometer for electric vans are $e^{use}_{electric} = 0.2$ kg CO_2 per km and for diesel vans $e^{use}_{diesel} = 0.5$ kg CO_2 per km. Diesel use-phase emissions reflect Euro 5 and 6 vehicles, while electric use-phase emissions are averaged across electricity-mix scenarios (Giordano et al., 2018).

Vehicle lifetime depends on utilization. While higher driving intensity shortens a vehicle's service life in years, it simultaneously increases the total kilometers driven before retirement (Morfeldt and Johansson, 2022). In other words, a vehicle that drives more kilometers per week is retired sooner, but covers a greater cumulative distance over its lifetime compared to a vehicle that is left idle or driven less intensively. This higher lifetime distance spreads manufacturing emissions over more kilometers, decreasing the emissions per kilometer. Following Morfeldt and Johansson's lifetime-intensity model, we express vehicle lifetime as a function of weekly distance D as

$$L(D) = \tau_0 \, (D_0/D)^{\gamma}$$

where τ_0 and D_0 are the average vehicle lifetime in weeks and average weekly distance, respectively. $L(D)$ denotes the vehicle's lifetime until retirement from use, not its operational or economic life in a single firm's fleet. We allocate manufacturing emissions over $L(D)$ to ensure fair attribution to the firm's period of use. In our analysis, we set $\gamma = 0.65$, as estimated by Morfeldt and Johansson (2022), $\tau_0 = 625.7$ weeks (or 12 years) for both electric and diesel vans, consistent with the lifetime reported in Giordano et al. (2018) and $D_0 = 130$ in line with the weekly driving distance per vehicle of the market-leading firm in our setting without postponement.

Based on the above, we can write the weekly emissions of a delivery firm operating F vehicles, each driving D kilometers per week, as

$$\left(\frac{E^{man}_{v}}{L(D)} + e^{use}_{v} D \right) F \text{ kg } CO_2,$$

where the subscript v indexes vehicle type, with $v \in \{\text{electric}, \text{diesel}\}$.

Fig. 8a and 8b illustrate weekly CO_2 emissions for fully electric and fully diesel delivery fleets under distance and fleet size minimization, with total emissions decomposed into manufacturing- and use-phase components. Looking at the solid lines in the graphs, we observe that the two objectives lead to opposing effects on manufacturing-phase and use-phase emissions for both fleet types. For distance minimization, postponement reduces route distance and increases fleet size. It therefore results in lower use-phase emissions but higher manufacturing-phase emissions. The increase in manufacturing-phase emissions dominates the reduction in use-phase emissions, leading to higher weekly CO_2 emissions when larger shares of deliveries can be postponed. For fleet size minimization, postponement reduces fleet size while slightly increasing route distance. It therefore results in lower manufacturing-phase emissions and slightly higher use-phase emissions. Overall, the reduction in manufacturing-phase emissions outweighs the modest increase in use-phase emissions. Hence, weekly CO_2 emissions initially decline as the share of deliveries that can be postponed increases and then level off once the minimum possible fleet size is achieved.

We derived our baseline value ($\gamma = 0.65$) from driving distance and retirement statistics for passenger cars in Sweden (Morfeldt and Johansson, 2022). Acknowledging that their model is not tailored to delivery vans, we test the sensitivity of our results to different values of γ by varying it across its full theoretical range. The shaded bands in Fig. 8a and 8b indicate this sensitivity. The endpoints of the bands correspond to $\gamma \in \{0, 1\}$. When $\gamma = 1$, vehicle lifetime is determined solely by cumulative distance driven, with no contribution from calendar aging. Since distance minimization reduces total kilometers, it extends vehicle lifetime and lowers weekly manufacturing-phase emissions by allocating them over more weeks. Fleet-size minimization, however, slightly increases weekly route distance, thereby reducing lifetime. In this setting, distance minimization yields lower total CO_2 emissions than fleet minimization only in the extreme case where 100% of deliveries can be postponed. In that case, distance minimization has maximal flexibility to reduce total distance, and the resulting reduction in use-phase emissions can outweigh the benefits of operating a smaller fleet. When $\gamma = 0$, lifetime is fixed at τ_0 and independent of cumulative distance driven, so manufacturing-phase emissions per week depend only on fleet size. In this case, distance minimization increases total emissions because it requires a larger fleet, while fleet size minimization reduces emissions. Overall, our qualitative conclusions are robust across the range of γ values that are plausible in practice.

Beyond γ, we verified that the results are not sensitive to the calibration of other key parameters. Specifically, we vary τ_0, D_0, E_v^{man}, and E_v^{use} by $\pm 50\%$ ceteris paribus. The only exception observed is a scenario in which diesel manufacturing emissions are reduced by 50%, in which case distance minimization becomes preferable over fleet minimization only if 98% of deliveries can be postponed. Otherwise, the patterns in Fig. 8a and 8b remain unchanged, confirming that the environmental benefit of the demand-smoothing approach is robust across plausible scenarios.

6. Discussion

The strategic insights from our analyses are broadly consistent with existing operational-level optimization studies on postponed delivery, while highlighting a clear trade-off between distance savings and fleet size reductions. Specifically, our analyses confirm that longer postponement windows and higher shares of deliveries that can be postponed increase flexibility and improve delivery operations (Muñoz-Villamizar et al., 2022; Voigt et al., 2025). As in Muñoz-Villamizar et al. (2024), benefits are larger in regions with lower baseline delivery densities and smaller in dense service regions. Moreover, we demonstrate that realized delays stay far below the full postponement window length, as also reported by Voigt et al. (2025).

Our study departs from prior studies by showing that, while postponed delivery can reduce total route distance when it is steered toward distance minimization, the resulting distance savings are modest in dense delivery contexts. Baseline delivery densities (i.e., prior to implementing postponed delivery) play a key role. When baseline delivery density is low, adding a few more deliveries to a route can noticeably shorten the average distance per delivery. When density is already high, the same increase tends to bring only very small additional reductions. In our setting, delivery densities are about 400 deliveries/day/km^2 in the urban region and 8.5 deliveries/day/km^2 in the rural region. By contrast, Muñoz-Villamizar et al. (2022) operates at much lower density, which we approximate at about 0.028 deliveries/day/km^2 based on 30 deliveries over four days and an estimated service-area size of 265 km^2. This lower density is consistent with the larger distance savings they find compared to our setting.

Another point of divergence from prior studies is that our strategic analysis explicitly introduces postponed delivery as a demand-smoothing lever. Operational-level optimization studies, such as (Voigt et al., 2025) and (Muñoz-Villamizar et al., 2024), focus on the average number of routes or vehicles per day, and report postponed delivery can improve these metrics. In our study, we analyze how using customers' willingness to wait to spread deliveries can reduce the number of delivery vehicles needed on the busiest day, under stylized time and capacity constraints. This peak-oriented measure captures the capital and staffing implications of postponed delivery. When deliveries are concentrated on fewer days, the number of routes over the delivery horizon or the average number of vehicles used per day may decrease, but the number of delivery vehicles needed on the busiest day increases.

In this study, we made several deliberate modeling choices that provide directions for future work. First, our empirical setting relies on operational data from a single logistics technology provider, which we scaled to approximate how total demand is distributed across space and weekdays. Although the dataset covers a large and diverse set of around 300 e-commerce retailers, it is not a complete census of Dutch e-commerce deliveries, and the scaled instances may not fully reflect the national demand mix. Moreover, scaling the data to represent the whole Netherlands through an approximate procedure may amplify underlying measurement and approximation errors. Future work could test robustness across multiple data sources.

Second, we chose to estimate total route distance using continuous approximation rather than explicitly modeling vehicle routes. This abstracts from operational detail, but is well suited for strategic analysis and keeps evaluating many different settings computationally feasible. Continuous approximation assumes that delivery locations are uniformly distributed within a service region. This is a near worst-case assumption for route length because it removes the clustering observed in real demand (Carlsson and Behroozi, 2017). In practice, clustering typically shortens travel distances and may increase the distance-saving potential of delivery postponement. A direction for future research is to replace the continuous approximation with a zone-based or threshold dispatch policy. Because this would shift the focus toward more tactical decision-making, it would form a complement to our present study.

Third, our analyses focus on a representative average week in which all deliveries are known in advance. This assumption is reasonable given relatively stable week-to-week demand volumes, except during specific holiday periods—most notably the surge from Black Friday through Christmas or other disruption-driven demand spikes. While the observed demand smoothing indicates the potential for a more balanced workload through reduced peak-day volumes, our model does not quantify associated social outcomes such as overtime, driver stress, or subcontracting, which may still occur under smoothed demand. Accordingly, these effects should be interpreted as managerial implications, rather than as outcomes quantified directly by the model. Future research can explore how delivery postponement performs under peak seasons with large demand surges and explicitly measure how it affects workload-related outcomes.

Fourth, we analyze each postponement window length separately to keep the results transparent and comparable across settings. In practice, however, customers differ in their willingness to wait and may respond to discounts and information, so postponement uptake is likely endogenous rather than fixed. Our model can accommodate richer representations of willingness to wait, including heterogeneous mixes of postponement windows in which each window corresponds to a different share of deliveries that can be postponed. Future work could allow heterogeneity in customers' willingness to wait within a single instance. A further step would be to introduce an economic component that links postponement adoption to incentive levels, enabling assessment of whether such incentives are financially viable. More generally, the managerial challenge is to design service offerings and incentives that motivate customers to choose postponed delivery and sustain uptake over time.

Fifth, our analysis assumes that postponable deliveries can be reallocated flexibly within the postponement window and that delivery operations can be adjusted accordingly. In practice, rolling-horizon planning and operational constraints (e.g., delivery promises, warehouse operations, route familiarity, and driver schedules) can limit this flexibility and reduce realizable gains. Other operational realities, such as depot throughput, dispatch waves, driver rosters, and cut-off time policies, may further constrain the extent to which willingness to wait can be translated into effective flexibility in practice. This holds especially for distance minimization, which typically requires high willingness-to-wait levels to be effective. By contrast, demand smoothing aimed at minimizing fleet size achieves its maximum potential already when only a modest share of customers is willing to wait. Since stated willingness to wait often exceeds these levels, there is likely a surplus of flexibility, which can serve as a buffer to absorb the operational inefficiencies described above.

7. Conclusion

This study examined how the willingness of customers to wait for their e-commerce order can be used to reduce the environmental impact of delivery operations. In a dense e-commerce delivery context, such as the Netherlands, using postponed delivery to smooth delivery volume peaks seems both realistic and effective. It enables substantial reductions in the number of delivery vehicles needed on the busiest day even with limited willingness to wait, and with only modest increases in route distance. This can considerably reduce total life-cycle emissions of the delivery fleet and may support more sustainable working conditions by reducing weekday peaks that are typically absorbed through intensified routes, longer shifts, and reliance on subcontractors, although such labor-related effects are not modeled directly. By contrast, using postponed delivery to concentrate demand yields meaningful distance savings only when willingness to wait is very high. Even then, savings are modest and come at the cost of larger delivery fleets, resulting in increased life-cycle emissions. Sensitivity analyses confirm that these conclusions are robust to variations in operational parameters, such as routing efficiency and service times, as well as in fleet composition and life-cycle parameters, although the exact environmental ranking depends on assumptions about vehicle technology and vehicle lifetime. Taken together, our analyses suggest that in dense e-commerce delivery systems, the main environmental potential of postponed delivery lies not in concentrating deliveries to minimize route distance, but in using customers' willingness to wait to smooth delivery volume peaks and reduce the size of the delivery fleet.

CRediT authorship contribution statement

Marith J. Zeelenberg: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization; **Onur A. Kilic:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization; **Paul Buijs:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was partially supported by the European Union's Horizon Europe research and innovation program under grant agreement No 101147942 (GreenTurn). We also thank the company Wuunder for providing the e-commerce delivery data and for their openness to discussion throughout the project, with special thanks to Noa, Jeroen, and Peter-Paul for their support and willingness to brainstorm.

References

Agatz, N., Campbell, A., Fleischmann, M., Savelsbergh, M., 2011. Time slot management in attended home delivery. Transp. Sci. 45 (3), 435–449. https://doi.org/10.1287/trsc.1100.0346

Agatz, N., Fan, Y., Stam, D., 2021. The impact of green labels on time slot choice and operational sustainability. Prod. Oper. Manage. 30 (7), 2285–2303. https://doi.org/10.1111/poms.13368

Allen, J., Bektas, T., Cherrett, T., Bates, O., Friday, A., McLeod, F., Piecyk, M., Piotrowska, M., Nguyen, T., Wise, S., 2018. The scope for pavement porters: addressing the challenges of last-mile parcel delivery in London. Transp. Res. Rec. 2672 (9), 184–193. https://doi.org/10.1177/0361198118794535

Arevalo-Ascanio, R., De Meyer, A., Gevaers, R., Guisson, R., Verbelen, G., Dewulf, W., 2025. Collaborative distribution network design for sustainable parcel deliveries: A strategic modelling approach. Transp. Res. Part D Transport Environ. 141, 104667. https://doi.org/10.1016/j.trd.2025.104667

Arslan, A.M., Agatz, N., Kroon, L., Zuidwijk, R., 2019. Crowdsourced delivery-a dynamic pickup and delivery problem with ad hoc drivers. Transp. Sci. 53 (1), 222–235. https://doi.org/10.1287/trsc.2017.0803

Bender, M., Kalcsics, J., Meyer, A., 2020. Districting for parcel delivery services–a two-stage solution approach and a real-world case study. Omega 96, 102283. https://doi.org/10.1016/j.omega.2020.102283

Buldeo Rai, H., Broekaert, C., Verlinde, S., Macharis, C., 2021. Sharing is caring: How non-financial incentives drive sustainable e-commerce delivery. Transp. Res. Part D Transport Environ. 93, 102794. https://doi.org/10.1016/j.trd.2021.102794

Buldeo Rai, H., Verlinde, S., Macharis, C., 2019. The "next day, free delivery" myth unravelled: Possibilities for sustainable last mile transport in an omnichannel environment. Int. J. Retail Distrib. Manage. 47 (1), 39–54. https://doi.org/10.1108/IJRDM-06-2018-0104

Carlsson, J.G., Behroozi, M., 2017. Worst-case demand distributions in vehicle routing. Eur. J. Oper. Res. 256 (2), 462–472. https://doi.org/10.1016/j.ejor.2016.03.047

Cauwelier, K., Buldeo Rai, H., Mommens, K., 2025. Thinking fast, slow delivery: nudging sustainable choices among Gen Z consumers. Res. Transp. Econ. 114, 101657. https://doi.org/10.1016/j.retrec.2025.101657

Cauwelier, K., Buldeo Rai, H., Puttemans, K., Macharis, C., Mommens, K., 2024. From cart to door: Unravelling consumer behaviour through attitudinal sustainability profiles. Transp. Res. Part D Transport Environ. 130, 104168. https://doi.org/10.1016/j.trd.2024.104168

Chien, T.W., 1992. Operational estimators for the length of a traveling salesman tour. Comput. Oper. Res. 19 (6), 469–478. https://doi.org/10.1016/0305-0548(92)90002-M

Daganzo, C.F., 1984. The distance traveled to visit n points with a maximum of c stops per vehicle: an analytic model and an application. Transp. Sci. 18 (4), 331–350. https://doi.org/10.1287/trsc.18.4.331

Dalla Chiara, G., Goodchild, A., 2025. The role of walking in last-mile urban deliveries. Transportation , https://doi.org/10.1007/s11116-025-10633-6.

Dietl, M., Voigt, S., Kuhn, H., 2024. From rush to responsibility: Evaluating incentives on online fashion customers' willingness to wait. Transp. Res. Part D Transport Environ. 133, 104280. https://doi.org/10.1016/j.trd.2024.104280

Giordano, A., Fischbeck, P., Matthews, H.S., 2018. Environmental and economic comparison of diesel and battery electric delivery vans to inform city logistics fleet replacement strategies. Transp. Res. Part D Transport Environ. 64, 216–229. https://doi.org/10.1016/j.trd.2017.10.003

Ignat, B., Chankov, S., 2020. Do e-commerce customers change their preferred last-mile delivery based on its sustainability impact? Int. J. Logist. Manage. 31 (3), 521–548. https://doi.org/10.1108/IJLM-11-2019-0305

Janjevic, M., Merchan, D., Winkenbach, M., 2021. Designing multi-tier, multi-service-level, and multi-modal last-mile distribution networks for omni-channel operations. Eur. J. Oper. Res. 294 (3), 1059–1077. https://doi.org/10.1016/j.ejor.2020.08.043

Klein, R., Neugebauer, M., Ratkovitch, D., Steinhardt, C., 2019. Differentiated time slot pricing under routing considerations in attended home delivery. Transp. Sci. 53 (1), 236–255. https://doi.org/10.1287/trsc.2017.0738

Kokkinou, A., Quak, H., Mitas, O., Mandemakers, A., 2024. Should I wait or should I go? Encouraging customers to make the more sustainable delivery choice. Res. Transp. Econ. 103, 101388. https://doi.org/10.1016/j.retrec.2023.101388

Kwon, O., Golden, B., Wasil, E., 1995. Estimating the length of the optimal TSP tour: An empirical study using regression and neural networks. Comput. Oper. Res. 22 (10), 1039–1046. https://doi.org/10.1016/0305-0548(94)00093-N

Lin, J., Zhou, W., Du, L., 2018. Is on-demand same day package delivery service green? Transp. Res. Part D Transport Environ. 61, 118–139. https://doi.org/10.1016/j.trd.2017.06.016

de Mello Bandeira, R.A., Goes, G.V., Gonçalves, D. N.S., de Almeida D'Agosto, M., de Oliveira, C.M., 2019. Electric vehicles in the last mile of urban freight transportation: a sustainability assessment of postal deliveries in Rio De Janeiro-Brazil. Transp. Res. Part D Transport Environ. 67, 491–502. https://doi.org/10.1016/j.trd.2018.12.017

Merchán, D., Winkenbach, M., 2019. An empirical validation and data-driven extension of continuum approximation approaches for urban route distances. Networks 73 (4), 418–433. https://doi.org/10.1002/net.21874

Morfeldt, J., Johansson, D. J.A., 2022. Impacts of shared mobility on vehicle lifetimes and on the carbon footprint of electric vehicles. Nat. Commun. 13 (1), 6400. https://doi.org/10.1038/s41467-022-33666-2

Muñoz-Villamizar, A., Velazquez-Martínez, J.C., Caballero-Caballero, S., 2024. A large-scale last-mile consolidation model for e-commerce home delivery. Expert Syst. Appl. 235, 121200. https://doi.org/10.1016/j.eswa.2023.121200

Muñoz-Villamizar, A., Velázquez-Martínez, J.C., Mejía-Argueta, C., Gámez-Pérez, K., 2022. The impact of shipment consolidation strategies for green home delivery: A case study in a Mexican retail company. Int. J. Prod. Res. 60 (8), 2443–2460. https://doi.org/10.1080/00207543.2021.1893852

Muriel, J.E., Zhang, L., Fransoo, J.C., Perez-Franco, R., 2022. Assessing the impacts of last mile delivery strategies on delivery vehicles and traffic network performance. Transp. Res. Part C Emerging Technol. 144, 103915. https://doi.org/10.1016/j.trc.2022.103915

Nguyen, D.H., De Leeuw, S., Dullaert, W., Foubert, B. P.J., 2019. What is the right delivery option for you? Consumer preferences for delivery attributes in online retailing. J. Bus. Logist. 40 (4), 299–321. https://doi.org/10.1111/jbl.12210

Perboli, G., Brotcorne, L., Bruni, M.E., Rosano, M., 2021. A new model for last-mile delivery and satellite depots management: The impact of the on-demand economy. Transp. Res. Part E Logist. Transp. Rev. 145, 102184. https://doi.org/10.1016/j.tre.2020.102184

Robusté, F., Daganzo, C.F., Souleyrette, II, R.R., 1990. Implementing vehicle routing models. Transp. Res. Part B Methodol. 24 (4), 263–286. https://doi.org/10.1016/0191-2615(90)90002-G

Robusté, F., Estrada, M., López-Pita, A., 2004. Formulas for estimating average distance traveled in vehicle routing problems in elliptic zones. Transp. Res. Rec. 1873 (1), 64–69. https://doi.org/10.3141/1873-08

Stokkink, P., Geroliminis, N., 2025. On the optimal micro-hub locations in a multi-modal last-mile delivery system. Transp. Res. Part E Logist. Transp. Rev. 203, 104344. https://doi.org/10.1016/j.tre.2025.104344

Stroh, A.M., Erera, A.L., Toriello, A., 2022. Tactical design of same-day delivery systems. Manage. Sci. 68 (5), 3444–3463. https://doi.org/10.1287/mnsc.2021.4041

Voigt, S., Frank, M., Kuhn, H., 2025. Last mile delivery routing problem with some-day option. Eur. J. Oper. Res. 324 (2), 477–491. https://doi.org/10.1016/j.ejor.2025.02.001

Waßmuth, K., Köhler, C., Agatz, N., Fleischmann, M., 2023. Demand management for attended home delivery—A literature review. Eur. J. Oper. Res. 311 (3), 801–815. https://doi.org/10.1016/j.ejor.2023.01.056

Yang, D., Hyland, M.F., 2024. Electric vehicles in urban delivery fleets: how far can they go? Transp. Res. Part D Transport Environ. 129, 104127. https://doi.org/10.1016/j.trd.2024.104127

Zhang, Y., Hu, X., Tian, Q., 2023. Order consolidation for the last-mile split delivery problem with green labels. Ann. Oper. Res. , 1–25. https://doi.org/10.1007/s10479-023-05697-5

Zhang, Y., Sun, L., Hu, X., Zhao, C., 2019. Order consolidation for the last-mile split delivery in online retailing. Transp. Res. Part E Logist. Transp. Rev. 122, 309–327. https://doi.org/10.1016/j.tre.2018.12.011