

Strategic Sizing of Collection and Delivery Point Networks for Urban Parcel Distribution

Benjamín Rojas, Homero Larrain, Mathias Klapp

Center for Advanced Transportation, Logistics, and Economic Competitiveness (CATLEC) and
Pontificia Universidad Católica de Chile, Santiago, Chile

[bgrojas@uc.cl, homero@uc.cl, maklapp@uc.cl]

October 9, 2026

Abstract

Collection and delivery points (CDPs) are fixed locations where customers can pick up parcels and/or drop off shipments or returns rather than requiring home delivery. CDPs allow logistics operators to consolidate multiple customer delivery requests on fewer vehicle stops, reducing distribution costs. However, for customers to adopt CDPs, they must be willing to travel to a nearby CDP to pick up their parcels. This choice depends on customers' personal preferences, the proximity of CDPs, and economic incentives. To strategically size and plan cost-efficient CDP networks, we propose a continuous approximation model that jointly determines the optimal number of CDPs and the economic incentives offered to customers who choose to be served at CDPs, based on aggregated instance data. Furthermore, we equip our model with a chance constraint that limits the probability that the random demand associated with each CDP exceeds its design capacity. In our base experiment, we find that CDPs alone can cut expected distribution and infrastructure costs by 16.9%, while offering incentives to customers may increase potential savings to 28.0%. As an insight, we find that CDPs are most beneficial in high-density regions, where they leverage economies of scale, whereas incentives are more effective in low-density regions. We also present a study inspired by the road network and population density of Santiago, Chile, showing that our continuous approximation model can reliably guide CDP network sizing decisions in a

realistic setting. This study provides managerial insights for understanding and designing cost-efficient urban parcel distribution systems.

Keywords: Last-mile distribution, Collection and delivery points, Home delivery, Continuous approximation.

1 Introduction

Urban logistics encompasses all the processes involved in the distribution of goods within cities, including inventory management and last-mile delivery operations. Among these activities, the last mile is widely recognized as accounting for a considerable share of total shipping costs (Savelsbergh and van Woensel, 2016; Boysen et al., 2021; Coppola, 2024). It also has significant implications for urban mobility and sustainability, as last-mile freight delivery contributes substantially to traffic congestion and environmental externalities in cities (Allen et al., 2018; Rai et al., 2021; Caspersen and Navrud, 2021). From a demand-side perspective, customers have increasingly raised their expectations for convenient and flexible last-mile delivery services. In particular, expectations regarding fast, reliable, and flexible delivery have become a key determinant of online purchasing decisions, with delivery service levels strongly influencing purchase intentions and customer satisfaction (Vakulenko et al., 2018; Olsson et al., 2023).

Logistics scientists and practitioners have proposed various innovations to reduce costs and increase customer service levels in last-mile distribution operations. As noted by Savelsbergh and van Woensel (2016) and Boysen et al. (2021), one such innovation is the introduction of collection and delivery points (CDPs), which are fixed locations where customers can opt to pick up their online purchases; these can be either self-attended, as in parcel lockers, or assisted by store personnel.

The use of CDPs can provide numerous benefits. From the operator's perspective, it may reduce last-mile distribution costs by consolidating the distribution of multiple customer orders into fewer delivery stops. Additionally, customers may find it more convenient to choose delivery to a nearby CDP as it lets them access orders easily without waiting at home or risking theft. Therefore, logistics providers could establish a CDP network to optimize last-mile distribution operations. Specifically, to attain a cost-efficient operation, they should carefully balance infrastructure investments and last-mile distribution expenses. A larger CDP network increases infrastructure costs but also reduces the distance between customers and CDPs, which in

turn increases the likelihood that customers will opt to self-pick up their orders rather than choose home delivery (Lemke et al., 2016; Lyu and Teo, 2022). Consequently, it may reduce the expected number of delivery stops and, thus, last-mile distribution costs.

Figure 1 illustrates this trade-off, presenting an example of a particular delivery operation attending 20 customer orders via vehicle routes dispatched from a distribution center. Each route's duration cannot exceed the maximum workday length. Figure 1a depicts a home delivery operation without CDPs; it requires

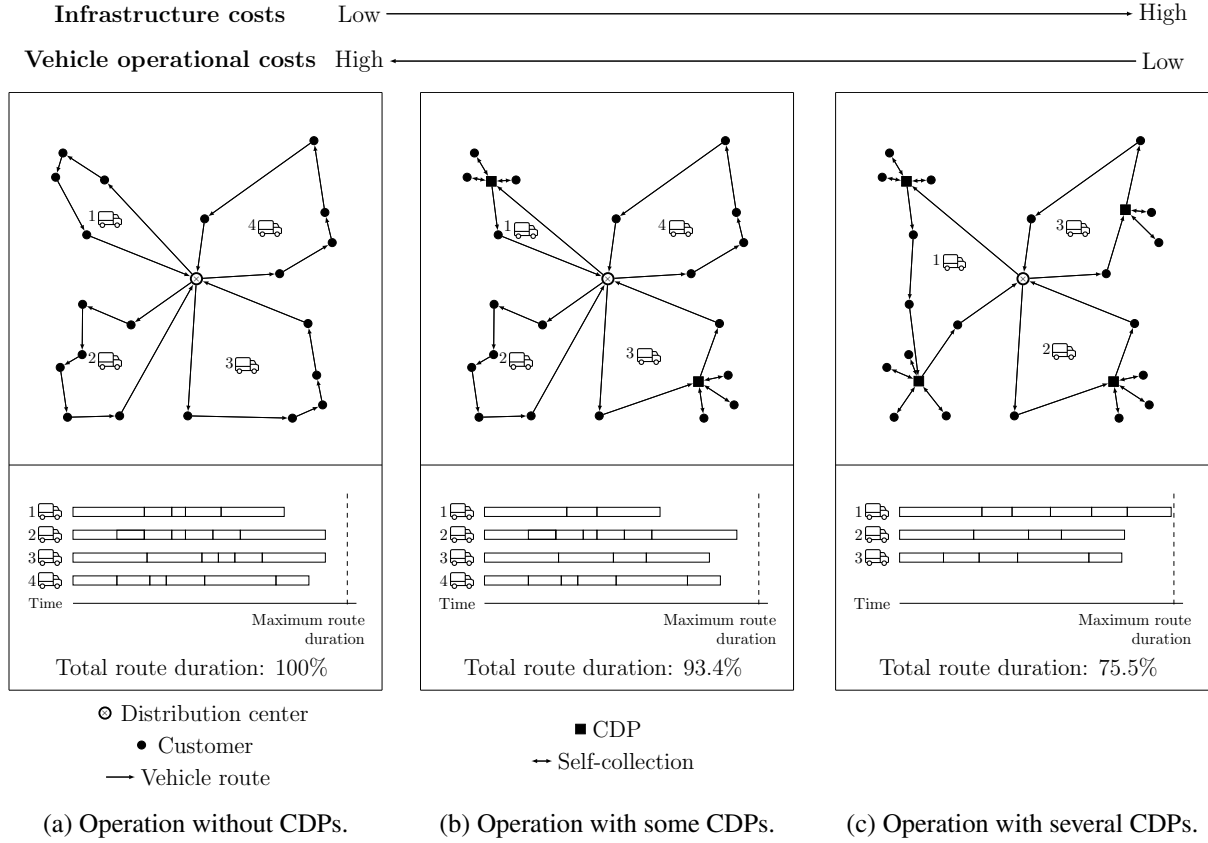


Figure 1: Operation without, and with some and several CDPs.

traditional routes to execute all delivery visits. Figure 1b presents the same setting but with two CDPs installed. In this setup, six customers are close enough to a CDP to prefer picking up their orders from this location; this reduces the number of delivery stops by four, and the total vehicle traveled time by 6.6%. We present a larger CDP network in Figure 1c. In this case, twelve customers choose to pick up their parcels at CDPs, reducing the number of delivery stops by eight and travel time by 24.5%.

Clearly, the potential benefits of installing CDPs depend on the number of customers who opt to use them

as order pickup stations. Therefore, one should understand customer preferences to predict the expected utilization of a given network size (Ma et al., 2022). Whether or not a customer chooses to pick up parcels themselves depends on the relative convenience of this option compared to home delivery. In this regard, decision-makers may encourage customers to use a CDP by (i) installing more, thereby increasing the network density and reducing the distance between customers and CDPs (Molin et al., 2022) or (ii) offering direct incentives to customers using these facilities, such as purchase discounts (*i.e.*, coupons) or delivery fee discounts (Yuen et al., 2018).

We study strategic decisions regarding the scale of a CDP network, anticipating their impact on customers' delivery service choices and downstream operational costs. Specifically, we consider decisions on the number of CDPs to install and the amount of economic incentive offered to customers willing to use them. In that sense, our objective is not to design a CDP network by determining specific locations, but rather to focus on its high-level CDP sizing decisions.

One possible approach to this problem is to rely on detailed stochastic discrete optimization. In that case, even evaluating a single network size and incentive configuration would require (i) determining the locations of the corresponding CDPs, (ii) simulating potential customer locations and delivery decisions, and (iii) solving a vehicle routing problem for each realization of customer demand and delivery service choices, with routing decisions made in the second stage and integrated with first-stage CDP location decisions. This simulation-based stochastic optimization approach becomes computationally intractable for reasonably sized instances, often making ad-hoc heuristics with limited analytical insight the only practical option. If our goal is to optimize the network size and incentive structure, an alternative is to leverage continuous approximation (CA) techniques, introduced in the seminal works of Beardwood et al. (1959) and Daganzo (1984), which provide tractable analytical expressions that capture the high-level trade-offs among the different costs, using only aggregated data.

Accordingly, we present a CA function that estimates the system's expected operational cost as the sum of infrastructure, distribution, and incentive costs. We then optimize the structural decision variables associated with the network size and the incentives offered to customers using CDPs. The types of tools we propose here are useful for high-level managers and decision-makers seeking to rapidly analyze multiple what-if scenarios, conduct pre-feasibility assessments, and support project budgeting decisions. In particular, our

models may be applied prior to detailed market research on potential sites, providing an initial estimate of an efficient network scale before making CDP location-specific decisions.

1.1 Contributions

We consider the following to be our main contributions.

1. We study a strategic CDP network sizing problem in which customers can be economically incentivized to use it. To the best of our knowledge, we are the first to propose a succinct CA function that evaluates infrastructure, distribution, and incentive costs, enabling tractable large-scale and what-if scenario analyses.
2. We build a stylized CA optimization model that jointly determines the optimal CDP network size and customer incentive levels, minimizing expected costs. The resulting model is computationally efficient, making it suitable for conducting strategic analyses and pre-feasibility studies.
3. We equip our CA-based optimization model with a chance constraint that limits the probability of CDP capacity violations. This constraint and capacity violations are subsequently evaluated in a computational study to assess practical feasibility.
4. We validate our approach on both synthetic and real-world road network instances, the latter departing from conditions under which CAs are typically known to perform well. In doing so, we provide a simple procedure to translate our continuous solutions into discrete CDP location decisions.
5. We conduct several sensitivity analyses on key parameters and derive managerial insights, summarized next.

1.2 Managerial Insights

Based on our approximations, sensitivity analysis, and validation experiments, we derive the following managerial insights.

1. In Section 4.5.1, we show that the economic value of CDP networks strongly depends on demand density. In particular, CDPs are especially effective in high-density settings, where shorter average

customer-to-CDP distances naturally increase adoption and improve delivery distribution efficiency.

2. The analysis in Section 4.5.1 also highlights the role of customer incentives in low-density settings, where incentives can partially compensate for limited network coverage by encouraging customers to travel longer distances to access CDPs.
3. CDP network sizing decisions critically depend on properly accounting for customer behavior, particularly customers' sensitivity to CDP distance (Section 4.5.2). We find that assuming deterministic customer choices based on simple distance thresholds may lead to significant estimation errors, especially when customers exhibit low distance sensitivity.
4. The optimal balance between CDP deployment and customer incentives changes with infrastructure costs (Section 4.5.3). As CDP infrastructure becomes cheaper, the model favors denser CDP networks and lower incentive levels.
5. As shown in Section 4.5.4, when the distribution center is located farther from the service region, stem travel times consume a larger fraction of each route's maximum duration, thereby increasing distribution costs. In such settings, the optimal strategy shifts from expanding the CDP network toward offering higher customer incentives.
6. There exists a trade-off between installing more CDPs and increasing each CDP's capacity to serve more customers (Section 4.5.5). Increasing capacity per CDP reduces the optimal number of installed CDPs while increasing the optimal customer incentive level for using them.
7. To derive our CA model, we assume uniformly distributed demand and metric travel cost. However, the model remains a useful decision-support tool even when these assumptions do not hold. Specifically, we show in Section 5 how it can be heuristically adapted to spatially non-uniform demand and non-metric travel costs.
8. Translating the CA model's continuous recommendations into a feasible and detailed design requires an additional discretization step and fine-tuning to account for operational and location-specific characteristics, particularly when capacity constraints are enforced. In Section 5, we propose two heuristic procedures for this purpose.

1.3 Paper Outline

The remainder of this paper is structured as follows. We revise the related literature in Section 2. In Section 3, we describe the studied problem and formulate our approximate model. In Section 4, we validate our model and derive managerial insights through multiple sensitivity analyses on key parameters. We further present a study inspired by real-world instance settings in Section 5. Finally, we conclude our work in Section 6.

2 Literature Review

We review the literature on the design and operation of CDP-based last-mile distribution systems. For comprehensive reviews on self-attended delivery and related operational challenges, as well as on the interaction between delivery operations and customer behavior, we refer to Janinhoff et al. (2024a) and Ma et al. (2022), respectively. To ensure broad coverage, our literature search covers several related terms used in the field, including collection and delivery points, parcel lockers, pick-up points, and out-of-home delivery systems.

We organize the literature into five main streams that structure the context of our study: (i) operational studies on routing and distribution decisions under an installed CDP network, (ii) strategic studies on CDP network design and planning, (iii) deployment of mobile CDPs, (iv) CA for related logistics problems, and (v) demand management in the last mile. We focus on the first two research streams, while the latter three provide complementary context to our work. Accordingly, Table 1 summarizes and compares operational and strategic studies based on the decisions considered, whether CDP capacity constraints are incorporated, the customer delivery service choice model adopted, and whether incentives are offered to customers who use CDPs. In our study, we add a tick in parentheses to indicate detailed facility location, as we also propose a simple procedure for translating continuous prescriptions into implementable solutions.

2.1 Operational Studies

At the operational level, the existing literature focuses on the efficient use of an installed CDP network and concentrates on last-mile distribution decisions. These studies largely address operational routing problems under a fixed CDP network infrastructure, without explicitly accounting for how strategic design choices

Table 1: A classification of operational and strategic studies involving CDPs.

Category	Reference	Decisions				CDP capacity constraints	Delivery choice				
		Facility location		Routing			Incentives	Endogenous (✓) / Exogenous (★)	Customer choice (✓) / Operator choice (★) / N.A. (–)	Probabilistic (✓) / Deterministic (★) / N.A. (–)	Incentive-sensitive (✓) / Insensitive (★) / N.A. (–)
		Discrete CA	Discrete CA								
Operational	Dumez et al. (2021)		✓			✓	✓	★	–	–	
	Tilk et al. (2021)		✓			✓	✓	★	–	–	
	Zang et al. (2023)		✓				✓	★	–	–	
	Dell’Amico et al. (2023)		✓			✓	★	–	–	–	
	dos Santos et al. (2022)		✓			✓	★	–	–	–	
	Janinhoff et al. (2024b)		✓			✓	✓	★	–	–	
	Mancini and Gansterer (2021)		✓			✓	✓	★	–	–	
	Galiullina et al. (2024)		✓		✓	✓	✓	✓	✓	✓	
	Akkerman et al. (2025)					✓	✓	✓	✓	✓	
Strategic	Deutsch and Golany (2018)	✓					✓	✓	★	–	
	Mancini et al. (2023)	✓				✓	✓	✓	★	–	
	Raviv (2023)	✓				✓	✓	✓	★	–	
	Kahr (2022)	✓				✓	✓	★	–	–	
	Zhou et al. (2024)	✓				✓	✓	★	–	–	
	Lin et al. (2020)	✓					✓	✓	✓	–	
	Lin et al. (2022)	✓					✓	✓	✓	–	
	Lyu and Teo (2022)	✓					✓	✓	✓	–	
	Xu et al. (2021)	✓				✓	✓	✓	✓	–	
	Enthoven et al. (2020)	✓		✓			✓	★	–	–	
	Ozyavas et al. (2025)	✓		✓		✓	✓	★	–	–	
	Zadeh et al. (2026)	✓		✓		✓	✓	★	–	–	
	Janinhoff and Klein (2023)	✓		✓		✓	✓	✓	✓	–	
	Janjevic et al. (2019)	✓		✓	✓	✓	✓	✓	★	–	
	Wang et al. (2026)	✓			✓		✓	✓	✓	–	
	Zhang et al. (2023)	✓			✓	✓	✓	✓	✓	✓	
		Our work	(✓)	✓		✓	✓	✓	✓	✓	✓

(such as the number or location of CDPs) may reshape downstream operational decisions and costs.

These problems typically relate to vehicle routing problem (VRP) models (see Toth and Vigo, 2014), and in particular to the VRP with delivery options (see Dumez et al., 2021; Tilk et al., 2021), in which each customer may be associated with multiple delivery locations, capturing the choice between home delivery and pickup at CDPs. As is common in the routing literature, several extensions of the VRP with delivery options that explicitly incorporate CDPs have been proposed, including variants with time windows (Zang et al., 2023), pickup and delivery settings (Dell’Amico et al., 2023), and two-echelon systems with occasional couriers (dos Santos et al., 2022).

More closely related to our work, Janinhoff et al. (2024b) use an extended VRP with delivery options model to analyze settings where customers face differentiated home-delivery prices. Similarly, Mancini and Gansterer (2021) propose a model in which customers can either be served at home or receive compensation to pick up their parcels at a CDP. However, neither Janinhoff et al. (2024b) nor Mancini and Gansterer (2021) explicitly treats these charges as decision variables that influence customers’ service choices, an aspect that we model in this paper.

In contrast, some papers examine the operational problem of planning vehicle routes while jointly offering incentives to customers using CDPs. While Galiullina et al. (2024) consider a setting in which incentives are offered after all customer realizations are known, Akkerman et al. (2025) study dynamically arriving customers to whom a menu of delivery options and incentives must be offered upon realization. Compared with Galiullina et al. (2024) and Akkerman et al. (2025), we investigate a setting in which incentives are determined prior to customer realization, *i.e.*, under pre-established incentive schemes.

2.2 Strategic Studies

At the strategic level, most studies focus on detailed CDP network design, *i.e.*, they build on facility location models. Given the vast literature on facility location problems (see Laporte et al. (2019) for a comprehensive review), we restrict our attention to those that specifically examine CDP locations. Within this stream, problems are studied under various modeling assumptions, including profit-maximizing objectives (Deutsch and Golany, 2018), uncertainty in customer demand and order processing capacity (Mancini et al., 2023), storage capacity constraints (Raviv, 2023), and facility layout decisions (Kahr, 2022). These works typically

assume a predefined set of candidate CDP locations and select a subset of them, whereas we aim to determine the optimal network size, *i.e.*, the number of CDPs

The research of Zhou et al. (2024) considers customers' satisfaction and proposes a fixed-incentive scheme that subsidizes their delivery costs; however, customer choice is treated as given to the system, as the logistics operator directly assigns demand to CDPs. More detailed studies integrate customer choice models into facility location problems, capturing the likelihood that customers opt to pick up their orders at CDPs rather than choose home delivery. These studies typically rely on discrete choice models in which the utility of a CDP decreases with the distance to the customer (Lin et al., 2020, 2022), or depends on additional factors such as customer location, CDP accessibility, and the availability of nearby alternatives (Lyu and Teo, 2022). Likewise, Xu et al. (2021) propose a data-driven approach to determine CDP locations based on historical demand data and purchase forecasts. All of these works treat customer CDP adoption as a predefined outcome and do not consider economic incentives as a decision lever to actively influence it, as we do in this paper.

Given the impact of strategic decisions on downstream distribution planning and costs, some strategic studies explicitly account for the interaction between CDP location and operational routing. These works relate to location-routing problems (see Nagy and Salhi, 2006) and, in particular, to the location-or-routing problem (Arslan, 2021), in which customers may either pick up their orders at nearby facilities or be served via delivery routes. Of these, some studies approximate the inherently stochastic customer arrival process by considering a deterministic set of customers (or service regions containing customers) as representative of daily operations and thereby solving a single routing scenario via metaheuristics (Enthoven et al., 2020) or detailed exact algorithms (Ozyavas et al., 2025; Zadeh et al., 2026).

Other works directly incorporate uncertainty in customer arrivals and service choices. For example, Janinhoff and Klein (2023) model a CDP location-routing problem as a two-stage stochastic integer program solved via a decomposition-based heuristic. Relatedly, Janjevic et al. (2019) study a multi-echelon network problem with CDPs, modeling demand changes induced by their placement. The resulting nonlinear integer program embeds CA formulas for routing costs and is solved using an ad-hoc heuristic. As an alternative to solving computationally hard, detailed integer programs, we propose a fully CA model to evaluate routing costs and jointly optimize the number of CDPs and the economic incentives offered to customers.

To our knowledge, the studies most closely related to ours are those of Wang et al. (2026) and Zhang et al. (2023). Wang et al. (2026) analyzes how CDP locations influence customers' choices of pickup points and how these choices, in turn, affect the total travel distance of last-mile distribution routes. In particular, they consider both settings in which CDPs and customer locations are served by separate routes and those in which they are integrated into a single route; our paper focuses on the latter structure. Regardless, their model does not account for infrastructure costs or the possibility of increasing CDP adoption through economic incentives, which are central features of our approach. Similarly, Zhang et al. (2023) studies joint location and service pricing decisions, modeling customer behavior through a user-equilibrium framework and approximating distribution costs via a CA approach. However, their analysis relies on a predefined set of candidate locations, where the utility perceived by a customer served at a CDP is driven by the service price and an exogenously defined attractiveness factor. In contrast, we treat the number of CDPs as a strategic decision variable and model incentives as a mechanism that directly modifies customers' distance-based utility, thereby shaping their service choices.

2.3 Mobile CDPs

Related studies complement both operational and strategic last-mile CDP models by introducing mobile CDPs, particularly mobile parcel lockers, as a flexible mechanism to improve delivery efficiency. These works are operational in nature, as they address day-to-day decision-making under dynamic demand conditions; however, they differ from classical routing studies in that they do operate over a mobile infrastructure over time.

Within this stream, some studies focus on routing mobile CDPs with static customer service preferences (Wang et al., 2020), while others incorporate adjustments that better align CDP locations and time-slot assignments with customer preferences (Schwerdfeger and Boysen, 2020). Additional contributions determine mobile CDP deployment in response to spatio-temporal demand patterns (Li et al., 2021), or use mobile CDPs both as delivery vehicles to customer homes and as pickup stations (Kötschau et al., 2023, 2025). In addition, Liu et al. (2023) examines how CDPs can work as mobile replenishment units transporting goods from depots to couriers serving customers.

In contrast to these studies, which primarily address operational and integrated deployment-routing

decisions, our work focuses on strategic-level infrastructure sizing decisions.

2.4 Continuous Approximation

Our study builds on the existing literature of CA models for vehicle routing costs. The seminal Beardwood–Halton–Hammersley (BHH) Theorem (Beardwood et al., 1959) states that the minimum length of a traveling salesman tour $L_n^\star$ over n independently and identically distributed (i.i.d.) locations scales as $\sqrt{n}$ almost surely as $n \rightarrow \infty$. Particularly, we can approximate $L_n^\star$ as

$$L_n^\star \approx \beta \cdot \sqrt{A \cdot n}, \quad (1)$$

where β is a proportionality constant and A is the covered service area. The study of Daganzo (1984) extends this approach to a multi-vehicle routing setting with stem distance costs. Recent applications of CA models include the design of same-day delivery systems (Stroh et al., 2022; Banerjee et al., 2023), the use of small unnamed vehicles (Jones et al., 2026; Zhang et al., 2026), and the development of dispatching and partitioning policies that account for fairness in customer service (Carlsson et al., 2024; Banerjee, 2026). For comprehensive overviews of the CA literature, we refer the reader to Franceschetti et al. (2017) and Ansari et al. (2018).

CA approaches have also been used in the analysis of logistics systems related to CDPs and facility location problems. For example, Dasci and Laporte (2005) use CA to analyze a location problem for competing retail firms that plan to operate within a specific geographic area. Li and Ouyang (2010) develop a model to address a reliable uncapacitated fixed charge location problem with the aim of minimizing initial investment costs and expected customer transportation costs. Ouyang and Daganzo (2006) apply a model to terminal design and propose a method to discretize the solution obtained with a small optimality gap. Pulido et al. (2015) define a model for locating warehouses in the context of same-day delivery operations. Carlsson and Jia (2015) present a model for a facility location problem in a two-echelon logistics distribution system, including fixed facility costs, backbone network costs to interconnect facilities, and transportation costs from facilities to end customers.

2.5 Demand Management in the Last-Mile

A related stream of research studies how customer behavior in last-mile logistics can be influenced through economic incentives and pricing mechanisms. This literature generally aims at creating temporal or spatial flexibility in customers, for instance, by encouraging customers to select convenient delivery time windows for attended home delivery (Klein et al., 2019; Yildiz and Savelsbergh, 2020), or by pricing delivery services across spatial zones to steer demand and improve system performance (Afsar et al., 2021; Afsar, 2022).

More recently, several studies have explored incentive mechanisms in crowd-based and platform-driven last-mile delivery systems using bilevel optimization formulations (Cerulli et al., 2024; Horner et al., 2024; Wang et al., 2024) and discrete choice models embedded into integer programming formulations (Çınar et al., 2024).

While these works differ in context and modeling assumptions, they consistently highlight the potential of customer incentives as a demand management tool in last-mile logistics.

3 The CDP Network Sizing Problem

We now introduce the CDP Network Sizing Problem (CDP-NSP), which aims to determine the optimal number of CDPs to install and the economic incentives offered to customers in order to minimize expected costs. In this CDP-based system, deliveries are made either to customers' locations or to CDPs, depending on the selected service option. Next, we formulate our CA cost function, propose customer service choice models, and integrate both into an optimization problem that determines the system's structural decisions subject to CDP capacity constraints.

3.1 Problem Description

The CDP-NSP arises at the strategic level to anticipate and balance infrastructure and routing costs of a last-mile distribution operation equipped with CDPs. We consider a fast delivery operation, such as next-day delivery, that serves each day a set of randomly distributed customers within a fixed service region who place their online orders on or before the previous day. Order deliveries are dispatched on vehicle routes from a distribution center to each customer's location of choice; specifically, customers may choose between having

their order delivered to their location (*i.e.*, home) or picking it up at a CDP. These customers' service choices can be influenced by economic incentives (such as cashback or delivery discounts), which, in turn, directly affect strategic sizing decisions, as encouraging customers to travel to a farther CDP may reduce the number of CDPs required to be installed. In summary, the decisions of the CDP-NSP involve determining the CDP network size and the economic incentive level so as to minimize the system's total expected cost, including CDP infrastructure expenses, last-mile routing costs, and incentive payments to customers who use CDPs.

In Figure 2, we illustrate our problem's dynamics. Figure 2a describes the operator's decisions; these are (i) sizing the CDP network over the service region and (ii) determining the economic incentive offered to customers picking up their parcels at CDPs. Here, incentive decisions are made prior to the realization of customer orders, since customers must be aware of the incentives when making a request. For ease of exposition, we consider circular CDP influence regions where customers located within these regions choose the corresponding CDP with certainty. The inner circles denote influence regions without economic incentives, while the outer circles represent expanded regions when economic incentives are offered to customers. Figure 2b presents one realization of customer requests and locations. Figure 2c depicts customers' delivery location choices. Finally, Figure 2d illustrates an order distribution plan for our example. The operator executes cost-efficient delivery routes starting and ending at the distribution center and visiting each home delivery location as well as each CDP serving at least one customer.

The CDP-NSP focuses on high-level decisions. We aim to anticipate expected costs, generate managerial insights, understand cost trade-offs, and evaluate structural decisions. Therefore, our approach intentionally avoids detailed CDP location and vehicle route planning. The detailed optimal location of CDPs is computationally hard and must account for all possible customer locations and service choice scenarios. Beyond location decisions, a detailed computation of expected last-mile distribution costs involves solving a hard vehicle routing problem for each customer demand scenario. Moreover, vehicle routing and location decisions are intertwined and depend on customer choices, defining where deliveries should occur.

3.2 Approximate Cost Model

In what follows, we present an approximate cost function for the last-mile distribution operation under study, depending on the number of installed CDP stations and the economic incentive offered to customers choosing

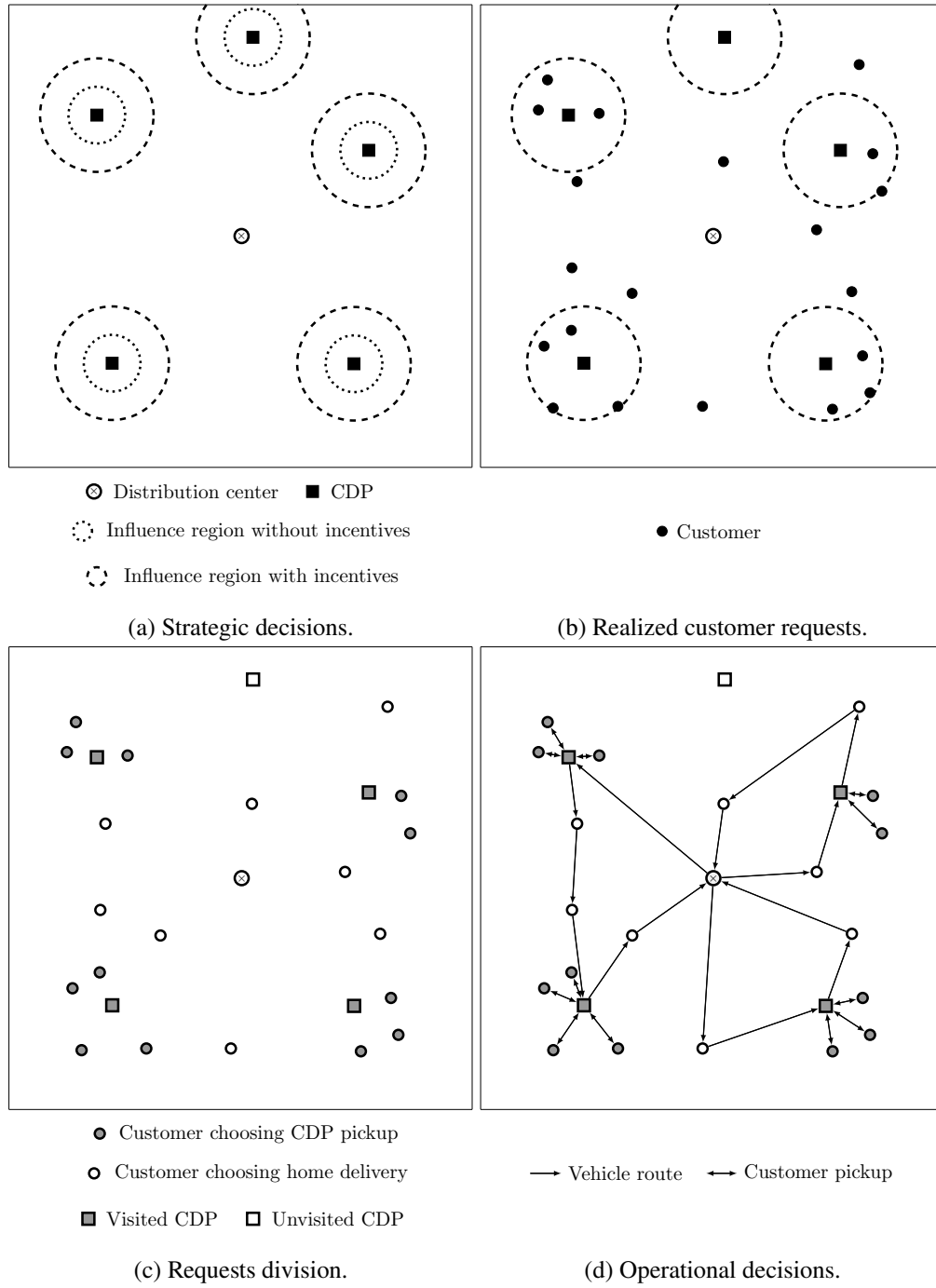


Figure 2: Dynamics of the CDP network sizing problem.

to pick up their requests at a CDP. To distinguish the discrete number of installed CDPs from its CA, we use $\bar{x} \in \mathbb{N}_0$ for the former and $x \geq 0$ for the latter. We also denote the incentive by $y \geq 0$. Our cost function only requires aggregated data, such as the expected number of customers n and the area A of the service region.

We assume that customer locations are uniformly and independently distributed over the service region, that each customer places a single order, and that each customer chooses between home delivery and picking up their order at the nearest CDP, depending on their relative distance to it and on y . This is consistent with empirical evidence indicating that approximately 79% of customers prefer CDPs located close to their residences (Lemke et al., 2016; Galiullina et al., 2024).

Moreover, we avoid setting the exact location of the CDP network and assume that the CDPs are evenly distributed over the service area, forming hexagonal influence regions, which is a reasonable assumption for a uniformly distributed demand (Newell, 1973; Papadimitriou, 1981; Carlsson and Jia, 2015). In the optimization model presented later in Section 3.4, we consider CDPs with a maximum capacity of q orders; the value of q may be either finite or unbounded, with the latter corresponding to the unconstrained-capacity setting. CDP capacity q represents the maximum design capacity in terms of daily customer orders dispatched to the CDP, accounting for customers' retrieval times. If customer pickup occurs within the delivery day, q is understood as the physical capacity of the CDP. Within our cost function, we also estimate last-mile distribution costs via CA instead of relying on detailed vehicle route planning tools.

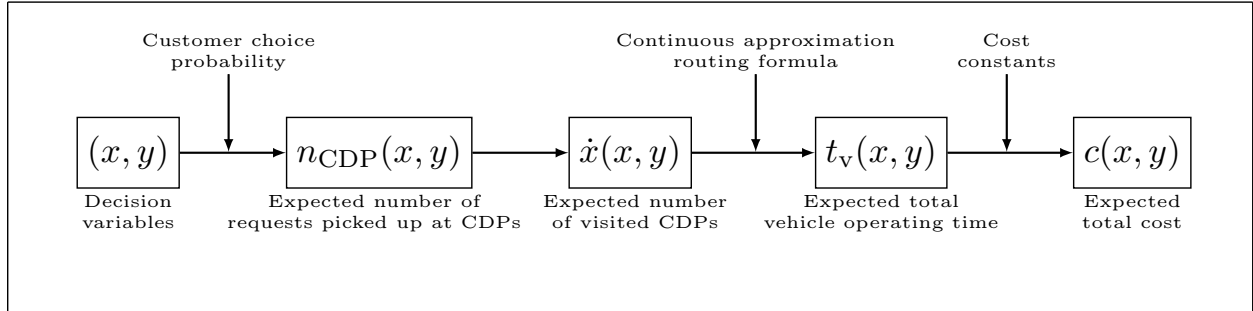
Our approximate function adds the following cost components: (i) a cost α_x per installed CDP, dependent on its capacity q , representing rent, maintenance, insurance, tax, and related infrastructure costs; (ii) a last-mile distribution cost α_v per unit time operated by each vehicle, representing vehicle lease and driver wage; and (iii) an economic incentive y paid to each customer collecting their goods at CDPs.

Regarding vehicle operating times, we consider that each vehicle incurs a fixed stop time t_{st} plus an additional service time t_{se} per delivered order. We further account for the fact that the distribution center may be located outside the service area; therefore, each route includes a stem time t_0 . We also impose that the vehicle's maximum workday length (denoted by $T_{\max}$) is the limiting resource, as we study the delivery of small-sized and lightweight parcels and disregard vehicle capacity constraints. Vehicles are assumed to travel at a constant speed of s .

Table 2 summarizes our notation, and Figure 3 provides an overview of the approximate cost model.

Table 2: Parameter and variable notation

Notation	Parameters	Unit
n	Expected number of requests.	requests/day
A	Service area.	km ²
t_0	Expected time between the distribution center and a customer.	hours
t_{se}	Variable service time per delivered request.	hours/request
t_{st}	Fixed time per stop.	hours/stop
T_{max}	Maximum vehicle workday length.	hours/day
s	Average vehicle speed.	km/hours
q	CDP capacity.	requests/CDP
α_x	CDP cost.	\$(CDP \cdot \text{day})
α_v	Vehicle operating time cost.	\$/hour
Decision Variables		
x	Number of installed CDPs.	CDPs
y	Incentive per customer choosing the CDP option.	\$/request
Auxiliary Variables		
$P_{CDP}(\omega, y)$	Customer CDP choice probability.	-
$n_{CDP}(x, y)$	Expected number of requests picked up at CDPs	requests/day
$\dot{x}(x, y)$	Expected number of visited CDPs.	CDPs
$t_v(x, y)$	Expected total vehicle operating time.	hours/day
$c(x, y)$	Expected total cost.	\$/day



x : Number of installed CDPs.

y : Incentive per customer choosing the CDP option.

Figure 3: Outline of the steps used to derive our approximate cost function.

The cost function, denoted as $c(x, y)$, is defined in a series of approximation steps starting from a pair of values $x, y \geq 0$. First, we estimate the expected number of requests picked up by customers at CDPs (referred to as $n_{\text{CDP}}(x, y)$) by assuming a customer-invariant probability function $P_{\text{CDP}}(\omega, y)$, which represents the likelihood of a customer choosing to pick up their order at the nearest CDP at a distance ω from their location when offered an incentive of y . Then, we estimate the expected number of CDPs visited per day $\dot{x}(x, y) \leq x$, as not all CDPs installed are expected to have demand each day. Next, we leverage CA for routing problems, together with the time components $t_0, t_{\text{se}}, t_{\text{st}}$, the time limit T_{max} , and the vehicle speed s , to estimate the total vehicle operating time $t_v(x, y)$ required to serve the $n - n_{\text{CDP}}(x, y)$ customers receiving home delivery and the $\dot{x}(x, y)$ selected CDP locations. Finally, we integrate the per-CDP cost $\alpha_x(q)$ and the per-hour cost α_v to compute $c(x, y)$. Each approximation step is detailed as follows.

Expected number of requests picked up at CDPs. Consider an integer number of CDPs $\bar{x} \in \mathbb{N}_0$ and let Ω_k denote the random variable representing the distance between the k^{th} customer and their nearest CDP. Further, define $B_k(\bar{x}, y) \sim \text{Bernoulli}(P_{\text{CDP}}(\Omega_k, y))$ a binary random variable that takes the value 1 if the k^{th} customer chooses to collect their parcels at the nearest CDP, and 0 otherwise. Then, the random number of requests picked up at CDPs, $N_{\text{CDP}}(\bar{x}, y)$, is represented as

$$N_{\text{CDP}}(\bar{x}, y) = \sum_{k=1}^n B_k(\bar{x}, y), \quad (2)$$

and its expected value is

$$n_{\text{CDP}}(\bar{x}, y) = \mathbb{E}[N_{\text{CDP}}(\bar{x}, y)] = \sum_{k=1}^n \mathbb{E}[B_k(\bar{x}, y)] = \sum_{k=1}^n \mathbb{E}[P_{\text{CDP}}(\Omega_k, y)]. \quad (3)$$

To estimate and extend $n_{\text{CDP}}(x, y)$ to the continuous domain $x \geq 0$, we propose an approximation in which the hexagonal influence region of each CDP is approximated as a circle of radius $r(x) = \sqrt{A/(\pi \cdot x)}$ if $x > 0$ and $r(x) = 0$ else; this follows from assuming that the total service area of all installed CDPs is equal

to the area of the whole service region, *i.e.*, $x \cdot \pi \cdot r(x)^2 = A$. Therefore, for $x \geq 0$, we obtain

$$n_{\text{CDP}}(x, y) \approx \begin{cases} x \cdot \frac{n}{A} \cdot \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\text{CDP}}(\omega, y) d\omega d\theta, & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where $n/A \cdot \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\text{CDP}}(\omega, y) d\omega d\theta$ estimates the expected number of customers served by a single CDP, under the assumption that each customer located within its service region chooses it with probability $P_{\text{CDP}}(\omega, y)$.

Expected number of visited CDPs. Only CDPs chosen as pickup stations by at least one customer should receive delivery visits. To estimate the expected number of visited CDPs, for all $x \geq 0$, we denote by $p(x, y)$ the probability that a customer chooses pickup at a CDP. Given that customer demand is evenly distributed across the installed CDPs and that customer choices are independent, we obtain

$$p(x, y) = \begin{cases} \frac{n_{\text{CDP}}(x, y)}{n \cdot x}, & \text{if } x, n > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

i.e., the probability that a customer chooses a particular CDP is the proportion of customers assigned to that CDP; by construction, $p(x, y)$ is valid for both continuous and integer values of the number of CDPs x .

Next, consider a fixed $\bar{x} \in \mathbb{N}_0$ and define $V_i(\bar{x}, y)$ for each $i \in \{1, \dots, \bar{x}\}$ as a binary random variable taking value 1 when the i^{th} CDP is left unused, and 0 otherwise. It follows that $V_i(\bar{x}, y) \sim \text{Bernoulli}((1 - p(\bar{x}, y))^n)$; in this expression, $1 - p(\bar{x}, y)$ represents the probability that a customer does not select a given CDP, and $(1 - p(\bar{x}, y))^n$ represents the probability that a given CDP is not selected by any of the n customers. Let the random variable representing the number of visited CDPs be

$$\dot{X}(\bar{x}, y) = \sum_{i=1}^{\bar{x}} (1 - V_i(\bar{x}, y)). \quad (6)$$

Its expected value is defined as

$$\dot{x}(\bar{x}, y) = \mathbb{E} [\dot{X}(\bar{x}, y)] = \bar{x} - \mathbb{E} \left[\sum_{i=1}^{\bar{x}} V_i(\bar{x}, y) \right]. \quad (7)$$

We finally approximate and extend $\dot{x}(x, y)$ to all $x \geq 0$ using the probability $p(x, y)$, obtaining

$$\dot{x}(x, y) \approx \begin{cases} x \cdot \left(1 - (1 - p(x, y))^n\right), & \text{if } x, n > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Expected total vehicle operating time. In our setting, for $\bar{x} \in \mathbb{N}_0$, the number of delivery stops $M(\bar{x}, y)$ is the sum of the number of used CDPs plus the number of home deliveries, *i.e.*, $M(\bar{x}, y) = \dot{X}(\bar{x}, y) + n - N_{\text{CDP}}(\bar{x}, y)$. Therefore, considering the service time per request, the fixed time per stop, and a square-root planning approximation for the routing time between stops, we estimate the vehicles' operating time (*i.e.*, excluding stem routing) as $T_{\text{op}}(\bar{x}, y) \approx \gamma \cdot t_{\text{se}} \cdot n + \delta \cdot t_{\text{st}} \cdot M(\bar{x}, y) + \eta/s \cdot \sqrt{M(\bar{x}, y) \cdot A}$. Constants γ , δ and η are later calibrated and validated in Section 4.2. To approximate the expected vehicle operating time $t_{\text{op}}(\bar{x}, y) = \mathbb{E}[T_{\text{op}}(\bar{x}, y)]$ and define it over all $x \geq 0$, we replace the number of stops $M(\bar{x}, y)$ by its expected value $m(x, y) \approx \dot{x}(x, y) + n - n_{\text{CDP}}(x, y)$. Therefore, we obtain

$$t_{\text{op}}(x, y) \approx \gamma \cdot t_{\text{se}} \cdot n + \delta \cdot t_{\text{st}} \cdot m(x, y) + \frac{\eta}{s} \cdot \sqrt{m(x, y) \cdot A}, \quad (9)$$

which overestimates the expected routing time as the square-root function is concave. We add stem routing times to Equation (9) following the approach proposed by Daganzo (1984) and estimate the expected total vehicle operating time as

$$t_v(x, y) \approx 2 \cdot t_0 \cdot \frac{t_v(x, y)}{T_{\text{max}}} + t_{\text{op}}(x, y), \quad (10)$$

which is defined as the sum of twice the expected travel time between the distribution center and a customer ($2 \cdot t_0$) times the estimated number of vehicle routes dispatched ($t_v(x, y)/T_{\text{max}}$) plus the remaining local operating time $t_{\text{op}}(x, y)$, as illustrated in Figure 4. If we define $\kappa = T_{\text{max}}/(T_{\text{max}} - 2 \cdot t_0)$ and isolate $t_v(x, y)$

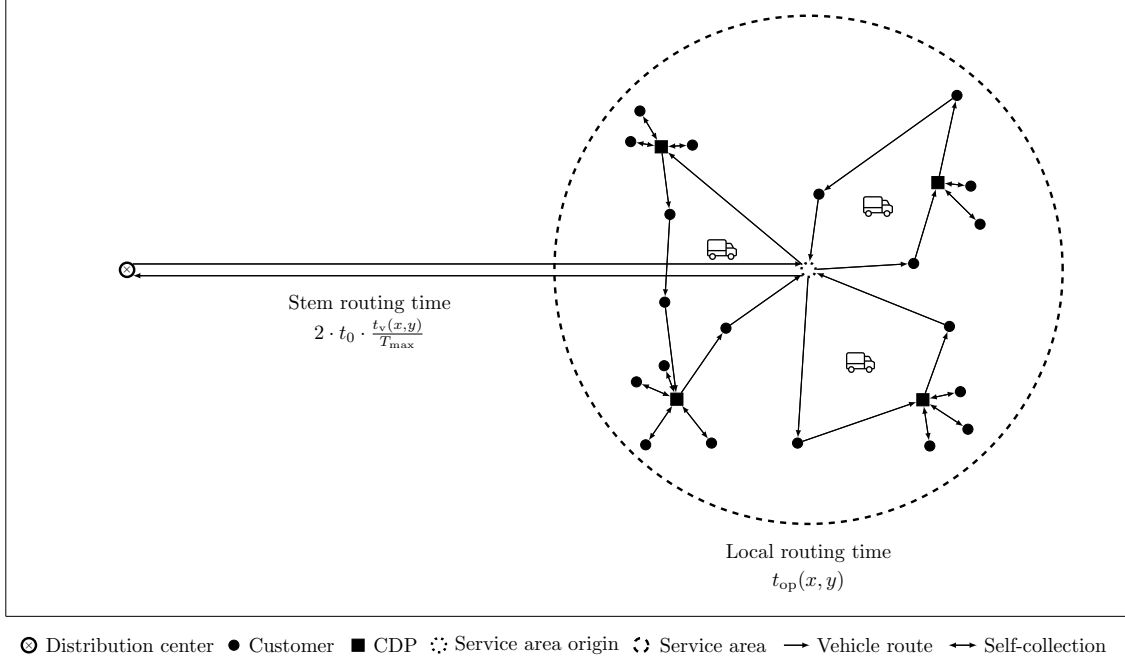


Figure 4: Expected total vehicle operating time approximation.

from Equation (10), then we obtain

$$t_v(x, y) \approx \gamma \cdot \kappa \cdot t_{\text{se}} \cdot n + \delta \cdot \kappa \cdot t_{\text{st}} \cdot m(x, y) + \eta \cdot \frac{\kappa}{S} \cdot \sqrt{A \cdot m(x, y)}. \quad (11)$$

From the definition of κ , the operation becomes infeasible and the total vehicle operating time diverges to infinity when the maximum vehicle workday length $T_{\max}$ approaches $2 \cdot t_0$ from above. In contrast, when $T_{\max} \rightarrow \infty$, then the total vehicle operating time $t_v(x, y)$ converges to $t_{\text{op}}(x, y)$, *i.e.*, the CA formula for the operating time of a traveling salesman tour. Note that a setting with capacitated vehicles can also be modeled with this framework by defining a proxy for the number of routes required as the total served demand divided by the capacity of each vehicle in this stem-routing cost component.

Expected total cost. Finally, for all $x, y \geq 0$, we approximate the logistics operator's total cost, including CDP infrastructure, vehicle distribution, and incentive costs as

$$c(x, y) \approx \alpha_x \cdot x + \alpha_v \cdot t_v(x, y) + y \cdot n_{\text{CDP}}(x, y). \quad (12)$$

3.3 CDP Choice Probability Function

In this section, we propose two possible CDP service choice probability functions $P_{\text{CDP}}(\omega, y)$ depending on the customer's distance ω to their closest CDP and the incentive y ; the first is a deterministic threshold rule, while the second is derived from a binomial logit choice model. As an auxiliary model, we define $\tau(y)$ as the distance a customer is willing to travel to pick up their parcels at their closest CDP given an incentive y . Therefore, we refer to $\tau(y)$ as the customers' indifference distance. The function $\tau(y)$ is assumed to be non-decreasing in y , as relatively more incentivized customers are more willing to travel longer distances.

Distance threshold choice (DTC). As presented in Equation (13), this choice model assumes that customers choose the CDP pickup option with certainty if this station is located within a distance less than or equal to $\tau(y)$ from their location; otherwise, they opt for home delivery.

$$P_{\text{CDP}}^{\text{DTC}}(\omega, y) = \begin{cases} 1, & \text{if } \omega \leq \tau(y), \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

If we replace this specific probability model in Equation (4), then the expected number of requests picked up at CDPs is approximated by

$$n_{\text{CDP}}(x, y) \approx \begin{cases} n \cdot \min \left\{ 1, \left(\frac{\tau(y)}{r(x)} \right)^2 \right\}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

A short derivation of this result is provided in Appendix A.1. Intuitively, Equation (14) states that the number of expected pickups at CDPs is proportional to the ratio between the customer's CDP acceptance area $\pi \cdot \tau(y)^2$ over the CDP's influence area $\pi \cdot r(x)^2$. This threshold-based model may be too simplistic, as customers' behavior is likely to show variability in practice. As follows, we generalize this choice model and account for random customer behavior via a logit choice model.

Binomial logit choice (BLC). We now consider a model in which customers choose to pick up their parcels at a CDP depending on a binomial logit choice model (McFadden, 1973) defined as

$$P_{\text{CDP}}^{\text{BLC}}(\omega, y) = \frac{1}{1 + e^{\lambda \cdot (\omega - \tau(y))}}, \quad (15)$$

where λ is the BLC model scale parameter that represents the customer-to-distance sensitivity, and $\tau(y) - \omega$ denotes a customer's perceived utility for choosing the CDP pickup option. If we replace this functional form of $P_{\text{CDP}}(\omega, y)$ in Equation (4) we obtain

$$n_{\text{CDP}}(x, y) \approx \begin{cases} n \cdot \left(1 + \frac{2 \cdot \text{Li}_2(-e^{-\lambda \tau(y)})}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \text{Li}_2(-e^{\lambda \cdot (r(x) - \tau(y))})}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \ln(1 + e^{\lambda \cdot (r(x) - \tau(y))})}{\lambda \cdot r(x)} \right), & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

where $\text{Li}_2(z) = -\int_0^z \ln(1-u)/u \, du$ is the dilogarithm function (Zagier, 2007) defined for $z \leq 1$. The detailed step-by-step algebraic procedure to obtain Equation (16) is presented in Appendix A.2.

The DTC model is a special case of the BLC model (Gallego and Topaloglu, 2019). In particular, the BLC model becomes an all-or-nothing DTC model when customers are highly sensitive to the difference $\tau(y) - \omega$ (*i.e.*, when $\lambda \rightarrow \infty$). In contrast, when customers are insensitive to $\tau(y) - \omega$ (*i.e.*, when $\lambda \rightarrow 0$), they choose either option with equal probability.

3.4 Total Cost Minimization and Capacity Constraints

We can minimize the total cost function by solving

$$\min_{x, y \geq 0} c(x, y), \quad (17)$$

to approximately optimize structural decisions (x, y) for our last-mile distribution operation equipped with a network of CDPs and incentives; Model (17) is equivalent to a profit-maximizing problem with fixed delivery prices, in which the incentives are considered as lost revenue from customers who choose CDP pickup.

At the strategic level, we intentionally allow the number of CDPs x to take fractional values. However, some of the model validations discussed below are conducted with an integer number of CDPs. In such

cases, our model can be evaluated for each integer value of x , optimizing the corresponding incentive level y for each x . Moreover, we may also seek to account for capacity shortages. To this end, we design the CDP-based last-mile delivery system so that the capacity of each CDP is not exceeded. Specifically, we now present a method to incorporate CDP capacity constraints into model (17).

Consider an integer number of CDPs $\bar{x} \in \mathbb{N}_0$, and let $N_{\text{CDP},i}(\bar{x}, y)$ denote the number of requests assigned to the i^{th} CDP for $i \in \{1, \dots, \bar{x}\}$. The capacity of CDP i is exceeded whenever $N_{\text{CDP},i} > q$. Therefore, for an integer number of CDPs, we aim to solve (17) subject to the following set of chance constraints

$$\mathbb{P}(N_{\text{CDP},i}(\bar{x}, y) > q) \leq \zeta \quad \forall i \in \{1, \dots, \bar{x}\}, \quad (18)$$

which ensures that the probability of exceeding the service capacity at each CDP is at most $\zeta \in (0, 1]$. Recall that $p(x, y)$ denotes the probability that a customer chooses pickup at a CDP, and that $p(x, y)$ is defined for both continuous and integer values of x . Consequently, $N_{\text{CDP},i}(\bar{x}, y) \sim \text{Binomial}(n, p(\bar{x}, y))$. Since $N_{\text{CDP},i}(\bar{x}, y)$ is defined for integer values $\bar{x} \in \mathbb{N}_0$, the corresponding Binomial chance constraint is discrete.

As follows, we extend this chance constraint to a continuous number of CDPs. From the properties of the Binomial distribution (Johnson et al., 2005), we have $\mathbb{P}(N_{\text{CDP},i}(\bar{x}, y) > q) = 1 - I_{1-p(\bar{x}, y)}(n - q, q + 1)$, where $I_z(a, b)$ is the regularized incomplete Beta function evaluated at z , with parameters a and b . Therefore, the continuous chance constraint for $x \geq 0$ can be written as

$$1 - I_{1-p(x, y)}(n - q, q + 1) \leq \zeta. \quad (19)$$

If we equip model (17) with constraint (19), the resulting formulation is a nonconvex continuous optimization problem. Nevertheless, we can impose $0 \leq x \leq x^{\max}$ and $0 \leq y \leq y^{\max}$ for appropriate upper bounds $x^{\max}$ and $y^{\max}$. With these bounds, the problem is two-dimensional and defined over a compact domain, and can therefore be quickly solved via a small enough grid search over (x, y) or by using any standard nonlinear optimization solver.

4 Analysis of Structural Decisions and Approximate Cost

In this section, we calibrate and validate our approximate cost function $c(x, y)$, comparing it against detailed vehicle-routing simulations. This validation has two complementary objectives: assessing how accurately the function estimates costs for given decisions and determining whether it leads to decisions whose simulated costs are close to the minimum over the evaluated decision grid. We then use the calibrated model to analyze the structure of its optimal decisions, x and y , across a range of last-mile distribution settings and to derive managerial insights through sensitivity analysis. A complementary validation on a real road network is presented in Section 5.

All experiments, including the solution of the CA optimization model and the associated VRP instances solved in each simulation, were conducted on a personal computer equipped with a 2.00 GHz AMD Ryzen 7 4700U processor and 8 GB of RAM.

In our numerical experiments, we set maximum number of CDPs $x^{\max} = n$, so that the number of CDPs is bounded above by the number of requests. Whenever $x > 0$, we also choose a maximum possible incentive level $y^{\max}$ such that $P_{\text{CDP}}^{\text{BLC}}(R, y^{\max}) \geq 1 - 10^{-6}$, ensuring that customers choose service at CDPs with high probability if $y = y^{\max}$. By construction of model (17) and constraint (19), if the number of CDPs is set to $x = 0$, then the total expected cost collapses to the expected distribution cost to all customer's locations, independently of incentive y . Without loss of generality, we report $y = 0$ when $x = 0$.

We begin by defining the parameter settings used throughout our study and calibrating the coefficients of our approximate model.

4.1 Base Parameter Values in our Study

Here we set the parameter values used to calibrate and test our approximate cost model for a base case, as shown in Table 3. We consider a daily operation expecting $n = 200$ requests within a circular service region of radius $R = 3$ km, which results in a request density $\rho \approx 7.07$ requests/(km² · day); this value is comparable to the 6.5 requests/(km² · day) presented by Janjevic et al. (2019). The average vehicle speed is assumed to be $s = 15$ km/hour, similar to the values suggested in the case study conducted by Allen et al. (2018). The distribution center is located at the center of the service region. Under this assumption, the expected distance to a uniformly distributed customer is $2R/3$, yielding an expected travel time to any customer of

Table 3: Parameter values used throughout our study.

Notation	Parameter	Value(s)	Unit
n	Expected number of requests.	200	requests/day
A	Service area.	$\pi R^2 \approx 28.27$	km ²
t_0	Expected time between the distribution center and a customer.	$2 \cdot R / (3 \cdot s) \approx 0.13$	hours
t_{se}	Variable service time per delivered request.	0.5/60	hours/request
t_{st}	Fixed time per stop.	3.5/60	hours/stop
$T_{\max}$	Maximum vehicle workday length.	5	hours/day
s	Average vehicle speed.	15	km/hour
q	CDP capacity.	∞	requests/CDP
α_x	CDP cost.	30	\$/ (CDP · day)
α_v	Vehicle operating time cost.	30	\$/hour

$t_0 = 2R/(3s) \approx 0.13$ hours. We define a service time of $t_{se} = 0.5$ minutes per parcel delivered. Also, the time spent per vehicle stop in a service route is $t_{st} = 3.5$ minutes. We consider a workday length of $T_{\max} = 5$ hours per day, which corresponds to the maximum driving time allowed for a worker under Chilean labor law (Law No. 20,271, Art. 25 bis., 2008). In the base case, we consider uncapacitated CDPs (*i.e.*, $q = \infty$). This simplifying assumption is later dropped in Section 4.5.5 and Section 5. The cost paid per installed CDP is set to $\alpha_x = \$30/(\text{CDP} \cdot \text{day})$, which lies within the \$16 to \$34 range presented by Xu et al. (2021). The hourly cost per operating vehicle is set to $\alpha_v = \$30/\text{hour}$ based on a \$28/hour driver salary and a \$0.14/km fuel cost.

We model the customers' indifference distance to a CDP as $\tau(y) = \tau_1 + \tau_2 \cdot y$, where $\tau_1 = 1$ km and $\tau_2 = 1$ km/\$. In our BLC model, such values set a 50% probability of choosing a CDP at a 1 km distance when no incentives are offered. Also, this distance value increases in one kilometer per \$1 incentive offered. The logit scale parameter is set to $\lambda = 3$, consistent with Janjevic et al. (2019), who state that the maximum distance a customer is willing to travel to pick up their parcels is within the 3 to 6 km range. In particular, Figure 5a plots the CDP choice probability $P_{\text{CDP}}^{\text{BLC}}(\omega, y)$ over different values of incentive y : our choice of λ is conservative as $P_{\text{CDP}}^{\text{BLC}}(\omega = 3, y = 0) \leq 0.25\%$. As expected, $P_{\text{CDP}}^{\text{BLC}}(\tau(y), y) = 50\%$. Figure 5b depicts $P_{\text{CDP}}^{\text{BLC}}(\omega, y)$ for different values of λ when $y = \$0/\text{request}$; a customer's CDP service choice becomes less variable and more predictable when λ increases.

4.2 Total Cost Function Calibration and Validation

In this subsection, we calibrate the coefficients of our approximate travel time model (*i.e.*, γ , δ , and η in Equation (10)) and then perform two validations. First, we assess the predictive accuracy of the CA cost

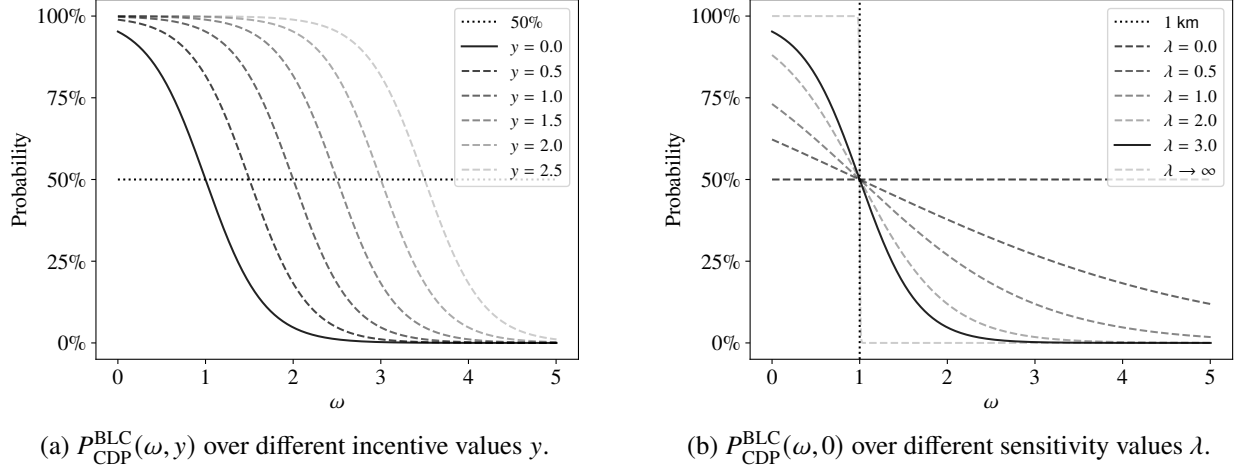


Figure 5: CDP choice probability $P_{\text{CDP}}^{\text{BLC}}(\omega, y)$.

function by comparing its estimates with simulated operating costs across in-sample and out-of-distribution settings. Second, we evaluate the model's decision-making performance by comparing the (x, y) solution recommended by the CA model with the minimum simulated cost over a discrete decision grid. We take the parameter values in Table 3 and build an instance set I containing tuples $(n, t_{\text{st}}, T_{\text{max}}, x, y)$, where $n \in \{100, 200, 400\}$, $t_{\text{st}} \in \{1/60, 3.5/60, 10/60\}$, and $T_{\text{max}} \in \{3, 5, 8\}$, representing various levels of request densities, stopping and parking times, and workday durations, respectively. Also, we consider the pair of decision variables $(x, y) \in \{(0, 0)\} \cup (\{1, \dots, 40\} \times \{0, 0.5, 3\})$, as it only makes sense to offer incentives when at least one CDP is installed.

For each instance $i = (n, t_{\text{st}}, T_{\text{max}}, x, y) \in I$, we evenly distribute the x CDPs over the circular region, generating a hexagonal grid of CDP influence regions, and simulate a set Ψ of 100 customer demand scenarios. For each demand scenario $\psi \in \Psi$, we draw n uniformly distributed customer requests over the service region and randomly generate each customer's service choice (*i.e.*, pickup at CDP or home delivery) depending on their distance ω to the closest CDP and incentive y as established in the BLC probability function $P_{\text{CDP}}^{\text{BLC}}(\omega, y)$. Consequently, for each instance $i \in I$ and demand scenario $\psi \in \Psi$, we generate a set of home-delivery requests and a set of pickup requests at specific CDPs. Given each pair (i, ψ) , we compute its last-mile distribution cost using the PyVRP heuristic solver (version 0.12.2) (Wouda et al., 2024) under its default parameter settings and a running time limit of 1 minute.

Of these 100 demand scenarios per instance, we use 80 as a calibration sample to estimate parameters γ ,

δ and η in the total travel time function $t_v(x, y)$ defined in Equation (11) via ordinary least squares regression, adjusting our model's output value $t_v(x, y)$ to each instance's average travel time over 80 demand scenarios. The resulting calibration yields an adjusted R^2 coefficient of 99.6%, indicating that the calibrated model explains nearly all of the variation in the observed costs. The calibrated coefficient values are $\gamma \approx 1.106$, $\delta \approx 1.043$, and $\eta \approx 0.751$, which is reasonable since γ and δ are close to 1 and η is similar to the theoretical and empirical values discussed in the CA literature (Franceschetti et al., 2017).

In addition to the remaining 20 test demand scenarios in instance set I , we construct two out-of-distribution sets, I_1 and I_2 . Our goal is to quantify the accuracy of our approximations when some of their underlying assumptions are deliberately violated. For I_1 , we drop the assumption of uniform spatial demand and instead consider a clustered spatial distribution with radial exponential decay. Specifically, we first generate five cluster centers uniformly within the service region. Then, each customer is assigned to one of these clusters and positioned at a random angle around its associated center. The radial distance r between the customer and its assigned cluster center is sampled from a truncated exponential distribution with density $f(r) = (0.05e^{-0.05r})/(1 - e^{-0.05R})$, for $0 \leq r \leq R$, where R denotes the maximum feasible distance before leaving the service region. This construction ensures that all customers remain within the service region and induces a higher concentration of customers near each cluster center, with probability mass decreasing exponentially as a function of distance. For I_2 , we drop the homogeneous customer behavior assumption and allow for heterogeneous utility parameters across customers. In particular, for each customer, we sample λ , τ_1 , and τ_2 from normal distributions truncated below at zero, with means and standard deviations of (3, 0.5), (1, 0.1), and (1, 0.1), respectively. For both I_1 and I_2 , we fix the parameters $n = 200$, $t_{st} = 3.5/60$, and $T_{max} = 5$ hours.

Table 4 presents the mean absolute percentage error (MAPE) for each component of our CA model compared to its empirically simulated counterpart. For the base instance set I , the MAPE of $n_{CDP}(x, y)$ and

Table 4: MAPE of the CA model component in our validation sample.

	$c(x, y)$	$n_{CDP}(x, y)$	$\dot{x}(x, y)$	$t_v(x, y)$
I	1.13%	2.89%	0.33%	4.04%
I_1	1.67%	5.07%	4.97%	5.84%
I_2	1.25%	3.35%	0.35%	4.28%

$t_v(x, y)$ are larger compared to the error of the total cost function $c(x, y)$, whereas the error of $\dot{x}(x, y)$ is

small in comparison to $c(x, y)$. This suggests that these three components' errors tend to offset each other, resulting in a more accurate overall fit of the model.

A similar pattern is observed for the out-of-distribution instance sets I_1 and I_2 ; however, these sets exhibit slightly higher average errors compared to I , reflecting the violation of some of our main modeling assumptions. Notably, the average cost error increases only marginally, from 1.13% to 1.67% in the case of I_1 and to 1.25% in the case of I_2 . The prediction error of $\hat{x}(x, y)$ remains relatively small for I and I_2 , at 0.33% and 0.35%, respectively, but increases to 4.97% in I_1 ; this suggests that the number of actually visited CDPs is more sensitive to customers' spatial distribution than to customer selection. The errors of $n_{\text{CDP}}(x, y)$ and $t_v(x, y)$ increase only moderately under both out-of-distribution settings.

So far, the validation process compares our CAs against simulations of the system using a heuristic vehicle routing solver. However, this may introduce a bias, as our continuous cost function is intended to approximate the true operational cost of the underlying routing problem rather than the value returned by a particular heuristic algorithm. In Appendix B.1, we revisit a subset of the experiments on instance set I , this time using a solver with optimality guarantees with a time limit of 1 minute, and considering a single demand scenario. The results remain consistent with those previously reported. In particular, our approximations deteriorate only marginally when evaluated against near-optimal vehicle routes, with increases of no more than a fraction of a percentage point and an overall total cost MAPE of 1.30%.

The previous validation empirically assessed the cost prediction accuracy of our CA model over different values of decisions (x, y) and parameter settings. However, it remains open whether our CA model can prescribe near-optimal (x, y) solutions. To this end, we construct a third instance set, I_3 , consisting of the 20 test demand scenarios from I , and fix the base-case parameter setting (*i.e.*, $n = 200$, $t_{\text{st}} = 3.5/60$, and $T_{\text{max}} = 5$). We then compare the solutions prescribed by the CA model with those obtained by minimizing the average simulated cost over the demand sample I , solving a vehicle routing problem for each demand scenario.

Figure 6a presents a heatmap of the CA cost function $c(x, y)$ over the decision space $(x, y) \in [0, 10] \times [0, 2]$, while Figure 6b presents the average simulated cost as a function of (x, y) over a discrete grid $x \in \{0, 1, \dots, 10\}$ and $y \in \{0, 0.25, \dots, 2\}$. Comparing Figures 6a and 6b supports the validity of our CA cost model: the CA- and simulation-based cost functions exhibit similar trends across the evaluated decision

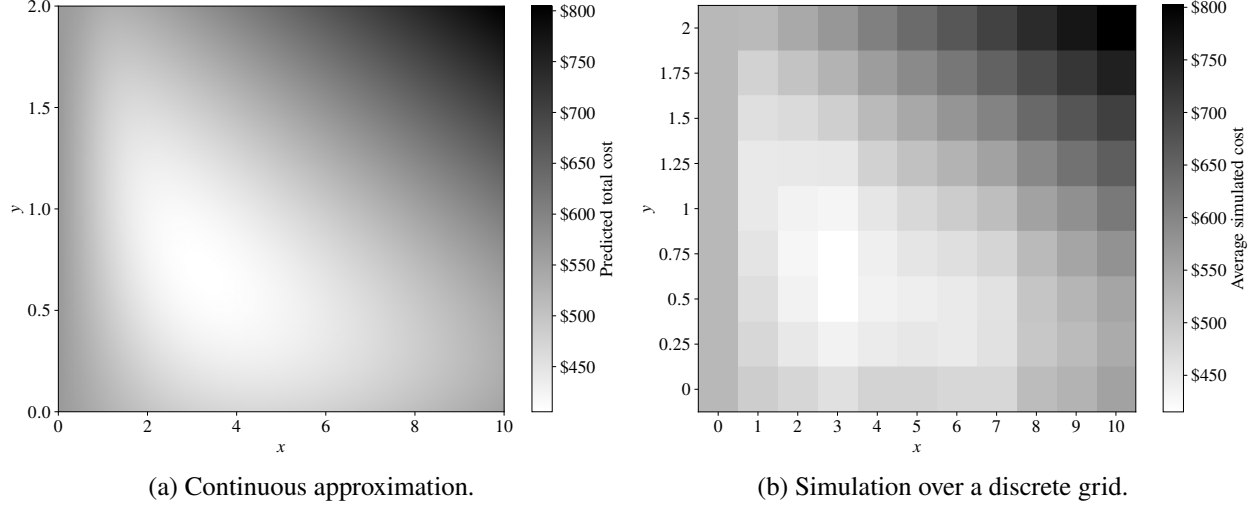


Figure 6: Comparison between the CA cost model and the averaged discrete cost over sampled demand scenarios.

space. The CA-based optimal solution is $(x^* \approx 3.25, y^* \approx \$0.69)$, resulting in an estimated cost of \$405.60. If we constrain the number of CDPs to be integer-valued ($x \in \mathbb{N}_0$), the CA model recommends 3 CDPs and an incentive of \$0.72 per customer, with a total estimated cost of \$406.07. In contrast, the simulation-based cost is minimized over the grid at $(x = 3, y = \$0.75)$ with an objective value \$415.21. If, instead, we set the prescribed CA-based integer recommendation $(x = 3, y = 0.72)$, then we obtain a simulated average cost of \$416.59 (0.33% higher than the simulation-based minimum). In Appendix B.2, we conduct an analogous validation with similar results on a real-world-inspired network.

Given that our CA model’s accuracy has empirically been validated, we now present different benchmark solutions to model 17 and compare these through sensitivity analysis over key problem parameters.

4.3 Benchmarks

In Table 5, we present different benchmark strategies to set values of structural decisions x and y in function $c(x, y)$. The first three are realistic strategies, while the last one is a lower bound on the minimum total expected cost. The first strategy does not install CDPs and its cost C^0 represents that of a traditional last-mile distribution service, which operates by dispatching delivery routes from the distribution center to each customer’s location. The second strategy represents a pure-CDP option, whose cost $\hat{C}$ is computed as the minimum possible cost of $c(x, 0)$ over $x \geq 0$ and rules out offering incentives to customers. The third

Table 5: Benchmark strategies

Benchmark	Cost notation	Var. notation	Problem solved
Cost without CDPs	$C^0 = c(0, 0)$	—	$x = y = 0$
Minimum cost without incentive	$\hat{C} = c(\hat{x}, 0)$	$\hat{x}$	Model (17) s.t. $y = 0$
Minimum cost	$C^* = c(x^*, y^*)$	x^*, y^*	Model (17)
Minimum cost with free CDPs	$C^F = c(x^F, y^F)$	x^F, y^F	Model (17) assuming $\alpha_x = 0$

strategy jointly optimizes both decisions x and y to obtain the minimum feasible cost C^* . The fourth option is an overly optimistic assumption used as a best-case scenario where infrastructure is free of cost for the decision-maker (*i.e.*, $\alpha_x = 0$). As optimization principles dictate, we must have $C^F \leq C^* \leq \hat{C} \leq C^0$.

Each solution of model (17) is obtained by selecting the best result between a grid search with step size 0.1 and the SHGO solver (Endres et al., 2018) available in the SciPy library (SciPy, 2026). Our procedure finishes in fractions of a second, so we do not report running times.

4.4 Base Instance Analysis

We begin studying results over the base instance parameters defined in Table 3. Figure 7 plots a heat map of function $c(x, y)$ with its level curves for $(x, y) \in [0, 6] \times [0, 2]$. Each color intensity corresponds to a specific total cost, with values ranging between \$405.7/day to \$671.3/day.

Also, Table 6 presents the base instance results for each benchmark strategy. It includes total expected cost, percentage savings compared to C^0 , all cost components (*i.e.*, infrastructure ($\alpha_x \cdot x$), distribution ($\alpha_v \cdot t_v(x, y)$) and incentive ($y \cdot n_{\text{CDP}}(x, y)$)

costs), the chosen number of CDPs (x), the chosen customer incentive (y), the resulting number of vehicle service stops ($m(x, y)$), and each CDP’s utilization measured as the expected number of requests picked up per CDP ($n_{\text{CDP}}(x, y)/x$) and the total number of requests per installed CDP ($v(x) = n/x$).

The logistics operator’s total cost without CDPs is $C^0 = \$563.6/\text{day}$ purely based on the distribution

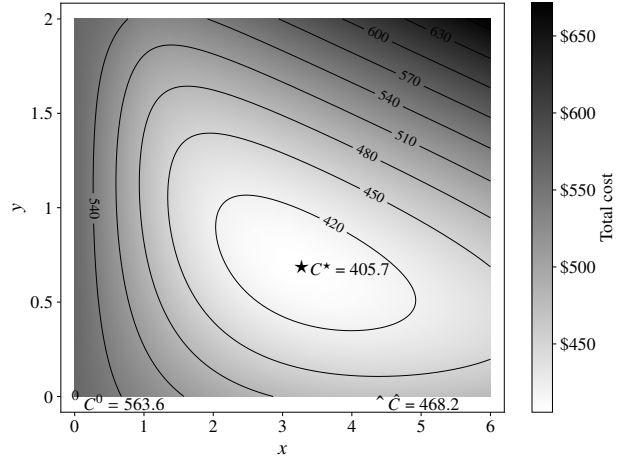
Figure 7: Heat map of $c(x, y)$

Table 6: Benchmark results

Benchmark				
	Cost without CDPs	Minimum cost without incentive	Minimum cost	Minimum cost with free CDPs
Cost	$C^0 = 563.6$	$\hat{C} = 468.2$	$C^* = 405.7$	$C^F = 222.4$
% savings	-	16.9%	28.0%	60.5%
Infrastructure cost	0	132.9	97.4	0
Distribution cost	563.6	335.3	198.9	203.8
Incentive cost	0	0	109.4	18.6
x	0	4.43	3.25	17.82
y	0	0	0.69	0.11
$m(x, y)$	200.0	99.8	43.9	45.8
$n_{\text{CDP}}(x, y)/x$	-	23.6	49.1	9.7
$v(x)$	-	45.1	61.6	11.2

component. If we install a cost-efficient CDP network without incentives, then we cut costs by 16.9% to $\hat{C} = \$468.2/\text{day}$ and require $\hat{x} = 4.43$ CDPs; this roughly yields $v(\hat{x}) = 45.1$ requests per CDP. This efficiency gain is mostly explained by service consolidation at CDPs, which cuts by over half the expected number of vehicle service stops and pushes distribution costs down to $\$335.3/\text{day}$.

As observed in Figure 7, for a fixed x value, we reduce costs if we begin increasing y starting from $\$0/\text{request}$. Thus, we can further reduce costs via customer incentives. If we jointly optimize both x and y and minimize $c(x, y)$, then we reach an expected cost $C^* = \$405.7/\text{day}$ (28.0% savings), cutting down distribution costs even more to $\$198.9$ and also infrastructure costs from $\$132.9$ to $\$97.4$, but adding an extra incentive cost of $\$109.4$ (about 27.0% of C^*). Compared to $\hat{x}$, the optimal number of CDPs reduces by 26.6% to $x^* = 3.25$ (roughly $v(x^*) = 61.6$ requests per CDP), but it requires setting a customer incentive of $y^* = \$0.69$ per request picked up at a CDP. Intuitively, incentivized customers are willing to be served at a farther CDP, allowing the decision-maker to reduce the number of installed CDPs. This increases each CDP utilization from 23.6 to 49.1 requests (*i.e.*, a 108.1% increase).

If the decision-maker assumes that infrastructure is free to use, then the expected cost can go down even more to $C^F = \$222.4$ (60.5% savings). Compared with the minimum cost benchmark, customer incentive decreases to $y^F = \$0.11$ per request (*i.e.*, an 84.1% reduction), as one can alternatively increase the density of the CDP network for free to encourage self-pickups. Moreover, the optimized number of installed CDPs is $x^F = 17.82$. Even if CDP stations are free, one should not install too many of them because an oversized network leads to an increased number of vehicle service stops. This may be particularly relevant for

municipalities or government agencies concerned with improving the competitiveness of local industry or emissions from urban logistics operations, which are often closely linked to distribution costs. However, a more comprehensive analysis should also consider customer pickup transportation mode and traveled miles, as these factors affect the overall carbon footprint associated with CDP-based delivery.

In what follows, we conduct a sensitivity analysis using the base-case parameter values reported in Table 3, systematically varying selected parameters while keeping all others fixed at their base-case values.

4.5 Sensitivity Analysis of Key Parameters

4.5.1 Cost Sensitivity as a Function of Request Density

We now investigate the system's performance as a function of request density $\rho = n/A$. Specifically, we keep all parameters fixed at their base-case values and vary n , thereby changing the request density ρ . Figure 8 plots the expected cost per request ($c(x, y)/n$), the number of requests per CDP ($v(x) = n/x$) and customer incentive (y) as a function of the request density (ρ), for each benchmark strategy defined in Table 5.

We observe economies of density as they occur in the BHH formula; this means that the cost per request over all benchmarks decreases as request density increases. When compared to the benchmark without CDPs (C^0/n), the strategies equipped with CDPs enhance these economies of density and

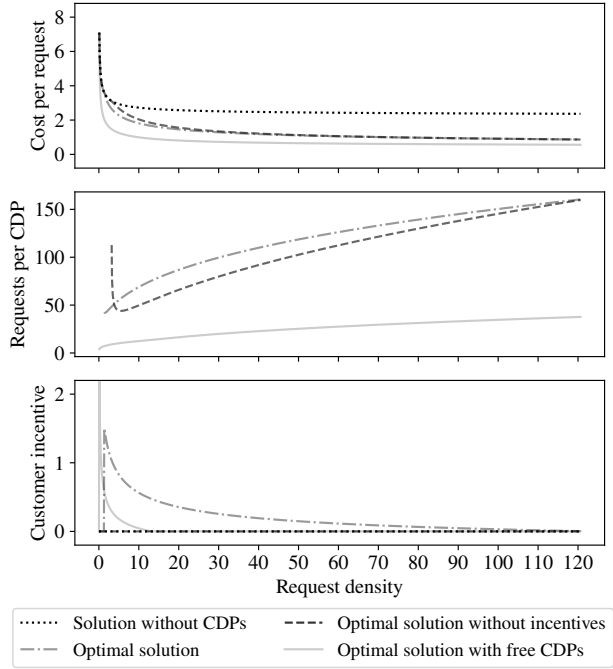


Figure 8: Benchmark strategies as a function of request density ($\rho = n/A$).

produce larger percentage savings as request density grows. Intuitively, a relatively higher request density favors the installation of more CDPs, leading to request consolidation and a smaller number of vehicle service stops. The number of requests per CDP increases as a function of ρ , which reciprocally indicates that fewer CDPs are needed to cover the same number of requests.

We also see that customer incentives are most valuable when request density is relatively low. Compared to the benchmark using CDPs without an incentive ($\hat{C}/n$), the optimal cost per request with incentives (C^*/n) produces the largest percentage cost savings when ρ is roughly 10 requests per square kilometer. Also, the optimal value of incentive decreases as a function of ρ and converges to zero when $\rho \rightarrow \infty$; the largest incentive offered to customers is $y^* \approx \$1.5$ per request picked up at a CDP and occurs when ρ is roughly 1.34 requests per square kilometer. The intuition behind it is that incentives serve as an effective strategy to encourage customers to travel longer distances, and thus, it is especially useful to increase CDP when demand is sparse without increasing infrastructure costs.

Regarding the optimistic case with free CDPs, relatively more CDPs are installed and fewer incentives are offered in this case. Counterintuitively, when the density is low, it is efficient to offer positive incentive values in this case, as they reduce distribution costs more effectively than installing more CDPs, without increasing the number of stops.

4.5.2 Cost Sensitivity as a Function of the Behavioral Parameters

In this section, we analyze how our CA model and its optimal solutions vary across different parameter settings in the behavioral model that describes how customers choose between CDPs and home delivery. We begin with the customer-to-distance sensitivity parameter λ .

Figure 9 plots the optimal number of requests per CDP ($v(x)$) and customer incentive (y) for the minimum cost benchmark as a function of λ (*i.e.*, the customer-to-distance sensitivity) assuming that customers choose their service option based on the binomial logit (BLC) and deterministic (DTC) models. It also presents the resulting expected cost per request for these decisions as dictated by our approximate cost model Equation (12) equipped with the more involved BLC model. We also plot the CDP service choice probabilities $P_{\text{CDP}}^{\text{DTC}}(\bar{\omega}, y)$ and $P_{\text{CDP}}^{\text{BLC}}(\bar{\omega}, y)$ evaluated at $\bar{\omega} = 2 \cdot r(x)/3$, which is the expected distance from any given customer to their nearest CDP.

As expected, if the decision-maker assumes that customers are deterministic entities, then their decisions become suboptimal for the more involved model. The paid cost difference is larger if customers are less sensitive to distance (*i.e.*, smaller λ) and tends to zero as λ grows. Also, decisions that assume deterministic customer choices are independent of λ . In contrast, an efficient decision-maker

adapts their actions as a function of λ and carefully selects the parameters that predict their customers' choices.

The additional cost is produced by overestimating the customer's CDP choice probability. The extreme case occurs when customers are insensitive to distance ($\lambda \rightarrow 0$). In this scenario, the CDP choice probability converges to 50% (equiprobable choice) as customers are indifferent to using CDPs independently of $\bar{\omega}$. In this case, an optimal decision should avoid increasing x or y , as customers do not make choices that strongly depend on these values. Conversely, customers become more sensitive to distance and, therefore, more predictable as λ grows. Specifically, both decisions converge when $\lambda \rightarrow \infty$.

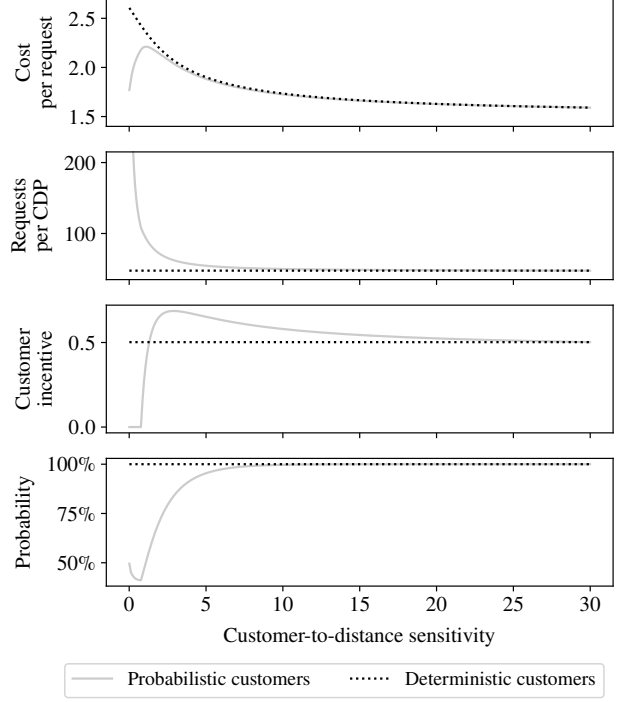


Figure 9: Optimal strategy for the BLC and DTC choice models as a function of customer-to-distance sensitivity (λ).

Next, we analyze how the specification of the customers' indifference distance function $\tau(y)$ impacts our results. Particularly, we retain the linear functional form $\tau(y) = \tau_1 + \tau_2 \cdot y$ and vary τ_1 and τ_2 between 0 and 3. The associated results are depicted in Figure 10. Installing CDPs or offering incentives is not beneficial

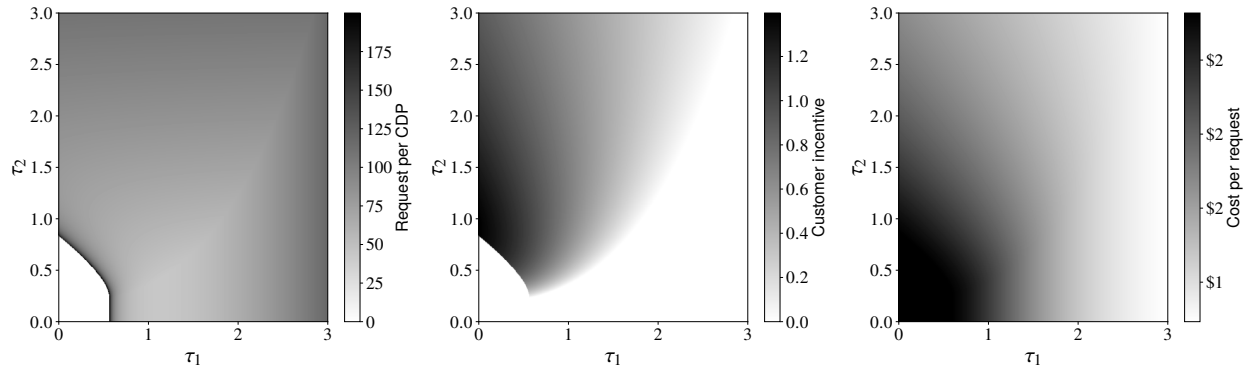


Figure 10: Optimal requests per CDP ($v(x^*)$), customer incentive (y^*), and cost per request (C^*/n) as a function of τ_1 and τ_2 .

when customers are not willing to travel to pick up their parcels (*i.e.*, when $\tau_1 \rightarrow 0$ and $\tau_2 \rightarrow 0$). Also, we observe that for each value of τ_1 , there is a threshold value of τ_2 below which the optimal incentive is $y = 0$. This is because, when τ_1 is sufficiently large, customers are sufficiently willing to pick up their parcels at CDPs that no incentive is needed to induce them to do so.

4.5.3 Cost Sensitivity as a Function of Infrastructure Cost

In this section, we study how the infrastructure cost α_x impacts expected costs and structural decisions. Figure 11 plots the expected cost per request, the chosen numbers of requests per CDP ($v(x)$), and the customer incentive (y) for all the benchmarks presented in Table 5 as functions of the infrastructure cost (α_x).

We observe that the optimal costs per request with and without incentives (C^*/n and $\hat{C}/n$, respectively) are bounded between the cost per request of the optimal solution with free CDPs (*i.e.*, C^F/n) and that of the solution without CDPs (*i.e.*, C^0/n). As expected, they converge to $C^F/n \approx \$1.1/\text{request}$ when $\alpha_x = 0$, and align with $C^0/n \approx \$2.8/\text{request}$ for a high-enough α_x value that completely discourages

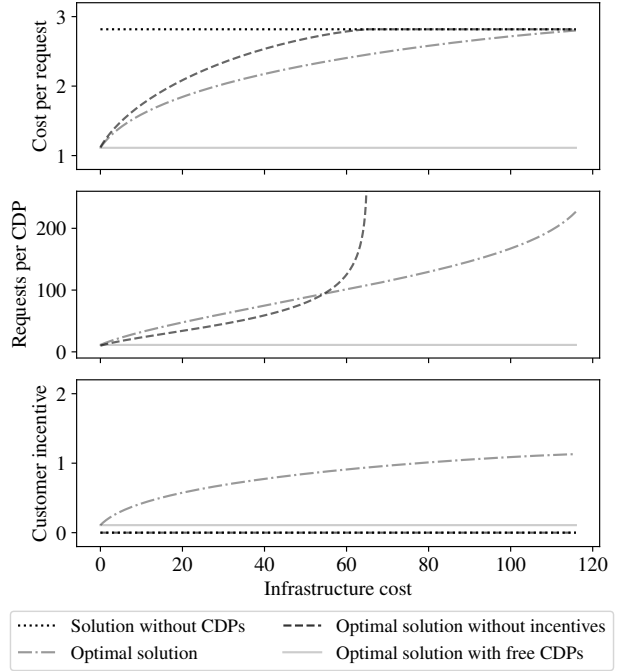


Figure 11: Benchmark strategies as a function of infrastructure cost (α_x).

the use of CDPs. Compared to the case without incentives, the use of incentives increases to almost twice this α_x value, as it allows for the reduction of distribution costs without incurring expensive infrastructure investments.

We also see that C^*/n and $\hat{C}/n$ increase sub-linearly as functions of α_x . This occurs because the optimal solutions adjust increasing y^* , $v(x^*)$, and $v(\hat{x})$ (*i.e.*, it reduces the number of CDPs) to adapt structural decisions to a more CDP-expensive setting.

As indicated by Xu et al. (2021), the infrastructure cost may range between \$16 and \$34/(CDP·day). In

this case, there are substantial cost savings (C^* and $\hat{C}$ compared to C^0) when using CDPs with and without customer incentives.

4.5.4 Cost Sensitivity as a Function of the Stem Travel Time

In our base case with the distribution center located at the center of the service region, we assumed the expected travel time between the distribution center and a customer to be $t_0 \approx 0.13$ hours. Now, we relax this assumption and explore scenarios in which the distribution center is located outside the service region, conducting a sensitivity analysis on the parameter t_0 .

Figure 12 plots the estimated minimum expected cost per request, and the optimal number of requests per CDP ($v(x)$) and customer incentive (y) for the four benchmarks as a function of $t_0 \in [0, 2]$ hours. As expected, we observe that the cost per request increases as a function of t_0 and diverges to infinity as $2 \cdot t_0 \rightarrow T_{\max}$. Nonetheless, the four benchmarks yield different optimal solution patterns with respect to t_0 .

The solution without incentives installs a larger number of CDPs ($\hat{x}$), which decreases the optimal number of requests per CDP ($v(\hat{x})$) as t_0 increases.

When economic incentives are allowed, the outcome changes. The optimal number of requests per CDP ($v(x^*)$) decreases until a given t_0 value (in our case, roughly 1.2 hours), after which it increases. In parallel, optimal incentives y^* rise gradually and then sharply beyond this point. Hence, expanding the CDP network with moderate incentives is effective only up to this t_0 threshold; beyond it, fewer CDPs and stronger incentives become preferable.

In contrast, the optimal solution with free CDPs increases the number of requests per CDP ($v(x^F)$) as t_0

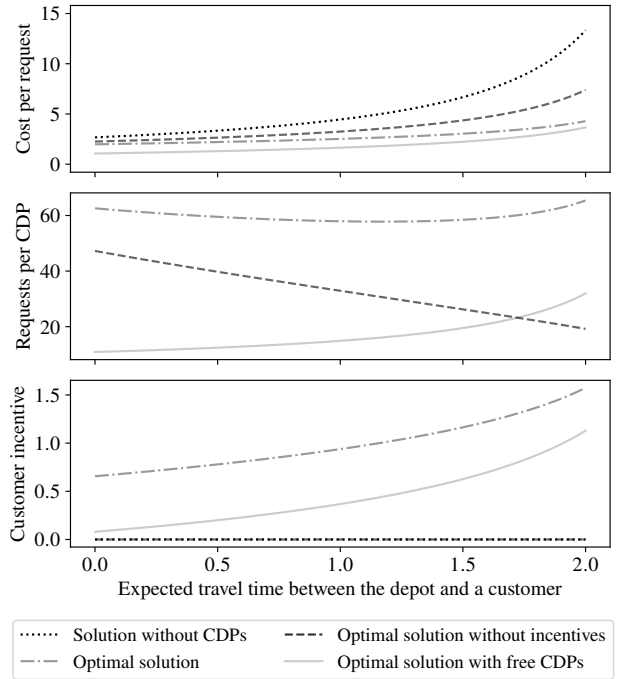


Figure 12: Benchmark strategies as a function of the expected travel time between the distribution center and a customer t_0 .

becomes significantly large. This results from the increase in the incentive y^F to attract more customers to CDPs and compensate for the longer stem travel time.

4.5.5 Cost Sensitivity as a Function of CDP Capacity

So far, our analyses assumed uncapacitated CDPs with a fixed CDP cost α_x . In this section, we drop this assumption and solve the four benchmarks defined in Table 5 subject to chance constraint (19), imposing that the average number of orders handled per CDP exceeds capacity q with probability less than or equal to $\zeta = 5\%$. We also consider a concave increasing cost per CDP $\alpha_x(q) = 6.708\sqrt{q}$ as a function of capacity q , so that a CDP with no capacity has no cost ($\alpha(0) = \$0$), while a CDP with a capacity of 20 has the same cost as in our base-case parameter setting ($\alpha(20) \approx \$30$). This functional form reflects the presence of economies of scale: larger facilities require higher total investment, but the marginal cost of additional capacity decreases with scale. Other increasing functional forms could also be considered within this modeling framework.

Figure 13 plots the estimated minimum expected cost per request, the optimal number of requests per CDP ($v(x)$) and customer incentive (y) for the four benchmarks as a function of q ranging between 0 and 100 requests per CDP. As expected, the minimum cost per request with free CDP (C^F/n) is non-increasing with capacity, since more capacity resources entail no additional cost ($\alpha_x(q) = 0$). Moreover, both the cost and solution (x^F, y^F) converge to a fixed value once the capacity constraint is no longer binding.

When a cost per CDP ($\alpha_x(q) > 0$) that increases with q is considered, the minimum cost per request C^*/n as a function of q follows a U-shape with a clear minimum value. For a relatively small value of q , CDP capacity is a limiting resource that limits decisions and increases cost. As q grows, the optimal

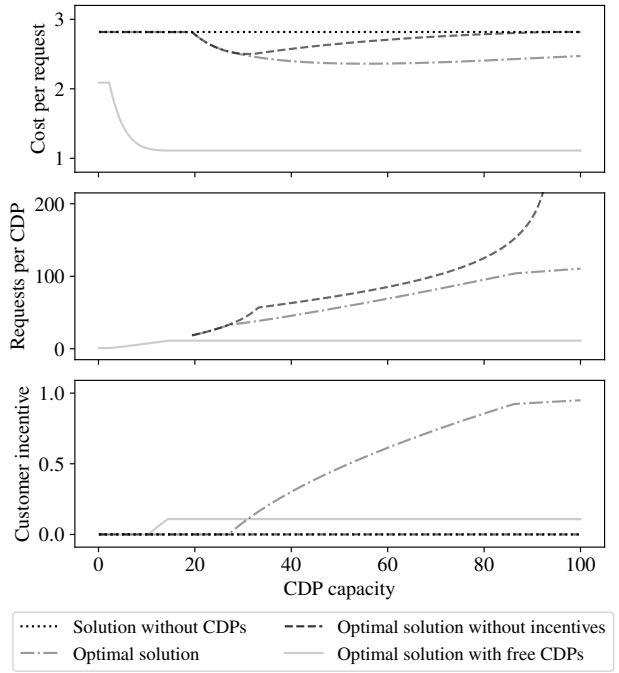


Figure 13: Benchmark strategies as a function of CDP capacity q .

cost per request decreases, since the effect of the capacity constraint diminishes until it becomes redundant. As q continues to increase, excessive CDP capacity ultimately adds unnecessary infrastructure costs. Also, the optimal number of CDPs x^* decreases (and the number of orders per CDP $v(x^*)$ increases) as q grows. This occurs because fewer CDPs with more capacity can serve the same number of customers at a lower cost. However, reducing the number of CDPs increases the average distance to the closest CDP for an average customer, requiring more economic incentives y^* .

5 Validation in a Real Network

In this section, we test our approximate cost model over an instance on a real road network with a realistic geographic request distribution set in the metropolitan area of Santiago, Chile (see Figure 14b). In this experiment, our CA model (17) is assessed in a setting where two of its assumptions are deliberately violated: (i) customers are not uniformly distributed over the service region, and (ii) road network distances are non-Euclidean and possibly asymmetric. Also, we develop and illustrate procedures to translate our CA prescriptions into a detailed, implementable set of CDP locations.

We start aggregating the instance’s data to parameterize our CA model. Our CA model is optimized to determine both structural decisions: the number of CDPs x to install (restricted to take integer values) and the incentive y offered to customers. Then, we heuristically set a discrete set of geographic locations for the x CDPs; for this discretization step, we consider two approaches described later in Section 5.1. For the chosen set of x discrete locations and incentive value y , we simulate a set Ξ of 500 demand scenarios. Each demand scenario contains n geographically sampled requests, drawn from the city’s spatial distribution of requests. For each customer request $k \in \{1, \dots, n\}$ in each demand scenario $\xi \in \Xi$, we simulate its service choice using the model $P_{\text{CDP}}^{\text{BLC}}(\Omega_k, y)$, where Ω_k denotes the shortest-path walking distance to the nearest CDP. Each scenario’s last-mile distribution cost is then computed by solving an *ad hoc* routing problem using the PyVRP heuristic solver on a graph in which arc weights correspond to shortest-path driving times. The expected cost is approximated by the sample average cost across all scenarios, which we compare with our CA model’s cost estimate, computing the corresponding MAPE. Appendix B.2 provides additional details on the validation of our CA model; our results report a total cost MAPE of 2.68%.

Specifically, our experiment considers a representative day of a logistics operator serving $n = 500$

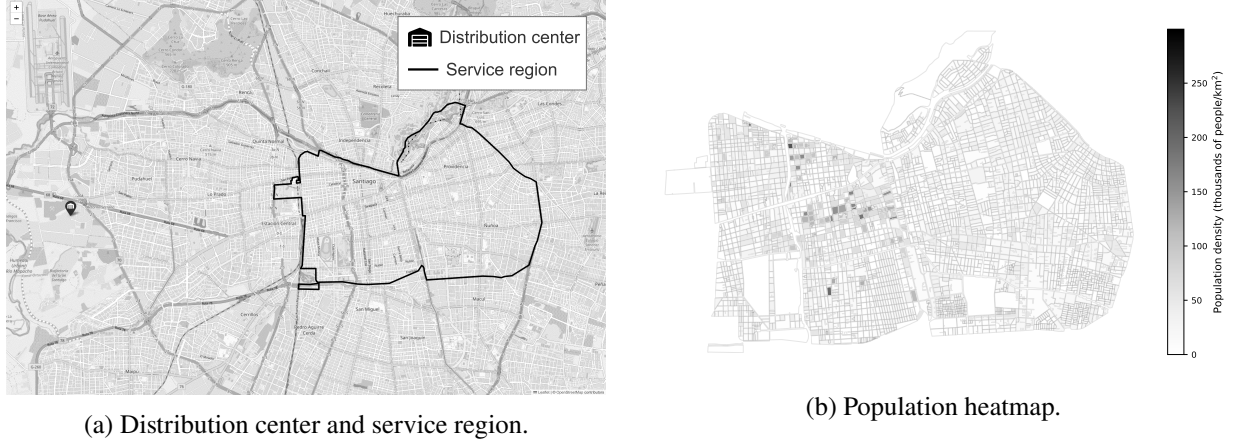


Figure 14: Road network for our real-world-inspired study.

customer requests within a 46.05 km^2 service region formed by three municipalities of the city (Santiago, Providencia and Ñuñoa) from a distribution center located 14.34 km west of the region’s centroid as depicted in Figure 14a; this yields a request density of $\rho \approx 10.86 \text{ requests/km}^2$. We retrieve the road network topology and arc distances from OpenStreetMap (OSM) via the OSMnx Python package (Boeing, 2025), using data from June 1, 2024. To capture the geographic heterogeneity of customer locations, we assign each city block a request probability proportional to the 2017 Chilean population census (INE, 2017), as illustrated in Figure 14b by population density; once requests are drawn at each city block, their specific locations are uniformly drawn within that block.

Our CA model assumes an Euclidean distance metric and thus is likely to underestimate vehicle travel times, as shortest-path network distances are longer than straight-line distances. We fix this distance-related issue by incorporating circuitry factors μ_w and μ_d for the walking and driving road networks, respectively. These factors represent the average ratio between the shortest-path network distance and the corresponding Euclidean distance for each origin-destination pair (see Merchán et al. (2020)). We use these factors to adjust the service area in the CA model as $A_w \approx \mu_w^2 \cdot A'$ and $A_d \approx \mu_d^2 \cdot A'$, where A' is the real-world area used to estimate the CA model’s area. Accordingly, we use A_w and A_d to estimate $n_{\text{CDP}}(x, y)$ and $t_v(x, y)$, respectively. The values of μ_w and μ_d are estimated as the sample averages of the ratio between shortest-path network distance and Euclidean distance, computed over 2,000 randomly sampled origin–destination pairs within the service region. This yields $\mu_d \approx 1.308$ and $\mu_w \approx 1.206$, with sample standard deviations of 0.405 and 0.146, respectively. We use vehicle speeds of $s_{\text{stem}} = 30 \text{ km/h}$ for stem travel and $s_{\text{local}} = 15 \text{ km/h}$ within

the service region. With $R_d = \sqrt{A_d/\pi} \approx 5.01$ km and the distribution center located 14.34 km from the service region, we obtain $t_0 = 14.34/s_{\text{stem}} + 2 \cdot R_d/(3 \cdot s_{\text{local}}) \approx 0.7$ hours. Furthermore, we use the calibrated values $\gamma \approx 1.106$, $\delta \approx 1.043$, and $\eta \approx 0.751$ obtained in Section 4.2 to test the robustness of our travel time model. Finally, we also consider different CDP capacity levels q , and infrastructure costs per CDP are set to $\alpha_x = 6.708\sqrt{q}$, as in Section 4.5.5. All remaining parameters are obtained from base case instance values in Table 3.

5.1 Detailed location of CDPs based on the CA solution

As follows, we present two alternative approaches to heuristically set the specific number of CDP locations prescribed by our CA optimization model. This step is not intended to be prescriptive, but rather a byproduct required to operationalize the CA-based solutions and enable comparison with their simulated performance.

The first approach solves an auxiliary k -means model, where $k = x$ corresponds to the number of CDPs to be installed. Each data point represents an OSM node in the city, with the total demand of each city block distributed evenly among its associated nodes. After solving the clustering problem, each CDP is placed at the OSM node closest to its corresponding cluster centroid. The k -means solution is computed using the k -means++ initialization method available in the scikit-learn library (Pedregosa et al., 2011), with 100 random restarts. This approach is simple, but may create CDP clusters with unbalanced demand, potentially violating each CDP’s capacity constraint.

Our second approach heuristically balances demand starting from the previous k -means solution and running a local search procedure over it described in Appendix A.3. Specifically, given an initial set of CDP locations, we compute the expected total demand absorbed by the CDPs, $n_{\text{CDP}}(x, y)$, by assigning each customer node to its closest CDP and weighting its contribution according to the binomial logit choice probability described previously. The algorithm seeks to maximize an objective function of the form $n_{\text{CDP}}(x, y) \cdot e^{-\text{cv}}$, where cv denotes the coefficient of variation of the demand allocated across CDPs. Hence, the objective favors solutions that capture a large CDP demand (by maximizing $n_{\text{CDP}}(x, y)$) while maintaining balanced CDP utilization (by minimizing cv). Starting from the initial k -means configuration, the algorithm iteratively attempts to relocate each CDP to alternative locations. Each tentative relocation is evaluated using the objective function above, and each local move is accepted whenever it increases the objective value. The

procedure terminates when no further improving relocation can be found, yielding a locally optimal set of CDP locations.

The CA estimate and discrete simulated costs are reported in Table 7; including the CA model's MAPE and a 95% confidence interval for simulated costs. The CA model (17) with the chance constraint (19) for a CDP capacity $q = 40$ requests and an allowable violation probability of $\zeta = 5\%$. The CA model recommends

Table 7: Comparison of detailed location methods

	CA	Simulation	
		<i>k</i> -means	Demand-balancing <i>k</i> -means
Total cost	\$1,393.8	\$1,388.9 $\pm$ 2.3	\$1,426.6 $\pm$ 2.2
MAPE	-	0.35%	2.30%
x^*		13	
y^*		\$0.44	
CDP capacity violation probability	#1	72.2%	3.8%
	#2	56.2%	1.6%
	#3	13.2%	1.4%
	#4	0.0%	0.2%
	#5	0.0%	3.8%
	#6	2.2%	2.0%
	#7	0.0%	1.6%
	#8	0.0%	1.8%
	#9	0.0%	3.6%
	#10	4.2%	2.4%
	#11	13.4%	0.8%
	#12	2.0%	1.4%
	#13	27.0%	1.8%
	Avg.	14.6%	2.0%

installing 13 CDPs and offering an incentive of \$0.44 per customer, resulting in an expected cost of \$1,393.8. The simulated probability that each of the 13 CDPs exceeds its capacity is also reported in Table 7, computed as the fraction of demand scenarios in which the assigned demand exceeds q . The *k*-means based CDP location procedure yields a solution with a sample-average cost of \$1,388.9 $\pm$ 2.3, *i.e.*, slightly below the cost of \$1,393.8 predicted by the CA model. However, this discrete solution is infeasible: five CDPs have capacity-violation probabilities above the allowable 5% threshold, with the worst case exceeding capacity in 72.2% of simulated scenarios. Moreover, the selected CDP locations result in a substantial imbalance in the demand assigned to each CDP, with several CDPs never experiencing a capacity violation.

In contrast, the demand-balancing *k*-means location approach produces a feasible solution. Specifically, no CDP has a violation probability exceeding the allowable 5% and these values are more evenly distributed

across CDPs, ranging between 0.2% and 3.8%. Enforcing this additional balancing results in a sample average cost of $\$1,426.6 \pm 2.2$, which is 2.3% above our CA cost prediction.

The associated discrete networks obtained using the plain and demand-balancing k -means placement heuristics are reported in Figure 15a and Figure 15b, respectively. Given the population heatmap presented

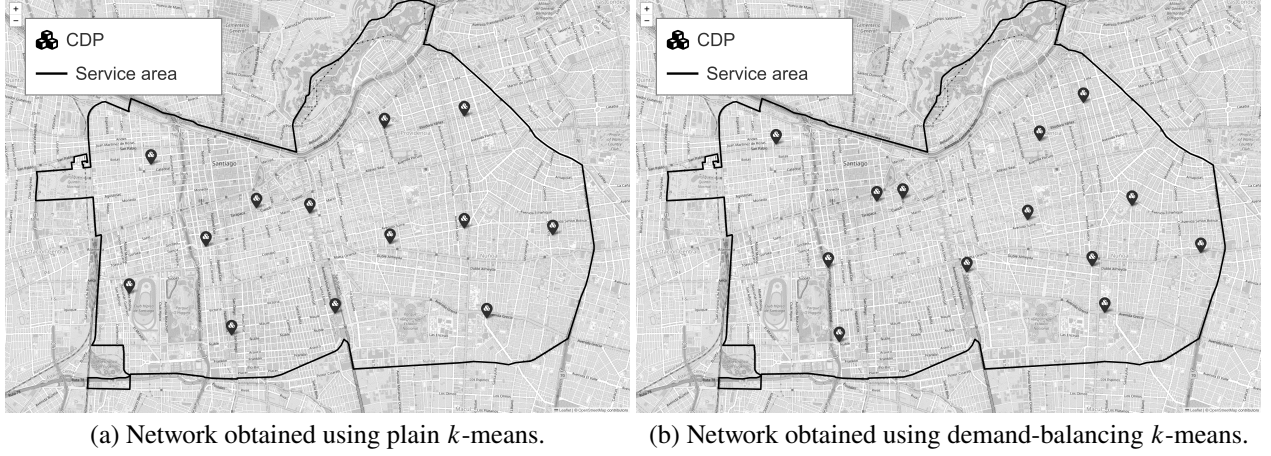


Figure 15: Resulting discrete CDP network designs under the two placement heuristics.

in Figure 8, the demand-balancing k -means procedure locates CDPs in more congested areas and distributes capacity more evenly across the city. Given that the demand-balancing k -means algorithm outperforms the plain k -means approach, we use the former as the baseline and the latter as a benchmark in what follows.

5.2 Impact of Chance Constraints on Cost and Solutions

In this section, we compare the results for the unconstrained CA optimization model with an analogous version using chance constraint (19) with $\zeta = 5\%$. We retain the notation C^* , x^* , and y^* for the unconstrained model solution and cost, and introduce C_P^* , x_P^* , and y_P^* to denote the corresponding solution and cost quantities for the constrained formulation.

In Table 8, we report the optimal solution and cost estimate of our CA model and compare it to the average simulated cost using the demand-balancing k -means CDP location approach and detailed last-mile routing solutions over all demand scenarios. This comparison is made for the constrained and unconstrained versions and for each CDP capacity $q \in \{30, 40, 50, 60\}$. In particular, the simulation results include the average sample cost and its 95% confidence interval, the number of CDPs (out of the x opened CDPs) whose

Table 8: Real-world-inspired study results (using the demand-balancing k -means location procedure)

q α_x		30 36.7	40 42.4	50 47.4	60 52.0
Unconstrained CA model					
CA	C^*	1,221.6	1,261.5	1,295.6	1,322.8
	MAPE	4.89%	0.78%	0.72%	0.67%
	x^*	8	7	6	6
	y^*	0.88	0.95	1.04	1.04
Sim.	Cost	$1,284.4 \pm 2.1$	$1,271.4 \pm 2.1$	$1,305.0 \pm 2.3$	$1,331.7 \pm 2.3$
	# Viol.	8	7	6	6
	Max. Viol.	100.0%	99.6%	99.0%	80.6%
	Avg. Viol.	99.8%	99.2%	98.5%	78.8%
CA model with chance constraint ($\zeta = 5\%$)					
CA	C_P^*	1,460.2	1,393.8	1,369.1	1,365.5
	MAPE	2.13%	2.24%	2.12%	2.57%
	x_P^*	18	13	10	9
	y_P^*	0.29	0.44	0.56	0.82
Sim.	Cost	$1,492.0 \pm 2.5$	$1,425.7 \pm 2.3$	$1,398.7 \pm 2.3$	$1,401.5 \pm 2.4$
	# Viol.	0	0	0	0
	Max. Viol.	3.8%	4.6%	1.4%	2.0%
	Avg. Viol.	2.3%	2.0%	1.0%	1.1%

simulated capacity constraint violation probability exceeds 5%, and the maximum and average violation percentage over all demand scenarios across all CDPs.

Our first observation is that, in all cases, the cost estimated by our CA model lies close to the 95% confidence interval of the simulated detailed costs, deviating by at most 4.89% from the simulated means. This shows that, in our real-world-inspired study, the CA model provides a good estimate of last-mile distribution costs, even when demand is spatially distributed according to the city's population and travel distances are non-Euclidean due to the underlying network structure.

As expected, we have $C^* \leq C_P^*$ for all values of q . Similarly, the optimal number of CDPs follows the same ordering, $x^* \leq x_P^*$, which is consistent with the additional capacity restrictions imposed by the chance constraint. In contrast, the optimal economic incentive moves in the opposite direction, *i.e.*, $y_P^* \leq y^*$, because relaxing capacity constraints allows the decision maker to rely on fewer and more congested CDPs, which increases customer travel distances and therefore requires higher incentives to make these CDPs attractive for the customers.

In the unconstrained case, the average capacity constraint violation ranges from 78.8% to 99.8%, with

between 6 and 8 CDPs experiencing demand overflows. In contrast, no CDPs violate the approximate chance constraint. Specifically, the average probability of capacity violation ranges between 1.0% and 2.3%, remaining below the imposed maximum allowable threshold of 5%.

For completeness, we also conduct an analogous experiment to the above, but using the simplest k -means-based location procedure; the results are reported in Appendix B.3.

6 Conclusions

In this paper, we present and validate a continuous approximation model depending on aggregate instance data that estimates the expected operating cost of a logistics last-mile distribution system delivering to both customers' locations and CDPs, considering customers' service choices. Our model examines the trade-offs between investing in infrastructure, covering last-mile distribution costs, and offering incentives to customers collecting their goods farther from their homes. Based on this cost function, we develop a stylized optimization model to minimize the expected cost by determining the system's structural decisions: (i) the number of CDPs to install and (ii) the incentive offered to customers.

From a managerial perspective, the proposed framework provides a fast strategic planning tool to assess the economic viability of CDP networks before engaging in detailed location-allocation analyses. The model can support strategic decisions regarding the appropriate network scale, the design of customer incentive policies, and the assessment of trade-offs between operational costs and service adoption.

Empirically, we assess the potential cost savings of last-mile distribution systems equipped with CDPs when compared to ones solely executing home deliveries. In a base experiment, the sole use of CDPs without customer incentives reaches cost savings of 16.9%. These percentage cost savings grow up to 28.0% when customer incentives are added. Our study in Santiago, Chile, further shows that the proposed CA model can effectively support CDP network sizing decisions in realistic urban settings. Even when some of its assumptions, such as uniformly distributed customers, are violated in practice, the model remains accurate in estimating costs and structural decisions.

Ours is an introductory study on the strategic design of CDP networks that anticipates customer service choices. We devise many future research directions to extend our model. Although we lack customer-specific information, our model may be extended to heterogeneous customers by allowing request density and CDP

choice probabilities to vary across locations. These spatial distributions could be estimated from historical demand data. Such an extension could then prescribe location-specific incentives or CDP densities

Also, customers could choose among multiple available CDPs, rather than only the closest one. In that sense, an interesting extension would be to replace the circular influence regions with more detailed Voronoi-based service regions, potentially allowing for heterogeneous or probabilistic customer assignment mechanisms as in discrete choice or spatial interaction models.

Another interesting direction for future research would be to consider a model extension in which incentives are second-stage decisions, adapted to a known customer demand distribution. We could also consider the dynamic version of this problem, incorporating time-varying customer incentives and the dynamic activation of CDPs.

Future research could also further study the detailed discrete optimization counterpart of our problem. In this regard, it is particularly interesting to explore how to add detailed CDP installation decisions over a continuous service region and how we could leverage our CA model for it. We envision the CA model as a strategic planning device capable of rapidly identifying promising network scales and incentive policies, whose outputs can subsequently be refined through more detailed discrete optimization procedures. Such an approach could help reduce computational effort, narrow the set of candidate designs, and provide high-quality initial solutions for strategic network design models. In this sense, tighter links between continuous approximation and discrete optimization could substantially enhance the effectiveness of strategic network design approaches.

Acknowledgements

Benjamín Rojas' work is supported by the Chilean National Agency for Research and Development (ANID) under PhD grant 21230934. Mathias Klapp and Homero Larrain thank ANID for its support via Fondecyt Regular Project Number 1251812.

References

- Afsar, H. M. (2022). Traveling salesperson problem with unique pricing and stochastic thresholds. *Computers & Industrial Engineering*, 173:108696.
- Afsar, H. M., Afsar, S., and Palacios, J. J. (2021). Vehicle routing problem with zone-based pricing. *Transportation Research Part E: Logistics and Transportation Review*, 152:102383.
- Akkerman, F., Dieter, P., and Mes, M. (2025). Learning dynamic selection and pricing of out-of-home deliveries. *Transportation Science*, 59(2):250–278.
- Allen, J., Piecyk, M., Piotrowska, M., McLeod, F., Cherrett, T., Ghali, K., Nguyen, T., Bektas, T., Bates, O., Friday, A., Wise, S., and Austwick, M. (2018). Understanding the impact of e-commerce on last-mile light goods vehicle activity in urban areas: The case of london. *Transportation Research Part D: Transport and Environment*, 61:325–338.
- Ansari, S., Başdere, M., Li, X., Ouyang, Y., and Smilowitz, K. (2018). Advancements in continuous approximation models for logistics and transportation systems: 1996–2016. *Transportation Research Part B: Methodological*, 107:229–252.
- Arslan, O. (2021). The location-or-routing problem. *Transportation Research Part B: Methodological*, 147:1–21.
- Banerjee, D. (2026). Dynamic delivery request acceptance with strict geographic fairness: a classical yield management approach. *Transportation Research Part E: Logistics and Transportation Review*, 205:104487.
- Banerjee, D., Erera, A. L., Stroh, A. M., and Toriello, A. (2023). Who has access to e-commerce and when? Time-varying service regions in same-day delivery. *Transportation Research Part B: Methodological*, 170:148–168.
- Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327.

- Boeing, G. (2025). Modeling and analyzing urban networks and amenities with osmnx. *Geographical Analysis*, 57(4):567–577.
- Boysen, N., Fedtke, S., and Schwerdfeger, S. (2021). Last-mile delivery concepts: a survey from an operational research perspective. *OR Spectrum*, 43(1):1–58.
- Carlsson, J. G. and Jia, F. (2015). Continuous facility location with backbone network costs. *Transportation Science*, 49(3):433–451.
- Carlsson, J. G., Liu, S., Salari, N., and Yu, H. (2024). Provably good region partitioning for on-time last-mile delivery. *Operations Research*, 72(1):91–109.
- Caspersen, E. and Navrud, S. (2021). The sharing economy and consumer preferences for environmentally sustainable last mile deliveries. *Transportation Research Part D: Transport and Environment*, 95:102863.
- Cerulli, M., Archetti, C., Fernández, E., and Ljubić, I. (2024). A bilevel approach for compensation and routing decisions in last-mile delivery. *Transportation Science*, 58(5):1076–1100.
- Coppola, D. (2024). Share of last-mile delivery costs out of total shipping costs in 2018 and 2023. Viewed 20 of February 2026, <https://www.statista.com/statistics/1434298/last-mile-share-of-total-shipping-costs/>.
- Daganzo, C. F. (1984). Distance traveled to visit n points with a maximum of c stops per vehicle: An analytical model and an application. *Transportation Science*, 18(4):331–350.
- Dasci, A. and Laporte, G. (2005). A continuous model for multistore competitive location. *Operations Research*, 53(2):263–280.
- Dell’Amico, M., Montemanni, R., and Novellani, S. (2023). Pickup and delivery with lockers. *Transportation Research Part C: Emerging Technologies*, 148:104022.
- Deutsch, Y. and Golany, B. (2018). A parcel locker network as a solution to the logistics last mile problem. *International Journal of Production Research*, 56(1-2):251–261.
- dos Santos, A. G., Viana, A., and Pedroso, J. P. (2022). 2-echelon lastmile delivery with lockers and occasional couriers. *Transportation Research Part E: Logistics and Transportation Review*, 162:102714.

- Dumez, D., Lehuédé, F., and Péton, O. (2021). A large neighborhood search approach to the vehicle routing problem with delivery options. *Transportation Research Part B: Methodological*, 144:103–132.
- Endres, S. C., Sandrock, C., and Focke, W. W. (2018). A simplicial homology algorithm for lipschitz optimisation. *Journal of Global Optimization*, 72(2):181–217.
- Enthoven, D. L., Jargalsaikhan, B., Roodbergen, K. J., uit het Broek, M. A., and Schrotenboer, A. H. (2020). The two-echelon vehicle routing problem with covering options: City logistics with cargo bikes and parcel lockers. *Computers and Operations Research*, 118:104919.
- Franceschetti, A., Jabali, O., and Laporte, G. (2017). Continuous approximation models in freight distribution management. *TOP*, 25(3):413–433.
- Galiullina, A., Mutlu, N., Kinable, J., and Woensel, T. V. (2024). Demand steering in a last-mile delivery problem with home and pickup point delivery options. *Transportation Science*, 58(2):454–473.
- Gallego, G. and Topaloglu, H. (2019). *Revenue Management and Pricing Analytics*, volume 279 of *International Series in Operations Research & Management Science*. Springer.
- Hexaly (2025). Hexaly on very large-scale instances of the Capacitated Vehicle Routing Problem (CVRP). Viewed 23 of April, 2026, <https://www.hexaly.com/benchmarks/large-scale-instances-capacitated-vehicle-routing-problem-cvrp>.
- Horner, H., Pazour, J., and Mitchell, J. E. (2024). Increasing driver flexibility through personalized menus and incentives in ridesharing and crowdsourced delivery platforms. *Naval Research Logistics*, 72(1):3–23.
- INE (2017). Open geodata. National Institute of Statistics. Viewed 20 of February 2026, <https://www.ine.gob.cl/herramientas/portal-de-mapas/geodatos-abiertos>.
- Janinhoff, L. and Klein, R. (2023). Stochastic location routing for out-of-home delivery networks. Preprint, URL https://papers.ssrn.com/abstract_id=4654115.
- Janinhoff, L., Klein, R., Sailer, D., and Schoppa, J. M. (2024a). Out-of-home delivery in last-mile logistics: A review. *Computers & Operations Research*, 168:106686.

- Janinhoff, L., Klein, R., and Scholz, D. (2024b). Multitrip vehicle routing with delivery options: a data-driven application to the parcel industry. *OR Spectrum*, 46(2):241–294.
- Janjevic, M., Winkenbach, M., and Merchán, D. (2019). Integrating collection-and-delivery points in the strategic design of urban last-mile e-commerce distribution networks. *Transportation Research Part E: Logistics and Transportation Review*, 131:37–67.
- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley, Hoboken, NJ, 3 edition.
- Jones, B., Xu, K., and Carlsson, J. G. (2026). Coordinated logistics with a truck and multiple sidekicks. *Transportation Science*, 60(4):527–544.
- Kahr, M. (2022). Determining locations and layouts for parcel lockers to support supply chain viability at the last mile. *Omega*, 113:102721.
- Klein, R., Neugebauer, M., Ratkovitch, D., and Steinhardt, C. (2019). Differentiated time slot pricing under routing considerations in attended home delivery. *Transportation Science*, 53(1):236–255.
- Kötschau, R., Scherr, N., Tilk, C., and Ehmke, J. F. (2025). Mobile home delivery parcel lockers. *Transportation Research Part E: Logistics and Transportation Review*, 193:103867.
- Kötschau, R., Soeffker, N., and Ehmke, J. F. (2023). Mobile parcel lockers with individual customer service. *Networks*, 82(4):506–526.
- Laporte, G., Nickel, S., and Saldanha da Gama, F., editors (2019). *Location Science*. Springer, 2 edition.
- Law No. 20,271, Art. 25 bis. (2008). Amends the chilean labor code, establishing the obligation to inform the working hours in employment contracts. Official Journal of the Republic of Chile.
- Lemke, J., Iwan, S., and Korczak, J. (2016). Usability of the parcel lockers from the customer perspective – the research in polish cities. *Transportation Research Procedia*, 16:272–287.
- Li, J., Ensafian, H., Bell, M. G., and Geers, D. G. (2021). Deploying autonomous mobile lockers in a two-echelon parcel operation. *Transportation Research Part C: Emerging Technologies*, 128:103155.

- Li, X. and Ouyang, Y. (2010). A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Research Part B: Methodological*, 44(4):535–548.
- Lin, Y., Wang, Y., Lee, L. H., and Chew, E. P. (2022). Profit-maximizing parcel locker location problem under threshold luce model. *Transportation Research Part E: Logistics and Transportation Review*, 157:102541.
- Lin, Y. H., Wang, Y., He, D., and Lee, L. H. (2020). Last-mile delivery: Optimal locker location under multinomial logit choice model. *Transportation Research Part E: Logistics and Transportation Review*, 142:102059.
- Liu, Y., Ye, Q., Escribano-Macias, J., Feng, Y., Candela, E., and Angeloudis, P. (2023). Route planning for last-mile deliveries using mobile parcel lockers: A hybrid q-learning network approach. *Transportation Research Part E: Logistics and Transportation Review*, 177:103234.
- Lyu, G. and Teo, C.-P. (2022). Last mile innovation: The case of the locker alliance network. *Manufacturing & Service Operations Management*, 24(5):2425–2443.
- Ma, B., Wong, Y. D., and Teo, C.-C. (2022). Parcel self-collection for urban last-mile deliveries: A review and research agenda with a dual operations-consumer perspective. *Transportation Research Interdisciplinary Perspectives*, 16:100719.
- Mancini, S. and Gansterer, M. (2021). Vehicle routing with private and shared delivery locations. *Computers and Operations Research*, 133:105361.
- Mancini, S., Gansterer, M., and Triki, C. (2023). Locker box location planning under uncertainty in demand and capacity availability. *Omega*, 120:102910.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.
- Merchán, D., Winkenbach, M., and Snoeck, A. (2020). Quantifying the impact of urban road networks on the efficiency of local trips. *Transportation Research Part A: Policy and Practice*, 135:38–62.
- Molin, E., Kosicki, M., and Duin, R. V. (2022). Consumer preferences for parcel delivery methods. *European Journal of Transport and Infrastructure Research*, 22(2):183–200.

- Nagy, G. and Salhi, S. (2006). Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2):649–672.
- Newell, G. F. (1973). Scheduling, location, transportation, and continuum mechanics; some simple approximations to optimization problems. *SIAM Journal on Applied Mathematics*, 25(3):346–360.
- Olsson, J., Hellström, D., and Vakulenko, Y. (2023). Customer experience dimensions in last-mile delivery: an empirical study on unattended home delivery. *International Journal of Physical Distribution & Logistics Management*, 53(2):184–205.
- Ouyang, Y. and Daganzo, C. F. (2006). Discretization and validation of the continuum approximation scheme for terminal system design. *Transportation Science*, 40(1):89–98.
- Ozyavas, P., Buijs, P., Ursavas, E., and Teunter, R. (2025). Designing a sustainable delivery network with parcel locker systems as collection and transfer points. *Omega*, 131:103199.
- Papadimitriou, C. H. (1981). Worst-case and probabilistic analysis of a geometric location problem. *SIAM Journal on Computing*, 10(3):542–557.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pulido, R., Muñoz, J. C., and Gazmuri, P. (2015). A continuous approximation model for locating warehouses and designing physical and timely distribution strategies for home delivery. *EURO Journal on Transportation and Logistics*, 4(4):399–419.
- Rai, H. B., Broekaert, C., Verlinde, S., and Macharis, C. (2021). Sharing is caring: How non-financial incentives drive sustainable e-commerce delivery. *Transportation Research Part D: Transport and Environment*, 93:102794.
- Raviv, T. (2023). The service points’ location and capacity problem. *Transportation Research Part E: Logistics and Transportation Review*, 176:103216.

- Savelsbergh, M. and van Woensel, T. (2016). City logistics: Challenges and opportunities. *Transportation Science*, 50(2):579–590.
- Schwerdfeger, S. and Boysen, N. (2020). Optimizing the changing locations of mobile parcel lockers in last-mile distribution. *European Journal of Operational Research*, 285(3):1077–1094.
- SciPy (2026). Optimization and root finding (scipy.optimize). Viewed 23 of April, 2026, <https://docs.scipy.org/doc/scipy/reference/optimize.html>.
- Stroh, A. M., Erera, A. L., and Toriello, A. (2022). Tactical design of same-day delivery systems. *Management Science*, 68(5):3444–3463.
- Tilk, C., Olkis, K., and Irnich, S. (2021). The last-mile vehicle routing problem with delivery options. *OR Spectrum*, 43(4):877–904.
- Toth, P. and Vigo, D., editors (2014). *Vehicle Routing*. Society for Industrial and Applied Mathematics.
- Vakulenko, Y., Hellström, D., and Hjort, K. (2018). What’s in the parcel locker? exploring customer value in e-commerce last mile delivery. *Journal of Business Research*, 88:421–427.
- Wang, Q., Lyu, G., He, L., and Teo, C. P. (2026). Does the locker alliance network improve last mile delivery efficiency? *Management Science*, 72(5):4124–4143.
- Wang, X., Arslan, O., and Delage, E. (2024). Crowdkeeping in last-mile delivery. *Transportation Science*, 58(2):474–498.
- Wang, Y., Bi, M., and Chen, Y. (2020). A scheduling strategy of mobile parcel lockers for the last mile delivery problem. *Promet - Traffic&Transportation*, 32(6):875–885.
- Wouda, N. A., Lan, L., and Kool, W. (2024). PyVRP: a high-performance VRP solver package. *INFORMS Journal on Computing*, 36(4):943–955.
- Xu, X., Shen, Y., Chen, W. A., Gong, Y., and Wang, H. (2021). Data-driven decision and analytics of collection and delivery point location problems for online retailers. *Omega*, 100:102280.

- Yildiz, B. and Savelsbergh, M. (2020). Pricing for delivery time flexibility. *Transportation Research Part B: Methodological*, 133:230–256.
- Yuen, K. F., Wang, X., Ng, L. T. W., and Wong, Y. D. (2018). An investigation of customers' intention to use self-collection services for last-mile delivery. *Transport Policy*, 66:1–8.
- Zadeh, A. S., Dayarian, I., and Bashiri, M. (2026). Strategic design of parcel locker networks for urban delivery. *Transportation Research Part E: Logistics and Transportation Review*, 206:104541.
- Zagier, D. (2007). The dilogarithm function. In Cartier, P., Moussa, P., Julia, B., and Vanhove, P., editors, *Frontiers in Number Theory, Physics, and Geometry II*. Springer, Berlin, Heidelberg.
- Zang, X., Jiang, L., Liang, C., and Fang, X. (2023). Coordinated home and locker deliveries: An exact approach for the urban delivery problem with conflicting time windows. *Transportation Research Part E: Logistics and Transportation Review*, 177:103228.
- Zhang, J., Yang, L., and Campbell, J. F. (2026). Economic and environmental impacts of integrated drone delivery systems. *Transportation Research Part E: Logistics and Transportation Review*, 209:104759.
- Zhang, W., Xu, M., and Wang, S. (2023). Joint location and pricing optimization of self-service in urban logistics considering customers' choice behavior. *Transportation Research Part E: Logistics and Transportation Review*, 174:103128.
- Zhou, L., Li, C., Hu, C., and Du, J. (2024). Parcel locker location problem with selectable volume sizes and heterogeneous customers in the last mile delivery. *Transportation Letters*, 16(9):1140–1154.
- Çınar, A. B., Dullaert, W., Leitner, M., Paradiso, R., and Waldherr, S. (2024). The role of individual compensation and acceptance decisions in crowdsourced delivery. *Transportation Research Part C: Emerging Technologies*, 169:104834.

Appendix A Additional Technical Details

Appendix A.1 Derivation of Equation (14)

We focus on the case $x > 0$, as with $x = 0$ the correspondence is trivial. Starting from the approximation of $n_{\text{CDP}}(x, y)$ provided in (4) and plugging the deterministic choice model defined in (13), we have that

$$n_{\text{CDP}}(x, y) \approx x \cdot \frac{n}{A} \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\text{CDP}}^{\text{DTC}}(\omega, y) d\omega d\theta \quad (20a)$$

$$= x \cdot \frac{n}{A} \int_0^{2\pi} \int_0^{r(x)} \omega \cdot \mathbf{1}\{\omega \leq \tau(y)\} d\omega d\theta \quad (20b)$$

$$= x \cdot \frac{n}{A} \int_0^{2\pi} \int_0^{\min\{r(x), \tau(y)\}} \omega d\omega d\theta \quad (20c)$$

$$= x \cdot \frac{n}{A} \int_0^{2\pi} \left[\frac{1}{2} \omega^2 \right]_0^{\min\{r(x), \tau(y)\}} d\theta \quad (20d)$$

$$= x \cdot \frac{n}{A} \int_0^{2\pi} \frac{1}{2} \min\{r(x), \tau(y)\}^2 d\theta \quad (20e)$$

$$= x \cdot \frac{n}{A} \cdot 2\pi \cdot \frac{1}{2} \min\{r(x), \tau(y)\}^2 \quad (20f)$$

$$= x \cdot \frac{n}{A} \cdot \pi \cdot \min\{r(x), \tau(y)\}^2. \quad (20g)$$

Using $A = x \cdot \pi \cdot r(x)^2$, we obtain the desired expression

$$n_{\text{CDP}}(x, y) = n \cdot \min \left\{ 1, \left(\frac{\tau(y)}{r(x)} \right)^2 \right\}. \quad (21)$$

Appendix A.2 Derivation of Equation (16)

We need to compute a closed-form formula for

$$x \cdot \frac{n}{A} \cdot \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\text{CDP}}^{\text{BLC}}(\omega, y) d\omega d\theta = 2 \cdot \pi \cdot x \cdot \frac{n}{A} \cdot \int_0^{r(x)} \omega \cdot P_{\text{CDP}}^{\text{BLC}}(\omega, y) d\omega, \quad (22)$$

with $P_{\text{CDP}}^{\text{BLC}}(\omega, y) = 1 / (1 + e^{\lambda \cdot (\omega - \tau(y))})$. Let us concentrate on the integral

$$\int_0^{r(x)} \omega \cdot P_{\text{CDP}}^{\text{BLC}}(\omega, y) d\omega = \int_0^{r(x)} \omega \cdot \frac{1}{1 + e^{\lambda \cdot (\omega - \tau(y))}} d\omega. \quad (23)$$

We first solve the indefinite counterpart of the integral. If we multiply this integral by $e^{-\lambda \cdot (\omega - \tau(y))} / e^{-\lambda \cdot (\omega - \tau(y))}$ and reorder it we obtain

$$\int \omega \cdot \frac{1}{1 + e^{\lambda \cdot (\omega - \tau(y))}} d\omega = \int \omega \cdot \frac{e^{-\lambda \cdot (\omega - \tau(y))}}{e^{-\lambda \cdot (\omega - \tau(y))} + 1} d\omega = e^{\lambda \tau(y)} \int \omega \cdot \frac{e^{-\lambda \omega}}{e^{-\lambda \cdot (\omega - \tau(y))} + 1} d\omega. \quad (24)$$

Now, we must solve

$$\int \omega \cdot \frac{e^{-\lambda \omega}}{e^{-\lambda \cdot (\omega - \tau(y))} + 1} d\omega, \quad (25)$$

which integrated by parts is equal to

$$\begin{aligned} \int \omega \cdot \frac{e^{-\lambda \omega}}{e^{-\lambda \cdot (\omega - \tau(y))} + 1} d\omega &= -\omega \frac{e^{-\lambda \tau(y)} \ln(e^{-\lambda \cdot (\omega - \tau(y))} + 1)}{\lambda} \\ &\quad + \frac{e^{-\lambda \tau(y)}}{\lambda} \int \ln(e^{-\lambda \cdot (\omega - \tau(y))} + 1) d\omega. \end{aligned} \quad (26)$$

Finally, we use the substitution $z = -e^{-\lambda \cdot (\omega - \tau(y))} \Rightarrow dz = \lambda e^{-\lambda \cdot (\omega - \tau(y))} d\omega$ to solve

$$\int -\frac{e^{-\lambda \tau(y)} \ln(e^{-\lambda \cdot (\omega - \tau(y))} + 1)}{\lambda} d\omega = -\frac{e^{-\lambda \tau(y)}}{\lambda^2} \int -\frac{\ln(1 - z)}{z} dz = -\frac{e^{-\lambda \tau(y)}}{\lambda^2} \text{Li}_2(z). \quad (27)$$

If we go back, we obtain

$$n_{\text{CDP}}(x, y) \approx \begin{cases} n \cdot \left(1 + \frac{2 \cdot \text{Li}_2(-e^{-\lambda \tau(y)})}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \text{Li}_2(-e^{\lambda \cdot (r(x) - \tau(y))})}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \ln(1 + e^{\lambda \cdot (r(x) - \tau(y))})}{\lambda \cdot r(x)} \right), & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

Appendix A.3 Demand-Balancing k -means Heuristic

Algorithm 1 Heuristic CDP Location Procedure

Require: Set of nodes $\mathcal{N}$, demand density d , number of CDPs x

Ensure: Locations of the selected CDPs L

```
1:  $L \leftarrow \text{WEIGHTEDKMEANS}(\mathcal{N}, d, x)$ 
2: Compute  $n_{\text{CDP}}(L)$  and  $\text{cv}(L)$ 
3:  $f(L) \leftarrow n_{\text{CDP}}(L) \cdot e^{-\text{cv}(L)}$ 
4: repeat
5:    $\text{improved} \leftarrow \text{False}$ 
6:   for each active CDP  $i \in L$  do
7:     for each candidate  $j \in \mathcal{N} \setminus L$  do
8:        $L' \leftarrow (L \setminus \{i\}) \cup \{j\}$ 
9:       Compute  $n_{\text{CDP}}(L')$  and  $\text{cv}(L')$ 
10:       $f(L') \leftarrow n_{\text{CDP}}(L') \cdot e^{-\text{cv}(L')}$ 
11:      if  $f(L') > f(L)$  then
12:         $L \leftarrow L'$ 
13:         $f(L) \leftarrow f(L')$ 
14:         $\text{improved} \leftarrow \text{True}$ 
15:      end if
16:    end for
17:  end for
18: until  $\text{improved} = \text{False}$ 
19: return  $L$ 
```

Appendix B Additional Computational Experiments

Appendix B.1 Model Validation with Near-Optimal Routing

In this section, we consider the instance set I and repeat the validation process described in Section 4.2, this time computing distribution costs using Hexaly as the routing solver. We use the same computing environment as in the previous validation and impose a one-minute time limit per instance; prior experiments show that this time limit yields solutions within less than 6% of the best-known solutions for instances with up to 30,000 nodes reported in the literature (Hexaly, 2025). However, rather than considering 100 scenarios per instance, we restrict ourselves to a single sampled scenario.

In our results, the vehicle routing optimality gap averages 2.4%, with a minimum of 0% and a maximum of 9.4%. Figure 16 shows the total cost function, both for the continuous approximations and the simulated discrete solutions. We again observe that the continuous approximation function is informative at the strategic level.

Figure 17 also depicts the other main components of our approximation model. The approximations accurately capture the observed trends in the simulation results, providing further support for the quality of the approximation.

The average MAPE between the simulation and the continuous approximation is reported in Table 9. As expected, our approximations slightly deteriorate when compared to results obtained from near-optimal

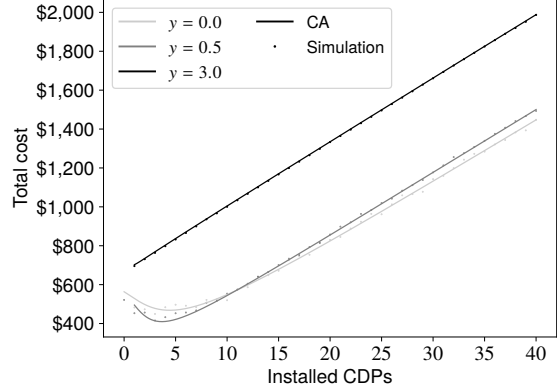


Figure 16: Approximate and simulated average total cost as function of x and y , simulations solved with Hexaly.

Table 9: MAPE of the CA model, simulations solved with Hexaly.

	$c(x, y)$	$n_{\text{CDP}}(x, y)$	$\dot{x}(x, y)$	$t_v(x, y)$
MAPE	1.30%	3.32%	0.29%	4.23%

vehicle routes; see Table 4 for a reference. However, these deteriorations are no more than a fraction of a

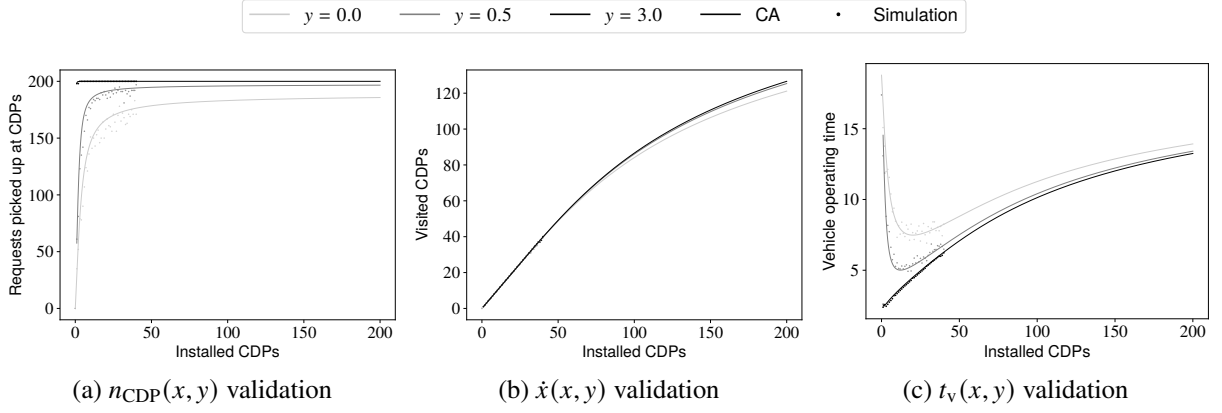


Figure 17: Validation of the main components of the CA model, simulations solved with Hexaly.

percentage point.

Appendix B.2 Further Validation on a Real Road Network

In this section, we use the CA model’s circuitry factor adjustment, simulation procedure, and parameter values introduced in Section 5 to conduct a detailed validation of the CA model on the real road network under study. Specifically, we consider the discrete decision grid formed by $x \in \{0, 1, \dots, 30\}$ and $y \in \{0, 0.25, \dots, 3\}$, and then compare the predictions of our CA model against those obtained from a detailed system simulation.

Figure 18 presents the CA- and simulation-based cost heatmaps as functions of the decision variables (x, y) . In Figure 18a, the CA’s suggested network consists of $x^* \approx 8.47$ CDPs, with an incentive of

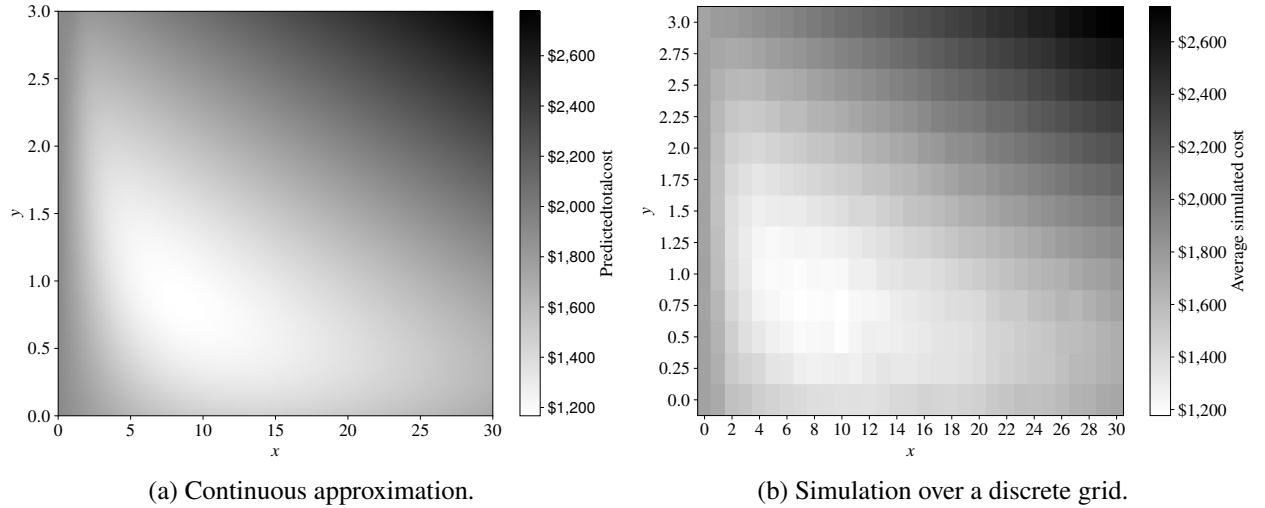


Figure 18: Comparison of the cost functions obtained from the continuous approximation and simulation.

$y^* \approx \$0.85$ per customer and a total cost of \$1,167.00. If we constrain the number of CDPs to be integer-valued ($x \in \mathbb{N}_0$), the suggested network consists of 8 CDPs with an incentive of \$0.88 per customer and a total cost of \$1,167.69. In Figure 18b, the pair of decisions over the aforementioned grid that minimizes the simulated cost consists of 7 CDPs and a customer incentive of \$0.75, yielding a total cost of \$1,177.02. Thus, in this case, our CA model restricted to an integer number of CDPs recommends one more CDPs than the simulation-based solution. To further assess the quality of the CA solution, we simulate the cost corresponding to 8 CDPs and an incentive of \$0.88 per customer; the resulting simulated cost is \$1,235.61, *i.e.*, 4.98% higher than the actual simulation-based minimum cost of \$1,177.02.

Table 10 reports the MAPE for all components of the CA model. In our real-world-inspired study, all

Table 10: MAPE of the CA model on the real-world-inspired study.

	$c(x, y)$	$n_{\text{CDP}}(x, y)$	$\dot{x}(x, y)$	$t_v(x, y)$
MAPE	2.68%	3.56%	0.42%	8.43%

errors on expected value estimations are less than or equal to 8.43%, demonstrating the model’s accuracy even when applied to real, complex road networks that deviate from idealized continuous assumptions.

Appendix B.3 Comparison of plain and demand-balancing k -means results

Comparing Table 8 and Table 11 further illustrates the trade-off between using the plain k -means and the more sophisticated demand-balancing k -means approach: while the plain k -means generates considerably lower simulated costs, most of its solutions are infeasible in practice. In particular, even when equipping our CA model with its capacity chance constraint, the solutions obtained using this capacity-unaware location procedure exhibit average capacity violations of at least 10.8% in simulation. Again, this highlights that the CA model should be viewed as a strategic design tool rather than as a direct prescription of implementable system configurations; in practice, the CA solution identifies the appropriate scale of the system and the corresponding incentive level, after which the specific CDP locations and other operational parameters should be carefully fine-tuned to ensure that the resulting system satisfies the desired operational requirements.

Table 11: Real-world-inspired study results (using the k -means location procedure)

q α_x		30 36.7	40 42.4	50 47.4	60 52.0
Unconstrained CA model					
CA	C^*	1,221.6	1,261.5	1,295.6	1,322.8
	MAPE	0.71%	0.53%	0.61%	0.61%
	x^*	8	7	6	6
	y^*	0.88	0.95	1.04	1.04
Sim.	Cost	1,230.4 $\pm$ 1.8	1,254.9 $\pm$ 1.6	1,287.0 $\pm$ 1.8	1,314.8 $\pm$ 1.9
	# Viol.	8	7	5	5
	Max. Viol.	100.0%	100.0%	100.0%	100.0%
	Avg. Viol.	94.2%	85.3%	78.1%	68.9%
CA model with chance constraint ($\zeta = 5\%$)					
CA	C_P^*	1,460.2	1,393.8	1,369.1	1,365.5
	MAPE	1.40%	0.23%	0.71%	0.09%
	x_P^*	18	13	10	9
	y_P^*	0.29	0.44	0.56	0.82
Sim.	Cost	1,440.1 $\pm$ 2.4	1,390.6 $\pm$ 2.4	1,359.4 $\pm$ 2.2	1,366.7 $\pm$ 1.7
	# Viol.	10	5	5	2
	Max. Viol.	66.0%	73.4%	80.0%	70.2%
	Avg. Viol.	16.0%	14.6%	15.1%	10.8%