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

Benjamín Rojas, Homero Larrain, Mathias Klapp
Pontificia Universidad Católica de Chile, Santiago, Chile
[bgrojas@uc.cl, homero@uc.cl, maklapp@uc.cl]

March 17, 2026

Abstract

Collection and delivery points (CDPs) allow logistics operators to consolidate multiple customer delivery requests on a single vehicle stop, 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 the customer's 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 decides on the network's optimal number of CDPs and the economic incentive offered to customers who choose to be served at CDPs. The incentive is strategically relevant because it induces customers to pick up their orders farther from home, thereby affecting the optimal CDP network size. Furthermore, we equip our model with a chance constraint limiting the probability that the average demand per CDP exceeds its capacity. In our experiments, we show that CDPs alone cut expected costs by 16.9%, while offering incentives to customers increases potential savings to 28.0%. As an insight, we find that CDPs are most beneficial in high-density areas, in which they leverage economies of scale, whereas incentives are more effective in low-density areas. Additionally, our analysis suggests that providing CDPs to logistics operators at no cost has the potential to reduce their distribution costs by up to 60%, making it a potential government policy to lower emissions. We also examine a case study in Santiago, Chile, showing that our continuous approximation model can reliably guide CDP network sizing decisions in a realistic setting. This study provides valuable insights for understanding and designing cost-efficient urban parcel distribution systems.

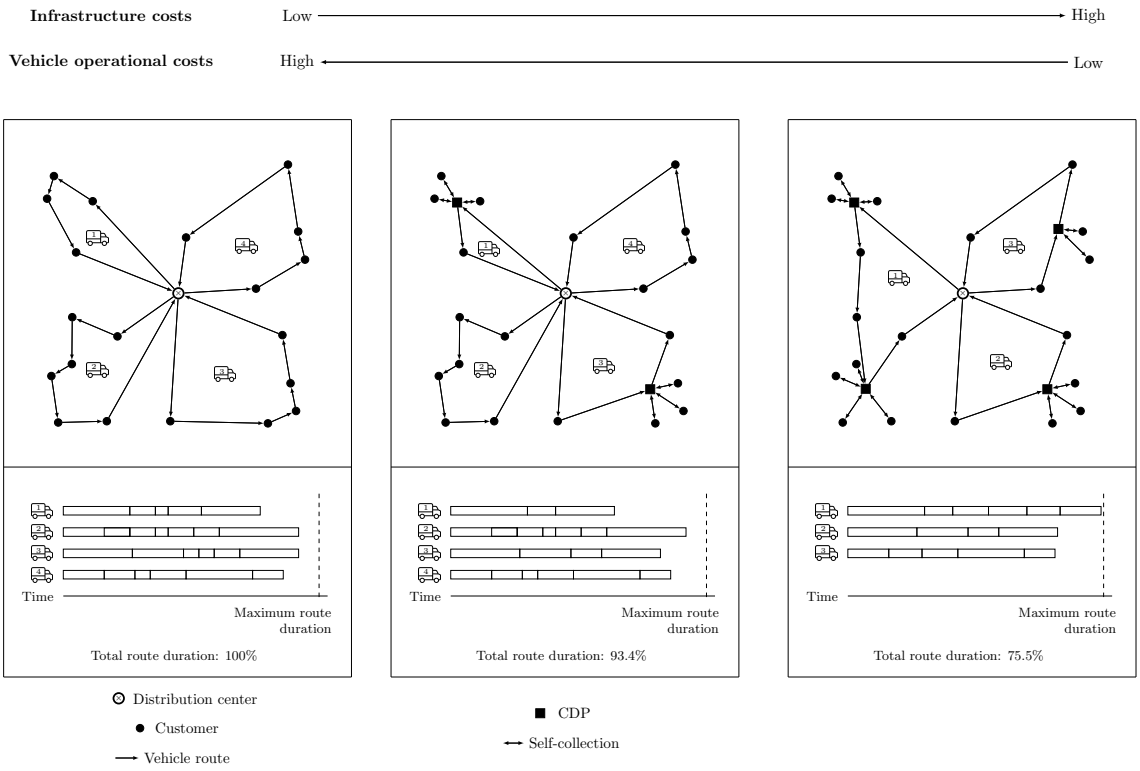
Keywords: Last-mile distribution, Collection and delivery points, Home delivery, Continuous approximation.

1 Introduction

Urban logistics encompasses all processes related to the distribution of goods within cities, including inventory management and last-mile operations. In recent years, the volume of last-mile operations has expanded significantly, with their global market size valued at 185.8 billion in 2025 and projected to surpass 487.2 billion by 2035, primarily fueled by the increase in online orders (Zoting and Shivarkar, 2026). However, the last-mile is often considered the most costly, socially detrimental, and environmentally damaging leg of the supply chain (Archetti and Bertazzi, 2021; Giuliano, 2023). According to Coppola (2024), the last-mile costs across all shipping expenses increased from 41% to 53% between 2018 and 2023. Furthermore, the World Economic Forum (2020) reports that the last-mile distribution operation will contribute to a 21% increase in traffic congestion and growth in pollutant emissions of up to 32% by 2030. From another point of view, customers have raised their expectations for convenient last-mile delivery services. In the US, the Baymard Institute (2026) estimates that approximately 70% of shopping carts are abandoned by online customers before completing a purchase due to unmet service requirements. Among these abandoned carts, 39% are due to high costs (e.g., shipping fees and taxes) and 21% to slow delivery times.

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) where customers can opt to pick up their online purchases; these CDPs 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. To attain a cost-efficient operation, they should carefully balance infrastructure investments and last-mile distribution expenses. A relatively larger CDP network increases infrastructure costs, but also the likelihood that customers will opt to pick up their orders at a CDP instead of choosing home delivery. Consequently, it reduces 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 traditional vehicle routing problem (VRP) routes to complete all delivery visits within each route duration limit. Figure 1b presents the same

instance 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 setting, twelve customers choose to pick up their parcels at CDPs; this reduces the number of delivery stops by eight and travel time by 24.5%. In this research, we aim to strategically optimize the trade-off between operating with a larger number of CDPs and incurring higher customer delivery costs.



(a) Operation without CDPs. (b) Operation with some CDPs. (c) Operation with several CDPs.

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

Within the transportation science and logistics literature, articles focused on designing and planning the operation of CDP networks are scarce. As our literature review reveals, most existing work either optimizes the operational details of these systems for a given CDP network or proposes detailed CDP location models that do not proactively account for operational routing costs or address strategic CDP network sizing questions. Most available studies rely on discrete linear optimization, which is effective for detailed facility location and route planning, but less suited for network sizing problems intended to anticipate expected future operational costs and investments, where only expected operational cost estimates are needed. To determine the size of a CDP network, it may be convenient to estimate downstream routing costs via continuous approximation (CA) techniques, introduced in the

seminal works of Beardwood et al. (1959) and Daganzo (1984), which aim to provide a clear understanding of the trade-offs among various logistical costs based on aggregated data. Such CA approaches have proven effective in other strategic logistics problems (Banerjee et al., 2022, 2025; Carlsson, 2012; Carlsson and Jia, 2015; Carlsson et al., 2024).

Moreover, the potential benefits of installing CDPs depend on the expected number of customers who opt to use them as order pickup locations. So, one should understand average customer preferences to predict the expected utilization of a given CDP network size (Ma et al., 2022). Despite this, there exists a research gap in studying how customer choices impact distribution costs and how these choices can be influenced via economic incentives. In particular, whether or not a customer chooses to pick up parcels at a CDP 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 (1) installing more CDPs, thereby increasing CDP network density and reducing the distance between customers and their nearest CDP (Molin et al., 2022) or (2) offering direct incentives to customers using these facilities, such as purchase discounts (*i.e.*, coupons) or delivery fee discounts (Yuen et al., 2018). Although recent research articles have explored how customer choice behavior impacts CDP system design (*e.g.*, Lin et al. (2020, 2022); Xu et al. (2021)), to our knowledge, only Galiullina et al. (2024) and Akkerman et al. (2025) have studied the impact of economic incentives on customers' delivery choices at the operational level, while Zhang et al. (2023) address this issue at the strategic level via a user equilibrium framework.

In this paper, we explore how to optimally determine the size of a CDP network anticipating its impact on customers' delivery service choices and downstream operational costs. To do so, we present a novel CA model that approximates the expected operational cost of the system as the sum of infrastructure, distribution, and incentive costs. We use this approximation to build a stylized cost minimization approach assuming that the decision-maker controls the structural variables associated with (1) the CDP network size and (2) the incentive offered to customers who opt to pick up their orders at CDPs. To further refine our model, we present two chance constraints to enhance our approximate model and ensure that the average demand per CDP exceeds its capacity with low probability. The first constraint relies on a concentration inequality to upper bound the cumulative distribution function of the number of customers selecting CDPs, yielding a conservative approximation. The second approximates the random number of customers choosing CDPs as a normal random variable.

We empirically validate our CA model by comparing it against average results from simulated scenarios, each with a detailed set of customer locations and service choices, as well as optimized delivery routes. Furthermore, we conduct a realistic case study in the

metropolitan area of Santiago, Chile, the results of which show that our CA model accurately predicts the distribution cost of an operation with CDPs and customer incentives. Based on our CA model's validation and case study, we derive the following insights:

- Compared to a home delivery operation, an operation with an optimally sized CDP network generates substantial savings, especially when request density is relatively high and infrastructure cost is relatively low.
- In low request-density settings, promoting CDP adoption through incentives becomes particularly valuable. Incentives allow the operator to reduce the number of installed CDPs while preserving high utilization levels, as incentivized customers are willing to travel longer distances.
- CDP network sizing decisions require advanced information on customer behavior, as they should account for customers' sensitivity to CDP distance. Errors from assuming deterministic customer choices increase as this sensitivity decreases.
- Total optimal expected costs increase sublinearly with infrastructure cost, as the system can adjust by reducing the number of CDPs installed and increasing incentives. If CDP infrastructure is provided free of charge, *e.g.*, by a government authority, then distribution costs can be reduced by over 60%, making it a potential government policy to lower emissions. In such cases, customer incentives become less essential.
- As the distribution center is located farther from the service region, distribution costs grow as maximum route durations are consumed by stem travel times. In such a case, the optimal strategy shifts from expanding the CDP network to offering higher customer incentives.
- There exists a trade-off between installing more CDPs and increasing each CDP's capacity to serve more customers. Adding more capacity to each CDP reduces the optimal number of installed CDPs and increases the optimal customer incentive level for using them.

The remainder of this paper is structured as follows. We revise related literature in Section 2. In Section 3, we describe the studied problem and formulate our approximate model. Then, we present the base-case parameters and detail the calibration and validation of our model. In Section 4, we derive managerial insights through multiple sensitivity analyses on key parameters. We further present a case study on a realistic road network setting in Section 5. Finally, we conclude our work in Section 6.

2 Literature Review

We present a review of the literature on the operation and design of CDP networks to support the home delivery process. For a comprehensive review on self-attended delivery operational challenges and on the interaction between delivery operations and customer behavior, refer to Janinhoff et al. (2024a) and Ma et al. (2022), respectively. Table 1 summarizes our literature review, categorizing each article as operational, strategic, or both based on the scope of the decisions involved. It also indicates if CDP capacity constraints are considered and who is assumed to select the delivery option (between home delivery and order pickup at a CDP), detailing the customer's choice model used.

At the operational level, the existing literature focuses on the efficient use of a given CDP network and concentrates on last-mile routing decisions relying on VRP models (see Toth and Vigo (2014)). Particularly, these articles relate to the VRP model with delivery options (see Dumez et al. (2021) and Tilk et al. (2021)), where each customer may offer more than one delivery location to represent the potential use of both options, home delivery and customer pickup at CDPs. Dumez et al. (2021) and Tilk et al. (2021) study a VRP with delivery options and service time windows in which a minimum percentage of customers must be served at each tier of their preferred delivery locations. Janinhoff et al. (2024b) extend the VRP with delivery options to a setting that allows multiple vehicle trips starting and ending at a depot. The authors analyze cases where customers are charged different prices for home delivery; however, they do not explicitly model these charges as decision variables influencing customers' service choices. Zang et al. (2023) examine an urban delivery problem with time windows, in which a subset of chosen deliveries are assigned to a CDP located within a maximum threshold distance to each customer. Mancini and Gansterer (2021) propose a VRP model where customers can either be served at home or receive some compensation to pick up their parcels at a CDP. Dell'Amico et al. (2023) present a pickup and delivery problem with CDPs, where customers can choose between home delivery, pickup at CDPs, or both options. dos Santos et al. (2022) define a two-echelon last-mile delivery problem with CDPs and occasional couriers, where CDPs can be used as both pickup points and transshipment facilities. Galiullina et al. (2024) model and solve the problem of simultaneously planning vehicle routes and offering incentives to customers using CDPs. Akkerman et al. (2025) study the problem of offering menus of CDPs and their respective incentives (or charges) to customers who arrive dynamically in the system.

At the strategic level, the available work focuses on detailed CDP network design, *i.e.*, facility location models. As the literature on facility location problems is vast (refer to Laporte et al. (2019) for a book on this topic), we focus on studies that examine CDP location decisions. Deutsch and Golany (2018) propose a model to optimally locate a profit

Table 1: A classification of studies involving CDPs.

Category	Reference	Decisions involved					CDP capacity constraints	Who selects between CDP pickup and home delivery?			
		Number of CDPs	Approximate routing costs	Incentive to customers	Detailed location of CDPs	Detailed route planning		Operator	Customer		
									Exogenous choice	Distance threshold choice	Discrete choice model
Operational	Dumez et al. (2021)					✓	✓	✓			
	Tilk et al. (2021)					✓	✓	✓			
	Janinhoff et al. (2024b)					✓	✓	✓			
	Zang et al. (2023)					✓		✓			
	Mancini and Gansterer (2021)					✓	✓	✓			
	Dell'Amico et al. (2023)					✓	✓	✓	✓		
	dos Santos et al. (2022)					✓	✓		✓		
	Galiullina et al. (2024)			✓		✓					✓
	Akkerman et al. (2025)			✓			✓				✓
Strategic	Deutsch and Golany (2018)	✓			✓					✓	
	Mancini et al. (2023)				✓		✓			✓	
	Raviv (2023)				✓		✓			✓	
	Kahr (2022)	✓			✓		✓	✓			
	Lin et al. (2020)				✓						✓
	Lin et al. (2022)				✓						✓
	Lyu and Teo (2022)				✓						✓
	Xu et al. (2021)				✓		✓				✓
Operational and strategic	Enthoven et al. (2020)				✓	✓		✓			
	Ozyavas et al. (2025)				✓	✓	✓	✓			
	Janjevic et al. (2019)		✓		✓	✓	✓			✓	
	Janinhoff and Klein (2023)				✓	✓	✓				✓
	Wang et al. (2025)		✓		✓						✓
	Zhang et al. (2023)		✓	✓			✓				✓
	Zadeh et al. (2026)				✓	✓	✓	✓			
	Our work	✓	✓	✓			✓			✓	✓

maximizing CDP network. Mancini et al. (2023) studies how to locate CDPs under uncertain customer demands and CDP order processing capacity. Raviv (2023) examines a CDP location problem that simultaneously sets each CDP's request storage capacity. Similarly, Kahr (2022) studies a detailed CDP network design problem, including CDP location and layout decisions. Lin et al. (2020) and Lin et al. (2022) investigate a detailed CDP location problem combined with customer discrete choice models to predict each customer's likelihood of picking up their goods at a CDP. The customer's utility within the choice models is assumed to be a decreasing function of the distance between the customer's location and CDPs. Lyu and Teo (2022) study a CDP location problem to maximize the proportion of orders delivered to CDPs. The authors propose a discrete choice model to predict CDP usage considering factors such as customer location, CDP accessibility, and the presence of other nearby CDPs. Xu et al. (2021) propose a data-driven optimization method to determine the detailed location of CDPs via historical customer demand data and customer purchase forecast models. All available strategic models aim to provide a detailed location for CDPs, selecting from a known set of location options. Moreover, they consider settings where the CDP service is the only option, disregarding the home-delivery service and its routing costs.

As strategic and operational levels are interdependent, some research efforts combine both the CDP network design problem and its impact on downstream operational routing costs. These works relate to location-routing problems (see Nagy and Salhi (2006)), and in particular to the location-or-routing problem (see Arslan (2021)), in which customers can pick up their goods at facilities if located within a given threshold distance or can be served by delivery routes. Enthoven et al. (2020) define a two-echelon VRP where the second echelon defines covering and satellite locations. While covering locations act as CDPs where customers pick up their parcels, satellite locations serve as transshipment facilities where goods are transferred to low-emission delivery vehicles. Similarly, Ozyavas et al. (2025) study a two-echelon location-routing problem with capacity constraints, where delivery bikes cover the second-echelon routes. Janjevic et al. (2019) define a multi-echelon problem that integrates the use of CDPs into a larger distribution network comprising a central hub and several satellite facilities. Janinhoff and Klein (2023) study a CDP location–routing problem under uncertainty in daily customer arrivals and in customers' service choices, which are influenced by CDP locations. Zadeh et al. (2026) study strategic decisions regarding CDP location, capacity, and vehicle fleet size, while accounting for aggregate operational decisions such as vehicle routing, CDP assignment, and replenishment planning. Wang et al. (2025) analyze how the geographic location of CDPs influences customers' choice of CDPs for picking up their goods. Also, the authors study how these customers' choices impact the total travel distance of last-mile distribution routes. Specifically, they examine both the

case where CDPs and customer locations are served by separate routes and the case where both are covered by a single route. However, the authors do not consider infrastructure costs or the option to enhance CDP adoption through economic incentives to customers. Zhang et al. (2023) study profit-maximizing CDP location and service pricing decisions. They model customer choices via a user equilibrium framework and distribution costs via CA.

Our study builds on the existing literature of continuous approximation (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 $n \to \infty$, meaning that we can generally approximate $L_n^\star$ as

$$L_n^\star \approx \beta \cdot \sqrt{n \cdot A}, \tag{1}$$

where β is a proportionality constant. The study of Daganzo (1984) extends this approach to a vehicle routing setting with multiple vehicles and a stem distance cost. For an overview 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. Specifically, Hazbún (2019) estimates the economic and social costs of using CDPs to aid last-mile distributions via a CA model. 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 CA 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 CA model to terminal design and propose a method to discretize the solution obtained with a small optimality gap. Pulido et al. (2015) define a CA model for locating warehouses in the context of same-day home delivery. Carlsson and Jia (2015) present a CA 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. Compared to this article, we focus on a single-echelon distribution system that explicitly incorporates customer choice, allowing customers to pick up their products at a CDP, and examine how economic incentives can influence their service selection.

In the last-mile logistics literature, there exist studies on how customer behavior can be influenced via economic incentives. For instance, there are research efforts focused on creating temporal and spatial flexibility in customers via incentives or pricing mechanisms. Examples include using incentives to encourage customers to select convenient time windows

for attended home delivery (Klein et al., 2019; Yildiz and Savelsbergh, 2020), and pricing delivery zones to distribute demand and maximize profits (Afsar et al., 2021; Afsar, 2022). Also, Cerulli et al. (2024), Çınar et al. (2024), Horner et al. (2024), and Wang et al. (2024) study the influence of customer incentive mechanisms in different crowd-based last-mile delivery settings.

3 The CDP Network Sizing Problem

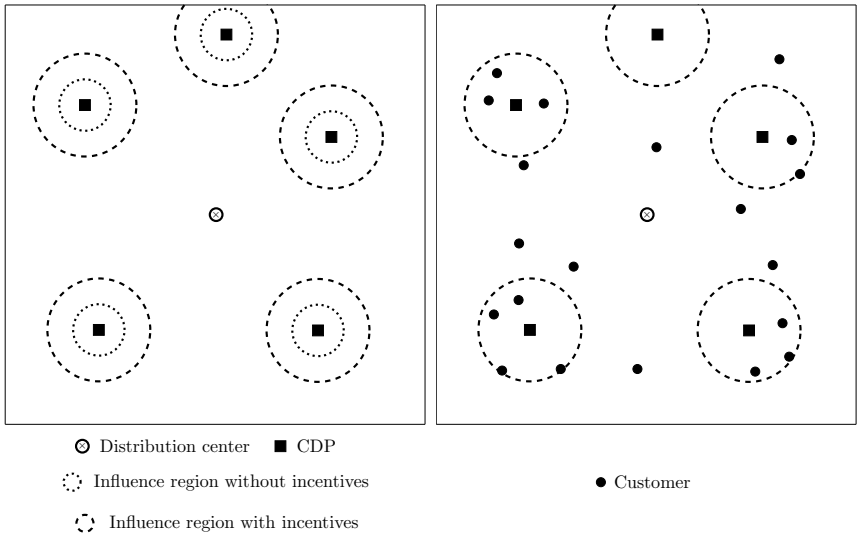
We now introduce the CDP Network Sizing Problem (CDP-NSP), which aims to determine the optimal number of CDPs to install and the economic incentives offered to customers in order to minimize expected costs. In this system, deliveries are made either to customers' locations or to CDPs, depending on customer choice. Next, we formulate a CA model and a customer service choice model to approximate the system's structural decisions and costs. Finally, we calibrate and validate the model.

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 area who placed 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, 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.

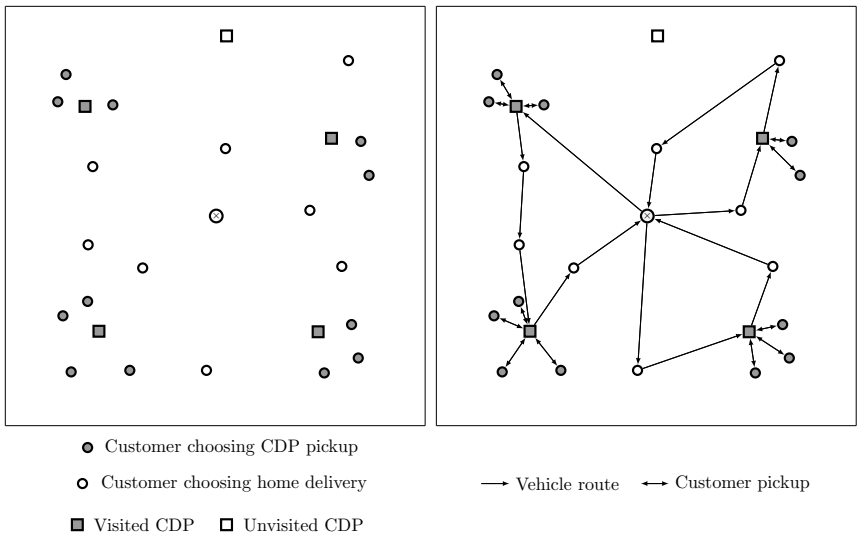
In Figure 2, we illustrate our problem's dynamics. Figure 2a describes the operator's decisions; these are (1) sizing the CDP network over the service region and (2) determining the economic incentive offered to customers picking up their parcels at CDPs. Here, incentive decisions are made prior to the realization of customer orders, since customers must be aware of the incentives when making a request. For simplicity, 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.



(a) Strategic decisions.

(b) Realized customer requests.



(c) Requests division.

(d) Operational decisions.

Figure 2: Dynamics of the CDP network sizing problem.

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 will intentionally avoid 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 CDP 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 CDP location decisions are intertwined and depend on customer choices, defining where deliveries should occur.

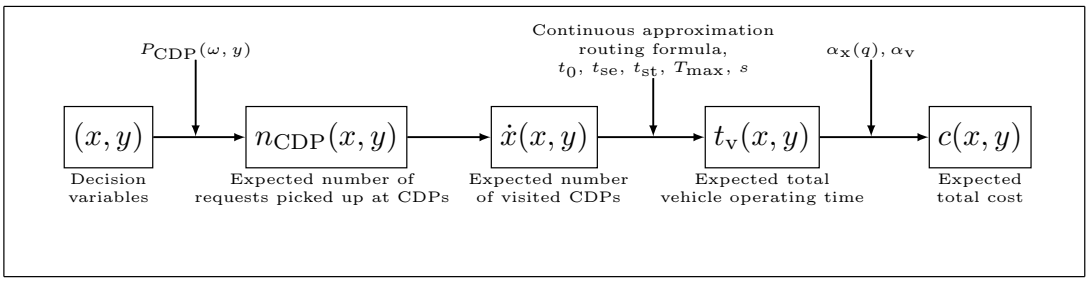
3.2 Approximate cost model

In what follows, we present a model to approximate the expected cost of a last-mile distribution operation equipped with CDPs as a function of the number x of installed CDPs and the economic incentive y offered to each customer choosing to pick up their requests at a CDP. Our model only requires aggregated data, such as the expected number of customers n and the service area of A km^2 during the operating period. We assume that customer locations are uniformly and independently distributed over the service area, and that each customer chooses between home delivery and picking up their goods at the nearest CDP depending on their relative distance to the CDP and y. This is consistent with empirical data suggesting that approximately 79% of customers prefer CDPs located close to their residences (Lemke et al., 2016; Galiullina et al., 2024). We avoid setting the exact location of the CDP network and assume that the x CDPs are evenly distributed over the service region, forming hexagonal influence regions, which is a reasonable assumption for a uniformly distributed demand (Newell, 1973; Papadimitriou, 1981; Carlsson and Jia, 2015). We consider CDPs with the capacity to serve up to q customers during the operational period, which may be unbounded. In our model, we estimate last-mile distribution costs via CA instead of relying on detailed vehicle route planning.

We consider the following cost components: (1) a cost α_{x} per installed CDP, possibly dependent on its capacity q, representing rent, maintenance, insurance, tax, and related infrastructure costs; (2) a last-mile distribution cost α_{v} per unit time operated by each vehicle, representing vehicle lease and driver wage; and (3) an economic incentive y paid to each customer collecting their goods at CDPs. At each stop, we assume that each vehicle incurs a fixed stop time t_{st} plus an additional service time t_{se} per delivered order. We also consider that the vehicle's maximum workday length (defined as $T_{\max}$) is the limiting resource, as we study the delivery of small-sized and lightweight parcels and disregard vehicle capacity constraints.

Figure 3 provides an overview of the approximate cost model. The cost function $c(x, y)$ is defined in a series of approximation steps starting from a pair of values x and y. First, we estimate the expected number of requests picked up by customers at CDPs (referred to as $n_{\mathrm{CDP}}(x, y)$) by assuming a customer-invariant probability function $P_{\mathrm{CDP}}(\omega, y)$, which

represents the likelihood of a customer choosing to pick up their orders 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 used per day $\dot{x}(x, y) \leq x$, as not all CDPs installed are expected to have demand each day. Next, we leverage on continuous approximation to estimate the required total vehicle operating time $t_{\mathrm{v}}(x, y)$ required to visit $n - n_{\mathrm{CDP}}(x, y)$ customers' homes plus $\dot{x}(x, y)$ CDP locations. Finally, we integrate all steps to approximate $c(x, y)$. Each step is detailed as follows, and Table 2 summarizes our notation.



x: Number of installed CDPs.
y: Incentive per customer choosing the CDP option.

Figure 3: Outline of the series of modeling steps.

Table 2: Parameter and variable notation

Notation	Parameter	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· day)
α_{v}	Vehicle operating time cost.	\$/hour
	Variable	
x	Number of installed CDPs.	CDPs
y	Incentive per customer choosing the CDP option.	\$/request
$P_{\mathrm{CDP}}(\omega, y)$	Customer CDP choice probability.	-
$n_{\mathrm{CDP}}(x, y)$	Expected number of requests picked up at CDPs	requests/day
$\dot{x}(x, y)$	Expected number of visited CDPs.	CDPs
$t_{\mathrm{v}}(x, y)$	Expected total vehicle operating time.	hours/day
$c(x, y)$	Expected total cost.	\$/day

Expected number of requests picked up at CDPs. To estimate the expected number of pickups at CDPs, let Ω_k denote the random variable representing the distance between the k^{th} customer and their nearest CDP, for $k \in \{1, ..., n\}$. Further, define

$B_k(x,y) \sim \text{Bernoulli}\left(P_{\text{CDP}}(\Omega_k, y)\right)$ as 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 $N_{\text{CDP}}(x,y)$ is represented as

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

and its expected value is

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

To estimate $n_{\text{CDP}}(x,y)$, we propose a continuous approximation in which the hexagonal influence region of each CDP is approximated as a circle of radius $r(x) = \sqrt{A/(\pi \cdot x)}$. 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, we obtain

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

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, define $V_i(x,y)$ for each $i \in \{1, ..., x\}$ as a binary random variable taking value 1 when the i^{th} CDP is left unused, and 0 otherwise. We assume that $V_i(x,y) \sim \text{Bernoulli}\left((1-1/x)^{N_{\text{CDP}}(x,y)}\right)$; in this expression, the term $1/x$ denotes the probability that a given customer selects a particular CDP, while $(1-1/x)$ represents the probability that the customer does not select it. Therefore, $(1-1/x)^{N_{\text{CDP}}(x,y)}$ represents the probability that a particular CDP is not selected by any of the customers who choose the CDP pickup option. Let the random variable representing the number of visited CDPs be

$$\dot{X}(x,y) = \sum_{i=1}^{x} \left(1 - V_i\left(x,y\right)\right), \tag{5}$$

its expected value is defined as

$$\dot{x}(x,y) = \mathbb{E}\left[\dot{X}(x,y)\right] = x - \mathbb{E}\left[\sum_{i=1}^{x} V_i(x,y)\right] = x \cdot \left(1 - \mathbb{E}\left[\left(1 - \frac{1}{x}\right)^{N_{\text{CDP}}(x,y)}\right]\right). \tag{6}$$

We make the conservative estimate $\mathbb{E}\left[(1-1/x)^{N_{\mathrm{CDP}}(x,y)}\right] \approx (1-1/x)^{n_{\mathrm{CDP}}(x,y)}$ as Jensen's inequality states that $\mathbb{E}[(1-1/x)^{N_{\mathrm{CDP}}(x,y)}] \geq (1-1/x)^{n_{\mathrm{CDP}}(x,y)}$ for $x \geq 1$. Moreover, we also estimate $(1-1/x)^{n_{\mathrm{CDP}}(x,y)} \approx \mathrm{e}^{-n_{\mathrm{CDP}}(x,y)/x}$ using the property of the exponential function $\lim_{x\to\infty}(1-1/x)^{x\cdot n_{\mathrm{CDP}}(x,y)/x} \approx \mathrm{e}^{-n_{\mathrm{CDP}}(x,y)/x}$. Then $\dot{x}(x,y) \leq x$ is approximated as

$$\dot{x}(x,y) \approx \begin{cases} x \cdot \left(1 - \mathrm{e}^{-\frac{n_{\mathrm{CDP}}(x,y)}{x}}\right), & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Expected total vehicle operating time. In our setting, the number of delivery stops $M(x,y)$ is the sum of the number of used CDPs plus the number of home deliveries, *i.e.*, $M(x,y) = \dot{X}(x,y) + n - N_{\mathrm{CDP}}(x,y)$. The BHH asymptotic square-root formula still applies, as the $M(x,y)$ delivery stops are i.i.d. over the service area, following a mixture distribution induced by the n uniformly and independently distributed customers and the customer choices $B_k(x,y)$. Therefore, considering the service time per request, the fixed time per stop, and the square-root approximation of the routing time between stops, we estimate the vehicles' operating time (excluding stem routing) as $T_{\mathrm{op}}(x,y) \approx \gamma \cdot t_{\mathrm{se}} \cdot n + \delta \cdot t_{\mathrm{st}} \cdot M(x,y) + \eta/s \cdot \sqrt{M(x,y) \cdot A}$, where γ, δ and η are constants to calibrate. To estimate its expected value $\mathbb{E}[T_{\mathrm{op}}(x,y)]$, we can replace the number of stops $M(x,y)$ by its expected value $m(x,y) = \mathbb{E}[M(x,y)] \approx \dot{x}(x,y) + n - n_{\mathrm{CDP}}(x,y)$, as proposed by Banerjee et al. (2025). Therefore, we obtain

$$t_{\mathrm{op}}(x,y) = \mathbb{E}[T_{\mathrm{op}}(x,y)] \approx \gamma \cdot t_{\mathrm{se}} \cdot n + \delta \cdot t_{\mathrm{st}} \cdot m(x,y) + \frac{\eta}{s} \cdot \sqrt{m(x,y) \cdot A}. \tag{8}$$

We add stem routing times to Equation (8) following the approach proposed by Daganzo (1984) and estimate the expected total vehicle operating time as

$$t_{\mathrm{v}}(x,y) \approx 2 \cdot t_0 \cdot \frac{t_{\mathrm{v}}(x,y)}{T_{\max}} + t_{\mathrm{op}}(x,y), \tag{9}$$

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 $(t_{\mathrm{v}}(x,y)/T_{\max})$ plus the remaining local operating time $t_{\mathrm{op}}(x,y)$, as illustrated in Figure 4.

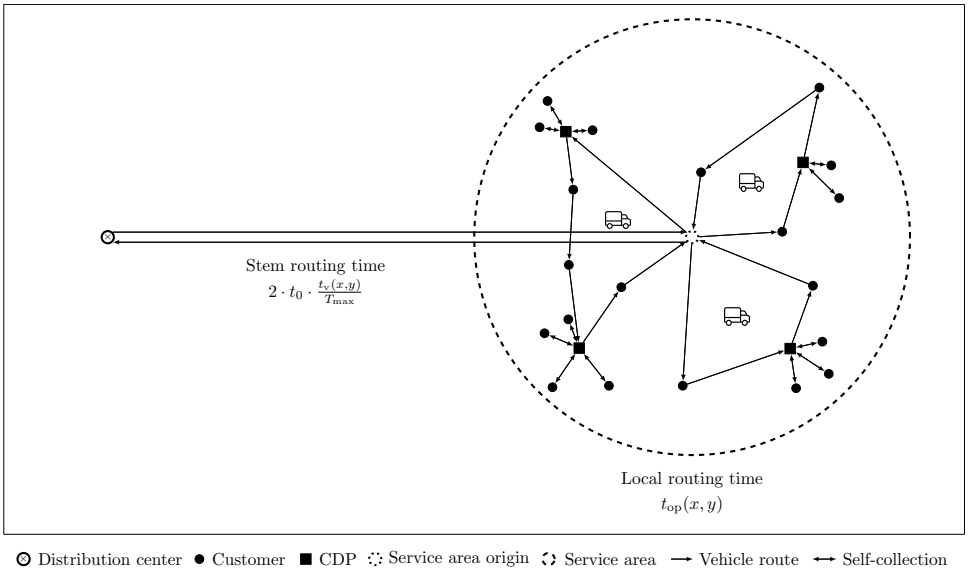


Figure 4: Expected total vehicle operating time approximation.

If we define $\kappa = T_{\max}/(T_{\max} - 2 \cdot t_0)$ and isolate $t_{\mathrm{v}}(x, y)$ from Equation (9), then we obtain

$$t_{\mathrm{v}}(x, y) \approx \gamma \cdot \kappa \cdot t_{\mathrm{se}} \cdot n + \delta \cdot \kappa \cdot t_{\mathrm{st}} \cdot m(x, y) + \eta \cdot \frac{\kappa}{s} \cdot \sqrt{m(x, y) \cdot A}. \tag{10}$$

As expected, when the maximum vehicle workday length $T_{\max} \to 2 \cdot t_0$ from above, the operation becomes infeasible and the total vehicle operating time diverges to infinity. In contrast, when $T_{\max} \to \infty$, then the total vehicle operating time $t_{\mathrm{v}}(x, y)$ converges to $t_{\mathrm{op}}(x, y)$, *i.e.*, the CA formula for the operating time of a traveling salesman tour.

Expected total cost. Finally, we approximate the system's total cost, including CDP infrastructure, vehicle distribution, and incentive costs as

$$c(x, y) \approx \alpha_{\mathrm{x}} \cdot x + \alpha_{\mathrm{v}} \cdot t_{\mathrm{v}}(x, y) + y \cdot n_{\mathrm{CDP}}(x, y). \tag{11}$$

Below, we present a theoretical result supporting the validity of Equation (11), which relates to the average cost per request function $c(x, y)/n$; its proof is detailed in Appendix A.1.

Property 1. The average cost per request $c(x, y)/n$ only depends on x, n and A through the request density $\rho = n/A$ and the number of requests received per installed CDP $\nu(x) = n/x$.

This property states that the expected cost per request is scale-independent and aligns with the classic BHH formula, where traveled distance per customer solely depends on customer density, *i.e.*, $L_n^\star/n \approx \beta \cdot \sqrt{n \cdot A}/n = \beta/\sqrt{\rho}$.

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. As an auxiliary model, we define $\tau(y)$ as the distance a typical 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. This 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 (12), 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.

$$P_{\text{CDP}}^{\text{DTC}}(\omega, y) = \begin{cases} 1, & \text{if } \omega \leq \tau(y), \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

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

$$n_{\text{CDP}}(x, y) \approx \begin{cases} n \cdot \min\left\{1, \left(\frac{\tau(y)}{r(x)}\right)^2\right\}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Intuitively, Equation (13) 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 unrealistic, as customers' behavior is likely to show variability in practice. We generalize this choice model and account for random customer behavior via the logit choice model we describe next.

Binomial logit choice (BLC). We now leverage the random utility theory (Block, 1974) and propose that customers choose to pick up their parcels at a CDP depending on a binomial logit choice model (McFadden, 1973) defined as

$$P_{\text{CDP}}^{\text{BLC}}(\omega, y) = \frac{1}{1 + \mathrm{e}^{\lambda \cdot (\omega - \tau(y))}}, \tag{14}$$

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

$$n_{\text{CDP}}(x,y) \approx$$
$$\begin{cases} n \cdot \left(1 + \frac{2\cdot \text{Li}_2\left(-\text{e}^{-\lambda\tau(y)}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2\cdot \text{Li}_2\left(-\text{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2\cdot \ln\left(1+\text{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda \cdot r(x)}\right), & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

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 (15) is presented in Appendix A.2. The following property, proven in Appendix A.3, establishes that the BLC model generalizes the DTC model.

Property 2. If $\omega \neq \tau(y)$, then the BLC and DTC models converge in probability as the scale parameter $\lambda \to \infty$.

This property suggests that the BLC model becomes an all-or-nothing model when customers are highly sensitive to the difference $\tau(y) - \omega$. In this case, customers' behavior becomes deterministic, except for those located exactly at distance $\tau(y)$ from the closest CDP; under our modeling assumptions this occurs with probability zero.

3.4 Total cost minimization

We can minimize the total cost function by solving

$$\min\left\{c(x,y) \quad \text{s.t.} \quad 0 \leq x \leq n,\ 0 \leq y \leq \tau^{-1}(R)\right\}, \tag{16}$$

and approximately optimize structural decisions (x, y) for our last-mile distribution operation equipped with a network of CDPs. In Model (16), the number of installed CDPs x is upper bounded by the number of requests n, while the maximum economic incentive y is limited to the amount necessary to induce customers to travel up to the service area radius R.

3.5 CDP capacity constraints

We now present an approximate method to equip Model (16) with the following chance constraint

$$\mathbb{P}\left(N_{\text{CDP}}(x,y) \geq q \cdot x\right) \leq \zeta, \tag{17}$$

imposing that the expected number of customers choosing service per CDP ($N_{\text{CDP}}(x,y)/x$) exceed CDP capacity (q) with a probability less than or equal to a prescribed value $\zeta \in (0, 1]$. This chance constraint is slightly weak, as it imposes that the average capacity across CDPs is

not violated with a given probability; a more precise constraint would enforce capacity limits at each individual CDP. However, our subsequent approximations either neglect variability or overestimate (17), so this approach remains reasonable in practice.

It is difficult to directly enforce constraint (17), as we do not have the cumulative distribution function of $N_{\text{CDP}}(x, y)$. Alternatively, we present three approaches to approximate it. A first simple approach is to impose that the expected number of customers choosing to pick up their goods at CDPs is less than or equal to q as

$$n_{\text{CDP}}(x, y) \leq q \cdot x. \tag{18}$$

However, this constraint only limits capacity for the expected number of services per CDP and is weaker than (17).

A more elaborate approach is building an upper bound $U(x, y)$ on the probability $\mathbb{P}\left(N_{\text{CDP}}(x, y) \geq q \cdot x\right)$ for each pair (x, y). Then, by enforcing the constraint $U(x, y) \leq \zeta$, we ensure satisfaction of the chance constraint, though this may overrestrict the feasible region depending on the tightness of the chosen upper bound. Concentration inequalities (Boucheron et al., 2013) are useful results to upper bound cumulative distribution functions (or equivalently, complementary cumulative distribution functions). Specifically, we use Cantelli's inequality, which states that

$$\mathbb{P}\left(N_{\text{CDP}}(x, y) \geq q \cdot x\right) \leq \frac{\sigma^2_{\text{CDP}}(x, y)}{\sigma^2_{\text{CDP}}(x, y) + \left(q \cdot x - n_{\text{CDP}}\left(x, y\right)\right)^2}, \tag{19}$$

where $\sigma^2_{\text{CDP}}(x, y)$ is the variance function of the random number of customers choosing CDPs $N_{\text{CDP}}(x, y)$. Then, we can upper bound the right-hand-side of (19) by ζ and derive

$$n_{\text{CDP}}(x, y) \leq q \cdot x - \sqrt{\frac{1 - \zeta}{\zeta}} \cdot \sigma_{\text{CDP}}(x, y), \tag{20}$$

which is more restrictive than the expected value constraint (18) and guarantees the satisfaction of the original chance constraint (17). This inequality requires function $\sigma^2_{\text{CDP}}(x, y)$,

which is formally defined as

$$\begin{aligned}\sigma^2_{\mathrm{CDP}}(x,y) = \mathbb{V}\left[N_{\mathrm{CDP}}(x,y)\right] &= \sum_{k=1}^{n} \mathbb{V}\left[B_k(x,y)\right] \\ &= \sum_{k=1}^{n} \left(\mathbb{E}\left[B_k(x,y)^2\right] - \mathbb{E}\left[B_k(x,y)\right]^2\right) \\ &= \sum_{k=1}^{n} \left(\mathbb{E}\left[B_k(x,y)\right] - \mathbb{E}\left[B_k(x,y)\right]^2\right) \\ &= \sum_{k=1}^{n} \left(\mathbb{E}[P_{\mathrm{CDP}}\left(\Omega_k, y\right)] - \mathbb{E}\left[P_{\mathrm{CDP}}\left(\Omega_k, y\right)\right]^2\right). \end{aligned} \tag{21}$$

Then, we approximate this variance via an analogous continuous approximation as $n_{\mathrm{CDP}}(x,y)$ in Equation (4) to obtain

$$\sigma^2_{\mathrm{CDP}}(x,y) \approx \begin{cases} n_{\mathrm{CDP}}(x,y) - \frac{n^2_{\mathrm{CDP}}(x,y)}{n}, & \text{if } x, n > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

An alternative capacity constraint can be derived using Markov's inequality; however, in our experiments, the use of Cantelli's inequality yielded a tighter approximation.

Additionally, we introduce an alternative proxy for the chance constraint (17), which assumes that $N_{\mathrm{CDP}}(x,y)$ follows an approximate normal distribution with mean $n_{\mathrm{CDP}}(x,y)$ and variance $\sigma^2_{\mathrm{CDP}}(x,y)$. This approximation is justified for sufficiently large values of n, since $N_{\mathrm{CDP}}(x,y)$ is the sum of n binary random variables. If we use the cumulative distribution function Φ of a standard normal random variable, we can approximate the probability function

$$\mathbb{P}\left(N_{\mathrm{CDP}} \geq q \cdot x\right) \approx 1 - \Phi\left(\frac{q \cdot x - n_{\mathrm{CDP}}(x,y)}{\sigma_{\mathrm{CDP}}(x,y)}\right), \tag{23}$$

and upper bound it by ζ to derive

$$n_{\mathrm{CDP}}(x,y) \leq q \cdot x - \Phi^{-1}(1-\zeta) \cdot \sigma_{\mathrm{CDP}}(x,y), \tag{24}$$

which is more restrictive than the expected value constraint, but less restrictive than (20) as $\Phi^{-1}(1-\zeta) \leq \sqrt{(1-\zeta)/\zeta}$ for all $\zeta \in (0,1)$. If we equip model (16) with constraints (18), (20) or (24), the resulting formulation is a nonconvex continuous optimization problem. Nevertheless, as the problem is two-dimensional and defined over a compact domain, it can be efficiently approximated through a bi-dimensional grid search.

3.6 Base parameter values in our study

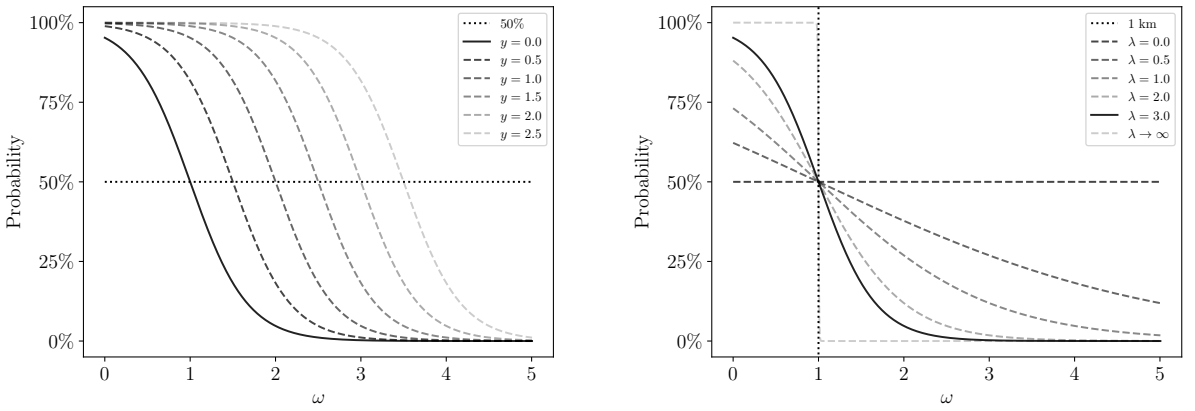
We next set the parameter values used to calibrate and test our approximate cost model for a base case, as shown in Table 3. We consider a daily operation expecting $n = 200$ requests within a circular service region of radius $R = 3$ km, which results in a request density $\rho \approx 7.07$ requests/$\left(\text{km}^2 \cdot \text{day}\right)$; this value is comparable to the 6.5 requests/$\left(\text{km}^2 \cdot \text{day}\right)$ presented by Janjevic et al. (2019). The average vehicle speed is assumed to be $s = 15$ km/hours, 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 $t_0 = 2R/(3s) \approx 0.13$ hours. We define a service time of $t_{\text{se}} = 0.5$ minutes per parcel delivered. Also, the time spent per vehicle stop in a service route is $t_{\text{st}} = 3.5$ 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 relaxed in Section 4.7 and Section 5. The cost paid per installed CDP is set to $\alpha_{\text{x}} = \$30/(\text{CDP}\cdot\text{day})$, which lies within the \$16 to \$34 range discussed by Xu et al. (2021). The hourly cost per operating vehicle is set to $\alpha_{\text{v}} = \$30/\text{hour}$ based on a \$28/hour driver salary and a \$0.14/km fuel cost.

Table 3: Parameter values in 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

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. Figure 5a plots the CDP choice probability $P^{\text{BLC}}_{\text{CDP}}(\omega, y)$ over different values of incentive y. Our choice

of λ is conservative as $P^{\text{BLC}}_{\text{CDP}}(\omega = 3, y = 0) \leq 0.1\%$. As expected, $P^{\text{BLC}}_{\text{CDP}}(\tau(y), y) = 50\%$. Figure 5b depicts $P^{\text{BLC}}_{\text{CDP}}(\omega, y)$ for different values of λ when $y = \$0$/request. As confirmed by Property 2, a customer's CDP service choice becomes less variable and more predictable when λ increases.



(a) $P^{\text{BLC}}_{\text{CDP}}(\omega, y)$ over different incentive values y.

(b) $P^{\text{BLC}}_{\text{CDP}}(\omega, 0)$ over different customer-sensitivity values λ.

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

3.7 Model calibration and validation

Now, we calibrate the coefficients of our approximate travel time model (*i.e.*, γ, δ and η) and validate its adjustment by comparing our expected cost estimate to the average operating cost of randomly simulated instances optimized with a detailed route planning model. We take the parameter values set in Table 3 and build an instance set I containing all tuples $(n, t_{\text{st}}, T_{\max}, x, y)$, where $n \in \{100, 200, 400\}$, $t_{\text{st}} \in \{1/60, 3.5/60, 10/60\}$, and $T_{\max} \in \{3, 5, 8\}$, representing various levels of request densities, stopping and parking times, and workday durations, respectively. Also, we consider the pair of decision variables $(x, y) \in \{(0, 0)\} \cup \{1, \ldots, 40\} \times \{0, 0.5, 3\}$, as it only makes sense to offer incentives when at least one CDP is installed. For each instance $i = (n, t_{\text{st}}, T_{\max}, x, y) \in I$, 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 $P^{\text{BLC}}_{\text{CDP}}(\omega, y)$. 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 running the PyVRP heuristic solver Wouda et al. (2024).

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 $t_{\mathrm{v}}(x, y)$ defined in Equation (10) via ordinary least squares regression, adjusting our model's output value $t_{\mathrm{v}}(x, y)$ to each instance's average travel time over 80 demand scenarios. The resulting calibration yields an adjusted R^2 coefficient equal to 99.6% and calibrated coefficient values $\gamma \approx 1.105$, $\delta \approx 1.043$ and $\eta \approx 0.752$, which is reasonable since γ and δ are close to 1 and η is similar to theoretical and empirical values discussed in the continuous approximation literature (Franceschetti et al., 2017).

The remaining 20 demand scenarios are used to validate our model's results. For the base case instance with $n = 200$, $t_{\mathrm{st}} = 3.5/60$ and $T_{\max} = 5$, Figure 6 plots our model's approximate expected cost $c(x, y)$ curve and compares it to the simulated sample average over the 20 demand scenarios as a function of x and the three different values of y. Empirically, we observe a good adjustment of our approximate cost model.

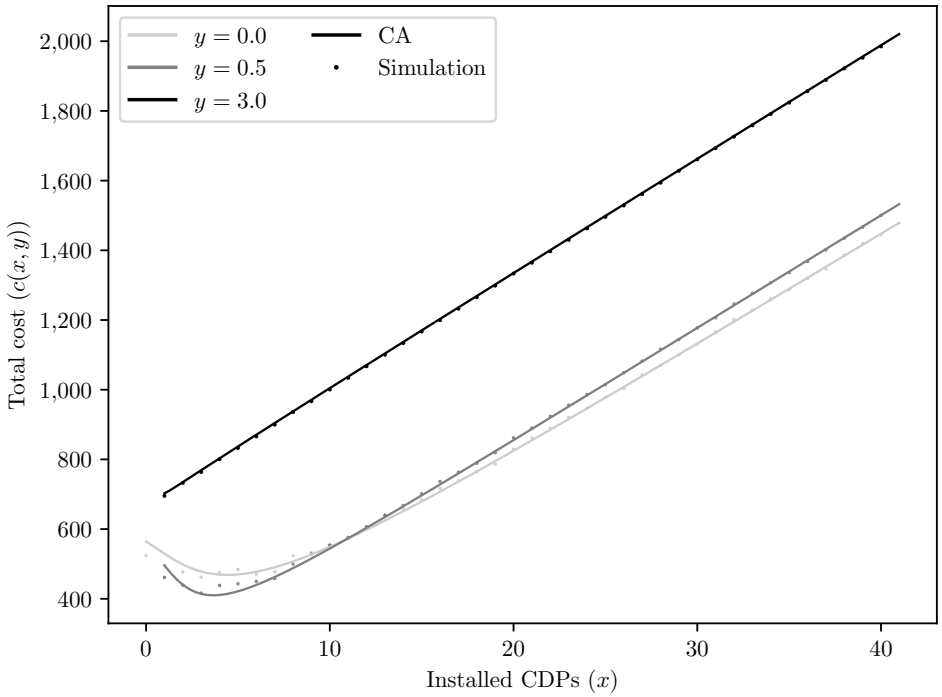
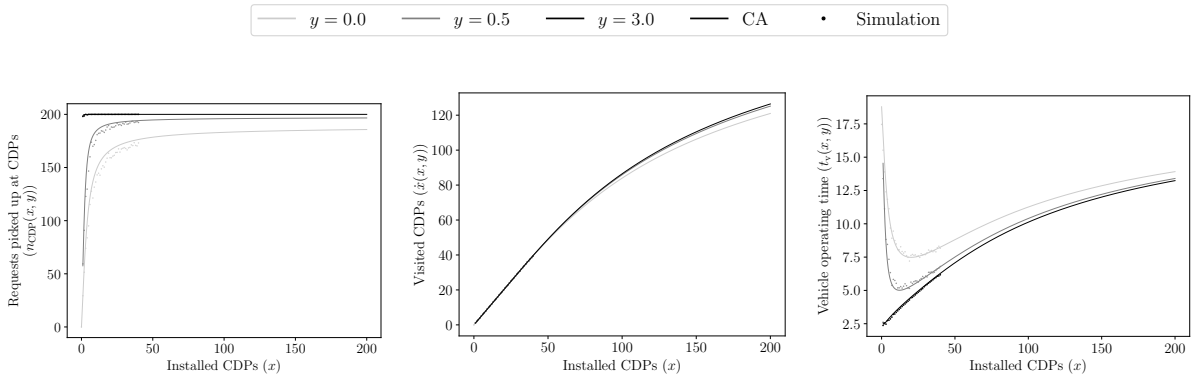


Figure 6: Approximate and simulated average total cost as a function of x and y given $n = 200, t_{\mathrm{st}} = 3.5/60$ and $T_{\max} = 5$.

Figure 7 explores further the quality of our proposed CA model and shows its estimate of the expected number of requests picked up at CDPs $n_{\mathrm{CDP}}(x, y)$, the expected number of visited CDPs $\dot{x}(x, y)$, and the total vehicle operating time $t_{\mathrm{v}}(x, y)$. Each of these values is compared to its corresponding analogous sample average value as a function of x and y. We also extrapolate the CA model values for $x \in [40, 200]$. Particularly, Figure 7a highlights how the use of customer incentives effectively increases the value of $n_{\mathrm{CDP}}(x, y)$. Figure 7b

illustrates the estimated and simulated values of $\dot{x}(x, y)$. When the number of installed CDPs grows up to $n = 200$, the number of visited CDPs does not exceed 127 and results in underutilized infrastructure. Finally, Figure 7c displays the corresponding values of $t_{\mathrm{v}}(x, y)$. We observe that relatively larger incentives reduce the required number of stops as more users pick up their parcels at CDPs. Therefore, the required operating time decreases.



(a) $n_{\mathrm{CDP}}(x, y)$ validation (b) $\dot{x}(x, y)$ validation (c) $t_{\mathrm{v}}(x, y)$ validation

Figure 7: Validation of the main components of the CA model as functions of x and y given $n = 200, t_{\mathrm{st}} = 3.5/60$ and $T_{\max} = 5$.

In Figure 8, we also compare the estimated standard deviation of the number of requests chosen to be picked up at a CDP in our model ($\sigma_{\mathrm{CDP}}(x, y)$) and its empirically simulated value over the 100 demand scenarios. Empirically, $\sigma_{\mathrm{CDP}}(x, y)$ fits and follows the same trend as the simulated value. As x increases, $\sigma_{\mathrm{CDP}}(x, y)$ attains a maximum and subsequently follows a non-increasing trend. This behavior is consistent with the intuition that installing additional facilities reduces the average customer distance to their closest CDP, thus leading to more homogeneous choices.

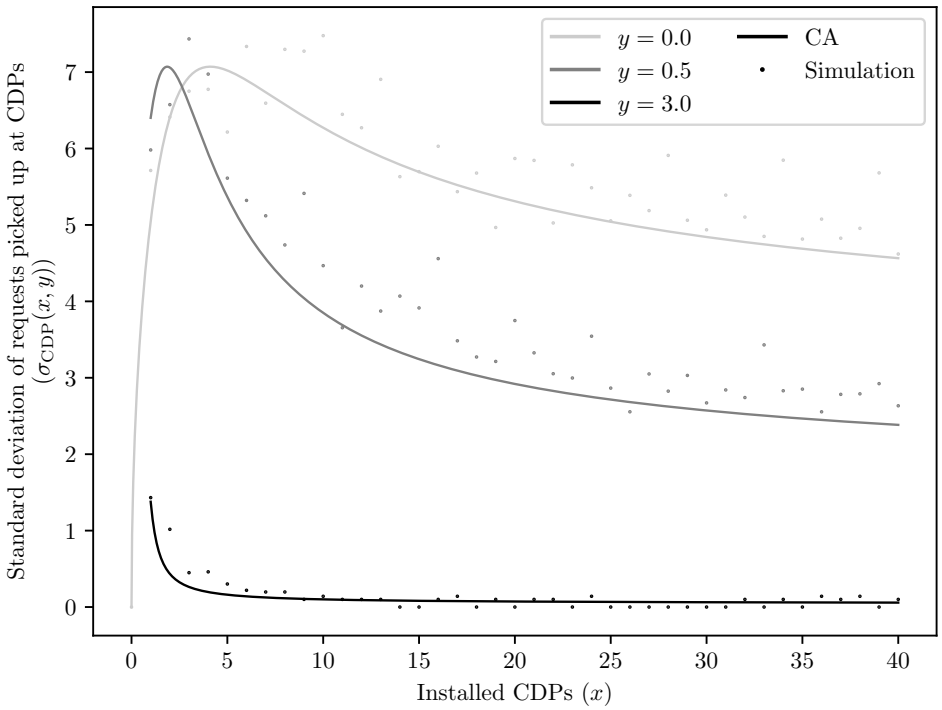


Figure 8: Approximate and simulated standard deviation of the number of requests chosen to be picked up at a CDP as a function of x and y given $n = 200, t_{\text{st}} = 3.5/60$ and $T_{\max} = 5$.

Table 4 presents the mean absolute percentage error (MAPE) for each component of our CA model compared to its empirically simulated counterpart. Specifically, the MAPE of $n_{\text{CDP}}(x, y)$ and $t_{\text{v}}(x, y)$ 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. The error of $\sigma_{\text{CDP}}(x, y)$ exceeds that of $n_{\text{CDP}}(x, y)$, as expected when estimating a standard deviation rather than a mean.

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

	$\boldsymbol{c(x,y)}$	$\boldsymbol{n_{\text{CDP}}(x,y)}$	$\boldsymbol{\dot{x}(x,y)}$	$\boldsymbol{t_{\text{v}}(x,y)}$	$\boldsymbol{\sigma_{\text{CDP}}(x,y)}$
MAPE	1.13%	2.89%	0.29%	4.03%	13.72%

4 Analysis of structural decisions and CA cost function

In this section, we analyze our approximate cost function $c(x, y)$ model and the structure of its optimal decisions x and y under varying parameter values representing a wide range of potential last-mile distribution settings. Our objective is to provide managerial insights for decision-makers.

4.1 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 strategy is the least-cost feasible option with objective $C^\star$ and considers that both decisions x and y are jointly optimized. 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.*, $\alpha_{\mathrm{x}} = 0$). As optimization principles dictates, we must have $C^{\mathrm{F}} \leq C^\star \leq \hat{C} \leq C^0$. Each solution of Model (16) 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 (Virtanen et al., 2020). Our procedure finishes in fractions of a second, so we do not report running times.

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 (16) s.t. $y = 0$
Minimum cost	$C^\star = c(x^\star, y^\star)$	$x^\star, y^\star$	Model (16)
Minimum cost with free CDPs	$C^{\mathrm{F}} = c(x^{\mathrm{F}}, y^{\mathrm{F}})$	$x^{\mathrm{F}}, y^{\mathrm{F}}$	Model (16) assuming $\alpha_{\mathrm{x}} = 0$

4.2 Base instance analysis

We begin studying results over the base instance parameters defined in Table 3. Figure 9 plots a heat map of function $c(x, y)$ with its level curves for $(x, y) \in [0, 6] \times [0, 2]$. Each color corresponds to a specific total cost, with values ranging between \$405.7/day to \$671.3/day. Also, Table 6 presents base instance results for each benchmark strategy. It includes total expected cost, percentage savings compared to C^0, all cost components (*i.e.*, infrastructure ($\alpha_{\mathrm{x}} \cdot x$), distribution ($\alpha_{\mathrm{v}} \cdot t_{\mathrm{v}}(x, y)$) and incentive ($y \cdot n_{\mathrm{CDP}}(x, y)$) 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 ($n_{\mathrm{CDP}}(x, y)/x$) and the total number of requests per installed CDP ($\nu(x) = n/x$).

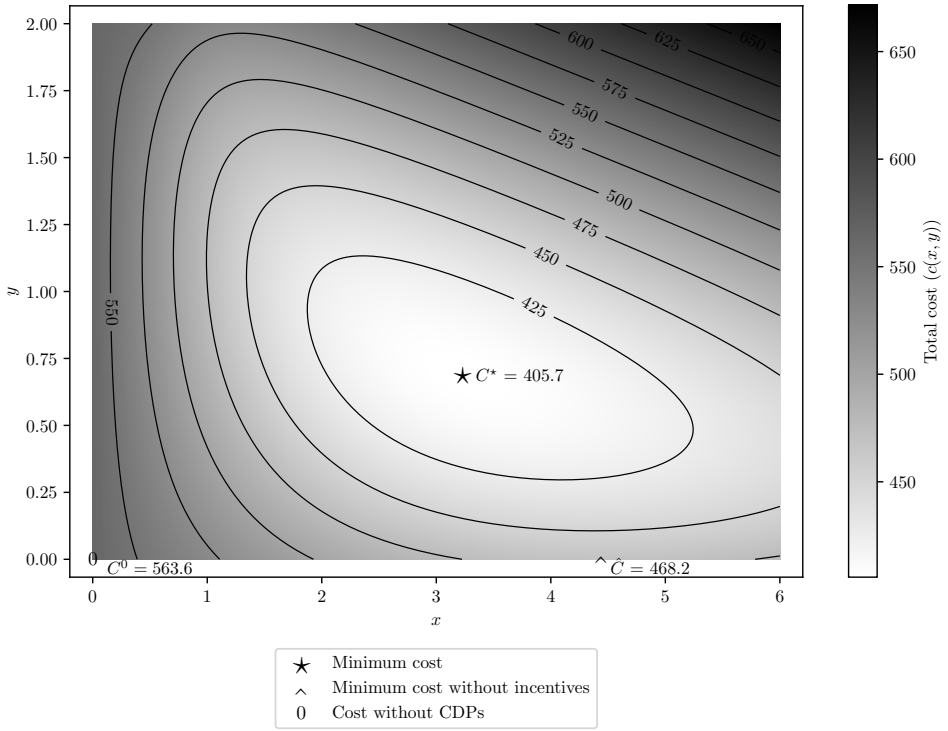


Figure 9: Heat map of $c(x, y)$

Table 6: Benchmark results

	Benchmark			
	Cost without CDPs	**Minimum cost without incentive**	**Minimum cost**	**Minimum cost with free CDPs**
Cost	$C^0 = 563.6$	$\hat{C} = 468.2$	$C^\star = 405.7$	$C^{\mathrm{F}} = 222.4$
% 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	17.83
y	0	0	0.69	0.11
$m(x, y)$	200.0	99.8	43.9	45.8
$n_{\mathrm{CDP}}(x, y)/x$	-	23.6	49.1	9.7
$\nu(x)$	-	45.1	61.5	11.2

As observed, the system's cost without CDPs is C^0 = \$563.6/day purely based on the distribution component. If we install a cost-efficient CDP network without incentives, then we cut costs by 16.9% to $\hat{C}$ = \$468.2/day and require $\hat{x} = 4.43$ CDPs; this roughly yields $\nu(\hat{x}) = 45.1$ 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/day. As observed in Figure 9 for a fixed x value, we reduce costs if we begin increasing y starting from \$0/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 $C^\star$ = \$405.7/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. Compared to $\hat{x}$, the optimal number of CDPs reduces by 26.6% to $x^\star = 3.25$ (roughly $\nu(x^\star) = 61.5$ requests per CDP), but it requires setting a customer incentive of $y^\star$ = \$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 208.1% 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 $x^{\mathrm{F}} = 17.83$. Even if CDP stations are free, one should not install too many of them because an over-sized network leads to an increased number of vehicle service stops. This is particularly interesting for municipalities or government agencies aiming to reduce emissions from urban logistics operations, which typically correlate with distribution costs. Subsidizing the use of CDPs could be a potential approach to achieving this goal.

As follows, we perform sensitivity analyses over key problem parameters to asses their impact and gain insight. In particular, we focus on the impact of request density, customer elasticity, and unit cost per installed CDP on optimal costs and structural solutions.

4.3 Cost sensitivity as a function of request density

We now investigate the system's performance as a function of request density $\rho = n/A$. Figure 10 plots the expected cost per request, the number of requests per CDP ($\nu(x) = n/x$) 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 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 CDPs per request increases as a function of ρ, which reciprocally indicates that fewer CDPs are needed to cover the same number of requests. We also observe 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 ($C^\star/n$) produces the largest percentage cost savings when $\rho = 10.0$ requests

per square kilometer. Also, the optimal value of incentive decreases as a function of ρ and converges to zero when $\rho \to \infty$; the largest incentive offered to customers is $y^\star = \$1.5$ per request picked up at a CDP and occurs when $\rho = 1.4$ 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.

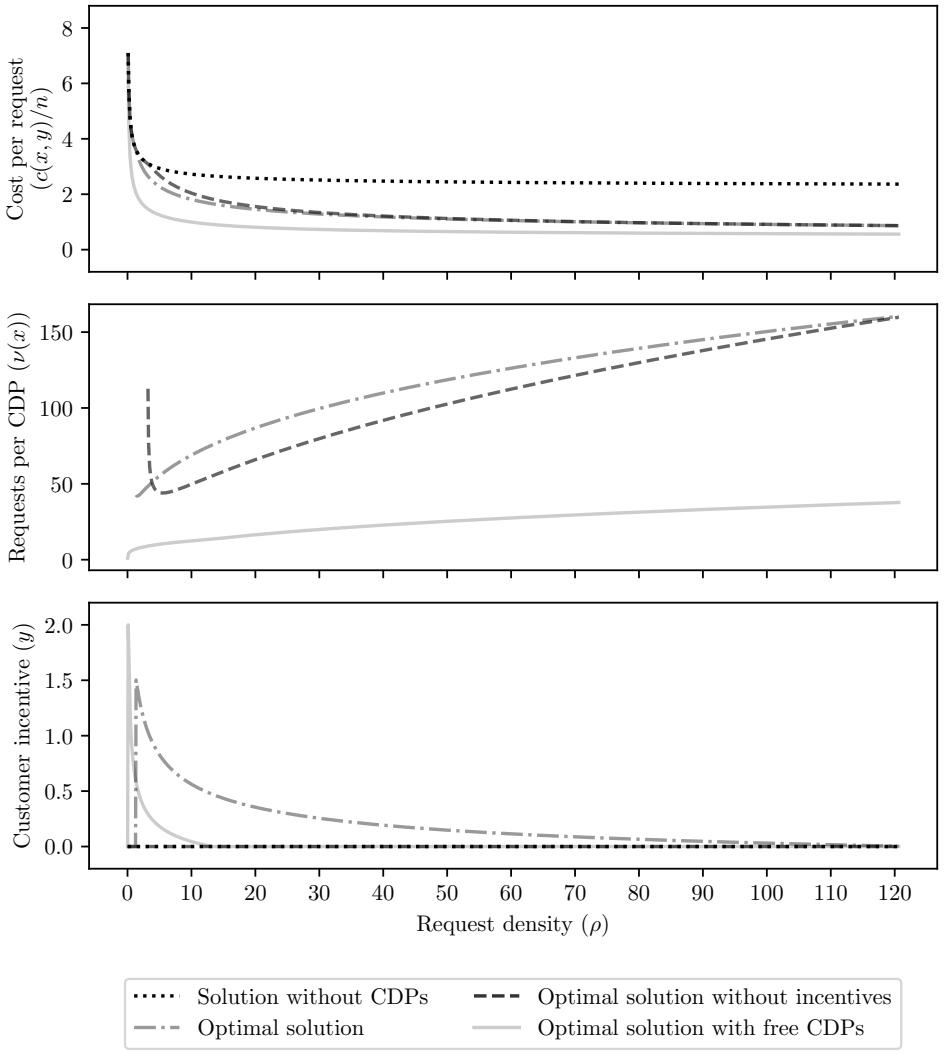


Figure 10: Benchmark strategies as a function of request density ($\rho = n/A$).

4.4 Cost sensitivity as a function of customer-to-distance sensitivity

Figure 11 plots the optimal number of requests per CDP ($\nu(x)$) and customer incentive (y) for the minimum cost benchmark as a function of λ (*i.e.*, the customer-to-distance sensitivity) assuming 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 in our approximate cost model Equation (11) equipped with the more realistic 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 realistic 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 ($\lambda \to 0$). 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 $\lambda \to \infty$ as suggested in Property 2.

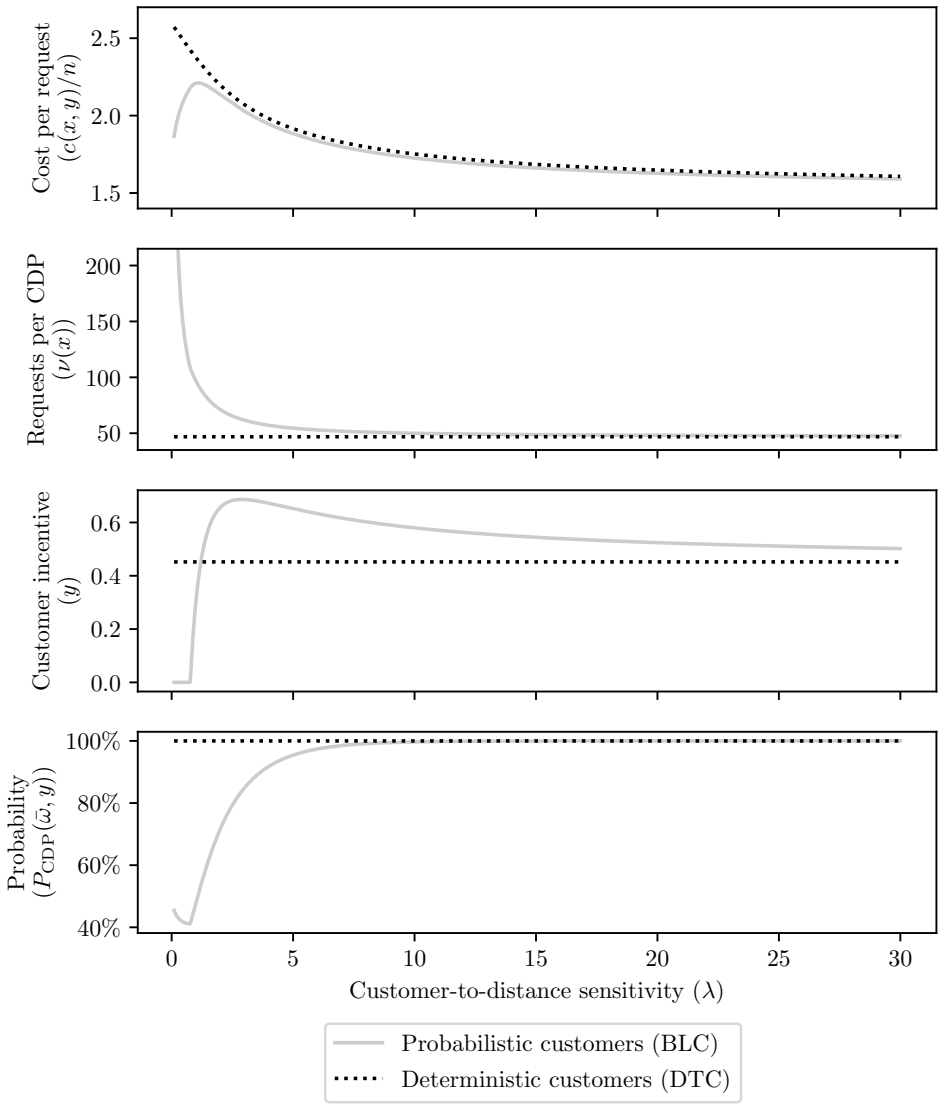


Figure 11: Optimal strategy for the BLC and DTC choice models as a function of customer-to-distance sensitivity (λ).

4.5 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 12 plots the expected cost per request, the chosen numbers of requests per CDP ($\nu(x)$), 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 ($C^\star/n$ 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 $C^{\text{F}}/n \approx \$1.1$/request when $\alpha_{\text{x}} = 0$, and align with $C^0/n \approx \$2.8$/request for a high-enough α_{x} value that completely discourages 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 $C^\star/n$ and $\hat{C}/n$ increase sub-linearly as functions of α_x. This occurs because the optimal solutions adjust increasing $y^\star$, $\nu(x^\star)$, and $\nu(\hat{x})$ (*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 ($C^\star$ and $\hat{C}$ compared to C^0) when using CDPs with and without customer incentives.

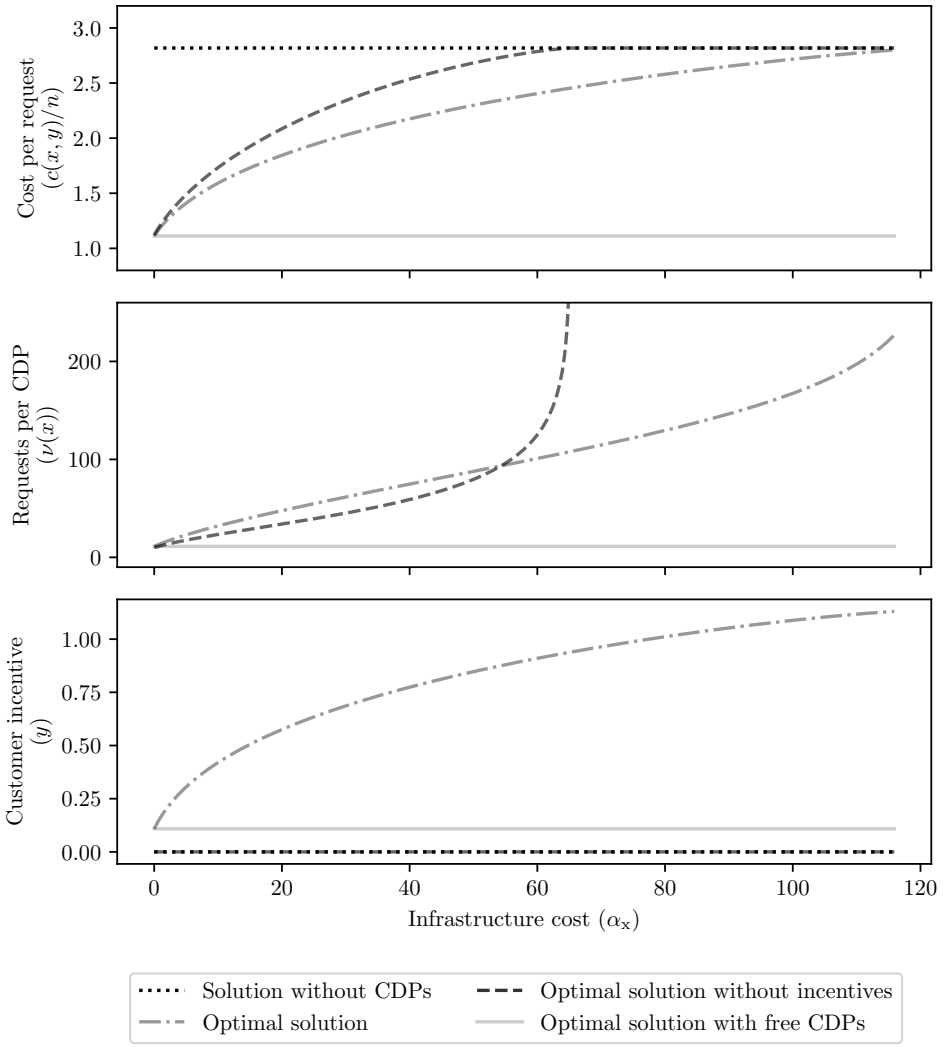


Figure 12: Benchmark strategies as a function of infrastructure cost (α_x).

4.6 Cost sensitivity as a function of the travel time between the distribution center and the service region

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 13 plots the estimated minimum expected cost per request, and the optimal number of requests per CDP ($\nu(x)$) 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 $2 \cdot t_0 \to T_{\max}$. 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 ($\nu(\hat{x})$) as t_0 increases. When economic incentives are allowed, the outcome changes. The optimal number of requests per CDP ($\nu(x^\star)$) decreases until a given t_0 value (in our case, roughly 1.2 hours), after which it increases. In parallel, optimal incentives $y^\star$ 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 ($\nu(x^{\mathrm{F}})$) as 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.

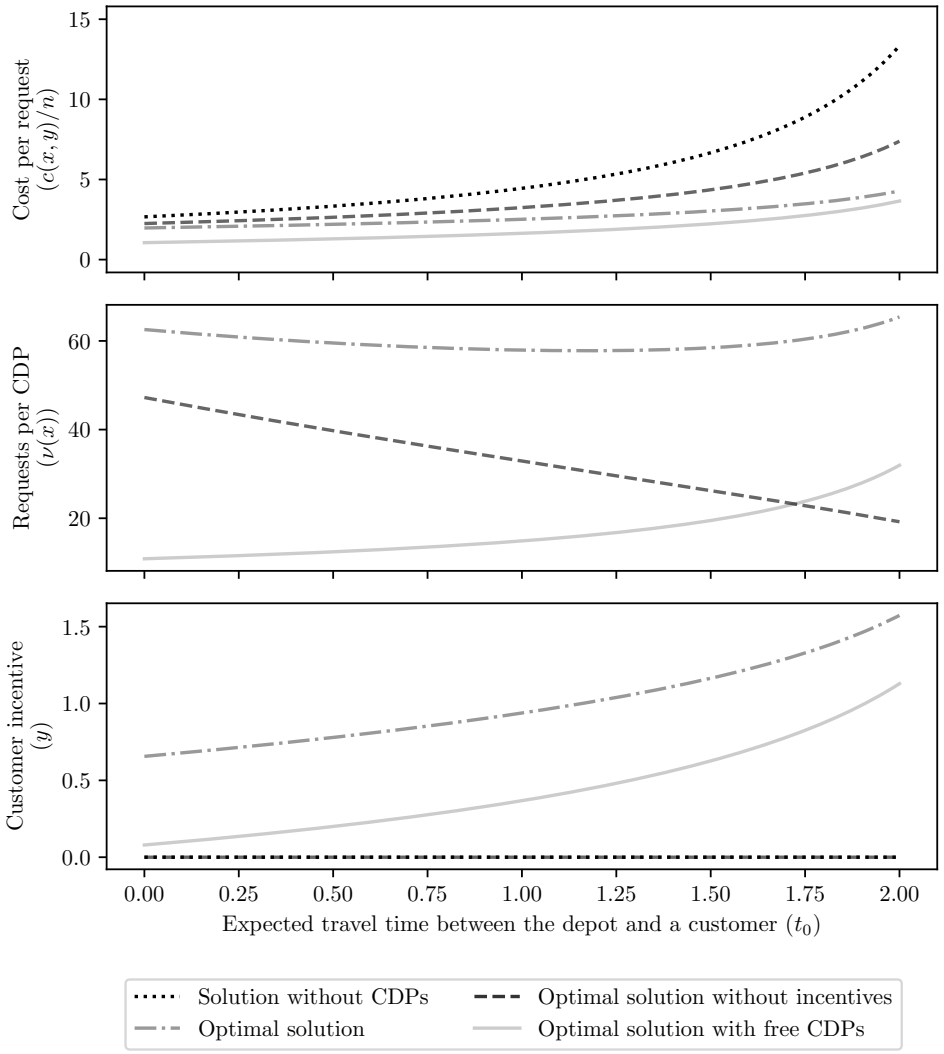


Figure 13: Benchmark strategies as a function of the expected travel time between the distribution center and a customer t_0.

4.7 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 relax this assumption and solve the four benchmarks defined in Table 5 subject to chance constraint (17) imposing that the average number of orders handled per CDP exceeds capacity q with probability less than or equal to $\zeta = 5\%$. Specifically, we equip model (16) with the normal chance constraint approximation defined in (24), as it yields tighter approximations. We also assume a concave increasing cost per CDP $\alpha_{\mathrm{x}} = 6.708\sqrt{q}$ as a function of capacity q, which is consistent to the presence of scale economies in larger facilities.

Figure 14 plots the estimated minimum expected cost per request, the optimal number of requests per CDP ($\nu(x)$) 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 ($x^{\mathrm{F}}, y^{\mathrm{F}}$) converge to a value once the capacity constraint is no longer binding. When a cost per CDP is paid ($\alpha_x(q) > 0$), then the minimum cost per request $C^\star/n$ follows a U-shape with a clear minimum value as a function of q. For a relatively small value of q, CDP capacity is a limiting resource that limits decisions and increases cost. As q grows, the 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. As expected, the optimal number of CDPs $x^\star$ decreases (and the number of orders per CDP $\nu(x^\star)$ 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 $y^\star$.

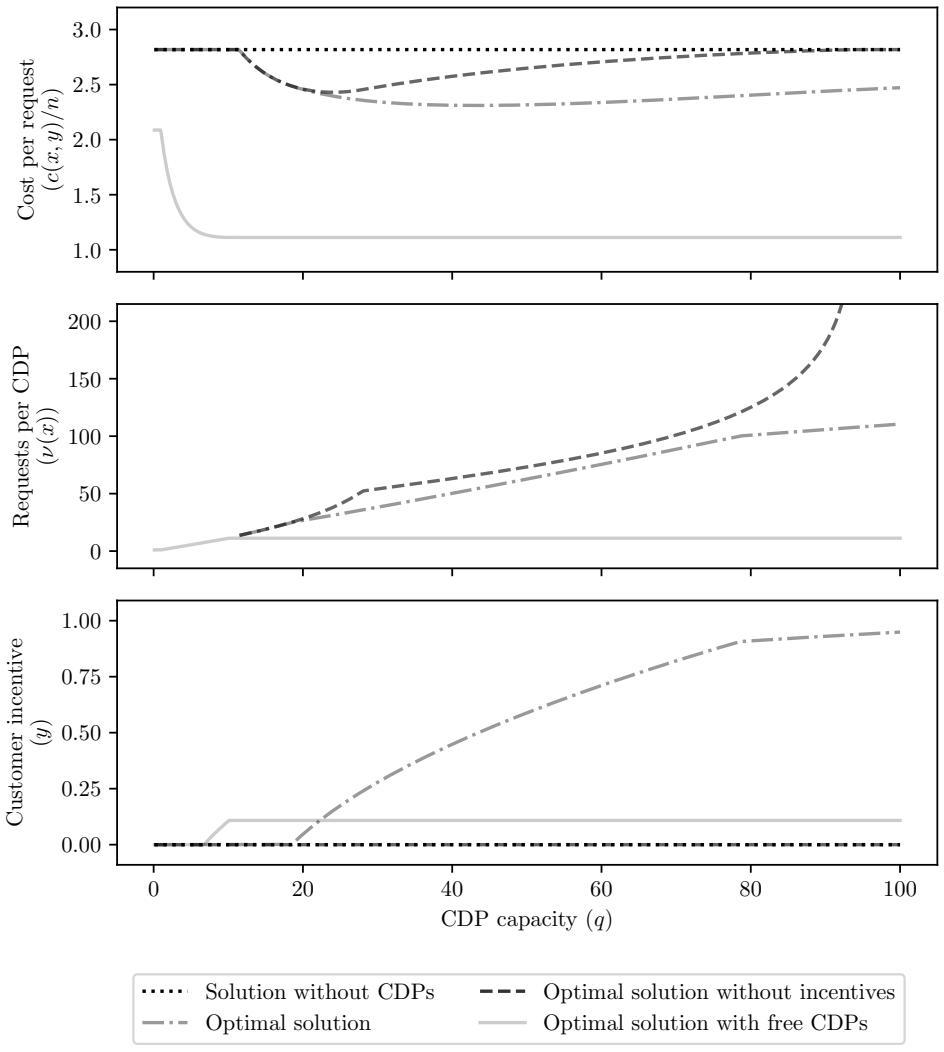


Figure 14: Benchmark strategies as a function of CDP capacity q.

5 Case study

In this section, we test the effectiveness of our approximate cost model through a case study on a real road network with a geographic request distribution inspired by the metropolitan area of Santiago, Chile. This raises two additional difficulties for our CA model: in reality, customers are not uniformly distributed over the service area, and road network distances need not be metric. With this, we aim to demonstrate the practical adaptability and effectiveness of our model, even when some of its assumptions are violated.

The road network topology and arc distance data are obtained from OpenStreetMaps (OSM) via the OSMnx python package (Boeing, 2025). We use our CA optimization model (16) equipped with chance constraint (24) and adapt its data to the case study. Then, we solve it to determine both structural decisions, *i.e.*, the number of CDPs x to install and the economic incentive y offered to each customer. Next, we heuristically choose a discrete set of geographic locations for the x CDPs, running a k-means clustering algorithm. For the chosen x CDP locations, we simulate a set of demand scenarios and compute the corresponding last-mile distribution routes and costs. The expected cost is then approximated by the sample average cost over all scenarios, which we compare to our CA model's cost estimate.

In the case study, we consider an average day of a logistics operator serving $n = 500$ 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 14.3 km west of the region's centroid, as depicted in Figure 15; this yields a request density $\rho \approx 10.86$ requests/km^2. The Euclidean distance metric assumed in our CA model is invalid for real road networks, as straight line distances are shorter than shortest path network distances. In this context, it is likely that our CA model will underestimate vehicle travel time. We fix the distance-related issue by setting a circuity factor $\mu \geq 1$ (see Merchán et al. (2020)), which assumes a constant ratio between the shortest path network distance and the Euclidean distance for each pair of origin-destination points. The μ factor is used in the CA model to scale its service area as $A \approx \mu^2 \cdot A'$, where A' is the real-world area used to estimate the CA model's area A. We calibrate the μ value by randomly sampling 2,000 origin-destination pairs within the service region, and then computing their shortest path network to Euclidean distance ratio in OSM, yielding an average $\mu \approx 1.308$. We also differentiate vehicle speeds, assuming 30 km/h for stem distances and 15 km/h within the service region; this sets $t_0 = 14.34/30 + 2 \cdot R/(3 \cdot s) \approx 0.7$ hours. Furthermore, we use the calibrated values $\gamma \approx 1.105$, $\delta \approx 1.043$, and $\eta \approx 0.752$ obtained in Section 3.7 to test the robustness of our travel time model. All remaining parameters are obtained from base case instance values in Table 3.

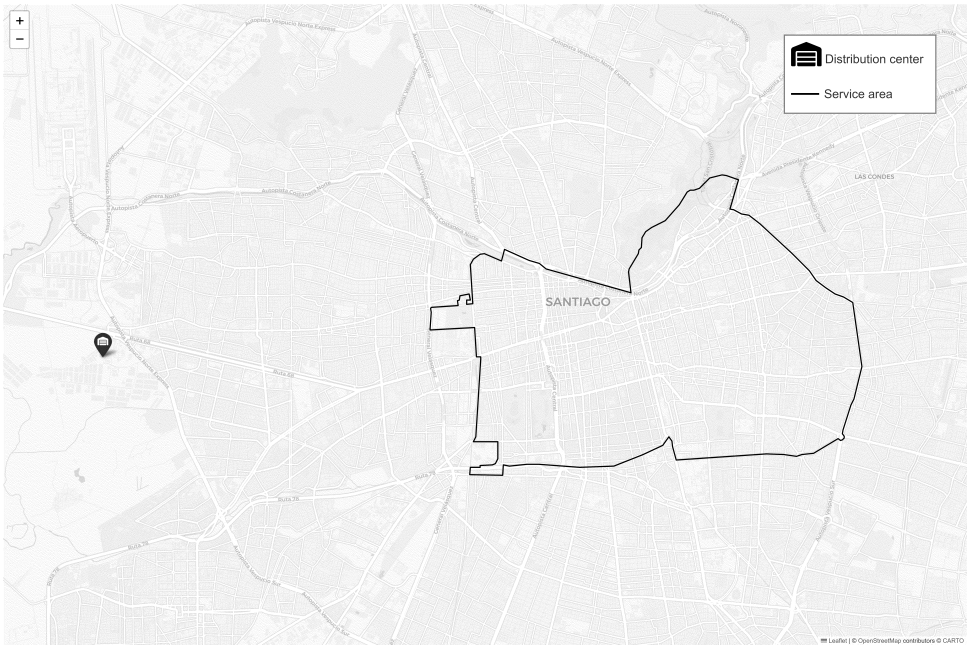


Figure 15: Geographic location of the distribution center and service region for our case study in Santiago, Chile.

To capture the geographic heterogeneity of customer locations, we use data from the 2017 Chilean census (INE, 2017) and assign each city block a request probability proportional to its census population, as illustrated in Figure 16.

Figure 16: Population heatmap for the service region defined by the municipalities of Santiago, Providencia, and Ñuñoa, Chile (INE, 2017).

Nonetheless, we solve our CA model as if customer requests were uniformly located within the service area and obtain a pair of structural decisions (x, y). The location of the set of x CDPs is determined by solving an auxiliary k-means model (with $k = x$), where each

data point of the instance is a city block centroid weighted by its population; the k-means solution is obtained via the scikit-learn library (Pedregosa et al., 2011). Then, each CDP is located at the node in the OSM graph closest to the corresponding cluster centroid from the k-means solution. We likewise assign the distribution center to the OSM graph node closest to its coordinates. Figure 17 illustrates an example that locates 11 CDPs within the service region; as expected, this location prioritizes populated areas. More sophisticated methods for discretizing solutions obtained via CA models can be found in Banerjee et al. (2022), Carlsson (2012), Carlsson and Jia (2015), Carlsson et al. (2024), Galvão et al. (2006), and Ouyang and Daganzo (2006).

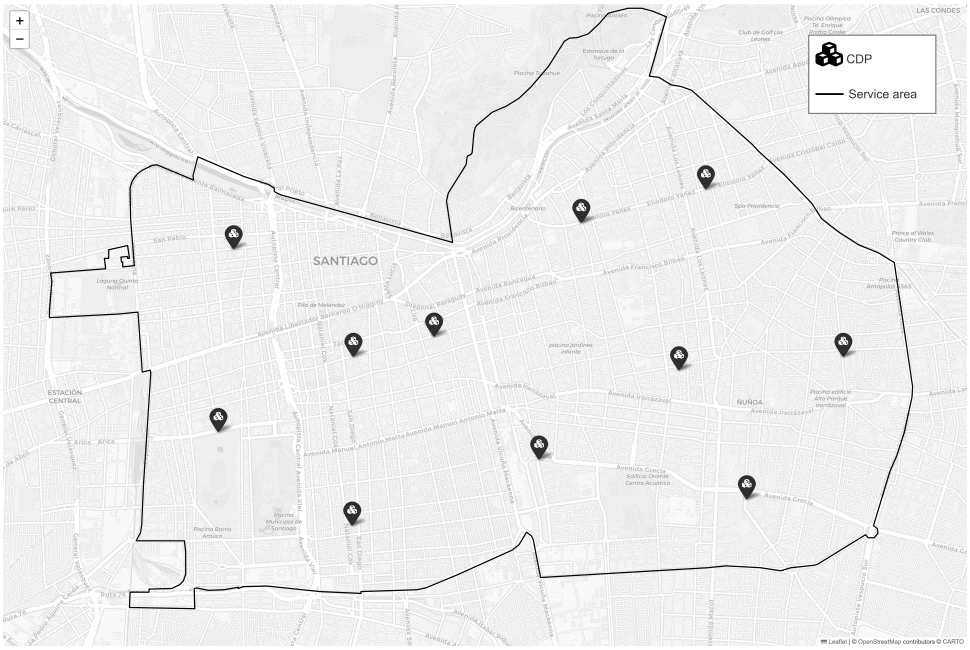


Figure 17: Example of a location of 11 CDPs in our case study.

Afterwards, we solve our CA model and test our cost estimate and structural decisions (x, y) via simulation of detailed demand scenarios following an analogous process as Section 3.7. Specifically, we generate a set Ξ of 20 demand scenarios, each consisting of 500 requests geographically located according to the spatial demand distribution in Figure 16, which specifies the probability of a request arising in each city block. Once requests are generated at each city block, their specific locations are uniformly drawn within that block. For each customer request $k \in \{1, \ldots, 500\}$ in each demand scenario $\xi \in \Xi$, we simulate its service choice using model $P^{\text{BLC}}_{\text{CDP}}(\Omega_k, y)$, where Ω_k denotes the shortest path walking distance to the nearest CDP. Furthermore, we compute each scenario's last-mile distribution cost via the PyVRP heuristic solver in a graph having shortest path driving times as arc travel times. Appendix A.4 presents more details of our CA model's validation over this road network; our cost estimate reports a MAPE of 2.09%.

Table 7 presents the three benchmarks tested for this case study, each studied with capacities $q \in \{30, 40, 50, 60\}$ requests per CDP. These include the solution to the unconstrained CA model, the CA model subject to the expected value constraint (18), and the CA model subject to chance constraint (24) with $\zeta = 5\%$.

Table 7: Case study's benchmarks

Benchmark	**Cost**	**Solution**	**Model solved**
Unconstrained CA model	$C^\star = c(x^\star, y^\star)$	$x^\star, y^\star$	Model (16)
CA model with expected constraint	$C^\star_{\mathbb{E}} = c(x^\star_{\mathbb{E}}, y^\star_{\mathbb{E}})$	$x^\star_{\mathbb{E}}, y^\star_{\mathbb{E}}$	Model (16) s.t. (18)
CA model with chance constraint	$C^\star_{\mathbb{P}} = c(x^\star_{\mathbb{P}}, y^\star_{\mathbb{P}})$	$x^\star_{\mathbb{P}}, y^\star_{\mathbb{P}}$	Model (16) s.t. (24) with $\zeta = 5\%$

We present our results in Table 8, which reports the CA estimates for costs and structural decisions, together with the sampled costs of detailed last-mile solutions over all demand scenarios, for all three benchmarks and CDP capacities tested. Infrastructure costs per CDP are set to $\alpha_{\mathrm{x}} = 6.708\sqrt{q}$, as in Section 4.7. Our first observation is that, in each case, the cost estimated by our CA model falls within the 95% confidence interval of the detailed cost derived from simulation. This shows that, in our case study, the continuous approximation model estimates the last-mile distribution cost of a real road network equipped with CDPs, even when demand is spatially distributed according to the city's population. As expected, we have $C^\star \leq C^\star_{\mathbb{E}} \leq C^\star_{\mathbb{P}}$ for all values of q. Similarly, the optimal number of CDPs follows the same ordering, $x^\star \leq x^\star_{\mathbb{E}} \leq x^\star_{\mathbb{P}}$, since a more restrictive capacity requires more CDPs. In contrast, the optimal economic incentive moves in the opposite direction, *i.e.*, $y^\star_{\mathbb{P}} \leq y^\star_{\mathbb{E}} \leq y^\star$, probably 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 number of CDPs decreases as q increases, since larger capacities augment infrastructure costs. However, in this case, the simulated average demand per CDP ($n^{\Xi}_{\mathrm{CDP}}(x, y)/x$) exceeds q. The same occurs with the sample average failure probability $p^{\Xi}(x, y)$, calculated as the proportion of demand scenarios $\xi \in \Xi$ in which the number of requests attended per CDP surpasses the available capacity q. If we impose the expected value constraint, then $n^{\Xi}_{\mathrm{CDP}}(x, y)/x$ lies below q, but the average failure probability $p^{\Xi}(x, y)$ still exceeds by far the $\zeta = 5\%$ probability allowed for the chance constrained model. In contrast, when the approximate chance constraint is imposed, $p^{\Xi}(x, y)$ falls below the admissible failure probability. For constrained problems, total cost first decreases when CDP capacity is binding, but as q increases, the marginal value of capacity declines and infrastructure costs outweigh the benefit, causing total cost to rise. Specifically, under the chance constraint, the minimum cost is $C^\star_{\mathbb{P}} = 1,344.7$ at a capacity of $q = 40$ requests per CDP, suggesting that mid-sized CDPs are an attractive option for this case study.

Table 8: Case study results

	q	30	40	50	60
	α_x	**36.7**	**42.4**	**47.4**	**52.0**
Unconstrained CA model					
CA	$C^\star$	1,251.9	1,294.3	1,328.9	1,358.1
	$x^\star$	7.82	7.15	6.66	6.28
	$y^\star$	0.95	1.00	1.05	1.08
Sim.	$c^\Xi(x,y)$	$1{,}245.5 \pm 14.9$	$1{,}305.9 \pm 14.3$	$1{,}341.6 \pm 13.0$	$1{,}372.0 \pm 18.2$
	$n^\Xi_{\text{CDP}}(x,y)/x$	51.9 ± 0.5	57.1 ± 0.4	58.1 ± 0.4	56.2 ± 0.6
	$p^\Xi(x,y)$	100.0%	100.0%	100.0%	100.0%
CA model with expected constraint					
CA	$C^\star$	1,349.1	1,337.4	1,344.8	1,361.7
	$x^\star$	13.85	10.48	8.42	7.02
	$y^\star$	0.52	0.70	0.85	0.99
Sim.	$c^\Xi(x,y)$	$1{,}357.0 \pm 21.1$	$1{,}329.3 \pm 20.4$	$1{,}330.8 \pm 19.6$	$1{,}377 \pm 15.7$
	$n^\Xi_{\text{CDP}}(x,y)/x$	28.8 ± 0.3	40.0 ± 0.5	50.0 ± 0.7	56.7 ± 0.6
	$p^\Xi(x,y)$	10%	40%	35%	0%
CA model with chance constraint ($\zeta = 5\%$)					
CA	$C^\star$	1,360.4	1,344.7	1,349.3	1,364.0
	x	14.38	10.85	8.70	7.24
	y	0.51	0.68	0.83	0.96
Sim.	$c^\Xi(x,y)$	$1{,}361.7 \pm 16.5$	$1{,}331.9 \pm 16.4$	$1{,}346.9 \pm 22.7$	$1{,}369.3 \pm 17.8$
	$n^\Xi_{\text{CDP}}(x,y)/x$	28.5 ± 0.2	37.4 ± 0.3	45.7 ± 0.6	56.3 ± 0.5
	$p^\Xi(x,y)$	0%	0%	0%	0%

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: (1) the number of CDPs to install and (2) the incentive offered to customers.

In our model, the expected cost per request depends only on the CDP network size, the request density, and the average number of requests per CDP, rather than on the total number of requests or the service area itself. This finding is consistent with the scale-independent distance per request described by the BHH formula (Beardwood et al., 1959).

Empirically, we also 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. Moreover, the use of CDPs enhances the economies of density observed in most last-mile distribution systems. As observed, the installation of CDPs is more effective in reducing costs in areas with a relatively higher number of requests per square kilometer. Conversely, the use of incentives to customers is most valuable when request density is relatively low. 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 utilization when demand is sparse without increasing infrastructure costs. Furthermore, if the CDP infrastructure is provided and operated for free (*e.g.*, by the city authorities), the distribution costs are cut down by 60.5%, which might be a good policy to significantly reduce emissions and other externalities produced by last-mile distribution operations. Even if CDPs are provided for free, installing these in excess is inefficient, as an oversized network leads to higher routing costs. In this case, customer incentives become less necessary, although these could be useful in sparse demand conditions.

We evaluate how important it is to have customers' behavior information in advance when planning structural decisions. Specifically, decisions should adapt to customers' sensitivity to the CDP distance, as errors from assuming a deterministic customer choice rather than a stochastic one are greater when customer sensitivity tends to zero.

We also observe that the system's cost is bounded between the benchmark cost that only delivers to customers' homes, and the optimistic case with free CDPs. Additionally, cost increases sublinearly as a function of the cost per CDP, as the system adapts by reducing the number of installed CDPs and increasing incentives.

Moreover, as the distribution center is located farther from the service region, costs rise because stem travel consumes valuable time. To mitigate this, operators without incentives must install more CDPs, raising infrastructure costs, while operators with incentives can consolidate demand at fewer CDPs by offering higher incentives to customers.

Cost trade-offs change when capacity constraints on CDP usage are introduced, as additional CDPs must be installed or incentives must be used to serve the same number of requests. The resulting cost behavior is non-monotonic as a function of capacity, as total cost first decreases as larger CDP capacities reduce congestion, but eventually rises again as infrastructure costs exceed the marginally less valuable additional capacity investments.

Our case 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 paper 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 precise information regarding customer-specific aspects, we could generalize our model to heterogeneous customers and assume a request density or customer CDP choice probability depending on the spatial location. We envision that these distributions could be estimated based on historical demand data, depending on geographical location. In such a setting, we could recommend customer incentives based on location or varying CDP density. One could work with multiple available CDP facilities in each customer's choice set (not only the closest one), as done in Lin et al. (2020, 2022), Akkerman et al. (2025), and Lyu and Teo (2022). 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 time-varying aspects and explore the dynamic version of this problem, incorporating time-varying customer incentives and the dynamic activation of CDPs. Future research could also 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.

Acknowledgements

Benjamín Rojas' work is supported by the Chilean National Agency for Research and Development (ANID) under PhD grant 21230934.

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*, 59:250–278.

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.

Archetti, C. and Bertazzi, L. (2021). Recent challenges in routing and inventory routing: E-commerce and last-mile delivery. *Networks*, 77:255–268.

Arslan, O. (2021). The location-or-routing problem. *Transportation Research Part B: Methodological*, 147:1–21.

Banerjee, D., Erera, A. L., and Toriello, A. (2022). Fleet sizing and service region partitioning for same-day delivery systems. *Transportation Science*, 56:1327–1347.

Banerjee, D., Erera, A. L., and Toriello, A. (2025). Pricing and demand management for integrated same-day and next-day delivery systems. *Transportation Science*, 59:279–300.

Baymard Institute (2026). 50 cart abandonment rate statistics 2026. Viewed 20 of February 2026, `https://baymard.com/lists/cart-abandonment-rate`.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55:299–327.

Block, H. D. (1974). Random orderings and stochastic theories of responses (1960). In *Economic Information, Decision, and Prediction. Theory and Decision Library*, volume 7. Springer, Dordrecht.

Boeing, G. (2025). Modeling and analyzing urban networks and amenities with osmnx. *Geographical Analysis*.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Boysen, N., Fedtke, S., and Schwerdfeger, S. (2021). Last-mile delivery concepts: a survey from an operational research perspective. *OR Spectrum*, 43:1–58.

Carlsson, J. G. (2012). Dividing a territory among several vehicles. *INFORMS Journal on Computing*, 24:565–577.

Carlsson, J. G. and Jia, F. (2015). Continuous facility location with backbone network costs. *Transportation Science*, 49:433–451.

Carlsson, J. G., Liu, S., Salari, N., and Yu, H. (2024). Provably good region partitioning for on-time last-mile delivery. *Operations Research*, 72:91–109.

Cerulli, M., Archetti, C., Fernández, E., and Ljubić, I. (2024). A bilevel approach for compensation and routing decisions in last-mile delivery. *Transportation Science*, 58:1076–1100.

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*, 18:331–350.

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*, 56:251–261.

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*, 72:181–217.

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*, 25:413–433.

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*, 58:454–473.

Galvão, L. C., Novaes, A. G., de Cursi, J. S., and Souza, J. C. (2006). A multiplicatively-weighted voronoi diagram approach to logistics districting. *Computers & Operations Research*, 33:93–114.

Giuliano, G. (2023). The challenges of freight transport in cities. In Edoardo Marcucci, V. G. and Pira, M. L., editors, *Handbook on City Logistics and Urban Freight*, pages 11–34. Edward Elgar Publishing.

Hazbún, C. (2019). Analysis of the impact of using smart lockers on home delivery. Master's thesis, Pontificia Universidad Católica de Chile, Santiago, Chile.

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. *SSRN Electronic Journal*.

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*, 46:241–294.

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.

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*, 53:236–255.

Laporte, G., Nickel, S., and da Gama, F. S., editors (2019). *Location Science*. Springer International Publishing.

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, X. and Ouyang, Y. (2010). A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Research Part B: Methodological*, 44:535–548.

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.

Lyu, G. and Teo, C.-P. (2022). Last mile innovation: The case of the locker alliance network. *Manufacturing & Service Operations Management*, 24:2425–2443.

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*, 25:346–360.

Ouyang, Y. and Daganzo, C. F. (2006). Discretization and validation of the continuum approximation scheme for terminal system design. *Transportation Science*, 40:89–98.

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*, 10:542–557.

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.

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.

Tilk, C., Olkis, K., and Irnich, S. (2021). The last-mile vehicle routing problem with delivery options. *OR Spectrum*, 43:877–904.

Toth, P. and Vigo, D., editors (2014). *Vehicle Routing*. Society for Industrial and Applied Mathematics.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wang, Q., Lyu, G., He, L., and Teo, C. P. (2025). Does the locker alliance network improve last mile delivery efficiency? *Management Science*, 0(0).

Wang, X., Arslan, O., and Delage, E. (2024). Crowdkeeping in last-mile delivery. *Transportation Science*, 58:474–498.

World Economic Forum (2020). The future of the last-mile ecosystem. Viewed 20 of February 2026, `https://www3.weforum.org/docs/WEF_Future_of_the_last_mile_ecosystem.pdf`.

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 Pierre Cartier, Pierre Moussa, B. J. 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, 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.

Zoting, S. and Shivarkar, A. (2026). Last mile delivery transportation market size, share and trends 2026 to 2035. Viewed 20 of February 2026, `https://www.precedenceresearch.com/last-mile-delivery-transportation-market`.

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 Omitted proofs

Appendix A.1 Proof of Property 1

We require to show that the expected cost formula divided by n solely depends on n, A, and x through the request density $\rho = n/A$ and the number of requests per CDP $\nu(x) = n/x$. We begin dividing the total cost by n and obtain

$$\frac{c(x,y)}{n} \approx \frac{\alpha_{\mathrm{x}}}{\nu(x)} + \alpha_{\mathrm{v}} \cdot \frac{t_{\mathrm{v}}(x,y)}{n} + y \cdot \frac{n_{\mathrm{CDP}}(x,y)}{n}. \tag{25}$$

So, we need to work on $n_{\mathrm{CDP}}(x,y)/n$ and $t_{\mathrm{v}}(x,y)/n$. The first term can be rewritten as

$$\frac{n_{\mathrm{CDP}}(x,y)}{n} \approx \begin{cases} \frac{\rho}{\nu(x)} \cdot \int_0^{2\pi} \int_0^{\sqrt{\frac{\nu(x)}{\pi \cdot \rho}}} \omega \cdot P_{\mathrm{CDP}}(\omega, y)\, d\omega\, d\theta, & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{26}$$

which satisfies the desired property. Also, the travel time per request can be written as

$$\frac{t_{\mathrm{v}}(x,y)}{n} \approx \gamma \cdot \kappa \cdot t_{\mathrm{se}} + \delta \cdot \kappa \cdot t_{\mathrm{st}} \cdot \frac{m(x,y)}{n} + \eta \cdot \frac{\kappa}{s} \cdot \sqrt{\frac{1}{\rho} \cdot \frac{m(x,y)}{n}}, \tag{27}$$

where the expected number of stops per request can be expressed as

$$\frac{m(x,y)}{n} \approx \frac{\dot{x}(x,y)}{n} + 1 - \frac{n_{\mathrm{CDP}}(x,y)}{n}. \tag{28}$$

The number of visited CDPs per request is

$$\frac{\dot{x}(x,y)}{n} \approx \begin{cases} \frac{1}{\nu(x)} \cdot \left(1 - \mathrm{e}^{-\frac{n_{\mathrm{CDP}}(x,y)}{x}}\right), & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{29}$$

and $n_{\mathrm{CDP}}(x,y)/x$ is defined as

$$\frac{n_{\mathrm{CDP}}(x,y)}{x} \approx \begin{cases} \rho \cdot \int_0^{2\pi} \int_0^{\sqrt{\frac{\nu(x)}{\pi \cdot \rho}}} \omega \cdot P_{\mathrm{CDP}}(\omega, y)\, d\omega\, d\theta, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

So, our desired property holds without loss of generality for all $x, y \geq 0$.

Appendix A.2 Proof of Equation (15)

We need to compute a closed-form formula for

$$x \cdot \frac{n}{A} \cdot \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\mathrm{CDP}}^{\mathrm{BLC}}(\omega, y)\, d\omega\, d\theta = 2 \cdot \pi \cdot x \cdot \frac{n}{A} \cdot \int_0^{r(x)} \omega \cdot P_{\mathrm{CDP}}^{\mathrm{BLC}}(\omega, y)\, d\omega, \tag{31}$$

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

$$\int_0^{r(x)} \omega \cdot P_{\text{CDP}}^{\text{BLC}}(\omega, y)\, d\omega = \int_0^{r(x)} \omega \cdot \frac{1}{1 + \mathrm{e}^{\lambda\cdot(\omega-\tau(y))}}\, d\omega. \tag{32}$$

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

$$\int \omega\cdot \frac{1}{1 + \mathrm{e}^{\lambda\cdot(\omega-\tau(y))}}\, d\omega = \int \omega\cdot \frac{\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))}}{\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1}\, d\omega = \mathrm{e}^{\lambda\tau(y)} \int \omega\cdot \frac{\mathrm{e}^{-\lambda\omega}}{\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1}\, d\omega. \tag{33}$$

Now, we must solve

$$\int \omega \cdot \frac{\mathrm{e}^{-\lambda\omega}}{\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1}\, d\omega, \tag{34}$$

which integrated by parts is equal to

$$\int \omega\cdot\frac{\mathrm{e}^{-\lambda\omega}}{\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1}\, d\omega = -\omega\frac{\mathrm{e}^{-\lambda\tau(y)}\ln(\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1)}{\lambda} - \int -\frac{\mathrm{e}^{-\lambda\tau(y)}\ln(\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1)}{\lambda}\, d\omega. \tag{35}$$

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

$$\int -\frac{\mathrm{e}^{-\lambda\tau(y)}\ln(\mathrm{e}^{-\lambda\cdot(\omega-\tau(y))} + 1)}{\lambda}\, d\omega = -\frac{\mathrm{e}^{-\lambda\tau(y)}}{\lambda^2}\int -\frac{\ln(1-z)}{z}\, dz = -\frac{\mathrm{e}^{-\lambda\tau(y)}}{\lambda^2}\mathrm{Li}_2(z). \tag{36}$$

If we go back, we obtain

$$n_{\text{CDP}}(x, y) \approx$$
$$\begin{cases} n\cdot\left(1 + \frac{2\cdot\mathrm{Li}_2\left(-\mathrm{e}^{-\lambda\tau(y)}\right)}{\lambda^2\cdot r(x)^2} - \frac{2\cdot\mathrm{Li}_2\left(-\mathrm{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda^2\cdot r(x)^2} - \frac{2\cdot\ln\left(1+\mathrm{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda\cdot r(x)}\right), & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

Appendix A.3 Proof of Property 2

To prove convergence in probability when $\omega \neq \tau(y)$, let F_λ be an infinite sequence of Bernoulli random variables, each with probability $p_\lambda = 1/\left(1 + \mathrm{e}^{\lambda\cdot(\omega-\tau(y))}\right)$ for $\lambda \geq 0$, *i.e.*, distributed according to the BLC model. Also, let $G = \mathbb{1}_{\{\omega\leq\tau(y)\}}$, be the corresponding indicator variable according to the DTC model.

As follows, we prove that $\lim_{\lambda\to\infty}\mathbb{P}(F_k = G) \to 1$, which is convergence in probability.

Indeed, we have that

$$\mathbb{P}(F_\lambda = G) = \begin{cases} \mathbb{P}(F_\lambda = 0) = 1 - p_\lambda, & \text{if} \quad \omega > \tau(y), \\ \mathbb{P}(F_\lambda = 1) = p_\lambda, & \text{if} \quad \omega < \tau(y), \\ 1/2, & \text{if} \quad \omega = \tau(y) \end{cases} \tag{38}$$

If $\lambda \to \infty$, then $p_\lambda \to 1$ when $\omega < \tau(y)$. Also, $p_\lambda \to 0$ when $\omega > \tau(y)$. So, we have that

$$\lim_{\lambda\to\infty} \mathbb{P}(F_\lambda = G) = \begin{cases} \lim_{\lambda\to\infty} \mathbb{P}(F_\lambda = 0) = 1, & \text{if} \quad \omega > \tau(y), \\ \lim_{\lambda\to\infty} \mathbb{P}(F_\lambda = 1) = 1. & \text{if} \quad \omega < \tau(y), \\ 1/2, & \text{if} \quad \omega = \tau(y) \end{cases} \tag{39}$$

which proves convergence in probability for all $\omega \neq \tau(y)$.

Appendix A.4 Model validation on the real road network

In this section, we use the CA model's circuity 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 construct a new instance set J by considering decision variables ranging as $(x, y) \in (0, 0) \cup \{1, \ldots, 40\} \times \{0, 0.5, 3\}$. Then, for each instance $j \in J$ and demand scenario $\xi \in \Xi$, we simulate the corresponding detailed daily operation. Figure 18 shows the average simulated costs for our case study in the Santiago metropolitan area, along with the costs predicted by our adjusted CA model $c(x, y)$. As with the analogous synthetic simulation presented in Figure 6, we observe that our approximate cost model performs well.

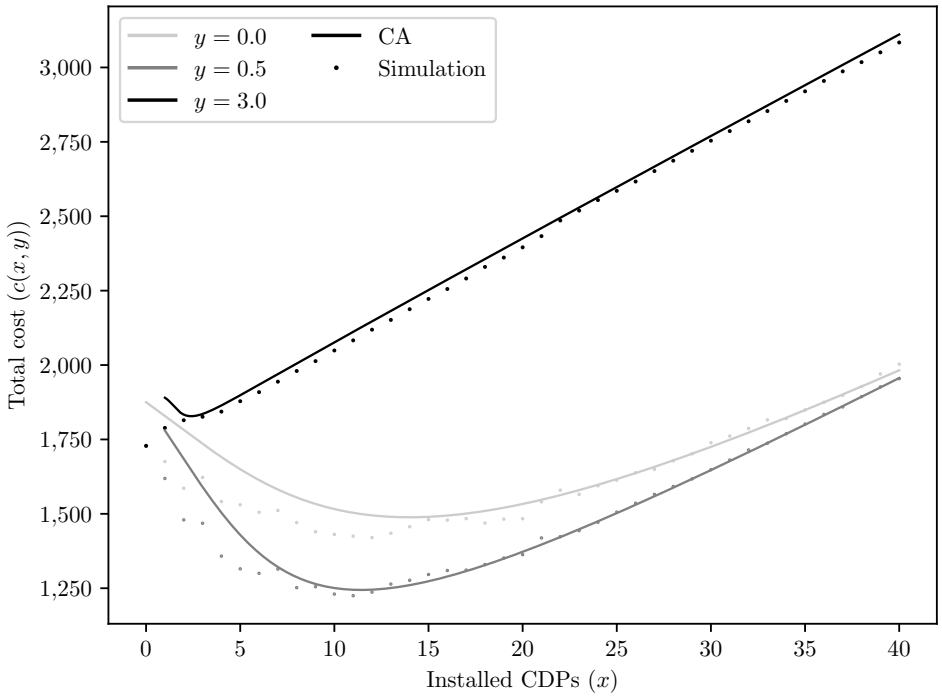
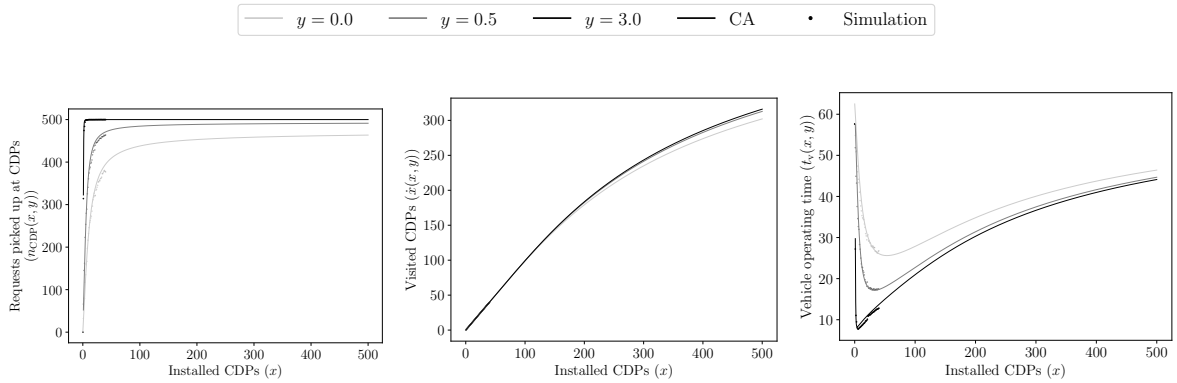


Figure 18: Approximate and simulated average total cost as function of x and y on the case study.

As in Figure 7, in Figure 19 we also present the validation of the expected number of

requests picked up at CDPs, $n_{\text{CDP}}(x, y)$, the expected number of visited CDPs, $\dot{x}$, and the expected total vehicle operating time, $t_{\text{v}}(x, y)$, for our real road network.



(a) $n_{\text{CDP}}(x, y)$ validation (b) $\dot{x}(x, y)$ validation (c) $t_{\text{v}}(x, y)$ validation

Figure 19: Validation of the main components of the CA model on the case study.

We add 80 more scenarios to Ξ and simulate the standard deviation of the number of requests picked up at CDPs. As in Figure 8 for our synthetic instances, Figure 20 shows both the simulated values of $\sigma_{\text{CDP}}(x, y)$ as well as those predicted by our CA model for the case study.

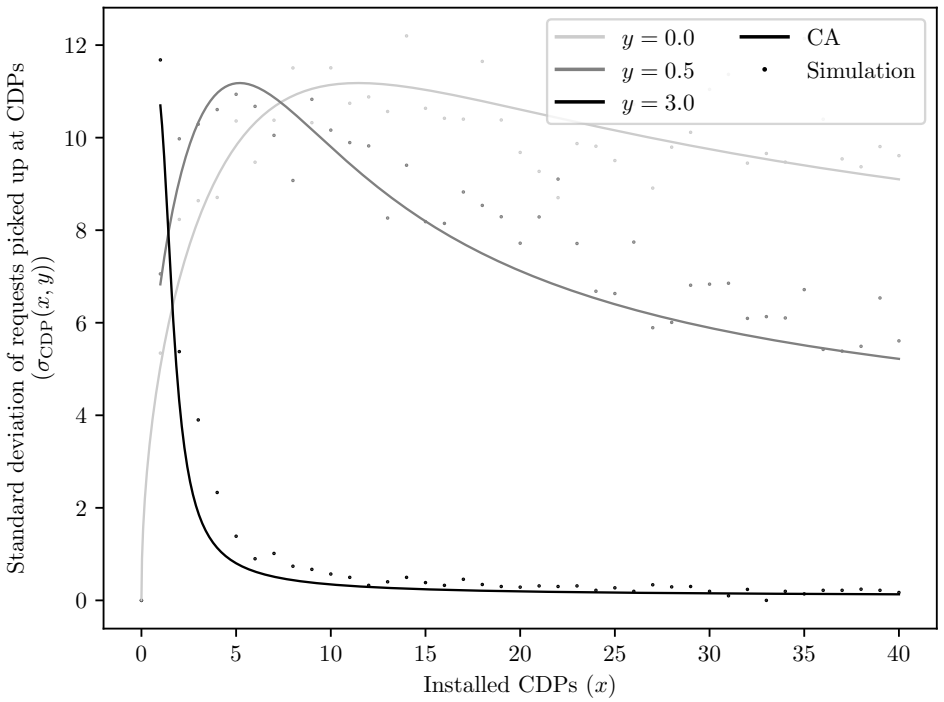


Figure 20: Approximate and simulated standard deviation of the number of requests chosen to be picked up at a CDP on the case study.

Table 9 reports the MAPE for all components of the CA model. In our case study, all errors on expected value estimations are less than or equal to 4.75%, demonstrating

the model's accuracy even when applied to real, complex road networks that deviate from idealized continuous assumptions. As expected, the error of $\sigma_{\mathrm{CDP}}(x, y)$ again exceeds that of $n_{\mathrm{CDP}}(x, y)$.

Table 9: MAPE of the CA model on the case study.

	$\boldsymbol{c(x,y)}$	$\boldsymbol{n_{\mathrm{CDP}}(x,y)}$	$\boldsymbol{\dot{x}(x,y)}$	$\boldsymbol{t_{\mathrm{v}}(x,y)}$	$\boldsymbol{\sigma_{\mathrm{CDP}}(x,y)}$
MAPE	2.09%	2.78%	0.21%	4.75%	16.2%