Strategic design 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 2, 2025

Abstract

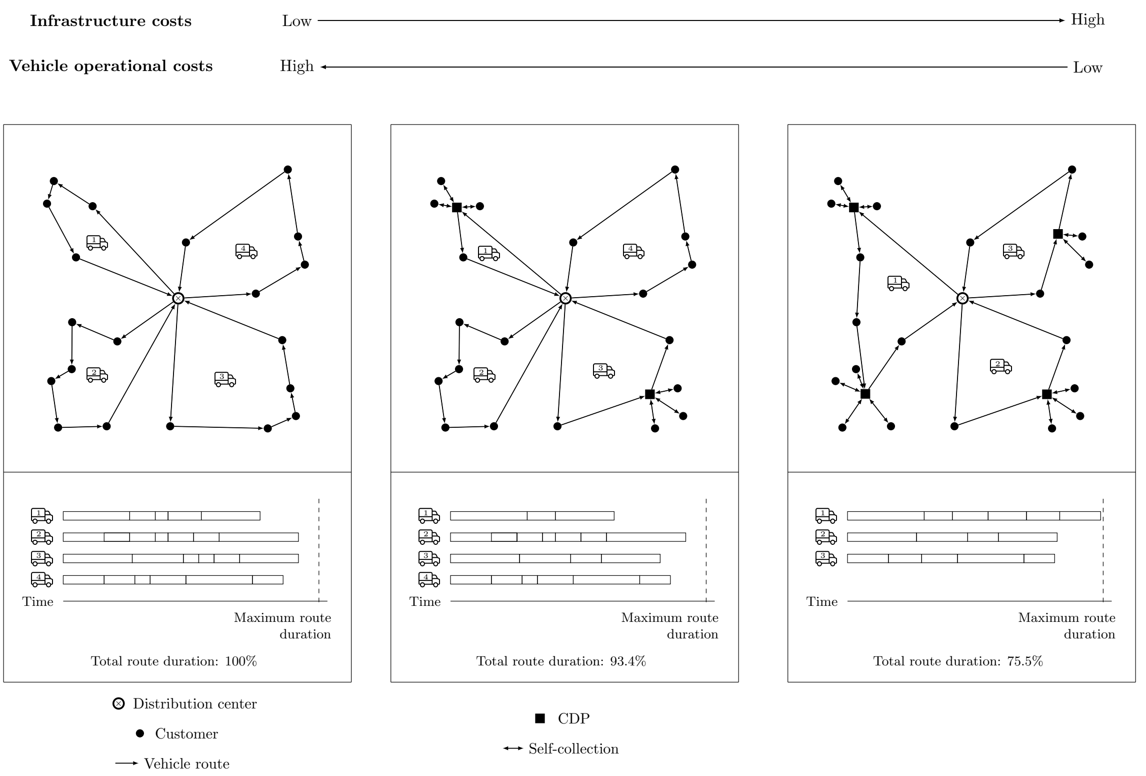
Collection and delivery points (CDPs) allow logistics operators to consolidate multiple customer request deliveries on a single vehicle stop, reducing distribution costs. However, for customers to adopt CDPs, they must be willing to travel to a nearby CDP to pick up their parcels. This choice depends on personal preferences, the proximity of CDPs, and economic incentives. In this study, we propose a strategic model that integrates CDP network design with customer incentives to minimize costs. Using a continuous approximation approach, we estimate operational costs and determine the optimal number of CDPs and customer incentive levels. In our experiments, we show that CDPs alone can reduce costs by 16.9% while incorporating incentives increases potential savings to 28.1%. We find that CDPs are most beneficial in high-density areas, leveraging economies of scale, whereas incentives are more effective in low-density regions. 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 161.2 billion in 2024 and projected to surpass 373.9 billion by 2033, primarily fueled by the increase in online orders (Banerjee, 2024). 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 (2024) 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, 48% are due to high costs (e.g., shipping fees and taxes) and 23% 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 a single delivery stop. 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 central depot. 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 optimize this balance between investing in CDPs and paying 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 efforts either optimize the operational details of these systems for a given CDP network or propose network design models that do not proactively account for operational routing costs. Typically, these studies rely on discrete linear optimization, which is effective for detailed route planning but less suited for strategic designs, where only expected operational cost estimates are needed. In strategic planning, it may be convenient to estimate routing costs via continuous approximation (CA) techniques, introduced in the seminal works of Beardwood et al. (1959) and Daganzo (1984), whose purpose is to provide a clear understanding of the trade-offs among various logistical costs and to offer tractable solutions based on aggregated data.

Moreover, the potential benefits of installing CDPs depend on the number of customers who opt to use them as pickup locations. So, we should understand customer preferences to accurately predict the utilization of a CDP network (Ma et al., 2022). In this sense, there exists a research gap in studying how customer choices impact operational costs and how these choices can be influenced via incentives. 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 might encourage customers to use a CDP by

(1) increasing the density of the CDP network to reduce average distances between CDPs and customer locations (Molin et al., 2022) and (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. (2024) have studied the impact of economic incentives on customers' delivery choices and only at the operational level.

In this paper, we explore how to determine an optimal CDP network design considering its impact on customers' delivery location choices and subsequent 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 incentives offered to customers that encourage them to pick up their delivery orders at CDPs. Also, we empirically calibrate and validate our CA model and compare it against detailed results obtained from simulated scenarios with random customer locations and choices, as well as optimized vehicle routes.

Based on our CA model, we derive the following insights regarding the use of CDPs coupled with customer incentives.

- An optimal installation of CDPs generates substantial savings, especially when request density is relatively high and infrastructure cost is relatively low. Furthermore, promoting the use of CDPs through incentives is particularly valuable in scenarios of relatively low request density. This is because incentives serve as an effective strategy to reduce the size of the installed network while ensuring high utilization, as incentivized customers are more willing to travel longer distances.
- It is important to have advanced customer behavior information for making structural decisions, as those decisions should account for customers' sensitivity to CDP distance. Errors from assuming deterministic customer choices are greater when sensitivity is lower.
- Total optimized cost increases sublinearly with the infrastructure cost, as the system adjusts by reducing CDP installations and increasing incentives. If CDP infrastructure is provided for free, 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, though they may still be useful in low-demand conditions.

The remainder of this paper is structured as follows. We review the related literature in

Section 2. In Section 3, we describe our problem and formulate our CA model, which we calibrate and validate in Section 4. We analyze the system's structural decisions in Section 5 to derive managerial insights. Finally, we conclude our work in Section 6.

2 Literature review

We present related literature on the operation and design of CDP networks to support the home delivery process. 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 who is assumed to select the delivery option (between home delivery and customer pickup at a CDP) and details the customer's choice model.

Table 1: A classification of studies involving CDPs.

		Decisions involved					**Who selects between CDP pickup and home delivery?**			
								Customer		
Category	**Reference**	**Number of CDPs**	**Approximate routing costs**	**Incentive to customers**	**Detailed location of CDPs**	**Detailed route planning**	**Operator**	Exogenous choice	Distance threshold choice	Discrete choice model
Operational	Dumez et al. (2021)					✓	✓			
	Tilk et al. (2021)					✓	✓			
	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. (2024)			✓						✓
Strategic	Deutsch and Golany (2018)	✓			✓				✓	
	Mancini et al. (2023)				✓				✓	
	Raviv (2023)				✓				✓	
	Lin et al. (2020)				✓					✓
	Lin et al. (2022)				✓					✓
	Xu et al. (2021)				✓					✓
Operational and strategic	Enthoven et al. (2020)				✓	✓	✓			
	Janjevic et al. (2019)		✓		✓	✓			✓	
	Our work	✓	✓	✓					✓	✓

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. 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. (2024) 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, most works focus on network design and CDP facility location models. The literature on facility location problems is vast within the operations research community (refer to Laporte et al. (2019) for a book on this topic). So, we focus on facility location articles dealing with CDP location decisions. Deutsch and Golany (2018) propose a model to set the optimal size and location of a CDP network that maximizes a logistics operator's profit. Mancini et al. (2023) study an extension of the capacitated facility location problem, where they choose how to locate CDPs under uncertain customer demands and CDP order processing capacities. Raviv (2023) examines a network design problem where they locate a CDP network and set each CDP's request storage capacity. 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. 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 strategic models aim to provide a detailed location of CDPs choosing their location from a known set of options. Moreover, they only consider settings where the CDP service is the only option available, disregarding the home delivery service option and its routing cost.

As strategic and operational levels are interdependent, some research efforts combine both the design and operation of a logistics distribution system via home delivery and customer pickups at CDPs. 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. 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.

Our study also builds on the existing literature on continuous approximation for delivery costs. The seminal work of Beardwood et al. (1959) states that the minimum length of a traveling salesman tour $L_n^\star$ over n uniformly distributed locations within a region of area A converges proportionally to $\sqrt{nA}$ with probability one as $n \to \infty$, meaning that we can approximate $L_n^\star$ as

$$L_n^\star \approx \beta\sqrt{nA}, \tag{1}$$

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). Continuous approximation 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.

In our problem, we influence customer's choices by offering them economic incentives. Such a setting has been explored in other last-mile problem contexts. For instance, there are research efforts focused on creating temporal and spatial flexibility in customers via incentives. 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, Çınar et al. (2024) and Horner et al. (2024) study how incentives can increase the performance of occasional couriers.

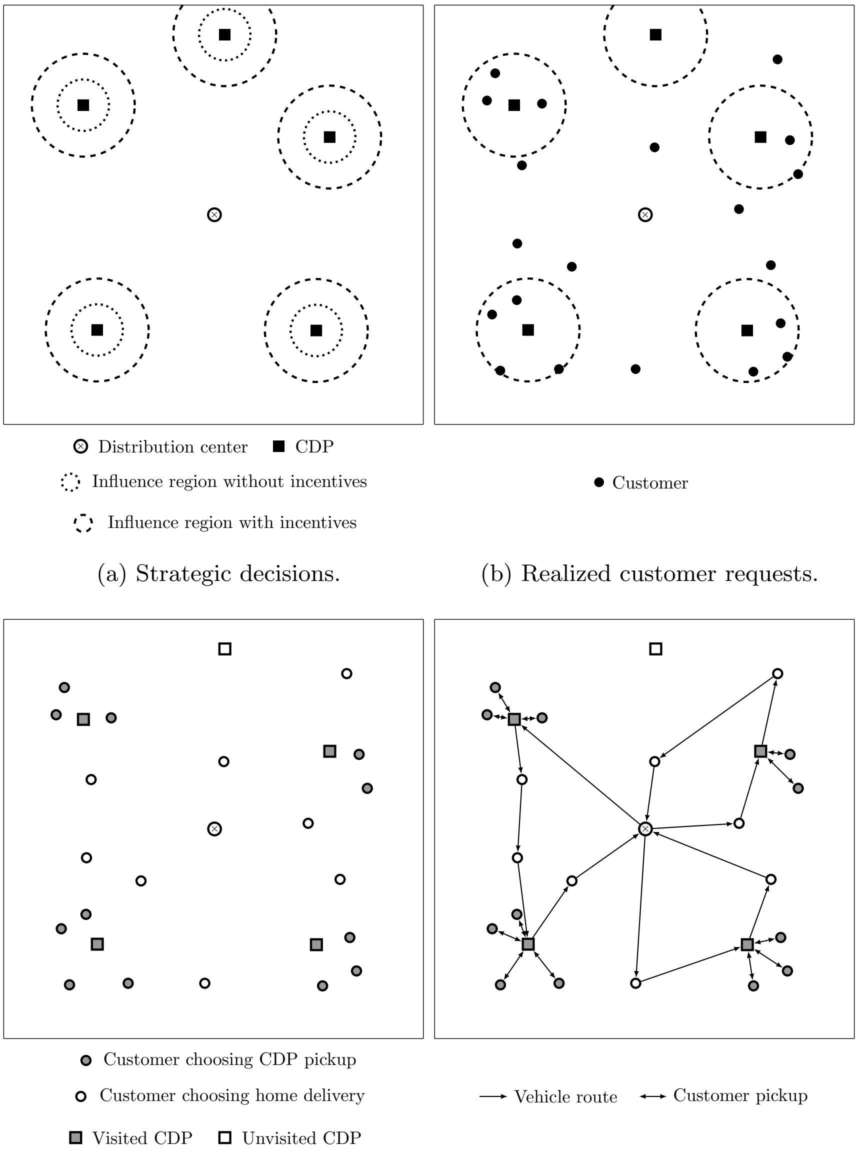
3 The CDP network design problem

We now introduce the CDP network design problem (CDP-NDP), which focuses on designing a cost-minimizing last-mile distribution system that can deliver orders either to the customers' homes or to the CDPs to be installed. Next, we formulate a continuous approximation model to approximate the system's cost and optimize its structural decisions. Finally, we introduce two different customer demand models that represent potential customer choice patterns.

3.1 Problem description

We study the design of a last-mile distribution operation that serves geographically distributed customers placing delivery orders online. These deliveries are dispatched in vehicle routes from a distribution center to each customer's location of choice, with response times as fast as those of next-day delivery. We consider that each customer can either receive their request at home or pick it up at one CDP chosen from those installed by the decision-maker. Also, we assume customers' delivery location choices are influenced by the economic incentives offered by the logistics operator. The CDP-NDP is a strategic planning problem focused on designing the CDP network that minimizes the system's total expected cost, including CDP expenses, last-mile distribution costs, and incentives paid to customers using CDPs.

In Figure 2, we present our problem's high-level dynamics. Figure 2a presents the operator's strategic decisions; these are (1) establishing the CDP network over a service region and (2) determining the economic incentive offered to customers picking up their parcels at CDPs. For simplicity, we depict the deterministic case where CDPs' influence regions are circular, and customers located within these regions use the corresponding CDP with certainty. Inner circles represent the coverage regions without incentives, while outer circles show expanded influence regions when economic incentives are offered to customers. In Figure 2b, we present a possible realization of customers' requests on a particular day. Then, we depict customers' delivery location choices in Figure 2c. Finally, Figure 2d illustrates the system's operational 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 chosen by at least one customer to pick up their goods.



(a) Strategic decisions. (b) Realized customer requests.

(c) Requests division. (d) Operational decisions.

Figure 2: Dynamics of the CDP network design problem.

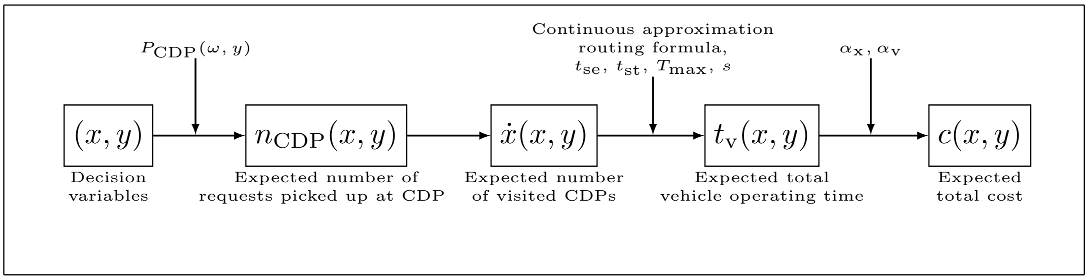
The CDP-NDP involves a complex facility location problem, which must account for expected future customer demand and learn how requests are revealed per unit time and location. Also, it should know and anticipate how customers choose between picking up their goods at CDPs and opting for home delivery. This choice is influenced by factors such as their proximity to the CDP network and the economic incentive offered for using the CDP pickup option. Such an incentive is a decision variable and impacts CDP location decisions, as incentivized customers may be willing to travel more and choose more distant CDPs. Beyond demand and customer choices, one must also account for expected last-mile distribution costs. This requires solving a vehicle routing problem for each customer set scenario and computing the expected vehicle operating time required. Moreover, this routing problem is intertwined with CDP location decisions and customer choices defining where deliveries should occur. The underlying detailed routing problem is challenging, even for a fleet of owned vehicles. Given these difficulties, we opt for a simpler approximate approach to obtain insight, measure the system's total cost, and evaluate its structural decisions.

3.2 Approximate cost model

In what follows, we approximate the cost of a last-mile distribution service 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. We assume knowledge of aggregated data, such as the expected number of n customers and the service area of A km^2 during the operating period. We also assume that customers are uniformly distributed over the service area, and that each customer's choice of delivery location depends on their relative distance to the nearest CDP and y. Also, we avoid choosing the exact location of the CDP network and assume that the x CDPs are evenly distributed over the service region, which is a reasonable assumption for a uniformly distributed demand. We also avoid detailed vehicle route planning and estimate last-mile distribution costs via continuous approximation. In our approximation, we consider the following cost components: (1) a cost α_{x} paid per installed CDP 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 takes a fixed time t_{st} plus an additional service time t_{se} per delivered order. We also assume 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 cost model. The approximate cost function $c(x, y)$ is defined in a series of approximation steps starting from a pair of values x and y. Then, 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 region.	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
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
$P_{\text{CDP}}(\omega, y)$	Customer CDP choice probability.	-
$n_{\text{CDP}}(x, y)$	Expected number of requests picked up at CDPs	requests/day
$\dot{x}(x, y)$	Expected number of visited CDPs.	CDPs
$t_{\text{v}}(x, y)$	Expected total vehicle operating time.	hours/day
$c(x, y)$	Expected total cost.	\$/day

Expected number of requests picked up at CDPs. If customers only consider their closest CDP, then evenly distributed CDPs define hexagonal influence regions. We approximate these influence regions by circles of radius $r(x) = \sqrt{A/\pi/x}$. Let $N_{\text{CDP}}(x, y)$ be a random variable representing the number of requests picked up at CDPs. Its expected value $n_{\text{CDP}}(x, y)$ is

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

In this expression, the term $n/A \cdot \int_0^{2\pi} \int_0^{r(x)} \omega \cdot P_{\text{CDP}}(\omega, y)\, d\omega\, d\theta$ estimates the expected daily demand of a single CDP considering that each customer inside the influence region will choose it with probability $P_{\text{CDP}}(\omega, y)$.

Expected number of visited CDPs. The operator should only schedule delivery visits to CDPs selected as pickup station by at least one customer. For n customers, the expected number of unused CDPs grows as a function of the total number installed x. For each CDP $k \in \{1, ..., x\}$, we define U_k as a binary random variable taking value 1 when the k^{th} CDP is not used and value 0 otherwise. We assume variable U_k follows a Bernoulli distribution with probability $(1 - 1/x)^{N_{\text{CDP}}(x,y)}$. To see it, consider that $1/x$ represents the likelihood that a single customer chooses a particular CDP. Consequently, $(1 - 1/x)$ is the probability that this individual does not select this CDP. Therefore, $(1 - 1/x)^{N_{\text{CDP}}(x,y)}$ represents the probability that a particular CDP is not selected by all customers choosing the CDP pickup option. So, the expected number of unused CDPs is

$$\mathbb{E}\left[\sum_{k=1}^{x} U_k\right] = \sum_{k=1}^{x} \mathbb{E}[U_k] = x \cdot \mathbb{E}\left[(1 - 1/x)^{N_{\text{CDP}}(x,y)}\right]. \tag{3}$$

Furthermore, we make a conservative estimate $\mathbb{E}\left[(1 - 1/x)^{N_{\text{CDP}}(x,y)}\right] \approx (1 - 1/x)^{n_{\text{CDP}}(x,y)} \approx \mathrm{e}^{(-n_{\text{CDP}}(x,y)/x)}$, given the properties of the exponential function and Jensen's inequality stating that $\mathbb{E}[(1 - 1/x)^{N_{\text{CDP}}(x,y)}] \geq (1 - 1/x)^{n_{\text{CDP}}(x,y)}$. Let $\dot{X}(x, y)$ be the random variable representing the number of visited CDPs. The expected number of visited CDPs $\dot{x}(x, y) \leq x$ is estimated as

$$\dot{x}(x, y) = \mathbb{E}\left[\dot{X}(x, y)\right] \approx \begin{cases} x \cdot \left(1 - \mathrm{e}^{-\frac{n_{\text{CDP}}(x,y)}{x}}\right), & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Expected total vehicle operating time. In our setting, the number of delivery stops $M(x, y)$ is the sum of all used CDPs plus the number of home deliveries, *i.e.*, $M(x, y) \approx \dot{X}(x, y) + n - N_{\text{CDP}}(x, y)$. Considering the service time per request, the fixed time per stop, and a square root routing time between stops, let $T_{\text{op}}(x, y) \approx \beta \cdot t_{\text{se}} \cdot n + \gamma \cdot t_{\text{st}} \cdot M(x, y) + \delta/s \cdot \sqrt{M(x, y) \cdot A}$ be the vehicle operating time without accounting for stem routing times, where β, γ and δ are the approximation constants to be calibrated. To calculate $\mathbb{E}[T_{\text{op}}(x, y)]$, the number of stops $M(x, y)$ can be replaced by its expected value $m(x, y) = \mathbb{E}[M(x, y)] \approx \dot{x}(x, y) + n - n_{\text{CDP}}(x, y)$, as proposed by Banerjee et al. (2024). Therefore, we obtain

$$t_{\text{op}}(x, y) = \mathbb{E}[T_{\text{op}}(x, y)] \approx \beta \cdot t_{\text{se}} \cdot n + \gamma \cdot t_{\text{st}} \cdot m(x, y) + \frac{\delta}{s} \cdot \sqrt{m(x, y) \cdot A}. \tag{5}$$

To include stem routing times, we follow the approach of Daganzo (1984) and estimate the expected total vehicle operating time as

$$t_{\mathrm{v}}(x,y) \approx 2 \cdot t_0 \cdot \frac{t_{\mathrm{op}}(x,y)}{T_{\max}} + t_{\mathrm{op}}(x,y), \tag{6}$$

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 ($t_{\mathrm{op}}(x,y)/T_{\max}$) plus the remaining operating time $t_{\mathrm{op}}(x,y)$. If we define $\kappa = (1+2\cdot t_0/T_{\max})$, we obtain

$$t_{\mathrm{v}}(x,y) \approx \beta \cdot \kappa \cdot t_{\mathrm{se}} \cdot n + \gamma \cdot \kappa \cdot t_{\mathrm{st}} \cdot m(x,y) + \delta \cdot \frac{\kappa}{s} \cdot \sqrt{m(x,y) \cdot A}. \tag{7}$$

As expected, when the maximum vehicle workday length $T_{\max} \to 0$, then the operation becomes infeasible and total vehicle operating time explodes. In contrast, when $T_{\max} \to \infty$, then total vehicle operating time $t_{\mathrm{v}}(x,y)$ converges to $t_{\mathrm{op}}(x,y)$ as it collapses to the expected cost 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

$$c(x,y) \approx \alpha_{\mathrm{x}} \cdot x + \alpha_{\mathrm{v}} \cdot t_{\mathrm{v}}(x,y) + y \cdot n_{\mathrm{CDP}}(x,y). \tag{8}$$

Having defined the cost function, we can optimize the problem

$$\min\{c(x,y) \quad \text{s.t.} \quad x,y \geq 0\} \tag{9}$$

and approximately optimize structural decisions (x,y) for our distribution system equipped with CDPs. Below, we present a result related to the average cost per request function; its proof is detailed in Appendix A.1.

Property 1. The average cost per request $c(x,y)/n$ only depends on x, n or A through the request density $\rho = n/A$ and the number of requests 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.*, $L_n^\star/n \approx \beta\sqrt{nA}/n = \beta/\sqrt{\rho}$.

3.3 CDP choice probability function

Our model requires a CDP choice probability function $P_{\text{CDP}}(\omega, y)$ that depends on the customer's distance ω to the closest CDP and the incentive y. Here, we propose two models representing different customer behaviors. In what follows, 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 $\tau(y)$ must 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 (10), this choice model assumes that customers always choose the CDP pickup option if this station is located within a distance less than or equal to $\tau(y)$ from their location; otherwise, they opt for home delivery.

$$P_{\text{CDP}}(\omega, y) = \begin{cases} 1, & \text{if } \omega \leq \tau(y) \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

If we replace this specific probability model in Equation (2) we obtain the following expected number of requests picked up at CDPs

$$n_{\text{CDP}}(x, y) \approx n \cdot \min \left\{ 1, \left(\frac{\tau(y)}{r(x)} \right)^2 \right\}. \tag{11}$$

Intuitively, Equation (11) 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

$$P_{\text{CDP}}(\omega, y) = \frac{1}{1 + \mathrm{e}^{\lambda \cdot (\omega - \tau(y))}}, \tag{12}$$

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 (2) we obtain

$$n_{\text{CDP}}(x,y) \approx$$
$$n \cdot \left(1 + \frac{2 \cdot \text{Li}_2\left(-\text{e}^{-\lambda\tau(y)}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \text{Li}_2\left(-\text{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \ln\left(1 + \text{e}^{\lambda\cdot(r(x)-\tau(y))}\right)}{\lambda \cdot r(x)}\right), \quad (13)$$

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 (13) 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. With probability 1, the BLC model converges to the DTC model 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.

4 Parameter setting, model calibration and validation

In this section, we set parameter values for our study representing a realistic last-mile distribution context. We then calibrate the coefficients of our cost model approximation defined in Equation (8) and validate its adjustment to the exact optimal cost of detailed computationally simulated instances.

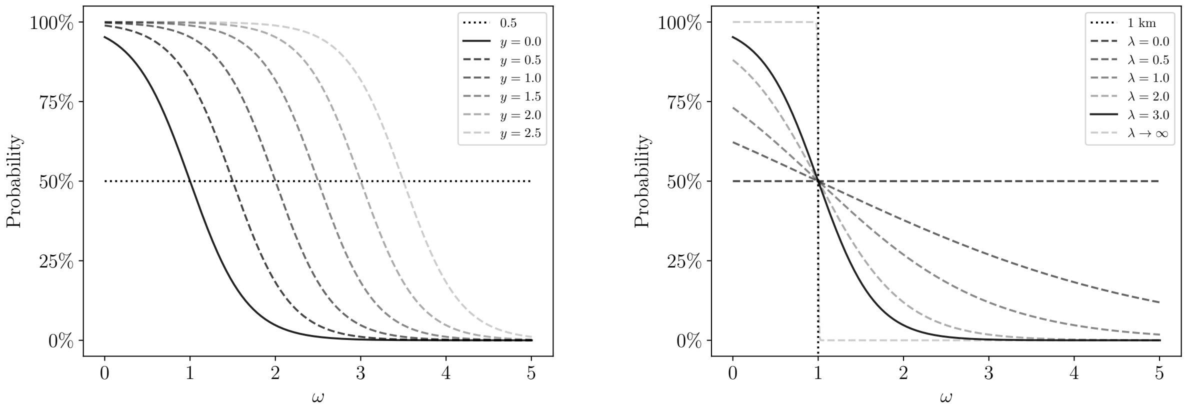
4.1 Parameter values in our study

We now set parameter values for our approximate cost model, as summarized in Table 3. We consider an operation expecting $n \in \{100, 200, 400\}$ requests/day within a circular service region with a radius of $R = 3$ km, which results in a request density ρ ranging between 3.54 and 14.14 requests/km^2/day; these values are comparable to the 6.5 requests/km^2/day 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). In this setting, the expected distance from the distribution center to a uniformly distributed customer location is $2R/3$ and, therefore, the expected travel time is $t_0 = 2R/3/s = 0.13$ hours. The time spent per vehicle stop in a service route is set to 1, 3.5, and 10 minutes to represent different urban distribution operations with varying levels of difficulty for stopping and parking. Also, we add an additional service time of 0.5 minutes per parcel delivered. We consider three different workday length values $T_{\max} \in \{3, 5, 8\}$ hours/day, which represent different labor laws such as the Chilean one (Law No. 20,271, Art. 25 bis., 2008) stating that workers cannot drive more than 5 consecutive hours. The daily cost paid per installed CDP is set to $\alpha_{\mathrm{x}} = \$30/\mathrm{CDP}/\mathrm{day}$, which lies within the \$16 to \$34 range discussed by Xu et al. (2021). The hourly cost per operating vehicle is set to $\alpha_{\mathrm{v}} = \$30/\mathrm{hour}$ 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.	$100, 200, 400$	requests/day
A	Service region.	$\pi R^2 \approx 28.27$	km^2
t_0	Expected travel time between the depot and a customer.	$2R/3/s \approx 0.13$	hours
t_{se}	Variable service time per delivered request.	$0.5/60$	hours/request
t_{st}	Fixed time per stop.	$1/60, 3.5/60, 10/60$	hours/stop
$T_{\max}$	Maximum vehicle workday length.	$3, 5, 8$	hours/day
s	Average vehicle speed.	15	km/hour
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. Our choice of λ is conservative as $P_{\mathrm{CDP}}(3, 0) \leq 0.001$. Figure 4a plots the CDP choice probability $P_{\mathrm{CDP}}(\omega, y)$ for our BLC model over different values of incentive y. As expected, $P_{\mathrm{CDP}}(\omega, y) = 50\%$ is reached when $\omega = \tau(y)$. Figure 4b depicts $P_{\mathrm{CDP}}(\omega, y)$ for different values of λ when $y = \$0/\mathrm{request}$. As confirmed by Property 2, a customer's CDP service choice becomes less variable and more predictable when λ increases.



(a) $P_{\text{CDP}}(\omega, y)$ over different incentive values y.

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

Figure 4: CDP choice probability $P_{\text{CDP}}(\omega, y)$.

4.2 Model calibration and validation

Now, we calibrate the coefficients of our approximate cost model and validate its adjustment by comparing it to the average operating cost of randomly simulated instances optimized with detailed route planning. 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 $T_{\max} \in \{3, 5, 8\}$. 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 there are installed CDPs. For each instance $i = (n, t_{\text{st}}, T_{\max}, x, y) \in I$, we evenly distribute the x CDPs over the circular region and simulate a set $\Psi = \{1, \ldots, 100\}$ of customer demand scenarios. For each demand scenario $\psi \in \Psi$, we draw n customer requests from a uniform distribution over the service region and randomly generate the service choice (*i.e.*, pickup at CDP or home delivery) for each request depending on its distance ω to the closest CDP and y based on their the BLC model $P_{\text{CDP}}(\omega, y)$ probability function. For each instance $i \in I$ and demand scenario $\psi \in \Psi$, we generate a set of customer home delivery requests and a set of CDPs with a specific number of requests to be picked up at their location. Given each pair (i, ψ), we compute its last-mile distribution cost running the PyVRP heuristic solver Wouda et al. (2024). We filter out any instance from set I yielding demand scenarios where the VRP cannot find a feasible solution.

Of these 100 demand scenarios per instance, we use 80 as a calibration sample to estimate the parameters β, γ and δ in the total travel time function $t_{\text{v}}(x, y)$ defined in Equation (7) via ordinary least squares regression, adjusting our model's $t_{\text{v}}(x, y)$ value 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 $\beta \approx 1.110$, $\gamma \approx 1.048$ and

$\delta \approx 0.755$.

We use the remaining 20 demand scenarios to validate our model's expected results. For the average instance case with $n = 200$, $t_{\text{st}} = 3.5/60$ and $T_{\max} = 5$, Figure 5 plots our model's approximate expected cost $c(x, y)$ curve and compares it to the simulated sampled 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.

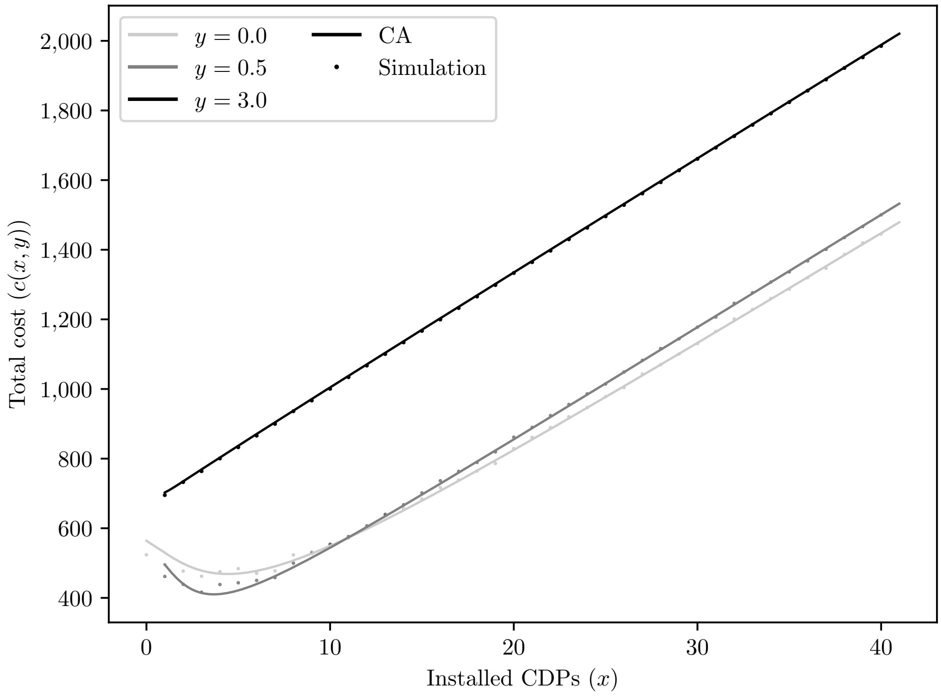
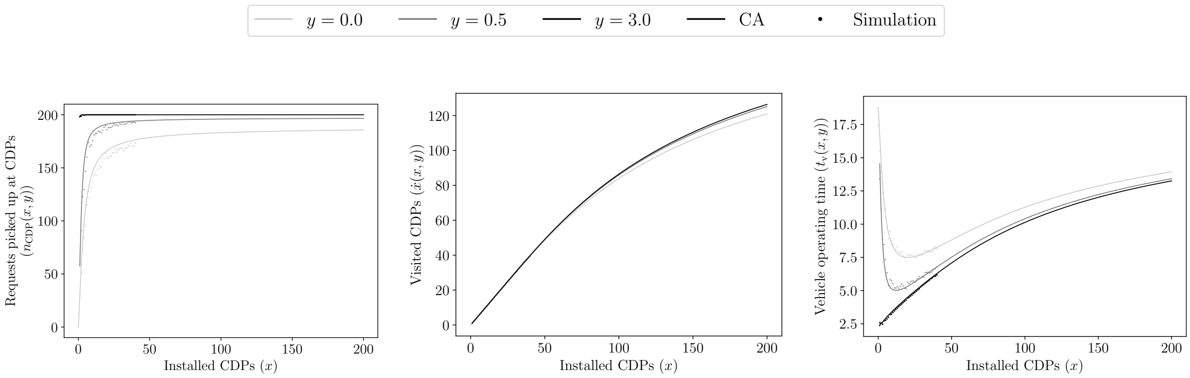


Figure 5: Approximate and simulated average total cost as a function of x and y given $n = 200, t_{\text{st}} = 3.5/60$ and $T_{\max} = 5$.

Figure 6 explores further the quality of our proposed CA model and shows its estimate of the expected number of requests picked up at CDPs $n_{\text{CDP}}(x, y)$, the expected number of visited CDPs $\dot{x}(x, y)$, and the total vehicle operating time $t_{\text{v}}(x, y)$. Each of these values is compared to the corresponding analogous sampled average value as a function of x and y. We also plot an extrapolation of our CA model for numbers of installed CDPs x above 40. Particularly, Figure 6b highlights how the use of customer incentives effectively increases the value of $n_{\text{CDP}}(x, y)$. Figure 6c 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 6d displays the corresponding values of $t_{\text{v}}(x, y)$. 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.



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

Figure 6: Validation of the main components of the CA model.

Table 4 presents the mean absolute percentage error (MAPE) for each component of the CA model compared to its simulated sampled average for all instances $i \in I$. Specifically, the errors 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.

Table 4: MAPE of the CA model.

	$\boldsymbol{c(x, y)}$	$\boldsymbol{n_{\text{CDP}}(x, y)}$	$\boldsymbol{\dot{x}(x, y)}$	$\boldsymbol{t_{\text{v}}(x, y)}$
MAPE	1.12%	2.89%	0.29%	3.97%

5 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.

5.1 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 depot 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 (9) is run via the SHGO solver (Endres et al., 2018) available in the Scipy library (Virtanen et al., 2020).

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 (9) s.t. $y = 0$
Minimum cost	$C^\star = c(x^\star, y^\star)$	$x^\star, y^\star$	Model (9)
Minimum cost with free CDPs	$C^\mathrm{F} = c(x^\mathrm{F}, y^\mathrm{F})$	$x^\mathrm{F}, y^\mathrm{F}$	Model (9) assuming $\alpha_\mathrm{x} = 0$

5.2 Base instance analysis

We begin studying results over instance parameters in Table 3 and fix $n = 200$, $t_\mathrm{st} = 3.5/60$ and $T_\mathrm{max} = 5$, which represents an average instance with a request density $\rho = 7.07$ requests per squared kilometer. Figure 7 plots a heat map of function $c(x, y)$ with $(x, y) \in [0, 6] \times [0, 2]$. Each color corresponds to a specific total cost, with values ranging between \$405.9/day to \$671.4/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 ($n_\mathrm{CDP}(x, y)/x$).

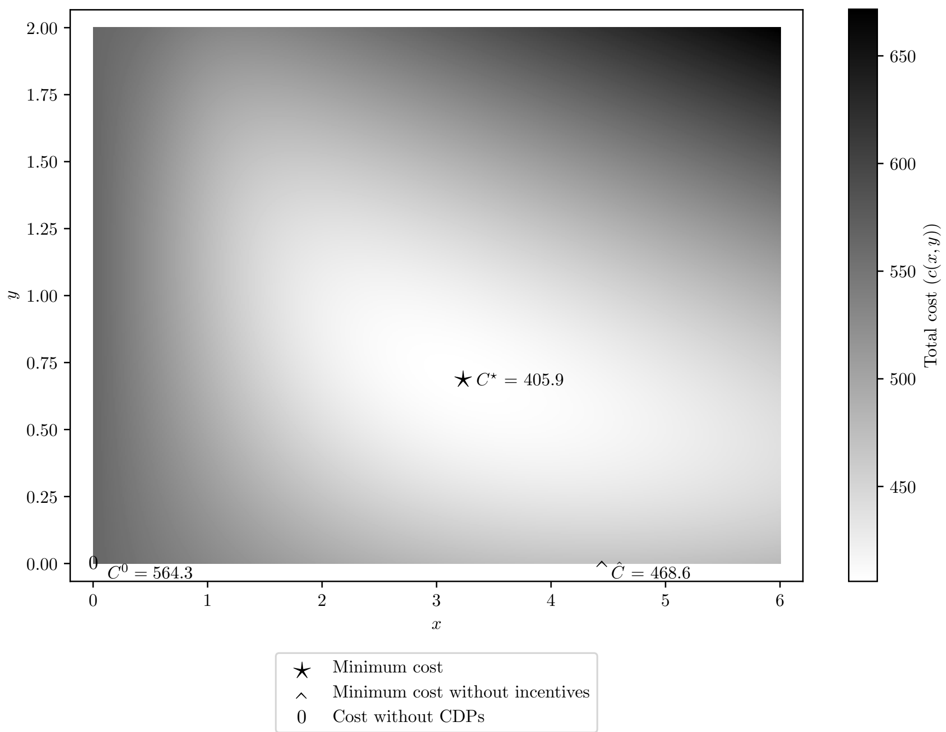


Figure 7: Heat map of $c(x, y)$ for $n = 200$, $t_{\text{st}} = 3.5/60$ and $T_{\max} = 5$.

Table 6: Benchmark results for $n = 200$, $t_{\text{st}} = 3.5/60$ and $T_{\max} = 5$.

	Benchmark			
	Cost without CDPs	**Minimum cost without incentive**	**Minimum cost**	**Minimum cost with free CDPs**
Cost	$C^0 = 564.3$	$\hat{C} = 468.6$	$C^\star = 405.9$	$C^{\text{F}} = 222.7$
% savings	-	16.9%	28.1%	60.5%
Infrastructure cost	0	133.0	97.4	0
Distribution cost	564.3	335.6	199.0	203.9
Incentive cost	0	0	109.6	18.7
x	0	4.43	3.25	17.81
y	0	0	0.69	0.11
$m(x, y)$	200.0	99.8	43.8	45.7
$n_{\text{CDP}}(x, y)/x$	-	23.6	49.1	9.7

As observed, the system's cost without CDPs is $C^0 = \$564.3$/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 $\hat{C} = \$468.6$/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 \$335.6/day. As observed in Figure 7 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 $C^\star = \$405.9$/day (28.1% savings), cutting down distribution costs even more to \$199.0 and also infrastructure costs from \$133.0 to \$97.4, but adding an extra incentive cost of \$109.6. Compared to $\hat{x}$, the optimal number of CDPs reduces by 26.8% to $x^\star = 3.25$ (roughly $\nu(x^\star) = 61.1$ 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 $C^{\mathrm{F}} = \$222.7$ (60.5% savings). Compared with the minimum cost benchmark, customer inventive downs to $y^{\mathrm{F}} = \$0.11$ 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 $x^{\mathrm{F}} = 17.81$. 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.

5.3 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 8 plots the expected cost per request, the chosen numbers of requests per CDP ($\nu(x)$) 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 squared 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 $y = \$1.5$ per request picked up at a CDP and occurs when $\rho = 1.4$ requests per squared 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.

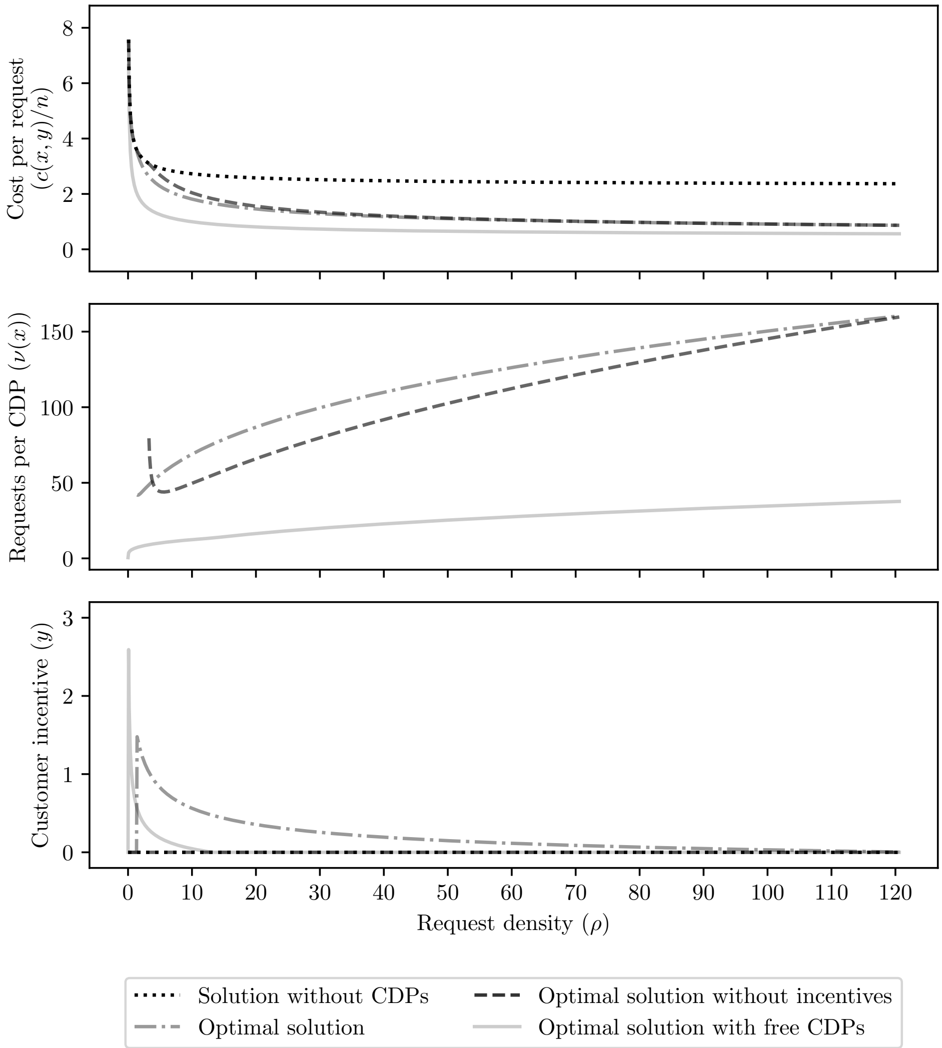


Figure 8: Benchmark strategies as a function of request density ($\rho = n/A$).

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

Figure 9 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' behave as the logit (BLC) and deterministic (DTC) models. It also presents the resulting expected cost per request for these decisions evaluated in Model Equation (8) using the more realistic BLC model. We also plot the CDP service choice probability $P_{\text{CDP}}(\bar{\omega}, y)$ 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.

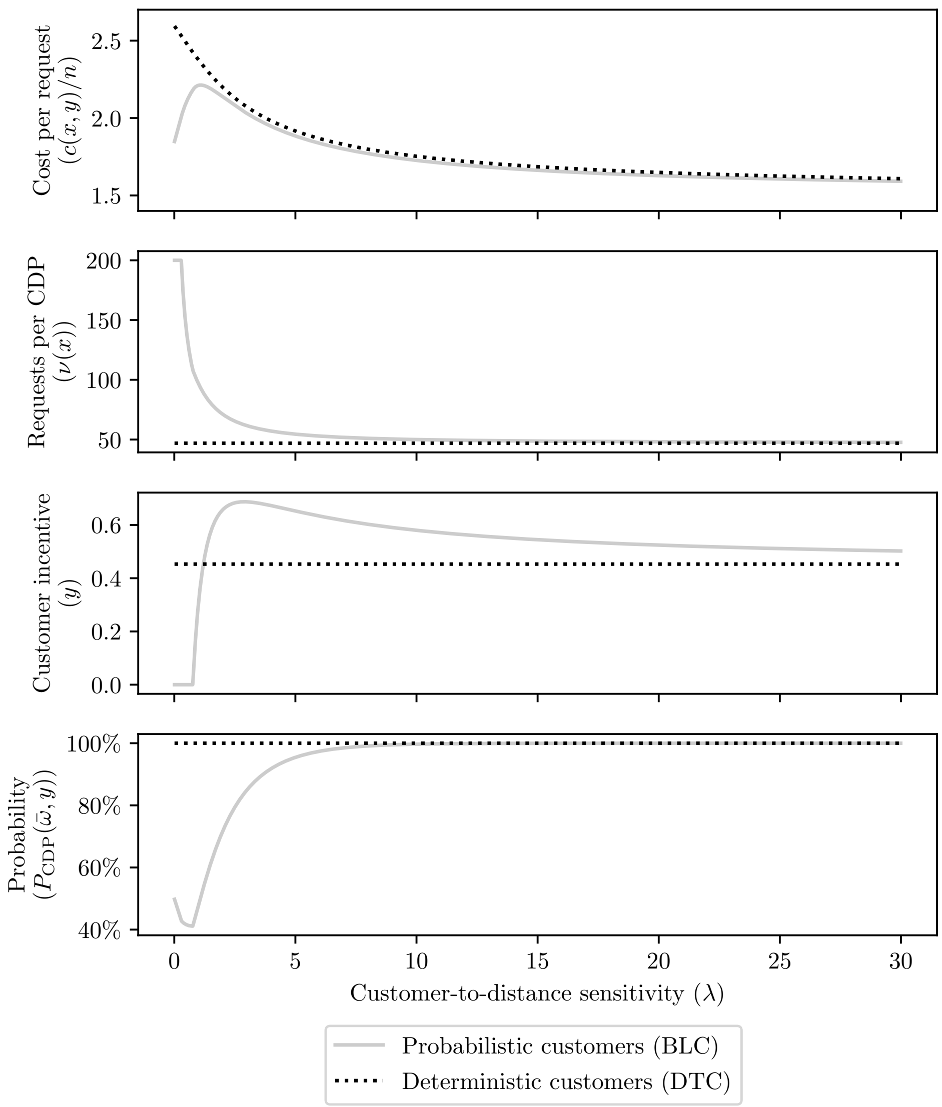


Figure 9: Optimal strategy for the BLC and DTC choice models as a function of customer-to-distance sensitivity (λ).

5.5 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 10 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 sufficiently 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.

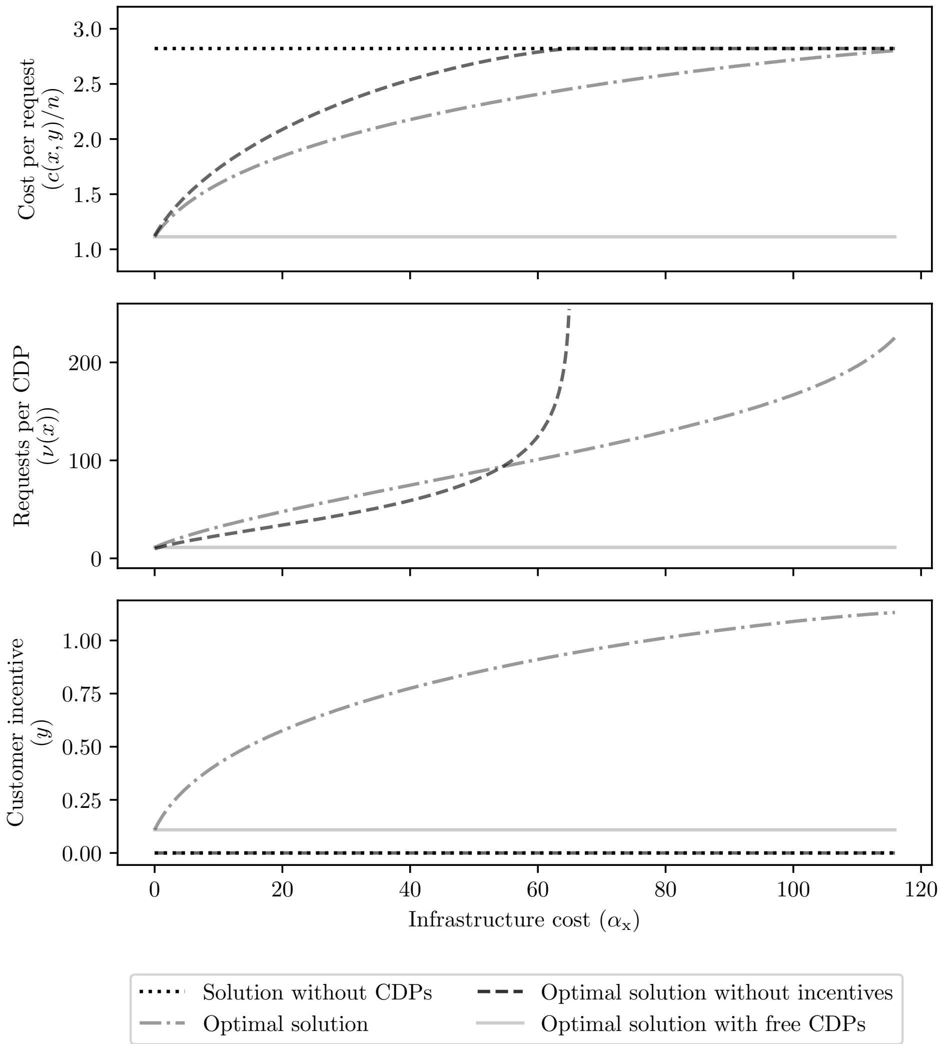


Figure 10: Benchmark strategies as a function of infrastructure cost (α_x).

6 Conclusions

In this paper, we present and validate a continuous approximation function depending on aggregate instance data, which estimates the expected operating cost of a logistics last-mile distribution system delivering to both customers' locations and CDPs, based on the customer's 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.

Our CA model shows that the expected cost per request is determined exclusively by the CDP network size, the request density, and the average number of requests per CDP, rather than by the total number of requests or the service area itself. This finding is consistent with the scale-independent average 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 28.1% 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 installment 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 also 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 which 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.

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. 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. (2024). Learning dynamic selection and pricing of out-of-home deliveries. *Transportation Science*, 0(0).

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, A. (2024). Last mile delivery market size. Viewed 28 of January 2025, `https://straitsresearch.com/report/last-mile-delivery-market`.

Banerjee, D., Erera, A. L., and Toriello, A. (2024). Pricing and demand management for integrated same-day and next-day delivery systems. *Transportation Science*, 0(0).

Baymard Institute (2024). 49 cart abandonment rate statistics 2024. Viewed 28 of January 2025, `https://www.statista.com/statistics/1286612/last-mile-delivery-market-size-worldwide/`.

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.

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

Coppola, D. (2024). Share of last-mile delivery costs out of total shipping costs in 2018 and 2023. Viewed 28 of January 2025, `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.

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.

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.

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.

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.

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.

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.

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

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.

World Economic Forum (2020). The future of the last-mile ecosystem. Viewed 28 of January 2025, `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.

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.

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

$$\frac{c(x,y)}{n} \approx \frac{\alpha_{\mathrm{x}}}{\nu(x)} + \alpha_{\mathrm{v}} \cdot \frac{t_{\mathrm{v}}(x,y)}{n} + y \cdot \frac{n_{\mathrm{CDP}}(x,y)}{n}. \tag{14}$$

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

$$\frac{n_{\mathrm{CDP}}(x,y)}{n} \approx \begin{cases} \frac{\rho}{\nu(x)} \cdot \int_0^{2\pi} \int_0^{\sqrt{\frac{\nu(x)}{\pi\cdot\rho}}} \omega \cdot P_{\mathrm{CDP}}(\omega, y)\, d\omega\, d\theta, & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

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

$$\frac{t_{\mathrm{v}}(x,y)}{n} \approx \beta \cdot \kappa \cdot t_{\mathrm{se}} + \gamma \cdot \kappa \cdot t_{\mathrm{st}} \cdot \frac{m(x,y)}{n} + \delta \cdot \frac{\kappa}{s} \cdot \sqrt{\frac{1}{\rho} \cdot \frac{m(x,y)}{n}}, \tag{16}$$

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

$$\frac{m(x,y)}{n} \approx \frac{\dot{x}(x,y)}{n} + 1 - \frac{n_{\mathrm{CDP}}(x,y)}{n}. \tag{17}$$

The number of visited CDPs per request is

$$\frac{\dot{x}(x,y)}{n} \approx \begin{cases} \frac{1}{\nu(x)} \cdot \left(1 - \mathrm{e}^{-\frac{n_{\mathrm{CDP}}(x,y)}{x}}\right), & \text{if } x > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{18}$$

and $n_{\text{CDP}}(x,y)/x$ satisfies

$$\frac{n_{\text{CDP}}(x,y)}{x} \approx \begin{cases} \rho \cdot \int_0^{2\pi} \int_0^{\sqrt{\frac{\nu(x)}{\pi \cdot \rho}}} \omega \cdot P_{\text{CDP}}(\omega, y)\, d\omega\, d\theta, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

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

Appendix A.2 Proof of Equation (13)

We need to compute a closed-form formula for

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

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

$$\int_0^{r(x)} \omega \cdot P_{\text{CDP}}(\omega, y)\, d\omega = \int_0^{r(x)} \omega \cdot \frac{1}{1 + \mathrm{e}^{\lambda \cdot (\omega - \tau(y))}}\, d\omega. \tag{21}$$

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

$$\int \omega \cdot \frac{1}{1 + \mathrm{e}^{\lambda \cdot (\omega - \tau(y))}}\, d\omega = \int \omega \cdot \frac{\mathrm{e}^{-\lambda \cdot (\omega - \tau(y))}}{\mathrm{e}^{-\lambda \cdot (\omega - \tau(y))} + 1}\, d\omega = \mathrm{e}^{\lambda \tau(y)} \int \omega \cdot \frac{\mathrm{e}^{-\lambda \omega}}{\mathrm{e}^{-\lambda \cdot (\omega - \tau(y))} + 1}\, d\omega. \tag{22}$$

Now, we must solve

$$\int \omega \cdot \frac{\mathrm{e}^{-\lambda \omega}}{\mathrm{e}^{-\lambda \cdot (\omega - \tau(y))} + 1}\, d\omega, \tag{23}$$

which integrated by parts is equal to

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

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

$$\int -\frac{\mathrm{e}^{-\lambda \tau(y)} \ln(\mathrm{e}^{-\lambda \cdot (\omega - \tau(y))} + 1)}{\lambda}\, d\omega = -\frac{\mathrm{e}^{-\lambda \tau(y)}}{\lambda^2} \int -\frac{\ln(1 - z)}{z}\, dz = -\frac{\mathrm{e}^{-\lambda \tau(y)}}{\lambda^2} \mathrm{Li}_2(z). \tag{25}$$

If we go back, we obtain

$$n_{\text{CDP}}(x,y) \approx$$
$$n \cdot \left(1 + \frac{2 \cdot \mathrm{Li}_2\left(-\mathrm{e}^{-\lambda \tau(y)}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \mathrm{Li}_2\left(-\mathrm{e}^{\lambda \cdot (r(x) - \tau(y))}\right)}{\lambda^2 \cdot r(x)^2} - \frac{2 \cdot \ln\left(1 + \mathrm{e}^{\lambda \cdot (r(x) - \tau(y))}\right)}{\lambda \cdot r(x)}\right). \tag{26}$$

Appendix A.3 Proof of Property 2

From Equation (12), we derive

$$\lim_{\lambda \to \infty} P_{\text{CDP}}(\omega, y) = \begin{cases} 1 & \text{if} \quad \omega < \tau(y) \\ \frac{1}{2} & \text{if} \quad \omega = \tau(y) \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

The case $\omega = \tau(y)$ occurs with probability 0 and can be discarded. So, we obtain the distance threshold model presented in Equation (10).