Article:

Zadeh, A.S., Dayarian, I. and Bashiri, M. orcid.org/0000-0002-5448-1773 (2026) Strategic design of parcel locker networks for urban delivery. Transportation Research Part E: Logistics and Transportation Review, 206. 104541. ISSN: 1366-5545

https://doi.org/10.1016/j.tre.2025.104541

Strategic design of parcel locker networks for urban delivery

Ali Sabzevari Zadeh[a], Iman Dayarian*[a] and Mahdi Bashiri[b]

[a]Culverhouse College of Business, The University of Alabama, Tuscaloosa, Alabama, USA

[b]Leeds University Business School, University of Leeds, Leeds, LS29JT, UK

Abstract

This paper focuses on the strategic design of a Parcel Locker (PL) network in the context of last-mile delivery. While PLs are emerging as a promising alternative to home delivery, their implementation poses practical challenges, including high deployment costs, inefficient capacity utilization, replenishment scheduling, and coordination with end users. To address this, we consider a network consisting of a depot, candidate locations for PLs, and service regions, each representing a group of customers with known aggregate daily demand distributed across multiple periods. Each period corresponds to a time window preferred by some customers for parcel pickup. The problem extends the classical location-routing problem by incorporating a multi-period horizon, a two-echelon network structure, and flexible replenishment planning. The objective is to minimize upfront strategic costs and anticipated operational costs over the network's economic life. Strategic decisions concern selecting locations and capacities for PLs, and the size of a fleet to replenish PLs from the depot. Operational decisions, which can be revisited after the network is deployed, involve planning replenishment and routes over a single-day, multi-period horizon. The problem is formulated as a mixed-integer linear program and solved using a variable neighborhood search embedded in a branch-and-cut framework, leveraging tailored valid inequalities and local search moves. Computational experiments, along with a case study in Chicago using real data from Amazon, demonstrate that incorporating expected operational costs into network design process leads to more cost-efficient PL networks. Moreover, flexibility in replenishment planning affects both operational cost and network configuration, leading to further improvements in overall system efficiency.

Keywords: Parcel locker network design, Location-routing problem, Flexible replenishment, Last-mile delivery, Variable neighborhood search

*Corresponding author

E-mail address: idayarian@ua.edu (I. Dayarian)

1 Introduction

In the last decade, global retail e-commerce sales have increased more than fourfold, growing from $1.34 trillion in 2014 to $6.00 trillion in 2024 (Statista, 2025). This surge reflects the intense competition among major retailers seeking to increase their online sales by offering fast, efficient and secure delivery services while ensuring affordability for a broader customer base. In the rapidly evolving landscape of e-commerce, last-mile delivery — the final point of contact between retailers and consumers — plays a pivotal role by directly impacting both customer satisfaction (Vakulenko et al., 2018) and delivery costs (Vanelslander et al., 2013). However, it remains one of the most challenging and costly stages, accounting for nearly 53% of total delivery expenses (Contimod, 2025) and up to 40% of the product price (Barenji et al., 2019). The rising demand for fast delivery services (Joerss et al., 2016) further intensifies these challenges, placing pressure on service providers to balance speed, cost-efficiency, and sustainability. According to an analysis by Muñoz-Villamizar et al. (2021), fast shipping options can increase total delivery costs and carbon emissions by up to 68% and 15%, respectively. This can be attributed to the need for a larger fleet and increased traffic congestion (Boysen et al., 2021). In light of these challenges, ensuring a positive customer experience requires convenient, cost-effective, fast, and secure delivery options that accommodate diverse consumer preferences.

A range of delivery modes exists for last-mile delivery (Savelsbergh and Van Woensel, 2016). On one end of the spectrum, the delivery company handles transportation entirely, as with home delivery (HD). However, the convenience of HD, preferred by most online customers (LaserShip, 2022), comes with a cost, making it one of the most expensive modes. A case study in Europe, for instance, indicates that the per-parcel delivery cost is approximately 9 to 12 times higher for HD than for in-store pickup (Hovi and Bo, 2024). Moreover, limited parking availability (Ranjbari et al., 2023), failed delivery attempts due to customer unavailability (Goodchild et al., 2019), and missing parcels left unattended on doorsteps (Guardian, 2023) have made this mode more costly and risky, especially for high-value items. At the other end of the spectrum, customers place online orders and travel to a store to collect their items, as in buy online, pick up in store (BOPIS) or curbside pickup (CP), offered by retailers such as Walmart, Target, and Best Buy. Note that in both BOPIS and CP, customers place orders online. In BOPIS, they enter the store to collect items from a designated pickup counter, while in CP, they remain in their vehicles until store staff bring the order to them outside, usually in a designated parking spot. Because of its cost efficiency, BOPIS has expanded rapidly: the share of stores offering the service grew from 44% in 2016 (Magana, 2019) to around 83% in 2023 (Digital Commerce 360, 2023). Although more profitable than other e-commerce channels (Ketzenberg and Akturk, 2021), BOPIS relies on in-store inventory and is not viable for retailers without a physical presence in the market, such as Amazon and eBay. Furthermore, BOPIS can shift part of the demand from the online channel to the store channel, where profit margins are often lower (Gao and Su, 2017; Akturk et al., 2018). In addition, the convenience of BOPIS largely depends on the number and geographic locations of stores offering this service within a city.

To address the challenges highlighted above, this paper explores an alternative delivery mode: parcel locker (PL) delivery. Like CP, customers place online orders, but instead of picking them up at stores, they retrieve them from self-service PLs. Indeed, if HD and CP represent two extremes, PL delivery lies in between, combining advantages of both while remaining feasible for e-retailers without physical storefronts. Implementing a PL network offers ample opportunities. Compared to HD, PL delivery improves cost efficiency through parcel consolidation (i.e., multiple parcels per drop), reduces shipping costs, eliminates failed deliveries, and enhances customer experience with secure, all-day pickup options (LJMGroup, 2024). Figure 1 presents illustrative examples of three delivery modes: home delivery, PL delivery, and BOPIS.

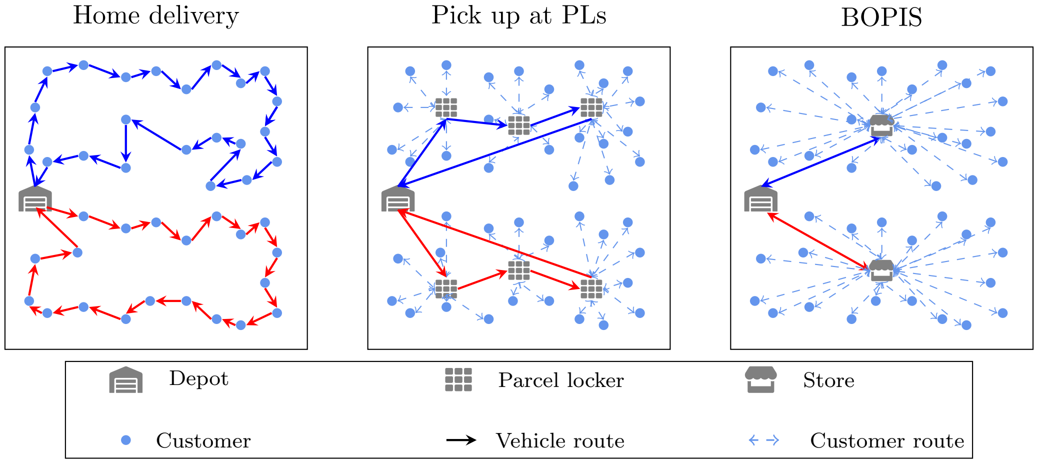


Figure 1: Illustrative examples of HD, delivery through PLs, and BOPIS.

Companies determining locations for PL installations rely heavily on location intelligence and consumer behavior analysis. They target high-foot-traffic areas such as supermarkets, transit hubs, shopping centers, schools, hospitals, and corporate zones to ensure visibility and convenience (DataAppeal, 2024). Accessibility factors (e.g., proximity to public transportation, pedestrian routes, and parking) are essential to maximize usage (Schaefer and Figliozzi, 2021). Demand forecasting and demographic profiling (e.g., residential density, internet usage, income levels) help tailor placements to areas with a higher likelihood of adoption (CleverMaps, 2024). Firms also assess competitive saturation and partnership possibilities by placing PLs near retail partners or in underserved spaces to gain a strategic advantage (Telling, 2025). Together, these considerations ensure that PL placements are not only convenient but optimized for both user adoption and operational efficiency.

Planning and implementing a PL delivery network presents a complex set of challenges for logistics companies. The growing volume and variety of parcels, rising consumer expectations (e.g., same-day delivery), and environmental concerns have driven the adoption of PLs as a more efficient last-mile solution (Peppel et al., 2024). However, successful implementation is far from straightforward. It requires overcoming significant barriers, including relatively high initial investment costs (a case study by Ozyavas et al. (2025) estimates the construction cost of a 60-locker PL at approximately \$23,000 in Groningen, the Netherlands), space limitations in densely populated areas, and the inability to accommodate large or irregularly shaped packages (LJMGroup, 2024). Designing the network also involves complex decisions regarding the number, configuration, and placement of lockers, all while managing a variety of operational expenses such as installation, maintenance, software, and connectivity. Furthermore, capacity utilization, demand forecasting, courier workflows, and user behavior must be considered to ensure optimal locker density and efficient performance (Rohmer and Gendron, 2020). Strategic site selection (i.e., balancing accessibility, safety, and demographic fit) is essential for maximizing usage. Beyond logistical hurdles, companies must also navigate regulatory constraints and slow consumer adoption, often influenced by entrenched preferences for home delivery (LJMGroup, 2024). Addressing these interconnected challenges holistically is vital to achieving a scalable and sustainable PL network.

Several major companies have implemented tailored versions of PL delivery services. Amazon, for example, operates its "Hub Locker" system in select U.S. cities. Given that Amazon mostly delivers small packages (< 5 lbs) (Guglielmo, 2013), utilizing PLs presents significant opportunities for cost savings and operational efficiency. UPS offers the "Access Point Locker" network, and FedEx provides a similar service through its "Delivery Manager". Other companies, such as DHL, Walmart, and SF Express have also introduced their own solutions, i.e., the "DHL Locker", "Walmart Pickup Tower", and "SF Locker Self-Drop". Table 1 presents a comparison of last-mile delivery modes, including BOPIS/CP, PL, and HD, based on several criteria from both the retailer and customer perspectives.

Table 1: Comparison of last-mile delivery modes.

Criterion	BOPIS/CP	PL	HD
Customer convenience	Low	Moderate	High
Distance from customers	Long	Short	Zero
Transportation cost (customer)	High	Low	Zero
Transportation cost (retailer)	Low	Moderate	High
Risk of missing parcel	Zero	Zero	Moderate
Risk of failed delivery	Zero	Zero	Low
Infrastructure cost	Low	High	Zero
Vehicle fleet size	Small	Moderate	Large
Requires physical store(s)?	Yes	No	No

Building on the growing trend of self-service delivery systems, this paper aims to design a smart PL network for a retailer serving a specific geographic area (e.g., a city or a state). A PL consists of several locker boxes, each assigned to a single order, which can be accessed using single-use PINs provided by the carrier. The network design problem, referred to as the Parcel Locker Location-Routing Problem (PL-LRP), consists of a two-echelon network comprising a central depot, a set of candidate PL locations, and a set of service regions (SRs). Each SR represents a part of the geographic area (e.g., one or more adjacent city blocks) and includes a number of customers with known aggregate daily demand distributed across multiple periods. Here, a period refers to a pickup time window during which a group of customers, based on their preferences and daily schedules, chooses to collect their parcels from assigned PLs. In practice, customers place online orders and select a pickup period from a predefined set of options. Parcels are then dispatched from the depot to the corresponding PLs before their respective pickup times, using a fleet of vehicles. During the scheduled periods, customers individually travel to their assigned PLs to retrieve their parcels.

The PL-LRP focuses on the strategic decisions involving PL locations, capacities, and vehicle fleet size. However, these strategic decisions are closely intertwined with operational ones made after network deployment, such as SR-to-PL assignments, PL replenishment plans, and vehicle routing over a single-day, multi-period horizon. Here, a PL's replenishment plan specifies both the schedule and quantity, that is, the periods it must be replenished and the number of orders to be delivered at each visit. Unlike classical (multi-period) LRPs, where replenishment plans are known a priori, our model allows flexibility in both dimensions: the retailer must decide when (i.e., in which periods) to replenish each active PL and how much to deliver in each visit. To reflect the long-term implications of strategic decisions, we incorporate a close approximation of operational costs over the economic life of the network, seeking an optimum balance between upfront investment and recurring future costs. It is worth noticing that the retailer is not required to adhere to the operational decisions made during network design. In practice, once the network is operational, these decisions can be revisited and adjusted in response to changes in the demand pattern that might occur over time. Thus, the consideration of operational decisions at this stage is mainly to guide the strategic decision-making process and, in turn, support the design of a cost-effective network. This paper presents a mathematical formulation and a solution approach that jointly consider strategic and operational aspects of the PL network. By accounting for the downstream impact of strategic decisions, our model ensures that the weighted sum of initial investments and estimated routing costs over the network's economic life cycle is minimized.

The contributions of this work are summarized as follows. (1) We introduce the PL-LRP as a new real-world variant of the LRP that incorporates flexibility in PL replenishment plans, enabling retailers to design efficient PL networks for last-mile delivery. Most existing publications focus on single-period LRPs (e.g., Janinhoff and Klein, 2023; Ozyavas et al., 2025), while those addressing multi-period LRPs often assume fixed replenishment plans (e.g., Grabenschweiger et al., 2022), unlike the PL-LRP, which allows flexibility in both. (2) We propose a mixed-integer linear programming model for the PL-LRP, which inherits key features from the flexible periodic vehicle routing problem, inventory routing problem, production-routing problem (PRP), and location-routing problem. In Section 2, we discuss the similarities and differences between the PL-LRP and each of these classical problems. (3) We develop an exact solution approach along with several families of valid inequalities, that effectively incorporates problem-specific local search moves into a branch-and-cut algorithm to improve the incumbent solution and best bound. (4) We demonstrate the performance of the proposed algorithm and the impact of its components through an extensive computational study, including a case study in the city of Chicago, IL using the real data from Amazon. To our knowledge, this is the first study to address a two-echelon, multi-period LRP with flexible replenishment planning, where demand at each PL depends on the SRs assigned to it. While the LRP literature is extensive, existing models require significant modifications to simultaneously handle all the decisions outlined. Although our approach focuses on minimizing strategic and estimated operational costs, we also quantify the social and environmental impacts of PL delivery in contrast to HD and CP.

The remainder of the paper is structured as follows. Section 2 summarizes relevant publications in a literature review. Section 3 outlines the formulation of the PL-LRP. Details of the solution methodology are discussed in Section 4. Section 5 presents results from numerical experiments and a case study in Chicago to evaluate the proposed model and solution approach. While our main focus is on costs, we also examine the social and environmental impacts of PL delivery compared to the HD and CP, which offers managerial insights. The paper concludes with final remarks, managerial insights, and suggestions for future research in Section 6.

2 Literature review

This section presents a wide-ranging review of the related studies, covering the inventory-routing problem (Section 2.1), the periodic vehicle routing problem (Section 2.2), and the location-routing problem (Section 2.3), including parcel locker network design problems (Section 2.3.1).

2.1 Inventory-routing problem

The inventory-routing problem (IRP), first introduced by Bell et al. (1983), is a class of multi-period vehicle routing problem where demand is known but it varies from period to period, with the stock at customers being replenished periodically by a supplier. The problem is to identify the replenishment plan to serve each customer, the quantity to deliver at each visit, and the vehicle route(s) to replenish customers, so that shipping and inventory holding costs are minimized while back-orders and stock-outs are avoided. For more details on IRP models, characteristics and variants, refer to the survey conducted by Coelho et al. (2014).

A majority of recent studies in IRP literature is focused on developing both exact and heuristic solution methods. Archetti et al. (2007) are pioneers in proposing an exact approach for the vendor-managed IRP. They developed a branch-and-bound (BB) algorithm and analyzed how distribution and holding costs are affected by inventory control policies, with the results demonstrating the superiority of the maximum stock level (ML) policy over the order-up-to level (OU) policy. Note that under the ML policy, the supplier determines when and how much to deliver to each customer over a multi-period horizon, subject to vehicle capacity and customer storage limit. The key feature of this policy is the flexibility in delivery quantities, which can range from zero up to the customer's maximum storage capacity. This allows for more efficient route planning and delivery consolidation. In contrast, the OU policy requires replenishing each customer's inventory to a fixed target level upon delivery, which restricts flexibility and can result in less efficient routing decisions. Since then, IRP scholars have placed increased emphasis on the multi-period replenishment policy, referred to as IRP-ML, due to its flexibility and cost-efficiency (e.g., Avella et al., 2018; Manousakis et al., 2021; Skålnes et al., 2022; Schenekemberg et al., 2024).

On the other hand, there has been growing interest in matheuristic algorithms for the IRP-ML in recent years. Su et al. (2020) investigated a practical multi-depot, multi-vehicle IRP-ML inspired by a real-world case, incorporating a heterogeneous fleet, heterogeneous drivers, and time windows. A route-first, cluster-second approach is proposed in Vadseth et al. (2021) for the classical IRP-ML. Touzout et al. (2022) addressed a variant, using a decomposition-based matheuristic, in which the travel time between each pair of nodes depends on the departure time. A two-phase matheuristic approach based on tabu search is developed by Charaf et al. (2024) for a two-echelon IRP-ML, where routes must be planned separately for each echelon. Najy et al. (2025) presented a matheuristic based on column generation with fixed routes and consistent driver assignments, aiming to enhance operational efficiency and delivery consistency.

Overall, the PL-LRP and IRP-ML are similar in periodicity: in both, replenishment plans are not predefined, i.e., there is no limit on the number of visits per customer (or PL in our setting), and the replenishment quantity per visit is also a decision variable. However, unlike IRP-ML, decisions regarding the location and capacity of PLs, as well as SR-to-PL assignments, must be made, adding complexity due to these additional layers of decision-making.

2.2 Periodic vehicle routing problem

The periodic vehicle routing problem (PVRP), originally proposed by Beltrami and Bodin (1974), extends the capacitated vehicle routing problem (VRP) by selecting a replenishment schedule for each customer—whose set of possible schedules is defined a priori—and replenishing them accordingly. Unlike in the IRP, the replenishment quantity for each customer remains fixed across visits. The objective is to assign a replenishment schedule to each customer and plan vehicle routes to minimize distribution costs. The interested reader is referred to the survey by Campbell and Wilson (2014) for further information on PVRP applications, solution approaches, variants, and extensions.

Several studies have applied or developed exact algorithms to solve different variants of the PVRP. Archetti et al. (2017) introduced the Flexible PVRP (FPVRP), where, similar to the IRP-ML but unlike the classical PVRP, replenishment plans are not fixed in advance. They used the commercial solver CPLEX to solve the problem instances. Rodríguez-Martín et al. (2019) studied a variant in which each customer must be served by the same vehicle and driver across all visits and proposed a branch-and-cut (BC) algorithm. This work was later extended by Rodríguez-Martín and Yaman (2022), who incorporated service times into the model and developed three BC algorithms, primarily based on Benders decomposition. More recently, Basir et al. (2024) conducted a comprehensive computational study using Gurobi to compare different formulations for the PVRP and its time window variant, strengthening the models with valid inequalities to improve performance. Other studies have focused on heuristic methods to address new variants of the PVRP. Motivated by a real-world application, Cantu-Funes et al. (2017) addressed a multi-depot PVRP with due dates and time windows, proposing a greedy randomized adaptive search procedure. Similarly, Asgharyar et al. (2025) studied a PVRP with cross-docking and developed both a variable neighborhood search (VNS) algorithm and a population-based VNS approach to solve the problem efficiently.

Review of the PVRP literature indicates that periodicity in previous studies mostly emerged from pre-defined replenishment schedules, with fixed replenishment quantities per visit. However, limited flexibility in replenishment plans can lead to sub-optimality in the PL-LRP, negatively impacting both fixed opening and routing costs. This occurs because, in our setting, replenishment quantities are allowed to vary across visits to each PL, extending the case where quantities delivered are fixed per visit. To the authors' knowledge, no existing work in the PVRP context, except for FPVRP, allows full flexibility in replenishment plans, requiring simultaneous planning of replenishment and routes.

2.3 Location-routing problem

In this subsection, we review the broad class of location-routing problems (LRPs), starting with publications closely related to parcel locker network design (PLND), followed by less related work including periodic LRPs and multi-echelon LRPs.

2.3.1 Parcel locker network design

The literature on PLND can be categorized into two main streams: studies that focus solely on locating PLs without accounting for routing costs from the depot to PLs, mostly resulting in facility location problems (e.g., Lin et al. 2020, 2022b,a; Kahr 2022; Mancini et al. 2023; Zhou et al. 2024; Ma et al. 2024); and studies that explicitly incorporate routing decisions into the PL network design, taking the form of LRPs (e.g., Enthoven et al. 2020; Wang et al. 2022; Grabenschweiger et al. 2022; Janinhoff and Klein 2023; Ozyavas et al. 2025). Since the latter more closely align with our work, we focus our review on LRPs addressing PLND. Enthoven et al. (2020) studied a 2E-LRP and proposed an daptive large neighborhood search (ALNS) for designing a network supporting HD and PL delivery, where parcels are delivered from a depot to either PLs or satellites. Location decisions involve selecting a subset of PLs and satellites to open. Each PL has a defined covering area; customers within it must pick up their orders from PL, while those outside are served by cargo bikes routed from satellites. A multi-period LRP is addressed in Grabenschweiger et al. (2022) to select a subset of PLs as an alternative to attended HD, where customers can be served either at home or at PLs during their desired time windows. If customers are served at PLs, they receive a fixed compensation, regardless of their distance from the assigned PL. The authors developed a MILP model and solved it using a heuristic approach. Janinhoff and Klein (2023) investigated a multiple trip LRP with delivery options and customer choice, where both the set of customers and their choices are uncertain. The problem is formulated as a two-stage stochastic program and solved using a two-phase hierarchical heuristic, which first solves a surrogate model to select a subset of PLs to open, and then applies an ALNS to solve the resulting VRPs. A PLND problem is formulated as a LRP by Ozyavas et al. (2025), where cargo bike routes serving HD customers start and end at PLs, but no routes are planned from the depot to PLs, unlike in the PL-LRP. The problem is solved using a branch-and-price algorithm.

Overall, except for Grabenschweiger et al. (2022), existing studies often consider single-period LRPs for PLND, distinguishing them from our multi-period setting. Moreover, although these studies consider routing costs in PL network design, none of them addresses fleet sizing, inventory, and replenishment, key factors closely linked to location and routing, and only two studies (i.e., Wang et al. 2022; Ozyavas et al. 2025) incorporate capacity and assignment decisions. Beyond solution approaches, the PL-LRP also differs from the multi-period LRP in Grabenschweiger et al. (2022), mainly because, unlike our setting but similar to the PVRP, replenishment is preplanned, inventory decisions are excluded, and the network is single-echelon. In fact, the PL-LRP develops a replenishment plan for each PL by determining the periods in which the PL is replenished and the corresponding replenishment quantities, whereas the problem in Grabenschweiger et al. (2022) assumes that replenishment plans are predefined input parameters. Table 2 summarizes related studies that address PLND using LRP variants. Note that in some studies (e.g., Naji-Azimi et al., 2012), pickup points are referred to as satellites, serving a role similar to PLs in our setting. Compared to other studies, the PL-LRP is the only one that simultaneously addresses a multi-echelon, multi-period LRP with unknown fleet size, while endogenously planning replenishment, making it the most comprehensive model for PLND. Later in the case study (Section 5.5), we demonstrate how flexibility in replenishment planning can reduce costs.

Table 2: Summary of closely related LRP studies on parcel locker network design.

Studies	Network structure		Decision variables						
	Multi-echelon	Multi-period	Location	Capacity	Fleet size	Routing	Assignment	Inventory	Replenishment
Naji-Azimi et al. (2012)	✓	-	✓	-	-	✓	✓	-	-
Veenstra et al. (2018)	✓	-	✓	-	-	✓	✓	-	-
Huang et al. (2019)	-	-	✓	-	-	✓	✓	-	-
Enthoven et al. (2020)	✓	-	✓	-	-	✓	-	-	-
Wang et al. (2022)	-	-	✓	✓	-	✓	✓	-	-
Grabenschweiger et al. (2022)	-	✓	✓	-	-	✓	-	-	-
Janinhoff and Klein (2023)	-	-	✓	-	-	✓	-	-	-
Ozyavas et al. (2025)	✓	-	✓	✓	-	✓	✓	-	-
This study	✓	✓	✓	✓	✓	✓	✓	✓	✓

2.3.2 Other network design problems

The periodic location-routing problem (PLRP) is a variant of the LRP that integrates elements of both the PVRP and LRP, incorporating additional decisions on replenishment schedules selected from customer-specified options. Hemmelmayr (2015) developed a large neighborhood search algorithm for the PLRP, which adopts operators arbitrarily irrespective of their performance. Later, an ALNS was developed by Hemmelmayr et al. (2017) for an extended version of the PLRP where a set of eligible customers are considered as candidate depots and customer visit frequency is not a priori and must be decided. Another PLRP was investigated by Savaşer and Kara (2022) in the context of mobile healthcare services in rural areas, and was solved using a cluster-first, route-second approach. Apart from their limited flexibility in replenishment plans, these multi-period models are generally single-echelon and exclude secondary or intermediate facilities; thus, they overlook SR-to-PL assignment decisions that itself falls in the class of location-allocation problem which is NP-hard.

Another related research area is the multi-echelon LRP, or N-echelon LRP (NE-LRP). As discussed by Mohamed et al. (2023), NE-LRPs can be approached hierarchically or simultaneously: the former uses crude routing cost approximations to make strategic decisions (e.g., location and capacity), requiring less computational effort, while the latter integrates routing decisions explicitly, allowing more accurate cost assessment but at higher computational cost. That said, hierarchical approaches help account for demand uncertainty during the network design phase. Janjevic et al. (2019) adopted a simultaneous method to design a distribution network, where customers can be served either at home or at collection-and-delivery points. A special case of 2E-LRP is studied in Faugère et al. (2022), where access hub locations and capacities are fixed, but modular capacities must be dynamically allocated over a multi-period horizon to address demand fluctuations. Escobar-Vargas and Crainic (2024) addressed a multi-attribute 2E-LRP that requires vehicle synchronization at intermediate facilities due to the lack of storage capacity, and proposed a discovery-based exact algorithm to solve it.

Figure 2 illustrates how the PL-LRP relates to classical, well-studied problems. First, compared to the lot-sizing problem with direct shipment, the PL-LRP does not involve decisions on production quantity and schedule; instead, it introduces additional complexity through PL location, replenishment planning and vehicle routing decisions. Second, beyond the complexity of the classical LRP, the PL-LRP also incorporates inventory and replenishment planning decisions for PLs. Third, it encompasses PL location and sizing decisions, as well as all the decisions addressed in the IRP. Lastly, the PL-LRP shares similarities with the PRP because, in both problems, replenishment plans are not predefined and must be decided; however, they differ in that the PL-LRP includes location and capacity decisions, while the PRP addresses decisions on production quantity and schedule. Therefore, although the PL-LRP inherits key features from these classical problems, it remains significantly different from each, underscoring its combinatorial complexity. In particular, the need to jointly optimize location, capacity, SR-to-PL assignment, inventory, replenishment planning, and routing decisions over multiple periods substantially expands the solution space, making the problem more challenging to solve than any of its classical counterparts.

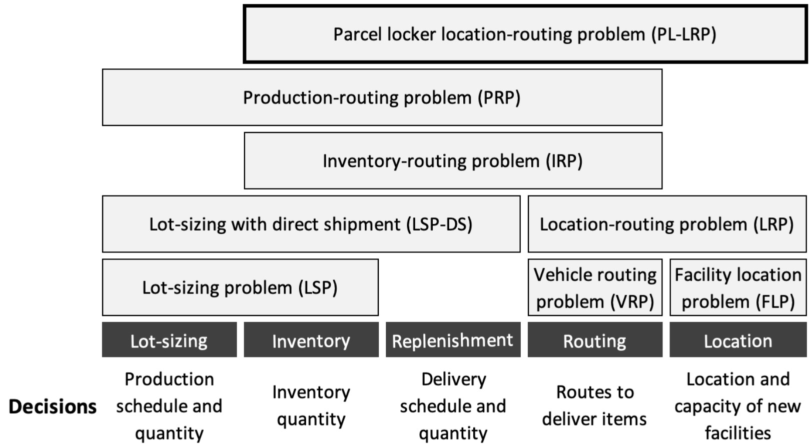


Figure 2: Relevant supply chain planning problems.

3 Problem description and modeling

The PL-LRP primarily focuses on strategic planning by designing a PL network, which includes decisions on the locations and capacities of PLs, as well as the vehicle fleet size. However, these strategic decisions are interconnected with operational decisions, such as customer assignments, replenishment planning, and routing. To account for this, we consider a reasonable approximation of operational costs based on a high-level aggregate demand over the PLs' economic life cycle, aiming to balance the tradeoff between one-time strategic costs and ongoing operational costs. In what follows, we describe the PL-LRP in detail and present its formulation as a mixed-integer linear programming model.

3.1 Problem description

The PL-LRP models a retailer offering last-mile delivery via PLs in an urban area. The planning horizon spans a day (or part of a day) divided into a set of non-overlapping periods (e.g., two-hour intervals), denoted by $\mathcal{T}$. The retailer guarantees next-day delivery, meaning all orders placed before a cutoff time are delivered to PLs before the customer-specified pickup periods the following day.

The urban area is divided into service regions (SRs), each potentially containing several customers. There is a set of candidate sites for PL installation. PLs have modular capacities, with module set $\mathcal{M}$; opening a PL of capacity K_m, $m \in \mathcal{M}$, at site i incurs a fixed cost F_{im}. Orders are prepared at a central depot (or fulfillment center) (node 0), then delivered by carriers to PLs ahead of scheduled pickup times. Define the node set $\mathcal{N} = \mathcal{N}_0' \cup \mathcal{N}_1$, where $\mathcal{N}_0' = 0 \cup \mathcal{N}_0$ includes the depot and candidate PLs, and $\mathcal{N}_1$ is the set of SRs. The objective is to decide which PLs to open, assign SRs to them, and plan vehicle routes so customers can retrieve orders on time from the assigned PLs.

A homogeneous vehicle fleet, whose size must be determined, starts and ends its multi-trip routes at the depot. Each vehicle has a capacity Q, incurs a fixed cost FV, and a variable cost α per unit of distance. Let E_{ij} denote the distance between nodes i and j. Routes and delivery quantities to PLs must be scheduled to ensure all orders are in place before the respective pickup periods. A long-enough travel time (e.g., one hour) between periods is assumed to account for transit and locker loading, adjustable based on traffic, distances, and locker/truck capacities. A route length limit can also be imposed to respect delivery time constraints. Early shipments are allowed, giving flexibility in scheduling. For example, in a three-period horizon, orders for period 3 can be delivered anytime before that period, subject to PL capacity. Distant PLs may be visited less frequently (e.g., once daily), while nearby ones can be restocked more often.

Demand is deterministic but time-varying and can be forecasted using SR-level data such as population, past order rates, and pickup preferences. Let D_{kt} represent demand from SR k in period t. SR-PL assignments depend on PL capacity, delivery frequency, and route scheduling. For example, a large but infrequently replenished PL may serve fewer SRs than a smaller one with more frequent deliveries. To encourage customers to choose PL delivery over attended home delivery, a small compensation proportional to their distance from the assigned PL is offered—an approach commonly adopted in last-mile delivery studies (e.g., Grabenschweiger et al., 2021; Mancini and Gansterer, 2021). The per-unit compensation is denoted by β.

The objective of the PL-LRP is to minimize total costs, including fixed costs for opening PLs, employing a fleet of vehicles, traveling costs of vehicles for PLs' replenishment, and total compensation paid to customers. To integrate strategic and operational decisions, we scale cost components over an economic life (e.g., five years) for PLs and vehicles, representing the duration they are expected to be economically viable and profitable. This involves multiplying transportation costs and compensation by the number of working days during this period of time. Thus, we consider α and β as the scaled parameters for shipping cost and compensation, respectively. Figure 3 illustrates a PL-LRP instance with 11 candidate PLs, 29 SRs, 2 periods, and 4 modules of capacity. Strategic decisions identify 7 PLs to be opened: one with 4 capacity modules, one with 3, three with 2, and two with 1. They also specify 2 vehicles to be employed for fleet setup. Operational decisions involve assigning SRs to opened PLs, planning replenishment schedules, designing vehicle routes to replenish PLs, and managing replenishment quantities and inventory levels. Notice that orders are customer-specific, and if placed in locker boxes ahead of time, they are considered as end-of-period inventory.

A solution to the PL-LRP supports strategic decisions, which guide the design of the PL network, while accounting for interconnected operational decisions that can be revisited after the network is deployed. The strategic design decisions include the location and capacity of PLs to be opened, as well as the vehicle fleet size, while the uncommitted operational decisions consist of SR-to-PL assignments, replenishment plans for PLs, and vehicle routes from the depot to replenish PLs in each period. These decisions are aimed at minimizing the total strategic and operational costs over the economic life of the PLs and vehicles. The modeling assumptions are: (a) backorder is not allowed, requiring orders to be delivered to PLs before customers' selected pickup periods; (b) each PL can be replenished by one vehicle in each period, but potentially by multiple vehicles over the planning horizon (i.e., a day in this paper); (c) each vehicle is allowed to perform at most one route per period, but potentially multiple routes throughout the single-day planning horizon; (d) locker boxes are large enough to accommodate any orders; (e) operating hours are consistent across all PLs, that is, they operate during periods with fixed and similar start and end times (e.g., Period 1 from 8:00-10:00, replenishment time from 10:00-11:00, Period 2 from 11:00-13:00, etc.).

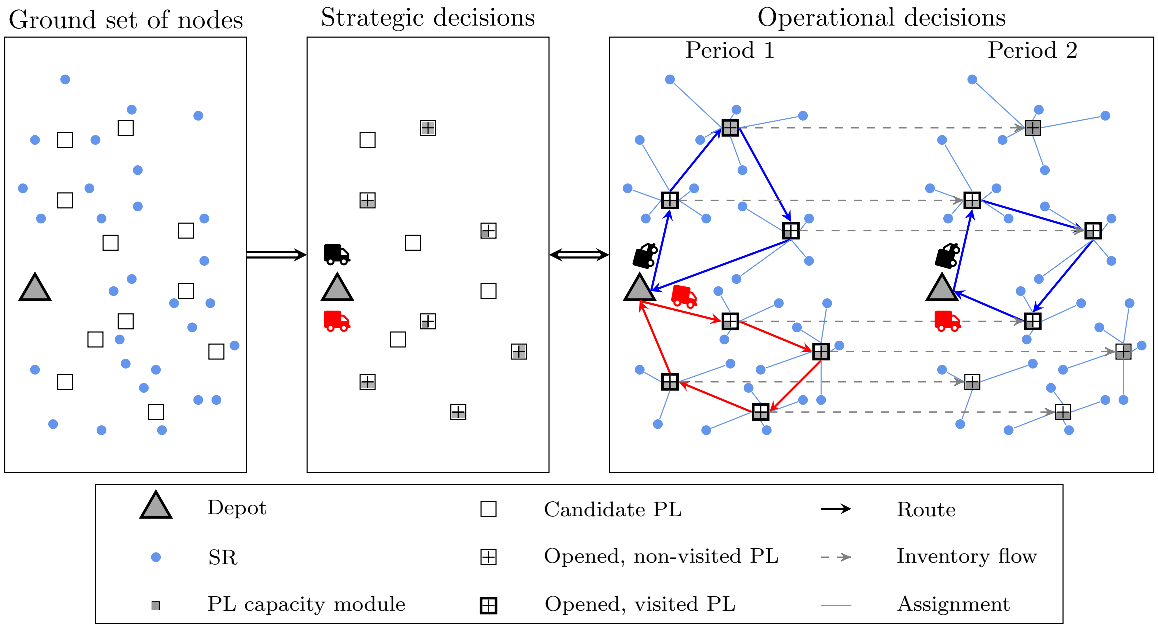


Figure 3: Graphical illustration of a PL-LRP instance.

3.2 Problem formulation

Let binary variable z_{im} take value 1 if a module m PL is opened at the candidate site i; 0 otherwise. Binary variable y_{ik} indicates whether SR k is assigned to PL i. Binary variable x_{ijt} takes value 1 if a vehicle travels from node i to j in period t; 0 otherwise $(i, j \in \mathcal{N}_0')$. Binary variable v_{it} indicates whether PL i is replenished in period t. Decision variable w_{it} represents the vehicle load just before visiting PL i in period t, while variable u corresponds to the number of vehicles used for shipping parcels from the depot to PLs (i.e., the fleet size). Decision variable q_{it} represents the replenishment quantity at PL i in period t, and variable I_{it} denotes the inventory at PL i at the end of period t. Table 3 summarizes the notation used in formulating the PL-LRP.

Table 3: List of notation

Sets	
$\mathcal{T}$	Set of (pickup) periods in the single-day planning horizon
$\mathcal{M}$	Set of PL capacity modules (e.g., {50, 100, 150} lockers)
$\mathcal{N}_0$	Set of candidate PLs
$\mathcal{N}_0'$	Set of first-echelon nodes; $\mathcal{N}_0' = \{0\} \cup \mathcal{N}_0$; 0 denotes the depot
$\mathcal{N}_1$	Set of SRs
$\mathcal{N}$	Set of all nodes; $\mathcal{N} = \mathcal{N}_0' \cup \mathcal{N}_1$
Parameters	
K_m	Number of lockers in capacity module m
Q	Capacity of a delivery vehicle
F_{im}	Fixed cost of opening a PL with capacity K_m at candidate site i
FV	Fixed cost of employing a delivery vehicle
α	Travel cost incurred by a delivery vehicle per unit distance (scaled over the economic life of PLs)
β	Compensation paid to a customer per unit of distance (scaled over the economic life of PLs)
E_{ij}	Travel distance between nodes i and j
D_{kt}	Demand from SR k in period t (in number of orders)
$\tilde{M}_t$	A large positive value; $\tilde{M}_t = \min\{Q\ ,\ K_M\ ,\ \sum_{k\in\mathcal{N}_1}\sum_{\tau=t}^{T} D_{k\tau}\}$
Decision variables	
Strategic	
z_{im}	Binary variable equal to 1 if a module m PL is opened at the candidate site i; 0 otherwise
u	Number of vehicles used for shipping parcels from the depot to PLs (i.e., the fleet size)
Operational	
y_{ik}	Binary variable equal to 1 if SR k is assigned to PL i; 0 otherwise
x_{ijt}	Binary variable equal to 1 if a delivery vehicle travels from node i to j in period t; 0 otherwise
v_{it}	Binary variable equal to 1 if PL i is replenished in period t; 0 otherwise
q_{it}	Number of parcels replenished at PL i in period t
w_{it}	Load of the delivery vehicle just before arriving at PL i in period t
I_{it}	Inventory level (number of parcels) at PL i at the end of period t

The mixed-integer linear programming (MILP) model formulated for the PL-LRP follows:

$$\text{Min} \quad \sum_{i\in\mathcal{N}_0}\sum_{m\in\mathcal{M}} F_{im}\, z_{im} + FV\ u \tag{1}$$

$$+ \sum_{i\in\mathcal{N}'_0}\ \sum_{j\in\mathcal{N}'_0\setminus i}\ \sum_{t\in\mathcal{T}} \alpha\, E_{ij}\, x_{ijt} + \sum_{i\in\mathcal{N}_0}\sum_{k\in\mathcal{N}_1}\sum_{t\in\mathcal{T}} \beta\, E_{ik}\, D_{kt}\, y_{ik}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{N}_0} y_{ik} = 1 \qquad \forall k \in \mathcal{N}_1 \tag{2}$$

$$\sum_{m\in\mathcal{M}} z_{im} \le 1 \qquad \forall i \in \mathcal{N}_0 \tag{3}$$

$$I_{it-1} + q_{it} = \sum_{k\in\mathcal{N}_1} D_{kt}\, y_{ik} + I_{it} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{4}$$

$$\sum_{k\in\mathcal{N}_1} D_{kt}\, y_{ik} + I_{it} \le \sum_{m\in\mathcal{M}} K_m\, z_{im} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{5}$$

$$v_{it} \le \sum_{m\in\mathcal{M}} z_{im} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{6}$$

$$q_{it} \le \tilde{M}_t\, v_{it} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{7}$$

$$\sum_{j\in\mathcal{N}_0} x_{0jt} \le u \qquad \forall t \in \mathcal{T} \tag{8}$$

$$\sum_{j\in\mathcal{N}'_0} x_{ijt} = \sum_{j\in\mathcal{N}'_0} x_{jit} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{9}$$

$$\sum_{j\in\mathcal{N}'_0} x_{ijt} = v_{it} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{10}$$

$$w_{it} - w_{jt} \ge q_{it} - \tilde{M}_t\,(1 - x_{ijt}) \qquad \forall i \in \mathcal{N}_0,\ j \in \mathcal{N}'_0,\ t \in \mathcal{T} \tag{11}$$

$$w_{it} \le Q\, v_{it} \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{12}$$

$$x_{ijt} \in \{0,1\},\ w_{it},\ u \ge 0 \qquad \forall i,j \in \mathcal{N}'_0,\ t \in \mathcal{T} \tag{13}$$

$$z_{im},\ y_{ik},\ v_{it} \in \{0,1\} \qquad \forall i \in \mathcal{N}_0,\ k \in \mathcal{N}_1,\ m \in \mathcal{M},\ t \in \mathcal{T} \tag{14}$$

$$I_{i0} = 0,\ I_{it},\ q_{it} \ge 0 \qquad \forall i \in \mathcal{N}_0,\ t \in \mathcal{T} \tag{15}$$

The objective function (1) minimizes strategic and estimated operational costs over the PLs' economic life, incorporating four cost components: the fixed costs of opening PLs, the fixed costs of employing vehicles, the (normalized) shipping costs of orders from the depot to PLs, and the (normalized) compensation paid to customers when picking up their orders from PLs. Here, the first two terms represent strategic costs, and the last two are estimated operational costs over the economic life. Notice that each SR contains multiple customers, some of whom may place orders in a given period; the inclusion of demand as a coefficient in the last term reflects the fact that each customer with an order needs to individually travel to the assigned PL to collect their order. Here, the demand in a given period refers to the number of customers expected to choose that period for order pickup and must therefore be compensated independently. Constraints (2) guarantee that each SR is assigned to exactly one PL. Constraints (3) ensure that no more than one module with certain capacity must allocate to each candidate PL. Note that different modules of capacity are available, yet only one module (e.g., with a capacity of 50 lockers) must be chosen to open a new PL. The inventory flow balance at PLs is represented in constraints (4). Constraints (5) formulate the well-known ML replenishment policy, restricting the total allocated demand and ending inventory to the capacity of PL. Constraints (6) enforce the rule that only opened PLs can be replenished. Constraints (7) state that replenishment can only occur if the corresponding PL is visited by a vehicle; otherwise, the replenishment quantity must be zero. Constraints (8) restrict the number of vehicle departures from the depot to the available fleet size. Constraints (9) and (10) ensure route continuity and restrict each PL to be visited by at most one vehicle, respectively. Constraints (11) keep track of loads on vehicles and, at the same time, impose sub-tour elimination. Constraints (12) limit the vehicle load to ensure it does not exceed the maximum vehicle capacity before visiting any PLs. Finally, constraints (13)-(15) define the domain of the decision variables. Clearly, decision variables v_{it} and q_{it} jointly determine the replenishment plan for PL i.

4 Solution approach

In this section, we present our exact solution method for the PL-LRP formulated in Section 3. The proposed Variable MIP Neighborhood Descent (VMND) algorithm, first introduced by Larrain et al. (2017), builds on VNS by embedding multiple customized neighborhood structures within a branch-and-cut (BC) framework. These neighborhoods, tailored to the problem's structural properties, guide the search toward promising regions of the solution space that the standard BC would not efficiently explore on its own. The key contribution lies in designing a problem-specific VNS and integrating it effectively with the BC algorithm. VNS is efficient for complex combinatorial structures (e.g., Olmez et al. 2022; Asgharyar et al. 2025) and can be seamlessly integrated into BC frameworks, combining its exploration strength with the optimality guarantees and pruning capabilities of BC. In the proposed VMND, this integration leverages the heuristic efficiency of VNS alongside the exact optimization power of BC, yielding a more efficient solution approach than a pure BC algorithm. This advantage is especially relevant given recent advances in commercial solvers like Gurobi and CPLEX. The VMND has been successfully applied to various routing-based problems, including the flexible two-echelon location-routing problem (Darvish et al., 2019), the multi-attribute inventory routing problem (Coelho et al., 2020), and the integrated planning of multimodal operations (Wagenaar et al., 2021).

We propose a VMND algorithm that alternates between two phases: the branch-and-cut phase (BCP) and the local search phase (LSP). The BCP solves the original MIP (1)–(15), strengthened with some valid inequalities, using a branch-and-cut algorithm initialized with a feasible solution. Under certain conditions, the BCP is paused and the incumbent solution is passed to the LSP to explore its neighborhoods. Each neighborhood is explored by solving a sub-MIP using a fix-and-optimize approach, where some binary variables are fixed to their incumbent values and the rest are optimized. Additional constraints may be added, such as enforcing a subset of PLs to remain open while allowing their capacities to be reoptimized. Each sub-MIP is solved within a time limit using the incumbent as an upper bound. If a local search (LS) improves the incumbent, the solution is updated and used to initialize the next LS operator; otherwise, the LSP stops. Once the LSP completes, the updated incumbent is passed back to the BCP, potentially pruning parts of the search tree. This cycle continues until an optimal solution is found or a time limit is reached. Figure 4 illustrates the framework of the proposed VMND algorithm.

In what follows, we first derive several valid inequalities based on the structure of the problem (Section 4.1). These inequalities are applied throughout the solution approach, from the construction of the initial solution to

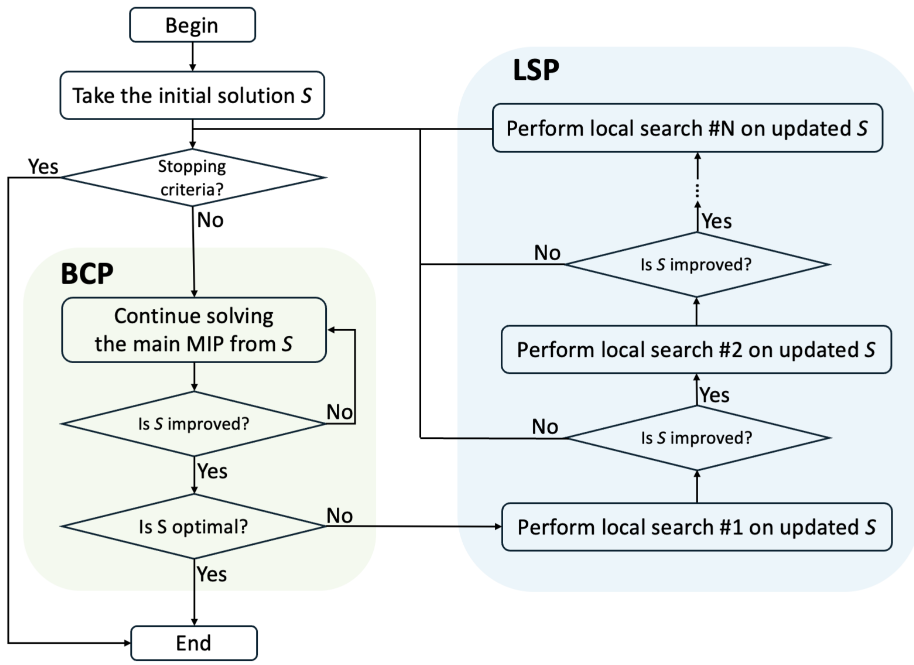


Figure 4: Framework of the proposed VMND algorithm.

the local search phase. Next, in Section 4.2, we describe our constructive heuristic for generating a high-quality initial solution. Note that while the VMND does not require an initial solution—as it starts from the BCP, which can construct one on its own—initializing the BCP with a high-quality solution can improve the overall efficiency of the solution approach. Finally, we present the VMND algorithm in Section 4.3.

4.1 Valid inequalities

Several families of valid inequalities (VIs) are derived to strengthen the MILP model (1)-(15). These VIs are used in the proposed exact solution approach to enhance the lower bound, reduce the search space, and expedite the solution process. The list of the VIs used in our implementations follows.

1. **Capacity inequalities:** The VIs (16), which aggregate constraints (5), place a lower limit on the overall capacity of delivery through PLs per period.

$$\sum_{i\in\mathcal{N}_0}\sum_{m\in\mathcal{M}} K_m\, z_{im} \geq \sum_{k\in\mathcal{N}_1} D_{kt} + \sum_{i\in\mathcal{N}_0} I_{it} \qquad \forall t\in\mathcal{T} \tag{16}$$

2. **Assignment inequalities:** The class of inequalities (17) is valid, imposing upper limits on the number of SRs that can be assigned to PLs. Here, $\underline{D}$ is the minimum demand across all SRs and periods, that is, $\underline{D} = \min_{k\in\mathcal{N}_1,\, t\in\mathcal{T}}\{D_{kt}\}$, and ϵ is a small positive value if $\underline{D} = 0$; 0 otherwise. These VIs state that if $z_{im} = 1$ for any $m \in \mathcal{M}$, then the maximum number of SRs that can be assigned to open PL i is $\min\{|\mathcal{N}_1|, \left\lfloor \frac{K_m}{\underline{D}+\epsilon} \right\rfloor\}$, where $\lfloor a \rfloor$ is the greatest integer less than or equal to a.

$$\sum_{k\in\mathcal{N}_1} y_{ik} \leq \sum_{m\in\mathcal{M}} \min\{|\mathcal{N}_1|, \left\lfloor \frac{K_m}{\underline{D}+\epsilon} \right\rfloor\} z_{im} \qquad \forall i \in \mathcal{N}_0 \tag{17}$$

The relationship between location and assignment variables is strengthened by VIs (18), taking advantage of the fact that no assignment to a PL can be accepted unless it is open.

$$y_{ik} \leq \sum_{m\in\mathcal{M}} z_{im} \qquad \forall i \in \mathcal{N}_0,\ k \in \mathcal{N}_1 \tag{18}$$

3. **Shipping inequalities:** The VIs (19) enforce a minimum number of cumulative vehicle departures from the depot up to each period t, based on the cumulative demand that must be fulfilled by that period and the vehicle capacity Q ($\lceil a \rceil$ is the smallest integer greater than or equal to a).

$$\sum_{j\in\mathcal{N}_0} \sum_{\tau=1}^{t} x_{0j\tau} \geq \left\lceil \frac{\sum_{k\in\mathcal{N}_1} \sum_{\tau=1}^{t} D_{k\tau}}{Q} \right\rceil \qquad \forall t \in \mathcal{T} \tag{19}$$

Valid inequalities (20), adopted from Archetti et al. (2015), impose upper limits on the total number of orders delivered to PLs, ensuring that the cumulative number of orders delivered until a certain period remains below the overall shipping capacity up to that period.

$$\sum_{i\in\mathcal{N}_0} \sum_{\tau=1}^{t} q_{i\tau} \leq t\, Q\, u \qquad \forall t \in \mathcal{T} \tag{20}$$

Valid inequalities (21) establish a direct link between first-period replenishment variables and assignment variables. Here, λ_k is an indicator parameter that is 1 if $D_{k1} > 0$ and 0 otherwise. These VIs ensure that if a SR with non-zero demand in first period is served by a PL, that PL must be replenished in the first period.

$$v_{i1} \geq \lambda_k\, y_{ik} \qquad \forall i \in \mathcal{N}_0,\ k \in \mathcal{N}_1 \tag{21}$$

4.2 Initial solution

The proposed approach is able to solve the PL-LRP without an initial solution, as it begins with a branch-and-cut phase that can generate a feasible solution on its own. However, starting from a good feasible solution can potentially lead to faster convergence, enhancing the algorithm's efficiency. To this end, a decomposition-based *location first-routing second* (LFRS) heuristic is developed to generate an initial solution. In doing so, a reduced version of the MILP (1)-(15) is solved, making all decisions except routing. This is followed by planning routes using the C&W algorithm (Clarke and Wright, 1964). In the reduced model, the fixed cost of employing vehicles is ignored and the routing costs are replaced with approximations in the objective function. Additionally, all the routing-related constraints (8)-(13) are disregarded.

Let parameter $\hat{R}_i$ approximate the distance traveled by a vehicle to replenish candidate PL i, based on the model proposed by Chien (1992), which estimates the TSP tour length as $\hat{L}_i \approx 0.98\sqrt{n\,A_i}$. Here, n denotes the total number of points in the tour. In our implementations, we set $n = \lceil Q/\max_m\{K_m\}\rceil + 1$, and accordingly define the set $\mathcal{P}_i$ to include the depot, PL i, and the $n-2$ closest candidate PLs to PL i. Denote A_i as the area of the smallest rectangle that encompasses all the points in set $\mathcal{P}_i$ (i.e., those included in the tour). The distance traveled by vehicle for feeding PLs from the depot are estimated at $\hat{R}_i = \hat{L}_i/(n-1)\ \forall i \in \mathcal{N}_0$. The **reduced MILP model** used for generating the initial solution is as follows:

$$\text{Min} \quad \sum_{i\in\mathcal{N}_0}\sum_{m\in\mathcal{M}} F_{im}\, z_{im} + \sum_{i\in\mathcal{N}_0}\sum_{t\in\mathcal{T}} \alpha\, \hat{R}_i\, v_{it} + \sum_{i\in\mathcal{N}_0}\sum_{k\in\mathcal{N}_1}\sum_{t\in\mathcal{T}} \beta\, E_{ik}\, D_{kt}\, y_{ik} \tag{22}$$

$$\text{s.t.} \quad (2)-(7),(14)-(16).$$

Preliminary experiments indicate that, of all the valid inequalities presented in Section 4.1, only the inclusion of capacity inequalities results in high-quality feasible solutions for large-sized instances within a short amount of time. Thus, these are the only valid inequalities used in the reduced model. Although less complicated than the original model, the reduced model is still computationally challenging to solve, especially for large-sized instances, as it involves the capacitated facility location problem, which itself is NP-hard. Therefore, it is solved using the branch-and-cut algorithm until either optimality or a time limit μ^I is reached. **Algorithm 1** shows the pseudocode of the LFRS heuristic developed to generate the initial solution, with its steps described below.

In Line 1, the reduced MILP model is solved, using the Gurobi's built-in branch-and-cut algorithm, to determine the locations and capacities of opened PLs ($\boldsymbol{z}$), SR-to-PL assignments ($\boldsymbol{y}$), replenishment schedules ($\boldsymbol{v}$), replenishment quantities ($\boldsymbol{q}$), and end-of-period inventory levels ($\boldsymbol{I}$). In Line 2, these grouped variables are fixed in model (1)-(15) to the values obtained from Line 1. The resulting sub-MIP is indeed a T-period VRP, which requires visiting a subset of opened PLs in each period according to the replenishment schedules (i.e., $\boldsymbol{v}$) and delivering a specified number of orders based on the replenishment quantities (i.e., $\boldsymbol{q}$), using a fleet of vehicle whose size must be determined. In Line 3, this multi-period VRP is decomposed into T regular VRPs (i.e., one VRP for each period) under the assumption that a sufficiently large fleet of capacitated vehicles is available. In Lines 4–6, all the regular VRPs are solved separately using the C&W algorithm—a savings-based heuristic that begins with one route per node and iteratively merges routes based on savings until no further improvement is possible (Clarke and Wright, 1964). Note that the C&W algorithm does not require the fleet size as an input parameter. In Line 7, along with the routing plan for the entire horizon ($\boldsymbol{x}$), the maximum number of vehicles required to replenish the opened PLs across all periods (u) is determined based on the solutions to the period-specific VRPs, that is, $u = \max_t\{u_t\}$, where u_t denotes the number of routes in the solution to the VRP corresponding to period t. In Line 8, all decisions made in Lines 1–7 are combined to generate a complete solution, including the location/capacity, assignment, replenishment plan and inventory decisions made in Line 1, along with the routing and fleet size decisions from Lines 4–7.

An analysis of the time complexity for solving VRPs in the LFRS heuristic and the first local search operator, θ_1, using the CW algorithm follows. Let $|\mathcal{N}_1^t|$ be the number of PLs that must be visited in the resulting VRP in period t. There are $|\mathcal{N}_1^t|(|\mathcal{N}_1^t|-1)/2$ pairs, with a constant time to calculate the saving for each pair, total time complexity for computing savings is $O(|\mathcal{N}_1^t|^2)$. Sorting all pairs according to savings takes $O(|\mathcal{N}_1^t|^2 \log |\mathcal{N}_1^t|)$. There are $O(|\mathcal{N}_1^t|^2)$ pairs and merging the routes can take up to $O(|\mathcal{N}_1^t|)$ per pair, leading to the time complexity of $O(|\mathcal{N}_1^t|^3)$. Hence, a single-period VRP gives the complexity of $O(|\mathcal{N}_1^t|^3)$ (Kučera et al., 2012). Assuming $\mathcal{N}_1^t = \mathcal{N}_1\ \ \forall t \in T$, the CW algorithm requires time $O(T|\mathcal{N}_1|^3)$ to solve a T-period VRP.

Algorithm 1 The constructive LFRS heuristic

$S_L(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{q}, \boldsymbol{I}) \leftarrow$ Solve the reduced model (22); (2)-(7); (14)-(16) within time limit μ^I
Substitute grouped variables $\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{q}$, and $\boldsymbol{I}$ by their values from $S_L(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{q}, \boldsymbol{I})$ in model (1)-(15)
Decompose the resulting sub-problem into T smaller VRPs, one VRP per period
for $t \in \mathcal{T}$ **do**
    $S_R^t(\boldsymbol{x}_t, \boldsymbol{w}_t, u_t) \leftarrow$ Solve VRP(t) using the C&W algorithm
end for
$S_R(\boldsymbol{x}, \boldsymbol{w}, u) \leftarrow \cup_{t=1}^{T} S_R^t(\boldsymbol{x}_t, \boldsymbol{w}_t, u_t)$
$S \leftarrow S_L(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{q}, \boldsymbol{I}) \cup S_R(\boldsymbol{x}, \boldsymbol{w}, u)$
return S

4.3 The VMND algorithm

The VMND algorithm alternates between two phases: BCP and LSP. The BCP solves the original MIP, strengthened with valid inequalities, while the LSP explores neighborhoods via fix-and-optimize sub-MIPs to improve the incumbent solution. Improved solutions from the LSP are fed back into the BCP, and this process repeats until optimality or a time limit is reached. Inspired by the VNS metaheuristic—which combines a local search phase for intensification with a shaking phase for diversification—the VMND adopts this structure by employing the LSP for intensification and the BCP for both intensification and diversification. The VNS relies on a predefined, ordered set of neighborhood structures explored systematically, followed by a random shaking phase to perturb the current solution and escape local optima when the neighborhood operators fail to improve it. This structured approach enables efficient and manageable integration within the BC algorithm, where fast, targeted local searches are essential for improving incumbents during node processing without excessive computational overhead. In our implementation, the BCP acts as the shaking phase, while the LSP performs neighborhood exploration.

Moreover, VNS's predictable neighborhood transitions align well with the incremental improvement strategy of BC, enabling rapid solution refinement and efficient pruning. This compatibility is supported by prior studies integrating VNS within exact methods (e.g., Larrain et al. 2017; Darvish et al. 2019; Coelho et al. 2020; Wagenaar et al. 2021), which demonstrate both its effectiveness and computational efficiency in solving complex routing-based problems. For the PL-LRP, which features a highly complex structure with several interconnected decision variables, VNS provides a practical balance between solution quality and solver integration complexity. Therefore, we identified VNS as the metaheuristic best suited for embedding within the BC framework.

The proposed VMND algorithm introduces several innovations that make it particularly well-suited for solving the PL-LRP. First, unlike existing VMND approaches (e.g., Darvish et al., 2019; Coelho et al., 2020), where the LSP stops after the first improvement, our method allows the LSP to continue as long as local searches yield better solutions, enabling more effective intensification. Second, to avoid inefficient switching between BCP and LSP, the algorithm employs a dynamic threshold for triggering the LSP, ensuring that local search is only invoked when the incumbent solution improves by at least a specified percentage. Indeed, when the incumbent solution improves only marginally (e.g., 0.001%), the likelihood of further improvement in the LSP is significantly lower than when the improvement is more substantial (e.g., 0.1%). This adaptive switching strategy is especially beneficial in early stages of optimization, when larger improvements are more likely. Third, the BCP is warm-started using a high-quality solution generated by a constructive heuristic, which accelerates search tree pruning during initial iterations. Lastly, sub-MIPs in the LSP are constructed adaptively by selectively fixing variables and adding constraints tailored to the structure of the incumbent solution, resulting in more targeted and efficient neighborhood exploration. These innovations, combined with the use of problem-specific valid inequalities and operators, make VMND particularly effective for the PL-LRP, which involves tightly interdependent location, assignment, inventory, replenishment planning, and routing decisions across multiple periods.

In the BCP, we employ Gurobi's built-in BC algorithm, which we enhance in several ways as outlined below. First, each local search operator creates a subproblem by fixing some binary variables while leaving others free, corresponding to a specific node in the search tree. At this stage, we pause the main BC process to solve the subproblem using either a heuristic or Gurobi's built-in BC. In other words, while the original BC is paused, we solve constructed subproblems that focus on promising parts of the solution space. This approach modifies the standard exploration strategy, effectively guiding the solver toward more promising nodes through targeted local search. Second, when the BCP resumes with a new incumbent from the LSP, we modify the built-in BC to update the upper bound and prune nodes without revisiting those already explored, which significantly improves efficiency. It is worth noting that solvers often do not natively support nested optimization procedures, that is, pausing a BC process to solve auxiliary subproblems and subsequently resuming the original BC process with complete state preservation, including restoration of the search tree, explored nodes, node bounds, and cuts. Third, we add valid inequalities (16)–(17) to strengthen the PL-LRP formulation and improve the lower bounds, particularly at the root node. Lastly, we warm-start the BCP with a high-quality solution generated by the constructive LFRS heuristic, which we propose as an alternative, standalone solution approach. These modifications not only improves convergence but also enables the solver to find better solutions within practical time limits, thereby extending the capabilities of the default BC.

Algorithm 2 outlines the main steps of the proposed VMND for the PL-LRP. Here, t^V is a timer that tracks the VMND's runtime. Lines 5-9 correspond to the BCP, while lines 10-19 represent the LSP. These two phases, including their corresponding parameters, are detailed in the following subsections.

4.3.1 Branch-and-cut phase

The VMND begins the optimization process by entering the BCP. An initial feasible solution is used to warm start the BCP procedure. Lines 5-9 in **Algorithm 2** represent the BCP. Here, t^B is a timer that tracks the BCP's runtime, resetting to zero when needed (Line 5). The BCP pauses upon meeting one of the following conditions (Line 6): (i) the incumbent solution improves by at least γ%, or (ii) a time limit μ^B is reached. To avoid frequent switches between BCP and LSP, parameters μ^B and γ are adaptively increased and decreased up to certain levels, respectively. When the BCP pauses, the incumbent solution is transferred to the LSP for neighborhood exploration. Then, two cases can occur: the LSP improves the incumbent solution, or it fails to do so. If the LSP produces a new incumbent solution, the BCP's upper bound is updated while the lower bound remains unchanged. If the LSP fails to produce a new incumbent solution, the BCP resumes exactly from its last termination point without any changes in the upper and lower bounds. Whenever the VMND is in the BCP, the solver keeps solving the main MIP (1)-(15), which is strengthened with valid inequalities (16)-(17),

Algorithm 2 The VMND for PL-LRP

1: $t^V \leftarrow 0$
2: $S \leftarrow$ Generate an initial solution using **Algorithm 1**
3: $S^* \leftarrow S$
4: **while** $t^V < \mu^V$ **and** S^* not optimal **do**
   //Branch-and-cut phase (Lines 5-9)
5: $\quad t^B \leftarrow 0$
6: $\quad$ **while** $t^B < \mu^B$ **or** $S > (1-\gamma)S^*$ **do**
7: $\quad\quad S \leftarrow$ Start/continue solving main MIP (1)-(15); (16)-(17) from S^*
8: $\quad$ **end while**
9: $\quad$ Update S^*
   //Local search phase (Lines 10-19)
10: $\quad$ **for** $\theta \in \mathbf{\Theta}$ **do**
11: $\quad\quad$ Construct sub-MIP(S^*, θ)
12: $\quad\quad S^\theta \leftarrow$ Solve sub-MIP(S^*, θ) within time limit μ^L
13: $\quad\quad$ **if** $f(S^\theta) < f(S^*)$ **then**
14: $\quad\quad\quad S^* \leftarrow S^\theta$
15: $\quad\quad$ **else**
16: $\quad\quad\quad$ **break**
17: $\quad\quad$ **end if**
18: $\quad$ **end for**
19: $\quad$ Update μ^B and γ
20: **end while**
21: **return** S^*

using the branch-and-cut algorithm (Line 7). We employ Gurobi's built-in BC algorithm to perform the BCP within the proposed VMND framework.

4.3.2 Local search phase

Lines 10-19 of **Algorithm 2** illustrate the LSP steps, where $\mathbf{\Theta}$ denotes the ordered set of local search operators (Line 10). Five local search structures are designed to explore the neighborhood of a solution in the LSP. Each operator fixes a subset of binary variables to their values in the incumbent solution, while leaving the remaining variables free (Line 11). If the resulting sub-MIP is constructed by $\boldsymbol{\theta_1}$, it is solved using the Clarke-and-Wright (C&W) algorithm (Clarke and Wright, 1964) as it can be decomposed into small VRPs; otherwise, it is solved using Gurobi's built-in branch-and-cut algorithm, strengthened by valid inequalities (Line 12). Here, a trade-off between computational time and the likelihood of finding a better solution must be managed because with fewer free variables, the sub-MIP is easier to solve, but more free variables increase the likelihood of improvement. Therefore, achieving the right balance between the number of fixed and free variables can significantly enhance the VMND algorithm's performance. Moreover, apart from the number of free variables, selecting which variables to remain free is crucial. If a local search operator improves the solution, the incumbent solution is replaced with the improved solution (Lines 13-14), and the next local search operator starts exploring the improved solution's neighborhood. The LSP ends when an LS fails to improve the incumbent (Lines 15-16). If the first local search operator (in the order list) does not produce a better solution, the time limit of BCP is updated as $\mu^B = \mu^V - t^V$ (Line 19). The neighborhood structures for the PL-LRP are as follows, detailing how to determine the sets of fixed and free variables.

1. **Re-planning period-based routes ($\boldsymbol{\theta_1}$):** This neighborhood, as proposed by Larrain et al. (2017), aims to improve routes within a period. To do so, it takes a period $t \in \mathcal{T}$ and allows all binary variables associated with routing and replenishment in that period (i.e., x_{ijt} and v_{it} $\forall i, j \in \mathcal{N}_0'$) to remain free, while setting the values of the other binary variables to their incumbent solution values. Since this local search operates on routes from one period at a time, the number and set of free variables associated with that period stay unchanged from one solution to another. The resulting sub-problem, which is a regular VRP, is then solved using the C&W algorithm. Hence, to find the best possible improvement, as many as T VRPs are solved in this neighborhood.

   For example, suppose our planning horizon consists of four periods. When applying θ_1 for period 2, all routing and replenishment variables related to that period are free, while those for periods 1, 3, and 4 remain fixed as per the current best solution. This approach allows us to refine the routes for period 2 without altering the structure of other periods.

2. **Reassessing PLs on the underutilized routes ($\boldsymbol{\theta_2}$):** This neighborhood focuses on replanning low-utilization truck routes. To this end, it chooses $\lfloor \xi_2 \, |\mathcal{R}_t| \rfloor$ routes with the lowest initial loads (i.e., vehicle loads after leaving the depot) from the incumbent solution, where ξ_2 is a user-set parameter within the range $(0, 1)$, and $\mathcal{R}_t$ represents the set of routes in period t according to the current solution. It then divides the binary variables into three different sets, labeled S_1, S_2, and S_3, as follows. Set S_1 includes all location (z_{im}), routing (x_{ijt}), and replenishment (v_{it}) variables related to the selected routes, as well as all assignment variables (y_{ik}). Set S_2 contains all binary variables with values 1 in the incumbent solution that are not in S_1. The remaining binary variables fall into S_3. Then, all binary variables in sets S_1 and S_3 are considered free, while those in S_2 are fixed at 1. The resulting sub-problem is a smaller PL-LRP that is solved using the branch-and-cut algorithm after being strengthened with valid inequalities (16)-(21). Since this sub-MIP is still difficult to solve, a problem-specific solution time limit μ^L is imposed.

   For example, if the incumbent solution includes 20 routes in total, and $\xi_2 = 0.25$, local search θ_2 selects the five routes with the lowest initial vehicle loads for reassessment. By freeing variables associated with these routes while fixing those in other routes, the method explores opening PLs that result in more cost-efficient routing.

3. **Reassessing location and capacity of PLs on the routes with high unused capacity-miles traveled ($\boldsymbol{\theta_3}$):** This neighborhood targets routing improvements by revisiting the location and capacity decisions of PLs that appear on low-quality routes, that is, routes characterized by high unused capacity-miles traveled. Here, unused capacity-miles refers to the product of the unutilized vehicle's capacity and the distance it travels, serving as an indicator of inefficient vehicle utilization. It is similar to θ_2, but instead of choosing routes with the lowest loads, it selects $\lfloor \xi_3 \, |\mathcal{R}_t| \rfloor$ routes with the highest unused capacity-miles

traveled. Here, $\xi_3 \in (0,1)$ is a user-set parameter. The unused capacity-miles on a route are calculated as follows:

$$U_r = \sum_{i,j \in r} (Q - \hat{w}_{jt})\, E_{ij}\, \hat{x}_{ijt} \qquad \forall r \in \mathcal{R}_t,\ \forall t \in \mathcal{T} \tag{23}$$

where $\hat{x}_{ijt}$ and $\hat{w}_{jt}$ are the current values of variables x_{ijt} and w_{jt} according to the incumbent solution. Then, a subset of routes with the highest U_r from the incumbent solution is chosen, based on which sets S_1, S_2 and S_3 are defined following the rules detailed for θ_2. Next, the neighborhood of the incumbent solution is searched by solving the resulting small PL-LRP using the branch-and-cut algorithm.

For example, suppose the incumbent solution contains 20 routes in total, and $\xi_3 = 0.25$. Local search θ_3 first computes the unused capacity–miles U_r for each route. Then, it selects the five routes with the highest U_r corresponding to vehicles that travel long distances while carrying relatively small loads, for reassessment. By freeing all binary variables related to these routes across entire periods while fixing others, the method can explore relocating or resizing PLs to reduce unused capacity–miles, which may lead to shorter distances traveled with better load utilization and lower total cost.

4. **Reassessing capacity of PLs on the routes with high unused capacity-miles traveled (θ_4):** This neighborhood revisits the capacity decisions of PLs located on low-quality routes, with the goal of constructing more cost-efficient routes through capacity adjustments. Similar to θ_3, this local search selects $\lfloor \xi_4 \,|\mathcal{R}_t|\rfloor$ routes with the highest unused capacity-miles from the incumbent solution and splits the binary variables into three disjoint sets S_1, S_2 and S_3, defined following the rules detailed in θ_2 and $\xi_4 \in (0,1)$. Then, all binary variables in sets S_1 and S_3 are considered free, and those in S_2 are fixed at 1. However, unlike θ_3, in this local search, PLs on the selected routes stay open, but their capacities and the routes visiting them may change. To do so, some new constraints are added to the sub-MIP as follows. Let $\mathcal{R}_s$ be the set of routes selected to be explored in this neighborhood and $\mathcal{N}_0^r$ represents the set of candidate PLs on route $r \in \mathcal{R}_s$. Then, constraints (24) ensure that all PLs on the selected routes remain open.

$$\sum_{m \in \mathcal{M}} z_{im} = 1 \qquad \forall i \in \mathcal{N}_0^r,\ r \in \mathcal{R}_s \tag{24}$$

The resulting sub-problem, which is solved using the branch-and-cut algorithm, is a special case of PL-LRP, where the subset of open PLs is already fixed, but decisions on PL capacities, SR-to-PL assignments, and routing must be made.

For example, suppose the incumbent solution contains 20 routes in total and $\xi_4 = 0.25$. Local search θ_4 then selects the five routes associated with vehicles that travel long distances while carrying relatively small loads. All PLs located on these routes must remain open, but their capacities, replenishment schedules, routing, and SR-to-PL assignments can be adjusted to improve efficiency. Also, PLs that are not on any routes in the incumbent solution can either remain closed or be opened if doing so yields more cost-efficient routes.

5. **Revisiting 1-closest unopened PLs on random routes (θ_5):** This local search operator improves solution quality by reoptimizing selected routes within a close neighborhood, allowing location, capacity, assignment, replenishment plan, routing, and inventory decisions to be revisited. In this neighborhood,

$\lfloor \xi_5 |\mathcal{R}_t| \rfloor$ routes are selected randomly, allowing all the variables related to PLs on the routes, along with their closest unopened neighbors to become free across all periods (here, ξ_5 is an input parameter between 0 and 1). More precisely, sets S_1, S_2 and S_3 are defined as follows. Set S_1 consists of all location, routing, and replenishment binary variables associated with PLs on the selected routes and their 1-closest unopened neighbors, as well as all assignment variables. Set S_2 contains all binary variables with values 1 in the incumbent solution that are not in S_1. The remaining binary variables fall into S_3. Then, all binary variables in sets S_1 and S_3 are freed, while those in S_2 are set to 1. The resulting sub-MIP, which is a smaller PL-LRP, is strengthened with valid inequalities (16)-(21) before being solved using the branch-and-cut algorithm.

For example, suppose the incumbent solution contains 20 routes, and $\xi_5 = 0.2$, so that four routes are selected randomly, e.g., routes 3, 5, 9, and 14. These routes visit a set of PLs, for instance, PLs 10, 12, 14, and 16. For each of these PLs, we identify its 1-closest unopened neighbor; for example, the closest unopened neighbors are PLs 11, 13, 15, 17, respectively. In this neighborhood search, all decisions related to the PLs 10-17 are freed. This allows (i) neighbor, closed PLs (e.g., PL 11) to be opened if beneficial, (ii) open PLs (e.g., PL 10) to adjust capacity or replenishment, and (iii) routes visiting these open PLs to be reoptimized for better efficiency.

5 Computational experiments

In this section, we present the results of computational experiments using both synthetic instances and real data from Amazon. To this end, we first describe the parameters used for generating random instances and those for the proposed exact algorithm in Section 5.1. Next, in Section 5.2, we compare the performance of the developed VMND with the LFRS heuristic and a branch-and-cut algorithm. Then, we discuss the implications of variations in key parameters, including the number of candidate PLs, capacity modules, and periods in Section 5.3 through a sensitivity analysis. We also assess the environmental impact of last-mile delivery in Section 5.4, evaluating locker parcel delivery (LD) against attended home delivery (HD) and curbside pickup (CP) based on carbon emissions as a sustainability metric. Finally, in Section 5.5, we conduct a case study in Chicago using real Amazon data to highlight the value of flexibility in replenishment planning and compare our approach with a hierarchical process in which strategic decisions precede operational ones.

The VMND and LFRS algorithms were coded in Python 3.9. All computational experiments were carried out on the main node of The University of Alabama High-Performance Computing (UAHPC) system, equipped with an Intel(R) Xeon(R) Gold 6126 CPU @ 2.60GHz, and running the Linux operating system. We used the built-in branch-and-cut algorithm of Gurobi 10.0.3 to solve the formulations presented in Section 4. Since constraints and valid inequalities in the proposed algorithms are all in polynomial number, no separation is required for them, and they are added to the corresponding models when needed.

5.1 Instances and parameter setting

We conduct extensive computational experiments to evaluate the cost-efficiency and sustainability of locker delivery system and assess the performance of the proposed algorithms. To the authors' knowledge, there is no standard benchmark instance for the PL-LRP in the literature. Hence, several instances with varied patterns and sizes were generated and divided into small-, mid- and large-sized sets. The small-sized set is mostly used to compare the LD with HD and CP, and to carry out a sensitivity analysis of key parameters, while the other two sets are used to evaluate the performance of the exact solution method.

The small-sized instance set consists of 5, 10 or 15 candidate PLs; 20, 30 or 50 SRs; 2, 3 or 4 capacity modules; and 3, 5 or 7 periods. Two distinct demand patterns with the same average are considered: *Uniform* and *Bimodal*. The goals are to (i) assess how demand patterns affect strategic decisions and, consequently, the PL network structure, and (ii) demonstrate that the performance of the proposed VMND is robust to the underlying data structure (i.e., the randomly generated instances under differen demand pattern). In practice, however, demand may follow a pattern different from either of these two. In the Uniform pattern, the average SR demand across periods remains constant. For this pattern, demand per period per SR is randomly generated from [0, 10]. In the *Bimodal* pattern, there are two periods of peak demand, such that the average demand initially decreases until the middle period(s) and then increases towards the last period. However, the overall average demand per period per SR remains the same at 5, as in the *Uniform* pattern. This makes both patterns comparable in terms of average demand per SR over the planning horizon, which is a day or part of a day in this paper. The mid-sized instance set includes 30 or 50 candidate PLs while the large-sized instance set consists of 50 or 100 candidate PLs. Both these instance sets include 100, 200 or 400 SRs; 2, 4 or 6 capacity modules; and 4, 6 or 12 periods. Other parameters are common to all three instance sets and are generated using the data presented in Table 4.

In line with common practice, we use a two-part cost structure for opening PLs: a location-based fixed space preparation cost and variable costs for manufacturing, shipping, and installation based on the capacity module. The fixed cost for establishing a fleet is set at \$50,000 per vehicle (CoolParcel, 2024), each with a capacity of 400 orders. A \$0.06 per mile compensation is considered for customers traveling to PLs, while vehicle travel cost for transporting orders from the depot to PLs is estimated at \$0.75 per mile (MotorCarrier, 2024). However, as mentioned earlier, the PL-LRP involves both strategic planning for establishing PLs and a fleet of vehicles and operational planning for routing. Therefore, we consider a five-year economic life for PLs and vehicles, and for normalization purposes, shipping costs and compensation in objective function (1) are multiplied by $5 \times 365 = 1,825$. All candidate PL and SR locations are randomly placed within a 25-mile square, with the depot located at the center.

The parameters required for the VMND are adjusted using a series of preliminary experiments on instances selected from all three sets. These parameters and their values are summarized in Table 5.

5.2 Performance of the proposed algorithms

In this section, we first evaluate the impact of valid inequalities (16)–(21) on improving the lower and upper bounds, as well as the performance of local search operators, in Section 5.2.1, which helps determine the efficient

Table 4: Parameter values for the PL-LRP instances.

Name	Parameter	Values	Unit
Fixed cost of PL	F_{im}	U[15(1+2 m), 20(1+2 m)]	$1,000
Fixed cost of vehicle	FV	50	$1,000
Shipping cost (after scaling)	α	1.37	$1,000/mile
Compensation (after scaling)	β	0.11	$1,000/mile
X coordinates (PLs & SRs)	X_i	U[0, 25]	mile
Y coordinates (PLs & SRs)	Y_i	U[0, 25]	mile
Depot location	(X_0, Y_0)	(12.5, 12.5)	-
Vehicle capacity	Q	400	order
PL capacity	K_m	50 m	box (or order)

Table 5: Input parameter values for the VMND algorithm.

Name	Parameter	Value
VMND maximum time	μ^V	10,800 s
BCP maximum time	μ^B	[120, 360] s
LS operator time limit	μ^L	300 s
Improvement step	γ	[0.0005, 0.01]
Order of LS operators	$\boldsymbol{\Theta}$	$[\theta_4, \theta_5, \theta_3, \theta_2, \theta_1]$
LFRS time limit	μ^I	600 s
Portion of routes selected in θ_2	ξ_2	0.35
Portion of routes selected in θ_3	ξ_3	0.35
Portion of routes selected in θ_4	ξ_4	0.50
Portion of routes selected in θ_5	ξ_5	0.35

sequential ordering of operators in the VMND. Then, in Section 5.2.2, we compare the performance of different variants of the proposed VMND method.

5.2.1 Efficiency of VIs and LS operators

To examine the impact of valid inequalities (16)-(21), computational experiments are performed using the uniform demand pattern. For this purpose, the lower bound (LB), upper bound (UB) and solution time obtained from model (1)-(15) are measured once without valid inequalities and once with all valid inequalities, within a one-hour time limit, using Gurobi B&C solver without warm start. Table 6 presents results for a subset of instances of varying sizes, each featuring a different number of candidate PLs. Column 1 displays the instance name. Columns 2-4 present the LB, UB, and computation time (in seconds) without valid inequalities, while columns 5-7 show the same results with valid inequalities included. The last three columns show percentage changes, with positive values indicating an increase and negative values indicating a decrease. For example, the percentage change for the LB is calculated as $(\mathrm{LB}_{\mathrm{VIs}} - \mathrm{LB}_{\mathrm{w/oVIs}})/\mathrm{LB}_{\mathrm{w/oVIs}} \times 100\%$, where $\mathrm{LB}_{\mathrm{w/oVIs}}$ and $\mathrm{LB}_{\mathrm{VIs}}$ are the LBs obtained from model (1)-(15) without and with valid inequalities, respectively. Observe that the valid inequalities increased the average LB by 6.6%, from 4352 to 4664, while reducing the UB by 10.5% and CPU time by 12.4% on average. These improvements help the VMND algorithm explore the search tree more efficiently.

Table 6: Impact of valid inequalities on lower bound, upper bound, and solution time.

Instance	Without VIs			With VIs			Change (%)		
(N_0,N_1,M,T)	LB	UB	Time(s)	LB	UB	Time(s)	LB	UB	Time
(15,50,2,3)	841.70	841.70	3601.19	841.70	841.70	87.71	0.00	0.00	-97.56
(15,50,3,5)	1180.12	1338.00	3601.41	1249.53	1313.91	3600.35	5.88	-1.80	-0.03
(30,200,4,6)	3290.53	4720.99	3606.24	3599.72	4210.19	3602.90	9.40	-10.82	-0.09
(30,200,6,12)	5313.28	7689.26	3609.70	5629.89	7321.06	3606.96	5.96	-4.79	-0.08
(50,200,4,6)	2956.23	7179.83	3609.72	3202.20	6874.95	3606.28	8.32	-4.25	-0.10
(50,400,6,12)	8637.33	13086.86	3632.39	9180.04	13170.79	3612.78	6.28	0.64	-0.54
(100,400,6,6)	4830.71	22468.42	3644.72	5309.87	13143.24	3628.65	9.92	-41.50	-0.44
(100,400,6,12)	7764.74	34546.98	3663.96	8302.42	26960.89	3652.54	6.92	-21.96	-0.31
Average	**4351.83**	**11484.00**	**3621.16**	**4664.42**	**9229.59**	**3174.77**	**6.59**	**-10.56**	**-12.39**

We also assess how each neighborhood structure impacts solution improvement by conducting experiments where only one local search operator is applied in the VMND algorithm at a time. Results for eight instances following the uniform demand pattern are shown in Table 7 within a three-hour time limit. Here, column 1 shows the instance name, and column 2 represents the solution value provided by the B&C algorithm. The following five columns show the percentage improvement in the B&C's solution value achieved by five variants of the VMND algorithm, labeled as VMND$-\theta_l$ $\forall l \in \{1, 2, 3, 4, 5\}$, where θ_l corresponds to the only local search operator applied in the algorithm.

Observe that all operators are capable of improving solutions, on average, by 13.3% to 19.6%. Most cost savings in the PL-LRP arise from strategic or cross-period decisions, such as location, capacity, or replenishment plan. As a result, operators that revisit location and capacity decisions tend to perform better, yet their effectiveness depends heavily on the quality and size of the neighborhoods they explore. For example, LS θ_1, which focuses solely on routing and replenishment adjustments within a single period, delivers the worst performance, with a 13.3% improvement. While such localized changes can yield improvements, their impact on total costs is limited, making θ_1 less influential than more comprehensive operators. Although θ_2 targets location/capacity and replenishment schedule decisions, its performance is not as strong as that of θ_3, θ_4, and θ_5. This is because the choice of neighborhood also plays a critical role; LS θ_2 selects PLs from routes with low initial load, whereas θ_3 and θ_4 target PLs on routes characterized by high unused capacity-miles traveled. LS θ_3 explores a larger neighborhood than θ_4 by revisiting location, capacity, and replenishment schedule decisions; however, its performance is inferior to θ_4, which is the best-performing operator. The reason is that the sub-MIPs generated in θ_3 are more computationally challenging to solve within the preset time limits, especially for large-sized instances. This also holds when comparing θ_3 and θ_5; by restricting the neighborhood to 1-closest unopened PLs, θ_5 yields sub-MIPs that are significantly easier to solve. These results are used to determine the priority order of the local search operators for sequential use in the VMND algorithm, leading to the order $[\theta_4, \theta_5, \theta_3, \theta_2, \theta_1]$.

Table 7: Performance of local search operators in the VMND algorithm.

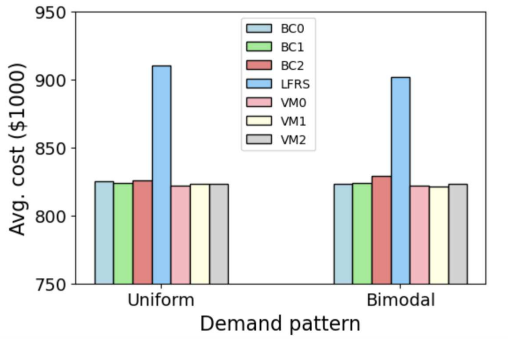
Instance	B&C	Improvement (%)				
(N_0,N_1,M,T)	(OFV)	VMND-θ_1	VMND-θ_2	VMND-θ_3	VMND-θ_4	VMND-θ_5
(15,50,2,3)	841.7	0.0	0.0	0.0	0.0	0.0
(15,50,3,5)	1344.2	3.0	2.3	2.2	2.9	3.1
(30,200,4,6)	4666.2	11.2	11.4	11.6	12.3	12.2
(30,200,6,12)	8277.6	16.7	16.3	16.3	15.3	16.4
(50,200,4,6)	6573.2	26.7	25.7	32.3	34.6	39.2
(50,400,6,12)	12195.3	-0.8	-0.3	3.9	2.7	5.3
(100,400,6,6)	13671.6	26.1	27.5	36.7	48.3	37.4
(100,400,6,12)	27340.4	23.2	28.8	33.9	40.6	41.7
Average		**13.3**	**14.0**	**17.1**	**19.6**	**19.4**

5.2.2 Performance of the VMND

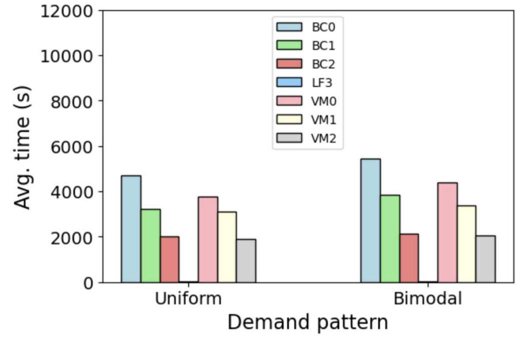
The performance of the proposed VMND method is compared against the LFRS heuristic and the B&C algorithm of Gurobi 10.0.3. We solved 260 random instances, comprising 80 small, 100 medium, and 80 large instances, using seven variants of solution algorithms, including three B&C variants, the LFRS, and three VMND variants. To ensure a fair comparison, the time limit for all solution methods was set to 3 hours. Figure 5 shows the average cost and computational time for both the **Uniform** and **Bimodal** demand patterns. Here, BC0, BC1, and BC2 correspond to Gurobi's B&C solver without a warm start, with a one-hour warm start, and with a two-hour warm start using the LFRS, respectively. The LF3 represents the LFRS heuristic with a three-hour time limit. The bars labeled VM0, VM1, and VM2 are associated with the VMND variants without a warm start, with a one-hour warm start, and with a two-hour warm start using the LFRS, respectively.

Results demonstrate that the VMND algorithm outperforms both the LFRS heuristic and the B&C algorithm overall, regardless of the demand pattern. In small-sized instances, the VMND algorithm produces solutions at lower costs and in less time, on average, compared to its B&C counterpart. Observe from Figure 5a that a warm start does not contribute to the final solution in small-sized instances, such that VM0 (i.e., the VMND without a warm start) has generally the best performance. Although the LFRS generates solutions very quickly (Figure 5b), its average costs are around 9%-10% higher than the other six variants (Figure 5a). However, as the instance sizes grow, the LFRS heuristic delivers better performance, such that for large-sized instances it outperforms most solution variants.

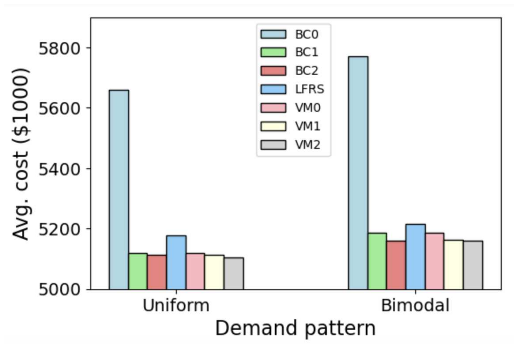
In mid-sized instances, Gurobi's built-in B&C solver (i.e., BC0) performs the worst, with costs nearly 9%-11% higher than the other six variants (Figure 5c). Observe that the VMND still yields better solutions than its equivalent B&C when warm-started with LFRS solutions, but with almost identical running times (Figures 5c-5d). Further, as the warm-start time increases, the VMND delivers better solutions, irrespective of the demand pattern. Similar trends are observed in large-sized instances, except for the LFRS heuristic, which produces solutions comparable to the VMND and B&C algorithms after being warm-started (Figure 5e). Despite this, VM2 can still generate slightly better solutions than LF3 for large-sized instances when demand pattern follows a uniform distribution. Overall, VM0 is best for small instances, while VM2 is superior for mid- and large-sized instances. More details are provided in Tables 14 and 15 in Appendix A. Table 14 shows average costs, and Table 15 displays average computational times for seven solution variants and two demand patterns. Note that



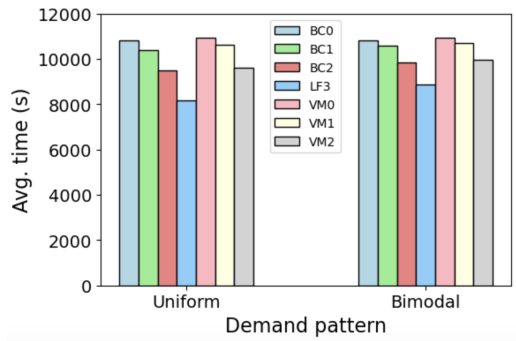
(a) Small-sized instances - cost



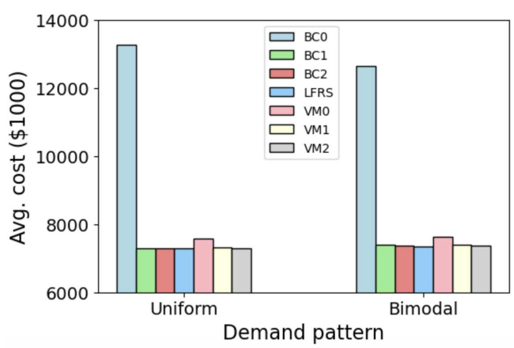
(b) Small-sized instances - time



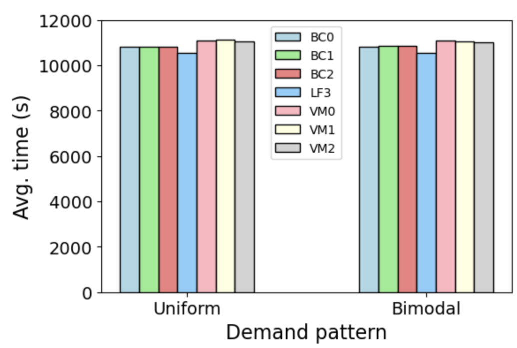
(c) Medium-sized instances - cost



(d) Medium-sized instances - time



(e) Large-sized instances - cost



(f) Large-sized instances - time

Figure 5: Comparison of B&C, LFRS, and VMND with two demand patterns.

each row in these tables represents the average cost and time across five instances of the same size.

5.3 Sensitivity analysis

In Section 5.3.1, we examine how key input parameters, including the number of candidate PLs, capacity modules, and periods, affect the strategic and operational cost components. Then, in Section 5.3.2, we discuss the implications of demand patterns and fluctuations on the PL network structure and costs.

5.3.1 Effects of key input parameters

We conduct a sensitivity analysis on three key input parameters: the number of candidate PLs, capacity modules, and time periods. The goal is to evaluate how each parameter impacts the cost components in objective function (1) as well as selected sustainability metrics. In each experiment, one parameter is varied while the others are kept constant. Table 8 reports the minimum, average, and maximum values of cost components across five optimally solved instances of equal size, denoted as (minimum, **average**, maximum), while Table 9 presents the average sustainability metrics for the same set of instances.

Number of candidate PLs (N_0): Increasing the number of candidate PLs from 5 to 15 increases PL fixed costs by 5.9% but reduces total costs by 14.6% on average. This is because opening more PLs reduces the overall distance traveled by customers, leading to a reduction in distance-based compensation and, as a result, total costs. However, the increase in the number of candidate PLs does not affect the fixed costs of vehicles and has only a slight impact on shipping costs. Note that the locations of candidate PLs remain unchanged; we incrementally add 5 new candidate PLs to the previous ones.

Number of capacity modules (M): Increasing the number of capacity modules from 1 to 2 affects shipping and compensation costs. This increase results in a 37.9% decrease in average shipping costs and a 5.3% rise in average compensation. The reason is that increasing the capacity can lead to opening PLs closer to the depot but further from customers. However, increasing the number of capacity modules from 2 to 3 has a modest impact on cost components, with a 9.2% reduction in shipping costs and a 1.7% increase in compensation.

Number of periods (T): To assess the implications of the number and length of periods, a 21-hour planning horizon with three possible scenarios is considered: (i) 3 periods of seven hours; (ii) 5 periods of 4.4 hours; and (iii) 7 periods of three hours each. A one-hour replenishment break is included between consecutive periods, and daily demand is carefully redistributed across scenarios to ensure comparability. Observe that increasing the number of periods from 3 (seven-hour each) to 7 (three-hour each) reduces PL fixed costs by 19.4%, vehicle fixed costs by 16.7%, and shipping costs by 11.1%, while increasing average compensation by 12%. This is because shorter pickup periods accelerate parcel turnover—parcels are retrieved sooner, reducing the total capacity required and thereby lowering upfront costs. This is achieved by opening fewer or smaller PLs, typically closer to the depot but farther from customers, which reduces fixed and shipping costs but increases compensation. Overall, if shorter periods do not negatively affect customer satisfaction or demand, increasing the number of periods from 3 to 7 yields an average cost saving of 2%.

Sustainability implications: The effects of each parameter is evaluated using three sustainability metrics: (i)

Table 8: Effects of input parameters on cost components.

Parameter	FC of PLs	FC of vehicles	Shipping costs	Compensation	Total costs
No. of PLs	Demand remains unchanged; $N_1 = 30,\ M = 3,\ T = 5$				
5	(212,**287**,362)	(50,**50**,50)	(112,**183**,262)	(692,**837**,979)	(1150,**1358**,1535)
10	(244,**294**,340)	(50,**50**,50)	(168,**177**,193)	(473,**662**,880)	(1017,**1184**,1438)
15	(262,**304**,348)	(50,**50**,50)	(159,**192**,238)	(470,**614**,880)	(1001,**1160**,1438)
No. of modules	Demand remains unchanged; $N_0 = 10,\ N_1 = 30,\ T = 5$				
1	(264,**294**,322)	(50,**50**,50)	(278,**314**,334)	(517,**618**,918)	(1065,**1276**,1547)
2	(264,**293**,331)	(50,**50**,50)	(170,**195**,227)	(473,**651**,880)	(1017,**1190**,1449)
3	(244,**294**,340)	(50,**50**,50)	(168,**177**,193)	(473,**662**,880)	(1017,**1184**,1438)
No. of periods	Total demand per SR remains unchanged; $N_0 = 10,\ N_1 = 30,\ M = 3$				
3	(304,**345**,404)	(50,**60**,100)	(118,**190**,228)	(455,**609**,792)	(1033,**1204**,1508)
5	(244,**294**,340)	(50,**50**,50)	(168,**177**,193)	(473,**662**,880)	(1017,**1184**,1438)
7	(230,**278**,315)	(50,**50**,50)	(97,**169**,206)	(467,**682**,889)	(999,**1180**,1432)

Distance traveled: This metric represents the daily distance traveled by both retailers' and customers' vehicles. (ii) Carbon emissions: This metric measures the total CO2 emissions produced by shipping trucks and customers' cars within a day. There is evidence that vehicles produce 19.57 and 22.42 lbs of CO2 emissions per gallon of gasoline and diesel, respectively (EPA, 2022). Hence, for a midsize gasoline-powered passenger car driving 25 miles and a five-ton shipping truck running 10 miles per gallon of diesel, the estimated CO2 emissions are 0.78 lb/mile for the car and 2.24 lb/mile for the truck. (iii) Traffic density: This metric measures the number of vehicles traveled to each PL per period. The following trends are observed:

(i) Increasing the number of candidate PLs makes the retailer's fleet travel 5 more miles daily, while customers' daily travel drops by over 3,000 miles. This raises the retailer's emissions by 11 lbs/day but cuts customers' emissions by about 2,404 lbs/day. Traffic density also falls by nearly 15 vehicles/PL/period.

(ii) Increasing the number of capacity modules reduces the retailer's fleet distance by nearly 100 miles per day but increases customers' combined distance by about 806 miles. This results in a reduction of 224 lbs of truck emissions per day but an increase of 631 lbs in customers' emissions. However, additional capacity modules could also lead to more traffic around PLs.

(iii) Expanding the number of periods from 3 to 7 reduces the retailer's emissions by 33lb/day but increases customers' emissions by 1,055lb/day. This occurs because it shortens the truck's travel distance by 15miles but lengthens the combined distance customers must travel by 1,330miles. Although less eco-friendly, increasing the number of periods helps reduce traffic density by nearly 42

5.3.2 Effects of demand

To evaluate the impact of demand patterns on the PL network, we performed a sensitivity analysis using six small instances with the same parameters except demand. Table 10 presents the results. The first column lists the indicators, and the next 12 columns provide their values for two demand patterns: uniform (U) and bimodal (B). Observe that although in two out of six instances, all the strategic decisions (i.e., PLs' locations/capacity and

Table 9: Effects of input parameters on sustainability metrics.

Parameter	Distance traveled (mile/day)		Carbon emissions (lb/day)		Traffic density (vehicles/PL/period)	
	Retailer	Customers	Retailer	Customers	Retailer	Customers
No. of PLs	Demand remains unchanged; $N_1 = 30,\ M = 3,\ T = 5$					
5	133.8	14325.7	300.2	11201.7	1.0	67.4
10	129.4	12042.1	290.1	9416.1	1.0	57.9
15	138.6	11251.4	310.8	8797.8	1.0	52.6
No. of modules	Demand remains unchanged; $N_0 = 10,\ N_1 = 30,\ T = 5$					
1	229.1	11235.9	513.6	8785.0	1.0	35.3
2	142.7	11841.9	319.8	9259.7	1.0	54.6
3	129.4	12042.1	290.1	9416.1	1.0	57.9
No. of periods	Total demand per SR remains unchanged; $N_0 = 10,\ N_1 = 30,\ M = 3$					
3	138.7	11073.1	311.0	8658.4	1.1	80.2
5	129.4	12042.1	290.1	9416.1	1.0	57.9
7	123.5	12403.9	277.1	9699.1	0.9	46.2

fleet size) remain unchanged when the demand pattern shifts from the uniform to bimodal, in three instances, this shift in demand pattern causes the number (and overall capacity) of PLs to increase, leading to a modest rise in total costs. In these cases, 40% to 75% of the selected PLs share the same locations and capacities across both demand patterns. However, in one instance, the number of PLs and total costs decrease, but the overall capacity goes up from 400 to 450. Overall, our analysis shows that the demand pattern can significantly impact the PL network structure and must be accurately determined to avoid poor strategic decisions.

Table 10: Effects of demand pattern on network structure and costs.

Indicator	Instance (N_0,N_1,M,T) & Demand pattern (U:Uniform, B:Bimodal)											
	(10,20,2,3)		(10,30,2,3)		(10,30,3,5)		(15,30,2,3)		(15,30,3,5)		(15,50,3,5)	
	U	B	U	B	U	B	U	B	U	B	U	B
Number of PLs opened	1	2	2	2	3	4	2	2	3	4	7	5
PL capacity installed	100	200	200	200	250	300	200	200	300	350	400	450
Same location & capacity	50%		100%		75%		100%		40%		43%	
Fleet size	1	1	1	1	1	1	1	1	1	1	1	1
Fixed costs of PLs	79.0	167.0	175.0	175.0	207.0	273.0	180.0	180.0	249.0	291.0	370.0	378.0
Fixed costs of fleet	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
Shipping costs	46.9	85.4	52.4	52.4	126.6	156.9	59.2	59.2	108.4	116.2	337.3	201.7
Compensation	243.9	196.9	324.7	354.2	434.5	379.7	308.4	314.5	391.5	363.6	548.4	654.4
Total costs	419.8	499.3	602.0	631.6	818.1	859.7	597.6	603.7	798.9	820.9	1305.7	1284.1

To assess the implications of demand fluctuations, four small-sized instances with uniform demand pattern are solved optimally under four different rates of demand perturbation: ±0%, ±10%, ±20%, and ±30%. Here, ±0% denotes the original instance, while the other rates represent instances in which period-specific SR demands are independently perturbed by up to 10%, 20%, and 30%, respectively, with all other parameters remaining unchanged. For example, suppose the demand of a given SR in the original instance is 8 in a certain period. In the ±20% instance, the demand of this SR in the given period equals the nearest integer to $8 \times \rho$, where ρ is a random value between 0.8 and 1.2. To assess how demand fluctuations impact the resulting PL network structure, each instance with perturbed demand is solved optimally twice: once with the strategic decision variables fixed at their values from the original instance, and once without any restrictions on strategic decision variables. The aim is to determine whether the strategic decisions made in the original instance remain feasible under demand perturbations, and if so, how much the total cost increases compared to the optimal solution of the perturbed instance without any restrictions on strategic decision variables. Table 11 summarizes the results for original instances (columns labeled ±0%) and perturbed-demand instances without restrictions on strategic decisions (columns labeled ±10%, ±20%, and ±30%). The results clearly show that demand fluctuations have a negligible impact on the network structure. Observe that the strategic decisions from the original instances remain not only feasible but also optimal under perturbations of up to 20%, with only one instance at 30% demand fluctuation becoming infeasible and requiring an additional PL to be opened. The reason is that the minimum PL capacity (i.e., the smallest capacity module) is far greater than the maximum possible demand from any SR, resulting in a total demand at each opened PL that remains relatively stable despite positive and negative fluctuations in the demands of its assigned SRs.

Table 11: Effects of demand fluctuations on network structure and costs.

Indicator	Instance (N_0,N_1,M,T) & Demand perturbation rates: {±0%, ±10%, ±20%, ±30%}															
	(10,20,2,3)				(10,30,2,3)				(10,30,3,5)				(15,30,2,3)			
	±0%	±10%	±20%	±30%	±0%	±10%	±20%	±30%	±0%	±10%	±20%	±30%	±0%	±10%	±20%	±30%
Variance of demand in a SR	8.3	8.4	8.8	9.3	8.3	8.4	8.8	9.3	8.3	8.4	8.8	9.3	8.3	8.4	8.8	9.3
Does original network remain optimal?	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Does original network remain feasible?	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Does original network remain optimal?	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of PLs opened	1	1	1	2	2	2	2	2	3	3	3	3	2	2	2	2
Same location & capacity	100%	100%	100%	50%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Fleet size	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PL capacity installed	100	100	100	150	200	200	200	200	250	250	250	250	200	200	200	200
Fixed costs of PLs	79.0	79.0	79.0	135.0	175.0	175.0	175.0	175.0	207.0	207.0	207.0	207.0	180.0	180.0	180.0	180.0
Fixed costs of fleet	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
Shipping costs	46.9	46.9	46.9	62.1	52.4	52.4	52.4	52.4	126.6	126.6	149.0	149.0	59.2	59.2	59.2	59.2
Compensation	243.9	242.7	240.1	191.8	324.7	327.0	316.2	329.5	434.5	433.0	393.9	404.7	308.4	312.6	306.2	293.4
Total strategic costs	129.0	129.0	129.0	185.0	225.0	225.0	225.0	225.0	257.0	257.0	257.0	257.0	230.0	230.0	230.0	230.0
Total operational costs	290.8	289.5	286.9	254.0	377.0	379.4	368.5	358.5	561.1	559.6	542.9	553.6	367.6	371.8	365.4	352.6
Total costs	419.8	418.5	415.9	439.0	602.0	604.4	593.5	583.5	818.1	816.6	799.9	810.6	597.6	601.8	595.4	582.6

5.4 Environmental impact of last-mile delivery

As previously mentioned, although our main focus is on delivery costs, we briefly discuss the environmental impact of last-mile delivery as well. To this end, three common delivery modes are compared based on the carbon emissions by the retailer's shipping trucks and customers' vehicles. First, we clearly define the setup for each delivery mode. Then, we discuss how the daily travel distance is estimated for the retailer and customers. For simplicity, the distance traveled within SRs is neglected for both shipping trucks and customers. To ensure fair comparisons, a set of small-sized instances is used with identical demands, periods, and SRs' locations.

1. Attended home delivery (HD): A service provider delivers orders directly from the depot to customers' doorsteps during specified periods chosen by the customers. To estimate the daily travel distance, a standard capacitated VRP is solved for each period, with the SRs being the nodes that must be visited by shipping trucks.

2. Delivery through PLs (LD): Based on the optimal solution of the PL-LRP, a set of PLs is established.

Customers then travel to designated PLs by their vehicles and collect their orders during specified periods. The provider replenishes the PLs using shipping trucks according to optimal replenishment plans.

3. Curbside pickup (CP): A few large stores, specifically at most one-fifth of the candidate PLs in LD, provide curbside pickup services. The locations of these stores are determined by solving an uncapacitated facility location problem with an additional constraint that imposes an upper limit on the number of stores that can be selected. Customers then travel to the the nearest store and collect their orders within selected periods. The provider replenishes the stores directly from the depot once a day, regardless of the number of periods.

Table 12 shows the average carbon emissions from the retailer and customers for HD, LD, and CP across five optimally solved instances of the same size. Results indicate that HD, followed by LD, is the greenest mode overall. However, HD creates the highest emissions from the retailer's perspective. Observe that LD can cut the retailer's carbon emissions by nearly 82% compared to HD, mainly due to the fleet traveling shorter distances. Therefore, one approach retailers can adopt to reduce their carbon footprint is to offer LD and CP services as widely and efficiently as possible. As opposed to CP, LD results in twice the emissions from the retailer's side but only one-third from the customers' side. This is due to the fleet traveling a longer distance, while customers travel a shorter distance overall.

Table 12: Comparing last-mile delivery modes with regard to carbon emissions.

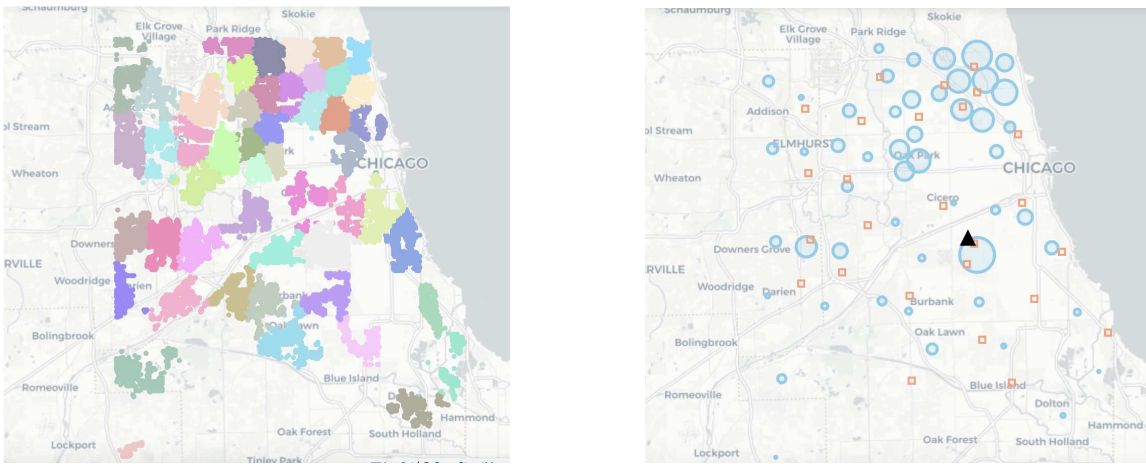
Instance	Carbon emissions (lb/day)								
(N_0,N_1,M,T)	Retailer			Customers			Retailer & Customers		
	HD	LD	CP	HD	LD	CP	HD	LD	CP
(10,20,2,3)	568	77	68	0	3468	6412	568	3545	6480
(10,20,3,5)	1165	93	77	0	5580	13608	1165	5673	13685
(10,30,2,3)	842	86	73	0	4616	12485	842	4702	12557
(10,30,3,5)	1250	198	88	0	5137	18330	1250	5336	18419
(15,30,2,3)	842	97	117	0	4384	9952	842	4481	10068
(15,30,3,5)	1250	252	95	0	4754	14535	1250	5006	14630
(15,50,2,3)	928	212	108	0	5644	14996	928	5856	15104
(15,50,3,5)	1708	477	115	0	7840	32557	1708	8318	32672
Average	**1069**	**187**	**93**	**0**	**5178**	**15359**	**1069**	**5365**	**15452**

HD: Attended home delivery directly from depot, LD: Delivery through PLs, CP: Curbside pickup at large stores.

5.5 Case study

To assess the practical applicability of the proposed PL-LRP, we conduct a case study on designing a PL network for Amazon in Chicago. Using data from the 2021 Amazon Last Mile Routing Research Challenge (Merchán et al., 2024), we provide a near-real evaluation of the PL-LRP's effectiveness and operational value in urban logistics. UPS and FedEx, as Amazon's delivery partners, operate multiple stores across Chicago. We select 25 locations—mainly UPS and FedEx stores—as candidate PL sites, with Amazon's Delivery Station DIL3 as the depot. Using k-means clustering, over 45,000 delivery stops from a 30-day period in 2018 are grouped into 50 SRs, each represented by its centroid and average daily demand. Given the rapid adoption of PLs (InPost, 2024), we assume PL pickup demand matches HD demand in 2018. Figure 6 shows the customer clusters, depot, candidate PLs, and SRs.

We consider a single-day horizon consisting of four periods: 10:00–12:00, 13:00–15:00, 16:00–18:00, and 19:00–21:00. Based on Amorim et al. (2024), we map customer home delivery time preferences to these periods, resulting in demand distributions of 26.4%, 26%, 20.1%, and 27.5%, respectively. We consider three capacity modules (50, 100, and 200 lockers) for each candidate PL, with fixed installation costs of $15,000, $25,000, and $30,000, respectively (Ozyavas et al., 2025). Space preparation costs vary by size and location: $5,000–$15,000 for small modules, $5,000–$20,000 for medium modules, and $10,000–$30,000 for large modules. UPS and FedEx deploy trucks with 400-package capacity and a fixed cost of $50,000. Travel costs are $0.75/mile for trucks (MotorCarrier, 2024), and customer compensation is $0.06/mile. Assuming a five-year economic life (with 1,825 business days), costs are normalized to $1,370/mile for trucks and $110/mile for customer compensation to balance strategic and operational costs in PL network design.



(a) Clusters of customers: each cluster represents a SR. (b) Depot (triangle), PLs (squares), and SRs (circles).

Figure 6: Clusters of customers and location of the depot, candidate PLs, and SRs in Chicago, IL.

To derive insights, we apply the proposed VMND algorithm (with a 30-minutes time limit) to solve the PLND problem under four different policies, namely (a) *Flexible replenishment*, (b) *Once-a-day replenishment*, (c) *Every-period replenishment*, and (d) *Hierarchical*. The problem *flexible-replenishment* policy is identical to the PL-LRP, as presented in model (1)-(15). The problem with *once-a-day replenishment* policy includes additional constraints (compared to the regular PL-LRP) that require PLs to be replenished only once. In the problem under *every-period replenishment* policy, each open PL is replenished every period. Finally, in the hierarchical policy, we first determine the location and capacity of PLs to be opened by solving a capacitated facility location problem (CFLP) based on maximum demand across periods to prevent infeasibility. Then, we determine the replenishment plans and routes by solving a flexible periodic vehicle routing problem (FPVRP) that results from fixing the location-capacity variables (i.e., z_{im}) in model (1)-(15). This allows us to quantify the value of concurrent decision-making as done in the regular PL-LRP compared to a sequential one.

Table 13 illustrates the results for the four problem variants, solved using the proposed VMND with the above modifications. Note that in this analysis, demand data and SRs remain unchanged. Observe that the regular PL-LRP with flexible replenishment is the most cost-efficient approach for designing a PL network in Chicago, with total costs that are 26.6%, 0.8%, and 3.7% lower than the once-a-day replenishment, every-period replenishment, and hierarchical policies, respectively. In particular, incorporating estimated shipping costs and compensation into location and replenishment planning decisions can reduce costs. In other words, a hierarchical decision-making process that first determines the location and capacity of PLs and then plans replenishment schedules and routes is less effective than a simultaneous approach. This shows that key decisions—location, capacity, fleet size, replenishment planning, and routing —are interconnected and should be made together to optimally design a PL network.

Table 13: The VMND results for four problem variants based on the Chicago case study.

	Flexible replenishment	Once-a-day replenishment	Every-period replenishment	Hierarchical
No. of PLs with capacity 50	4	1	6	12
No. of PLs with capacity 100	5	5	2	1
No. of PLs with capacity 150	0	8	0	0
Total number of open PLs	9	14	8	13
Total PL capacity installed	700	1750	500	700
Fleet size	2	5	2	2
Fixed costs of PLs ($1000)	257.0	564.0	205.0	295.0
Fixed costs of fleet ($1000)	100.0	250	100.0	100.0
Shipping costs ($1000)	323.6	234.9	308.3	414.4
Compensation ($1000)	620.7	598.3	698.8	540.1
Total costs ($1000)	1301.3	1647.2	1312.1	1349.5

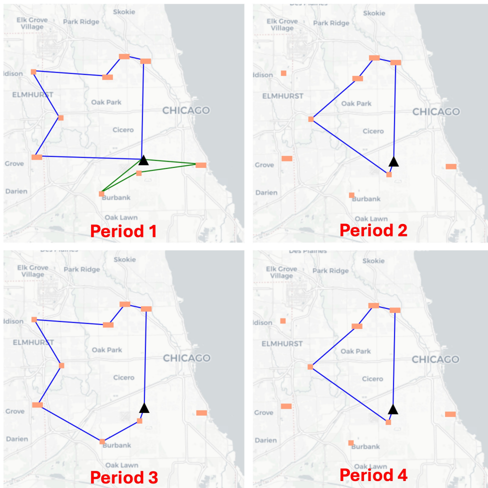
Figure 7 illustrates the location and capacity of open PLs and routes in each period for the problem variants discussed above. Here, the triangle represents the depot, while the rectangles indicate the selected PLs, with bigger rectangles denoting larger PLs. The results show that although the flexible replenishment policy (i.e., the regular PL-LRP in model (1)-(15)) does not strictly dominate other policies in terms of PL fixed costs, fleet fixed costs, shipping, or compensation, it effectively balances all cost components, achieving the lowest overall cost. In particular, this case study provides two actionable insights for logistics planners:

(i) **Flexible replenishment significantly enhances cost efficiency**: compared to the flexible replenishment policy in the PL-LRP, adopting a once-a-day replenishment policy can increase total costs by nearly 27% over the network's economic life, despite reducing shipping costs by 27%;

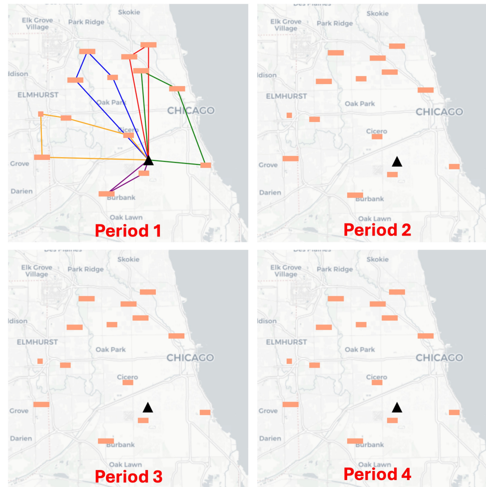
(ii) **Integrating expected routing costs into strategic decisions matters**: a simultaneous approach that incorporates expected operational costs in the network design phase can reduce total costs by approximately 4% compared to a hierarchical approach where location and sizing decisions are made independently of routing considerations.

6 Discussion and managerial insights

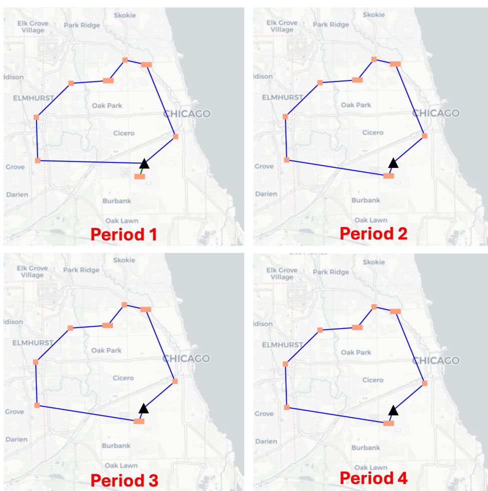
In this paper, we introduced a novel variant of multi-period location-routing problem for parcel locker network design, refeeed to as the Parcel Locker LRP (PL-LRP). The problem, formulated as a mixed-integer linear programming model, determines the locations and capacities of PLs, as well as the fleet size, while taking into



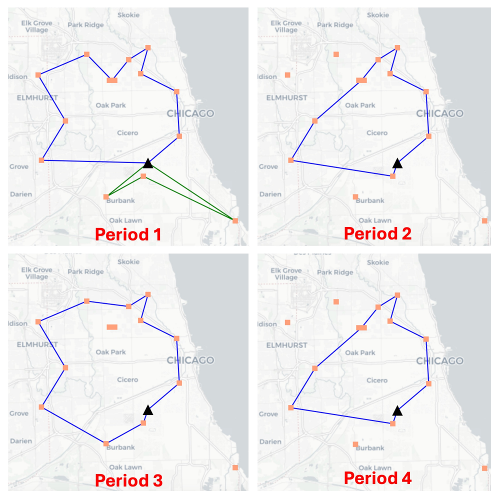
(a) Flexible replenishment ($1,301,257).



(b) Once-a-day replenishment ($1,647,245).



(c) Every-period replenishment ($1,312,086).



(d) Hierarchical ($1,349,545).

Figure 7: Solution obtained by the VMND algorithm for the Chicago case.

account a close approximation of future operational costs over the network's economic life. Operational decisions, which can be revisited after the network is deployed, include replenishment planning, service region (SR) assignments, and vehicle routing for distributing parcels from a depot to opened PLs, all aimed at minimizing strategic and operational costs. To solve the PL-LRP, we developed an exact variable MIP neighborhood descent (VMND) algorithm, which incorporates several valid inequalities and local search operators into a branch-and-cut algorithm. Local search operators are applied to enhance the upper bound, which, together with valid inequalities, expedite the search tree exploration process. To test the proposed algorithm, we used several randomly generated synthetic instances with two different demand patterns: uniform and bimodal. We also conducted a case study in Chicago using real data from Amazon. Results show that the VMND algorithm by far outperforms Gurobi's built-in B&C algorithm, particularly for mid-sized and large-sized instances.

Sensitivity analysis of key input parameters offers three recommendations for configuring a PL network.

First, including as many candidate PLs as possible enhances the cost-efficiency of the resulting network. Specifically, adding more candidate PLs offers greater flexibility in PL placement and SR-to-PL assignments, helping reduce customer travel distances (and consequently, emissions from customers), compensation costs, and traffic density. Second, increasing the number of capacity modules allows for better alignment between PL capacity and demand, which can reduce travel distances and emissions from delivery vehicles, though it may lead to increased traffic density around PLs due to more customers being assigned to larger PLs. Third, given a certain daily demand at each SR, incorporating more but shorter periods (e.g., four periods of two hours each) can reduce the total cost compared to fewer but longer periods (e.g., two periods of four hours each). In fact, shorter pickup periods increase parcel turnover, that is, parcels stay in PLs for less time and are retrieved more quickly, thereby reducing the total capacity needed to meet demand, which in turn lowers upfront fixed costs (see Table 8). That said, offering fewer but longer pickup periods, while more costly, can improve customer experience, as customers often prefer broader time windows for greater flexibility. Therefore, as long as pickup periods are not so narrow that they cause customer dissatisfaction or demand loss, increasing the number of periods is advantageous for the service provider. However, these configuration recommendations increase the problem size, leading to higher computational complexity, and may not continue to considerably improve cost-efficiency beyond certain levels.

Regarding demand implications, the results confirm that demand patterns significantly impact the PL network configuration. Our analyses show that bimodal demand patterns generally require opening more PLs with higher capacities than uniform patterns, leading to higher total costs. Therefore, urban logistics planners must accurately anticipate demand patterns to design efficient PL networks. However, the impact of demand fluctuations on the network structure is negligible as long as each PL has a capacity significantly higher than the demand from any SR. Demand fluctuation analysis shows that designing the PL network based on estimated demand yields an optimal solution that remains robust under moderate uncertainty. The reason is that the total demand at each opened PL remains relatively stable despite positive and negative fluctuations in the demands of its assigned SRs. Thus, urban logistics planners can confidently use the proposed model to design a reliable PL network. However, large demand fluctuations may necessitate adding backup PLs to maintain feasibility.

Our study has two managerial insights to offer. First, replenishment planning should not be viewed solely as an operational afterthought. Flexibility in replenishment plans (i.e., deciding when and how much to replenish each PL) can significantly enhance the cost-efficiency of PL networks. For example, in our Amazon–Chicago case study, flexible replenishment led to substantial savings compared to fixed once-a-day and every-period replenishment plans. This suggests that incorporating replenishment planning into the early stages of network design can improve both strategic and operational efficiency. Second, separating strategic decisions (e.g., which PLs to open and at what capacities) from operational ones (e.g., routing, assignments and replenishment) can result in less cost-efficient networks. Our results show that hierarchical decision-making approaches, which make strategic decisions first and operational decisions afterward, may lead to network designs that are significantly more expensive to establish and operate than those produced by integrated approaches that make both strategic and (uncommitted) operational decisions simultaneously. For planners, this underscores the importance of adopting integrated decision-making tools for designing PL networks, such as PL-LRP and the proposed VMND, particularly in complex urban delivery settings.

This study can be extended in several directions. First, extending the PL-LRP to allow flexible periods, where each PL operates on its own schedule, could enhance capacity utilization and delivery efficiency. Second, incorporating demand and/or travel time uncertainty would improve robustness and real-world applicability. Finally, exploring alternative solution methods, such as metaheuristics or decomposition-based exact methods, could help handle large-sized instances more efficiently.

References

Akturk, M. S., Ketzenberg, M., and Heim, G. R. (2018). Assessing impacts of introducing ship-to-store service on sales and returns in omnichannel retailing: A data analytics study. *Journal of Operations Management*, 61:15–45.

Amorim, P., DeHoratius, N., Eng-Larsson, F., and Martins, S. (2024). Customer preferences for delivery service attributes in attended home delivery. *Management science*, 70(11):7559–7578.

Archetti, C., Bertazzi, L., Laporte, G., and Speranza, M. G. (2007). A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation science*, 41(3):382–391.

Archetti, C., Fernández, E., and Huerta-Muñoz, D. L. (2017). The flexible periodic vehicle routing problem. *Computers & Operations Research*, 85:58–70.

Archetti, C., Jabali, O., and Speranza, M. G. (2015). Multi-period vehicle routing problem with due dates. *Computers & Operations Research*, 61:122–134.

Asgharyar, M., Farmand, N., and Shetab-Boushehri, S. N. (2025). A novel mathematical modeling approach for integrating a periodic vehicle routing problem and cross-docking system. *Computers & Operations Research*, 180:107048.

Avella, P., Boccia, M., and Wolsey, L. A. (2018). Single-period cutting planes for inventory routing problems. *Transportation Science*, 52(3):497–508.

Barenji, A. V., Wang, W., Li, Z., and Guerra-Zubiaga, D. A. (2019). Intelligent e-commerce logistics platform using hybrid agent based approach. *Transportation research part E: Logistics and transportation review*, 126:15–31.

Basir, S. A., Şahin, G., and Özbaygın, G. (2024). A comparative study of alternative formulations for the periodic vehicle routing problem. *Computers & Operations Research*, 165:106583.

Bell, W. J., Dalberto, L. M., Fisher, M. L., Greenfield, A. J., Jaikumar, R., Kedia, P., Mack, R. G., and Prutzman, P. J. (1983). Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer. *Interfaces*, 13(6):4–23.

Beltrami, E. J. and Bodin, L. D. (1974). Networks and vehicle routing for municipal waste collection. *Networks*, 4(1):65–94.

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

Campbell, A. M. and Wilson, J. H. (2014). Forty years of periodic vehicle routing. *Networks*, 63(1):2–15.

Cantu-Funes, R., Salazar-Aguilar, M. A., and Boyer, V. (2017). Multi-depot periodic vehicle routing problem with due dates and time windows. *Journal of the operational research society*, pages 1–12.

Charaf, S., Taş, D., Flapper, S. D. P., and Van Woensel, T. (2024). A matheuristic for the two-echelon inventory-routing problem. *Computers & Operations Research*, 171:106778.

Chien, T. W. (1992). Operational estimators for the length of a traveling salesman tour. *Computers & operations research*, 19(6):469–478.

Clarke, G. and Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations research*, 12(4):568–581.

CleverMaps (2024). Parcel lockers expansion. https://www.clevermaps.io/solutions/use-case/parcel-lockers-expansion. Accessed: 2025-08-14.

Coelho, L. C., Cordeau, J.-F., and Laporte, G. (2014). Thirty years of inventory routing. *Transportation Science*, 48(1):1–19.

Coelho, L. C., De Maio, A., and Laganà, D. (2020). A variable mip neighborhood descent for the multi-attribute inventory routing problem. *Transportation Research Part E: Logistics and Transportation Review*, 144:102137.

Contimod (2025). 20+ last-mile delivery statistics must know in 2025. https://www.contimod.com/last-mile-statistics/.

CoolParcel (2024). How much does a ups truck cost? https://coolparcel.com/blog/how-much-does-a-ups-truck-cost.

Darvish, M., Archetti, C., Coelho, L. C., and Speranza, M. G. (2019). Flexible two-echelon location routing problem. *European Journal of Operational Research*, 277(3):1124–1136.

DataAppeal (2024). Choosing the best locations for parcel lockers using location data. https://datappeal.io/choosing-the-best-locations-for-parcel-lockers-using-location-data/. Accessed: 2025-08-14.

Digital Commerce 360 (2023). Consumers return to stores, maintaining bopis and curbside pickup capabilities.

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 & Operations Research*, 118:104919.

EPA (2022). Greenhouse gases equivalencies calculator - calculations and references. https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references.

Escobar-Vargas, D. and Crainic, T. G. (2024). Multi-attribute two-echelon location routing: Formulation and dynamic discretization discovery approach. *European Journal of Operational Research*, 314(1):66–78.

Faugère, L., Klibi, W., White III, C., and Montreuil, B. (2022). Dynamic pooled capacity deployment for urban parcel logistics. *European Journal of Operational Research*, 303(2):650–667.

Gao, F. and Su, X. (2017). Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science*, 63(8):2478–2492.

Goodchild, A., Kim, H., and Ivanov, B. (2019). Final 50 feet of the urban goods delivery system: Pilot test of an innovative improvement strategy. Technical report.

Grabenschweiger, J., Doerner, K. F., Hartl, R. F., and Savelsbergh, M. W. (2021). The vehicle routing problem with heterogeneous locker boxes. *Central European Journal of Operations Research*, 29:113–142.

Grabenschweiger, J., Dörner, K. F., and Hartl, R. F. (2022). The multi-period location routing problem with locker boxes. *Logistics research*, 15(1):1–25.

Guardian (2023). Parcel thefts from uk doorsteps soaring, new figures reveal. https://www.theguardian.com/money/2023/jul/08/parcel-thefts-from-uk-doorsteps-soaring-new-figures-reveal?

Guglielmo, C. (2013). Turns out amazon, touting drone delivery, does sell lots of products that weigh less than 5 pounds. https://www.forbes.com/sites/connieguglielmo/2013/12/02/turns-out-amazon-touting-drone-delivery-does-sell-lots-of-products-that-weigh-less-than-5-pounds/. Accessed: 2024-09-05.

Hemmelmayr, V., Smilowitz, K., and de la Torre, L. (2017). A periodic location routing problem for collaborative recycling. *IISE Transactions*, 49(4):414–428.

Hemmelmayr, V. C. (2015). Sequential and parallel large neighborhood search algorithms for the periodic location routing problem. *European Journal of Operational Research*, 243(1):52–60.

Hovi, I. B. and Bo, E. (2024). Unlocking the potential: How can parcel lockers drive efficiency and environmental friendliness in e-commerce? *Sustainable Futures*, 7:100189.

Huang, Z., Huang, W., and Guo, F. (2019). Integrated sustainable planning of self-pickup and door-to-door delivery service with multi-type stations. *Computers & Industrial Engineering*, 135:412–425.

InPost (2024). 70 https://www.talkingretail.com/news/industry-news/70-of-gen-z-and-millennials-opt-for-parcel-lockers-02-05-2024/.

Janinhoff, L. and Klein, R. (2023). Stochastic location routing for out-of-home delivery networks. *Available at SSRN 4654115*.

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.

Joerss, M., Neuhaus, F., and Schröder, J. (2016). How customer demands are reshaping last-mile delivery. *The McKinsey Quarterly*, 17:1–5.

Kahr, M. (2022). Determining locations and layouts for parcel lockers to support supply chain viability at the last mile. *Omega*, 113:102721.

Ketzenberg, M. and Akturk, M. S. (2021). How 'buy online, pick up in-store'gives retailers an edge. *Harvard Business Review*, 25(05):2021.

Kučera, P. et al. (2012). Different versions of the savings method for the time limited vehicle routing problem. *Acta univ. agric. et silvic. Mendel. Brun*, 60(7):47–56.

Larrain, H., Coelho, L. C., and Cataldo, A. (2017). A variable mip neighborhood descent algorithm for managing inventory and distribution of cash in automated teller machines. *Computers & Operations Research*, 85:22–31.

LaserShip (2022). Consumers' growing demands for fast and free home delivery over buy online, pick up in-store. https://www.ontrac.com/lasership-study-highlights-consumers-growing-demands-for-fast-and-free-home-delivery-over-buy-online-pick-up-in-store/.

Lin, Y., Wang, Y., Lee, L. H., and Chew, E. P. (2022a). Profit-maximizing parcel locker location problem under threshold luce model. *Transportation Research Part E: Logistics and Transportation Review*, 157:102541.

Lin, Y. H., Tian, Q., and Zhao, Y. (2022b). Locating facilities under competition and market expansion: Formulation, optimization, and implications. *Production and Operations Management*, 31(7):3021–3042.

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.

LJMGroup (2024). The parcel locker phenomenon: Transforming the future of delivery. https://myljm.com/the-parcel-locker-phenomenon-transforming-the-future-of-delivery/.

Ma, B., Teo, C.-C., and Wong, Y. D. (2024). Location analysis of parcel locker network: Effects of spatial characteristics on operational performance. *Transportation Research Part E: Logistics and Transportation Review*, 192:103776.

Magana, G. (2019). The omnichannel fulfillment report: Why the death of brick-and-mortar has been greatly exaggerated. https://www.businessinsider.com/omnichannel-fulfillment-report.

Mancini, S. and Gansterer, M. (2021). Vehicle routing with private and shared delivery locations. *Computers & 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.

Manousakis, E., Repoussis, P., Zachariadis, E., and Tarantilis, C. (2021). Improved branch-and-cut for the inventory routing problem based on a two-commodity flow formulation. *European Journal of Operational Research*, 290(3):870–885.

Merchán, D., Arora, J., Pachon, J., Konduri, K., Winkenbach, M., Parks, S., and Noszek, J. (2024). 2021 amazon last mile routing research challenge: Data set. *Transportation Science*, 58(1):8–11.

Mohamed, I. B., Labarthe, O., Bouchery, Y., Klibi, W., and Stauffer, G. (2023). Multi-echelon urban distribution networks: Models, challenges and perspectives. *The Routledge Handbook of Urban Logistics*, pages 208–224.

MotorCarrier (2024). Cost-per-mile calculator. https://www.motorcarrierhq.com/tools/cost-per-mile-calculator/.

Muñoz-Villamizar, A., Velázquez-Martínez, J. C., Haro, P., Ferrer, A., and Mariño, R. (2021). The environmental impact of fast shipping ecommerce in inbound logistics operations: A case study in mexico. *Journal of Cleaner Production*, 283:125400.

Naji-Azimi, Z., Renaud, J., Ruiz, A., and Salari, M. (2012). A covering tour approach to the location of satellite distribution centers to supply humanitarian aid. *European Journal of Operational Research*, 222(3):596–605.

Najy, W., Archetti, C., and Diabat, A. (2025). A column generation-based matheuristic for an inventory-routing problem with driver-route consistency. *European Journal of Operational Research*.

Olmez, O. B., Gultekin, C., Balcik, B., Ekici, A., and Özener, O. (2022). A variable neighborhood search based matheuristic for a waste cooking oil collection network design problem. *European Journal of Operational Research*, 302(1):187–202.

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.

Peppel, M., Spinler, S., and Winkenbach, M. (2024). Integrating mobile parcel lockers into last-mile delivery networks: an operational design for home delivery, stationary, and mobile parcel lockers. *International Journal of Physical Distribution & Logistics Management*, 54(4):418–447.

Ranjbari, A., Diehl, C., Dalla Chiara, G., and Goodchild, A. (2023). Do parcel lockers reduce delivery times? evidence from the field. *Transportation Research Part E: Logistics and Transportation Review*, 172:103070.

Rodríguez-Martín, I., Salazar-González, J.-J., and Yaman, H. (2019). The periodic vehicle routing problem with driver consistency. *European Journal of Operational Research*, 273(2):575–584.

Rodríguez-Martín, I. and Yaman, H. (2022). Periodic vehicle routing problem with driver consistency and service time optimization. *Transportation Research Part B: Methodological*, 166:468–484.

Rohmer, S. and Gendron, B. (2020). *A guide to parcel lockers in last mile distribution: Highlighting challenges and opportunities from an OR perspective*, volume 11. Cirrelt Montreal.

Savaşer, S. K. and Kara, B. Y. (2022). Mobile healthcare services in rural areas: an application with periodic location routing problem. *OR Spectrum*, pages 1–36.

Savelsbergh, M. and Van Woensel, T. (2016). 50th anniversary invited article—city logistics: Challenges and opportunities. *Transportation science*, 50(2):579–590.

Schaefer, J. S. and Figliozzi, M. A. (2021). Spatial accessibility and equity analysis of amazon parcel lockers facilities. *Journal of Transport Geography*, 97:103212.

Schenekemberg, C. M., Guimarães, T. A., Chaves, A. A., and Coelho, L. C. (2024). A three-front parallel branch-and-cut algorithm for production and inventory routing problems. *Transportation Science*, 58(4):687–707.

Skålnes, J., Andersson, H., Desaulniers, G., and Stålhane, M. (2022). An improved formulation for the inventory routing problem with time-varying demands. *European Journal of Operational Research*.

Statista (2025). Retail e-commerce sales worldwide from 2014 to 2026. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/.

Su, Z., Lü, Z., Wang, Z., Qi, Y., and Benlic, U. (2020). A matheuristic algorithm for the inventory routing problem. *Transportation Science*, 54(2):330–354.

Telling, O. (2025). Delivery groups rush to open parcel lockers as demand surges. https://www.ft.com/content/e896727f-0a20-4256-83d8-c46fc3535771. Accessed: 2025-08-14.

Touzout, F. A., Ladier, A.-L., and Hadj-Hamou, K. (2022). An assign-and-route matheuristic for the time-dependent inventory routing problem. *European Journal of Operational Research*, 300(3):1081–1097.

Vadseth, S. T., Andersson, H., and Stålhane, M. (2021). An iterative matheuristic for the inventory routing problem. *Computers & Operations Research*, 131:105262.

Vakulenko, Y., Hellström, D., and Hjort, K. (2018). What's in the parcel locker? exploring customer value in e-commerce last mile delivery. *journal of Business Research*, 88:421–427.

Vanelslander, T., Deketele, L., and Van Hove, D. (2013). Commonly used e-commerce supply chains for fast moving consumer goods: comparison and suggestions for improvement. *International Journal of Logistics Research and Applications*, 16(3):243–256.

Veenstra, M., Roodbergen, K. J., Coelho, L. C., and Zhu, S. X. (2018). A simultaneous facility location and vehicle routing problem arising in health care logistics in the netherlands. *European Journal of Operational Research*, 268(2):703–715.

Wagenaar, J., Fragkos, I., and Zuidwijk, R. (2021). Integrated planning for multimodal networks with disruptions and customer service requirements. *Transportation Science*, 55(1):196–221.

Wang, M., Zhang, C., Bell, M. G., and Miao, L. (2022). A branch-and-price algorithm for location-routing problems with pick-up stations in the last-mile distribution system. *European Journal of Operational Research*, 303(3):1258–1276.

Zhou, L., Li, C., Hu, C., and Du, J. (2024). Parcel locker location problem with selectable volume sizes and heterogeneous customers in the last mile delivery. *Transportation Letters*, 16(9):1140–1154.

Appendix

A Performance evaluation

Tables 14 and 15 respectively show the average total cost and computational time for both Uniform and Bimodal demand patterns discussed in Section 5.2. Each row represents the average value of five randomly generated instances of the same size. In both tables, instances are labeled as (N_0,N_1,M,T) in the second column, where N_0, N_1, M and T correspond to the number of candidate PLs, SRs, capacity modules and periods, respectively. The columns labeled "Uniform demand pattern" show results for seven solution method variants with uniform distribution demand, while those labeled "Bimodal demand pattern" show results for bimodal demand. To ensure a fair comparison, the time limit for all solution method variants was set to 3 hours. The columns labeled BC0, BC1, and BC2 correspond to Gurobi's built-in B&C solver without a warm start, with a one-hour warm start, and with a two-hour warm start using LFRS, respectively. The LF3 columns show results from the LFRS heuristic with a three-hour time limit. The columns labeled VM0, VM1, and VM2 are associated with the VMND variants without a warm start, with a one-hour warm start, and with a two-hour warm start using LFRS, respectively.

Table 14: Comparison of solution algorithm variants by average total cost (in thousands of dollars).

Instance		Uniform demand pattern							Bimodal demand pattern						
Size	(N_0,N_1,M,T)	BC0	BC1	BC2	LF3	VM0	VM1	VM2	BC0	BC1	BC2	LF3	VM0	VM1	VM2
Small	(10,20,2,3)	482.8	482.8	482.8	511.6	482.8	482.8	482.8	501.8	501.8	501.8	529.0	501.8	501.8	501.8
	(10,20,3,5)	702.0	702.0	702.0	770.8	702.0	702.0	702.0	677.8	677.8	677.8	743.8	677.8	677.8	677.8
	(10,30,2,3)	622.4	622.4	622.4	685.6	622.4	622.4	622.4	666.8	666.8	666.8	717.0	666.8	666.8	666.8
	(10,30,3,5)	953.2	949.8	949.6	1089.4	949.6	949.8	949.8	928.8	930.6	930.2	1027.6	928.8	928.8	929.4
	(15,30,2,3)	636.4	636.4	636.4	725.6	636.4	636.4	636.4	652.6	652.6	652.6	719.8	652.6	652.6	652.6
	(15,30,3,5)	909.8	913.4	915.8	973.6	906.8	914.2	915.2	899.2	903.8	909.2	1005.4	897.8	898.0	903.6
	(15,50,2,3)	922.4	922.6	924.0	988.2	922.4	922.4	922.4	942.2	942.4	945.0	1035.0	940.0	940.0	944.2
	(15,50,3,5)	1370.0	1359.6	1376.8	1534.8	1353.0	1353.6	1352.0	1311.8	1318.6	1346.4	1434.8	1310.6	1299.2	1308.4
	Avg.	**824.9**	**823.6**	**826.2**	**910.0**	**821.9**	**823.0**	**822.9**	**822.6**	**824.3**	**828.7**	**901.6**	**822.0**	**820.6**	**823.1**
Medium	(30,100,2,4)	2004.6	1904.4	1929.6	1991.8	1892.4	1910.4	1922.2	1952.2	1924.6	1925.6	1995.2	1906.2	1901.8	1925.6
	(30,100,4,6)	2699.4	2586.4	2606.8	2663.0	2533.4	2574.8	2589.0	2671.8	2563.8	2553.8	2623.4	2538.8	2518.4	2556.4
	(30,200,4,6)	4573.8	4206.0	4166.6	4234.6	4151.6	4146.2	4158.8	4612.2	4287.2	4298.0	4375.8	4282.6	4285.0	4303.6
	(30,200,6,12)	7730.4	7038.2	7047.2	7135.2	7056.2	7027.2	7041.4	7871.8	7300.0	7248.0	7306.0	7311.6	7217.0	7238.2
	(30,400,6,6)	7862.2	7122.8	7127.8	7194.8	7090.6	7146.0	7134.8	7779.6	7341.4	7305.2	7370.2	7298.0	7302.2	7299.4
	(30,400,6,12)	13963.4	12655.0	12555.4	12708.4	12598.0	12674.0	12618.6	14932.2	12612.0	12513.0	12604.2	12671.6	12681.6	12534.0
	(50,100,2,4)	2024.8	1942.8	1980.6	1985.6	1931.4	1883.0	1893.6	2071.2	1963.0	1987.8	1987.4	1883.0	1895.6	1911.4
	(50,100,4,6)	2887.4	2530.4	2541.8	2585.0	2513.0	2529.8	2524.4	2880.0	2548.8	2547.2	2619.6	2490.8	2548.0	2519.8
	(50,200,4,6)	4875.4	4102.0	4137.6	4191.6	4170.8	4137.2	4142.6	4752.6	4270.8	4259.8	4247.4	4306.4	4236.2	4262.4
	(50,200,6,12)	7990.6	7104.0	7033.0	7092.2	7239.8	7098.2	7026.4	8176.8	7033.6	6954.8	7028.8	7161.0	7048.8	7057.2
	Avg.	**5661.2**	**5119.2**	**5112.6**	**5178.2**	**5117.7**	**5112.7**	**5105.2**	**5770.0**	**5184.5**	**5159.3**	**5215.8**	**5185.0**	**5163.5**	**5160.8**
Large	(50,400,6,6)	8351.0	6890.2	6920.4	6937.0	7151.6	6895.4	6903.6	8490.4	7088.8	7091.4	7096.4	7287.4	7106.0	7107.2
	(50,400,6,12)	13136.0	11899.2	11895.4	11924.8	12329.4	11995.0	11902.6	13531.2	11870.6	11790.4	11830.6	12271.4	11889.6	11826.6
	(100,200,2,4)	3714.8	3162.0	3164.6	3192.6	3261.0	3177.2	3130.4	3691.6	3241.6	3235.2	3253.8	3265.4	3230.0	3232.0
	(100,200,4,6)	6482.0	4127.0	4140.8	4172.4	4260.4	4136.8	4117.2	6678.0	4240.0	4233.2	4192.2	4270.0	4251.4	4242.4
	(100,200,6,12)	18343.2	7019.2	6976.2	6933.6	7223.8	6990.8	6966.4	11893.6	7001.6	6886.2	6916.0	7314.2	6995.8	6875.2
	(100,400,4,6)	13490.6	6799.8	6779.6	6775.6	7030.2	6797.0	6798.2	13388.8	6996.0	6964.8	6942.2	7234.4	6955.2	6955.4
	(100,400,6,6)	13375.0	6856.4	6862.2	6843.0	7126.4	6888.0	6847.2	14830.8	7000.2	7024.0	6966.8	7291.6	7015.4	7011.8
	(100,400,6,12)	29220.6	11651.6	11649.8	11600.4	12225.4	11651.2	11645.4	28706.6	11645.8	11679.2	11584.4	12200.6	11644.8	11604.4
	Avg.	**13264.2**	**7300.7**	**7298.6**	**7297.4**	**7576.0**	**7316.4**	**7288.9**	**12651.4**	**7385.6**	**7363.0**	**7347.8**	**7641.9**	**7386.0**	**7356.9**

Table 15: Comparison of solution algorithm variants based on average computational time (in seconds).

Instance		Uniform demand pattern							Bimodal demand pattern						
Size	(N_0,N_1,M,T)	BC0	BC1	BC2	LF3	VM0	VM1	VM2	BC0	BC1	BC2	LF3	VM0	VM1	VM2
Small	(10,20,2,3)	49.0	80.4	52.6	9.3	58.4	18.4	21.2	34.8	38.4	36.6	3.9	13.6	16.8	15.8
	(10,20,3,5)	1631.4	1251.8	1199.8	9.1	2917.8	2117.4	1616.6	1945.6	1269.0	1042.2	5.9	2289.8	1424.6	1300.2
	(10,30,2,3)	81.8	67.4	104.8	5.2	18.6	18.0	15.8	284.4	233.4	318.6	4.8	324.8	227.0	227.6
	(10,30,3,5)	6276.2	3794.6	2419.8	23.9	3795.8	3652.2	2348.8	5170.2	5244.2	2764.2	13.4	4028.0	3270.2	2340.4
	(15,30,2,3)	1455.6	1102.0	1422.8	17.2	904.4	1070.6	1255.2	5004.2	2671.4	2200.6	11.8	1747.2	1654.4	1653.0
	(15,30,3,5)	10800.8	7228.6	3634.0	24.7	8432.8	6992.4	3654.6	10800.4	7229.2	3627.6	20.2	9400.4	7231.2	3711.6
	(15,50,2,3)	6600.8	5025.2	3628.2	13.1	3266.2	3751.0	2515.6	9449.2	6635.0	3417.0	51.9	6563.8	5653.8	3359.6
	(15,50,3,5)	10801.4	7272.0	3686.8	54.2	10637.4	7300.0	3792.2	10800.8	7338.6	3744.4	126.4	10805.4	7423.8	3769.6
	Avg.	**4712.1**	**3227.8**	**2018.6**	**19.6**	**3753.9**	**3115.0**	**1902.5**	**5436.2**	**3832.4**	**2143.9**	**29.8**	**4396.6**	**3362.7**	**2047.2**
Medium	(30,100,2,4)	10805.8	8419.4	4929.6	921.5	10909.8	8514.6	4574.4	10803.0	9324.6	5445.6	2002.1	10857.4	9160.8	6000.6
	(30,100,4,6)	10802.6	10805.0	9036.6	7196.5	10846.6	10686.6	9231.8	10802.2	10803.2	10856.4	10503.2	10856.4	10954.6	10938.4
	(30,200,4,6)	10802.8	10810.2	9999.6	8763.5	10952.8	11093.8	10581.4	10802.2	10804.8	10806.8	10504.3	10910.4	10922.4	10882.4
	(30,200,6,12)	10806.4	10808.6	10808.2	10515.8	10967.0	10922.6	10914.2	10806.4	10809.6	10809.4	10509.6	11032.8	10994.2	10874.8
	(30,400,6,6)	10808.2	10811.4	10808.0	10511.9	11132.2	11016.0	10860.4	10808.2	10813.2	10808.6	10509.3	11015.2	11088.0	10864.2
	(30,400,6,12)	10808.4	10818.4	10812.4	10512.9	10858.2	10925.2	10855.8	10808.4	10819.4	10819.0	10518.4	10929.2	10950.8	10842.6
	(50,100,2,4)	10803.6	9006.2	5997.4	2295.6	10831.2	10273.0	6382.8	10802.8	10128.0	7263.8	4221.3	10833.6	10025.8	6847.4
	(50,100,4,6)	10805.6	10807.4	10806.2	10066.6	10949.4	10865.6	10875.0	10823.6	10807.2	10235.8	9057.9	10940.4	10942.6	10640.2
	(50,200,4,6)	10818.0	10808.2	10808.2	10508.5	11041.4	10940.6	10834.4	10807.0	10810.4	10808.6	10509.8	11060.0	10929.0	10837.0
	(50,200,6,12)	10820.0	10822.6	10815.8	10512.7	10929.0	10869.4	10860.6	10809.8	10820.0	10819.2	10519.4	10959.2	10898.6	10887.6
	Avg.	**10808.1**	**10391.7**	**9482.2**	**8180.6**	**10941.8**	**10610.7**	**9597.1**	**10807.4**	**10594.0**	**9862.3**	**8885.5**	**10939.5**	**10686.7**	**9961.5**
Large	(50,400,6,6)	10812.8	10816.4	10820.6	10534.6	10995.6	10921.4	10875.8	10811.6	10820.2	10827.0	10513.9	11007.2	10873.4	10859.2
	(50,400,6,12)	10820.8	10822.8	10834.8	10535.5	10994.0	10876.0	10919.2	10818.2	10841.8	10836.0	10550.9	11020.8	10946.8	10892.0
	(100,200,2,4)	10812.6	10821.0	10817.8	10515.1	10880.6	11186.8	10851.2	10815.2	10831.4	10816.4	10521.9	10910.4	10885.4	10870.0
	(100,200,4,6)	10823.4	10826.4	10823.6	10521.2	10992.2	11012.0	10973.4	10825.0	10835.8	10836.6	10522.6	10916.8	10885.6	10870.6
	(100,200,6,12)	10843.6	10844.4	10834.6	10533.8	11295.0	11481.6	11271.4	10836.8	10857.4	10839.8	10539.6	11320.8	11389.2	11281.0
	(100,400,4,6)	10832.4	10836.8	10842.4	10544.0	11186.6	10977.4	11073.4	10836.2	10843.2	10842.2	10545.7	11094.8	11004.0	10990.8
	(100,400,6,6)	10830.4	10840.2	10832.8	10547.2	11038.6	11004.2	11005.6	10835.2	10858.4	10849.4	10559.7	11072.8	11005.8	11111.4
	(100,400,6,12)	10867.0	10866.6	10869.2	10595.3	11479.8	11540.0	11422.4	10858.0	10868.4	10880.6	10563.9	11444.4	11459.2	11165.4
	Avg.	**10830.4**	**10834.3**	**10834.5**	**10540.8**	**11107.8**	**11124.9**	**11049.0**	**10829.5**	**10844.6**	**10841.0**	**10539.8**	**11098.5**	**11056.2**	**11005.0**