Optimisation of Last-Mile Delivery Costs in Mumbai Using Transportation Models and Mixed-Integer Linear Programming

Pranav Damaraju[1], Sarthak Jain[2*], Yash Choube[3], Sarthak Jain[4], Shrey Mittal[5]
[1, 2, 3, 4, 5]B.Sc. (Finance) Students, Anil Surendra Modi School of Commerce, SVKM's NMIMS, Mumbai, India
[*]Corresponding author: Sarthak Jain[2] (e-mail address)

Abstract: This paper develops and tests a method for minimising last-mile delivery costs in Mumbai by combining a transportation model with a mixed-integer linear programme (MILP). The problem studied is how a delivery firm with several depots and a capacitated van fleet should allocate delivery zones to depots, group and sequence the zones on routes, and choose the number of vans so that all daily demand is met at minimum cost. Allocation and routing interact, compact exact models scale poorly, and open, verifiable city-level examples are scarce; an illustrative case with clearly labelled synthetic demand is therefore used. A transportation linear programme is formulated as an allocation benchmark, and a multi-depot capacitated routing MILP with lifted Miller–Tucker–Zemlin constraints is solved to proven optimality with the SCIP solver through Google OR-Tools for a two-depot, twelve-zone Mumbai network serving 1,000 parcels per day. Compared with a nearest-depot, nearest-neighbour dispatch rule, the optimal plan reduces daily cost by 20.4% (from ₹6,443.68 to ₹5,131.94), distance by 14.5% and the fleet from four vans to three; 68.6% of the saving comes from eliminating one under-filled route. Across seven re-solved sensitivity scenarios the saving ranges from 13.6% to 23.1%. The plan is robust to changes in cost rates but sensitive to vehicle capacity and demand through the minimum fleet size. The paper also documents a structured EBSCO bibliometric search, profiles the core literature and derives managerial implications for urban parcel operators.

Keywords: Last-mile delivery, transportation problem, mixed-integer linear programming, multi-depot vehicle routing, operations research, Mumbai.

I. INTRODUCTION

A. Background and Motivation

Online retail, quick commerce and on-demand services have turned the final leg of a parcel's journey — from a local depot to the customer — into a central planning problem for logistics firms. Reviews describe last-mile logistics as spanning distribution, fulfilment, transport and delivery decisions [1] and stress that each delivery concept creates decision problems that operations research (OR) methods can address [2]. City-logistics research has long recognised that urban distribution is fragmented, congested and costly, and that it generates externalities such as emissions and traffic [3], [4]. Because the last mile is typically the least efficient stage of B2C e-commerce distribution, efficiency-raising solutions have attracted sustained attention [5].

Mumbai is a useful setting for studying these decisions. It is a dense, congested coastal city whose delivery zones are strung along long north–south corridors, so the choice of which depot serves which zone, and in what sequence, directly determines the distance travelled and the number of vehicles required. The city is used here as a setting for a methodological illustration; no claim is made that the numbers describe any actual operator.

B. Problem Statement

A delivery firm operates a small number of depots and a homogeneous fleet of capacitated vans. Each day a known number of parcels must reach a set of delivery zones. The firm must decide (i) which depot serves each zone, (ii) which zones share a van and in what order they are visited, and (iii) how many vans to use, so that every zone is served exactly once, no van exceeds its capacity, and total distance-based and per-route cost is minimal. In practice dispatch is often done by rule of thumb — for example, "serve each zone from its nearest depot and chain the nearest unvisited zone" — and the cost of such rules is rarely measured against an optimum.

Three features make the problem worth studying. First, depot allocation and routing interact, yet they are often treated separately. Second, compact exact formulations have weak linear relaxations and scale poorly, which is why larger instances are usually solved with heuristics [6], [7]. Third, real company data are rarely public, so any open example must be built on assumptions that are stated explicitly.

C. Approach and Contributions

The paper formulates a transportation model as an allocation benchmark and a multi-depot capacitated routing MILP, solves the MILP exactly for a small Mumbai-style case, and compares the result with a rule-based baseline on cost, distance, fleet size, utilisation and cost per parcel under seven scenarios. Its contributions are:

1) a transparent two-stage formulation that links a transportation-type allocation decision to an integrated multi-depot routing decision;
2) an exact solution with a decomposition of the saving into distance-driven and fleet-driven components, which shows where the value of optimisation lies;
3) a sensitivity analysis that identifies which parameters change the optimal plan and which only change its cost; and
4) a fully specified, reproducible dataset, model and bibliometric protocol suitable for teaching and for extension with real data.

The remainder of the paper is organised as follows. Section II states the research questions and objectives; Sections III–V present the literature review, bibliometric protocol and research gap; Sections VI–IX describe the methodology, data, model and solution approach; Sections X–XI report results and sensitivity analysis; and Sections XII–XIV cover managerial implications, limitations and conclusions.

II. RESEARCH QUESTIONS AND OBJECTIVES

Main research question. How can transportation models and mixed-integer linear programming be used to minimise last-mile delivery costs in Mumbai while satisfying customer demand, vehicle capacity and operational constraints?

Supporting questions. *RQ1:* How does a transportation model allocate zone demand to depots, and how consistent is that allocation with the integrated routing solution? *RQ2:* How much cost, distance and fleet reduction does an exact MILP achieve relative to a rule-based dispatch baseline? *RQ3:* Which cost components account for the saving? *RQ4:* How robust is the optimal plan to changes in vehicle capacity, demand and cost parameters?

Objectives. The study aims to: (O1) review the literature on last-mile delivery, transportation problems, LP/MILP and vehicle routing and identify a defensible gap; (O2) specify a reproducible bibliometric protocol; (O3) build a transparent, clearly labelled illustrative Mumbai dataset; (O4) formulate a transportation-based allocation model and a multi-depot capacitated routing MILP; (O5) solve the MILP exactly and compare it with a baseline on cost, distance, fleet size, utilisation and cost per parcel; (O6) test the solution under capacity, demand, cost and fleet scenarios and derive managerial implications; and (O7) provide code that reproduces every reported number.

III. LITERATURE REVIEW

A. Transportation Problem and Linear Programming

The classical transportation problem, first posed by Hitchcock [8], minimises the cost of shipping a homogeneous good from sources with limited supply to destinations with fixed demand. It is a linear programme and, with integer supplies and demands, has integer optimal flows [9]. Its weakness for last-mile work is structural: costs are linear in the quantity shipped, and the model has no vehicles, per-trip capacities or visiting sequence. It therefore answers which depot supplies which zone, but not how vans are loaded and routed — a limitation that motivates the two-stage design of this study.

B. Vehicle Routing and MILP Formulations

Dantzig and Ramser [10] introduced the truck dispatching problem, the origin of the vehicle routing problem (VRP). Laporte [6] traces fifty years of exact and heuristic methods, and Toth and Vigo [7] provide the standard modern treatment of VRP formulations and algorithms. For small and medium instances, compact MILP formulations remain attractive because they are transparent and solvable with general-purpose solvers. Subtours are commonly eliminated with the ordering constraints of Miller, Tucker and Zemlin [11], whose weak linear relaxation is strengthened by the lifted version of Desrochers and Laporte [12]; both ideas are used in Section VIII. The price of compactness is solution time as instances grow, which is why larger problems are normally solved with heuristics [6], [7].

Solution methods fall into two broad families. Exact methods — branch-and-bound, branch-and-cut and column generation implemented in general-purpose or specialised solvers — prove optimality and provide a lower bound against which any plan can be judged, but their running time grows rapidly with the number of customers. Heuristics and metaheuristics, from classical constructive and improvement procedures to large-neighbourhood search, return good solutions for large instances quickly but without a guarantee of optimality [6], [7]. A rule-based dispatch procedure, such as the nearest-depot, nearest-neighbour rule used as the baseline here, is the simplest constructive heuristic and approximates what many operators do in practice. Comparing such a rule with an exact optimum on the same data is therefore a direct way to measure the value of optimisation.

C. Location, Multi-Depot and Two-Echelon Distribution

Where freight is held and from which facility each customer is served are classical location–allocation decisions. Melo, Nickel and Saldanha-da-Gama [13] review how facility-location models with capacities and single sourcing are embedded in supply-chain network design. For cities specifically, Crainic, Ricciardi and Storchi [14] propose two-tier systems in which freight is consolidated at peripheral platforms and transferred to small vehicles at urban satellites, and Perboli, Tadei and Vigo [15] formalise the two-echelon capacitated VRP with math-based heuristics. Montoya-Torres et al. [16] survey the multi-depot VRP (MDVRP) from 1988 to 2014, covering time windows, split deliveries, heterogeneous fleets and other variants; the MDVRP is the natural structure for a firm with several depots, as in the present case. Vidal, Laporte and Matl [17] describe how VRP research has moved towards richer

objectives and integration with tactical decisions, while Pillac et al. [18] review dynamic VRPs in which information arrives during execution.

D. Emerging Last-Mile Delivery Concepts

A growing stream models new delivery modes. Murray and Chu [19] introduced the flying sidekick travelling salesman problem, a MILP in which a truck and a drone cooperate, and Agatz, Bouman and Schmidt [20] develop exact and heuristic approaches for the travelling salesman problem with drone. Ostermeier, Heimfarth and Hübner [21] study multi-vehicle truck-and-robot routing, and Arnold et al. [22] use simulation to show that cargo bikes and delivery points can reduce cost and externalities in Antwerp. At the network level, Stokkink and Geroliminis [23] use a MILP for micro-hub location, capacity and fleet assignment in a multi-modal system and illustrate it with a Madrid case in which operating cost and pollution fall. Closer to the present problem, He et al. [24] formulate a mixed-integer model for multi-depot last-mile delivery with roaming customers and report a 16.9% cost reduction for a multi-depot set-up relative to a single depot, with larger reductions when depots collaborate. Beatrici et al. [25] propose a two-stage stochastic model for fleet size and mix with consistent routes, together with a heuristic for larger instances.

E. Reviews of Last-Mile Logistics

Several reviews structure the field. Savelsbergh and Van Woensel [3] set out the challenges and opportunities of city logistics; Ranieri et al. [4] review last-mile innovations through the lens of external-cost reduction; Mangiaracina et al. [5] classify solutions that raise B2C delivery efficiency; Olsson, Hellström and Pålsson [1] review 155 peer-reviewed publications and propose a five-component framework (last-mile logistics, distribution, fulfilment, transport and delivery); Boysen, Fedtke and Schwerdfeger [2] survey established and novel delivery concepts and the OR methods used for their decision problems; and Jazemi et al. [26] group the urban last-mile VRP literature into crowdshipping, parcel lockers, delivery by sidekicks and delivery to optional points. Table I synthesises these streams.

TABLE I
SYNTHESIS OF LITERATURE STREAMS RELEVANT TO THIS STUDY

Stream	Representative sources	Typical method	Limitation relevant to this study
Transportation problem	[8], [9]	LP, network flow	No vehicles, trips or sequencing
Classical VRP / MILP	[10], [11], [12], [6], [7]	Compact MILP; exact and heuristic algorithms	Compact models scale poorly; mostly single-depot, generic instances
Location, multi-depot and two-echelon	[13], [14], [15], [16]	MILP, math-based heuristics, surveys	Allocation and routing rarely reported separately for a city case
VRP variants and dynamics	[17], [18]	Reviews and frameworks	Broad; no small reproducible cost decomposition
Emerging last-mile concepts	[19], [20], [22], [21], [23], [24], [25]	MILP, stochastic MIP, simulation, heuristics	New modes, location and uncertainty; not a small reproducible case for one city
Last-mile reviews	[3], [4], [5], [1], [2], [26]	Systematic review, OR survey	Map the field; do not solve a specific city case

Source: authors' synthesis of the cited works.

F. Synthesis

Three patterns emerge. First, methodology has moved from single-depot, static, capacitated routing towards multi-attribute, dynamic and concept-specific problems [17], [2], but the exact MILP remains the reference model against which heuristics and rules are judged. Second, transportation-type allocation and routing are usually treated as separate topics, even though in a multi-depot firm the allocation decision largely determines routing cost. Third, much of the review literature maps the field rather than demonstrating a small, fully transparent computation that a practitioner or student can replicate. Recent studies add a fourth pattern: strategic network decisions (location, fleet mix) and uncertainty are increasingly modelled [23], [25], which the deterministic case in this paper does not capture. The review is intentionally selective rather than systematic and is complemented by the structured database search described in Section IV.

IV. BIBLIOMETRIC ANALYSIS

A. Design

Bibliometric analysis complements a narrative review by quantifying publication trends, influential sources and the intellectual structure of a field [27], [28]. The study follows the performance-analysis and science-mapping workflow of Donthu et al. [27], with reference management in Mendeley and analysis in Bibliometrix/Biblioshiny [29] or VOSviewer [30]. Table II states the search protocol.

TABLE II
BIBLIOMETRIC SEARCH PROTOCOL

Item	Specification
Database	EBSCOhost (institutional access); Scopus, Web of Science or Google Scholar as supplements
Compulsory keywords	"operations" AND "research"
Search string	operations AND research AND ("last mile" OR "last-mile" OR "vehicle routing" OR "transportation problem" OR "mixed integer")
Filters	Academic (peer-reviewed) journals; English; fixed start and end year
Screening	Each exclusion logged with reason and document type (e.g., case studies, introductions, theses, out-of-scope articles)
Export and management	RIS export (≤150 records per batch) into a Mendeley project folder; metadata verified; full texts attached where available; .ris and .bib exported
Analysis	BibTeX loaded into Biblioshiny; Excel output enriched with Google Scholar citation counts

B. Indicators

The EBSCO search, executed on 6 October 2026 with the filters in Table II, returned 108 records, which were exported to Mendeley for screening and analysis. Table III lists the indicators defined for these records, the procedure for each and the tool used. This paper reports the descriptive profile of the core literature (Section IV-C); full performance analysis and science mapping of the 108 records is reserved for an extended study.

TABLE III
BIBLIOMETRIC INDICATORS AND TOOLS

Indicator	Procedure	Tool / output
Publications by year	Count records per year; fit growth trend	Biblioshiny: annual scientific production
Most-cited papers	Rank by global citations and Google Scholar counts	Biblioshiny: most global-cited documents
Leading journals	Count documents per source; apply Bradford's law	Biblioshiny: most relevant sources
Leading authors	Count documents and fractional authorship	Biblioshiny: most relevant authors
Frequent keywords	Count author and indexed keywords after thesaurus merge	Biblioshiny: word frequencies
Keyword co-occurrence	Co-occurrence network with a minimum-occurrence threshold	VOSviewer map; Biblioshiny network
Themes and evolution	Thematic map and thematic evolution across time slices	Biblioshiny: thematic map / evolution

C. Profile of the Reviewed Literature

As a first, descriptive step, the scholarly works cited in Section III were profiled by publication period and outlet (Table IV). The thirty works span 1941 to 2025 and are concentrated after 2010, with a visible shift in recent years from general routing towards concept-specific and network-design problems. *Transportation Science* and the *European Journal of Operational Research* are the most frequent outlets, reflecting the centrality of transport-focused OR journals. Because this is a selective reading list rather than a sample of the field, the profile should not be read as evidence of publication trends; that role is reserved for the database analysis specified above.

TABLE IV
PROFILE OF THE 30 SCHOLARLY WORKS REVIEWED

Period	Works	Most frequent outlets	Works
Before 1990	3	Transportation Science	5
1990–2009	4	European Journal of Operational Research	4
2010–2016	7	Sustainability	2
2017–2021	11	Journal of Business Research	2
2022–2025	5	Other journals (one each), books and preprints	17
Total	**30**	**Total**	**30**

Source: authors' count of the works cited in Section III; software references excluded.

V. RESEARCH GAP

Existing studies have (a) established the transportation and VRP formulations and their exact and heuristic solution methods, (b) surveyed multi-depot, two-echelon and other variants, and (c) mapped last-mile concepts and research themes. For the present purpose, their limitation is that general surveys and benchmark-instance studies do not show, in one small and fully transparent example, how a depot-allocation (transportation) view relates to an integrated routing solution; they rarely decompose the saving from optimisation into distance-driven and fleet-driven parts; and they say little about how a city-specific plan responds to capacity, demand and cost changes. What remains insufficiently addressed, at least among the sources reviewed, is a compact, reproducible, city-level illustration that links these elements.

This study addresses that gap for a Mumbai-style setting: it combines a transportation allocation benchmark with a multi-depot MILP, decomposes the cost saving and tests seven scenarios, with all code and labelled synthetic data supplied. It does not claim to be the first study of Mumbai last-mile delivery, nor to provide empirical findings about real operators.

VI. RESEARCH METHODOLOGY

The study follows a standard OR modelling cycle, summarised in Fig. 1 and Table V: the problem is defined from the literature, a case is constructed, the allocation and routing models are formulated and solved, the solution is compared with a baseline, and its robustness is tested.

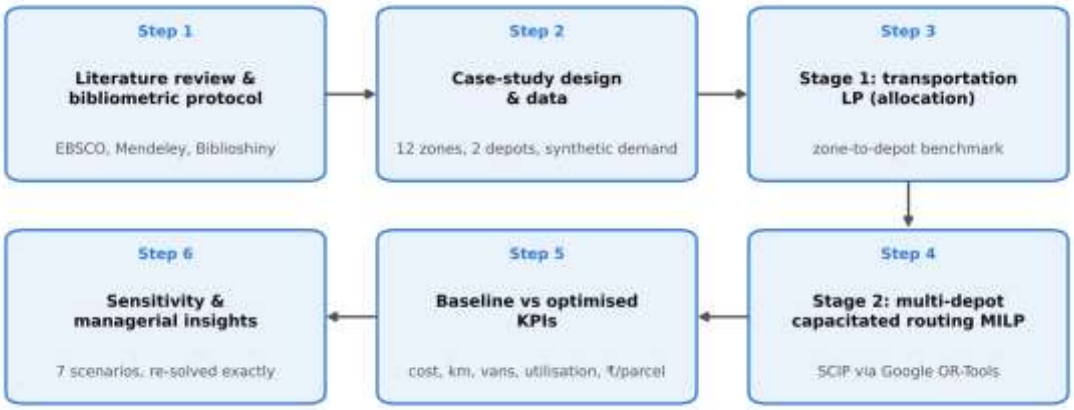


Fig. 1. Research framework

Justification of the design. Three design choices follow from the research questions. An *exact* MILP is used rather than a heuristic because RQ2 and RQ3 require a proven optimum against which the baseline can be measured; any saving reported by a heuristic would mix the value of optimisation with the quality of the heuristic. A *transportation model* is retained alongside the MILP because it is the textbook allocation tool and answering RQ1 requires comparing its allocation with the integrated solution. *Synthetic, clearly labelled demand* is used because company delivery data are confidential, and an openly stated dataset allows every number in the paper to be reproduced and challenged. The rule-based *baseline* represents a plausible current practice rather than a deliberately weak comparator, so that the measured saving is not inflated.

TABLE V
METHODOLOGY STEPS AND STATUS

#	Step	Activity	Status
1	Literature review	Selective review of transportation, LP/MILP, VRP and last-mile sources (Section III)	Completed
2	Bibliometric analysis	EBSCO search, Mendeley screening and indicator design (Section IV)	Search executed (108 records)
3	Case-study design	Two depots, twelve zones, homogeneous fleet, one-day horizon	Completed
4	Data preparation	Haversine distances scaled by a circuity factor; synthetic demand and cost parameters	Completed (synthetic)
5	Mathematical modelling	Transportation LP benchmark and multi-depot capacitated routing MILP (Section VIII)	Completed
6	Optimisation	Exact solution with SCIP via OR-Tools; rule-based baseline	Completed
7	Results analysis	Cost, distance, vans, utilisation, cost per parcel, saving decomposition	Completed
8	Sensitivity analysis	Capacity, demand, cost-rate and fleet-size scenarios, each re-solved exactly	Completed

VII. DATA AND CASE STUDY

A. Data Classification

To avoid any confusion between real and constructed inputs, every input is classified. **Real (approximate):** the latitude and longitude of twelve Mumbai-region localities and two depot locations (Bhiwandi and Andheri East), used only to compute distances. **Assumed:** road distance equals 1.30 times the great-circle (haversine) distance; variable cost ₹14 per km; fixed cost ₹900 per van route; van capacity 350 parcels; three vans available at each depot; depot stock capacity 700 parcels per day each. **Synthetic:** the daily parcel demand of each zone. None of these figures comes from a company or from measured road data, and the results must not be presented as such. Tables VI and VII list the inputs.

TABLE VI
SYNTHETIC DAILY DEMAND BY DELIVERY ZONE (PARCELS PER DAY; TOTAL 1,000)

Zone	Parcels	Zone	Parcels	Zone	Parcels
Colaba	45	Bandra	110	Chembur	80
Worli	70	Malad	105	Ghatkopar	90
Dadar	85	Borivali	75	Powai	100
Kurla	95	Mulund	60	Thane	85

Source: synthetic assumption constructed for illustration. Depots: D1 Bhiwandi, D2 Andheri East.

TABLE VII
MODEL PARAMETERS AND THEIR CLASSIFICATION

Parameter	Value	Classification / how obtained
Depot and zone coordinates	12 zones, 2 depots	Real (approximate public locality coordinates), used only for distances
Road distance d_{ij}	1.30 × haversine km	Assumption (circuity factor); not measured road distance
Zone demand q_j	45–110 parcels/day	Synthetic (Table VI)
Van capacity Q	350 parcels	Assumption
Vans per depot K_d	3	Assumption
Depot stock s_d	700 parcels/day each	Assumption (non-binding in the base case)
Variable cost c	₹14 per km	Assumption
Fixed cost F	₹900 per van route	Assumption

Source: authors' assumptions; no company data used.

B. Scope Simplifications

The one-day horizon, a single vehicle type and deterministic demand keep the problem at a tractable scale. Each zone is treated as a single delivery node, so a zone's parcels occupy one position in a route sequence — a simplification of door-to-door delivery that captures the depot-to-zone and zone-to-zone movements that dominate van kilometres.

VIII. MATHEMATICAL MODEL

A. Notation

TABLE VIII
SETS, PARAMETERS AND DECISION VARIABLES

Symbol	Definition
D, Z	Set of depots ($\|D\| = 2$) and set of delivery zones ($\|Z\| = 12$)
$N = D \cup Z$, A	Set of all nodes; set of arcs (i, j) between nodes, excluding depot-to-depot arcs
q_j	Daily demand of zone j (parcels)
d_{ij}	Road distance from node i to node j (km)
s_d, K_d	Stock capacity (parcels/day) and number of available vans at depot d
Q, c, F	Van capacity (parcels), variable cost (₹/km), fixed cost per van route (₹)
u_{dj}	Unit cost of supplying zone j from depot d in the transportation model: $2c\, d_{dj}/Q$
$y_{dj} \geq 0$	Parcels allocated from depot d to zone j (Stage 1)
$x_{ij} \in \{0,1\}$	1 if a van travels directly from node i to node j (Stage 2)
$a_{jd} \in \{0,1\}$	1 if zone j is assigned to depot d (Stage 2)
$L_i \geq 0$	Cumulative load delivered on the route up to and including zone j (Stage 2)

B. Stage 1: Transportation Model (Allocation Benchmark)

The transportation model allocates each zone's demand to depots at minimum cost, where the unit cost u_{dj} is the van-kilometre cost per parcel of a full round trip:

$$\min \ \Sigma_{d\in D} \Sigma_{j\in Z} u_{dj}\, y_{dj} \tag{1}$$

$$\Sigma_{d\in D}\, y_{dj} = q_j, \quad \Sigma_{j\in Z}\, y_{dj} \leq s_d, \quad y_{dj} \geq 0 \qquad \forall\, j \in Z,\, d \in D \tag{2}$$

C. Stage 2: Multi-Depot Capacitated Routing MILP

The routing model chooses arcs, depot assignments and route loads jointly:

$$\min \ \Sigma_{(i,j)\in A}\, c\, d_{ij}\, x_{ij} + F \Sigma_{d\in D} \Sigma_{j\in Z}\, x_{dj} \tag{3}$$

$$\Sigma_{i\in N,\, i\neq j}\, x_{ij} = 1 \qquad \forall\, j \in Z \tag{4}$$

$$\Sigma_{k\in N,\, k\neq j}\, x_{jk} = 1 \qquad \forall\, j \in Z \tag{5}$$

$$\Sigma_{d\in D}\, a_{jd} = 1 \qquad \forall\, j \in Z \tag{6}$$

$$\Sigma_{j\in Z}\, x_{dj} = \Sigma_{j\in Z}\, x_{jd} \leq K_d \qquad \forall\, d \in D \tag{7}$$

$$\Sigma_{j\in Z}\, q_j\, a_{jd} \leq s_d \qquad \forall\, d \in D \tag{8}$$

$$x_{dj} \leq a_{jd}, \quad x_{jd} \leq a_{jd} \qquad \forall\, d \in D, j \in Z \tag{9}$$

$$a_{id} - a_{jd} \leq 1 - x_{ij}, \quad a_{jd} - a_{id} \leq 1 - x_{ij} \qquad \forall\, i \neq j \in Z,\, d \in D \tag{10}$$

$$L_j \geq L_i + q_j - Q(1 - x_{ij}) + (Q - q_i - q_j)\, x_{ji} \qquad \forall\, i \neq j \in Z \tag{11}$$

$$q_j \leq L_j \leq Q \qquad \forall\, j \in Z \tag{12}$$

$$\Sigma_{d\in D} \Sigma_{j\in Z}\, x_{dj} \geq \lceil \Sigma_{j\in Z}\, q_j \,/\, Q \rceil \tag{13}$$

$$x_{ij} \in \{0,1\}, \ a_{jd} \in \{0,1\}, \ L_j \geq 0 \tag{14}$$

D. Interpretation of the Model

Equation (3) is the objective: total distance-based cost plus a fixed cost for every route that leaves a depot, so the number of vans used is priced. Constraints (4) and (5) require each zone to be entered and left exactly once, so every demand is served and no zone

is skipped or split. Constraint (6) assigns each zone to one depot — the allocation, or transportation, element. Constraint (7) ensures that vans leaving a depot return to it and that no more than the K_d available vans are used. Constraint (8) limits the parcels drawn from each depot to its stock. Constraints (9) and (10) link routes and assignment: a van may use a depot arc only for a zone assigned to that depot, and two consecutive zones on a route must share a depot. Constraints (11) and (12) are the vehicle-capacity and route-flow constraints: L_j accumulates the load along a route and never exceeds Q, and the same ordering logic eliminates subtours; (11) is the lifted Miller–Tucker–Zemlin form of Desrochers and Laporte [12], [11]. Constraint (13) is a valid inequality that bounds the number of vans from below and tightens the linear relaxation without cutting off any feasible plan. Constraint (14) states the variable domains.

E. Model Size

With 14 nodes, the instance has 180 binary arc variables (132 zone-to-zone arcs and 48 depot–zone arcs), 24 binary assignment variables and 12 continuous load variables. Although small, this compact formulation already requires minutes of solver time, which illustrates the scaling behaviour discussed in Section III.

IX. SOLUTION APPROACH

A. Implementation

The model is implemented in Python 3. The transportation LP (Stage 1) is solved with SciPy's HiGHS interface; the routing MILP (Stage 2) is built with the Google OR-Tools linear-solver wrapper [31] and solved by SCIP [32] on a single CPU core with a 420-second limit per instance. Each equation (3)–(14) corresponds to one labelled block of code. SCIP returns both the objective value and a proven lower bound, so "optimal" below means an optimality gap of zero.

B. Baseline Dispatch Rule

The baseline mimics rule-of-thumb dispatch and uses the same distances and cost parameters as the optimised plan, so that differences arise only from the plan itself:

1) assign every zone to its nearest depot;
2) from each depot, start a van and repeatedly visit the nearest unvisited assigned zone whose demand still fits in the van's remaining capacity;
3) when no remaining zone fits, return the van to the depot and start a new van; and
4) stop when all zones assigned to the depot are served.

C. Performance Indicators

Plans are compared on five key performance indicators (KPIs): total cost = c × distance + F × vans; total distance; number of vans; average utilisation = parcels delivered ÷ (vans × Q); and cost per parcel = total cost ÷ parcels. The saving is 1 − (optimised cost ÷ baseline cost). A replication file contains the dataset, baseline, both models and the scenario runner and prints every figure reported below. The optimal objective value is unique; where several optimal route sets exist, the particular routes printed may differ between solver versions.

X. RESULTS AND ANALYSIS

All figures in this section are calculated from the dataset in Section VII and the model in Section VIII; they are properties of the illustrative case, not of Mumbai operations. The base-case MILP was solved to proven optimality in roughly 100 to 130 seconds on a single core across several runs.

A. Optimal Routes

The optimal plan uses three vans (Table IX). Bhiwandi serves a northern and eastern loop (Borivali, Malad, Mulund, Thane), while Andheri East splits the remaining eight zones across two vans loaded to 350 and 325 parcels.

TABLE IX
OPTIMAL ROUTES (BASE CASE)

No.	Sequence	Load / capacity	Distance (km)	Cost (₹)
1	Bhiwandi → Borivali → Malad → Mulund → Thane → Bhiwandi	325 / 350	67.95	1,851.30
2	Andheri East → Ghatkopar → Chembur → Dadar → Kurla → Andheri East	350 / 350	35.98	1,403.72
3	Andheri East → Powai → Colaba → Worli → Bandra → Andheri East	325 / 350	69.78	1,876.92
	Total	**1,000 / 1,050**	**173.71**	**5,131.94**

Source: authors' calculation. Cost = ₹14 × km + ₹900 per route.

B. Baseline versus Optimised Plan

Table X compares the two plans and Fig. 2 maps their routes. The optimised plan saves ₹1,311.74 per day (20.4%), travels 29.41 fewer kilometres (14.5%) and needs one van fewer, raising average utilisation from 71.4% to 95.2%.

TABLE X
BASELINE VERSUS OPTIMISED PLAN (BASE CASE, PER DAY)

KPI	Baseline	Optimised	Change
Total delivery cost (₹)	6,443.68	5,131.94	−20.4%
of which distance cost (₹14/km)	2,843.68	2,431.94	−411.74
of which fixed route cost (₹900/van)	3,600.00	2,700.00	−900.00
Total distance (km)	203.12	173.71	−14.5%
Vans / routes used	4	3	−1
Average van utilisation	71.4%	95.2%	+23.8 pts
Cost per parcel (₹)	6.44	5.13	−20.4%

Source: authors' calculation. Parcels delivered: 1,000 per day in both plans.

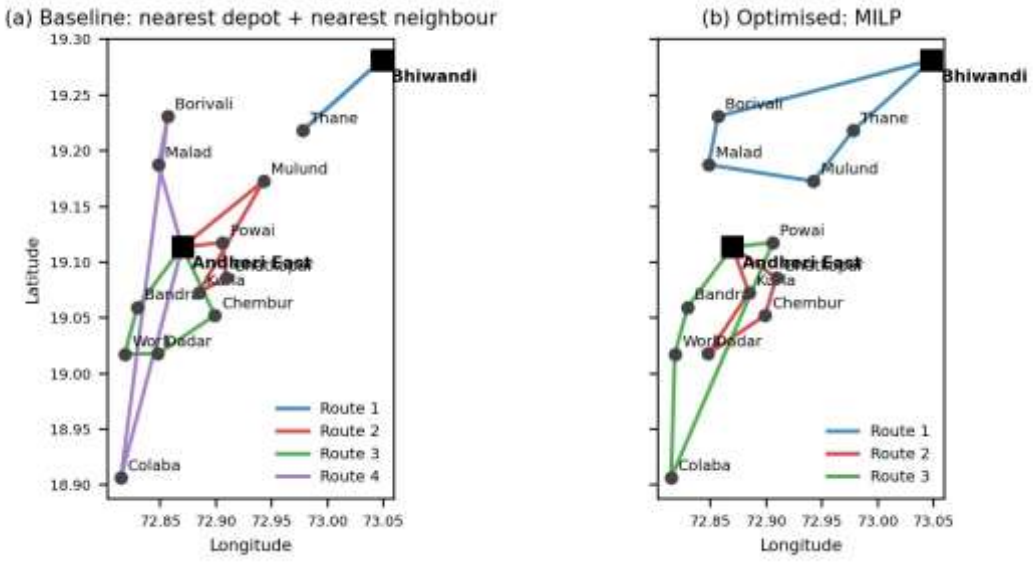


Fig. 2. Baseline and optimised routes (schematic straight-line links between locations, not road paths; ■ = depot)

Worked KPI check. For the baseline, distance cost is ₹14 × 203.12 km = ₹2,843.68 and fixed cost is ₹900 × 4 vans = ₹3,600.00, giving ₹6,443.68; utilisation is 1,000 ÷ (4 × 350) = 71.4% and cost per parcel ₹6.44. For the optimum, ₹14 × 173.71 km = ₹2,431.94 plus ₹900 × 3 = ₹2,700.00 gives ₹5,131.94; utilisation is 1,000 ÷ (3 × 350) = 95.2% and cost per parcel ₹5.13. The saving is 1 − 5,131.94 ÷ 6,443.68 = 20.4%.

C. Sources of the Saving (RQ2, RQ3)

Fig. 3 decomposes the saving. Of the ₹1,311.74 saved per day, ₹411.74 (31.4%) comes from 29.41 fewer kilometres and ₹900.00 (68.6%) from running three vans instead of four. The baseline's nearest-depot rule sends only Thane (85 parcels, 24% of a van) to Bhiwandi, creating a nearly empty route. The optimum instead uses Bhiwandi for a fuller northern and eastern loop. This challenges the assumption that the closest depot is the right allocation rule: Malad is closer to Andheri East, yet it is served from Bhiwandi because doing so fills the Bhiwandi van and removes a route.

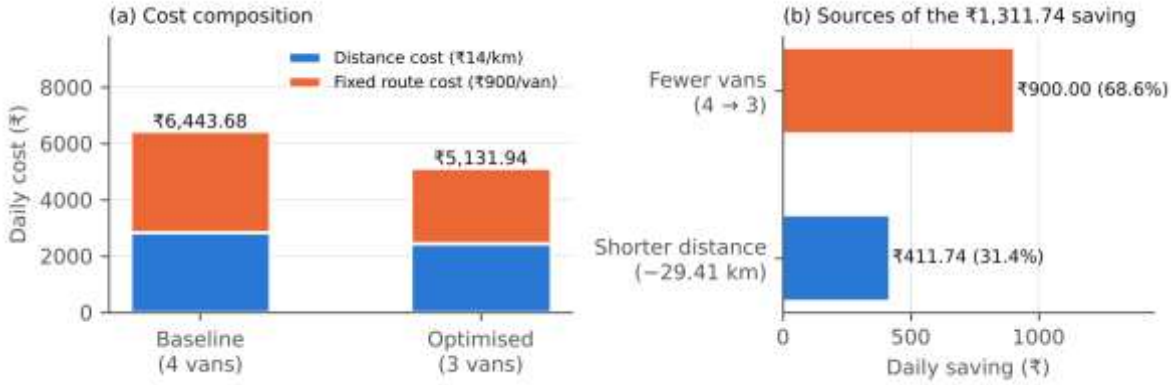


Fig. 3. Cost composition of the two plans (a) and sources of the daily saving (b)

D. Transportation Benchmark (RQ1)

The Stage 1 LP, which ignores vans and sequencing, allocates Borivali, Mulund and Thane entirely and 80 of Malad's 105 parcels to Bhiwandi, and everything else — including the remaining 25 Malad parcels — to Andheri East (Table XI). Its zone-to-depot allocation matches the MILP exactly except that it splits Malad, which the routing model cannot do because each zone is visited

once. The transportation view is therefore a good guide to allocation but cannot see the fleet saving. Its objective (₹1,148.88) is a per-parcel round-trip proxy and is not comparable with the routing cost.

TABLE XI
DEPOT ALLOCATION: TRANSPORTATION LP VERSUS ROUTING MILP

Depot	Transportation LP (Stage 1)	Parcels	Routing MILP (Stage 2)	Parcels
Bhiwandi	Borivali, Mulund, Thane, Malad (80 of 105)	300	Borivali, Malad, Mulund, Thane	325
Andheri East	Remaining eight zones, Malad (25 of 105)	700	Remaining eight zones	675

Source: authors' calculation. Both allocations respect the 700-parcel depot stock limit.

E. Verification of the Optimal Solution

The reported solution was checked against every constraint class. All twelve zones appear exactly once across the three routes (constraints 4–5); every route starts and ends at the same depot (7); route loads of 325, 350 and 325 parcels do not exceed $Q = 350$ (11–12); depot draws of 325 and 675 parcels are within the 700-parcel stock (8); and the three-van fleet equals the lower bound $\lceil 1{,}000/350 \rceil = 3$ in constraint (13). Route costs recomputed as ₹14 × km + ₹900 sum exactly to the objective value of ₹5,131.94.

F. Discussion

Two findings stand out. First, allocation and routing decisions are linked: the transportation view alone recommends a sensible depot split but cannot see that one under-filled route is the main avoidable cost. Second, the compact MILP, although exact, needed roughly 100–130 seconds for the base case and up to about 290 seconds in the scenarios for only twelve zones, consistent with the scaling caution in Section III and with the usual move to heuristics for larger networks [6], [7]. The 20.4% saving is of the same order as the 16.9% multi-depot saving reported by He et al. [24], although the settings and baselines differ and the figures are not directly comparable.

XI. SENSITIVITY ANALYSIS

Seven scenarios change one parameter at a time. Each was re-solved exactly, and the baseline rule was re-applied to the same data. All seven solved to proven optimality in 37 to 286 seconds. Table XII and Fig. 4 report the results.

TABLE XII
SENSITIVITY SCENARIOS (COST IN ₹ PER DAY)

Scenario	Baseline cost	Optimised cost	Saving	Vans (base→opt)	Opt. km	Opt. utilisation	Opt. ₹/parcel	Parcels
Base case	6,443.68	5,131.94	20.4%	4→3	173.71	95.2%	5.13	1,000
Capacity 300	7,678.98	6,541.26	14.8%	5→4	210.09	83.3%	6.54	1,000
Capacity 400	6,113.14	4,863.28	20.4%	4→3	154.52	83.3%	4.86	1,000
Demand −20%	6,320.90	4,863.28	23.1%	4→3	154.52	76.2%	6.08	800
Demand +20%	7,678.98	6,631.14	13.6%	5→4	216.51	85.7%	5.53	1,200
Per-km cost +25%	7,154.60	5,739.93	19.8%	4→3	173.71	95.2%	5.74	1,000
Fixed cost ×2	10,043.68	7,831.94	22.0%	4→3	173.71	95.2%	7.83	1,000
Fleet cap 3	6,443.68	5,131.94	20.4%	4→3	173.71	95.2%	5.13	1,000

Source: authors' calculation. Capacity scenarios use 300 and 400 parcels per van; demand scenarios scale every zone by 0.8 or 1.2 (rounded up); per-km cost +25% = ₹17.50; fixed cost doubled = ₹1,800; fleet cap = at most three vans in total, under which the four-van baseline would not be feasible.



Fig. 4. Daily cost under baseline and optimised plans across scenarios (labels show the saving)

A. Vehicle Capacity

At 300 parcels per van the minimum fleet rises from three to four ($\lceil 1{,}000/300 \rceil = 4$) and optimal cost rises 27.5% to ₹6,541.26. At 400 parcels the fleet cannot fall below three ($\lceil 1{,}000/400 \rceil = 3$), so the extra capacity only shortens routes: distance falls 11.0% to 154.52 km and cost 5.2% to ₹4,863.28. Capacity is therefore valuable only while it changes the minimum van count.

B. Demand

With 20% less demand (800 parcels) three vans are still needed; total cost falls 5.2%, but cost per parcel rises 18.5% to ₹6.08 because fixed route costs are spread over fewer parcels. With 20% more demand (1,200 parcels) a fourth van is required; total cost rises 29.2% while cost per parcel rises only 7.7%, a route-density effect. At the base demand the plan is already at the fleet lower bound with 95.2% utilisation, so growth immediately forces a new van.

C. Cost Rates and Fleet Size

Raising the per-km rate by 25% raises cost by 11.8% but leaves the optimal routes unchanged (173.71 km). Doubling the fixed route cost raises cost by 52.6%, again with identical routes, because three vans is already the minimum feasible fleet and the fixed cost cannot be avoided by further consolidation. Capping the fleet at three vans leaves the optimal plan unchanged (₹5,131.94); a cap of two is infeasible ($2 \times 350 = 700 < 1{,}000$ parcels), and the four-van baseline could not run under the cap. Closing either depot is also infeasible, because each depot holds only 700 parcels of stock against total demand of 1,000. The percentage saving versus the baseline is largest (22.0%) in the high-fixed-cost scenario because the baseline carries a fourth van. Overall, the optimal plan is robust to cost-rate changes but sensitive to capacity and demand (RQ4).

XII. MANAGERIAL IMPLICATIONS

The following points apply to the illustrative case and indicate what a manager should test with real data:

- **Cost efficiency.** The largest lever was fleet size (68.6% of the saving), not shorter driving. Managers should look first at consolidation — the number of routes — before fine-tuning sequences.
- **Vehicle utilisation.** In the baseline one van ran at only 24% load; the optimum raises average utilisation to 95.2%. High utilisation also leaves almost no buffer: a 20% rise in demand adds a van.
- **Customer allocation.** Nearest-depot allocation is not necessarily optimal; allocating by route loop and load, as the transportation benchmark and MILP together show, can remove under-filled routes.
- **Fleet planning.** The lower bound ⌈total demand ÷ capacity⌉ is a quick check on fleet requirements. Larger vans pay off only when they lower that bound; otherwise they mainly shorten routes.
- **Delivery planning.** Because cost per parcel rises when volume falls, low-volume days may justify merging zones or deferring deliveries within service promises.
- **Decision support.** An exact model of this kind is practical for roughly a dozen zones per planning run; larger operations would need heuristics [6], [7] or decomposition, with the exact model used to benchmark them.

A. Implementation Roadmap for Operators

A firm wishing to apply the approach can proceed in four steps. First, assemble a daily demand file by delivery zone and a depot list with stock and fleet limits. Second, replace the circuity-factor distances with road-network distances or travel times from an open routing engine such as OpenStreetMap/OSRM. Third, run the transportation model as a quick allocation check and the routing MILP (or a heuristic benchmarked against it) each evening for the next day's dispatch. Fourth, track the KPIs used here — cost per parcel, van utilisation and vans per day — against the current rule, so that the value of optimisation is measured rather than assumed.

B. Theoretical Implications

For OR teaching and research, the case illustrates three points. The transportation problem remains a useful first approximation of depot allocation, but its linear cost structure hides the fixed cost of vehicles, which is where most of the saving lies in small urban networks. The lower bound ⌈total demand ÷ capacity⌉ in constraint (13) is not only a computational device but also a managerial indicator of whether the current fleet is excessive. Finally, sensitivity results that change cost but not routes (cost rates) should be distinguished from those that change the structure of the solution (capacity and demand); only the latter call for re-planning.

XIII. LIMITATIONS AND FUTURE RESEARCH

Limitations. The dataset is small, and demand, capacities and cost rates are synthetic; distances are scaled great-circle estimates rather than measured road or traffic data; costs are linear in distance with one fixed charge per route; only one vehicle type, one day and twelve zones are modelled; zones are single nodes; the baseline is a simple rule designed for this study, so the percentage savings are specific to it; the compact exact model will not scale to large instances; and solver times were obtained on a single core and are indicative only.

Future research. Natural extensions include real company or open-data demand and road-network travel times (for example from OpenStreetMap/OSRM); time windows, real-time traffic and dynamic re-routing [18]; multiple vehicle types and fleet-mix decisions under stochastic demand [25]; driver-hour limits; micro-hub location and multi-modal delivery [23], [14]; and heuristics for larger networks. Table XIII organises these directions into a research agenda; a full science-mapping analysis of the 108 records retrieved in Section IV would further test the research gap stated in Section V.

TABLE XIII
FUTURE RESEARCH AGENDA

Theme	Research question	Suggested method
Real data	How large is the saving for an actual Mumbai operator's daily demand?	Firm data or open demand proxies; road-network distances (OSRM)
Time and traffic	How do time windows and congestion change optimal routes and fleet size?	VRP with time windows; time-dependent travel times
Uncertainty	How much fleet buffer is needed when demand varies day to day?	Two-stage stochastic or robust MILP [25]
Network design	Where should micro-hubs be located across the city's corridors?	Facility location and two-echelon models [13], [14], [23]
New delivery modes	When do electric two-wheelers, lockers or drones reduce cost per parcel?	Concept-specific MILPs and simulation [19], [22], [2]
Scalability	How close do heuristics come to the exact optimum as zones increase?	Metaheuristics benchmarked against exact MILP solutions [6], [24]
Literature mapping	How has last-mile OR research evolved, and where are Indian studies placed?	Bibliometric analysis per Section IV [27], [29]

XIV. CONCLUSION

This paper combined a transportation allocation benchmark with a multi-depot capacitated routing MILP to plan last-mile deliveries in an illustrative Mumbai case. Solved to proven optimality, the model cut daily cost by 20.4% and distance by 14.5% relative to a nearest-depot, nearest-neighbour baseline, with most of the gain coming from running one van fewer. The transportation model reproduced the optimal depot allocation almost exactly but could not see the fleet saving, which shows why allocation and routing should be optimised together. Scenario testing showed that the plan is robust to cost-rate changes but responds sharply to capacity and demand through the minimum fleet size. Because the data are synthetic, the contribution is a transparent, reproducible method and a set of managerial insights to be tested on real data, not an empirical finding about Mumbai's delivery sector.

ACKNOWLEDGMENT

The authors thank Dr. Vishal Pradhan, faculty for the Operations Research course, for his guidance and feedback during this project.

DATA AND CODE AVAILABILITY

The dataset, replication code and Mendeley library export are available from the corresponding author on reasonable request.

REFERENCES

[1] J. Olsson, D. Hellström, and H. Pålsson, "Framework of last mile logistics research: A systematic review of the literature," *Sustainability*, vol. 11, no. 24, Art. no. 7131, 2019, doi: 10.3390/su11247131.

[2] N. Boysen, S. Fedtke, and S. Schwerdfeger, "Last-mile delivery concepts: A survey from an operational research perspective," *OR Spectrum*, vol. 43, no. 1, pp. 1–58, 2021, doi: 10.1007/s00291-020-00607-8.

[3] M. Savelsbergh and T. Van Woensel, "50th anniversary invited article—City logistics: Challenges and opportunities," *Transportation Science*, vol. 50, no. 2, pp. 579–590, 2016, doi: 10.1287/trsc.2016.0675.

[4] L. Ranieri, S. Digiesi, B. Silvestri, and M. Roccotelli, "A review of last mile logistics innovations in an externalities cost reduction vision," *Sustainability*, vol. 10, no. 3, Art. no. 782, 2018, doi: 10.3390/su10030782.

[5] R. Mangiaracina, A. Perego, A. Seghezzi, and A. Tumino, "Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: A literature review," *International Journal of Physical Distribution & Logistics Management*, vol. 49, no. 9, pp. 901–920, 2019, doi: 10.1108/IJPDLM-02-2019-0048.

[6] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, no. 4, pp. 408–416, 2009, doi: 10.1287/trsc.1090.0301.

[7] P. Toth and D. Vigo, Eds., *Vehicle Routing: Problems, Methods, and Applications*, 2nd ed. Philadelphia, PA, USA: SIAM, 2014, doi: 10.1137/1.9781611973594.

[8] F. L. Hitchcock, "The distribution of a product from several sources to numerous localities," *Journal of Mathematics and Physics*, vol. 20, no. 1–4, pp. 224–230, 1941, doi: 10.1002/sapm1941201224.

[9] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*, 11th ed. New York, NY, USA: McGraw-Hill Education, 2021.

[10] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959, doi: 10.1287/mnsc.6.1.80.

[11] C. E. Miller, A. W. Tucker, and R. A. Zemlin, "Integer programming formulation of traveling salesman problems," *Journal of the ACM*, vol. 7, no. 4, pp. 326–329, 1960, doi: 10.1145/321043.321046.

[12] M. Desrochers and G. Laporte, "Improvements and extensions to the Miller–Tucker–Zemlin subtour elimination constraints," *Operations Research Letters*, vol. 10, no. 1, pp. 27–36, 1991.

[13] M. T. Melo, S. Nickel, and F. Saldanha-da-Gama, "Facility location and supply chain management – A review," *European Journal of Operational Research*, vol. 196, no. 2, pp. 401–412, 2009, doi: 10.1016/j.ejor.2008.05.007.

[14] T. G. Crainic, N. Ricciardi, and G. Storchi, "Models for evaluating and planning city logistics systems," *Transportation Science*, vol. 43, no. 4, pp. 432–454, 2009, doi: 10.1287/trsc.1090.0279.

[15] G. Perboli, R. Tadei, and D. Vigo, "The two-echelon capacitated vehicle routing problem: Models and math-based heuristics," *Transportation Science*, vol. 45, no. 3, pp. 364–380, 2011, doi: 10.1287/trsc.1110.0368.

[16] J. R. Montoya-Torres, J. López Franco, S. Nieto Isaza, H. Felizzola Jiménez, and N. Herazo-Padilla, "A literature review on the vehicle routing problem with multiple depots," *Computers & Industrial Engineering*, vol. 79, pp. 115–129, 2015.

[17] T. Vidal, G. Laporte, and P. Matl, "A concise guide to existing and emerging vehicle routing problem variants," *European Journal of Operational Research*, vol. 286, no. 2, pp. 401–416, 2020, doi: 10.1016/j.ejor.2019.10.010.

[18] V. Pillac, M. Gendreau, C. Guéret, and A. L. Medaglia, "A review of dynamic vehicle routing problems," *European Journal of Operational Research*, vol. 225, no. 1, pp. 1–11, 2013, doi: 10.1016/j.ejor.2012.08.015.

[19] C. C. Murray and A. G. Chu, "The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 86–109, 2015, doi: 10.1016/j.trc.2015.03.005.

[20] N. Agatz, P. Bouman, and M. Schmidt, "Optimization approaches for the traveling salesman problem with drone," *Transportation Science*, vol. 52, no. 4, pp. 965–981, 2018, doi: 10.1287/trsc.2017.0791.

[21] M. Ostermeier, A. Heimfarth, and A. Hübner, "The multi-vehicle truck-and-robot routing problem for last-mile delivery," *European Journal of Operational Research*, vol. 310, no. 2, pp. 680–697, 2023, doi: 10.1016/j.ejor.2023.03.031.

[22] F. Arnold, I. Cardenas, K. Sörensen, and W. Dewulf, "Simulation of B2C e-commerce distribution in Antwerp using cargo bikes and delivery points," *European Transport Research Review*, vol. 10, Art. no. 2, 2018, doi: 10.1007/s12544-017-0272-6.

[23] P. Stokkink and N. Geroliminis, "On the optimal micro-hub locations in a multi-modal last-mile delivery system," *Transportation Research Part E: Logistics and Transportation Review*, vol. 203, Art. no. 104344, 2025, doi: 10.1016/j.tre.2025.104344.

[24] Y. He, M. K. Lim, F. Zhou, and S. Dai, "An improved adaptive large neighborhood search algorithm on collaborative last mile delivery with roaming customers," *International Journal of Industrial Engineering Computations*, vol. 16, no. 4, pp. 1013–1024, 2025, doi: 10.5267/j.ijiec.2025.8.003.

[25] P. Beatrici, S. Birolini, F. Maggioni, and P. Malighetti, "Stochastic fleet size and mix consistent vehicle routing problem for last mile delivery," arXiv preprint arXiv:2512.09764, 2025.

[26] R. Jazemi, E. Alidadiani, K. Ahn, and J. Jang, "A review of literature on vehicle routing problems of last-mile delivery in urban areas," *Applied Sciences*, vol. 13, no. 24, Art. no. 13015, 2023, doi: 10.3390/app132413015.

[27] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, 2021, doi: 10.1016/j.jbusres.2021.04.070.

[28] S. Verma and A. Gustafsson, "Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach," *Journal of Business Research*, vol. 118, pp. 253–261, 2020, doi: 10.1016/j.jbusres.2020.06.057.

[29] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975, 2017, doi: 10.1016/j.joi.2017.08.007.

[30] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: 10.1007/s11192-009-0146-3.

[31] L. Perron and V. Furnon, "OR-Tools," Google. [Online]. Available: https://developers.google.com/optimization

[32] K. Bestuzheva *et al.*, "The SCIP Optimization Suite 8.0," arXiv preprint arXiv:2112.08872, 2021.