# COST SAVINGS USING LOAD CONSOLIDATION AND MULTI-STOP ROUTING

**R.W. Oltmann[1] and J.H. Bührmann[2*]**

1 School of Industrial Engineering, North-West University, South Africa | rudi.oltmann@gmail.com

2 School of Industrial Engineering, North-West University, South Africa | joke.buhrmann@nwu.ac.za

## ABSTRACT

The research investigates opportunities for cost reduction in a secondary distribution for a fourth-party logistics provider operating in East London, South Africa. The logistics provider uses a transport management system that effectively minimises procurement costs through reverse auctions but does not fully optimise multi-stop routing and load utilisation. Historical shipment, loading, and vehicle data were analysed to solve this gap and evaluate different load consolidation and routing approaches. Among these viable solutions were clustering, load consolidation using a linear programming model, and a multi-stop vehicle routing problem using Open Door Logistics Studio software. The final design incorporated a multi-stop consolidation, allowing deliveries to multiple destinations per route while accounting for vehicle capacity and cost constraints. The results showed significant improvements by reducing the total number of trucks used by more than 50%. These findings highlight the value of vehicle routing-based approaches for logistics providers in complex distribution environments and how they can be applied in similar logistics networks.

**Keywords:** 4PL, vehicle routing, logistics, multi-stop routing

* Corresponding author

# 1 INTRODUCTION

Efficient logistics have become increasingly critical as customers demand faster and more reliable deliveries. Traditional in-house logistics functions are no longer sufficient to meet modern demand, prompting many companies to outsource their logistics to third-party logistics (3PL) providers. These firms typically own their own fleets and distribution networks, allowing their client companies to focus on their core activities while outsourcing any transportation tasks [1].

More recently, fourth-party logistics providers (4PLs) have emerged, focusing on strategic coordination and optimisation rather than asset ownership. 4PLs integrate multiple 3PLs and supply chains and support the optimisation of 3PLs' logistics. Many well-known international companies have adopted this model, allowing businesses to benefit from streamlined operations without the complexity of managing their assets directly [2].

Central to the operations of 4PLs is the Transport Management System (TMS), which integrates data from multiple sources to plan, optimise, and execute logistics operations. A TMS provides significantly improved visibility across networks and enables collaboration among supply chain partners [3].

This study focuses on a 4PL company headquartered in Johannesburg, South Africa. The company provides logistics strategy, 3PL procurement, routing and execution services. The company heavily relies on its TMS, which allows visibility of distance travelled, costs, weight, utilisation and trip details. The system integrates with 3PLs' Economic Resource Planning (ERP) systems. A key feature of the company's TMS is the use of a reverse auction system, where 3PLs competitively bid for shipments offered by shippers, ensuring that transport is secured at near-minimum base rates [4].

Despite these advantages provided by the TMS, the company continues to face cost-related challenges in the East London region. Most importantly, a large variation in types of customers is seen in this area. These customers range from large retailers who order full truckloads of 34-36 tons, to small spaza shops in remote areas, which typically order small shipments of less than 500kg at once.

This variability in shipment sizes results in an increased volume of empty space within trucks and necessitates the deployment of a greater number of trucks required to fulfil all shipments. Consequently, this escalates operational costs. The combination of these factors within the company's current circumstances indicates significant potential for finding hidden cost savings within the TMS.

The overall aim of this research therefore is to investigate any untapped cost savings by improving how the TMS manages load consolidation and route planning. To complete this, the following deliverables are required by the end of this research:

1. Conduct a root cause analysis on historical TMS data;
2. Identify various solution approaches to solve the root causes; and
3. Compare the proposed solution approaches against historical data to determine the best approach.

It should be noted that statistical analysis (significance, sensitivity or runtimes) falls outside the scope of this project.

# 2 LITERATURE REVIEW

A comprehensive literature review was conducted to explore potential strategies for enhancing the current state of the TMS. The objective of this literature review was to underscore the significance of evaluating multiple solution approaches from external sources. The investigation included various methodologies, principles, and engineering tools that could be employed to further refine the TMS and address the current challenges.

## 2.1 Load consolidation strategies

Load consolidation fundamentally involves the combination of multiple smaller shipments into single loads for dispatch in a singular vehicle. This approach optimises any unutilised capacity, while balancing quality, cost, and transport time requirements [5], [6].

Hall [7] identified three primary consolidation forms: inventory consolidation, vehicle consolidation, and terminal consolidation, each addressing different operational challenges. The most relevant of the three for this research, vehicle consolidation, involves the loading and unloading of items at various origin and destination locations along a vehicle's route. Mutlu et al. [8] further categorised vehicle consolidation into three policies:

1. Time-based policies (consolidating shipments within predetermined intervals),
2. Quantity-based policies (holding orders until a target load accumulates), and
3. Hybrid time-quantity policies (a combination of the previous two approaches).

However, the effectiveness of consolidation depends significantly on contextual factors. Wu [9] demonstrated that cargo consolidation proves especially viable when incoming container utilisation is low, or when unit vehicle costs are high. Hernández et al. [10] further demonstrated that carrier collaboration through load consolidation under dynamic capacity conditions can significantly reduce empty miles and improve overall network efficiency, particularly in less-than-truckload operations. When utilisation falls below a certain threshold, opportunities emerge to combine partial loads and reduce costs. However, consolidation may lead to delayed deliveries and additional costs if not properly balanced against the timing requirements [11].

Within the specific context of auctioning by use of TMSs or other methods, Mesa-Arango and Ukkusuri investigated combinatorial auctions within less-than-truckload operations [4]. This method of auction allows shippers to bid on segments of a truckload, as opposed to bidding on an entire truck. This method of consolidated bidding substantially outperformed non-consolidated approaches by increasing profits for both shippers and carriers while reducing the overall costs.

Within the solution approach of this research, load consolidation effectively corresponds with the existing load planning challenges faced by the company. Combinatorial auctions also align well with the existing TMS platform in its use of reverse auctions but fell outside the scope of this investigation.

## 2.2 Multi-stop consolidation

Multi-stop truckload optimisation, also termed milk run logistics due to its origins in the dairy industry, represents a flexible approach to delivery consolidation [12]. Unlike traditional milk run models with repeating routes, multi-stop truckload optimisation allows for dynamic addition of pickups, deliveries, and stops based on operational requirements.

Stenger et al. identified key variables impacting transportation cost and carrier acceptance rates, including, but not limited to distance, number of stops, lead time, stop-off charges, and more [13]. While additional charges may apply for multi-stop deliveries, overall costs remain lower than dispatching multiple vehicles for individual shipments, particularly when sophisticated routing algorithms account for practical operational constraints.

But despite the drawback of possible carrier rejection, multi-stop truckload optimisation offers multiple benefits, as documented [14]. These include a 20-37% reduction in transportation cost, a 10-35% improvement in vehicle loading rates, shortened travel distances, enhanced flexibility, and significant inventory reductions. Further case studies also continue to show these significant improvements. In this study, the implementation of multi-stop routing, in conjunction with load consolidation, improved the current system's performance significantly.

## 2.3 Vehicle Routing Problem applications

The Vehicle Routing Problem (VRP) represents one of the most studied topics in operational research, addressing the challenge of determining optimal delivery routes for vehicle fleets [15]. Recent comprehensive reviews have documented substantial evolution of VRP methodologies, with over 30 distinct VRP variants identified that extend the classical formulation to address increasingly complex real-world operational constraints [16].

More recent advances in VRP formulations incorporate multiple criteria, including vehicle capacity constraints, time restrictions, multiple depot considerations, and 2-dimensional space requirements, among others [17], [18]. These VRP formulations are significantly more advanced than their original counterparts, with new variants continuously being developed to incorporate additional criteria and constraints to better model real-world operations. Of special interest to this study is the advances to incorporate vehicle capacity restrictions in the capacitated VRP (CVRP),and the inclusion of multiple vehicle types with varying capacities and costs in the Vehicle Fleet Mix Problem (VFMP) [17].

For large-scale problems, clustering approaches provide a computationally efficient way of grouping delivery locations before optimising. Various clustering algorithms each offer different advantages: K-means/K-medians performs well for equally sized clusters, Sweep/ Sweep-Nearest naturally account for any capacity constraints (within CVRPs) and hybrid methods such as the DBSCAN clustering based on density and K-means with demand balancing provide a combined benefit [19], [20].

# 3 METHODOLOGY

## 3.1 Research design

The research employs a mixed-method approach combining a root cause analysis with quantitative optimisation modelling. This methodology comprises three distinct phases: diagnostic analysis to identify inefficiencies, solution development comparing various solution approaches and further testing of the final selected solution approach with historical data.

## 3.2 Data collection and preparation

This study utilised two primary datasets provided by the company, covering a 13-month period from January 2024 to February 2025. The load-level dataset contained 23 000 entries representing overall truck performance, while the shipment-level dataset comprises over 46 000 entries documenting individual shipments sent during this period. After applying data filtering focused on the operations in East London, it reduced the loadings dataset to approximately 2 000 load entries, and approximately 2 200 shipment entries.

## 3.3 Root cause analysis framework

Root cause analysis employed three different techniques to identify the main problems:

1. **Data analysis:** The systematic examination and thorough exploration of the data to identify potential issues and comprehend the feasible and infeasible approaches within the data.
2. **Fishbone diagram:** Systematic categorisation of potential causes across multiple domains.
3. **Pareto analysis:** Quantification of the defect frequencies as identified by the fishbone diagram to prioritise improvement opportunities.

## 3.4 Optimisation model development

The solution approach integrates insights from the root cause analysis alongside potential solutions identified in the literature review to define the design requirements and determine feasible implementation options. It is important to note that direct implementation is not possible due to the lack of access to the TMS's source code. Instead, various suggested solution approaches were tested as proof-of-concepts to demonstrate which solution would deliver the best results within the company's specific context.

### *3.4.1 Scalability considerations*

While the direct multi-stop VRP approach demonstrated superior optimisation results, it is important to note the trade-off between solution quality and computational scalability. The clustered VRP approach, though somewhat less optimal, offers greater advantages for larger-scale implementations. By decomposing the routing problem into smaller, more manageable subproblems, the clustering significantly reduces the computational complexity, making it more suitable for large datasets. Conversely, the direct multi-stop VRP faces increasing computational time as the problem size grows. Organisations aiming to scale this framework across multiple regions (or on a general larger scale) should consider this trade-off.

# 4 ROOT CAUSE ANALYSIS

The root cause analysis presented below encompasses all relevant factors that inform the final design solution. The investigation covered various aspects, including the types of trucks used, their utilisation rates, the quantity of each truck type used within the datasets, shipping and loading costs (considering distance, weight, and volume), distance travelled dependant on truck type, and delivery frequency during singular trips. For the sake of conciseness, this study focused exclusively on highlighting the potential problems identified.

## 4.1 Data analysis

### *4.1.1 Fleet composition & utilisation analysis*

Figure 1 indicates smaller vehicles (3-ton, 16-ton, 24-ton flatdecks) demonstrate consistently poor utilisation, with each operating below 80% utilisation (otherwise known as load factor) on average, compared to bigger vehicles (34-ton -36-ton tautliners).

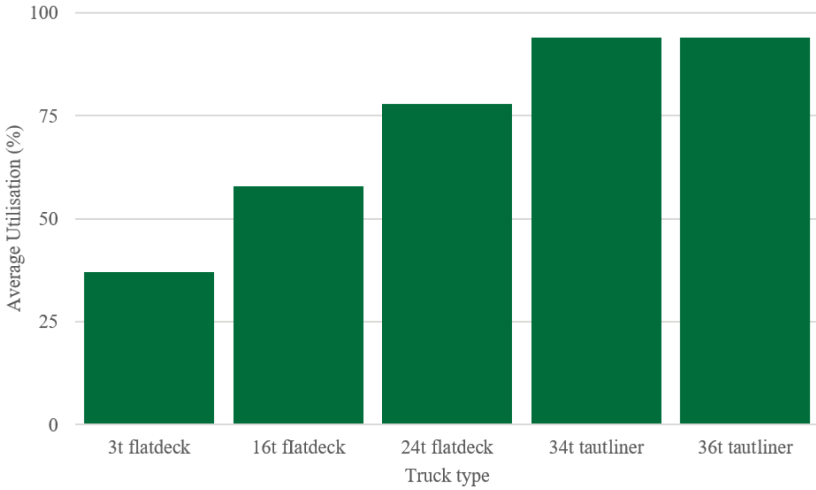


**Figure 1: Average utilisation of each vehicle type**

### *4.1.2 Delivery pattern analysis*

The loadings dataset contained data on the number of shipments, number of deliveries, number of stops, and number of pickups during a particular trip. The analysis is shown in Figure 2. The investigation revealed a critical inefficiency: more than 85% of trucks involved single deliveries with one shipment per truck. This pattern indicated minimal to no exploitation of consolidation opportunities, with trucks consistently not being well utilised and returning empty after singular deliveries.

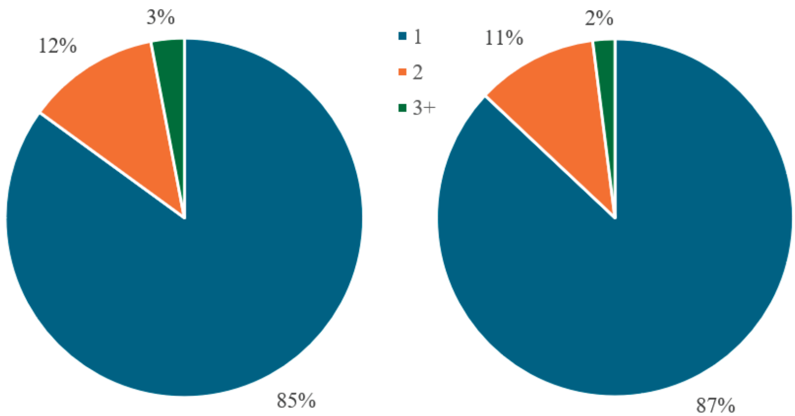


Figure 2a. Number of shipments per trip

Figure 2b. Number of deliveries per trip

**Figure 2: Number of shipments (Fig 2.a) and deliveries (Fig 2.b) made during one truck trip**

## 4.2 Fishbone diagram

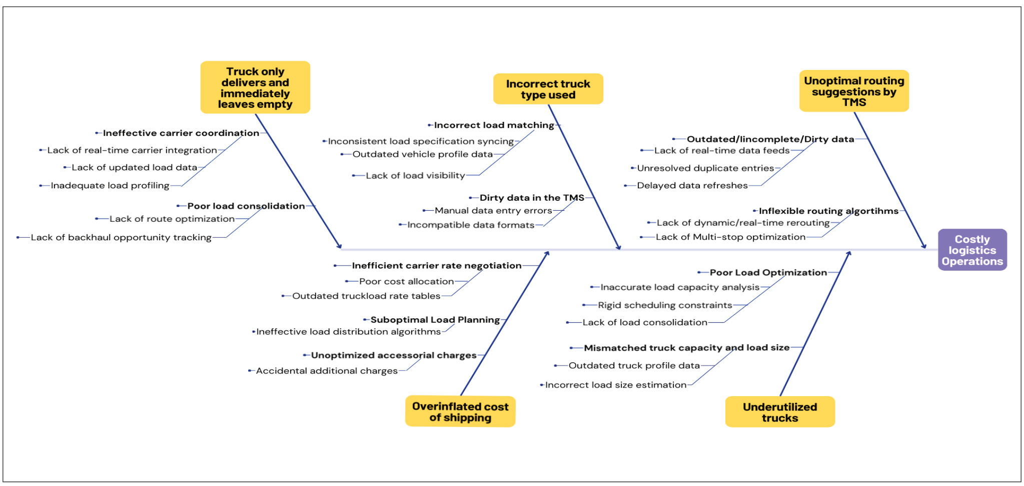


**Figure 3: Fishbone diagram for the root cause analysis**

Figure 3 presents a fishbone diagram illustrating five main areas that may contribute to costly logistics operations. Each area branches out to identify the underlying root causes of each of these issues. This tool was used to gain a comprehensive understanding of the problem, assess which root causes are addressable, which may be beyond control (within the scope of this research), and to verify whether historical data supports the existence of these root causes.

The historical data analysis, in conjunction with the fishbone diagram, revealed five key root causes that significantly impact the logistics efficiency:

1. **A lack of backhauling**
   The scarcity of backhaul trips – return journeys that carry cargo – leads to increased empty miles and higher operational costs;
2. **Mismatched truck capacity & load size**
   Inefficient pairing of truck sizes with shipment volume results in underutilised capacity, negatively affecting cost and performance;
3. **Ineffective load distribution algorithms**
   Current algorithms for allocating loads to trucks fail to optimise for cost and efficiency, causing suboptimal routing and resource use;
4. **Lack of multi-stop optimisation**
   The lack of route planning that consolidates multi-stops into a single trip diminishes delivery efficiency and increases travel time and expenses;
5. **Poor cost allocation**
   Poorly designed cost allocation practices obscure the true shipping expenses, leading to misinformed decisions and financial inefficiency.

## 4.3 Pareto Analysis

To effectively use a Pareto chart, the causes of defects were identified. The key issues were identified based on the fishbone diagram and data analysis. These are outlined below along with the criteria used to test their occurrence in the historical dataset:

1. **Lack of backhauling opportunity**
   A truck returning empty is flagged when the number of pick-ups equals zero.
2. **Low truck utilisation**
   Total utilisation per truck (load factor) is lower than the capacity of a vehicle with smaller capacity, indicating that a smaller truck could have been chosen with lower travelling costs to improve efficiency.
3. **Single shipments causing low utilisation**
   When a truck is less than 80% utilised, but only carries a single shipment, where shipments to the same delivery location could have been consolidated to improve efficiency.
4. **Single deliveries causing with low utilisation**
   When a truck is less than 80% utilised, but only carries items to a single delivery location, suggesting a missed opportunity to combining deliveries to improve efficiency.
5. **Cost-distance mismatch**
   A truck travelling a certain distance should incur costs within an expected range; deviations from this indicate possible inefficiencies or misallocations.

Figure 4 depicts the Pareto chart that was constructed from the available data. The figure highlights that 80% of the logistical issues stem from three main issues: low truck utilisation, single shipment with low utilisation, and single delivery with low utilisation. Therefore, the proposed solution should primarily focus on addressing these three problems.

Furthermore, the TMS did not record empty trips meaning the impact of the lack of backhauling could not be analysed effectively. Consequently, it was difficult to determine with certainty whether backhauling was actually taking place. However, the company indicated that carriers typically adjust pricing to offset empty return trips, effectively managing this concern in practice. As a result, backhauling was considered a lower priority that had zero number of defects and did not require further focus.

In summary, the root cause analysis suggested that the key problems to resolve were:

- Low utilisation of trucks, where vehicles with smaller load capacity were available,
- A single shipment occurred when truck utilisation was low and shipments to the same delivery point could have been combined, and
- A single delivery occurred when truck utilisation was low and multiple deliveries to locations in close proximity, for example, could have been combined.

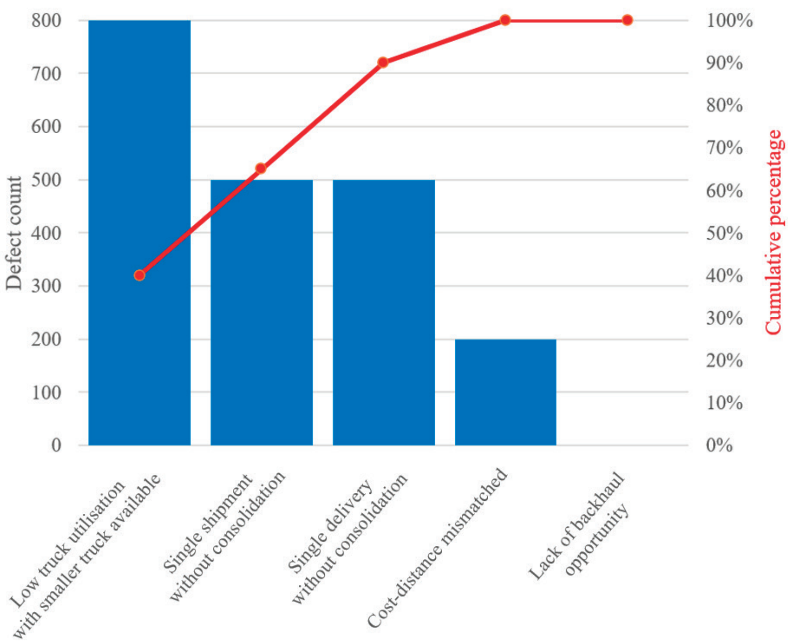


**Figure 4: Pareto chart analysis showing different defects within historical data**

# 5 FINAL DESIGN

The proposed solution approach identified inefficiencies through integrated multi-stop consolidation and routing optimisation. Unlike simple load consolidation targeting same-destination shipments, the design enabled deliveries along common routes, maximising utilisation while minimising total distance. Shipping and loading costs are highly variable, depending on season, market conditions (high or low volume movement), and truck availability. For these reasons, the costs in the historical datasets were not suitable for comparison with the solution results. Instead, distances and the number of vehicles used were provided to allow the reader to find the cost savings in comparison to the current market conditions.

Three load consolidation strategies were evaluated:

1. **A simplified Linear Programming Model (LP Model):** consolidating loads to the same delivery location without considering vehicle routing.
2. **Clustered multi-stop VRP:** grouping delivery locations followed by vehicle routing optimisation.
3. **Multi-stop VRP:** direct optimisation across all delivery points without upfront clustering.

The three consolidation strategies represent a progressive hierarchy of optimisation complexity. The simplified LP model did not consider any vehicle routing and served as the baseline benchmark, representing the most fundamental consolidation approach achievable without sophisticated routing algorithms. This delivery location consolidation strategy deliberately excluded routing considerations, consolidating only shipments destined for the same delivery location. It required minimal technical expertise to implement manually. A logistics coordinator could theoretically execute this strategy using basic spreadsheet tools by grouping shipments by destination location.

To enable consolidation across the dataset, it was essential to segment the data by capacity and time, as discussed in the literature review. Given the historical nature of the dataset, splitting by capacity or exact timing could produce impractical solutions, such as consolidating shipments that are months apart. Therefore, a practical approach was to divide the dataset into week-long intervals.

This approach assumed that shipments within the same week were eligible for consolidation, which simplified solution testing without handling over 2 000 shipments simultaneously. Additionally, for comparison purposes, the number of shipments in the historical dataset was assumed to be equal to the number of trucks used. This provided a straightforward benchmark to evaluate improvements from optimised vehicle assignments.

For both VRP-based approaches, Open Door Logistics (ODL) Studio software [21] was used as the optimisation routing platform. ODL Studio is a free and practical tool for VRP analysis that supports multiple vehicle types, considers vehicle capacity constraints, incorporates per-kilometre cost variations by vehicle type, and allows precise location mapping using latitude and longitude data. It also enables modelling of multi-stop deliveries along a single route, making it well-suited for evaluating historical data in the East London region.

ODL Studio was set to allow for an infinite fleet size (no data or information is available showing any limit on trucks), using the 'Great Circle' distance setting. Otherwise, the base vehicle routing algorithm was used, and any non-relevant columns were left empty.

## 5.1 Linear Programming Model

The simplified LP model assumed that two or more shipments travelling to the same delivery location were eligible for consolidation. This approach did not yet incorporate VRP routing but was used to test whether simple load consolidation could reduce costs before considering more advanced solutions.

The LP model used an exact method to find the most optimal answer, with the following objective function and constraints:

$$Minimise \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \quad (1)$$

subject to:

$$\sum_{j=1}^{n} x_{ij} \geq 0 \ \forall i = 1,2,\dots,m \quad (2)$$

$$\sum_{j=1}^{n} x_{ij} * c_j \geq W_i \;\; \forall\, i = 1,2, \dots, m \qquad (3)$$

$$\sum_{i=1}^{m} x_{ij} \leq 1 \; \forall\, j = 1,2, \dots, n \qquad (4)$$

$$x_{ij} \in (0\,,1) \; \forall\, i = 1,2, \dots, m; \; \forall\, j = 1,2, \dots, n \quad (5)$$

where $i$ represents the delivery locations, $m$ the total number of locations, $j$ the individual vehicles available and $n$ the total number of trucks. $x_{ij}$ represents the binary decision variable indicating if vehicle j was allocated to location $i$. Furthermore, $c_j$ represents for the capacity of vehicle $j$ – varying based on vehicle type for the 3–36-ton trucks that were available, and $W_i$ the total weight required to be delivered to a specific location $i$.

The minimisation function in Eq. (1) aims to use the least number of trucks possible. Note that the LP model does not use latitude and longitude values; it consolidates loads based only on shipment weight and location name.

The constraint in Eq. (2) ensures that at least one truck was assigned to every delivery location $i$. Eq. (3) ensures the total weight to be delivered to a location $i$ is limited to the sum of the total capacity of all vehicles allocated to the same location. Eq. (4) ensured that a vehicle $j$ could only be allocated to a single delivery location and Eq. (5) restricted the $x_{ij}$ decision variables to binary values.

As an example, the first week of the dataset was extracted. We refer to this as the first test set, containing 91 shipments requiring deliveries to various locations within the East London region. Table 1 provides a list of the model solution results based on this test set, as well as a sample of ten destinations from the test set. The LP model's results, as seen in Table 1, end up being a significant improvement over the original number of trucks used in the historical dataset, with an improvement from 91 trucks used and shipments sent, to only 36 trucks used to complete the deliveries. However, one may notice that a large majority of trucks were still only carrying one shipment to their given destination, and additionally the utilisation of these trucks was extremely low (only the first ten entries are shown).

A summary of the total number of vehicles per truck type for the solution is given in Table 2, indicating that only 3-ton and 36-ton trucks were used within the LP's solution. This is likely due to very few lone shipments being within the 3-ton to 34-ton range. This solution was likely to improve significantly during the following two solution approaches, where lone shipments could be combined based on delivery location groupings.

**Table 1: The LP Model solution results for the first test set, listing the first ten destinations**

| Objective value: 36 | | | |
|---|---|---|---|
| **Total shipments: 91** | | | |
| **Delivery Location** | **Total Weight** | **Number of Shipments** | **Assigned trucks** |
| Aliwal North | 372 | 1 | 1 |
| Amatola View | 1 216 | 1 | 1 |
| Amatoleville | 1 223 | 5 | 1 |
| Beacon Bay | 361 | 2 | 1 |
| Cambridge | 2 293 | 3 | 1 |
| East London | 120 446 | 42 | 4 |
| Flagstaff | 64 512 | 2 | 2 |
| Fort Beaufort | 50 | 1 | 1 |
| Gately | 367 | 2 | 1 |
| Grahamstown | 32 256 | 1 | 1 |

**Table 2: A summary of the total number of vehicles per truck type for the LP model solution for the first test set**

| Total trucks: 36 | |
|---|---|
| **Truck Type** | **Total number per truck type** |
| 3t flatdeck | 13 |
| 16t flatdeck | 0 |
| 24t flatdeck | 0 |
| 34t tautliner | 0 |
| 36t tautliner | 23 |

## 5.2 Clustered multi-stop VRP

To reduce the computational complexity and improve the scalability of large-scale VRPs, clustering methods are often used as a preprocessing step to group delivery points prior to optimisation. Clustering creates geographically cohesive sets of delivery locations that can then be assigned to vehicles in a second stage or route planning. For more information on clustering and its effects please refer to Bührmann and Bruwer [22], where an in-depth explanation is given and exploration done of clustering and its effects.

Several clustering algorithms were initially tested in this study, including DBSCAN, Sweep, Sweep-Nearest, K-means and K-medians. Both K-means and K-medians performed relatively well in grouping delivery locations spatially. These are shown in Figures 7 and 8 respectively. However, these algorithms produced clusters that were highly imbalanced in terms of shipment weight. Such imbalance results in impractical or infeasible routing solutions, as overloaded clusters need to be serviced by multiple vehicles, while underloaded clusters lead to low utilisation and inefficiencies. This is a common limitation of purely distance-based clustering approaches, which tend to minimise spatial dispersion without accounting for logistical constraints such as vehicle capacity [19].

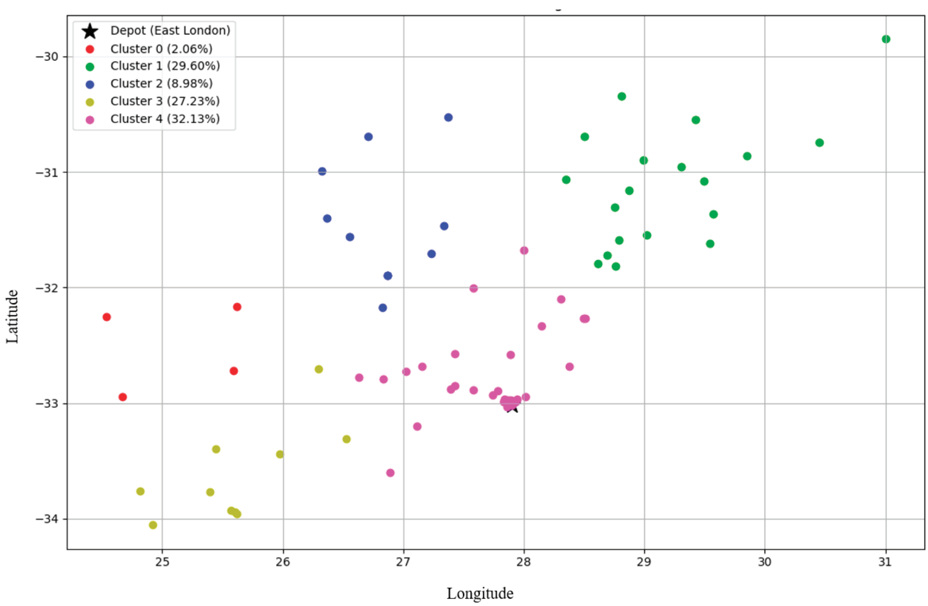


**Figure 7: K-means clustering results the first test set, with destinations in the East London region**

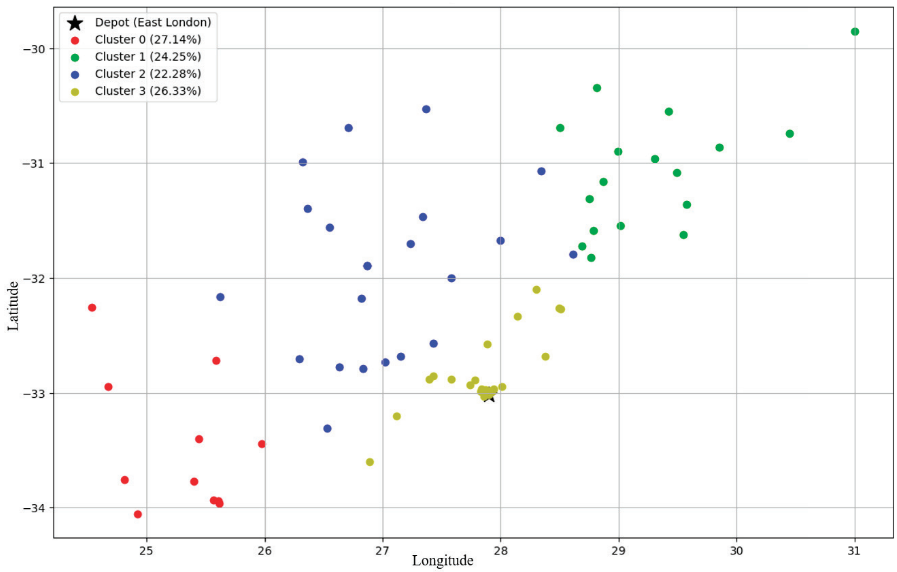


**Figure 8: K-medians clustering results the first test set, with destinations in the East London region**

To overcome this challenge, a demand-balanced variation of K-means clustering was implemented. This selection of demand-balanced K-means was driven by the specific operational constraints of the East London distribution network. Standard K-Means clustering, while computationally efficient, fails to account for the heterogeneous shipment weights characteristic of the company's customer base. The demand-balanced variant addressed this limitation by incorporating capacity constraints directly into the clustering algorithm. This modification maintained the spatial efficiency of traditional K-Means while producing clusters that were operationally feasible for assignment to the company's fleet. As shown in Figure 9, this method achieved a significantly more balanced allocation compared to standard K-means or K-medians. This clustering allocation to group delivery locations was then applied to ODL Studio to create a final routing solution, using the same initial test set with 91 shipments.

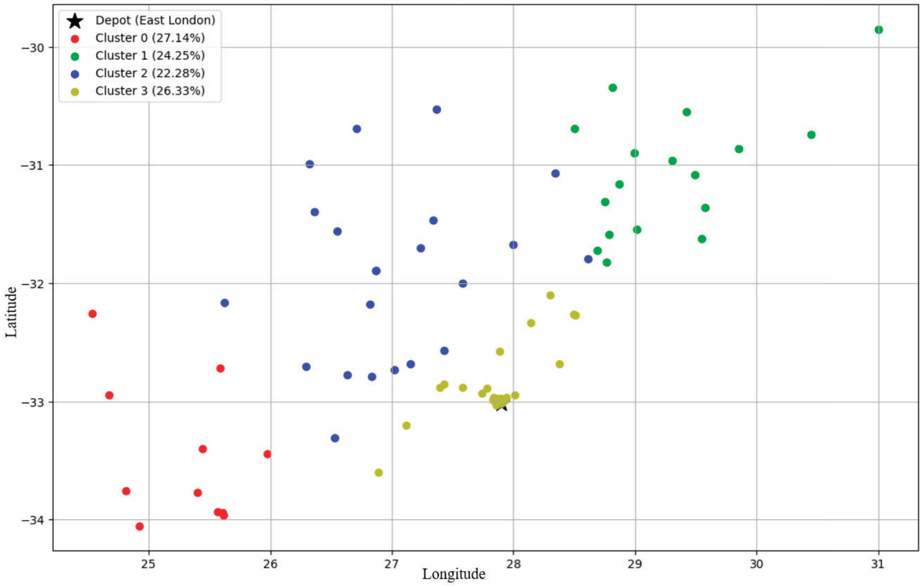


**Figure 9: Demand-balanced clusters for the first test set, with destinations in the East London region**

Table 3 presents the ODL Studio results for the demand-balanced clustered solution. Compared to the LP model, the clustered approach improved the vehicle utilisation and further reduced the total fleet size from 39 for the LP solution to only 29 for the clustered VRP solution. Nonetheless, clustering introduced some inefficiencies of its own. For example, one truck was assigned only a single 129 kg delivery. This motivated a direct comparison with a full multi-stop VRP approach, where clustering was omitted and ODL Studio assigned shipments more directly.

**Table 3: ODL Studio solution for the demand balanced-clustered solution for the first test set**

| Vehicle name | Capacity (kg) | Deliveries count | Delivered quantity (kg) | Kilometres travelled |
|---|---|---|---|---|
| 3t flatdeck | 2700 | 1 | 129 | 294 |
| 3t flatdeck | 2700 | 7 | 2551 | 536 |
| 3t flatdeck | 2700 | 4 | 1641 | 7 |
| 3t flatdeck | 2700 | 14 | 2627 | 32 |
| 16t flatdeck | 14400 | 12 | 12211 | 302 |
| 16t flatdeck | 14400 | 2 | 8434 | 215 |
| 16t flatdeck | 14400 | 8 | 4541 | 12 |
| 16t flatdeck | 14400 | 14 | 8549 | 20 |
| 16t flatdeck | 14400 | 8 | 9701 | 9 |
| 24t flatdeck | 21600 | 1 | 21579 | 152 |
| 34t tautliner | 30600 | 1 | 28224 | 243 |
| 36t tautliner | 32400 | 1 | 31940 | 233 |
| 36t tautliner | 32400 | 1 | 32256 | 243 |

## 5.3 Multi-stop VRP

Building on the clustered approach, the multi-stop VRP solution was tested using ODL Studio without any prior clustering. In this setup, all shipments were directly assigned to routes while considering vehicle capacity, cost, and distance simultaneously. This eliminates the potential inefficiencies caused by rigid cluster boundaries, while still leveraging the computational power of ODL Studio to solve the problem at scale.

The results for the same 91 shipment initial test set, for the multi-stop VRP model, are listed in Table 4. Compared to both the LP model and the demand-balanced clustered VRP, the direct multi-stop VRP approach produced superior outcomes. The required fleet size was reduced to just 27 trucks - outperforming the demand-balanced clustered VRP (29 trucks) and the LP model (36 trucks). Smaller vehicles carried most shipments, while the largest trucks were allocated only to high-volume single deliveries. This is consistent with operational expectations. The progression from clustered VRP to direct VRP highlights a key finding: while clustering can provide balanced and computationally simplified subproblems, direct optimisation captures the full problem structure more effectively, leading to further reductions in fleet size and travel distance. The next section verifies these findings against multiple weekly test datasets.

**Table 4: ODL Studio multi-stop VRP solution details for the first test set**

| Vehicle name | Capacity (kg) | Deliveries count | Delivered quantity (kg) | Kilometres travelled |
|---|---|---|---|---|
| 3t flatdeck | 2 700 | 8 | 2 603 | 714 |
| 3t flatdeck | 2 700 | 6 | 2 600 | 9 |
| 16t flatdeck | 14 400 | 9 | 11 887 | 300 |
| 16t flatdeck | 14 400 | 18 | 11 360 | 16 |
| 16t flatdeck | 14 400 | 24 | 13 099 | 39 |
| 16t flatdeck | 14 400 | 5 | 8 835 | 150 |
| 24t flatdeck | 21 600 | 1 | 21 579 | 152 |
| 34t tautliner | 30 600 | 1 | 28 224 | 243 |
| 36t tautliner | 32 400 | 1 | 32 256 | 152 |
| 36t tautliner | 32 400 | 1 | 32 256 | 233 |

## 5.4 Comparison between different solutions

Table 3 gives a comparison of the three solutions for the initial test set, showing the important metrics as well as the improvements gained by each solution.

**Table 5: Comparison between different solutions' results for the initial test set**

| Metric | Original | LP model | Clustered VRP | Multi-stop VRP |
|---|---|---|---|---|
| Number of trucks | 91 | 36 | 29 | 26 |
| Shipments consolidated | 0 | 55 | 70 | 70 |
| Average utilisation | ~50% | ~58% | ~65% | ~85% |
| Improvement in utilisation | N/A | 8% | 15% | 35% |

# 6 RESULTS

To verify that the various solution approaches consistently reduced costs, nine more random test datasets (each one week in size) were evaluated. These test sets were randomly chosen throughout the year and contain between 16 and 99 shipments during a particular week. A total of 525 shipments were investigated.

Three key metrics were used to validate the solution against the identified root causes:

1. Truck utilisation rates,
2. Frequency of single delivery/shipment trips, and
3. Total distance travelled.

Table 6 summarise the results of the ten test sets, including the initial and nine new test sets. The table demonstrates that the multi-stop consolidation approach achieved a 61.3% reduction in trucks used overall (from 525 to only 203), and a further 38.5% reduction in total distance travelled overall (from 55 982 km to only 34 396 km). All test sets furthermore showed both a reduction in the number of trucks and total distance travelled using the multi-stop consolidation VRP approach. These improvements translate directly to cost savings through reduced procurement needs and lower per-kilometre operational expenses.

The solution successfully transformed the delivery pattern from 87% single-delivery trips to a multi-stop approach where smaller vehicles averaged four to eight deliveries per route (depending on the number of shipments available to consolidate). The fundamental shift in operational approach directly addressed the second critical root cause of inefficient single-shipment dispatches.

**Table 6: Multi-stop consolidation solution details for ten test datasets in comparison to the base case**

| | BASE CASE/BENCHMARK CASE | | MULTI-STOP CONSOLIDATION SOLUTION | |
|---|---|---|---|---|
| **Date** | **No. of trucks used** | **Total distance travelled (km)** | **No. of trucks used** | **Total distance travelled(km)** |
| 2024/01/05 | 91 | 6 850 | 27 | 4 383 |
| 2024/02/05 | 41 | 5 329 | 21 | 3 892 |
| 2024/03/11 | 35 | 3 488 | 7 | 1 304 |
| 2024/04/08 | 44 | 6 523 | 25 | 5 120 |
| 2024/05/06 | 99 | 10 743 | 37 | 5 904 |
| 2024/06/10 | 16 | 1 583 | 6 | 955 |
| 2024/10/07 | 85 | 9 782 | 34 | 5 754 |
| 2024/11/11 | 62 | 8 627 | 36 | 6 206 |
| 2025/01/27 | 25 | 2 175 | 6 | 687 |
| 2025/02/24 | 27 | 882 | 4 | 191 |
| **Total** | **525** | **55 982** | **203** | **34 396** |

While the multi-stop consolidation VRP solution demonstrated substantial improvements, several limitations should be noted:

- The model assumed static demand patterns based on historical data;
- Time windows and service duration constraints were not incorporated due to data limitations;

- As more destination locations are added to the problem, the computational complexity and time required to calculate the solution increase; and
- Backhauling opportunities could not be evaluated as the TMS did not record empty return trips.

# 7 CONCLUSION AND FUTURE RESEARCH

This research successfully demonstrates that implementing multi-stop consolidation through VRP optimisation can achieve substantial cost reductions for 4PL providers operating in complex distribution environments. The solution reduced vehicle requirements by over 60%, while improving utilisation rates and reducing total travel distance by nearly 40%.

The key contribution of this work lies in providing a practical framework for 4PLs to enhance their current TMS capabilities without requiring complete system overhauls. By focusing on multi-stop consolidation rather than simple same destination grouping, the approach unlocks previously hidden efficiency gains that are particularly valuable in markets with diverse customer profiles.

In the case of this company, implementation of this approach in the East London region alone could generate significant annual savings. The framework is scalable to other regions and adaptable to different market conditions, making it valuable for the broader 3PL and 4PL industry facing similar challenges with mixed customer bases.

Several opportunities exist to extend this research, including the development of dynamic, real-time optimisation systems capable of responding to incoming orders rather than relying on weekly batch processing. The recording of empty return trips in the TMS could further allow for the study of incorporating backhaul consolidation opportunities that could lead to further cost savings. Integration with the existing TMS through an API function also represents a valuable path, enabling seamless adoption without replacing the core platform. Testing the approach in other South African regions where there are varying customer densities and infrastructure would provide insights into further scalability and adaptability.

# 8 REFERENCES


[1] C. Caplice and Y. Sheffi, "Optimization-based procurement for transportation services," *Journal of Business Logistics*, vol. 24, no. 2, pp. 109–128, Sep. 2003, doi: 10.1002/j.2158-1592.2003.tb00048.x.

[2] L. Saglietto, "Towards a classification of fourth party logistics (4PL)," *Universal Journal of Industrial and Business Management*, vol. 1, no. 3, Jan. 2013, doi: 10.13189/ujibm.2013.010305.

[3] S. E. Griffis and T. J. Goldsby, "Transportation management systems: an exploration of progress and future prospects," *Journal of Transportation Management*, vol. 18, no. 1, pp. 18–32, Apr. 2007, doi: 10.22237/jotm/1175385780.

[4] R. Mesa-Arango and S. V. Ukkusuri, "Benefits of in-vehicle consolidation in less than truckload freight transportation operations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 60, pp. 113–125, Dec. 2013, doi: 10.1016/j.tre.2013.05.007.

[5] M. A. Ülkü, "Dare to care: Shipment consolidation reduces not only costs, but also environmental damage," *International Journal of Production Economics*, vol. 139, no. 2, pp. 438 – 446, Oct 2012, doi: 10.1016/j.ijpe.2011.09.015.

[6] C. Yang, Y. Lee and C. Lee, "Data-Driven Order Consolidation with Vehicle Routing Optimization," *Sustainability*, vol. 17, no. 3, pp. 848, Jan 2025, doi: 10.3390/su17030848.

[7] R. W. Hall, "Consolidation strategy: inventory, vehicles and terminals," *Journal of Business Logistics*, vol. 8, no. 2, 1987, [Online]. Available: https://www.proquest.com/openview/9726cbee08678a8c7559da17eb5e458d/ [Accessed: 23 May 2026]..

[8] F. Mutlu, S. Çetinkaya, and J. H. Bookbinder, "An analytical model for computing the optimal time- and-quantity-based policy for consolidated shipments," *IIE Transactions*, vol. 42, no. 5, pp. 367–377, Feb 2010, doi: 10.1080/07408170903462368.

[9] A. Wu, "Cargo consolidation in intermodal container transport," M. A. Thesis, Delft University of Technology, Delft, The Netherlands, 2013. [Online]. Available: https://repository.tudelft.nl/file/File_d754b36d-5bb9-4c5f-b94a-747b9debb92f [Accessed 21 May 2026].

[10] S. Hernández, S. Peeta, and G. Kalafatas, "A less-than-truckload carrier collaboration planning problem under dynamic capacities," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 933–946, Novs 2011, doi: 10.1016/j.tre.2011.03.001.

[11] P. Berling and F. Eng-Larsson, "Pricing and timing of consolidated deliveries in the presence of an express alternative: Financial and environmental analysis," *European Journal of Operational Research*, vol. 250, no. 2, pp. 590 – 601, Apr 2016, doi: 10.1016/j.ejor.2015.09.041.

[12] 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, Sep. 2020, doi: 10.1007/s00291-020-00607-8.

[13] A. Stenger, D. Vigo, S. Enz, and M. Schwind, "An adaptive variable neighborhood search algorithm for a vehicle routing problem arising in small package shipping," *Transportation Science*, vol. 47, no. 1, pp. 64–80, Feb. 2013, doi: 10.1287/trsc.1110.0396.

[14] G. Singh Brar and G. Saini, "Milk run logistics: Literature review and directions," *in Proceedings of the World Congress on Engineering (WCE)*, vol. 1, pp. 6–8, Jul 2011, [Online]. Available: https://www.researchgate.net/profile/Mohamed-Mourad-Lafifi/post/Which-supply-chain-management-tools-are-essential-for-optimaziation/attachment/5a6f645f4cde266d5887339e/AS%3A588201156681733%401517249492554/download/Milk+Run+Logistic+_+Literature+Review+and+Directionx.pdf [Accessed 23 May 2023].

[15] G. D. Konstantakopoulos, S. P. Gayialis, and E. P. Kechagias, "Vehicle routing problem and related algorithms for logistics distribution: a literature review and classification," *Operational Research*, vol. 22, pp. 2033 – 2062, Sep 2020, doi: 10.1007/s12351-020-00600-7.

[16] K. Braekers, K. Ramaekers, and V. Nieuwenhuyse, "The vehicle routing problem: State of the art classification and review," *Computers & Industrial Engineering*, vol. 99, pp. 300–313, Sep 2016, doi: 10.1016/j.cie.2015.12.007.

[17] C. Prins, “Two memetic algorithms for heterogeneous fleet vehicle routing problems,” *Engineering Applications of Artificial Intelligence*, vol. 22, no. 6, pp. 916–928, Sep 2009, doi: 10.1016/j.engappai.2008.10.006.

[18] S. Lin, V. F. Yu, and S. Chou, “A note on the truck and trailer routing problem,” *Expert Systems with Applications*, vol. 37, no. 1, pp. 899–903, Jan 2010, doi: 10.1016/j.eswa.2009.06.077.

[19] T. Kodinariya and R. P. Makwana, “Review on determining number of clusters in K-Means clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90–95, Nov 2013, [Online]. Available: https://www.researchgate.net/profile/Trupti-Kodinariya/publication/313554124_Review_on_Determining_of_Cluster_in_K-means_Clustering/links/5789fda408ae59aa667931d2/Review-on-Determining-of-Cluster-in-K-means-Clustering.pdf [Accessed 23 May 2026].

[20] D. Deng, “DBSCAN Clustering Algorithm Based on Density,” *in 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), Hefei, China,* Sep 2020, pp. 949-953, doi: 10.1109/IFEEA51475.2020.00199.

[21] ODL Studio, “Open door logistics studio,” 2026 [Online]. Available: https://www.opendoorlogistics.com/ [Accessed: 23 May 2026].

[22] J. Bührmann and F. Bruwer, “K-Medoid petal-shaped clustering for the capacitated vehicle routing problem,” *The South African Journal of Industrial Engineering*, vol. 32, no. 3, pp. 33–41, Nov. 2021, doi: 10.7166/32-3-2610.