# Enhancing Vehicle Routing with Time Windows Solutions via K-means Clustering: A Comparative Study of Elbow and Truck Utilization Methods

**Kanokporn Boonjubut[1], Prat Boonsam[2], and Sirichai Yodwangjai[3*]**
[1,2]Department of Engineering Management, Faculty of Industrial Technology,
Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand
[3]Department of Industrial Engineering Technology, College of Industrial Technology,
King Mongkut's University North Bangkok, Bangkok, Thailand
E-mail: kanokporn.b@nrru.ac.th, prat.b@nrru.ac.th, sirichai.y@cit.kmutnb.ac.th*

Received: May 10, 2025 / Revised: September 29, 2025 / Accepted: October 2, 2025

***Abstract*—The vehicle routing problem with time windows is important in optimizing logistics distribution. For VRPTW optimization, a strategy is used to classify and optimize routes using artificial intelligence methods. Therefore, an improved two-phase algorithm is required to find a solution. Namely, a customer group can be divided into several regions using the K-means algorithm in the first phase, and each region can be decomposed into smaller subgroups according to certain constraints. In the second phase, local search from OR-tools solves the routing problem. In this experiment, two different methods of determining the number of clusters, namely, the elbow method and the truck utilization method, are compared by experimenting with a total of 26 standard instances. The results show that the truck utilization ratio outperforms the elbow method for the K-means algorithm in terms of overall results. The results from this experiment can be highly beneficial for routing, particularly when handling huge amounts of data that need to be subdivided ahead.**

***Index Terms—Clustering, K-means, Routing Problem, Time Windows***

## I. Introduction

Nowadays, people are using the internet to purchase items more and more because of these factors. The primary issue for the businesses in this situation is to offer more demanding clients a more effective distribution service. A number of large online merchants have begun searching for more creative and efficient ways to expedite last-mile and same-day delivery.

The rapid economic growth at both national and global levels has intensified business competition. To remain competitive, companies must enhance their logistical and transportation systems, as these are crucial components of efficient business operations. In supply chain management, one of the fundamental challenges is optimizing vehicle selection and routing to minimize shipping costs. This problem aligns with the Vehicle Routing Problem (VRP), which addresses transportation system issues such as route optimization, vehicle capacity management, transportation time windows, and accommodating diverse customer requirements [1].

The emergence of the Internet of Things (IoT) and big data analytics has significantly improved logistics operations. These technologies enable precise and effective monitoring of transportation processes through cloud computing, enhancing logistics efficiency and reducing operational costs [2]. However, many companies still rely on outdated logistics management systems that require significant manual effort. As businesses handle increasing order volumes, the need for a robust and efficient decision-support system for logistics distribution becomes critical.

A widely adopted approach in solving vehicle routing problems is the Cluster-First Route-Second (CFRS) strategy. This method consists of two main stages: first, decomposing shipping points into smaller clusters, followed by optimizing routes within each cluster [3]. Clustering plays a vital role in managing transportation route problems, as it helps break down complex logistics challenges into manageable subgroups.

Several clustering techniques have been applied in logistics, with K-means being one of the most prominent methods. K-means effectively clusters groups of delivery points based on proximity and shared transportation constraints, forming distinct clusters before applying route optimization algorithms [4]. Transportation route optimization can be approached by grouping delivery points into clusters based on factors such as distance or time windows, allowing each cluster to be assigned a specific vehicle for efficient deliveries [5].

The CFRS strategy is particularly useful in solving the Vehicle Routing Problem with Time Windows (VRPTW), where deliveries must adhere to strict time constraints. By first clustering delivery points and then optimizing routes within each cluster, computational time is reduced, and solution quality is improved, particularly in large-scale logistics operations [6]. Therefore, clustering should be effectively executed to maximize its impact on logistics efficiency and overall supply chain performance.

While the Cluster-First, Route-Second (CFRS) strategy has demonstrated effectiveness in addressing the Vehicle Routing Problem with Time Windows (VRPTW), a significant research gap persists in the optimal determination of the number of clusters. Existing literature predominantly employs the elbow method to determine the appropriate number of clusters in K-means clustering. However, this technique does not sufficiently account for operational constraints such as vehicle capacity utilization, which is critical for practical logistics applications.

The research question explored in this study is: How does the elbow method for cluster determination compare with optimization based on truck utilization ratio in terms of overall logistics efficiency and cost-effectiveness? This comparison is particularly significant because while the elbow method aims to minimize within-cluster variance as a statistical measure of clustering quality, it may not align with the practical objective of maximizing vehicle capacity utilization, which directly impacts operational expenses and resource efficiency. Current literature predominantly employs the elbow method for determining optimal cluster numbers in vehicle routing problems with time windows without a systematic evaluation of its operational effectiveness. This creates a research gap where statistically optimal clustering is assumed to translate into operationally efficient routing without empirical validation.

This study addresses this gap by making several key contributions to logistics optimization. First, this work develops the first systematic comparison framework between statistical clustering validation and operational optimization approaches for vehicle routing problems with time windows. Second, it provides empirical validation demonstrating the conditions under which each clustering approach delivers superior performance in terms of cost reduction and resource utilization. These contributions advance both theoretical understanding of clustering effectiveness in transportation systems and provide practical guidance for logistics optimization in real-world applications.

## II. Literature Review

The Vehicle Routing Problem (VRP) is a critical issue frequently addressed by managers in logistics and transportation. A VRP involves the distribution of goods from an origin as a single depot or multiple depots to multiple destinations [7]. Solving VRPs provides significant benefits to businesses, including optimized routes and travel times, reduced vehicle requirements, and improved vehicle utilization. These efficiencies lead to lower overall logistics costs and increased profitability for organizations [7].

An extension of the VRP is the Vehicle Routing Problem with Time Windows (VRPTW), which not only focuses on minimizing route costs but also incorporates time constraints to ensure that customers are serviced within predefined time windows, bounded by the earliest and latest permissible times [8], [9].

In the VRPTW framework, the optimization model must account for specific time-based service requirements designated by each customer within the distribution network. These temporal constraints are formally defined through a dual-parameter system: ($e_i$) represents the earliest permissible service initiation time, while ($l_i$) denotes the latest allowable service completion time for each customer. This temporal framework introduces an additional dimension of complexity to the traditional spatial routing optimization problem.

A notable characteristic of VRPTW implementations is the concept of soft time windows, which introduces operational flexibility while maintaining service quality standards. Under this constraint structure, delivery vehicles are permitted to arrive at customer locations before the designated earliest start time ($s_i$). However, in such instances, service commencement must be delayed until the customer's specified time window officially begins. This waiting time, while not subject to direct penalties, must be incorporated into the overall route optimization calculations as it impacts operational efficiency and resource utilization. Fig. 1 shows an example of a VRPTW problem where each customer has their specific service time window.

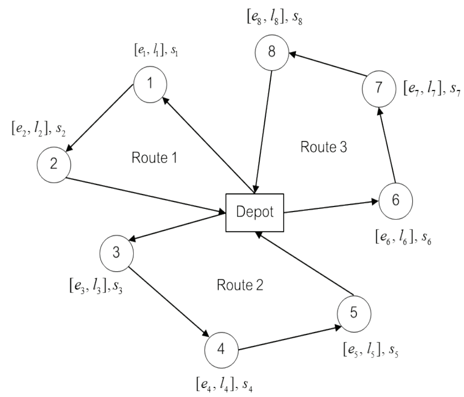


Fig. 1. Eight customers and one depot network

This study employs a "one-to-all" routing model, where each vehicle departs from and returns to a designated depot. The warehouse serves as the storage site and the departure point for vehicles. Vehicle routes are visually depicted with directional arrows, and polygons in various colors represent customer groups assigned to distinct routes. Products are loaded onto vehicles at the depot and delivered to their respective customers.

The model also considers vehicle storage constraints, ensuring that each vehicle's capacity accommodates the weight and quantity of transported goods. Two key assumptions underpin this model: (1) the total number of available vehicles is sufficient to meet the daily demand of all customers, and (2) each customer's daily order does not exceed the capacity of a single vehicle. Under these conditions, each vehicle is assigned a route to serve multiple customers and deliver various goods. While a single vehicle can cater to multiple customers along its route, only one vehicle is allocated to serve a particular customer at any given time.

The Vehicle Routing Problem with Time Windows (VRPTW) represents a complex combinatorial optimization problem that seeks to determine optimal routes for a vehicle fleet serving customers within specified time windows while satisfying capacity constraints. A prominent solution methodology for VRPTW is the two-phase Cluster-First Route-Second (CFRS) approach, which decomposes the problem into computationally tractable subproblems. The clustering phase partitions customers into geographically coherent groups using algorithms such as K-means, ensuring that vehicle capacity and temporal constraints are maintained within each cluster through appropriate parameter adjustments. Subsequently, the routing phase applies heuristic algorithm improvements to optimize vehicle routes within individual clusters while minimizing travel distance and satisfying time window constraints. This decomposition strategy reduces computational complexity from exponential to polynomial time for large-scale instances, enabling near-optimal solutions for practical applications. The CFRS methodology has demonstrated effectiveness in various logistics domains, including perishable goods distribution and urban delivery systems, where it successfully balances solution quality with computational efficiency. The approach's scalability and adaptability make it particularly suitable for large-scale VRPTW instances where exact methods become computationally prohibitive, thereby providing a practical framework for real-world logistics optimization applications.

While various clustering methodologies exist for addressing VRPTW applications, including hierarchical clustering, heuristic-based clustering approaches, and spectral clustering, K-means clustering has emerged as the most widely adopted approach in various domains due to its computational efficiency and scalability. Although K-means has not yet gained widespread adoption specifically in logistics optimization, its proven effectiveness in partitioning data points into coherent groups and its ability to handle large datasets efficiently make it a promising candidate for vehicle routing applications. Consequently, this research employs K-means clustering as the primary clustering methodology to investigate the comparative effectiveness of different cluster determination approaches in vehicle routing optimization.

### *A. K-means Clustering Algorithm*

One well-known clustering technique is K-means clustering, which divides $n$ observations into $k$ clusters, where $k$ is prioritized. The client's cluster is determined by the K-means clustering; a second clustering is then applied to each of the K-means clusters. As far as possible, the method aims to segregate each compact class and minimize the goal function. According to Khan and Ahmad [10], K-means clustering can be explained as follows.

First, the K-means algorithm selects $k$ objects at random, each of which represents a mass of grouping relationships. Based on each grouping's mass to object distance, each object is then assigned to the most similar clustering. Next, figure out each clustering's new mass. Continue doing the aforementioned till the guideline function is put together. In general, the square error guideline function [11] is the adopted guideline function.

Assume that $X = \{x_1, x_2 \ldots, x_n\}$ is a set of observations, with $d$ being the dimension of each real vector. $c_k$ is the formula for a set of $K$ centers: $K = \{1,2,...,K\}$. With sj = {d | d is a member of the cluster $k$}, the set of samples that belong to the $k^{th}$ the cluster is displayed. Euclidean distance is the distance between a point and the cluster center $c_k$, as given in (1).

$$\sum_{i=1}^{n} dist(d_i, c_k) \tag{1}$$

Step 1: Random sampling was used to generate a set of $c_k$.

Step 2: Decide each cluster's members according to the requirements for the minimum distance from the cluster center.

Step 3: Equation (2) will be used to calculate $c_k$. $|S_k|$ Represents the number of data elements in the $k^{th}$ cluster.

$$c_k = \frac{\sum_{d_i \in S_k}^{n} d_i}{|S_k|} \tag{2}$$

Step 4: Until the objective is optimal, steps 2 and 3 could be repeated. When evaluating the K-means clustering procedure, the Sum of Squares Error (SSE) criterion is most frequently utilized. The greatest results are obtained from the clustering result with the lowest SSE value. Equation (3) computes the sum of the squares of the object distances from the cluster center points [12].

$$SSE = \sum_{i=1}^{K} \sum_{x \in c_i} dist^2(m_i, x) \tag{3}$$

*B. Determine the Optimal Number of K-Means Clusters*

From the relevant research literature, multiple methodologies for determining the optimal number of clusters (k) in vehicle routing applications have been identified, each with distinct theoretical foundations and practical implications. Among the prominent approaches, Comert *et al.* [13] proposed a mathematical formulation for calculating the number of clusters based on operational constraints, specifically incorporating truck capacity and customer demand parameters. Their methodology determines cluster numbers by optimizing the truck utilization ratio, thereby directly linking clustering decisions to vehicle capacity management and operational efficiency. This approach represents a paradigm shift from traditional statistical clustering validation methods by prioritizing operational feasibility over purely geometric or statistical cluster quality measures. Additionally, the elbow method constitutes another widely recognized technique for cluster number determination, which employs a different theoretical framework based on the analysis of the within-cluster sum of squares, with details presented as follows:

1) Therefore, to solve a capacitated VRP problem using K-means clustering, Comert et al. [14] provided an equation for calculating the number of clusters. Employing (4) as the calculation formula, the number of clusters is determined by truck capacity and demand, or the truck utilization ratio.

$$\text{number of clusters} = \frac{\text{total demand}}{\text{truck capacity}} \tag{4}$$

2) The elbow approach uses the K-means clustering algorithm to plot the explained variations against the number of clusters. The elbow curve is then used to determine the number of clusters. The quality of aggregation within a cluster and the separation between clusters are represented by summing the squared errors of all data points within the cluster. Thorndike came up with the elbow method [14], which uses the variation graph (data dispersion) as a function of the number of clusters to compute the ideal number of clusters to utilize. The elbow curve is the result of this method. Bertagnolli [15] states that the Elbow approach can be expressed mathematically (5), (6) as follows:

$$W_k = \sum_{r=1}^{k} \frac{1}{n_r} D_r \tag{5}$$

$$D_r = \sum_{i=1}^{n_r-1} \sum_{j=1}^{n_r} \left\| d_i - d_j \right\|_2 \tag{6}$$

Where $W_k$ is the average internal sum of squares, and $D_r$ are the sum of the distances between each point in a cluster, $k$ is the number of the cluster, and $n_r$ is the number of points in the cluster.

For different values of $k$ (number of clusters), calculate the Sum of Squared Distances (SSE) between each data point and its nearest centroid. Plot the number of clusters ($k$) versus the SSE values. Then, locate the point on the graph where the SSE decline begins to level out and take the shape of an elbow. This is the ideal number of clusters.

## III. METHODOLOGY

This research presents an experimental study focused on VRPTW applications, specifically investigating the comparative effectiveness of two fundamentally different approaches for determining the optimal number of clusters (k) in K-means clustering algorithms. This work conducts experimental investigations on VRPTW instances to evaluate two distinct methodologies for determining the optimal number of clusters (k) in K-means clustering: The truck utilization ratio approach and the elbow method. The objective of this study is to assess the comparative performance and effectiveness of these two clustering determination techniques in optimizing logistics operations and overall system efficiency within vehicle routing optimization scenarios.

The CFRS approach represents a systematic two-phase methodology for solving VRPTW by decomposing the complex optimization problem into computationally manageable subproblems. This methodology leverages K-means clustering in the first phase to partition customers into geographically coherent groups, followed by route optimization within each cluster in the second phase, as shown in Fig. 2.

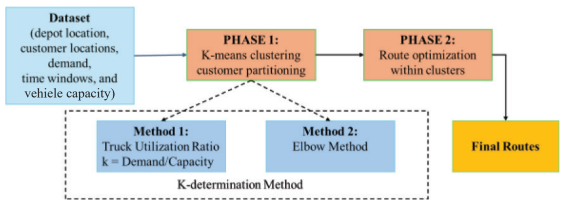


Fig. 2. CFRS methodology with different K-determination methods

In this study, standard VRPTW instances were employed, selected randomly to ensure unbiased representation. The analysis focused on comparing different subgroup initialization strategies for the K-means clustering method in the first phase of the solution process. In the subsequent routing phase, vehicle capacity and time window constraints were imposed to reflect realistic operational conditions.

The performance of the designed algorithm is compared to models from the existing literature in this section through computational studies. Using Solomon's instances [16], six different types of problems that impact the behavior of algorithms for VRPTW are identified. They are divided into three categories: Random (R), Clustered (C), and Random-Clustered (RC). Of these, there are two types: Type 1, which has a short scheduling horizon (R1, C1, RC1), and type 2, which has a long scheduling horizon (R2, C2, RC2). The variables, which are drawn from Solomon [17], include the number of customers, geographic location, demand level, time windows (ready and due times), service time, and vehicle capacity. These benchmarks enable thorough testing of VRPTW algorithms in various scenarios.

VRPTW algorithms are still evaluated using the Solomon benchmark problems. These instances are divided into six groups according to operational characteristics, time window restrictions, and customer spatial distribution, with each problem's details presented as shown in Figs. 3 to Fig. 8.

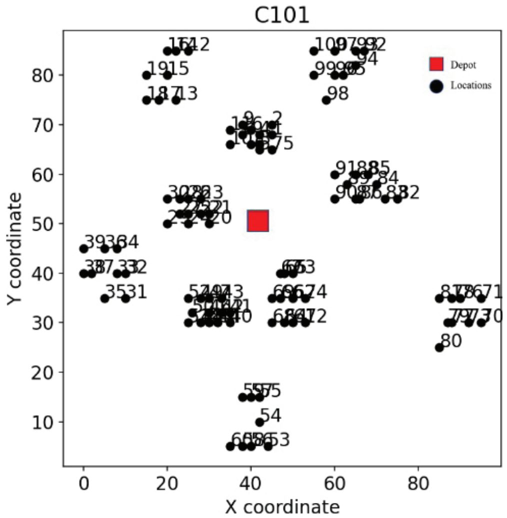


Fig. 3. Distribution of instance C101

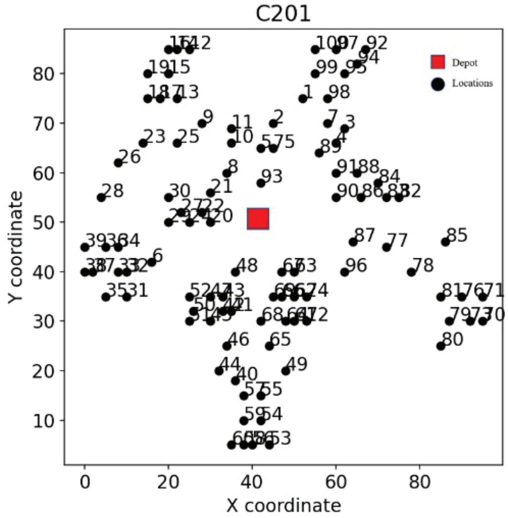


Fig. 4. Distribution of instance C201

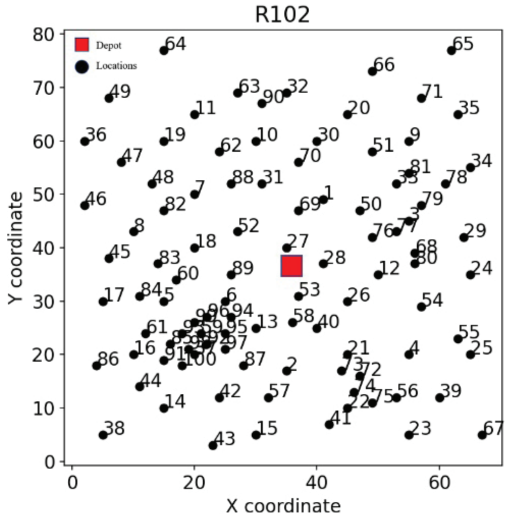


Fig. 5. Distribution of instance R102

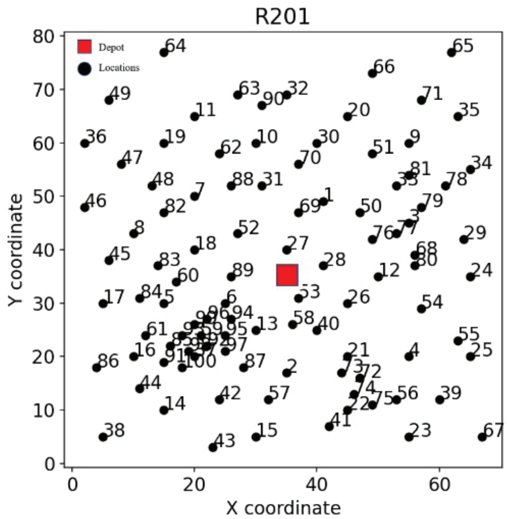


Fig. 6. Distribution of instance R201.

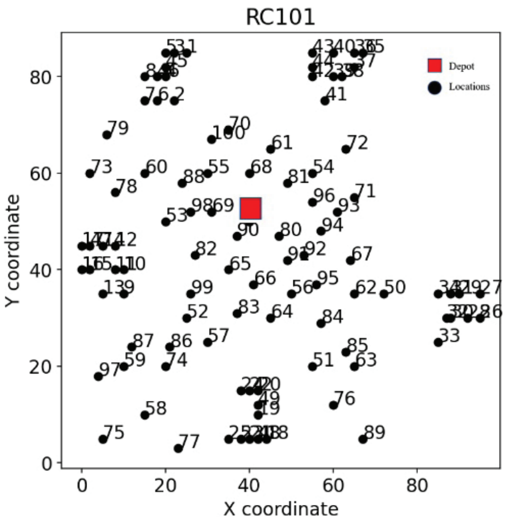

Fig. 7. Distribution of instance RC101

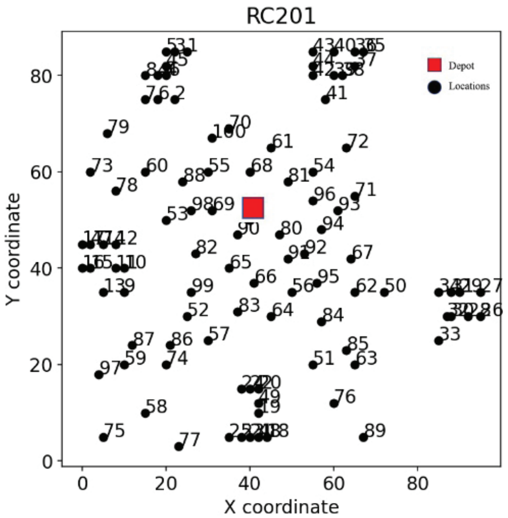

Fig. 8. Distribution of instance RC201

The strategy in this approach involves utilizing the K-means algorithm. Determining the appropriate number of clusters for a given dataset is a crucial step in this algorithm. The elbow method identifies the optimal number of clusters by examining the curve representing the within-cluster sum of squares plotted against the number of clusters. The elbow point on this curve indicates where adding additional clusters yields diminishing returns in variance reduction. Thus, the optimal number of clusters identified through the elbow method is then utilized in the K-means algorithm to perform clustering.

In this research, the experiment is conducted on the Solomon benchmark problem, which involves 26 problems: 5 problems of C1, 5 problems of C2, 5 problems of R1, 5 problems of R2, 3 problems of RC1, and 3 problems of RC2. The experiment was designed to compare the efficiency of routing with different K-means clustering methods, with each cluster being routed using OR-Tools' local search method. Employing local search to address vehicle routing problems is a pragmatic decision owing to its capacity to effectively traverse complex, large-scale solution spaces efficiently. Local search heuristics iteratively refine a candidate solution by exploring its nearby "neighborhood" of solutions, making small, incremental adjustments that often yield quick improvements in route cost or feasibility. This method is particularly valuable for VRP, which is NP-hard and difficult to solve optimally within a reasonable time, especially as the number of customers grows.

## IV. Results

Based on Solomon's problem classification, the six problem types, R1, R2, C1, C2, RC1, and RC2, are distinguished by customer distribution and vehicle constraints. For R1 problems, which involve narrow time windows and low vehicle capacity, the elbow method produces a shorter total distance than the clustering approach proposed by Comert *et al.* [13], as measured by the truck utilization ratio. However, for R2 problems, which feature wider time windows and higher vehicle capacity, the clustering method by Comert *et al.* [13] achieves better performance in terms of total distance.

According to the findings of 26 experiments, the clustering by using the truck utilization ratio outperforms the elbow approaches for all 14 of the 26 instances, as shown in Table I and Fig. 9. The following are examples of the experimental findings from applying the CFRS approach in routing utilizing K-means with various techniques for determining the number of groups (*k*), namely the elbow method and truck utilization ratio: C101, and C203.

TABLE I
FESIBLE SOLUTION

| Instance | Elbow Method | | | Truck Utilization Ratio | | |
|---|---|---|---|---|---|---|
| | Distance | Car | *k* | Distance | Car | *k* |
| C101 | **1880.82** | 20 | 5 | 1976.79 | 66 | 10 |
| C102 | 1441.24 | 16 | 5 | **1422.62** | 41 | 10 |
| C105 | **1277.93** | 15 | 5 | 1291.30 | 40 | 10 |
| C108 | 892.50 | 11 | 5 | **879.05** | 21 | 10 |
| C109 | **723.78** | 11 | 5 | 742.66 | 16 | 10 |
| C201 | **629.94** | 8 | 5 | 648.81 | 4 | 3 |
| C203 | 634.33 | 8 | 5 | **568.05** | 4 | 3 |
| C205 | 623.53 | 7 | 5 | **567.24** | 4 | 3 |
| C206 | 1449.01 | 11 | 5 | **1350.08** | 5 | 3 |
| C207 | 840.62 | 8 | 5 | **567.84** | 4 | 3 |
| R102 | **1039.85** | 12 | 4 | 1103.9 | 15 | 8 |
| R103 | **905.37** | 11 | 4 | 930.43 | 12 | 8 |
| R105 | 1031.43 | 12 | 4 | **972.16** | 15 | 8 |
| R109 | **838.95** | 10 | 4 | 947.31 | 12 | 8 |
| R112 | 756.22 | 9 | 4 | **750.25** | 9 | 8 |

TABLE I
FESIBLE SOLUTION (CON.)

| Instance | Elbow Method | | | Truck Utilization Ratio | | |
|---|---|---|---|---|---|---|
| | **Distance** | **Car** | *k* | **Distance** | **Car** | *k* |
| R201 | 1438.99 | 9 | 4 | **1373.70** | 6 | 2 |
| R202 | 1302.69 | 8 | 4 | **1203.58** | 6 | 2 |
| R203 | 1069.22 | 9 | 4 | **1054.87** | 4 | 2 |
| R208 | 740.19 | 6 | 4 | **710.58** | 3 | 2 |
| R210 | **979.25** | 7 | 4 | 985.45 | 5 | 2 |
| RC101 | **1126.71** | 12 | 4 | 1176.68 | 17 | 9 |
| RC103 | **981.57** | 10 | 4 | 1003.06 | 13 | 9 |
| RC104 | **857.34** | 10 | 4 | 916.37 | 12 | 9 |
| RC201 | 1584.65 | 10 | 4 | **1507.91** | 8 | 2 |
| RC206 | **1149.22** | 8 | 4 | 1175.81 | 4 | 2 |
| RC208 | 687.92 | 4 | 4 | **671.76** | 4 | 2 |

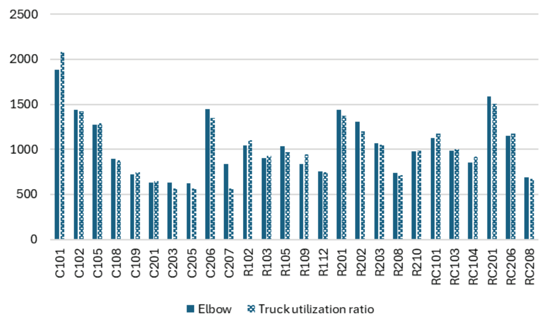


Fig. 9. Comparison results of methods for determining the number of groups

TABLE II
PAIRED T-TEST RESULTS

| Dependent Variable | Mean | | t-statistic | p-value |
|---|---|---|---|---|
| | **Group 1 (Elbow Method)** | **Group 2 (Truck Utilization Ratio)** | | |
| Distance | 1033.97 | 1019.16 | .991 | .331 |
| Car | 10.08 | 13.46 | -1.47 | .152 |

The results from the t-test indicate that there is no statistically significant difference between the two methods for either variable. The paired sample t-test showed no significant difference between the Elbow Method and the Truck Utilization Ratio for both distance (Elbow Method: Mean = 1033.97, Truck Utilization Ratio: Mean = 1019.16; t = 0.991, p = 0.331) and the number of trucks (Elbow Method: Mean = 10.08, Truck Utilization Ratio: Mean = 13.46; t = -1.47, p = 0.152). These findings indicate that both methods yield statistically comparable outcomes, suggesting that either approach can be applied without compromising routing efficiency or fleet utilization, as shown in Table II.

Establish the answer by the elbow method. For the example instance of C101 and C203. Start by dividing the data to determine the number of groups, which shows the results as shown in Figs. 10 and 11. After that, considering the characteristics of the points that form the elbow angle, which indicates that there should be how many groups (*k*) from the result, a total of five groups are divided based on the characteristics of the data in this problem. Next, divide the groups, which shows the results as shown in Figs. 12 and 13. Finally, work on finding the route sequence for each group using the OR tool by the local search method, which displays the results as shown in Tables III and IV.

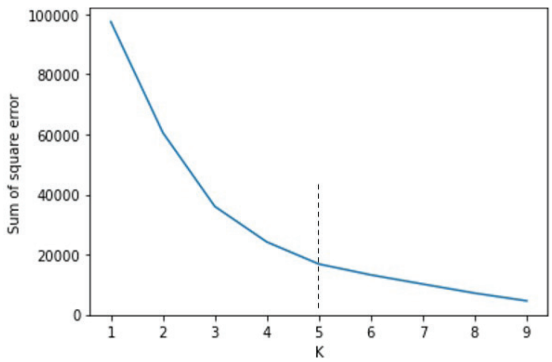


Fig.10. The result of determining *k*, for instance, C101, by the elbow method

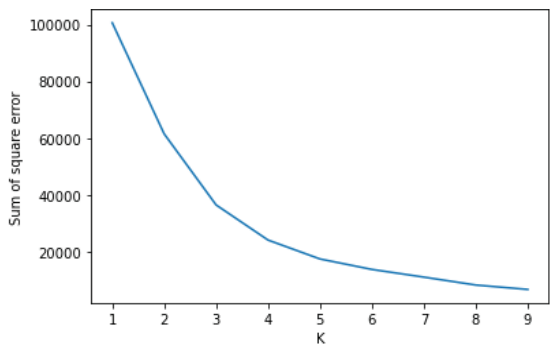


Fig. 11. The result of determining *k*, for instance, C203, by the elbow method

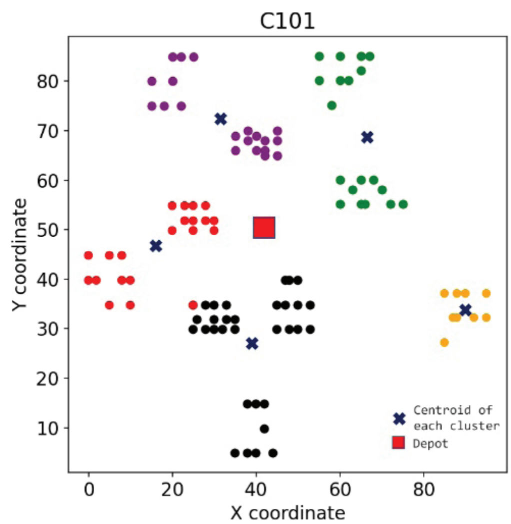


Fig. 12. Clustering by the elbow method result for instance C101

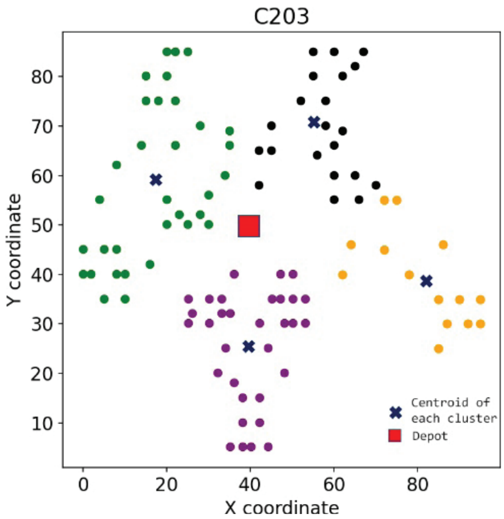

Fig. 13. Clustering by the elbow method result for instance C203

TABLE III
FEASIBLE SOLUTION OF C101 BY THE ELBOW METHOD

| Number of Clusters | Route | Sequencing Order | Demand | Distance |
|---|---|---|---|---|
| $k = 1$ | 1 | 0-90-98-96-87-86-95-83-94-92-82-0 | 160 | 187.41 |
| | 2 | 0-84-93-85-97-100-88-89-99-91-0 | 200 | 188.73 |
| $k = 2$ | 3 | 0-61-64-59-68-51-50-66-0 | 70 | 102.39 |
| | 4 | 0-69-0 | 10 | 15.81 |
| | 5 | 0-67-43-42-57-55-65-0 | 100 | 76.17 |
| | 6 | 0-49-47-0 | 20 | 21.21 |
| | 7 | 0-41-40-62-44-56-72-46-58-60-45-48-0 | 190 | 191.35 |
| | 8 | 0-63-54-53-74-0 | 160 | 81.90 |
| $k = 3$ | 9 | 0-20-24-32-33-25-31-27-35-29-37-0 | 200 | 166.47 |
| | 10 | 0-30-38-28-39-26-36-23-34-22-52-21-0 | 180 | 251.06 |
| $k = 4$ | 11 | 0-70-73-0 | 40 | 61.52 |
| | 12 | 0-76-71-0 | 30 | 57.20 |
| | 13 | 0-80-0 | 10 | 51.48 |
| | 14 | 0-77-79-0 | 20 | 53.00 |
| | 15 | 0-81-78-0 | 50 | 50.43 |

| Number of Clusters | Route | Sequencing Order | Demand | Distance |
|---|---|---|---|---|
| $k = 5$ | 16 | 0-18-19-8-0 | 50 | 66.30 |
| | 17 | 0-5-3-13-17-7-0 | 90 | 65.83 |
| | 18 | 0-2-0 | 30 | 20.62 |
| | 19 | 0-10-15-11-16-9-14-12-6-4-0 | 170 | 150.34 |
| | 20 | 0-1-75-0 | 30 | 21.60 |

TABLE IV
FESIBLE SOLUTION OF C203 BY THE ELBOW METHOD

| Number of clusters | Route | Sequencing order | Demand | Distance |
|---|---|---|---|---|
| $k = 1$ | 1 | 0-50-51-52-0 | 30 | 30.04 |
| | 2 | 0-21-0 | 20 | 11.66 |
| | 3 | 0-20-22-24-27-30-29-6-32-33-31-35-37-38-39-36-34-28-0 | 310 | 87.76 |
| $k = 2$ | 4 | 0-83-82-85-76-71-70-73-80-79-81-78-77-87-96-0 | 240 | 124.47 |
| $k = 3$ | 5 | 0-26-23-18-19-16-14-12-15-17-13-25-9-11-10-8-0 | 300 | 116.34 |
| $k = 4$ | 6 | 0-90-0 | 10 | 20.62 |
| | 7 | 0-93-5-75-2-1-99-100-97-92-94-95-98-7-3-4-89-91-88-84-86-0 | 370 | 106.55 |
| $k = 5$ | 8 | 0-67-63-62-74-72-61-64-66-69-68-65-49-55-54-53-56-58-60-59-57-40-44-46-45-47-43-42-41-48-0 | 530 | 136.89 |

When applying the by-truck utilization ratio technique to solve sample instances C101 and C103, the number of groups ( $k$ ) is determined by dividing the problem's overall demand by the total number of vehicles available. Fig. 14, which is the solution to instance C101, then shows how members of each group are found using clustering. Instance C203 has an answer in Fig.15. Then, using OR Tools and the local search method, the routes are determined. Tables IV and V, respectively, display the results.

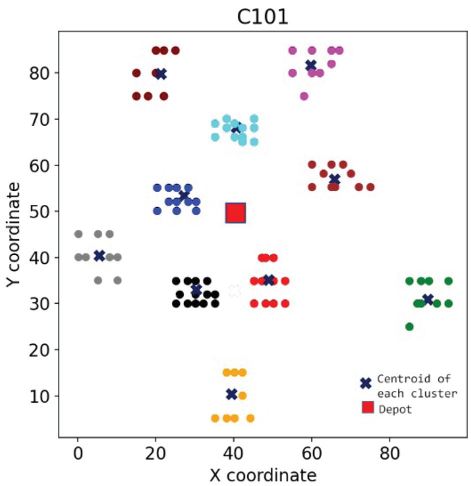


Fig. 14. Clustering by truck utilization ratio result for instance C101

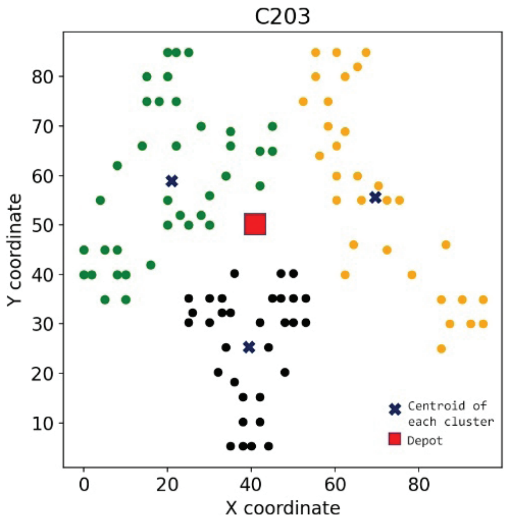


Fig. 15. Clustering by truck utilization ratio result for instance C203

TABLE V
FESIBLE SOLUTION OF C101 BY TRUCK UTILIZATION RATIO

| Number of Clusters | Route | Sequencing Order | Demand | Distance |
|---|---|---|---|---|
| $k = 1$ | 1 | 0-42-43-0 | 30 | 22.31 |
| | 2 | 0-49-47-0 | 20 | 21.21 |
| | 3 | 0-50-52-0 | 20 | 25.96 |
| | 4 | 0-48-51-0 | 20 | 26.32 |
| | 5 | 0-46-45-0 | 40 | 22.59 |
| | 6 | 0-41-40-0 | 20 | 20.68 |
| | 7 | 0-44-0 | 10 | 21.54 |
| $k = 2$ | 8 | 0-70-73-0 | 40 | 61.52 |
| | 9 | 0-76-71-0 | 30 | 57.20 |
| | 10 | 0-80-0 | 10 | 51.48 |
| | 11 | 0-77-79-0 | 20 | 53.00 |
| | 12 | 0-81-78-0 | 50 | 50.43 |
| $k = 3$ | 13 | 0-30-0 | 10 | 20.62 |
| | 14 | 0-29-0 | 10 | 20.00 |
| | 15 | 0-27-0 | 10 | 17.12 |
| | 16 | 0-28-26-0 | 30 | 19.72 |
| | 17 | 0-24-25-0 | 50 | 17.00 |
| | 18 | 0-23-0 | 10 | 13.00 |
| | 19 | 0-22-21-0 | 40 | 14.17 |
| | 20 | 0-20-0 | 10 | 10.00 |
| $k = 4$ | 21 | 0-83-82-0 | 30 | 35.39 |
| | 22 | 0-84-85-0 | 50 | 33.88 |
| | 23 | 0-87-86-0 | 30 | 26.50 |
| | 26 | 0-90-0 | 10 | 20.62 |
| $k = 5$ | 27 | 0-58-60-0 | 50 | 48.04 |
| | 28 | 0-56-0 | 30 | 45.00 |
| | 29 | 0-54-53-0 | 60 | 45.44 |
| | 30 | 0-59-0 | 10 | 35.06 |
| | 31 | 0-57-55-0 | 50 | 37.00 |
| $k = 6$ | 32 | 0-93-0 | 40 | 43.01 |
| | 33 | 0-94-92-0 | 30 | 44.22 |
| | 34 | 0-97-100-0 | 50 | 45.31 |
| | 35 | 0-95-0 | 30 | 37.20 |
| | 36 | 0-99-0 | 10 | 33.54 |
| | 37 | 0-98-96-0 | 30 | 36.20 |
| $k = 7$ | 38 | 0-38-0 | 30 | 41.23 |
| | 39 | 0-39-0 | 20 | 40.31 |
| | 40 | 0-37-0 | 20 | 39.29 |
| | 41 | 0-35-0 | 10 | 38.08 |
| | 42 | 0-36-0 | 10 | 35.36 |
| | 43 | 0-31-0 | 20 | 33.54 |
| | 44 | 0-34-0 | 20 | 32.39 |
| | 45 | 0-32-33-0 | 70 | 33.62 |
| $k = 8$ | 46 | 0-72-61-0 | 20 | 26.85 |
| | 47 | 0-64-0 | 10 | 21.54 |
| | 48 | 0-68-0 | 10 | 20.62 |
| | 49 | 0-74-0 | 50 | 19.85 |
| | 50 | 0-62-0 | 20 | 18.03 |
| | 51 | 0-66-0 | 10 | 16.55 |
| | 52 | 0-69-0 | 10 | 15.81 |
| | 53 | 0-63-0 | 50 | 14.14 |
| | 54 | 0-67-65-0 | 20 | 13.21 |
| $k = 9$ | 55 | 0-16-0 | 40 | 40.31 |
| | 56 | 0-14-12-0 | 30 | 42.36 |
| | 57 | 0-15-0 | 40 | 36.06 |
| | 58 | 0-18-19-0 | 30 | 40.36 |
| | 59 | 0-13-17-0 | 50 | 34.81 |

TABLE V
FESIBLE SOLUTION OF C101 BY TRUCK UTILIZATION RATIO (CON.)

| Number of Clusters | Route | Sequencing Order | Demand | Distance |
|---|---|---|---|---|
| $k = 10$ | 60 | 0-2-0 | 30 | 20.62 |
| | 61 | 0-11-9-0 | 20 | 22.81 |
| | 62 | 0-6-4-0 | 30 | 21.24 |
| | 63 | 0-10-0 | 10 | 16.76 |
| | 64 | 0-7-8-0 | 40 | 18.83 |
| | 65 | 0-1-75-0 | 30 | 21.68 |
| | 66 | 0-5-3-0 | 20 | 16.13 |

TABLE VI
FESIBLE SOLUTION OF C203 BY TRUCK UTILIZATION RATIO

| Number of Clusters | Route | Sequencing Order | Demand | Distance |
|---|---|---|---|---|
| $k = 1$ | 1 | 0-93-75-2-5-0 | 100 | 26.70 |
| | 2 | 0-20-22-24-27-30-29-6-32-33-31-35-37-38-39-36-34-28-26-23-18-19-16-14-12-15-17-13-25-9-11-10-8-21-0 | 630 | 183.64 |
| $k = 2$ | 3 | 0-67-63-62-74-72-61-64-66-69-68-65-49-55-54-53-56-58-60-59-57-40-44-46-45-51-50-52-47-42-41-43-48-0 | 560 | 146.91 |
| $k = 3$ | 4 | 0-1-99-100-97-92-94-95-98-7-3-4-89-91-88-86-84-83-82-85-76-71-70-73-80-79-81-78-77-96-87-90-0 | 520 | 210.80 |

The C1 and C2 categories include customers clustered in groups, where C1 has narrow time windows and low vehicle capacity. Results from C1-type problems are arranged using the elbow method at a smaller distance. Regarding C2, wide time windows and large vehicle capacity, along with the vehicle routing by truck utilization ratio [1], provide better results.

The RC1 and RC2 categories blend randomly distributed and clustered customers, with RC1 having narrow time windows and low vehicle capacity. Although RC1 has low vehicle capacity and short windows, it performs better when implemented with the elbow method. RC2 has broad time frames and a high vehicle capacity. The truck utilization ratio [13] was used to improve the results.

The sensitivity analysis was carried out by comparing the distance outputs obtained from the Elbow Method and the Truck Utilization Ratio across multiple vehicle routing instances. For each instance, the absolute difference was calculated as the absolute deviation between the two methods, while the percentage difference was derived by normalizing the deviation against the Elbow Method baseline. These measures allowed for both direct and relative comparisons of the two approaches. Subsequently, summary statistics, including mean deviations and the identification of maximum and minimum discrepancies, were computed to assess overall trends. To further illustrate the findings, bar charts and scatter plots were employed, with the percentage differences specifically presented in Fig. 16, providing visual insights into the degree of sensitivity across all instances.

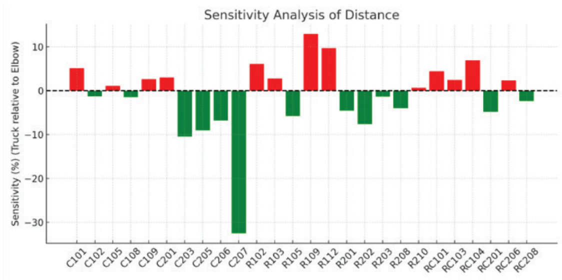


Fig. 16. Percentage differences between the Elbow Method and the Truck Utilization Ratio across all instances

The experimental results demonstrate comparative performance between the Elbow Method and the Truck Utilization Ratio across 26 benchmark instances. In terms of routing distance, the Elbow Method produced shorter routes in approximately 60% of the cases, whereas the Truck Utilization Ratio outperformed in the remaining 40% of the instances. The average sensitivity, defined as the relative percentage change in total distance, was +1.2%, suggesting a slight overall advantage for the Elbow Method. However, the variability was substantial, ranging from –33% (favoring Truck Utilization Ratio) to +13% (favoring Elbow Method).

The graphical comparison (Fig. 9) shows that while most instances present only minor differences, certain cases reveal significant improvements when applying the Truck Utilization Ratio. Fig. 16 highlights these differences by illustrating sensitivity values for each instance, clearly indicating which method yields shorter routes. These findings highlight that the reliability of either method is not uniform across all instances. The sensitivity appears to be influenced by spatial and demand heterogeneity within the dataset. Consequently, practitioners should interpret results with caution, particularly when applying these methods to diverse or irregular problem instances.

## V. Conclusion

The location of the logistics center is a significant strategic consideration for logistics system optimization. The purpose of this work was to implement a method for data analysis that optimizes the vehicle routing process by applying ensemble approaches and a two-phase algorithm. Two different approaches were set up to accomplish this goal, each of which used a clustering model to organize delivery points (customers) into clusters. Effective customer grouping was made possible using the elbow method and truck utilization ratio for clustering, which supplied essential information for the development of routing techniques. The adaptability of the suggested methodology was shown by the application of two different methods, each focused on a particular clustering model.

In summary, while both approaches can be applied effectively to vehicle routing problems, the Elbow Method provides a more stable baseline, whereas the Truck Utilization Ratio represents a practical alternative that should be validated against specific dataset characteristics. Further investigation into the underlying factors contributing to high-sensitivity cases is recommended to enhance methodological reliability and improve decision-making accuracy in logistics planning.

According to the results of the experiments, the K-means and truck utilization ratio are used in the strategies. The truck utilization ratio outperforms the elbow method for the K-means algorithm in terms of overall results, and the strategy performs satisfactorily in the total distance driven by trucks while maintaining a balanced distribution of the distance traveled. This study addresses existing issues and suggests future research areas by offering useful information on fleet routing through heuristic approaches in routing and data analysis in clustering. This effort will make a significant impact on the area of logistics and transportation operations.

## References


[1] P. Toth and D. Vigo, *Vehicle Routing: Problems, Methods, and Applications*, 2nd ed, Philadelphia, PA: SIAM, 2014, pp. 1-8.

[2] M. Haonan, Y. C. He, M. Huang, Y. Wen, Y. Cheng, and Y. Jin, "Application of K-means clustering algorithms in optimizing logistics distribution routes," in *Proc. 6th Int. Conf. Systems and Informatics (ICSAI)*, 2019, pp. 1466-1470.

[3] R. G. Dondo and J. Cerda, "A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows," *Eur. J. Oper. Res.*, vol. 176, no. 3, pp. 1478-1507, Feb. 2007, https://doi.org/10.1016/J.EJOR.2004.07.077

[4] C. Ren, "Research on VRPTW optimizing based on k-means clustering and IGA for electronic commerce," in *Proc. 3rd IEEE Conf. Ind. Electron. Appl.*, 2008, pp. 61-66.

[5] M. Schneider, A. Stenger, and D. Goeke, "The electric vehicle-routing problem with time windows and recharging stations," *Transp. Sci.*, vol. 48, no. 4, pp. 500-520, Mar. 2014, https://doi.org/10.1287/trsc.2013.0490

[6] A. F. L. Villalba and E. C. G. L. Rotta, "Clustering and heuristics algorithm for the vehicle routing problem with time windows," *Int. J. Ind. Eng. Comput.*, vol. 13, no. 2, pp. 165-184, 2022, https://doi.org/10.5267/j.ijiec.2021.12.002

[7] G. D. Konstantakopoulos, S. P. Gayialis, and E. P. Kechagias, "Vehicle routing problem and related algorithms for logistics distribution: A literature review and classification," *Oper. Res. (Lond.)*, vol. 22, no. 3, pp. 2033-2062, Sep. 2020, https://doi.org/10.1007/s12351-020-00600-7

[8] L. Moccia, J.-F. Cordeau, and G. Laporte, "An incremental tabu search heuristic for the generalized vehicle routing problem with time windows," *J. Oper. Res. Soc.*, vol. 63, no. 2, pp. 232-244, Feb. 2012, https://doi.org/10.1057/jors.2011.25

[9] Y. Wang, L. Wang, G. Chen, Z. Cai, Y. Zhou, and L. Xing, "An improved ant colony optimization algorithm to the periodic vehicle routing problem with time window and service choice," *Swarm Evol. Comput.*, vol. 55, p. 100675, Jun. 2020, https://doi.org/10.1016/j.swevo.2020.100675

[10] S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for K-means clustering," *Pattern Recognit. Lett.,* vol. 25, no. 11, pp. 1293-1302, Aug. 2004, https://doi.org/10.1016/j.patrec.2004.04.007

[11] R. Chunyu and W. Xiaobo, "Research on VRP optimizing based on hierarchy clustering and IGA under common distribution," in *Proc. 2006 Int. Conf. Comput. Intell. Secur*., 2006, pp. 143-146. https://doi.org/10.1109/ICCIAS.2006.294108

[12] K. C. Tan, L. H. Lee, Q. L. Zhu, and K. Ou, "Heuristic methods for vehicle routing problem with time windows," *Artif. Intell. Eng.*, vol. 15, no. 3, pp. 281-295, Jul. 2001, https://doi.org/10.1016/s0954-1810(01)00005-x

[13] S. E. Comert, H. R. Yazgan, S. Kır, and F. Yener, "A cluster first-route second approach for a capacitated vehicle routing problem: A case study," *Int. J. Procurement Manage.*, vol. 11, no. 4, p. 399, Jan. 2018, https://doi.org/10.1504/ijpm.2018.092766

[14] R. L. Thorndike, "Who Belongs in the Family?," *Psychometrika*, vol. 18, no. 4, pp. 267-276, Dec. 1953, https://doi.org/10.1007/bf02289263

[15] N. Bertagnolli, "Elbow method and finding the right number of clusters." *Nicolas Bertagnolli*, 2015. [Online]. Available: https://www.nbertagnolli.com/jekyll/update/2015/12/10/Elbow.html [Accessed: Apr. 2, 2025].

[16] M. M. Solomon, "Algorithms for the vehicle routing and scheduling problems with time window constraints," *Operations Research*, vol. 35, no. 2, pp. 254-265, Apr. 1987, https://doi.org/10.1287/opre.35.2.254

[17] E. Queiroga, "CVRPLIB: Capacitated vehicle routing problem library," *Pontifical Catholic University of Rio de Janeiro*, 2025. [Online]. Available: https://vrp.galgos.inf.puc-rio.br/index.php/en/ [Accessed: Sep 20, 2025].


**Kanokporn Boonjubut** is a lecturer in the Department of Engineering Management at Nakhon Rachasima Rajabhat University, Thailand. She received her Ph.D. degree in Functional Control Systems from Shibaura Institute of Technology (SIT), Japan. Her research interests include optimization, logistics, and supply chain management.

**Prat Boonsam** is an Assistant Professor of the Department of Engineering Management at Nakhon Rachasima Rajabhat University, Thailand. He received his Ph.D. in Logistics from the University of the Thai Chamber of Commerce, Thailand. His interests include truck distribution planning and industrial productivity improvement.

**Sirichai Yodwangjai** is an Assistant Professor in the Department of Industrial Engineering Technology, College of Industrial Technology, King Mongkut's University of Technology North Bangkok, Thailand. He received a Ph.D. from Naresuan University, Phitsanulok, Thailand. His research interests are optimization, metaheuristics, logistics, and supply chain.