INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 32 (2025) 863–887
DOI: 10.1111/itor.13282

Clustered vehicle routing problem for waste collection with smart operational management approaches

Jungmin Kim[a], Apurba Manna[b], Arindam Roy[a,c] and Ilkyeong Moon[a,d,*]

[a] *Department of Industrial Engineering, Seoul National University, Seoul 08826, South Korea*
[b] *Research Centre in Natural Sciences, Raja N. L. Khan Women's College, Medinipur 721102, India*
[c] *Department of Computer Science & Application, Prabhat Kumar College, Contai 721404, India*
[d] *Institute of Engineering Research, Seoul National University, Seoul 08826, South Korea*
E-mail: andykim14@snu.ac.kr [Kim]; apurba_sukumar@rnlkwc.ac.in [Manna]; arcontai@gmail.com [Roy]; ikmoon@snu.ac.kr [Moon]

Received 29 July 2022; received in revised form 18 November 2022; accepted 27 February 2023

Abstract

Waste collection is one of the essential tasks in a smart city. The Internet of Things (IoT) is a promising technology that offers potential solutions for transforming traditional systems. An IoT-based smart bin is a modern technology that offers real-time fill level information to a cleaning authority. However, high uncertainty associated with the smart bin's fill levels and improper operation hinder efficient waste collection. In order to tackle the uncertainty in a smart bin and improve the waste collection operation, the IoT sensor's usage must be combined with optimization procedures. The present work introduced two operational management approaches to define dynamic optimal routes and combined ant colony optimization with a k-means clustering algorithm to solve the clustered vehicle routing problem for waste collection on a large scale. Operational management approaches reflect practical constraints when using IoT-based smart bins. A hybrid metaheuristic is proposed and performed with these approaches thereby showing the potential of building a smart waste collection system.

Keywords: clustering; ant colony optimization; smart bin; waste collection; vehicle routing

1. Introduction

The Internet of Things (IoT) is gaining attention in numerous research areas because it has a good effect on the way operations are managed (Atzori et al., 2017). To take advantage of this technology, governments and companies are developing capabilities to maximize their utilities to work in tandem with the IoT. Doing so will potentially boost the efficiency and better effectiveness of their

*Corresponding author.

operations while streamlining decision-making within the organizations (Fayoumi and Loucopoulos, 2016).

Waste has increased drastically with rapid urbanization, and the outbreak of COVID-19 has accelerated mainly plastic waste production (Peng et al., 2021). According to recent findings, more than eight million tons of pandemic-associated plastic waste have been generated globally, with more than 25,000 tons entering the global ocean. While dealing with this unexpected surge in waste, conventional incineration facilities and landfills have reached their limit. Hence, the waste management industry faces intense pressure over handling hazardous waste generated from COVID-19. The utilization of IoT technologies can pave the way for better monitoring and control of waste.

Some researchers have studied different areas within waste management. Del Pia and Filippi (2006) considered moving depots for waste collection and proposed a heuristic to reduce the total route duration. Zhao and Zhu (2016) considered planning tours and vehicle acquisitions for waste collection. Sahebjamnia et al. (2018) proposed a multiobjective model with environmental considerations in the context of a closed-loop supply chain. Coban et al. (2018) addressed the multicriteria decision-making methods to dispose of municipal waste and indicated the prominence of recycling and landfill technologies in developing countries. López-Sánchez et al. (2018) addressed the multiobjective routing problem for waste collection and proposed a hybrid heuristic to solve a real-world problem. Ramos et al. (2018) addressed the recyclable waste collection problem using smart bins and developed optimization procedures. Lin et al. (2020) addressed a robust facility location model for waste transfer stations under uncertain environments. Eryganov et al. (2020) presented possible approaches to modeling cooperation between the waste producers in a certain location with limited or banned landfilling using the cooperative game theory. Adeleke and Ali (2021) suggested point-of-collection sorting for efficient collection. In this paper, we focus on the vehicle routing problem for waste collection using a smart bin.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature. Section 3 provides a detailed problem description and two operational management approaches with mathematical models. In Section 4, a hybrid metaheuristic is presented. Computational experiments are given in Section 5. Section 6 presents the contributions and summary of this paper.

2. Literature review

Increasing waste has become one of the more concerning issues in a smart city because of rapid urbanization. To tackle this problem, many authorities spotlight the importance of smart bins. The smart bin equipped with an IoT sensor provides real-time fill level information to the authorities to aid them in making proper decisions to collect waste. Huang and Lin (2015) made decisions on optimal route planning and scheduling for collecting waste. However, they did not consider IoT-based smart bins. Ramos et al. (2018) presented operational management approaches using a smart bin. They introduced some parameters to estimate the fill level of bins. López-Sánchez et al. (2018) proposed a multiobjective problem for waste collection with a hybrid metaheuristic and a case study in Spain is presented. Ferrer and Alba (2019) elaborated on the mechanism of the IoT-based smart bin but failed to deliver a mathematical model. Ríos-Mercado et al. (2023) presented a dispersion territory design problem considering legal constraints and proposed an advanced metaheuristic algorithm to solve this problem. Delgado-Antequera et al. (2020) proposed a multiobjective approach for waste collection and proposed a hybrid heuristic, while Haque et al. (2020) presented an IoT-based efficient waste collection system with smart bins. Nonetheless, their works are confined to a systematic approach to collecting waste. Salamirad et al. (2023) performed a multicriteria decision-making analysis to treat industrial wastewater. They also proposed a new hybrid heuristic to suggest managerial insights in a case study. Sarkar et al. (2022a) focused on food waste with a circular economy concept. Exhaustive experimental results are supported with extensive supply chain analysis. Roy et al. (2022) considered IoT-based smart bins with multiple waste types and utilized vehicle compartments for each waste type. Sarkar et al. (2022b) proposed a green supply-chain management setup for biodegradable products with an outsourcing strategy and transportation modes. To the best of our knowledge, this paper is the first study that considers the usage of IoT-based smart bin in a mathematical model and suggests operational management approaches for waste collection.

The clustered vehicle routing problem (CluVRP) generalizes the capacitated vehicle routing problem (CVRP). Sevaux et al. (2008) introduced CluVRP in the context of a real-world application where containers are employed to carry goods. To the best of our knowledge, the literature describes only two exact approaches for the CluVRP. Pop et al. (2012) presented two compact formulations but did not show computational results. Battarra et al. (2014) developed two exact algorithms and provided results for a set of benchmark instances.

In the CluVRP, numerous researchers suggested different criteria for generating clusters. Ghiani et al. (2005) allocated arcs and edges to vehicles for satisfying service deadlines and container-to-vehicle compatibility constraints. Defryn et al. (2016) investigated the effect of the cost allocation of a partner's strategy on nondelivery penalties and the properties of its customer locations (distance to the depot, degree of clustering). Fernández et al. (2018) introduced the shared customer collaboration in a vehicle routing problem (VRP). They assumed that some customers are shared by different vehicles, in the sense that they have demand from more than one vehicle. Defryn and Sörensen (2017) presented an improved two-level heuristic to solve the CluVRP. And they introduced a new variant of the CluVRP, the CluVRP with weak cluster constraints. Hintsch and Irnich (2018) decomposed CluVRP into three subproblems (i.e., the assignment of clusters to routes, the routing inside each cluster, and the sequencing of the clusters in the routes). In this paper, we follow the decomposing method presented by Hintsch and Irnich (2018). For the assignment of clusters, the k-means clustering algorithm is used. Furthermore, an ant colony optimization (ACO) is applied for calculating routing inside each cluster. Finally, for the sequencing of the clusters in the route, smart bins equipped with IoT sensors are used so that vehicles start their journey when they receive signals from smart bins. Al-Refaie et al. (2021) considered clusters for efficient waste collection. Hubs are allocated in each cluster, and a vehicle visits these hubs instead of visiting whole bins. In this paper, bins are segmented into each cluster, and a vehicle can enter and leave a cluster. Some previous research in the waste collection are compared with the present study in Table 1.

The CluVRP-WC is a generalization of the CVRP in which smart bins are segmented by clusters. Furthermore, because determining the optimal solution to VRP is an $\mathcal{NP}$-hard, as per Toth and Vigo (2002), and the CluVRP-WC is a special case of the VRP, the CluVRP-WC is also an $\mathcal{NP}$-hard. To tackle this problem, this paper suggests two operational management approaches. In the first experiment, this paper extends the experiment presented by Kim et al. (2020) to compare the operational management approaches. In the next experiment, the proposed hybrid metaheuristic was implemented to solve this problem on a large scale.

Table 1
Contribution of different authors

Authors (year)	Model	Smart bin	Clustering	Waste type	Solution methodology
Zhao and Zhu (2016)	MDVRP			Recyclable	Lexicographic weighted Tchebycheff method
Defryn and Sörensen (2017)	CluVRP		✓		Variable neighborhood search
Ramos et al. (2018)	DVRP	✓		Recyclable	Heuristic
López-Sánchez et al. (2018)	WCP			Solid	Variable neighborhood descent
Hintsch and Irnich (2018)	CluVRP		✓		Neighborhood search, Variable neighborhood descent
Ríos-Mercado et al. (2023)	MDTDP			Electronics	Metaheuristic
Delgado-Antequera et al. (2020)	MCDM			General	Heuristic
Salamirad et al. (2023)	CVRP			Water	Variable neighborhood search
Jorge et al. (2022)	VRPP	✓		Recyclable	Simulated annealing, Neighborhood search
Sarkar et al. (2022a)	CEM			Food	Stochastic dual coordinate ascent
Roy et al. (2022)	CVRP	✓		Food, Recyclable, General	Variable neighborhood search
Sarkar et al. (2022b)	GSC			Biodegradable	Operational management approaches
This paper	CluVRP	✓	✓	General	Operational management approaches

Abbreviations: CEM, circular economic model; DVRP dynamic VRP; GSC, green supply chain; MCDM, multicriteria decision-making; MDTDP, maximum dispersion territory design problem; MDVRP, multi depot VRP; VRP, vehicle routing problem; VRPP, VRP profit; WCP, waste collection problem.

3. Mathematical model

3.1. Motivation

In this section, we present the motivation for the smart waste collection problem occurring in a smart city where IoT-based smart bins are placed around the city. Because of irregular cleaning and improper waste management, most waste bins get overfilled frequently and consequently create unhealthy and nonhygienic situations. In the proposed model, these problems can be solved by setting up IoT-based smart bins, which send alerts to the cleaning authority when they are filled up to a predefined level, say 80%. During routing, a driver cannot check all bins fill level information status because of the limitation of data transmission. According to Kim (2015), the IoT sensor range is classified by its length (i.e., short-range for Bluetooth mesh networking, Wi-Fi, radio frequency identification (RFID); medium-range for LTE and 5G; and long-range for low-power wide-area networking and very small aperture terminals. In this paper, it is assumed that smart bins are equipped with a short-range (maximum 10 m) IoT sensor. In addition, each vehicle is equipped with an IoT sensor receiver so that neighboring bins locations and fill level information are sent to the vehicle during routing. This paper introduces a neighboring bound to represent the radius of the IoT sensor. To the best of our knowledge, this paper is the first study that suggests a set concept for certainly filled bins, penalty bins, and neighboring bins in the mathematical model.

3.2. Problem description

The problem addressed in this paper is clustered vehicle routing problem for waste collection (CluVRP-WC), considering the use of real-time information about bin fill levels to define dynamic routes. This problem can be defined as follows: A cluster set is defined as M, based on the k-means clustering algorithm and elbow method. For simplicity, the k-means algorithm is used to generate clusters. A set of the depot and all bins is defined as B. A set of bins in each cluster is defined as $(B_m, m \in M)$. Each bin i's fill level is stated with L_i. A complete undirected graph is considered to connect all bins and a depot, with the distance, d_{ij}, for each edge (i, j) in the graph. K is assumed to be a set of homogeneous vehicles with a maximum capacity of L each. All vehicles depart and return to the depot. The maximum number of vehicles is less than or equal to the number of clusters to utilize the vehicle fully. Following Expósito-Izquierdo et al. (2016), the CluVRP-WC can be defined by the mathematical model below, which requires the definition of some additional variables. Suppose Z to be any subset of B that is different from B. Then, let $\delta^+(Z)$ be the set of edges $(i, j) \in Z \times B\backslash Z$, and $\delta^-(Z)$ be the set of edges $(i, j) \in B\backslash Z \times Z$. This paper considers several sets to represent the CluVRP-WC model. These are cluster sets, filled bin sets, penalty bin sets, and neighboring bin sets. If the fill levels of bins are above the threshold fill level (TFL), these bins are defined as filled bin set, B_f. The penalty is imposed on overflowing bins (i.e., a penalty bin set is defined when the fill level is more than 100%) and is defined as B_p. Sets of the filled bins and penalty bins in each cluster are defined as B_{mf} and B_{mp}. Following the definition by Jorge et al. (2022), these filled bins are included in "must-go bins." Considering the IoT sensor range, which is defined as aneighboring bound, R, a neighboring bin set is introduced. When a bin is given, neighboring bins are within the neighboring bound from the bin. These bins are included in the neighboring bin set. The total time horizon is considered as T. At the beginning of each day, $t(\in T)$, in the morning, the sensors located inside the waste bins transmit information on the bins' fill levels. The problem in hand selects the waste bins to be visited (must-go bins) and the optimal visiting sequence in each day t for each vehicle k, which will minimize the summation of routing cost throughout the clusters while satisfying the vehicles' fixed capacity. Routing cost and penalty cost are included in the total cost. The assumptions of the mathematical model are as follows.

Assumptions:

1. Smart bins are equipped with IoT sensors and identical with the same capacity. The working principle is taken from Roy et al. (2022).
2. Vehicles are equipped with IoT sensor receivers and homogeneous with the same capacity. During routing, a vehicle can receive fill level information within a neighboring bound from its current location.
3. A general type of waste is considered.
4. If the fill level in a bin reaches the TFL, then an alert message is transmitted to the cleaning authority so that the bin can be visited.
5. Every bin in each cluster will be visited by the same vehicle during collection time and will be visited if the fill level is above TFL or the bin is within the neighboring bound from the filled bin.
6. All filled bins (above TFL) must be visited. If not, a penalty is imposed in proportion to the fill level of the filled bin.

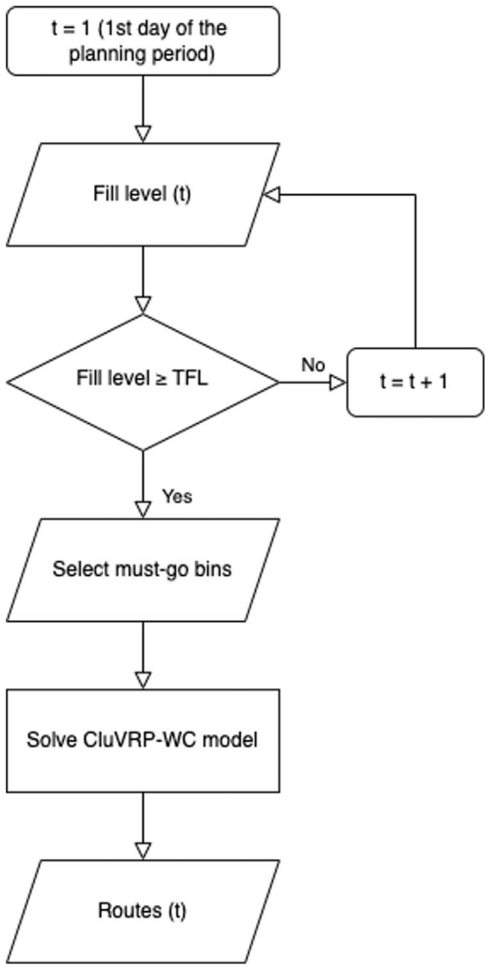

Fig. 1. Operational management approaches.

3.3. Operational management approaches

Considering the problem description above, two operational management approaches to define dynamic collection routes are studied. Figure 1 shows the outline of operational management approaches. On the first day of the planning period, a cleaning authority checks the fill level information from all bins. If the bins' fill levels are above TFL, an alert message is sent to the cleaning authority so that a vehicle can depart from a depot. In the selection process for must-go bins, two operational management approaches are introduced. After one of the approaches is decided, then solve the CluVRP-WC model and waste collection proceeds.

(1) *estimation-based collection approach*, based on Phase 1 heuristic by Jorge et al. (2022) in which fill level estimation is considered to select the must-go bins at each day. This is coupled with a

CluVRP-WC model that optimizes each vehicle's route, defining the best sequence to visit the selected bins.

(2) *Neighborhood-based collection approach*, in which a neighboring bin is proposed to select the must-go bins to perform the collection operation, and the revised mathematical model is presented.

The first approach is based on the method followed by Jorge et al. (2022), where the selection of the bins to be visited in each day is based on currently filled bins (above TFL) and future filled bins considering a fill level estimation. Once the set of bins to be visited is defined, each vehicle route is optimized through a CluVRP-WC model. In the neighborhood-based collection approach, a neighboring bin is introduced. A feasible set that a vehicle visit is reduced to the set of filled bins and their neighboring bins. While visiting filled bins, a vehicle will visit neighboring bins from the filled bin if the vehicle has enough space to collect waste. Between these approaches, the only difference is the must-go bin selection. The notation used to formulate the CluVRP-WC is as follows:

Sets

M	Set of m clusters, $M = \{1, 2, \dots, m\}$
B_m	Set of bins in cluster m, $b_{mi} \in B_m$, $b_{mi} =$ bin i in cluster m, $\forall m \in M$
$\{0\}$	Depot
B	Set of the depot and all bins, $B = \{0\} \cup B_1 \cup B_2 \cup \cdots \cup B_m$
B_{mf}	Set of the filled bins in cluster m, fill level$\geq TFL$, $B_{mf} \subseteq B_m$, $\forall m \in M$
B_f	Set of all filled bins, $B_f = B_{1f} \cup B_{2f} \cup \cdots \cup B_{mf}$
B_{mp}	Set of penalty bins in cluster m, fill level$\geq$ 100%, $B_{mp} \subseteq B_{mf}$, $\forall m \in M$
B_p	Set of all penalty bins, $B_p = B_{1p} \cup B_{2p} \cup \cdots \cup B_{mp}$
K	Set of vehicles, $\lvert K\rvert \leq m$
$\delta^+(Z)$	Set of edges $(i, j) \in Z \times B\backslash Z$
$\delta^-(Z)$	Set of edges $(i, j) \in B\backslash Z \times Z$
$N[b_{mi}]$	Neighbor of b_{mi}, $N[b_{mi}] = \{b_{mi}\} \cup \{b_{mx} : b_{mx} \in B_m ; i \neq x, \ \forall m \in M, \forall i, x = 1, 2, \dots, \lvert B_m\rvert$ and $d(b_{mx}, b_{mi}) \leq R$, d denotes the euclidean distance$\}$
T_m	$B_{mf} \cup N[b_{ml}]$, $b_{ml} \in B_{mf}$

Parameters

TFL (%)	Threshold fill level
d_{ij} (km)	Travel cost required to move a vehicle from node i to node j
L_i (kg)	Amount of waste at bin i
P_i ($)	Penalty if bin i's fill level is more than 100%
L (kg)	Maximum capacity of a vehicle
T	Total time horizon
R (m)	Neighboring bound

Variables

$$x_{ijk} = \begin{cases} 1 & \text{if a vehicle } k \text{ traverses from node } i \text{ to node } j \text{ in the undirected graph} \\ 0 & \text{otherwise} \end{cases}$$

$i, j \in B, k \in K$

$$y_i = \begin{cases} 1 & \text{if a bin} i \text{is visited in the solution} \\ 0 & \text{otherwise} \end{cases}$$

$i \in B\backslash\{0\}$

To build the CluVRP-WC model, routing cost for collecting waste is considered in the objective function and penalty cost is added to the total cost throughout all clusters. The total cost is given by

$$\sum_{k\in K}\sum_{i\in B}\sum_{j\in B} x_{ijk}d_{ij} + \sum_{i\in B_p} P_i. \tag{1}$$

The followings are constraints. Constraint (2) enforces that all vehicles start from a depot and return to the depot.

$$\sum_{i\in B\setminus\{0\}} x_{0ik} = \sum_{j\in B\setminus\{0\}} x_{j0k} = 1, \quad \forall k \in K. \tag{2}$$

Constraint (3) establishes that only one vehicle can enter and leave a cluster.

$$\sum_{k\in K}\sum_{(i,j)\in\delta^+(B_m)} x_{ijk} = \sum_{k\in K}\sum_{(i,j)\in\delta^-(B_m)} x_{ijk} = 1, \quad \forall m \in M \tag{3}$$

Constraint (4) verifies the same vehicle can enter and leave a cluster.

$$\sum_{(i,j)\in\delta^+(B_m)} x_{ijk} = \sum_{(i,j)\in\delta^-(B_m)} x_{ijk}, \quad \forall m \in M, \quad \forall k \in K. \tag{4}$$

Constraint (5) indicates that each vehicle cannot exceed its maximum capacity.

$$\sum_{i\in B\setminus\{0\}}\sum_{j\in B} x_{ijk}L_i \le L, \quad \forall k \in K, \quad i \ne j. \tag{5}$$

Constraint (6) guarantees that every bin is visited at most once in the solution.

$$\sum_{k\in K}\sum_{i\in B\setminus\{0\}} x_{ijk} = y_j, \quad \forall j \in B\setminus\{0\}, \quad i \ne j. \tag{6}$$

Constraint (7) ensures the connectivity of the path of a single vehicle.

$$\sum_{i\in B\setminus\{0\}} x_{ipk} = \sum_{j\in B\setminus\{0\}} x_{pjk}, \quad \forall k \in K, \quad \forall p \in B\setminus\{0\}. \tag{7}$$

Constraint (8) represents the classic subtour elimination constraints.

$$\sum_{i\in S}\sum_{j\in S} x_{ijk} \le |S| - 1, \quad \forall k \in K, \quad \forall S \subseteq B\setminus\{0\}, \quad |S| \ge 2. \tag{8}$$

Constraint (9) represents all filled bins (above TFL) that must be visited.

$$\sum_{i\in B_f} y_i = |B_f|. \tag{9}$$

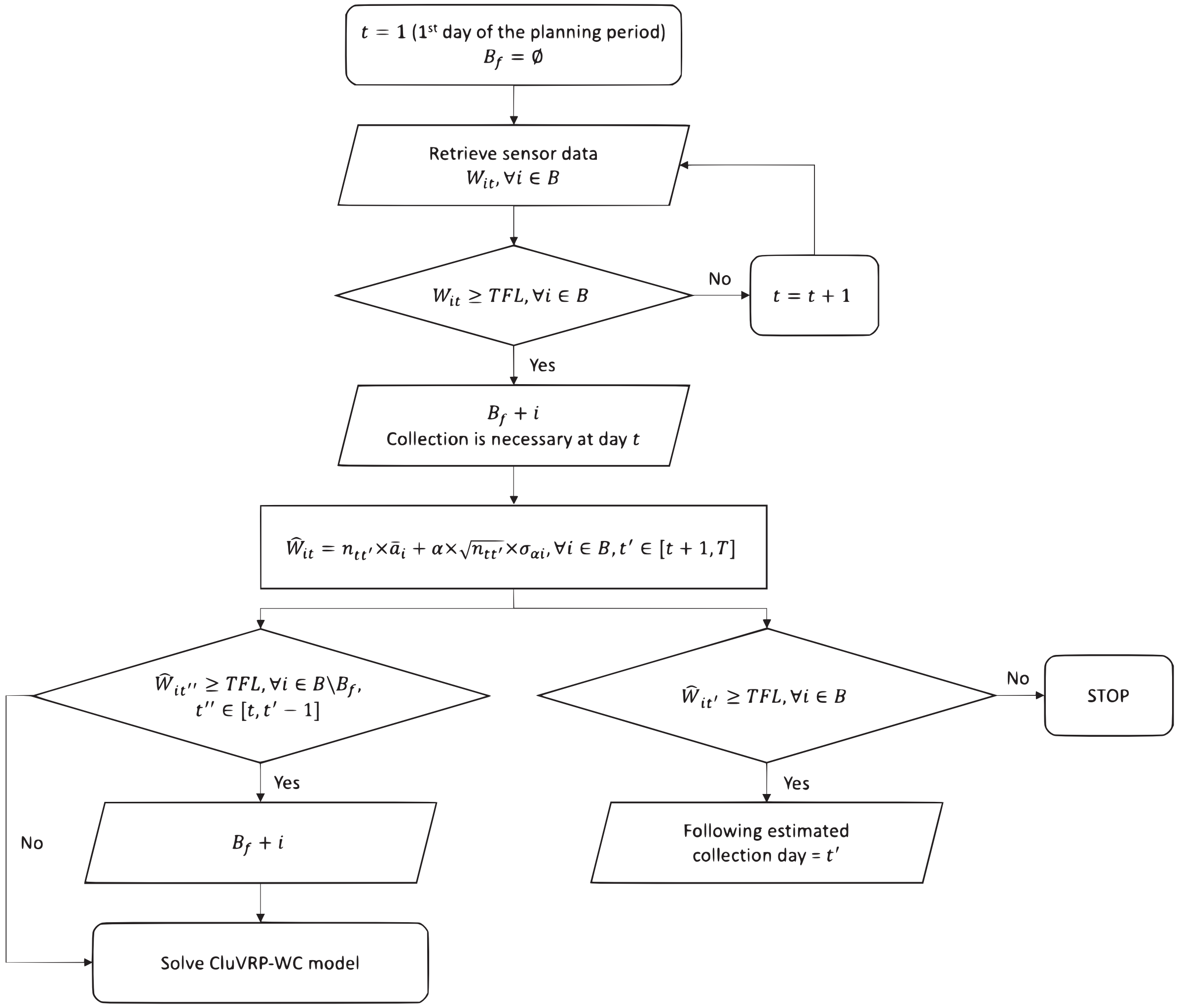

Fig. 2. Must-go bins selection algorithm.

Constraints (10) and (11) present binary variables.

$$x_{ijk} \in \{0, 1\}, \ \forall i \in B, \ \forall j \in B, \ \forall k \in K, \tag{10}$$

$$y_i \in \{0, 1\}, \ \forall i \in B. \tag{11}$$

3.3.1. Estimation-based collection approach

The fill level estimation and "must-go bins" selection algorithm are the ones proposed by Jorge et al. (2022). We followed the notations, parameters, and the outline of the must-go bins selection algorithm by Jorge et al. (2022), and the differences are described in Fig. 2. While they considered full or overflowing bins for selecting must-go bins, we considered a certain threshold fill level for selecting bins to avoid overflowing bins. The bins to be visited are decided in the morning of each day t, after receiving the real-time information on the bins' fill level (L_i). "Must-go bins" are defined as the combination of bins in which the fill level is more than TFL at day t and the fill level is estimated to be above TFL from $t+1$ to t' ($t' \in (t+1, T)$). TFL indicates a certain threshold fill level of a bin. The first collection day is the day on which the bins' fill level is above TFL at day t. These bins are included in the must-go bins ($B_f + i$). The following estimated collection day is the day in which the bins that were above TFL on the first collection day are predicted to become above TFL again ($\hat{W}_{it'} \geq TFL$). Let us assume the following estimated collection day is t'. Between day t and day t', check bins' fill level status which is above TFL and is not included in the must-go bins on the first collection day. If the fill level is estimated to be above TFL during this time period, these bins are included on the first collection day to avoid potential overflowing. The next step is to solve the CluVRP-WC model. To build the cost minimization model for waste collection, the following sets, parameters, and variables are defined. The number of clusters, m, is decided using a k-means clustering algorithm and the elbow method. The following parameters are adopted from Jorge et al. (2022).

Parameters

W_{it} (kg)	Amount of waste at bin i at day t
$\hat{W}_{it'}$ (%)	Estimated waste fill level in bin i at day t'
$n_{tt'}$	Number of days between collection day t and day t'
$\bar{a}_i$ (%)	Average daily accumulation rate of bin i
α	Value for a probability retrieved from the normal distribution depending on the desired level of confidence for $\hat{W}_{it}$
σ_{ai}	Standard deviation of the daily accumulation rates of bin i

3.3.2. Neighborhood-based collection approach

In this approach, this paper introduced that a vehicle visits neighboring bins from the filled bins instead of using fill level estimation. All the filled bins are included in the must-go bins as in the estimation-based collection approach. Furthermore, the must-go bins selection algorithm is revised in Fig. 3. A vehicle will visit all the filled bins (above TFL) and their neighboring bins. If the vehicle has enough space to hold waste after visiting a filled bin ($L - \sum_{i \in B_f} W_{it} > 0$), then the vehicle searches neighboring bins from the filled bin ($B_f + N[b_{mi}]$). For the mathematical modeling part, only the differences are presented below. A subset of filled bins and their neighboring bins in each cluster is defined as T_m. Parameter R represents the radius of the IoT sensor on bins. A feasible area that a vehicle can visit is reduced to the set of the filled bins and their neighboring bins. Therefore, B_m in constraints (3) and (4) is replaced with T_m in constraints (12) and (13). Other constraints are the same as the estimation-based collection approach.

$$\sum_{k \in K} \sum_{(i,j) \in \delta^+(T_m)} x_{ijk} = \sum_{k \in K} \sum_{(i,j) \in \delta^-(T_m)} x_{ijk} = 1, \quad \forall m \in M, \tag{12}$$

$$\sum_{(i,j) \in \delta^+(T_m)} x_{ijk} = \sum_{(i,j) \in \delta^-(T_m)} x_{ijk}, \quad \forall m \in M, \quad \forall k \in K. \tag{13}$$

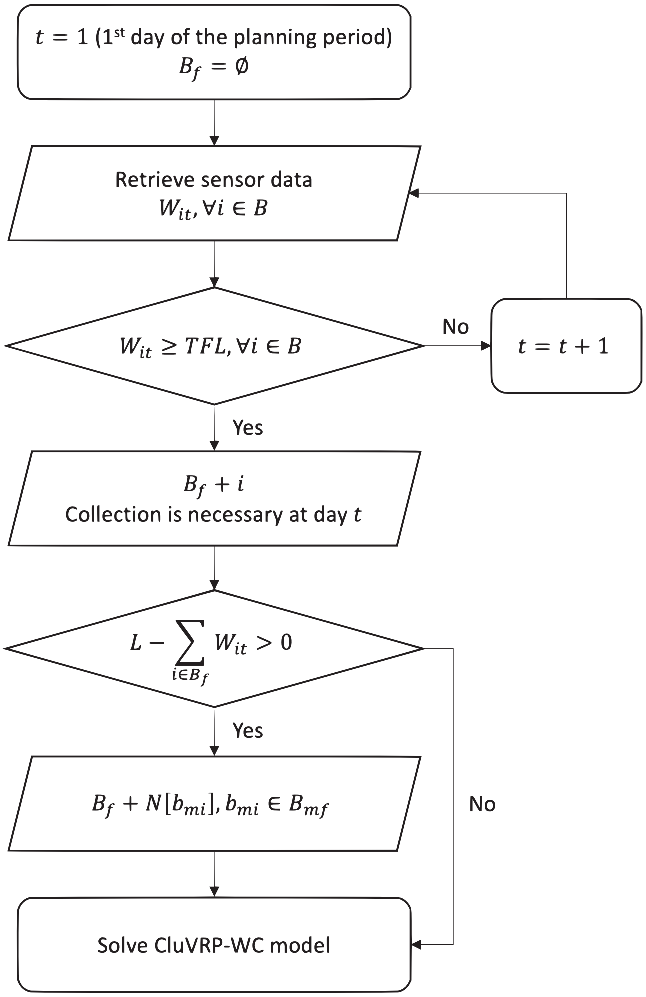

Fig. 3. Revised must-go bins selection algorithm.

The CluVRP-WC is a generalization of CVRP. For this reason, this problem can be solved in a limited instance size. To tackle this problem, this paper presented a hybrid metaheuristic combining the k-means clustering algorithm and ACO algorithm.

4. Hybrid metaheuristic

In this section, this paper proposed an algorithm that combines the k-means clustering algorithm (Algorithm 1) and ACO. The waste bins are segmented based on the k-means clustering algorithm, and paths in each cluster are optimized by using ACO. The methodology is described as follows.

4.1. Cluster computation and optimal path design

Numerous methods for generating clusters (e.g., distribution-based, density-based, graph-based, prototype-based, etc.) are proposed by Tan et al. (2016). In this problem, waste bins are distributed in distinct geographical locations. For simplicity, the k-means clustering algorithm is used, one of the prototype-based clustering methods, to group the bins. This section uses the clustering technique to improve the performance of the evolutionary computation algorithm.

The k-means clustering algorithm is a distance-based clustering technique from which centroids of each cluster are calculated. This algorithm performs an iterative alternating fitting process to form a specific number of clusters. Initially, c points are randomly selected to be a first guess of the centroids of the clusters. Each node is assigned to the nearest centroid to form temporary clusters. In the next step, the centroids are replaced by means of clusters, and the nodes and centroids are reassigned. This process continues until no further changes occur in the clusters. The k-means clustering algorithm is one of the partition-based clustering techniques proposed by Hartigan and Wong (1979). The classical k-means clustering algorithm is described in Algorithm 1. In this proposed CluVRP model, all the waste bins are grouped using the k-means clustering algorithm. The number of clusters is selected by the elbow method, the distance is based on Euclidean types, and mean square errors examine its consistency. The k-means clustering algorithm is given below:

Algorithm 1. k-means clustering algorithm

Require: Set of bins and clusters
Ensure: Unique set of bins in each cluster; centroid of each cluster
Step 1: Initialize a positive integer c ($c = |M|$) using the elbow method
Step 2: Generate bins in each cluster (an initial set) randomly
Step 3: Determine the centroid of each cluster
Step 4: Allocate each bin to the cluster where it is nearest to the centroid
Step 5: Update the centroid in each cluster
Step 6: Repeat Steps 5 and 6 until the centroids no longer move

After allocating the bins to each cluster, the next aim is to find the optimum path in each cluster. In the exact approach, it is solved in a limited instance size and difficult to determine the best path that considers the routing cost in a feasible time. Thus, this paper presents the evolutionary computation method. With this heuristic, all nodes are clustered into $|M|$ groups, and an optimal path is calculated in each cluster.

4.2. k-Means-ant colony optimization

The proposed ACO and its procedures in each cluster are given as follows.

4.2.1. Representation

In each cluster, a complete cycle of vehicle depot, filled bins, and neighboring bins represent a solution of ants. Therefore, an $N(p+1)$-dimensional integer vector, $X_i = (\{0\}, b_{m1}, b_{m2}, \ldots, b_{mp})$, $m = 1, 2, \ldots, |M|$, is used to represent a vehicle depot and p the number of must-go bins. To be specific,

Algorithm 2. Path construction

Require: A depot and the set of bins of a cluster
Ensure: The optimum tour for the rth ant of the cluster;
Step 1: Set $E = \{1, 2, \ldots, N\},\ l = 1$ and $E' = \{2, 3, \ldots, N\}$. Node 1 to N in E represent a depot and must-go bins, respectively
Step 2: x_{rl}= a random element from the set E'
Step 3: $i = x_{rl}$
Step 4: Set WL=0; WL means waste load
Step 5: while $\{((L - WL) >= L_i)$ *and* $(l < N)\}$
$\{WL = WL + L_i$
Set E'=E'-$\{i\}$
Let bin i be the present position of an ant. Then the next bin $j \in E'$ is selected by the ant, with probability p_{ij} given by the formula, $p_{ij} = \frac{\tau_{ij}^{\delta_1}}{\sum_{j \in E'} \tau_{ij}^{\delta_1}}$, where δ_1 is a user-defined parameter that controls the relative importance of pheromone concentration. A roulette-wheel selection process is used for this parameter.
$l = l + 1, i = j\}$
Step 6: Print the optimal tour of the cluster

$\{0\}$ represents a vehicle depot and $b_{m1}, b_{m2}, \ldots, b_{mp}$ represent a set of must-go bins in cluster m in a complete cycle. In the proposed ACO, τ_{ij} represents the amount of pheromone that lies on the path between bin i and bin j.

4.2.2. *Pheromone initialization*

As the aim of the CluVRP-WC is to minimize the total routing cost on the proposed model, it is assumed that in each cluster, an initial value of pheromone between bin i and bin j is $\tau_{ij} = \frac{1}{\sqrt{d_{ij}}}$.

4.2.3. *Path construction*

In each cluster, the following are required to construct a path X_r for the rth ant and presented in Algorithm 2. n paths are constructed for n different ants and all vehicles, $\forall k \in K$.

4.2.4. *Pheromone evaporation*

In each cluster, the following formula is used between bin i and bin j for the evaporation of pheromone:

$$\tau_{ij} = (1 - \rho)\tau_{ij}, \tag{14}$$

where ρ lies between [0, 1]. The constant ρ specifies the pheromone evaporation rate, causing ants to forget previous decisions.

4.2.5. *Pheromone updating*

After the completion by all ants, a pheromone is increased on the paths through which the ants have traveled. Depending upon the nature of the present problem, a pheromone is updated using the

Algorithm 3. k-means-ACO

Require: Problem data, ACO parameters, number of clusters ($|M|$)
Ensure: The optimum tour in each cluster
Step 1: Initialize a set of bins and the number of clusters ($|M|$)
Step 2: Execute a k-means clustering algorithm (subsection 4.1)
Step 3: Start ACO
Step 4: Identify the fill level information above TFL
Step 5: Distribute the total number of ants
Step 6: Determine pheromone initialization (subsection 4.2.2)
Step 7: Execute path construction (subsection 4.2.3)
Step 8: Perform pheromone evaporation (subsection 4.2.4)
Step 9: Execute pheromone updating (subsection 4.2.5)
Step 10: If pheromone updating occurred, go to Step 7
Step 11: Find the optimal tour in each cluster

Table 2
Parameters of the k-means-ACO algorithm

Parameter	α	β	ρ	n	δ_1
Value	1	2	0.5	300	1

following rules, where ρ represents the rate of evaporation and n is the number of ants. Algorithm 3 shows the outline of the proposed hybrid heuristic.

$$\tau_{ij} = (1-\rho)\tau_{ij} + \frac{\rho}{n}\sum_{i=1}^{n}\tau_{ij}^{best}. \tag{15}$$

5. Computational experiments

In this section, this paper presents the comparison results of two operational management approaches. At first, experiments were carried out in a small size with an exact method and the proposed heuristic. These experiments extend the study presented by Kim et al. (2020). Their study showed the practical usage of a smart bin and presented the outline of optimal waste collection. However, the number of bins in their experiments is restricted to only five and they considered one vehicle. In the small-size experiments, we extended this experiment with 15 bins and four vehicles. In addition, 20 different instances are used to represent the diverse random fill level information. This paper also considered penalty costs for overflowing bins and applied two operational management approaches. The limitation of the estimation-based collection approach is presented, and large-size experiments were conducted to support this result. To judge the effectiveness and feasibility of the proposed hybrid algorithm, a dispersion test was conducted based on the traveling salesman problem library (TSPLIB) presented by Reinelt (1991). In the large-size experiment, this problem is hard to solve with the exact method because of exponentially increased computation time. In this regard, this paper applied a hybrid metaheuristic to solve this problem. The parameter setting for the hybrid metaheuristic is described in Table 2. In Tables 4–6, a deviation is calculated between

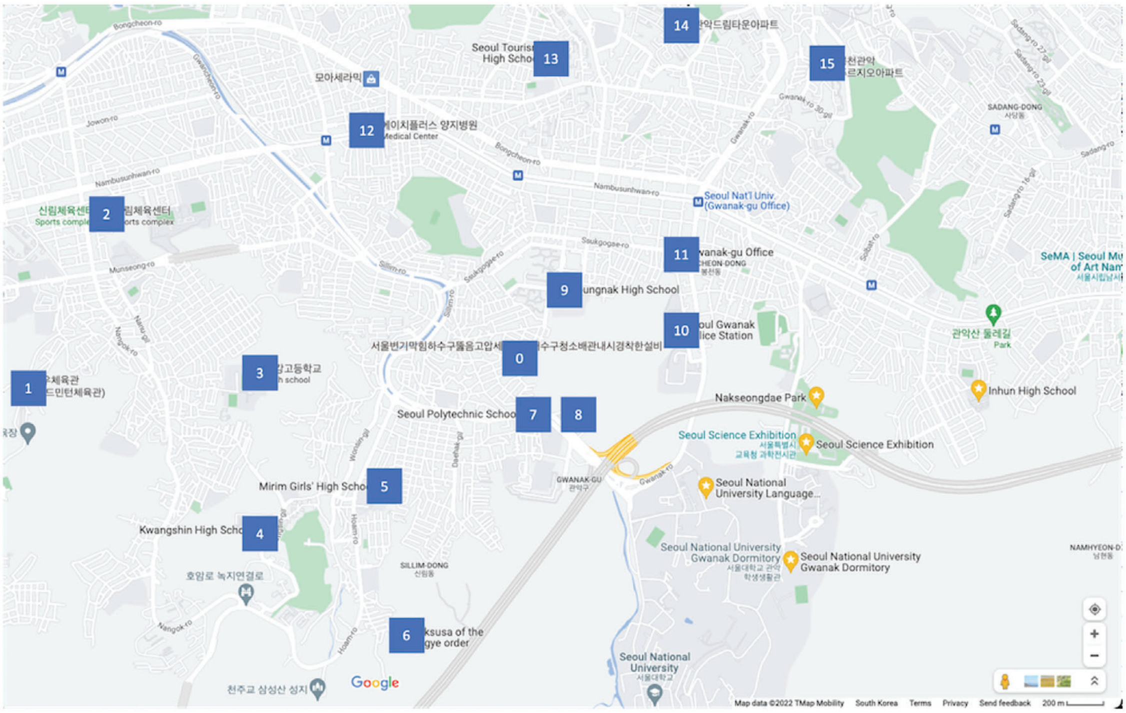

Fig. 4. Gwanak district in Seoul, Korea.

operational management approaches. EBCA stands for estimated-based collection approach, and NBCA stands for neighborhood-based collection approach. In Tables 7 and 8, an error is calculated between the average solution and the best solution.

$$\text{Deviation}(\%) = \frac{\text{NBCA} - \text{EBCA}}{\text{EBCA}} \times 100,$$

$$\text{Error}(\%) = \frac{\text{average solution} - \text{best solution}}{\text{best solution}} \times 100.$$

5.1. Computational results on the small-size problem

An exact method was coded in Python 3.7.9 with CPLEX Optimizer 20.1, and the hybrid metaheuristic was coded in Python 3.6.8 on an Intel Core i7 Macintosh 2.6 GHz. In the small-size problem, the Gwanak district (Fig. 4) in Seoul, Korea, is chosen for the location of IoT-based smart bins. Kim et al. (2020) randomly chose five bins in Seoul. In this experiment, this paper also selected 15 places randomly in Seoul. Each location corresponds to a bin. A depot is set in the middle of these places and denoted as 0. All these places have been taken from Google Maps, and their distances are written in Table 3. For the waste generation, 20 instances are used and each instance corresponds to a different random seed in the Python code. The computation times of the k-means-ACO were calculated as the average values from 10 experiments. In these experiments, the unit for

Table 3
Distance matrix of bins and a depot: Gwanak district, Seoul, Korea

(*i*, *j*)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	2.41	2.16	1.13	1.34	0.69	1.38	0.36	0.58	0.74	1.17	1.40	1.46	1.89	2.31	2.63
1	2.41	0	1.03	1.28	1.53	2.05	2.47	2.77	3.02	2.99	3.58	3.66	2.36	3.42	4.10	4.71
2	2.16	1.03	0	1.13	1.85	2.07	2.77	2.50	2.72	2.48	3.14	3.10	1.50	2.57	3.31	4.01
3	1.13	1.28	1.13	0	0.84	0.92	1.63	1.49	1.75	1.71	2.29	2.40	1.44	2.37	2.98	3.52
4	1.34	1.53	1.85	0.84	0	0.67	9.94	1.56	1.79	2.08	2.49	2.72	2.24	3.02	3.57	3.97
5	0.69	2.05	2.07	0.92	0.67	0	0.78	0.86	1.08	1.45	1.79	2.04	1.97	2.54	3.01	3.31
6	1.38	2.47	2.77	1.63	9.94	0.78	0	1.35	1.50	2.04	2.21	2.54	2.74	3.26	3.65	3.83
7	0.36	2.77	2.50	1.49	1.56	0.86	1.35	0	0.21	0.72	0.93	1.20	1.80	1.99	2.29	2.49
8	0.58	3.02	2.72	1.75	1.79	1.08	1.50	0.21	0	0.65	0.70	1.03	1.85	1.95	2.16	2.32
9	0.74	2.99	2.48	1.71	2.08	1.45	2.04	0.72	0.65	0	0.63	0.68	1.33	1.28	1.58	1.90
10	1.17	3.58	3.14	2.29	2.49	1.79	2.21	0.93	0.70	0.63	0	0.46	2.01	1.66	1.67	1.64
11	1.40	3.66	3.10	2.40	2.72	2.04	2.54	1.20	1.03	0.68	0.46	0	1.80	1.28	1.2	1.25
12	1.46	2.36	1.50	1.44	2.24	1.97	2.74	1.80	1.85	1.33	2.01	1.80	0	1.04	1.78	2.50
13	1.89	3.42	2.57	2.37	3.02	2.54	3.26	1.99	1.95	1.28	1.66	1.28	1.04	0	0.72	1.51
14	2.31	4.10	3.31	2.98	3.57	3.01	3.65	2.29	2.16	1.58	1.67	1.20	1.78	0.72	0	0.81
15	2.63	4.71	4.01	3.52	3.97	3.31	3.83	2.49	2.32	1.90	1.64	1.25	2.50	1.51	0.81	0

cost is a dollar, and the unit for computation time is a second. The distance unit is a kilometer, and the maximum bin capacity is 100 kg. For the penalty cost, this paper imposed penalty cost in proportion to a bin's fill level so as to represent the relative amount of penalty.

In the small-size problem, the following parameters are used. $|K| = 4$, $L = 625$ kg, $T = 7$, $P_i = L_i \times 15\%$, $R = 10$ m. We assumed the unit traveling cost of a vehicle as \$1 so that the distance matrix in Fig. 4 corresponds to the traveling cost in the distance. Table 4 shows the exact method results for operational management approaches. In both collection approaches, total routing costs in the neighborhood-based collection approach were lower than those of the estimation-based collection approach with an average 3.87%. However, there were significant differences in the computation times. In the neighborhood-based collection approach, computation times were almost half on average compared to those in the estimation-based collection approach.

In the estimation-based collection approach, this paper used the must-go bins selection algorithm inspired by Jorge et al. (2022) and it had some limitations. First, because each bin has a different fill rate, there is a possibility of unnecessarily including most of the bins in the must-go bins so that a vehicle visits all bins, thereby increasing computation times. To be specific, suppose a bin is filled up above TFL at the first time period but the fill rate is slow. Then the time when this bin is filled up above TFL again is set far from the first time period so that most of the bins are included in the must-go bins in the first time period. Second, they assumed normal distribution for the fill rate, which is a too naive approach for tackling the uncertainty of fill level information. Therefore, the must-go bins selection algorithm is revised and combined with a hybrid metaheuristic to solve the problem on a large scale.

In the small-size problem, this paper also applied a proposed heuristic to compare the performance with the exact algorithm. Table 5 presents the comparison of operational management approaches. The parameter setting is the same as in the exact algorithm. In both collection approaches, total routing costs in the neighborhood-based collection approach were lower than

Table 4
Comparison of operational management approaches with an exact algorithm in a small size

	Estimation-based collection approach				Neighborhood-based collection approach				Deviation (%)			
Instance	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost	Penalty cost	Total cost	Computation times
Instance 1	947.87	70.20	1018.07	222.97	984.60	70.20	1054.80	174.01	3.87	0.00	3.61	−21.96
Instance 2	1263.83	0.00	1263.83	332.88	1159.37	0.00	1159.37	148.27	−8.27	0.00	−8.27	−55.46
Instance 3	704.33	15.00	719.33	213.47	684.07	15.00	699.07	38.75	−2.88	0.00	−2.82	−81.85
Instance 4	686.41	0.00	686.41	202.68	624.73	0.00	624.73	73.87	−8.99	0.00	−8.99	−63.55
Instance 5	1068.38	45.00	1113.38	249.08	1083.28	45.00	1128.28	90.72	1.39	0.00	1.34	−63.58
Instance 6	1197.18	45.00	1242.18	320.06	1091.43	45.00	1136.43	207.16	−8.83	0.00	−8.51	−35.28
Instance 7	907.70	33.30	941.00	247.92	892.84	32.10	924.94	170.17	−1.64	−3.60	−1.71	−31.36
Instance 8	1190.91	0.00	1190.91	303.34	1100.72	0.00	1100.72	88.53	−7.57	0.00	−7.57	−70.82
Instance 9	943.39	98.55	1041.94	274.62	1036.75	98.85	1135.60	132.05	9.90	0.30	8.99	−51.92
Instance 10	1201.21	0.00	1201.21	302.41	1008.78	0.00	1008.78	116.30	−16.02	0.00	−16.02	−61.54
Instance 11	899.16	112.20	1011.36	243.75	946.94	127.20	1074.14	49.27	5.31	13.37	6.21	−79.79
Instance 12	907.41	45.60	953.01	322.86	846.60	45.60	892.20	37.41	−6.70	0.00	−6.38	−88.41
Instance 13	912.62	101.70	1014.32	277.03	947.87	101.70	1049.57	192.14	3.86	0.00	3.48	−30.64
Instance 14	889.90	32.40	922.30	237.25	782.64	32.40	815.04	38.47	−12.05	0.00	−11.63	−83.79
Instance 15	863.65	34.20	897.85	235.48	846.60	34.20	880.45	110.57	−2.02	0.00	−1.94	−53.05
Instance 16	911.09	31.50	942.59	319.49	932.45	31.50	963.95	218.05	2.34	0.00	2.27	−31.75
Instance 17	1164.76	0.00	1164.76	304.74	1032.78	0.00	1032.78	106.67	−11.33	0.00	−11.33	−65.00
Instance 18	906.17	99.90	1006.07	295.73	785.33	99.90	885.23	70.70	−13.34	0.00	−12.01	−76.09
Instance 19	912.93	116.70	1029.63	262.64	947.87	116.70	1064.57	169.15	3.83	0.00	3.39	−35.60
Instance 20	918.15	15.00	933.15	299.10	828.91	15.00	843.91	79.51	−9.72	0.00	−9.56	−73.42
Average											−3.87	−57.74

Table 5
Comparison of operational management approaches with the proposed heuristic in a small size

	Estimation-based collection approach				Neighborhood-based collection approach				Deviation (%)			
Instance	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost	Penalty cost	Total cost	Computation times
Instance 1	951.60	70.20	1021.80	49.49	988.06	70.20	1058.26	37.32	3.83	0.00	3.57	−24.59
Instance 2	1265.71	0.00	1265.71	64.59	1253.91	0.00	1253.91	61.47	−4.88	0.00	−4.88	−4.83
Instance 3	771.60	15.00	786.60	29.63	689.57	15.00	704.57	29.88	−4.44	0.00	−4.35	0.84
Instance 4	702.75	0.00	702.75	29.05	639.68	0.00	639.68	29.93	−8.98	0.00	−8.98	3.03
Instance 5	1160.65	45.00	1205.65	50.72	1100.28	45.00	1145.28	42.16	−0.93	0.00	−0.90	−16.88
Instance 6	1198.01	45.00	1243.01	54.63	1092.73	45.00	1137.73	52.97	−8.79	0.00	−8.47	−3.04
Instance 7	912.27	33.30	945.57	48.76	935.06	15.00	950.06	52.71	2.50	−54.95	0.47	8.10
Instance 8	1196.28	0.00	1196.28	63.88	1132.27	0.00	1132.27	53.60	−5.35	0.00	−5.35	−16.09
Instance 9	943.96	98.55	1042.51	52.57	1058.85	114.15	1173.00	54.64	12.17	15.83	12.52	3.94
Instance 10	1208.05	0.00	1208.05	61.43	1039.86	0.00	1039.86	43.66	−13.92	0.00	−13.92	−28.93
Instance 11	908.21	112.20	1020.41	45.50	987.20	88.25	1075.45	43.02	8.70	−21.35	5.39	−5.45
Instance 12	911.39	45.60	956.99	43.98	855.39	49.50	904.89	43.88	−6.14	8.55	−5.44	−0.23
Instance 13	918.23	101.70	1019.93	50.49	955.84	94.85	1050.69	41.19	4.10	−6.74	3.02	−18.42
Instance 14	890.23	32.40	922.63	46.14	790.82	48.00	838.82	43.39	−11.17	48.15	−9.08	−5.96
Instance 15	865.25	34.20	899.45	47.21	855.25	30.60	885.85	45.20	−1.16	−10.53	−1.51	−4.26
Instance 16	913.53	31.50	945.03	49.67	948.11	47.10	995.21	44.98	3.78	49.52	5.31	−9.44
Instance 17	1167.25	0.00	1167.25	59.76	1040.13	0.00	1040.13	41.43	−10.89	0.00	−10.89	−30.67
Instance 18	909.06	99.90	1008.96	49.50	822.08	99.90	921.98	45.53	−9.57	0.00	−8.62	−8.02
Instance 19	917.84	116.70	1034.54	48.45	962.68	102.50	1065.18	49.62	4.89	−12.17	−2.96	2.41
Instance 20	920.59	15.00	935.59	47.22	855.10	15.00	870.10	45.82	−7.11	0.00	−7.00	−2.96
Average											−2.81	−8.07

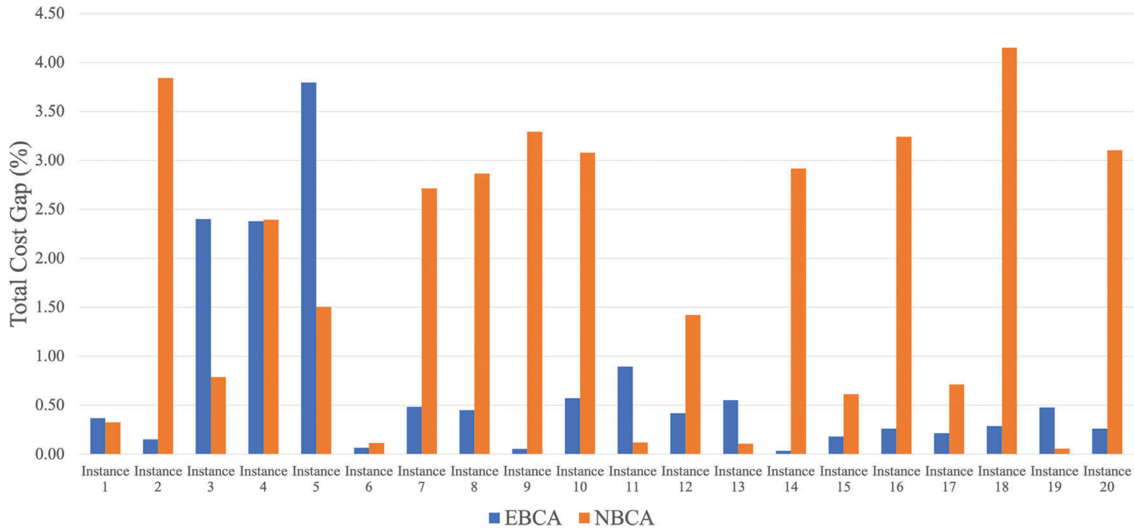

Fig. 5. Performance gaps with respect to operational management approaches.

those of the estimation-based collection approach with an average 2.81%. In addition, computation time is reduced to 8.07% on average. To judge the effectiveness of the heuristic, we calculated the total cost gap between the exact method and the heuristic with operational management approaches in a small size, and the results are described in Fig. 5. BFS stands for the total cost in Table 4, while HEU stands for the total cost in Table 5. Through all instances, these gaps are calculated within 5%. As for computation time, the neighborhood-based collection approach outperforms the estimation-based collection approach. However, as for total cost, the estimation-based collection approach shows stable performances when using the exact method and heuristic.

$$\text{Total cost gap } (\%) = \frac{\text{HEU} - \text{BFS}}{\text{BFS}} \times 100.$$

5.2. *Computational results on the large-size problem*

In the large-size problem, the eil51 instance is used from the TSPLIB for general location information. In this instance, 51 nodes correspond to 51 locations of smart bins, and a depot is set in the middle of the nodes. The following parameters are used for the experiment. $|K| = 5$, $L = 1000$ kg, $T = 7$, $P_i = L_i * 15\%$, $R = 10$ m. This paper assumed the unit traveling cost of a vehicle as \$1 so that the distance matrix in the TSPLIB corresponds to the traveling cost in the distance. We consider eil51, a TSPLIB instance of 51 nodes corresponding to 51 bins. A total of 51 bins were indexed from 1 to 51 and were segmented into five clusters. The number of clusters was selected using the elbow method. In this experiment, total routing costs in the neighborhood-based collection approach were lower than those of the estimation-based collection approach with an average 2.97%. However, there were also significant differences in the computation times. In the neighborhood-based collection approach, computation times were reduced 10.61% on average compared to those

Table 6
Comparison of operational management approaches with the proposed heuristic in a large size

Instance	Estimation-based collection approach				Neighborhood-based collection approach				Deviation (%)			
	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost ($)	Penalty cost ($)	Total cost ($)	Computation times (s)	Routing cost	Penalty cost	Total cost	Computation times
Instance 1	2429.97	15.00	2444.97	228.29	2459.76	15.00	2474.76	169.67	1.23	0.00	1.22	−25.68
Instance 2	2455.41	45.00	2500.41	213.81	2282.37	60.60	2342.97	172.60	−7.05	34.67	−6.30	−19.27
Instance 3	1102.06	15.00	1117.06	124.71	1121.72	15.00	1136.72	103.60	1.78	0.00	1.76	−16.93
Instance 4	1113.96	15.00	1128.96	117.87	998.19	15.00	1013.19	112.60	−10.39	0.00	−10.25	−4.47
Instance 5	2268.96	60.00	2328.96	188.40	2245.60	47.55	2293.15	171.85	−1.03	−20.75	−1.54	−8.78
Instance 6	2176.97	135.00	2311.97	186.76	1805.62	120.60	1926.22	185.18	−17.06	−16.68	−16.68	−0.85
Instance 7	2291.84	15.00	2306.84	191.46	2102.51	46.95	2149.46	159.05	−8.26	213.00	−6.82	−16.93
Instance 8	2402.72	45.00	2447.72	202.83	2201.15	45.00	2246.15	180.09	−8.39	0.00	−8.24	−11.21
Instance 9	1843.00	48.00	1891.00	182.62	1788.20	30.00	1818.20	178.98	−2.97	−37.50	−3.85	−1.99
Instance 10	2357.46	15.00	2372.46	196.63	2316.25	32.55	2348.80	169.77	−1.75	117.00	−1.00	−13.66
Instance 11	2344.09	15.00	2359.09	204.28	2494.84	15.00	2509.84	172.50	6.43	0.00	6.39	−15.56
Instance 12	1754.29	45.00	1799.29	154.49	1743.31	45.00	1788.31	171.86	−0.63	0.00	−0.61	11.24
Instance 13	2365.67	31.95	2397.62	196.48	2242.91	30.60	2273.51	179.35	−5.19	−4.23	−5.18	−8.72
Instance 14	2360.13	15.00	2375.13	197.99	2335.97	30.60	2366.57	172.48	−1.02	104.00	−0.36	−12.88
Instance 15	1620.14	15.00	1635.14	148.42	1642.46	15.00	1657.46	151.77	1.38	0.00	1.36	2.26
Instance 16	2299.61	15.00	2314.61	189.05	2202.78	15.00	2217.78	164.29	−4.21	0.00	−4.18	−13.10
Instance 17	2272.27	0.00	2272.27	191.19	2165.50	0.00	2165.50	160.18	−4.70	0.00	−4.70	−16.22
Instance 18	2255.77	15.00	2270.77	197.90	2267.08	15.00	2282.08	164.59	0.50	0.00	0.50	−16.83
Instance 19	2297.73	15.00	2312.73	194.31	2358.98	15.00	2373.98	164.96	2.67	0.00	2.65	−15.10
Instance 20	2296.59	15.00	2311.59	199.18	2215.36	15.00	2230.36	184.06	−3.54	0.00	−3.51	−7.59
Average											−2.97	−10.61

Table 7
Dispersion results of k-means-ACO for different TSPLIB instances

Instance	Clusters	Best	Worst	Average	SD	Error
bayg29	2	2022	2188	2086.7	56.90	3.20
	3	2049	2275	2189	80.16	6.83
	4	2060	2196	2131.5	44.60	3.47
	5	2135	2329	2239.6	55.61	4.90
dantzig42	2	714	851	792.6	40.58	11.01
	3	711	847	793.3	41.69	11.58
	4	741	817	773.8	20.34	4.43
	5	841	898	861.1	16.37	2.39
eil51	2	615	694	649.5	27.90	5.61
	3	782	846	809.4	23.14	3.50
	4	881	909	896.6	7.88	1.77
	5	1019	1070	1046.2	17.38	2.67
berlin52	2	7846	8416	8197.5	212.92	4.48
	3	8150	8668	8464.3	166.44	3.86
	4	8401	8963	8766.7	156.12	4.35
	5	8377	9570	9024.7	497.06	7.73
st70	2	1058	1181	1143.7	39.40	8.10
	3	1252	1439	1368.4	57.69	9.30
	4	1597	1702	1653.2	40.30	3.52
	5	1719	1780	1756.4	19.20	2.18
rat99	2	2068	2353	2172.3	96.77	5.04
	3	2077	2392	2285.7	111.26	10.05
	4	2192	2456	2315.7	91.62	5.64
	5	2345	2648	2526.2	93.69	7.73
eil101	2	976	1088	1035.9	38.39	6.14
	3	1158	1342	1234.5	47.31	6.61
	4	1382	1491	1439.3	31.20	4.15
	5	1519	1659	1572.8	47.56	3.54

in the estimation-based collection approach. Table 6 shows results for operational management approaches with the proposed hybrid metaheuristic in a large size.

5.3. Performance test of k-means-ACO

The performance of the proposed hybrid algorithm was statistically tested by running it 10 times. An average value, standard deviation (SD), and error are calculated according to the optimal solution against seven TSPLIB instances. To judge the effectiveness of the clustering algorithm, different numbers of clusters are provided. The results are given in Table 7. To compare the dependency of the clustering solution with other existing algorithms, an advanced GA is selected. The parameter setting for GA is exactly the same from the paper. The modified GA is the combination of roulette wheel selection, comparison crossover, and random mutation proposed by Maity et al. (2016). This GA is combined with the k-means algorithm to compare with the k-means-ACO. The results are presented in Table 8.

Table 8
Performance test of k-means-ACO and k-means-GA on standard TSPLIB instances

			Total cost($)								Computation time(s)							
			Average		Best		SD		Error(%)		Average		Best		SD		Error(%)	
Instance	Nodes	Clusters	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA	k-means-ACO	k-means-GA
bayg29	29	3	1846.5	1676.6	1646	1628	96.81	53.57	12.18	2.99	9.66	16.74	8.98	15.67	0.77	0.79	7.53	6.83
bays29	29	3	2252.7	2167.7	2033	2022	121.68	201.71	10.81	7.21	9.75	17.77	9.17	15.63	0.77	3.35	6.25	13.68
dantzig42	42	2	818.6	795.4	765	763	35.31	32.75	7.01	4.25	13.46	37.23	12.81	35.63	0.39	1.25	5.06	4.48
eil51	51	5	1027.8	956.8	982	894	35.93	54.93	4.66	7.02	15.16	44.01	14.87	42.02	0.23	1.93	1.97	4.75
berlin52	52	5	8597.8	9061.9	8267	8306	162.24	544.83	4.00	9.10	15.54	45.50	14.77	43.27	0.89	2.66	5.20	5.16
st70	70	4	1659.1	1544.1	1619	1389	29.49	96.85	2.48	11.17	20.82	60.76	19.64	57.60	0.97	2.87	6.00	5.50
eil76	76	4	1201.6	1134.6	1146	1081	39.35	46.11	4.85	4.96	22.33	67.14	21.52	64.47	1.03	2.01	3.77	4.14
pr76	76	5	138434.7	138493.7	123620	131538	8535.55	5005.62	11.98	5.29	21.72	67.85	21.06	64.31	0.52	1.76	3.15	5.50
rat99	99	3	2229.8	2584.7	2160	2369	40.74	162.24	3.23	9.11	45.46	113.59	44.63	92.11	1.03	26.76	1.87	23.32
eil101	101	4	1332.6	1415.1	1282	1325	33.38	61.91	3.95	6.80	49.58	101.30	46.04	93.76	3.80	8.34	7.68	8.04
pr107	107	4	47638.6	53725.9	45243	47540	1656.90	5115.19	5.29	13.01	65.95	113.65	64.59	98.51	1.41	12.26	2.11	15.37
pr124	124	4	70681.7	85259	67418	77200	2136.93	5414.06	4.84	10.44	76.06	135.20	75.28	122.38	0.54	10.82	1.03	10.47
pr136	136	4	120233.3	144288.5	115226	128099	3332.05	8293.75	4.35	12.64	84.35	159.94	82.70	140.00	1.81	19.12	2.00	14.24
pr144	144	4	80285.5	101878.6	75160	89002	4370.23	8710.08	6.82	14.47	90.98	185.87	86.99	162.14	2.59	23.93	4.59	14.64
kroA150	150	3	63218.7	83864.8	59283	78461	2896.72	2691.36	6.64	6.89	181.06	229.83	167.37	212.42	36.04	17.38	8.18	8.20
kroB150	150	3	60198.4	81029.8	54562	78160	3095.18	2463.38	10.33	3.67	160.77	249.30	157.41	214.25	3.52	33.72	2.14	16.36
pr152	152	4	87998.8	110102.3	81851	88351	3280.86	15575.98	7.51	24.62	98.20	184.31	97.13	168.95	0.77	17.32	1.10	9.09
u159	159	5	54682.9	68031.9	51632	62686	2094.34	3064.84	5.91	8.53	166.74	186.80	160.36	162.26	7.27	17.14	3.97	15.13
rat195	195	4	4448.3	5979.9	4233	5284	126.72	342.42	5.09	13.17	204.25	311.82	175.01	275.05	14.28	29.03	16.71	13.37
tsp225	225	5	5045.3	6816.8	4649	6417	196.00	275.41	8.52	6.23	234.74	332.42	232.73	291.29	16.92	37.60	7.74	14.12

6. Conclusions

6.1. Managerial insights

This work consists of three research questions: *Why are clusters needed for waste collection?*, *Why is optimization difficult, and how might we best suggest a heuristic?*, and *What are the contributions of this research?'* First and foremost, vehicle routing can be carried out on a cluster basis instead of on a bin basis by grouping bins into clusters, which reduces the complexity of the VRP by orders of magnitude. However, applying only the k-means clustering algorithm has some limitations in that this algorithm is based on only the location between nodes. Therefore, this paper introduced sequential clustering using filled bin sets and their neighboring sets to reduce the subset that a vehicle will visit. Second, it seemed possible to get an optimal solution to this problem. According to Ramos et al. (2018) and Jorge et al. (2022), it took a long computation time to get an optimal solution, which is far from a practical point of view. However, the proposed hybrid heuristic suggests a good solution in a short time. Third, this paper first proposed the neighboring bin concept considering the radius limitation of IoT sensors and this concept is implemented in a mathematical model. In addition, this paper presented two operational management approaches for smart waste collection. One of them is based on fill level estimation and the other one is based on neighborhood. The first one showed good performance as for dynamic fill level information, while the second one presented fast solutions as for computation times and considered practical constraints.

6.2. Future study

This paper represents a vehicle routing problem for smart waste management. The above system is configured with IoT-based smart bins. However, this paper has some limitations. First, we consider only a single type of waste for simplicity. This may be extended to several types of waste within the municipal waste system (e.g., food waste, recyclable waste, biodegradable waste). Second, multiple compartments in a vehicle will facilitate collection of several types of waste. Furthermore, special types of waste could be considered, such as medical waste, chemical waste, and industrial waste, to broaden .CluVRP-WC model. The best waste management system has two aspects. One is efficient waste collection and savings in transportation costs. The other is maximizing utilities for reusing waste. Proper waste management is necessary to build a pollution-free smart city.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning under Grant No. NRF-2019R1A2C2084616; and the Brain Pool Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning under Grant No. NRF-2020H1D3A2A01085443.

We thank the associate editor and the two anonymous referees for their helpful comments.

References

Adeleke, O.J., Ali, M.M., 2021. An efficient model for locating solid waste collection sites in urban residential areas. *International Journal of Production Research* 59, 3, 798–812.

Al-Refaie, A., Al-Hawadi, A., Fraij, S., 2021. Optimization models for clustering of solid waste collection process. *Engineering Optimization* 53, 12, 2056–2069.

Atzori, L., Iera, A., Morabito, G., 2017. Understanding the internet of things: definition, potentials, and societal role of a fast evolving paradigm. *Ad Hoc Networks* 56, 122–140.

Battarra, M., Erdoğan, G., Vigo, D., 2014. Exact algorithms for the clustered vehicle routing problem. *Operations Research* 62, 1, 58–71.

Coban, A., Ertis, I.F., Cavdaroglu, N.A., 2018. Municipal solid waste management via multi-criteria decision making methods: a case study in Istanbul, Turkey. *Journal of Cleaner Production* 180, 159–167.

Defryn, C., Sörensen, K., 2017. A fast two-level variable neighborhood search for the clustered vehicle routing problem. *Computers & Operations Research* 83, 78–94.

Defryn, C., Sörensen, K., Cornelissens, T., 2016. The selective vehicle routing problem in a collaborative environment. *European Journal of Operational Research* 250, 2, 400–411.

Del Pia, A., Filippi, C., 2006. A variable neighborhood descent algorithm for a real waste collection problem with mobile depots. *International Transactions in Operational Research* 13, 2, 125–141.

Delgado-Antequera, L., Caballero, R., Sánchez-Oro, J., Colmenar, J.M., Martí, R., 2020. Iterated greedy with variable neighborhood search for a multiobjective waste collection problem. *Expert Systems with Applications* 145, 113101.

Eryganov, I., Šomplák, R., Nevrlỳ, V., Smejkalová, V., Hrabec, D., Haugen, K.K., 2020. Application of cooperative game theory in waste management. *Chemical Engineering Transactions*, 81, 877–882.

Expósito-Izquierdo, C., Rossi, A., Sevaux, M., 2016. A two-level solution approach to solve the clustered capacitated vehicle routing problem. *Computers & Industrial Engineering* 91, 274–289.

Fayoumi, A., Loucopoulos, P., 2016. Conceptual modeling for the design of intelligent and emergent information systems. *Expert Systems with Applications* 59, 174–194.

Fernández, E., Roca-Riu, M., Speranza, M.G., 2018. The shared customer collaboration vehicle routing problem. *European Journal of Operational Research* 265, 3, 1078–1093.

Ferrer, J., Alba, E., 2019. Bin-ct: urban waste collection based on predicting the container fill level. *Biosystems* 186, 103962.

Ghiani, G., Guerriero, F., Improta, G., Musmanno, R., 2005. Waste collection in southern Italy: solution of a real-life arc routing problem. *International Transactions in Operational Research* 12, 2, 135–144.

Haque, K.F., Zabin, R., Yelamarthi, K., Yanambaka, P., Abdelgawad, A., 2020. An IoT based efficient waste collection system with smart bins. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*. IEEE, Piscataway, NJ, pp. 1–5.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28, 1, 100–108.

Hintsch, T., Irnich, S., 2018. Large multiple neighborhood search for the clustered vehicle-routing problem. *European Journal of Operational Research* 270, 1, 118–131.

Huang, S.H., Lin, P.C., 2015. Vehicle routing–scheduling for municipal waste collection system under the keep trash off the ground policy. *Omega* 55, 24–37.

Jorge, D., Antunes, A.P., Ramos, T.R.P., Barbosa-Póvoa, A.P., 2022. A hybrid metaheuristic for smart waste collection problems with workload concerns. *Computers & Operations Research* 137, 105518.

Kim, C., Yang, R., Lee, J., Jang, H., Ha, S., Kim, W., 2020. A study to find the optimal waste collection path. In *Proceedings of the Korea Information Processing Society Conference*, Korea Information Processing Society, Seoul, pp. 216–219.

Kim, J.S., 2015. The introduction of IoT (internet of things) technologies and policy directions. *Review of Korea Contents Association* 13, 1, 18–24.

Lin, Z., Xie, Q., Feng, Y., Zhang, P., Yao, P., 2020. Towards a robust facility location model for construction and demolition waste transfer stations under uncertain environment: The case of Chongqing. *Waste Management* 105, 73–83.

López-Sánchez, A., Hernández-Díaz, A.G., Gortázar, F., Hinojosa, M.A., 2018. A multiobjective GRASP–VND algorithm to solve the waste collection problem. *International Transactions in Operational Research* 25, 2, 545–567.

Maity, S., Roy, A., Maiti, M., 2016. An imprecise multi-objective genetic algorithm for uncertain constrained multi-objective solid travelling salesman problem. *Expert Systems with Applications* 46, 196–223.

Peng, Y., Wu, P., Schartup, A.T., Zhang, Y., 2021. Plastic waste release caused by covid-19 and its fate in the global ocean. *Proceedings of the National Academy of Sciences* 118, 47.

Pop, P.C., Kara, I., Marc, A.H., 2012. New mathematical models of the generalized vehicle routing problem and extensions. *Applied Mathematical Modelling* 36, 1, 97–107.

Ramos, T.R.P., de Morais, C.S., Barbosa-Póvoa, A.P., 2018. The smart waste collection routing problem: Alternative operational management approaches. *Expert Systems with Applications* 103, 146–158.

Reinelt, G., 1991. Tspliba traveling salesman problem library. *ORSA Journal on Computing* 3, 4, 376–384.

Ríos-Mercado, R.Z., González-Velarde, J.L., Maldonado-Flores, J.R., 2023. Tabu search with strategic oscillation for improving collection assignment plans of waste electric and electronic equipment. *International Transactions in Operational Research* 30, 1002–1030.

Roy, A., Manna, A., Kim, J., Moon, I., 2022. IoT-based smart bin allocation and vehicle routing in solid waste management: a case study in South Korea. *Computers & Industrial Engineering* 171, 108457.

Sahebjamnia, N., Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., 2018. Sustainable tire closed-loop supply chain network design: hybrid metaheuristic algorithms for large-scale networks. *Journal of Cleaner Production* 196, 273–296.

Salamirad, A., Kheybari, S., Ishizaka, A., Farazmand, H., 2023. Wastewater treatment technology selection using a hybrid multicriteria decision-making method. *International Transactions in Operational Research* 30, 1479–1504.

Sarkar, B., Debnath, A., Chiu, A.S., Ahmed, W., 2022a. Circular economy-driven two-stage supply chain management for nullifying waste. *Journal of Cleaner Production* 339, 130513.

Sarkar, B., Ganguly, B., Pareek, S., Cárdenas-Barrón, L.E., 2022b. A three-echelon green supply chain management for biodegradable products with three transportation modes. *Computers & Industrial Engineering* 174, 108727.

Sevaux, M., Sörensen, K., et al.2008. Hamiltonian paths in large clustered routing problems. In *Proceedings of the EU/MEeting 2008 Workshop on Metaheuristics for Logistics and Vehicle Routing, EU/ME*, Vol. 8, pp. 411–417.

Tan, P.N., Steinbach, M., Kumar, V., 2016. *Introduction to Data Mining*. Pearson Education India, New Delhi.

Toth, P., Vigo, D., 2002. The Vehicle Routing Problem. *Monographs on Discrete Mathematics and Applications*. SIAM, Philadelphia, PA.

Zhao, J., Zhu, F., 2016. A multi-depot vehicle-routing model for the explosive waste recycling. *International Journal of Production Research* 54, 2, 550–563.