



Balancing the cardinality of clusters with a distance constraint: a fast algorithm

Herminia I. Calvete¹ · Carmen Galé² · José A. Iranzo³

Received: 30 May 2023 / Accepted: 17 April 2024
© The Author(s) 2024

Abstract

Logistics companies partition the customers they serve into delivery zones as a tactical decision and manage the customers assigned to each zone as a cluster for the purpose of routing, workload allocation, etc. Frequently, this partition is made in accordance with customers' geographical location, which can result in very unbalanced clusters in terms of the number of customers they include. In addition, in the day-to-day operations, not necessarily all customers need to be served every day so, even if the clusters originally created are balanced, daily needs may lead to unbalanced clusters. Given an a priori assignment of customers to clusters, improving the balance between clusters in advance of workload management is therefore a key issue. This paper addresses the problem of balancing clusters, when there is a distance constraint that prevents reassigning customers to clusters far away from their original pre-assignment. This problem is formulated as a lexicographic biobjective optimization model. The highest priority objective function minimizes the variance of the number of customers in the clusters. The second ranked objective function minimizes the total distance resulting from all reassignments. A fast and effective heuristic algorithm is developed, based on exploring customer reassignments, either by comparing clusters two by two or by extending the search to allow for sequential customer swaps among clusters. Both the quality of the solution and the computational time required encourage the use of this algorithm by logistics companies to balance clusters in real scenarios.

Keywords Balancing clusters · Distance constraint · Heuristic algorithm

All of the authors have contributed equally to this work.

✉ Herminia I. Calvete
herminia@unizar.es

Carmen Galé
cgale@unizar.es

José A. Iranzo
joseani@unizar.es

¹ Departamento de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA), Universidad de Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain

² Departamento de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA), Universidad de Zaragoza, María de Luna 3, 50018 Zaragoza, Spain

³ Departamento de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA), Universidad de Zaragoza, Violante de Hungría 23, 50009 Zaragoza, Spain

1 Introduction

Grouping users of a service (from now on customers) in geographical areas (districting) is a common practice in different areas. Kalcsics and Ríos-Mercado (2019) state that there are four major areas of application: political districting, sales territory design, service districting and distribution districting, and provides relevant literature on these topics. District mapping is a common strategy adopted by logistics companies to better manage their resources and workload. To partition the distribution area generally allows each zone, subarea or cluster of customers to be regularly served by the same set of drivers, which improves customer service and reduces routing complexity by handling smaller sets of customers (Vidal et al, 2020). As a strategic decision, clusters of customers are created a priori, usually based on their geographical location (zip code zones). This form of grouping customers can result in clusters that are very unbalanced in terms of the number of customers they include. In addition, in the day-to-day operations of companies, not necessarily all customers need to be served every day so, even if the clusters originally created are balanced, daily needs may lead to unbalanced clusters. This is of concern to logistics companies, for whom a balanced workload is an important issue (Matl et al, 2018, 2019).

This paper addresses the problem of constructing new clusters, by removing some customers from their pre-assigned cluster and reassigning them to a different one, so that the number of customers in the new clusters be as balanced as possible. In practical applications like logistics and supply chain management, balanced clusters help optimize resource utilization, streamline operations, and ultimately reduce costs. This problem was brought to our attention by Alerce <https://www.alerce-group.com/>, a consultancy company which provides services to the logistics industry. The aim was to make changes in such a way that the allocation of the workload or the calculation of routes would not be impacted significantly. Broadly speaking, these changes should affect customers who are close to the zone to whose cluster they are to be reassigned. In addition, this reassignment should be done in a very short computing time so that it can be used when the daily workload is planned. Hence, this work focuses primarily on the strategic aspect of balancing customer clusters, implicitly assuming that the logistics company has sufficient resources to accommodate the reallocation of customers. By assuming adequate resources, we aim to isolate the impact of customer clustering strategies on workload distribution and operational efficiency.

Constructing clusters with the same or nearly the same number of customers can be easily solved by assigning to each cluster a number of customers equal to (or close to) the average. However, from a company operations point of view, it may be impractical to group geographically distant customers into the same cluster. Therefore, it is necessary to take into account a distance constraint which prevents the reassignment of customers to clusters in a distant zone. It is this distance constraint that makes it difficult to deal with cluster balancing as it may be impossible to assign almost the same number of customers to all clusters. From now on, the above problem will be called the Balancing customer Clusters with a Distance constraint Problem (BCDP). It is worth emphasizing that there is no general agreement on how to address what is meant by balancing. Depending on the area of study and on the objective looked for, there have been different manners of approaching the balancing problems. This issue can be considered close to that of equity measurement, which has been the subject of multiple economic studies for a long time (Atkinson, 1970; Young, 1994; Sen and Foster, 1997). Karsu and Morton (2015) point out that incorporating equitability in the decision process depends on the structure of the problem as well as on what is understood by a fair distribution. Taking into account the BCDP characteristics, in this paper we propose

to measure the imbalance by the variance of the number of customers in the clusters. Since the distance constraint does not, in general, allow clusters with a cardinal equal to the mean number of customers, the variance provides a measure of the degree of spread around the mean. Moreover, since it is also important to make changes that have the least possible impact on the company's initial organization, a goal related to the total distance resulting from customers' reassignments will be considered.

Both issues, balancing the number of customers and handling the distance, are modeled as a biobjective mixed integer optimization problem. In general, when dealing with multiobjective optimization, there is no single feasible solution which minimizes both objective functions at the same time. Hence, it has been studied from different points of view (Ehrgott, 2005). When an order among the objectives can be established, the lexicographic approach is appropriate since it allows determining an optimal solution by applying the lexicographic order induced among the feasible solutions (Romero, 2001). This approach assigns pre-emptive priorities to the objective functions in order to minimize them in a lexicographic order. Thus, first the highest priority objective function, related to balancing, is minimized over the feasible region defined by the constraints. Then, from the set of optimal solutions to this single-objective problem, an optimal solution which minimizes the second objective function, referring to distance, is selected.

The paper describes the development of a specially tailored heuristic for solving the BCDP that is accurate and fast. This heuristic consists of four steps. The first and the second steps aim to obtain balanced clusters by exploring the existence of customers which can be reassigned to reduce the imbalance. This is made by comparing either clusters by pairs or addressing sequential customer exchanges. From the solution obtained at the end of these steps, and without modifying its highest priority objective function value, the third and fourth steps of the algorithm aim to explore if a better reassignment is possible from the point of view of the total distance of reassignment. Compared to a general purpose solver, this algorithm provides outstanding results in very short computing times. Thus, it can be successfully implemented in real-life scenarios and so this methodology can help logistic companies when the daily workload is known and balancing is needed.

In summary, the main contributions of this paper are the following:

- We introduce a new problem, the BCDP.
- We formally model the BCDP as a lexicographic biobjective nonlinear integer optimization problem.
- We propose a heuristic algorithm based on reassigning customers among clusters.
- We perform extensive computational experiments on instances derived from real-world data that show that the algorithm provides high-quality solutions in short computing times.

The paper is organized as follows. Following the Introduction section, Sect. 2 summarizes related work focusing on areas in which districting or balancing are relevant topics. Section 3 establishes the mathematical model for the BCDP. Section 4 describes the characteristics of the heuristic algorithm developed for solving the BCDP. In Sect. 5, we analyze the performance of the algorithm by comparing the results provided with a general purpose solver, using a set of large instances based on real-world data. Finally, in Sect. 6 we summarize the key insights and present some concluding remarks and future work.

2 Related work

In this section, without being exhaustive, we present some relevant papers in different fields in which balancing plays an important role. These papers are relevant in their own right, but also because they provide a number of references that we do not reproduce here for the sake of conciseness. Balancing problems have been studied mainly in connection with district design, clustering, routing or facility location problems.

2.1 Districting

Related to districting problems, Kalcsics and Ríos-Mercado (2019) provide a state-of-the-art review in different areas of application. They also describe the problems associated with quantifying different characteristics that are desirable, because of their different focus depending on the area in which the problem is being formulated. Tasnádi (2011) and Goderbauer and Winandy (2018) surveys focus on political districting to avoid unequal representation of citizens. Liberatore et al (2020) and Samanta et al (2022) refer to the police districting problem in which the purpose is to improve the ability of police agencies to stop and prevent crime. Benzarti et al (2013) analyze the role of districting in home health care services. The criteria considered are compactness, accessibility, conformity to administrative boundaries, indivisibility and workload balance. Sandoval et al (2022) address the problem of delivery district design in postal and last mile logistics from a real-world problem. The model is based on the classic district design problem but focuses on the quality of service rather than on workload balancing. Based on a real-world problem of a major dairy company, Zhou et al (2021) develop a heuristic technique for dealing with districting, aiming to minimize the total operational cost computed as a function of the fixed costs of the districts and the routing costs. Finally, the following papers take into consideration the existence of some random component. Diglio et al (2020) study a stochastic districting problem in which demand is assumed to be random. The focus is on overall compactness while meeting balancing constraints defined as average demand per district. Haugland et al (2007) focus on designing districts for vehicle routing problems with stochastic demands. The purpose is to partition a set of customers in contiguous districts in such a way that customers in the same districts are served by the same vehicle. In the proposed approach, first customers are allocated to districts and this allocation remains fixed whatever the demand pattern that occurs. In the second stage, vehicle routing problems (VRP) associated to each district are solved.

2.2 Clustering

Clustering is a methodology used in unsupervised learning to categorize data points based on their similarity and differences. Widely utilized in exploratory data analysis, it holds significant importance across various applications, including pattern recognition and market and customer segmentation, among others. The k -means algorithm is a widely used method for clustering data points into distinct groups, or clusters, based on similarity. The algorithm iteratively partitions the data into k clusters by minimizing the sum of squared distances between data points and the centroid of their assigned cluster. The algorithm k -means is computationally efficient and easy to implement, making it a popular choice for clustering tasks in various domains. However, it is sensitive to the initial placement of centroids and may converge to local optima, requiring multiple runs with different initializations to obtain the

optimal clustering solution. Moreover, in some cases, it can lead to clusters of very different sizes.

Ensuring balanced clusters holds significant importance across various fields and applications. Balanced clusters facilitate more effective decision-making processes, enhance the interpretability of results, and enable fair resource allocation. In fields such as machine learning and data analysis, balanced clusters contribute to the accuracy and robustness of models, leading to improved predictive performance. Additionally, in social sciences and market research, balanced clusters ensure representative samples, leading to more accurate insights and informed decision-making. Therefore, various algorithms have been presented in the literature to construct balanced clusters. These algorithms aim to address the challenge of creating clusters with roughly equal numbers of data points. Malinen and Fränti (2014) propose what they call a balanced k -means which differs from the k -means in the assignment phase which forces all clusters to be of the same size. Lin et al (2019) address the problem of balancing clusters using a regularization term which acts on the cluster sizes to control their balance. This regularization technique is similar to that used in machine learning models to prevent overfitting. De Maeyer et al (2023) offer a comprehensive review of existing approaches to balanced clustering and introduce an alternative method based on the k -means algorithm. Central to this approach is the inclusion of an escalating penalty term into the assignment function of the k -means algorithm. This augmented function encourages the assignment of objects to smaller clusters, thus promoting a more balanced distribution. Ding (2020) assumes that the k cluster centers are provided and proceeds to construct a spatial partition framework which serves as the foundation for developing an algorithm tailored to the k -means clustering problem, wherein the size of each cluster is constrained by predetermined lower and upper bounds

2.3 Balancing in VRP and location problems

Vidal et al (2020) provide a recent review of existing and emerging variants of VRP in which balancing is one of the emerging objectives, measured as workload balance, service equity or collaborative planning. VRP with route balancing aims to preserve equity among drivers through a good balance of their workload. This problem is usually modeled as a biobjective problem in which both the distance traveled and a measure of the workload imbalance are minimized. As mentioned by Matl et al (2018), who provide an extensive survey on workload equity in VRP, most models and methods proposed in the literature measure workload by the route distance and aim to minimize the longest route length (Golden et al, 1997; Jozefowicz et al, 2009) or the difference between the longest route length and the shortest route length (Lacomme et al, 2015). They conclude by emphasizing the importance of properly selecting what is to be balanced as this determines the solutions that will be obtained. Lee and Ueng (1999) minimize the distance traveled and what they call the best working time balance defined as the sum of the working time difference between each vehicle and the vehicle with the shortest working time. Jozefowicz et al (2007) approach the balancing problem from a biobjective point of view considering the distance traveled by the vehicles and the difference between the longest route length and the shortest route length as minimization objectives. Matl et al (2019) supplement their previous work. They analyze, among other aspects, how the obtained solutions represent the preference structure of decision makers or what is the degree of agreement on what is meant by well-balanced VRP solutions. Halvorsen-Weare and Savelsbergh (2016) carry out an exhaustive review of papers dealing with workload balancing in VRP. Moreover, they highlight how the selection of the objective functions affects

the resulting Pareto front. Lehuédé et al (2020) propose to refine the min-max approach in the sense that, when a minimal value has been found for the longest route, the second longest route is considered, then the third longest route, and so on, until all ties have been broken, and develop a multi-directional local search approach. Nikolakopoulou et al (2004) aim to balance the time utilization of the vehicles used by partitioning the distribution network into subnetworks each served from a single depot. Bektaş et al (2019) study the Balanced Vehicle Routing Problem, where each route is required to visit a maximum and a minimum number of customers. They describe several families of facet-inducing inequalities which are used in a branch-and-cut algorithm. Mancini et al (2021) focus on carriers which collaborate to better use the available resources and introduce the collaborative VRP with time and service consistency and workload balance for which they develop an efficient matheuristic. Linfati et al (2020) indicate that traditional approaches to measuring the workload balance have some problems and so propose considering customers' compactness and visual attractiveness by using several objective functions.

The literature on facility location problems has also devoted attention to the issues of equity and fairness, aiming to measure to what extent facilities cover homogeneously the people they serve (see, for instance, Marsh and Schilling (1994); Eiselt and Laporte (1995); Bélanger et al (2019) and references therein). In general, these papers deal with equity from the customer's point of view, since it is important that the coverage of customers by open facilities has a similar quality. Ogryczak (2000) analyzes efficiency and inequality measures through the properties of solutions to location models. Marín (2011) considers a discrete facility location problem where the difference between the maximum and minimum number of customers allocated to every plant has to be minimized. Two different formulations of the problem are proposed and valid inequalities are developed. Related to humanitarian logistics, Liu et al (2021) focus on the location problem in the preparedness phase of disasters, using distributionally robust chance constraints to characterize uncertainties.

2.4 Reassigning customers in VRP

The problem of reassigning customers has been addressed in the literature relating to the well-known cluster first - route second approach for the VRP. In this approach, logistics companies establish delivery zones as a tactical decision. To partition the distribution area generally allows each zone, subarea or cluster of customers to be regularly served by the same set of drivers, which improves customer service and reduces routing complexity by handling smaller sets of customers. In this regard, some studies have considered the possibility of allowing reassignments among zones in the process of computing the routes. Wong and Beasley (1984) propose a routing strategy based upon the division of the depot area into subareas where a single vehicle is assigned to each subarea and develop a heuristic algorithm which generates an initial partition and, by exchanging customers between subareas, attempts to minimize the cost of the partition. Janssens et al (2015) assume a given partition of the distribution region into smaller microzones that are assigned to a preferred vehicle. Then, they develop a metaheuristic to address the problem of reassigning the microzones to vehicles, aiming to balance the workload of the different vehicles while the total distance traveled is minimized. Schneider et al (2015) realize that pre-assigning customers to drivers improves service consistency but worsens route efficiency as measured by total distance traveled, especially when time windows constraints are present. Hence, they develop a procedure that selects a set of seed customers from which they generate the service territories which are served by specific drivers by adding a predefined percentage of customers. Then, the

daily routes are designed based on these service territories. From a real-world problem arising in parcel delivery, Bender et al (2020) propose a two-stage solution approach which establishes delivery districts in the first stage and computes routes adapted to the daily demand realizations in the second stage, allowing for some reassignments of basic areas. They aim to establish a compromise between service consistency and daily demand fluctuations. They also present a complete review of districting problems for VRP with demand uncertainty. Note that the aforementioned papers address customer reassignment in the route construction process, but they do not have in mind balancing customer assignment across zones.

3 Problem formulation

In this section, the BCDP is described and the mathematical model is formulated. We assume a logistics company with a portfolio of customers that have been grouped, generally based on geographic proximity, into a set of clusters. Every day, the company serves those customers who request its services, organizing the workload or the distribution routes in each cluster. The clusters obtained as a result of the customers to be served each day can be very different in terms of the number of customers they group together, which can lead to an undesirable imbalance. The goal is to reassign the customers served each day, henceforth referred to as customers, to provide balanced new clusters. On the other hand, it is often impossible to reassign a customer to a cluster that is far away from their current one. Therefore, we also assume the existence of a measure indicating the distance between each customer and every cluster, along with a designated maximum distance that allows for reassignment. When a customer is pre-assigned to a cluster, their distance to that cluster is zero. Given that the distance constraint typically prevents clusters from having a cardinality equal to the mean number of customers, we assume the variance of the number of customers in the clusters as the measure of the imbalance. In addition, if there exist several optimal solutions with respect to the imbalance objective, one with the least total distance should be preferred. The model is particularly well suited for addressing the challenges encountered by logistics companies involved in the distribution of small parcels, particularly within densely populated urban areas. In such scenarios, where a considerable volume of customers must be served daily from one or more warehouses, the model offers the possibility of achieving clusters as balanced as possible, taking into account the distance constraint. This strategic approach may enable them, for instance, to optimize service delivery by assigning a single driver, thereby enhancing customer satisfaction and streamlining routing complexities. It is worth noting that while the model is tailored for these specific logistics scenarios, its adaptable nature allows for potential application in a broader range of contexts.

3.1 The mathematical model

The following notations are used to formulate the BCDP:

Indexes

- i Index of customer
- j, k Index of cluster

Parameters

- n Number of customers
- m Number of clusters
- d_{ij} Distance from customer i to cluster j

D Maximum distance to allow a reassignment

Sets:

I Set of customers, $I = \{1, \dots, n\}$

J Set of clusters, $J = \{1, \dots, m\}$

Variables:

x_{ij} If the customer i is reassigned to the cluster j , $x_{ij} = 1$; otherwise $x_{ij} = 0$

The BCDP can be formulated as the following lexicographic biobjective nonlinear integer optimization problem:

$$\text{lex min} \quad \left(Z_1 = \frac{1}{m} \sum_{j \in J} \left(\sum_{i \in I} x_{ij} - \frac{1}{m} \sum_{k \in J} \sum_{i \in I} x_{ik} \right)^2, \quad Z_2 = \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij} \right) \quad (1)$$

subject to

$$\sum_{j \in J} x_{ij} = 1, \quad i \in I \quad (2)$$

$$d_{ij} x_{ij} \leq D, \quad i \in I, j \in J \quad (3)$$

$$x_{ij} \in \{0, 1\}, \quad i \in I, j \in J \quad (4)$$

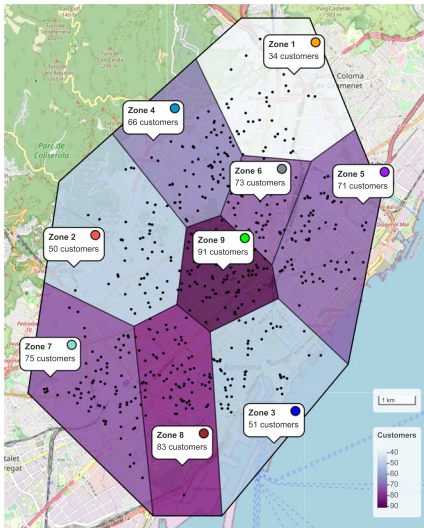
The objective function (1) lexicographically optimizes the two ranked objectives. The objective function with the highest priority, Z_1 , minimizes the imbalance, i.e., it minimizes the variance of the number of customers in the clusters. Notice that $\sum_{i \in I} x_{ij}$ provides the number of customers which are finally assigned to the cluster j . Furthermore, by expanding the summation and removing the constant terms, Z_1 is equivalent to $\sum_{j \in J} (\sum_{i \in I} x_{ij})^2$. The objective function with the second priority, Z_2 , refers to the total distance resulting from customers' reassignments. Constraints (2) ensure that each customer is assigned to exactly one cluster. Constraints (3) refer to the maximum reassignment distance. Finally, the integrality requirements for the decision variables x_{ij} are expressed by constraints (4).

Next, we illustrate the behavior of the BCDP with the help of two examples.

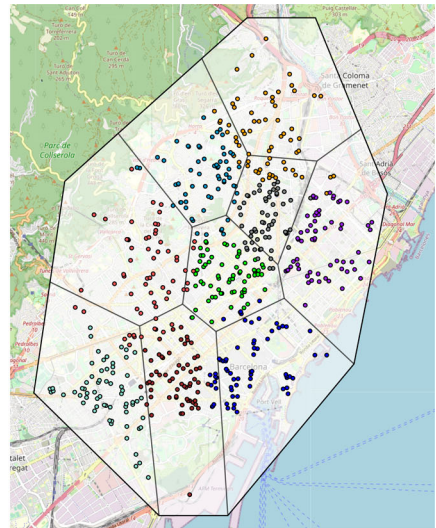
3.2 Example with full balance

This real-world example arose during the early stages of the COVID pandemic. The consultancy company which provided the data collaborated with an initiative to provide free home-delivered meals for the elderly in Barcelona (Spain). Figure 1a displays a map showing the location of the 594 customers needing to be served. Note that the customers are fairly evenly distributed in the geographical area. As an initial step, the company established the nine delivery zones represented in Fig. 1a by different shades of purple, the more intense the color the greater the number of customers in the zone. The cardinality of the corresponding clusters ranges from 34 to 91 customers, in contrast with having about 66 customers which correspond to fully balanced clusters. Taking into account the distances involved, the maximum distance to allow the reassignment of a customer has been set to $D = 1$ km.

The optimal solution of BCDP provides fully balanced clusters, i.e. clusters with 66 customers each, with the least total distance resulting from reassignments. Figure 1b displays this optimal solution. Each color represents a cluster. For instance, the optimal cluster of zone 1, whose customers are drawn as orange circles, includes the original 34 customers in this zone plus 11 customers reassigned from zone 4, 3 customers reassigned from zone 5 and 18 customers reassigned from zone 6. Note that only 94 out of the 594 customers need to be reassigned.



(a) Zones and pre-assigned clusters



(b) Optimal solution of BCDP. All clusters have 66 customers.

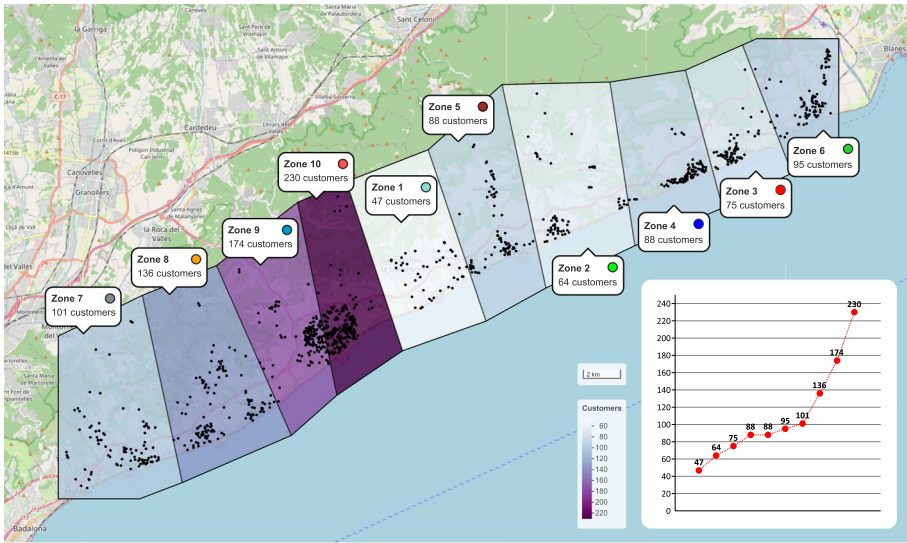
Fig. 1 Fully balance clusters

3.3 Example with incomplete balance

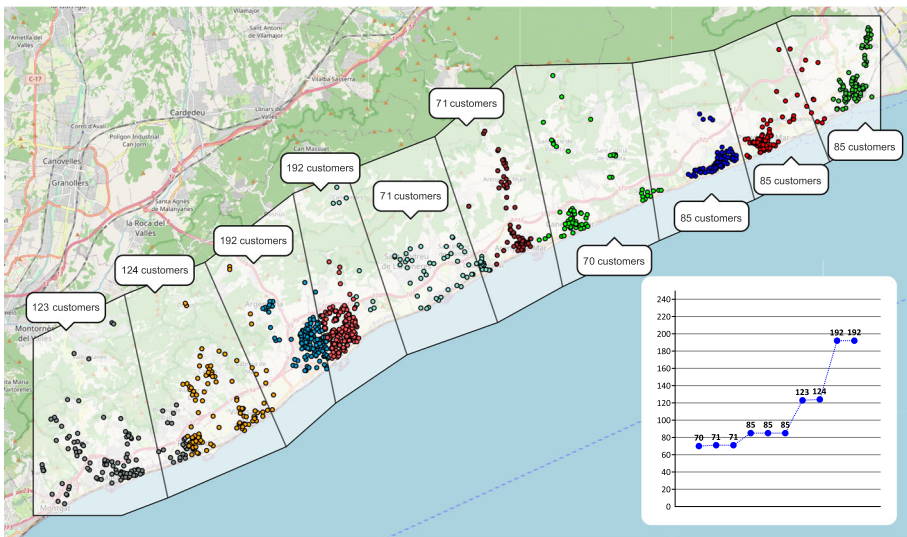
The second example corresponds to a particular distribution day of a logistics company in the area surrounding Barcelona. In this case, there are 10 zones and 1098 customers. Figure 2a displays the distribution of customers and the pre-assignment of customers to clusters. If the clusters were balanced, each of them should have about 110 customers. However, the cardinality of the pre-assigned clusters ranges from 47 to 230 customers. Note that the distribution of customers in this example is quite heterogeneous, with the number of customers assigned to each of the clusters being very different. Moreover, due to the geographical distribution of zones, customers are only close to one or two clusters, which makes balancing the problem more difficult due to the distance constraint. The maximum distance to allow the reassignment of a customer has been set to $D = 2$ km. Figure 2b shows the optimal solution provided by BCDP. As above, each color represents a cluster. In this solution, 100 out of the 1098 customers are reassigned. Now, the cardinality ranges from 70 to 152, with clusters of four sizes: three clusters with 70–71 customers, three clusters with 85 customers, two clusters with 123–124 customers and two more clusters with 192 customers. Despite the complexity of balancing this problem due to restrictions on customer movement between clusters, this model achieves the company’s goal of having more balanced customer clusters.

4 HBCDP: a heuristic for solving the BCDP

The HBCDP algorithm is based on exploring customer reassignments by pairwise cluster comparisons or by extending the search to include sequential customer exchanges between clusters. It consists of four steps which are described below. Roughly speaking, the first two stages focus on minimizing the objective function Z_1 , whereas the other two focus



(a) Zones and pre-assigned clusters



the algorithm, it is worth noting that the variance of the number of customers in the clusters can be alternatively expressed as $\sum_{j \in J} \sum_{k \in J} (y_j - y_k)^2 / (2m^2)$. Thus, in order to decrease the objective function Z_1 , it is of interest to reduce the value of the difference between y_j and y_k . Additionally, we set $x_{ij} = 1$ if the customer $i \in I$ is included in the incumbent cluster $j \in J$; otherwise, $x_{ij} = 0$. These values will be updated as the algorithm progresses through the steps. The algorithm applies the steps in the order in which they are explained in the paper.

4.1 Step 1: Improving Z_1 with direct reassigning

At this stage, the clusters are selected in pairs. Let j be the cluster to be examined. This cluster is paired with the remaining clusters $k \in J$. The cluster j is selected in increasing order of the number of customers in the pre-assigned clusters C_j . The clusters k is selected in increasing order of the distance Δ_{jk} , where Δ_{jk} measures the distance between the clusters j and k .

Reassigning a customer from the cluster j to the cluster k decreases Z_1 only if $y_j > y_k + 1$. Let this be the case and let m_{jk} be the number of customers currently assigned to the cluster j which can be reassigned to the cluster k . This means that the distance from any of these customers to the cluster k is less than or equal to D . Then, r customers are reassigned from the cluster j to the cluster k , where

$$r = \min \left\{ m_{jk}, \left\lfloor \frac{y_j - y_k}{2} \right\rfloor \right\}$$

and $\lfloor a \rfloor$ denotes the greatest integer value which is less than or equal to a . These r customers are selected from among those i who can be reassigned from the cluster j to the cluster k in increasing order of $d_{ik} - d_{ij}$. With this selection the purpose is to take into account the second objective function Z_2 .

This step is repeated while there exist pairs of clusters j and k for which $y_j > y_k + 1$ and $m_{jk} > 0$. Algorithm 1 displays the pseudocode of Step 1.

4.2 Step 2: Improving Z_1 with path reassigning

This stage extends Step 1 in the sense that it looks for reassignments which involve several clusters. For this purpose, we construct a directed graph that enables the identification of paths through which customers can be transferred among clusters.

Let $G = (J, A)$ be a directed graph where $A = \{(j, k) : m_{jk} > 0, j, k \in J\}$. Given a cluster j , we are looking for a path $\{j = j_0 \rightarrow j_1 \rightarrow j_2 \rightarrow \dots \rightarrow j_{i-1} \rightarrow j_i\}$ in G finishing in a cluster j_i such that $y_j > y_{j_i} + 1$. The existence of this path guarantees that at least one customer can be reassigned from cluster j_0 to j_1 , from j_1 to j_2 , ..., and from j_{i-1} to j_i , thus decreasing the value of Z_1 . For the purpose of computing the path (if any exists) we apply the Breadth-First Search (BFS) algorithm (Cormen et al, 2009) in the graph G starting at node j . After finding the path, r customers are reassigned along it, where

$$r = \min \left\{ m_{j_0 j_1}, m_{j_1 j_2}, \dots, m_{j_{i-1} j_i}, \left\lfloor \frac{y_{j_0} - y_{j_i}}{2} \right\rfloor \right\} \tag{5}$$

The selection of the cluster j as well as the selection of the customers i which are reassigned from cluster j_i to cluster j_{i+1} is made as explained in Sect. 4.1. This step is repeated while

Algorithm 1: Pseudocode of Step 1

```

1  $x_{ij} = 1$  for  $i \in C_j$ ;
2  $x_{ij} = 0$  for  $i \notin C_j$ ;
3  $y_j = |C_j|$  for each  $j \in J$ ;
4  $m_{jk} = |\{i \in I_k : x_{ij} = 1\}|$  for each  $j, k \in J$ ;
5 do
6   for  $j \in J$  selected in increasing order of  $|C_j|$  do
7     for  $k \in J$  selected in increasing order of  $\Delta_{jk}$  do
8       if  $y_j - y_k > 1$  then
9          $r = \min \left\{ m_{jk}, \left\lfloor \frac{y_j - y_k}{2} \right\rfloor \right\}$ ;
10        for  $r' = 1, \dots, r$  do
11           $i' = \operatorname{argmin}_{i \in I_k} \{d_{ik} - d_{ij} : x_{ij} = 1\}$ ;
12          Update  $x, y$  and  $m$  based on the reassignment of the customer  $i'$  from cluster  $j$  to cluster  $k$ ;
13        end
14      end
15    end
16  end
17 while some reassignment occurs;

```

there exist pairs of clusters j and k for which $y_j > y_k + 1$ and a path from j to k exists in G . Algorithm 2 displays the pseudocode of Step 2.

Algorithm 2: Pseudocode of Step 2

```

1 do
2   for  $j \in J$  selected in increasing order of  $|C_j|$  do
3     Let  $V$  be the set of vertices connected with a path starting at  $j$  over the graph  $G = (J, A)$ , with  $A = \{(j, k) : m_{jk} > 0\}$ , computed by using the BFS algorithm;
4     if  $V \cap \{k \in J : y_j - y_k > 1\} \neq \emptyset$  then
5       Let  $p = (j_0 = j, j_1, \dots, j_t)$  be a path in  $G$  with  $y_{j_0} - y_{j_t} > 1$ ;
6        $r = \min \left\{ m_{j_0 j_1}, \dots, m_{j_{t-1} j_t}, \left\lfloor \frac{y_{j_0} - y_{j_t}}{2} \right\rfloor \right\}$ ;
7       for  $t' = t, \dots, 2, 1$  do
8         for  $r' = 1, \dots, r$  do
9            $i' = \operatorname{argmin}_{i \in I_{j_{t'}}} \{d_{i j_{t'}} - d_{i j_{t'-1}} : x_{i j_{t'-1}} = 1\}$ ;
10          Update  $x, y$  and  $m$  based on the reassignment of the customer  $i'$  from cluster  $j_{t'-1}$  to cluster  $j_{t'}$ ;
11        end
12      end
13    end
14  end
15 while some reassignment occurs;

```

At the end of Step 2, the value of the objective function with the highest priority, Z_1 is fixed and, in the following steps, the algorithm aims to improve, if possible, the reassignment of the customers from the point of view of the second ranked objective function Z_2 , without worsening Z_1 .

4.3 Step 3: Improving Z_2 without modifying the number of customers in the clusters

At this stage, the cardinality of the current clusters remains unchanged. As previously mentioned, we have developed two variants of Step 3, which differ in the way in which the exchange of customers is made. These variants will be evaluated in Sect. 5.

4.3.1 Variant 1: Improving Z_2 by exchanging pairs of customers

The variant 1 explained in this section involves looking for pairs of customers which can be exchanged between clusters in order to reduce the value of Z_2 . For $j, k \in J$ such that $m_{jk} > 0$, we define

$$\begin{aligned}
 g_{jk} &= \min_{i \in I_k} \{d_{ik} - d_{ij} : x_{ij} = 1\} \\
 i_{jk} &= \arg \min_{i \in I_k} \{d_{ik} - d_{ij} : x_{ij} = 1\}
 \end{aligned}
 \tag{6}$$

If it is possible to reassign customers from j to k , g_{jk} measures the minimum impact in Z_2 for assigning a customer of the cluster j to the cluster k . This customer is i_{jk} . Hence, in this step we look for pairs of clusters j and k such that $g_{jk} + g_{kj} < 0$ and, if this is the case, the customers i_{jk} and i_{kj} are exchanged.

This step is repeated while there exist pairs of clusters j and k for which the above-mentioned conditions are verified. The selection of the clusters j and k is made as explained in the pseudocode of Step 3 variant 1, which is displayed in Algorithm 3.

Algorithm 3: Pseudocode of Step 3 - variant 1

```

1  $g_{jk} = \min_{i \in I_k} \{d_{ik} - d_{ij} : x_{ij} = 1\}$  for each  $j, k \in J$  with  $m_{jk} > 0$ ;
2  $g_{jk} = \infty$  for each  $j, k \in J$  with  $m_{jk} = 0$ ;
3 do
4   for  $j \in J$  selected in increasing order of  $|C_j|$  do
5      $k' = \operatorname{argmin}_{k \in J} \{g_{jk} + g_{kj}\}$ ;
6     if  $g_{jk'} + g_{k'j} < 0$  then
7        $i_{jk'} = \operatorname{argmin}_{i \in I_{k'}} \{d_{ik'} - d_{ij} : x_{ij} = 1\}$ ;
8        $i_{k'j} = \operatorname{argmin}_{i \in I_j} \{d_{ij} - d_{ik'} : x_{ik'} = 1\}$ ;
9       Update  $x, y, m$  and  $g$  based on the reassignments of the customer  $i_{jk'}$  from cluster  $j$  to
       cluster  $k'$  and of the customer  $i_{k'j}$  from cluster  $k'$  to cluster  $j$ ;
10    end
11  end
12 while some reassignment occurs;

```

4.3.2 Variant 2: Improving Z_2 by using an optimization model

The variant 2 of Step 3 considers multiple exchanges of customers simultaneously to reduce the value of Z_2 . It seeks a better assignment of customers which does not change the cardinality of the current clusters which, as mentioned before, is given by $y_j, j \in J$, by solving a transportation problem. For this problem, the sources are the customers, each supplying

1 unit, and the demand points are the clusters, each demanding its current cardinality units. The cost c_{ij} associated with the source i and the demand point j is d_{ij} if $d_{ij} < D$ and M otherwise, where M is a big enough constant.

Let the decision variable f_{ij} be equal to 1 if customer i is assigned to cluster j , and 0 otherwise. The transportation problem can be formulated as:

$$\begin{aligned} \min_{f_{ij}} \quad & \sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij} \\ \text{subject to} \quad & \sum_{j \in J} f_{ij} = 1, \quad i \in I \\ & \sum_{i \in I} f_{ij} = y_j, \quad j \in J, \\ & f_{ij} \geq 0, \quad i \in I, j \in J. \end{aligned} \quad (7)$$

After solving problem (7), the customers are reassigned according to the optimal solution f_{ij}^* , i.e. $x_{ij} = f_{ij}^*$, $i \in I$, $j \in J$.

4.4 Step 4: Improving Z_2 by exchanging multiple customers

The above variants of Step 3 do not modify the number of customers in the clusters obtained at the end of Step 2. The purpose of Step 4 is to reassign customers (possibly modifying the value of y_j , $j \in J$, but not the value of Z_1) in such a way as to achieve a reduction in the value of Z_2 .

For this purpose, let us consider the graph $G = (J, A)$ introduced in Sect. 4.2 and let g_{jk} defined in (6) be the parameter cost of the arc $(j, k) \in A$. Given a cluster j , we are looking for a path $\{j = j_0 \rightarrow j_1 \rightarrow j_2 \rightarrow \dots \rightarrow j_{t-1} \rightarrow j_t\}$ with negative cost such that $y_j = y_{j_t} + 1$ or a cycle $\{j_0 \rightarrow j_1 \rightarrow j_2 \rightarrow \dots \rightarrow j_{t-1} \rightarrow j_t = j_0\}$ with negative cost. This path or cycle is computed by applying the Bellman-Ford algorithm (Cormen et al, 2009). If a negative cost cycle is obtained, then the customers $i_{j_0 j_1}, i_{j_1 j_2}, \dots, i_{j_{t-1} j_0}$ (see (6)) are reassigned to the corresponding cluster. If a path with negative cost from j to j_t is obtained and $y_j = y_{j_t} + 1$ then the customers $i_{j j_1}, i_{j_1 j_2}, \dots, i_{j_{t-1} j_t}$ are reassigned to the corresponding cluster. In both cases, the objective function Z_2 is reduced.

This step is repeated while it is possible to find cost negative cycles or paths for which the above-mentioned conditions are verified. The selection of the cluster j is made as explained in the pseudocode of Step 4, which is displayed in Algorithm 4.

5 Computational experiment

The purpose of the computational experiment was to analyze the performance of the heuristic algorithm HBCDP. Since there are no benchmark instances in the literature for this problem, we built a set of large instances from real-world data provided by a consultancy company. These instances vary both in the geographical distribution of customers and the number of partitions established, which influences the pre-assignment of customers to clusters. These data correspond to four large areas surrounding the Spanish cities of Barcelona (BCN), Madrid (MAD), Valencia (VLC) and Zaragoza (ZAZ), which have 13,319, 14,210, 5518 and 6138 customers, respectively (see Fig. 3). The company provided three instances for each set which partition the customers in 10, 25 and 100 clusters corresponding to zones defined by polygons. Moreover, following the suggestions given by the company, a maximum distance

Algorithm 4: Pseudocode of Step 4

```

1 do
2   for  $j \in J$  selected in increasing order of  $|C_j|$  do
3     Apply Bellman-Ford algorithm to compute the shortest paths, if any, from  $j$  to the remaining
4     vertices on the graph  $G = (J, A)$ , with  $A = \{(j, k) : m_{jk} > 0\}$  and cost  $g_{jk}$  for each arc;
5     if there is some negative cycle then
6       Let  $p = (j_0, j_1, \dots, j_t = j_0)$  be a negative cycle in  $G$ ;
7       for  $t' = 1, 2, \dots, t$  do
8          $i_{j_{t'-1}j_{t'}} = \operatorname{argmin}_{i \in I_{j_{t'}}} \{d_{ij_{t'}} - d_{ij_{t'-1}} : x_{ij_{t'-1}} = 1\}$ ;
9         Update  $x, m$  and  $g$  based on the reassignment of the customer  $i_{j_{t'-1}j_{t'}}$  from cluster
10         $j_{t'-1}$  to cluster  $j_{t'}$ ;
11      end
12    else
13      Let  $V$  be the set of vertices such that the shortest path from vertex  $j$  has negative cost;
14      if  $V \cap \{k \in J : y_j - y_k = 1\} \neq \emptyset$  then
15        Let  $p = (j_0 = j, j_1, \dots, j_t)$  a path in  $G$  with  $y_{j_0} - y_{j_t} = 1$  and the least cost;
16        for  $t' = t, \dots, 2, 1$  do
17           $i_{j_{t'-1}j_{t'}} = \operatorname{argmin}_{i \in I_{j_{t'}}} \{d_{ij_{t'}} - d_{ij_{t'-1}} : x_{ij_{t'-1}} = 1\}$ ;
18          Update  $x, y, m$  and  $g$  based on the reassignment of the customer  $i_{j_{t'-1}j_{t'}}$  from
19          cluster  $j_{t'-1}$  to cluster  $j_{t'}$ ;
20        end
21      end
22    end
23  end
24  while some reassignment occurs;

```

to allow reassignment of $D = 10$ km and $D = 20$ km was established as these values are within the working standards of the logistics companies. This results in 24 instances.

The experiments were carried out on a PC Intel Core i7-6700 with 3.4 GHz, having 32.0 GB of RAM and Windows 10 64-bit as the Operating System. The BCDP was solved using IBM ILOG CPLEX 22.1.1. Although we had a multi-processor computer at hand, only one processor was used in our tests. In addition, a preprocessing step was performed to eliminate all variables x_{ij} such that $d_{ij} > D$, since these reassignments are not possible. The CPLEX stopping criterion was set at 7200s. The problem was solved in two stages. In the first stage, the BCDP was solved with the objective function Z_1 . Let Z_1^* be its optimal value. If there was remaining time within the 7200s limit, it was utilized in the second stage, which involves solving the BCDP with the objective function Z_2 and the additional quadratic constraint $Z_1 = Z_1^*$. Furthermore, in the second stage, the solution obtained at the end of the first stage was provided as the initial solution. The code of HBCDP was written in C++. For solving the transportation problem (7), we selected the simplex algorithm for Transportation Problems (MacDonald, 2015). Throughout the experiment, it is assumed that the clusters were created by partitioning the distribution area into zones defined by polygons. Hence, the distance d_{ij} is the Euclidean distance of customer i to the nearest line segment of the polygon defining zone j and Δ_{jk} measures the distance between the polygons which define zones j and k .

Table 1 displays the results of the computational experiment. The first to third columns show the characteristics of the instance (area, value of D and number of zones). The fourth and sixth columns display in bold the optimal value of the first and the second ranked objective

Table 1 Results of the computational experiment: objective function values and CPU times in seconds

Area	D	J	Z ₁		Z ₂		CPU time (s)		HBCDP ₁	HBCDP ₂
			CPLEX	HBCDP	CPLEX	HBCDP	CPLEX	HBCDP ₁		
BCN	10	10	578,876.49	=	43596.17	24127.98	7200.00	16.69	12.85	
		25	45,450.10	=	58418.10	=	7200.00	27.96	33.66	
	20	100	4856.51	=	79886.08	49259.70	7200.00	48.14	199.93	
		10	149,988.09	=	72089.68	72089.63	7200.00	25.48	17.75	
		25	4876.10	=	144522.44	83532.34	7200.00	50.39	46.53	
MAD	10	100	1862.19	=	169133.04	113276.84	7200.00	141.29	257.02	
		10	211,248.40	=	31,998.59	=	234.81	16.10	15.34	
	20	25	64,473.84	=	68242.28	35832.47	7200.00	26.91	42.52	
		100	4007.59	=	84408.15	44336.04	7200.00	76.62	228.80	
		10	41,074.00	=	51,786.93	=	2104.70	77.56	16.14	
VLC	10	25	15,530.72	=	156912.73	79116.58	7200.00	74.47	46.60	
		100	969.49	=	177516.80	92547.58	7200.00	142.16	262.68	
	20	10	189,754.16	=	16127.44	16127.39	7200.00	0.87	2.69	
		25	14,904.84	=	19707.16	=	7200.00	3.98	7.23	
		100	766.49	=	32166.91	19733.80	7200.00	27.87	48.17	
ZAZ	10	10	101,379.56	=	33439.50	33439.40	7200.00	6.10	3.11	
		25	8761.24	=	56929.37	36220.46	7200.00	8.02	8.32	
	20	100	372.25	=	62638.44	37734.19	7200.00	37.56	48.86	
		10	283,014.56	=	13661.61	13661.59	7200.00	0.52	2.52	
		25	59,315.69	=	24694.55	24693.32	7200.00	3.33	7.74	
MAD	10	100	2580.30	=	35540.39	26176.53	7200.00	20.74	47.41	
		10	112,782.96	=	48707.32	34897.86	7200.00	9.93	3.39	
	20	25	37,686.41	=	63311.80	46092.25	7200.00	17.27	8.98	
		100	1042.78	=	73084.64	59135.17	7200.00	28.15	62.35	

HBCDP₁ and HBCDP₂ refers to variants 1 and 2 in Step 3 of the algorithm

Numbers in bold refer to the optimal value of the first and second ranked objective functions

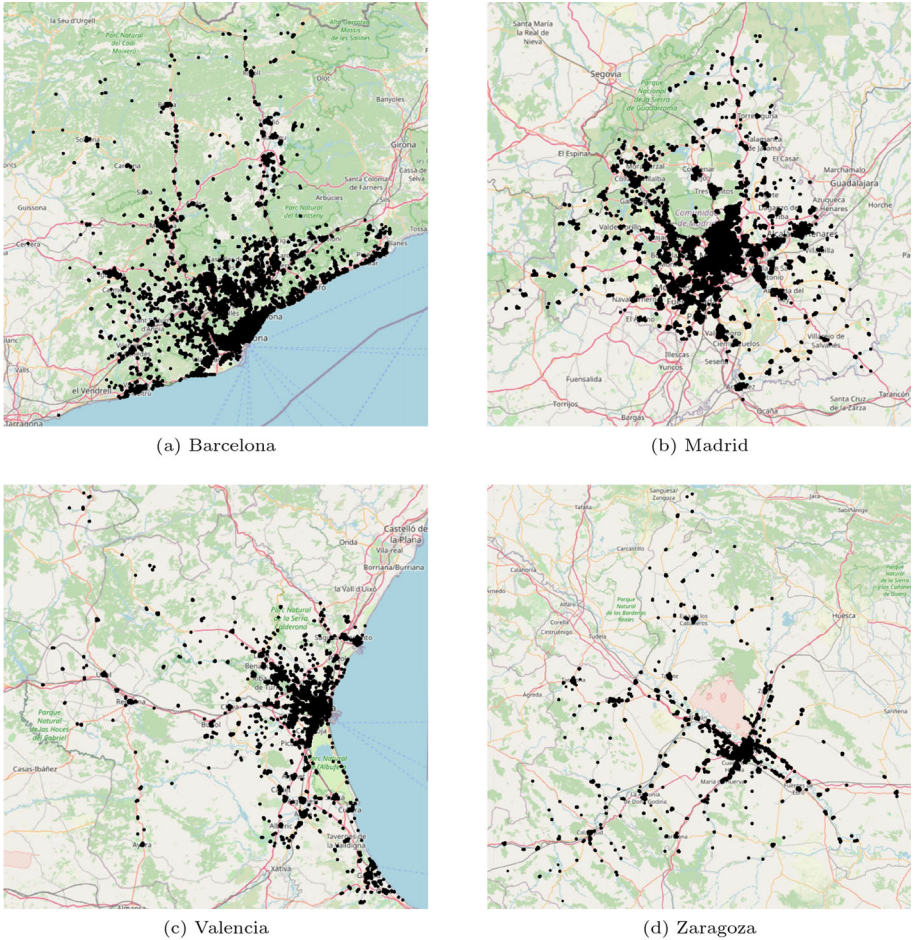


Fig. 3 Customer distribution in the benchmark instances

functions Z_1 and Z_2 when the instance is solved to optimality by CPLEX; otherwise, they show the best objective function value provided by CPLEX. Considering the values taken by Z_1 , we have selected MIP gap tolerance = $1e-06$ to ensure a more accurate comparison between the results obtained from CPLEX and the algorithm HBCDP. Since both variants of the algorithm yield identical values for Z_1 and Z_2 , the table presents a single column corresponding to each objective. Thus, the fifth and seventh columns display the best value of both objective functions provided by the algorithm HBCDP. The symbol ‘=’ indicates that HBCDP provides the same value as CPLEX. It is worth mentioning that HBCDP always provides the same value of Z_1 as CPLEX. Regarding Z_2 , HBCDP always provides values which are less than or equal to those given by CPLEX.

CPLEX is only capable of completely solving two instances in the allotted time, those corresponding to MAD with $D = 10$ and $D = 20$, and $|J| = 10$. In 14 out of the remaining 22 instances, CPLEX is only able to yield the optimal solution with respect to the first ranked objective function Z_1 , but terminates due to the stopping criterion while solving the problem with the second objective function Z_2 . In the remaining 8 examples, CPLEX stops when

Table 2 Percentage decrease of Z_1 and Z_2 through the steps of the algorithm

Area	D	$ J $	$PD_{Z_1}^{s1}$	$PD_{Z_1}^{s2}$	HBCDP ₁		HBCDP ₂	
					$PD_{Z_2}^{s3}$	$PD_{Z_2}^{s4}$	$PD_{Z_2}^{s3}$	$PD_{Z_2}^{s4}$
BCN	10	10	64.95	0.00	7.38	0.37	7.68	0.05
		25	80.35	0.00	9.39	0.39	9.75	0.00
		100	69.62	0.01	8.87	0.48	9.28	0.04
	20	10	90.92	0.00	8.39	0.04	8.43	0.00
		25	97.89	0.01	9.98	0.29	10.23	0.01
		100	88.34	0.08	7.06	0.64	7.62	0.04
MAD	10	10	86.61	0.00	4.15	0.03	4.16	0.02
		25	78.72	0.00	11.36	0.16	11.47	0.04
		100	79.04	0.02	12.75	1.53	13.97	0.13
	20	10	97.40	0.00	16.54	0.26	16.74	0.02
		25	94.87	0.00	12.57	0.97	13.39	0.03
		100	94.93	0.02	10.76	0.84	11.37	0.16
VLC	10	10	76.97	0.00	0.03	0.06	0.04	0.06
		25	88.08	0.00	4.70	0.02	4.71	0.02
		100	80.41	0.14	8.63	0.52	9.03	0.08
	20	10	87.69	0.00	7.92	0.05	7.93	0.05
		25	92.99	0.01	6.09	0.24	6.30	0.02
		100	90.49	0.08	9.12	0.46	9.41	0.14
ZAZ	10	10	74.18	0.00	0.00	0.00	0.00	0.00
		25	78.21	0.00	1.40	1.01	2.35	0.05
		100	76.41	0.02	4.83	0.44	5.06	0.20
	20	10	89.71	0.00	3.65	0.04	3.65	0.04
		25	86.16	0.00	5.52	0.12	5.53	0.11
		100	90.47	0.05	4.18	0.20	4.30	0.07
Mean			84.81	0.02	7.30	0.38	7.60	0.06

solving the problem with the objective function Z_1 . In these cases, the table shows the value of Z_2 corresponding to the solution provided by CPLEX when it terminates. In summary, HBCDP matches the optimal solution when it is available and provides a better solution when CPLEX stops at the time limit.

In order to show the impact on the variance value of the clustering provided by the algorithm compared to the initial clustering, the fourth column of Table 2 displays the percentage decrease in the value of Z_1 after step 1 calculated as:

$$PD_{Z_1}^{s1} = \frac{Z_1^{init} - Z_1^{step1}}{Z_1^{init}} \times 100$$

Notice that the reduction ranges from 64.95 to 97.89, with an average of 84.81. As expected, the greater the distance, the greater the reduction. The fifth column displays the percentage decrease in the value of Z_1 after step 2 computed as:

$$PD_{Z_1}^{s2} = \frac{Z_1^{step1} - Z_1^{step2}}{Z_1^{step1}} \times 100 \quad (8)$$

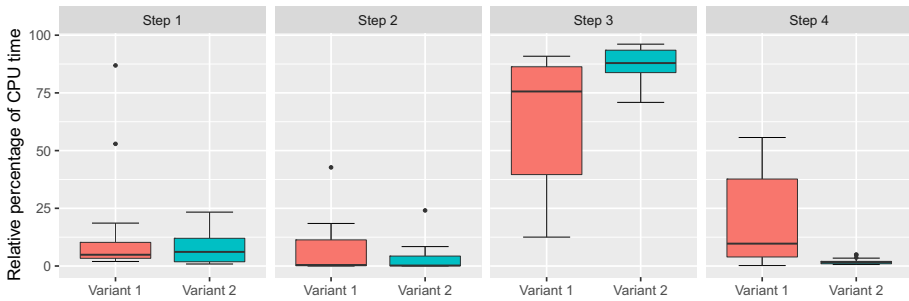


Fig. 4 Boxplot of the relative percentage of CPU time according to the variant used in Step 3 of the algorithm

This step results in smaller decreases, since, in many cases, the optimal solution is achieved following step 1. The subsequent four columns illustrate the percentage decrease in the second objective function value after steps 3 and 4, respectively, depending on the variant. This percentage reduction has been computed in a manner analogous to the expression (8), using Z_2 and adjusting the step accordingly. Both variants yield comparable reductions.

Regarding the computing time, the eighth column shows the CPU time in seconds spent by CPLEX in solving the instances. The ninth and tenth columns display HBCDP₁ and HBCDP₂, the CPU time needed by HBCDP with the first and the second variant in step 3, respectively. It can be observed that the algorithm HBCDP requires significantly less time compared to CPLEX, regardless of the variant used. It is worth pointing out that when variant 1 of Step 3 is used, the algorithm HBCDP consistently takes less than three minutes. Similarly, when variant 2 is utilized, it takes less than five minutes. When comparing the CPU times according to the variant, we conclude that, in 15 out of the 24 instances, variant 1 spends less time than variant 2. Moreover, variant 1 generally performs better when the instance is larger. Finally, to gain an insight into how each variant of the algorithm HBCDP uses the CPU time, for each instance and variant we have computed the relative percentage of CPU time invested in each step. Figure 4 shows the corresponding boxplots. Note that variant 2, i.e. solving the transportation problem (7), leads to less time in step 4, although in general step 3 consumes more CPU time. Moreover, the variability in steps 3 and 4 increases when using variant 1.

6 Conclusions

This paper addresses a problem often faced by logistics companies, that of establishing clusters of their customers that allow the companies to efficiently manage their resources and workload. The difficulty arises when the clusters cannot be formed by customers that are too far apart, as this hinders the daily management of the logistics company. For this reason, logistics companies generally divide their geographical distribution area into zones and group their customers according to their geographical location. Depending on the characteristics of the zones, this can lead to very unbalanced clusters. In addition, the number of customers to be served varies from day to day, which can also contribute to this imbalance. Therefore, the problem dealt with in this paper is that of reassigning customers to achieve balanced clusters formed by customers who are reasonably close.

The first issue to be addressed in order to solve this problem is to establish how the balance, in terms of the number of customers, between clusters should be measured when

there is a distance constraint that prevents reassigning customers to clusters that are too far apart. We propose to measure the imbalance through the minimization of the variance of the number of customers in the clusters. Hence, the lexicographic biobjective nonlinear integer problem proposed to model the BCDP, considers this imbalance measure as the highest priority objective function. Moreover, looking for the optimal solution with the least total distance resulting from customers' reassignments, this distance is taken as the second ranked objective function.

The large instances faced by logistics companies are very time-consuming to solve by using commercial software. This prevents cluster balancing from being incorporated into the daily work routine. Hence, we have developed a tailored algorithm that is accurate and fast. The algorithm reassigns customers by pairs or sequentially as long as the objective functions improve. The performance of the algorithm has been tested in a set of large instances provided by the consultancy company. For all of them, the solution provided by the algorithm is better than or equal to the solution provided by the commercial software CPLEX, and the CPU times are smaller. In fact, none of the instances need more than 2.5 min of CPU time when solved by the algorithm.

From a managerial point of view, it is worth mentioning that the BCDP and thus the approach taken to handle the cluster imbalance both allow the companies to construct groups of clusters with a fairly homogeneous cardinality. Moreover, the accuracy and speed of the developed algorithm allows it to be used in real time. Hence, both provide a useful decision support tool which can be used by logistics companies in their day-to-day operations, prior to workload allocation or routing.

Although particularly effective in addressing challenges faced by logistics firms distributing small parcels, especially in densely populated urban areas, it is worth considering potential future applications in other areas. One possible direction for future research involves extending this procedure to address other problems that incorporate balancing in user clusters, such as districting. This extension could involve the consideration of additional constraints beyond distance, which would help to maintain the desired properties for a district. Also, at the clustering level, it would be interesting to investigate the possibility of modifying the algorithm to generate balanced clusters from scratch. The modeling and algorithm presented here offer a different perspective from those commonly found in the literature, which often prioritize the objective defining clusters by altering the phase of introducing elements into clusters in an attempt to provide balanced clusters. The natural progression of the study about the use of balanced clusters in the logistics environment presented in this paper leads to consider additional constraints. These constraints may involve scenarios where customers possess multiple packages or there are capacity restrictions on the number of customers that can be assigned to a cluster.

Acknowledgements This research has been funded by the Spanish Ministry of Economy, Industry and Competitiveness under grants PID2019-104263RB-C43 and PID2022-139543OB-C43, and by the Gobierno de Aragón under grants E41-20R and E41-23R. The authors gratefully acknowledge the consultancy company Alerce for posing the problem and providing the instances.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2, 244–263. [https://doi.org/10.1016/0022-0531\(70\)90039-6](https://doi.org/10.1016/0022-0531(70)90039-6)
- Bektaş, T., Gouveia, L., Martínez-Sykora, A., et al. (2019). Balanced vehicle routing: Polyhedral analysis and branch-and-cut algorithm. *European Journal of Operational Research*, 273, 452–463. <https://doi.org/10.1016/j.ejor.2018.08.034>
- Bélanger, V., Ruiz, A., & Soriano, P. (2019). Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. *European Journal of Operational Research*, 272, 1–23. <https://doi.org/10.1016/j.ejor.2018.02.055>
- Bender, M., Kalcsics, J., & Meyer, A. (2020). Districting for parcel delivery services: A two-stage solution approach and a real-world case study. *Omega*, 96(102283), 1–21. <https://doi.org/10.1016/j.omega.2020.102283>
- Benzarti, E., Sahin, E., & Dallery, Y. (2013). Operations management applied to home care services: Analysis of the districting problem. *Decision Support Systems*, 55, 587–598. <https://doi.org/10.1016/j.dss.2012.10.015>
- Cormen, T., Leiserson, C., Rivest, R., et al. (2009). *Introduction to algorithms* (3rd ed.). Cambridge, MA: The MIT Press.
- De Maeyer, R., Sieranoja, S., & Fränti, P. (2023). Balanced k-means revisited. *Applied Computing and Intelligence*, 3(2), 145–179. <https://doi.org/10.3934/aci.2023008>
- Diglio, A., Nickel, S., & Saldanha-da-Gama, F. (2020). Towards a stochastic programming modeling framework for districting. *Annals of Operations Research*, 292(1), 249–285.
- Ding, H. (2020). Faster balanced clusterings in high dimension. *Theoretical Computer Science*, 842(1), 28–40.
- Ehrgott, M. (2005). *Multicriteria Optimization* (2nd ed.). Berlin: Springer.
- Eiselt, H., & Laporte, G. (1995). Objectives in location problems. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 151–179). Berlin: Springer.
- Goderbauer, S., & Winandy, J. (2018). Political districting problem: Literature review and discussion with regard to federal elections in Germany https://www.or.rwthachen.de/researchpublications/LitSurvey_PoliticalDistricting_Goderbauer_Winandy_20171123.pdf
- Golden, B., Laporte, G., & Taillard, E. (1997). An adaptive memory heuristic for a class of vehicle routing problems with minmax objective. *Computers & Operations Research*, 24(5), 445–452. [https://doi.org/10.1016/S0305-0548\(96\)00065-2](https://doi.org/10.1016/S0305-0548(96)00065-2)
- Halvorsen-Weare, E., & Savelsbergh, M. (2016). The bi-objective mixed capacitated general routing problem with different route balance criteria. *European Journal of Operational Research*, 251, 451–465. <https://doi.org/10.1016/j.ejor.2015.11.024>
- Haugland, D., Ho, S., & Laporte, G. (2007). Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, 180, 997–1010.
- Janssens, J., Van den Bergh, J., Sörensen, K., et al. (2015). Multi-objective microzone-based vehicle routing for courier companies: From tactical to operational planning. *European Journal of Operational Research*, 242, 222–231. <https://doi.org/10.1016/j.ejor.2014.09.026>
- Jozefowicz, N., Semet, F., & Talbi, E. (2007). Target aiming pareto search and its application to the vehicle routing problem with route balancing. *Journal of Heuristics*, 13(3), 455–469. <https://doi.org/10.1007/s10732-007-9022-6>
- Jozefowicz, N., Semet, F., & Talbi, E. (2009). An evolutionary algorithm for the vehicle routing problem with route balancing. *European Journal of Operational Research*, 195(3), 761–769. <https://doi.org/10.1016/j.ejor.2007.06.065>
- Kalcsics, J., & Ríos-Mercado, R. (2019). Districting problems. In G. Laporte, S. Nickel, F. Saldanha da Gama (Eds) *Location science*. Springer, pp. 705–743. https://doi.org/10.1007/978-3-030-32177-2_25
- Karsu, O., & Morton, A. (2015). Inequity averse optimization in operational research. *European Journal of Operational Research*, 245, 343–359. <https://doi.org/10.1016/j.ejor.2015.02.035>
- Lacomme, P., Prins, C., Prodhon, C., et al. (2015). A multi-start split based path relinking (MSSPR) approach for the vehicle routing problem with route balancing. *Engineering Applications of Artificial Intelligence*, 38, 237–251. <https://doi.org/10.1016/j.engappai.2014.10.024>

- Lee, T., & Ueng, J. (1999). A study of vehicle routing problems with load-balancing. *International Journal of Physical Distribution & Logistics Management*, 29(10), 646–657. <https://doi.org/10.1108/09600039910300019>
- Lehuédé, F., Péton, O., & Tricoire, F. (2020). A lexicographic minimax approach to the vehicle routing problem with route balancing. *European Journal of Operational Research*, 282, 129–147. <https://doi.org/10.1016/j.ejor.2019.09.010>
- Liberatore, F., Camacho-Collados, M., & Vitoriano, B. (2020). Police districting problem: Literature review and annotated bibliography. In R. Ríos-Mercado (Ed) *Optimal districting and territory design*. Springer, Cham, pp. 9–29. https://doi.org/10.1007/978-3-030-34312-5_2
- Lin, W., He, Z., & Xiao, M. (2019). In *Proceedings of the twenty-eighth international joint conference on artificial intelligence (IJCAI-19)*. AAAI Press, Macao, China, pp. 2987–2993.
- Linfati, R., Yáñez-Concha, F., & Escobar, J. (2020). Mathematical models for the vehicle routing problem by considering balancing load and customer compactness. *Sustainability*, 14(12937). <https://doi.org/10.3390/su141912937>.
- Liu, K., Zhang, H., & Zhang, Z. (2021). The efficiency, equity and effectiveness of location strategies in humanitarian logistics: A robust chance-constrained approach. *Transportation Research Part E*, 156, 102521. <https://doi.org/10.1016/j.tre.2021.102521>
- MacDonald, D. T. (2015). C++ implementation of the transportation simplex algorithm. Available online <https://github.com/engine99/transport-simplex>
- Malinen, M., & Fránti, P. (2014). Balanced k-means for clustering. In P. Fránti, G. Brown, M. Loog, et al. (Eds.) *Structural, syntactic, and statistical pattern recognition*. S+SSPR 2014. Lecture Notes in Computer Science, vol. 8621. Springer, pp. 32–41.
- Mancini, S., Gansterer, M., & Hartl, R. (2021). The collaborative consistent vehicle routing problem with workload balance. *European Journal of Operational Research*, 293, 955–965. <https://doi.org/10.1016/j.ejor.2020.12.064>
- Marín, A. (2011). The discrete facility location problem with balanced allocation of customers. *European Journal of Operational Research*, 210, 1–17. <https://doi.org/10.1016/j.ejor.2010.10.012>
- Marsh, M., & Schilling, D. (1994). Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research*, 74, 1–17. [https://doi.org/10.1016/0377-2217\(94\)90200-3](https://doi.org/10.1016/0377-2217(94)90200-3)
- Matl, P., Hartl, R., & Vidal, T. (2018). Workload equity in vehicle routing problems: A survey and analysis. *Transportation Science*, 52(2), 239–260. <https://doi.org/10.1287/trsc.2017.0744>
- Matl, P., Hartl, R., & Vidal, T. (2019). Workload equity in vehicle routing: The impact of alternative workload resources. *Computers & Operations Research*, 110, 116–129. <https://doi.org/10.1016/j.cor.2019.05.016>
- Nikolakopoulou, G., Kortesis, S., Synefaki, A., et al. (2004). Solving a vehicle routing problem by balancing the vehicles time utilization. *European Journal of Operational Research*, 152, 520–527. [https://doi.org/10.1016/S0377-2217\(03\)00042-0](https://doi.org/10.1016/S0377-2217(03)00042-0)
- Ogryczak, W. (2000). Inequality measures and equitable approaches to location problems. *European Journal of Operational Research*, 122, 374–391. [https://doi.org/10.1016/S0377-2217\(99\)00240-4](https://doi.org/10.1016/S0377-2217(99)00240-4)
- Romero, C. (2001). Extended lexicographic goal programming: A unifying approach. *Omega*, 29, 63–71. [https://doi.org/10.1016/S0305-0483\(00\)00026-8](https://doi.org/10.1016/S0305-0483(00)00026-8)
- Samanta, S., Sen, G., & Ghosh, S. (2022). A literature review on police patrolling problems. *Annals of Operations Research*, 316, 1063–1106. <https://doi.org/10.1007/s10479-021-04167-0>
- Sandoval, M., Álvarez-Miranda, E., Pereira, J., et al. (2022). A novel districting design approach for on-time last-mile delivery: An application on an express postal company. *Omega*, 113, 102687. <https://doi.org/10.1016/j.omega.2022.102687>
- Schneider, M., Stenger, A., Schwahn, F., et al. (2015). Territory-based vehicle routing in the presence of time-window constraints. *Transportation Science*, 49(4), 732–751. <https://doi.org/10.1287/trsc.2014.0539>
- Sen, A., & Foster, J. (1997). *On economic inequality* (Enlarged). New York: Oxford University Press.
- Tasnádi, A. (2011). The political districting problem: A survey. *Society and Economy*, 33(3), 543–554.
- Vidal, T., Laporte, G., & Matl, P. (2020). A concise guide to existing and emerging vehicle routing problem variants. *European Journal of Operational Research*, 286, 401–416. <https://doi.org/10.1016/j.ejor.2019.10.010>
- Wong, K., & Beasley, J. (1984). Vehicle routing using fixed delivery areas. *Omega*, 12(6), 591–600. [https://doi.org/10.1016/0305-0483\(84\)90062-8](https://doi.org/10.1016/0305-0483(84)90062-8)
- Young, H. (1994). *Equity in theory and practice*. Princeton, NJ: Princeton University Press.
- Zhou, L., Zhen, L., Baldacci, R., et al. (2021). A heuristic algorithm for solving a large-scale real-world territory design problem. *Omega*, 103(102442), 1–28. <https://doi.org/10.1016/j.omega.2021.102442>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.