JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
2020, VOL. 71, NO. 11, 1740–1759
https://doi.org/10.1080/01605682.2019.1621217

ORIGINAL ARTICLE

A novel optimization model for designing compact, balanced, and contiguous healthcare districts

Hiwa Farughi[a], Madjid Tavana[b,c], Sobhan Mostafayi[a] and Francisco J. Santos Arteaga[d]

[a]Department of Industrial Engineering, University of Kurdistan, Sanandaj, Iran; [b]Business Systems and Analytics Department, La Salle University, Philadelphia, PA, USA; [c]Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, Paderborn, Germany; [d]Faculty of Economics and Management, Free University of Bolzano, Bolzano, Italy

ABSTRACT
In this study, we propose a new multi-objective mathematical model for designing compact, balanced, and contiguous districts in healthcare systems. The objective functions minimize districting heterogeneity and the implementation cost of monitoring plans for promoting hygiene and public health. Expert teams perform these plans periodically by maximizing the coverage area. The purpose of the districting problem is to specify how teams are formed and allocated based on their service provision capacity and type of expertise. Improper team allocation and unsuitable service provision cause an increase in time and costs, fostering an adverse effect on the promotion of public health in the area. Compliance of the plan with the specified requirements and its implementation cost are the main factors impacting the districting problem decisions. We define two meta-heuristic algorithms including a multi-objective genetic algorithm II (NSGAII) and a multi-objective grey wolf optimizer (MOGWO) for solving the mathematical model in a real-scale because districting problems are NP-hard problems. We also present a case study taking place at a university medical centre in Iran to demonstrate the applicability and efficacy of the proposed model.

ARTICLE HISTORY
Received 11 April 2018
Accepted 10 May 2019

KEYWORDS
Health monitoring plan; districting problem; bi-objective integer mathematical model; NSGAII; MOGWO

1. Introduction

Generally defined, the districting problem refers to the grouping of small areas (basic units) into larger groups (districts). Optimally, the resulting districts should exhibit features such as balance, contiguity, compactness, and absence of embedded districts. This latter condition is met whenever districts are not located inside one another. In a districting problem, the inter-area relations must be shown in a network represented as an undirected graph $G = (V, E)$, where V represents the set of vertices, and E represents the communication paths between the vertices. In other words, the set $V = \{v_i; i = 1, 2, \dots, |V|\}$, with a cardinality of $|V|$, represents the vertices of the graph. Each vertex within this set, v_i, is represented by vertical and horizontal vector components, (x_i, y_i). Similarly, the set $E = \{e_{ij}; i, j = 1, 2, \dots, |V|; i \neq j; e_{ij} = e_{ji}\}$, containing a total of $|E|$ elements, accounts for the graph edges, with e_{ij} representing the relation between vertices v_i and v_j. The districting problem has many applications, such as political, school, and social facility districting, as well as urban waste collection. Healthcare constitutes one of the newest applications of the districting problem (Datta, Figueira, Gourtani, & Morton, 2013).

One of the main activities performed by most health systems is the implementation of specific plans aimed at supporting the provision of a wider range of services, which advances hygiene and public health levels. These plans are generally developed with the help of expert medical teams allocated and dispatched to populated areas (Shortell & Kaluzny, 2000). The allocation of a team to a specific area requires coordination with government centres, which usually involves administrative bureaucracy, as well as time and money. Coordination activities facilitate the allocation of resources and reduce operating costs. Moradi-Lakeh and Vosoogh-Moghaddam (2015) show that plan compliance, namely, the effectiveness of the plans, and operational costs are the key criteria for their successful implementation.

The allocation of teams and populated areas to a partition is based on the type of services needed by the areas and aimed at maximizing the coverage level of the needs. When a team responsible for the implementation of a specific number of services is allocated to a partition, it may happen that some of the services provided by the team are not demanded in several areas covered by this particular partition. In this regard, the concept of accurate implementation does not denote the quality of service implementation provided by the team, since this latter

CONTACT Madjid Tavana tavana@lasalle.edu Business Systems and Analytics Department, La Salle University, Philadelphia, PA 19141, USA

depends on the experience and scientific levels of the team members. This concept denotes accuracy in the proper fulfilment of the services needed by each sector by the team(s) allocated to it. For example, if plan compliance is a vital factor, the allocation of any area to a partition should aim at its fulfilment in an optimal state. In this case, allocating only one area to each partition leads to a heterogeneity value of zero and the maximum accuracy possible. The second criterion, the implementation cost, is not limited to the financial costs of implementing the plan, but also includes costs derived from the coordination activities required to form and transfer teams as well as those arising from the different interactions with the administrative bureaucracy. Therefore, if the implementation cost of the plan is one of the main priorities, the optimal scenario would consist of the creation of a single partition to which all the areas are allocated. Therefore, as can be intuitively inferred, the two factors considered, plan compliance and implementation costs, are in direct conflict since increasing one of them leads to a decrease in the other.

The problem investigated in this paper is that of districting populated areas and allocating the services needed by the areas to the relevant districts using a bi-objective mathematical model with the objectives of maximizing plan compliance and minimizing implementation cost. The objectives are achieved, respectively, by minimizing the heterogeneity of the services required by the populated areas composing the partitions and the number of team displacements taking place across the different partitions. Furthermore, since teams must be established in a given partition, the formation of partitions results in the formation and establishment of the teams being allocated to each one of them. Therefore, the number of partitions defined must aim at reducing the costs of formation and establishment of teams according to the objectives stated in the problem.

The paper proceeds as follows. Section 2 reviews the related literature and provides intuition on the research being performed. The mathematical model along with the main assumptions on which it is built are presented in Section 3. The solution method proposed and the corresponding algorithms are detailed in Section 4. The results obtained are presented and analysed in Section 5. Furthermore, the monitoring plan implemented by the health authorities of the South Khorasan Province in Iran is investigated as a case study in Section 6. Finally, Section 7 concludes and suggests future research directions.

2. Literature review

Several methods have been suggested to solve districting problems focusing on their different applications. However, a comprehensive mathematical model has not yet been presented due to the difficulties involved in defining constraints such as the necessity of contiguity and the absence of embedded districts (Boulle, 2004; Kalcsics, 2015; Lewis, Kochenberger, & Alidaee, 2008; Salazar-Aguilar, Ríos-Mercado, & Cabrera-Ríos, 2011; Steiner, Datta, Neto, Scarpin, & Figueira, 2015). As a result, heuristic and meta-heuristic algorithms have been used for solving partitioning problems (e.g. Baruch, Cret, & Pusztai, 1999; Datta & Figueira, 2011; Kim, Hwang, Kim, & Moon, 2011; Steiner, Datta, Neto, Scarpin, & Figueira, 2015). Some of the most important studies describing the structure and potential applications of the districting problem are examined below.

Baruch, Cret, and Pusztai (1999) applied the meta-heuristic genetic algorithm for the first time to solve partitioning problems. In recent years, researchers have focused on the behaviour of algorithms in solving the graph partitioning problem. Even though there are just a few of these papers, their results have impressively promoted the power of algorithms in solving several problems in the field. Kim and Moon (2004) investigated the solution space structure of graph partitioning problems. Based on their results, the space around the global optimal solution is convex. Thus, if the crossover operator is used to search the space, the performance of the algorithm would increase, since the operator leads the solutions to transfer more quickly from marginal regions to central ones in the space, which makes it more probable to find a proper solution. The crossover operator has therefore been identified as an important one in designing solution algorithms for graph partitioning problem. Refer to Kim, Hwang, Kim, and Moon (2011) to familiarize with how algorithms, the genetic one in particular, operate in solving graph partitioning problems.

Since the genetic algorithm lacks proper performance when doing proximity searches around the available solutions, hybrid algorithms based on the genetic one have been developed. The main purpose of designing these combinatorial algorithms is to exploit the advantages of other algorithms together with the crossover operator of the genetic one. Local search algorithms are an instance of this type of algorithms, which usually obtain feasible solutions within an acceptable time, though such solutions are often of improper quality. In this regard, the memetic algorithm refers to a group of algorithms that incorporate the environmental conditions of the problem within the genetic one. Inayoshi and Manderick (1994) applied the memetic algorithm for the first time to solve the graph partitioning problem. Table 1 presents some of the main studies

Table 1. Genetic algorithm used in solving the districting problem.

		Constraints			Objective function	
Author(s)	Area of study	Contiguity	Compactness	Lack of holes	Single-objective	Multi-objective
Maini, Mehrotra, Mohan, and Ranka (1994)	Structural study	✓			✓	
Rummler and Apetrei (2002)	Structural study	✓				✓
Datta, Figueira, Fonseca, and Tavares-Pereira (2008)	Structural study	✓	✓	✓		✓
Tavares-Pereira, Figueira, Mousseau, and Roy (2007)	Partitioning and analysing the urban transport system in Paris	✓	✓			✓
Datta, Fonseca, and Deb (2008)	Urban land segmentation	✓	✓			✓
Datta and Figueira (2011)	Partitioning populated areas	✓	✓	✓		✓
Datta, Malczewski, and Figueira (2012)	Political partitioning of populated areas	✓	✓	✓		✓
Datta, Figueira, Gourtani, and Morton (2013)	Health system partitioning	✓	✓	✓		✓
Steiner, Datta, Neto, Scarpin, and Figueira (2015)	Health system partitioning	✓	✓	✓		✓
Knight, Harper, and Smith (2012)	Locating emergency medical services	✓			✓	
Lin, Sir, and Pasupathy (2013)	Specification of resource levels in surgical services	✓				✓

in the different research areas where the genetic algorithm has been applied to solve the graph partitioning problem.

Besides the genetic algorithm, several other meta-heuristic ones, such as simulated annealing (Brooks & Morgan, 1995), tabu search (Bozkaya, Erkut, & Laporte, 2003), hybrids of simulated annealing and tabu search (Baños, Gil, Paechter, & Ortega, 2007), particle swarm (Wang, Wu, & Mao, 2007), and differential evolution (Datta & Figueira, 2011), have been applied to solve partitioning problems in different application areas. Table 2 describes several main research opuses that have used algorithms other than the genetic one for solving the partitioning problem. Moreover, Table 3 describes several recent studies that have also investigated the population area partitioning problem.

Based on the literature reviewed, we have not found any research dealing with the districting problem together with the organization of expert teams when implementing health system plans. Furthermore, only three studies have investigated multi-objective problems when analysing healthcare districting (Benzarti, Sahin, & Dallery, 2013; Datta, Figueira, Gourtani, & Morton, 2013; Steiner, Datta, Neto, Scarpin, & Figueira, 2015). This is the case despite the fact that in developed and developing countries a large part of social health is evaluated using the implementation of such plans. To the best of our knowledge, the simultaneous evaluation of the healthcare districting problem together with the organization of expert teams in the implementation of health system plans is investigated for the first time in the current paper.

Despite the large number of publications, only a small number of papers have studied districting problems in practice. Kalcsics (2015) has argued that there is very little consensus on the suitability of the districting criteria, their importance and measurement in real-world problems. This author has suggested that instead of developing yet another meta-heuristic for districting problems, researchers should focus on a common and generic framework. Therefore, in the current paper, a generic framework for districting problems is investigated for the first time using integer mathematical modelling. Furthermore, the model proposed considers the compactness and contiguity of partitions as well as the absence of embedded districts constraints. The contiguity and absence of embedded districts constraints are among the fundamental ones of area partitioning and can be used in other partitioning problems as well. The proposed mathematical model has two objectives, namely, the maximization of the plan compliance factor – achieved through the minimization of the heterogeneities included in each partition – and the minimization of the implementation cost factor – achieved through the minimization of the number of displacements of the expert teams across partitions.

3. Statement of the problem

This paper investigates the districting problem designed to manage the implementation plans of organizations whose goal is to improve the general hygiene level of society. In these plans, the cities demanding health services are divided into specific districts. Unlike previous research, the number of districts is not given as input data, and the proposed mathematical model specifies the optimal number of districts based on the objectives being considered. Furthermore, the fundamental constraints of the districting problem (contiguity, compactness, and

Table 2. Meta-heuristic algorithms used in solving the districting problem.

Author(s)	Case study	Method	Domain of usage
De Assis, Franca, and Usberti (2014)	São Paulo, Brazil	GRASP	Classifying urban areas for classifying electric meter readers
Li, Church, and Goodchild (2014)	Southern California	TS-SA	Partitioning urban lands, transport in Southern California
Benzarti, Sahin, and Dallery (2013)	–	MILP	Partitioning population areas, domestic health system
Shirabe (2012)	–	Geographic map analysis-heuristic method	School bus problem
Salazar-Aguilar, Ríos-Mercado, and González-Velarde (2011)	Monterrey, Mexico	MILP	Partitioning population areas, mineral water distribution
Ricca, Scozzari, and Simeone (2013)	Italy	TS & SA	Partitioning population areas, polling places
Haugland, Ho, and Laporte (2007)	–	TS-heuristic	Partitioning and planning transport, goods delivery
Galvão, Novaes, De Cursi, and Souza (2006)	São Paulo, Brazil	Weighted frequency diagram	Partitioning areas, mail package delivery
Bozkaya, Erkut, and Laporte (2003)	Edmonton, Canada	TS-heuristic	Partitioning population areas, polling places
D'Amico, Wang, Batta, and Rump (2002)	Buffalo	SA	Partitioning population areas, police centres, urban management
Ríos-Mercado and Salazar-Acosta (2011)	Food product company	GRASP	Partitioning population areas, drink bottle collection
Yamada (2009)	Japan	GA, SA, TS	Political partitioning
García-Ayala, González-Velarde, Ríos-Mercado, and Fernández (2016)	Structural study	MILP	Arc routing districting
Ríos-Mercado and Escalante (2016)	Coke bottle distribution company	GRASP-CTDP	Commercial districting
Alawadhi and Mahalla (2015)	Kuwait	4 meta-heuristic	Political partitioning
Contreras, Fernández, and Reinelt (2012)	Benchmark instances	MILP	Combined facility location and network design
Ríos-Mercado and López-Pérez (2013)	Monterrey, Mexico	MILP	Commercial districting
Liu, Xie, and Garaix (2014)	Benchmark instances	TS	Periodic vehicle routing problem
Salazar-Aguilar, Ríos-Mercado, González-Velarde, and Molina (2012)	Business company	TS-Scatter Search & NSGAII	Commercial border specification

Table 3. Recent research on the districting problem.

Author(s)		Description
Camacho-Collados, Liberatore, and Angulo (2015)	Subject	Police districting – Spain
	Innovation	It is the first research to mention the physical characteristics of the area, risk, compactness, and insurance coverage level with regard to urban police partitioning. The decision-maker can specify his desirability value from the features based on workload balance and effectiveness
	Solution method	Heuristic methods of the greedy algorithm, the local search algorithm, and the random search algorithm
Butsch, Kalcsics, and Laporte (2014)	Subject	Partitioning – arc routing
	Innovation	It is the title of the first research that has investigated the combinatorial problem of partitioning and arc routing with a consideration of the contiguity and proximity constraints as well as some other new criteria
	Solution method	A heuristic method based on the operation of the tabu search algorithm
Jovanovic, Tuba, and Voss (2016)	Subject	Commercial partitioning – sending and delivering goods from customers
	Innovation	It is the title of the first research that has used the ant colony algorithm for the area partitioning problem
	Solution method	The ant colony algorithm
Datta, Figueira, Gourtani, and Morton (2013)	Subject	Partitioning populated areas in health system management – the southern part of England
	Innovation	Presentation of a five-objective problem with a consideration of the contiguity and absence of embedded districts constraints
	Solution method	The NSGAII algorithm
Steiner, Datta, Neto, Scarpin, and Figueira (2015)	Subject	Partitioning populated areas in treatment networks – Paraná State, Brazil
	Innovation	Presentation of a five-objective problem with a consideration of the contiguity and absence of embedded districts constraints
	Solution method	The NSGAII algorithm
Benzarti, Sahin, and Dallery (2013)	Subject	Partitioning populated areas in the domestic health system
	Innovation	Presentation of the multi-objective mathematical model of manpower allocation (without consideration of the contiguity and absence of embedded districts constraints in the partitions)
	Solution method	Solving the mathematical model and investigating with the help of randomly generated data

absence of holes) have been formalized and accurately implemented for the first time in the current model. As emphasized through the previous section, no integrated, complete mathematical model has been presented so far in the literature due to the difficulty involved in designing these constraints (Kalcsics, 2015; Salazar-Aguilar, Ríos-Mercado, & Cabrera-Ríos, 2011).

The main objective of the problem is to classify populated areas (cities and villages) into districts (hygiene networks), as well as to allocate expert teams in the provision of health services to each district based on the needs of the areas composing the district. It is noteworthy that, since this problem is a planning decision, the time frame through which demand occurs is about 2–5 years. The objective functions accounting for the heterogeneity of the services provided in the districts and the number of displacements of the expert teams across districts are simultaneously minimized. An expert team has been considered for the provision of each of the services required. After the optimal districting is specified, each expert team is located in the district with the largest number of cities demanding its services. In order to serve the other demanding districts, the team in question needs to travel from the district where it is located to all those demanding its expertise. It will be assumed that an expert team can move between the cities located in a district without incurring into any interactions with the administrative bureaucracy. In contrast, the displacement of a team across districts requires dealing with the administrative bureaucracy and spending considerable amounts of time and money.

3.1. Satisfying the contiguity and absence of embedded districts constraints

One of the prominent characteristics of the current model is that it considers the set of fundamental constraints characterizing the districting problem (contiguity, compactness, and absence of embedded districts). To the best of our knowledge, no integrated, specific mathematical model has been presented so far in the literature, since the corresponding constraints are difficult to define and implement within an optimization model (Kalcsics, 2015). For the constraints to be met, the shortest path available between every two points in a district must be located inside the very same district. This requires all the areas on the shortest path between two points to be placed within the same district. Besides assuring contiguity and avoiding unusual district allocations, this constraint generates compact partitions. Therefore, the final districts generated are expected to be convex and prevent the inclusion of any unusual allocation such as a hole.

The points on the shortest path between two nodes can be specified using standard algorithms in the field such as Dijkstra's algorithm. This algorithm is structured so that the shortest-path tree is formed if it is run for all the points in the area under investigation. The relations between points will be defined in the form of a graph network, as can be observed in the case study presented. It should be noted that there is not necessarily one communication path between each two points (corners) within the graph network used in the current research. A question that may be raised is how path selection works if there is more than one shortest path between two different nodes. To answer this question, we must consider the structure of Dijkstra's algorithm and note that the shortest-path tree is generated after running the algorithm for all the nodes in the area. Therefore, two shortest paths cannot be identified between two specific nodes, since the resulting structure would then involve a cycle, which is no longer a tree, contradicting the structure of the algorithm. It therefore follows that the algorithm implemented identifies the points on the shortest path between two nodes. The mathematical model of the problem is presented in Section 3.2 based on the requirements described.

3.2. Notations

3.2.1. Subscripts

- i, j, r — City indices
- c, c' — Expert team indices providing services of type c, c'
- p — Partition indices

3.2.2. Set

- P — Set of districts
- V — Set of cities
- T — Set of services
- H_{ij} — The set of points on the shortest path between cities i and j

3.2.3. Parameters

- V_{ic} — 1 if city i needs the expert team that provides service c, and zero otherwise
- M — A positive, sufficiently large number

3.2.4. Decision variables

- x_{ip} — 1 if city i is allocated to district p, and zero otherwise
- S_{cp} — 1 if the expert team that provides service type c is allocated to district p, zero otherwise

3.2.5. Mathematical model

$$\text{Min } Z_1 = \sum_c \sum_p S_{cp} \sum_i x_{ip} \, (1 - V_{ic}) \qquad (1(a))$$

$$\text{Min } Z_2 = \sum_c \left(\sum_p S_{cp} - 1 \right) \; s.t. \qquad (1(b))$$

$$\sum_p x_{ip} = 1 \;\; \forall i \in V \qquad (2)$$

$$\sum_i x_{ip} \ V_{ic} \leq M \ S_{cp} \quad \forall p \in P \ , \ \forall c \in T \tag{3}$$

$$x_{ip} + x_{jp} - 1 \leq x_{rp} \quad \forall i \neq j \in V, \ r \in H_{ij}, \ p \in P \tag{4}$$

$$x_{ip}, \ S_{cp} \ \in \ \{0,1\} \ \forall i \in V, \ c \in T, \ p \in P \tag{5}$$

The first objective function minimizes the value of the heterogeneity arising between the services needed and offered in the districts. In this objective function, the number of heterogeneous city needs–-service provided pairs is considered across all districts as the criterion for heterogeneity. The second objective function minimizes the total number of displacements between districts. As stated before, there is an expert team per service type provided and each team must be located in one of the districts. Thus, the number of times that each expert team needs to travel among districts is obtained by subtracting one (which accounts for the district where the team is located) from the total number of districts demanding the services provided by the team in question. The second objective function counts the number of times that an expert team must travel from the district where it is located to the other districts demanding its services.

Constraint (2) guarantees that each city is allocated to only one district. Constraint (3) sets the values of the decision variables S_{cp} as determined by the demand for services across the different districts. Constraint (4) guarantees the absence of embedded districts together with the contiguity and convexity of the partitions. According to this constraint, if two nodes i and j belong to a given district, then all the nodes located on the shortest path between them must also be assigned to the same district. Besides assuring contiguity and absence of holes, this constraint gives place to compact districts. Constraint (5) defines the range of the decision variables of the mathematical model.

We provide now a more detailed description of the intuition on which the two objective functions are built and emphasize their complementarity. Consider the first objective function. It has been designed to minimize the heterogeneity of the services demanded by the cities composing the different districts. That is, note that whenever an expert team providing a given service is allocated to a district, that is, when $S_{cp} = 1$, a set of potential requirements for this service emerges across the cities composing the district. The diversity of these requirements is captured by the expression $\sum_i x_{ip} \ (1- \ V_{ic})$, which accounts explicitly for the cities within the district that do not demand a given service. More precisely, this expression lists all the cities within the district and assigns them a value of zero if they require a given service and a value of one if they do not.

Whenever the model adds cities to a given district, it is constrained by the fact that their needs will have to be fulfilled, as required by Equation (3). As a result, a large diversity of cities requiring a unique service each would create a highly heterogeneous set, leading to a substantial value of Equation (1(a)). Ideally, all the cities within a given district should require all the services being assigned to it. Note how the objective function defined in (1(a)) focuses on the homogeneity of the services that must be provided within a given district. As a result, this equation does not impose any restriction on the distribution of expert teams across districts. Equation (1(b)) has been defined to tackle this second requirement. In this case, we focus on the distribution of expert teams across districts. Once again, Equation (1(a)) does not impose any constraint on their distribution as long as the needs of the cities composing the districts are satisfied. This implies that a distribution where all the expert teams are allocated to a unique district and another one where they are evenly distributed across all the districts would deliver identical minimization results. In order to prevent such a logistic shortcoming, which could be quantified in terms of transportation costs, we have introduced Equation (1(b)) as the second objective function of the problem.

In this case, we account explicitly for the travel requirements of each team of experts across districts. That is, the team of experts located in a given district will have to travel to all of those districts where cities requiring their services have been allocated. The travel requirements of a team across districts are captured by the expression $\sum_p S_{cp} - 1$ within Equation (1(b)), which delivers a value equal to the total number of districts minus one if a team has to travel to all the districts (including the one where it has been located) in order to provide its service. On the other hand, if the team does not have to travel outside its own district, the value of this expression will be equal to zero, namely, the district where the team is located minus one. Note that this equation has been introduced in order for the model to favour distributions of cities across districts that constrain the number of trips that the teams have to make. That is, between two distributions of expert teams across the same number of districts, this constraint favours the distribution leading to a lower number of displacements of the teams across districts.

Note that Equation (1(b)) does not only account for the distribution of teams across districts. The distribution of cities across districts is also implicitly considered since it determines the trips that must be performed by the different teams. Thus, Equation (1(b)) complements the constraint imposed by

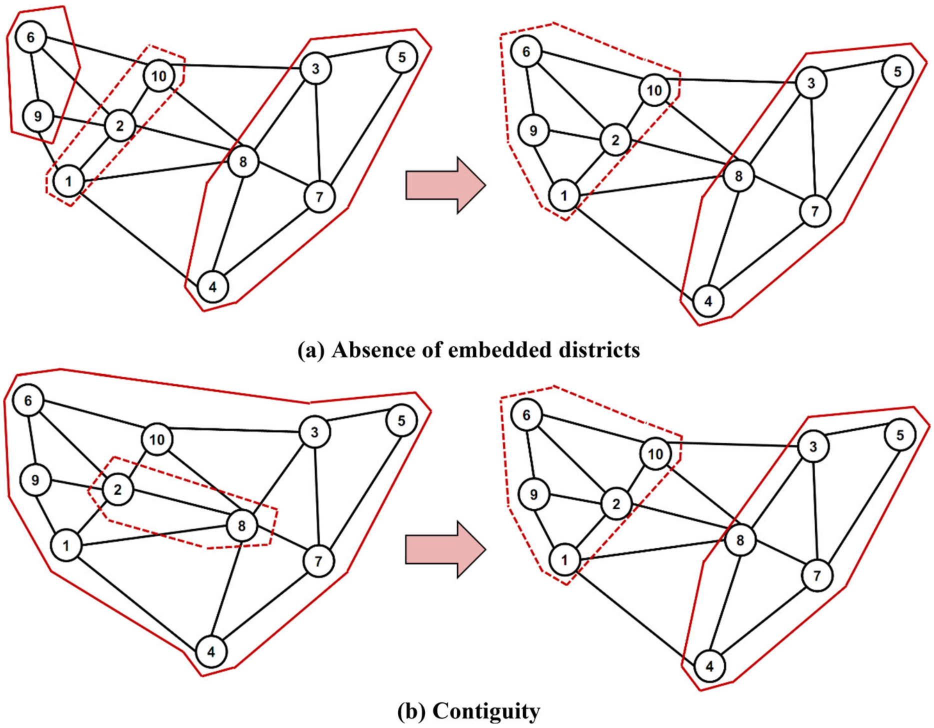


Figure 1. Contiguity and absence of embedded districts constraints.

Equation (1(a)) when defining the trips that must be performed by each team across districts. The functions described in both equations are required to guarantee that cities and teams simultaneously minimize the potential heterogeneity of the districts, which would lead to an inefficient use of expert teams' displacements, and the number of trips that must be performed by each team, which is determined by the location selected for each team and the distribution of cities across districts.

The complementary structure defined by Equations (1(a)) and (1(b)) determines the distribution of cities and teams across districts. However, the main novelty of the current paper is given by Equation (4), which distributes cities across districts by considering the shortest path between every two city nodes within each district. Based on the literature review performed, Constraint (4) constitutes – to the best of our knowledge – the first mathematical form presented to assure contiguity, compactness, and absence of embedded districts. Figure 1 provides a representation of how this constraint operates. As illustrated in Figure 1(a), a district cannot be created inside another since it would block the communication paths between districts and violate the convexity requirement. Similarly, Figure 1(b) describes the lack of contiguity between the areas of a district, a potential scenario prevented by the implementation of Constraint (4).

A potential problem regarding the practical implementation of this constraint is that too many constraints will actually be generated given its current structure. More precisely, assume that there are 10 points on the shortest path between points A and B. Consequently, the set $\{H_{AB} | A, B \in V\}$ will contain 10 members and Equation (4) will generate one constraint for each of the members. However, the description of this equation seems to imply the existence of a unique constraint. The structure of Constraint (4) has therefore been improved and presented as Constraint (6)

$$\sum_{r \in H_{ij}} x_{rp} \geq SP_{ij} - M\big(2-(x_{jp}+x_{ip})\big), \ \ \forall i \neq j \in \ V, \ \ p \in P \tag{6}$$

where SP_{ij} denotes the number of points on the shortest path between i and j. This equation generates a constraint for the whole set of points on the shortest path instead of generating a constraint for each of the points on the path. Thus, the mathematical model of the problem consists of the first and second objective functions, (1(a)) and (1(b)), together with Constraints (2), (3), and (6). Note that the first objective function is nonlinear due to an expression containing the product of two binary variables. It is possible to linearize the corresponding equation by defining a new decision variable.

We conclude this section by noting that the formalization closest to the model presented in the current paper is that of Steiner, Datta, Neto, Scarpin, and Figueira (2015). One of the main features on which the current model improves upon theirs is the fact that in their model the number of districts is a parameter while in the current framework it is selected as part of the optimization process. Moreover, even though the objectives defined are similar – due to the fact that both models formalize a districting problem – Steiner et al. (2015) focus on average differences in the heterogeneity of districts and the services provided within them. Furthermore, the set of constraints on the districting process are all introduced separately, requiring a complex amount of additional equations, while we are able to incorporate all the constraints within a unique compact expression. Finally, the mathematical structure of the current model considers the displacement of teams across districts, a feature that combined with the traveling distance minimization objective of Steiner et al. (2015) could deliver interesting results in future research.

Regarding recent developments in districting and allocation problems, we describe the main related opuses below while remarking that none of them follows a formalization process similar to the one just introduced in the current paper.

Steiner Neto et al. (2017) apply the same formal approach as Steiner et al. (2015) to determine the location of grain silos in the Brazilian state of Paraná. In particular, they minimize the difference between storage capacity and production from their mean values per region and the product transportation cost across region municipalities. A different approach is applied by De Barros Franco & Arns Steiner (2018) to categorize abandoned locations within two clusters based on their potential to host facilities to capture solar energy. These authors use a hybrid fuzzy c-means algorithm initialized by three meta-heuristics to distribute different locations between both clusters.

Among the recent multi-objective location–allocation environments analysed in the literature, we would like to highlight the research of Yanık, Sürer, and Öztayşi (2016) and that of Khodaparasti et al. (2017). The former authors apply a multi-objective genetic algorithm to cluster geographic areas so that their energy requirements match the capacity of the corresponding clusters. They follow an iterative location–allocation strategy where the location of the clusters is selected using a genetic algorithm while the allocation follows from a binary integer programming model that minimizes the total distance between basic units and region centres. On the other hand, Khodaparasti et al. (2017) focus on the ease of access to a facility as one of the main factors conditioning the commitment to the programs offered by community based organizations. Accessibility is determined by the number, type and location of the first and second level facilities available. These authors design a location–allocation model within a multi-objective framework to account for the social welfare features of equity, accessibility, and efficiency.

Finally, Zhou, Geng, Jiang, and Wang (2018) tackle a queuing model for a hospital with stochastic arrivals and lengths of stay among patients, who must be allocated to the different wards without decreasing revenues. Zhou et al. (2018) design a multi-objective stochastic programming model to maximize revenue and equity, focusing on the performance of the system and its patient admission and allocation capacities.

4. Solution method

Due to the conflict existing between the objective functions, it is impossible to obtain a solution that simultaneously optimizes both objectives. Therefore, the main purpose is to obtain a set of efficient or non-dominant Pareto points. Several methods have been introduced to solve multi-objective mathematical models, such as simple additive weighting, goal programming, as well as the epsilon constraint and LP metric methods.

In the current paper, the epsilon method is used to solve the proposed mathematical model. In order to obtain the Pareto front applying this method, the first objective function is considered as the main objective of the problem, and the second one as a constraint restricted to a specific ε value. The model can therefore be rewritten as follows:

$$\text{Min } Z_1 = \sum_c \sum_p S_{cp} \sum_i x_{ip} \ (1 - V_{ic}), \ s.t. \tag{1(a)}$$

$$\sum_c \left(\sum_p S_{cp} - 1 \right) \leq \varepsilon, \tag{7}$$

$$Z_2^* \leq \varepsilon \leq Z_2\left(Z_1^*\right) \tag{8}$$

(2, 3, 5, 6).

All the Pareto front points can be identified iteratively after solving the above mathematical model aided by a standard optimization software.

4.1. Proposed algorithms

Since the districting problem is an NP-hard one (Steiner, Datta, Neto, Scarpin, & Figueira, 2015), the optimal solution cannot be obtained within the proper time for large-scale problems. Meta-heuristic algorithms are used to overcome this problem and obtain acceptable solutions within proper times. Most studies dealing with districting problem across different fields have applied the genetic algorithm (Kim, Hwang, Kim, & Moon, 2011). Similarly, in this paper, two meta-heuristic algorithms, namely, a multi-objective genetic algorithm II (NSGAII) and a multi-objective grey wolf optimizer (MOGWO), are implemented to solve the problem in large scale instances.

4.2. Chromosome representation and initialization

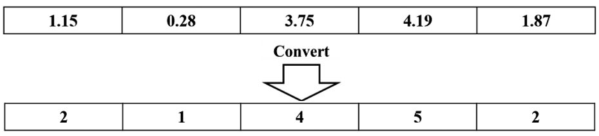
The algorithm solution chosen for the current healthcare system districting problem is an array of basic units. The value of a solution element consists of the districts where the represented basic units belong. More precisely, the structure of the chromosomes is shown below.

1	4	2	1	3	2	4	3

Each gene identifies a district number while the position of a gene represents a basic unit in a chromosome. The length of a chromosome corresponds to the number of basic units. For example, the above chromosome (string) displays a four-district solution with the following basic units in each district:

District 1: Basic units 1 and 4
District 2: Basic units 3 and 6
District 3: Basic units 5 and 8
District 4: Basic units 2 and 7

To impose a continuous representation, the value of each gene is generated from $[1, |P|+1)$ and converted into an integer number using a ceiling function. For example, let $|P| = 5$. The corresponding representation generates chromosomes as follows:



We apply a greedy algorithm to initialize the algorithm solution since it is rather difficult to obtain a feasible solution in graph partitioning problems using random assignments. In addition, a labelling mechanism, which relabels a disconnected portion of a district as a new district, is applied if the contiguity constraint is violated when implementing the proposed crossover operator. However, the remaining constraints could be violated at various stages through the initialization process or when generating a given solution. Consequently, a constructive/repairing mechanism is applied to ensure that those constraints are sufficiently satisfied according to the guidelines of the algorithm proposed by Steiner, Datta, Neto, Scarpin, and Figueira (2015).

4.3. Multi-objective genetic algorithm (NSGAII)

Many evolutionary algorithms (especially from the genetic algorithm family) have been developed so as to solve the multi-objective optimization problems tackled in different research areas. One of the most common multi-objective meta-heuristic algorithms is the second version of the multi-objective genetic algorithm based on non-dominance sorting (NSGAII) proposed by Deb, Pratap, Agarwal, and Meyarivan (2002). The most important feature of this algorithm is the sorting solution designed based on the dominance criteria. In each generation, the chromosomes that are not dominated by other chromosomes are labelled as the first level of dominance. The process of forming dominance levels, the temporary removal of chromosomes, and the search for the non-dominated chromosomes continues until all the solutions are graded.

The NSGAII algorithm is structured as Pseudo-code1 below. According to the chromosome structure, the initial population is generated randomly or systematically. In each generation, parents are categorized for crossover and mutation using binary selection operators. Two parents are selected randomly each time and compared *via* the dominate level or crowding distance. A specified number of chromosomes is selected to generate offspring by repeating this operator through each generation. The final step consists of selecting the population for the next iteration, a strategy based on elitist selection. Following to this strategy, the first chromosomes in the mating pool are sorted based on the dominate level or crowding distance. This operation continues until the stopping criteria are satisfied.

Pseudo-code1: NSGAII

```
Initialize Population
Generate N feasiable solution and insert into Population
  While Stopping criteria not met Do
    Generate ChildPopulation of Size N
    Select Parents from Population
    Create Children from Parents
    Mutate Children
    Repare Solution using repair mechanism
    Merge Population and ChildPopulation with size 2N
      For each individual in CurrentPopulation Do
      Assign rank based on Pareto-Fast non-dominates sort
      end
    Generate sets of non-dominated vector along PF_known
    Loop (inside) by adding solution to the next generation of
    Population starting from the best front
    Until N solution are found and determine crowding distance
    between points on each front
  end
Report results
```

4.4. Multi-Objective GWO (MOGWO)

Two new processes are applied to perform multi-objective optimization with GWO. The first one involves archiving the non-dominated Pareto optimal solutions. The second process involves selecting a leader to choose alpha, beta, and delta solutions from the archived list. An archive controller is used to control the archive. The non-dominated solutions are compared against the archive through the course of iterations (Mirjalili, Saremi, Mirjalili, and Coelho, 2016).

Pseudo-code2: multi-objective gray wolf optimization algorithm (MOGWO)

Initialize the grey wolf population
Initialize a, A, and C
Calculate the objective values for each search agent
Find the non-dominated solutions and initialized the archive with them
X_α = *Select Leader (archive)*
Exclude alpha from the archive temporarily to avoid selecting the same leader
X_β = *Select Leader (archive)*
Exclude beta from the archive temporarily to avoid selecting the same leader
X_δ = *Select Leader (archive)*
Add back alpha and beta to the archive
$t = 1$
    While *($t <$ Max number of iterations)*
        For *each search agent*
        Update the position of the current search agent
        Repare Solution using repair mechanism
        End
    Update a, A, and C
    Calculate the objective values of all search agents
    Find the non-dominated solutions
    Update the archive with respect to the obtained non-dominated solutions
        If *the archive is full*
        Run the grid mechanism to omit one of the current archive members
        Add the new solution to the archive
        End
            If *any of the new added solutions to the archive is located outside the hypercube*
            Update the grids to cover the new solution(s)
            End
    X_α = *Select Leader (archive)*
    Exclude alpha from the archive temporarily to avoid selecting the same leader
    X_β = *Select Leader (archive)*
    Exclude beta from the archive temporarily to avoid selecting the same leader
    X_δ = *Select Leader (archive)*
    Add back alpha and beta to the archive
    $t = t + 1$
    End while
Return ***archive***

5. Numerical example

In this section, several numerical examples are examined to validate the mathematical model and assess the performance of the proposed algorithms. The quality of the solutions generated by the algorithms is compared to the quality of those obtained from CPLEX 12.5.1. Due to the lack of sample cases, numerical examples have been generated randomly. The proposed meta-heuristic algorithms, coded in the C# environment, were run using a personal computer with a central Core i7 series processor with a processing power of 3.2 GHz and 16 GB RAM.

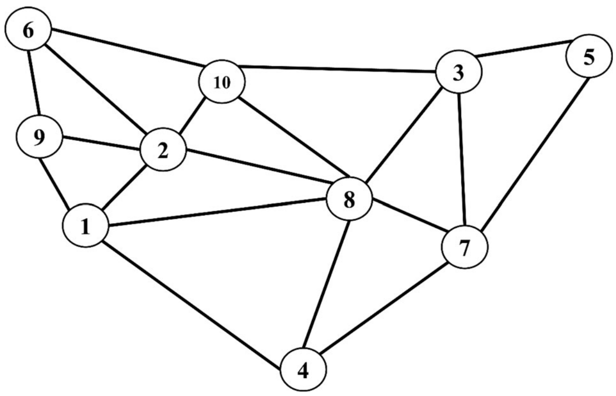


Figure 2. Geographic locations of the cities in the numerical example.

Table 4. Types of services needed in the cities within the first example.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
C1	1	1	0	0	1	0	0	0	1	0	1	0	0	0	1
C2	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0
C3	1	1	1	1	1	1	0	0	0	0	1	0	0	0	1
C4	1	0	0	0	0	0	1	1	0	0	0	1	1	1	1
C5	0	0	1	1	0	1	1	0	0	1	0	1	1	1	1
C6	0	1	1	1	0	1	1	0	1	0	1	1	0	0	0
C7	1	0	1	1	0	0	0	0	0	1	1	1	0	0	0
C8	0	1	1	1	0	0	0	0	0	0	1	1	0	0	0
C9	1	1	1	0	1	1	0	0	0	0	1	1	0	1	0
C10	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0

5.1. Validation of the mathematical model

In order to validate the mathematical model, an example area with 10 cities demanding a total of 15 medical service types is examined. The geographic location of the cities is illustrated in Figure 2 and the type of services needed in these cities is described in Table 4. These needs are assumed to be determined by the viewpoints of experts from organizations in charge of public health.

After solving the example using the CPLEX solver, the results are presented in accordance with the structure of the heuristic method defined in Section 4 to obtain the Pareto front. In order to allow for a more accurate examination of the model performance and the behaviour of Constraint (6), the Pareto front is presented in Figure 3 and the solutions described in each of the front points are examined afterwards.

As shown in Figure 3, none of the Pareto front members dominates another. A new non-dominated point is created as the number of partitions is modified in each iteration of the solution method. This result may indeed reflect the efficiency of the proposed heuristic method for finding the Pareto points. It is also clear that as one objective function is improved, the other one is worsened. Another important feature of the model regards its run time per iteration. Figure 4 illustrates how the solution time of the model increases in the number of

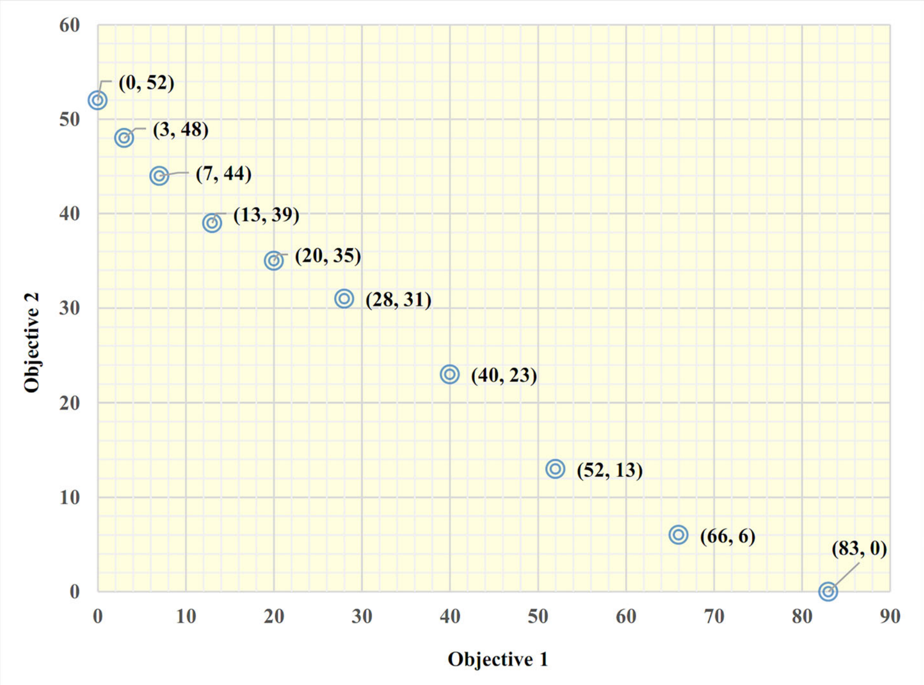


Figure 3. Non-dominated solutions obtained by CPLEX.

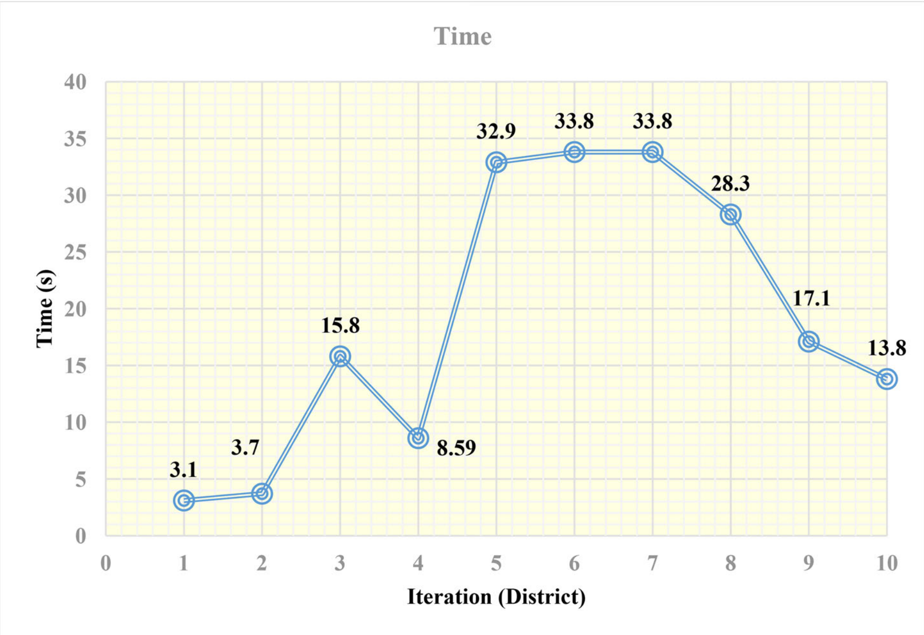


Figure 4. CPU time for each iteration (district).

partitions through consecutive iterations. This is of course predictable since an increment in the number of partitions increases the complexity of the problem when determining the optimal allocations. On the other hand, as the number of partitions converges to the number of cities, the solution time decreases, reflecting a decrease in the complexity level of the problem with a very large or very small number of partitions.

In order to provide additional intuition regarding the performance of the model and the behaviour of Constraint (6), the solution defined by point (13, 52) in the Pareto front of Figure 3 – and consisting of a total of three districts – is explicitly presented in Figure 5. As can be observed, cities 1, 2, 6, 9, and 10 have been allocated to District 1, cities 4, 7, and 8 to District 2, and cities 3 and 5 to District 3. The features of compactness, contiguity, and absence of embedded districts can also be identified. Given the objective functions of the problem, the cities have been allocated to the corresponding districts so that heterogeneity (namely, the first objective function) and displacement (the second objective function) have the lowest possible values. It can therefore be stated that Constraint (6) is sufficiently efficient to assure the creation of compact, contiguous districts

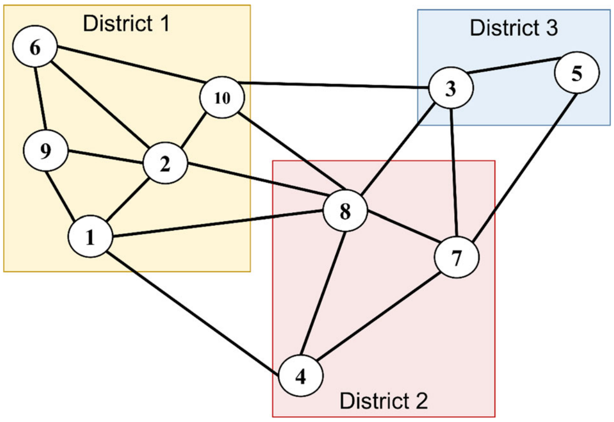


Figure 5. Optimal partitioning with three districts (for Point (13, 52) in Figure 3).

Table 5. Allocation of services to the partitions.

District 1	{1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14}
District 2	{1, 2, 7, 8, 12, 13, 14, 15}
District 3	{1, 2, 3, 4, 10, 11, 12}

without uncommon allocations (i.e. guaranteeing the absence of embedded districts). The other important feature of the model concerns the way in which services are allocated to the districts, as described in Table 5.

Note how the number of expert teams allocated to District 1 is greater than those assigned to the other two districts. At the same time, the first district contains a larger number of cities. The model has minimized the value of the objective functions by locating cities with a demand for common services in one district, while assigning the expert teams to the districts so that the lowest possible number of displacements takes place. Clearly, the partitioning of districts and the allocation of expert teams will differ if a different Pareto front point is examined.

5.2. Evaluation of the efficiency of the algorithms

In this section, several numerical examples are randomly generated to evaluate the effectiveness of the algorithms. The proposed algorithms are run 10 times for each test problem and the best solutions obtained in the executions are compared.

5.2.1. Evaluation metrics

In order to perform comparisons regarding the quality of the different Pareto-optimal fronts produced by the optimization algorithms, we consider the following performance assessment metrics.a. *Convergence measure:* it calculates the distance between the Pareto-algorithm front and the Pareto-optimal front as follows:

$$\gamma = \frac{\sum_{i=1}^{|Q|} d_i}{|Q|}, \tag{9}$$

where d_i is the minimum Euclidean distance between solution $i \in Q$ and the best solution, and Q is the final Pareto front (Datta & Figueira, 2012).b. *Dispersion measure:* it is considered a measure of the distribution around the Pareto-optimal fronts and is calculated through the following equation:

$$\Delta = \frac{\sum_{m=1}^{M} d_m^e + \sum_{i=1}^{|Q|} \left| d_i - \bar{d} \right|}{\sum_{m=1}^{M} d_m^e + |Q|\bar{d}}, \tag{10}$$

where Q is the final Pareto front, m is the number of solutions in the Pareto-optimal front, d_i is the minimum Euclidean distance between solution $i \in Q$ and the best solution, $\bar{d}$ is the average of the d_i values, d_m^e is the distance between the best Pareto front value and solution $i \in Q$ in the objective function m. We should note that this measure works best when the distribution is uniform (Coello, Lamont, & Van Veldhuizen, 2007).c. *Diversification metric:* it is used to measure the spread of a non-dominated solution set as follows:

$$DM = \sqrt{\sum_{i=1}^{n} \max\left(\left| \left| x_i' - y_i' \right| \right| \right)}, \tag{11}$$

where n is the number of Pareto front members, and $||x_i' - y_i'||$ is the Euclidean distance between the non-dominated solutions x_i and y_i (Coello, Lamont, & Van Veldhuizen, 2007).d. *Spread of non-dominant solution (SNS)*: it is considered as a diversity measure and evaluates the standard deviation of the distance from a non-dominated set to an ideal point. The SNS is calculated as:

$$SNS = \sqrt{\frac{\sum_{i=1}^{n} (\bar{c} - c_i)^2}{n-1}}. \tag{12}$$

e. *Data envelopment analysis (DEA)*: it is usually applied to evaluate the performance of different choices according to some pre-specified attributes. In the current paper, we consider each Pareto-optimal solution as a decision-making unit and use DEA to measure the efficiency of the non-dominated solutions obtained by each method. The values of the objective functions are used as inputs and outputs. Subsequently, all the non-dominated solutions are combined and DEA is implemented to measure their efficiency (Coello, Lamont, & Van Veldhuizen, 2007).

5.2.2. Comparisons among the algorithms

In this section, the efficiency of each algorithm is assessed using the above evaluation metrics to determine whether there is a significant difference among their performances. As shown in Table 6, the MOGWO algorithm is superior to the NSGAII one in all the evaluation metrics. This superiority is

Table 6. The metrics obtained for the performance of the algorithms.

		Algorithm 1: NSAGII					Algorithm 2: MOGWO				
Instance	Size	γ	Δ	DM	SNS	DEA	γ	Δ	DM	SNS	DEA
Ins.1	Small	0	N/A	127.13	94.31	0.66	0	N/A	137.43	112.11	0.69
Ins.2		0	N/A	136.89	92.76	0.53	0	N/A	151.99	111.26	0.61
Ins.3		0.00867	0.8749	85.26	105.46	0.53	0	N/A	103.86	125.86	0.59
Ins.4		0.00892	0.6902	157.18	82.93	0.61	0	N/A	172.58	106.73	0.65
Ins.5		0.00992	0.7584	109.01	87.64	0.59	0	N/A	124.81	107.74	0.65
Ins.6	Medium	0.01238	0.7689	160.4	65.27	0.64	0.00511	0.4488	172	78.87	0.68
Ins.7		0.01413	0.8293	146.77	57.06	0.59	0.00625	0.5947	163.97	70.36	0.65
Ins.8		0.01491	0.8091	156.96	101.45	0.51	0.0065	0.4599	167.76	115.75	0.54
Ins.9		0.01067	0.6524	90.15	57.58	0.57	0.00602	0.4970	108.85	68.68	0.6
Ins.10		–	–	96	71.76	0.64	–	–	112.2	86.66	0.72
Ins.11		–	–	114.25	105.71	0.5	–	–	129.15	122.21	0.55
Ins.12		–	–	126.25	72.26	0.66	–	–	141.35	83.56	0.69
Ins.13		–	–	158.7	100.52	0.45	–	–	180.6	121.72	0.53
Ins.14		–	–	73.57	96.73	0.68	–	–	95.37	108.03	0.71
Ins.15		–	–	159.4	105.83	0.53	–	–	180	125.73	0.58
Ins.16	Large	–	–	122.8	65.14	0.49	–	–	140.4	86.04	0.58
Ins.17		–	–	157.66	56.87	0.47	–	–	173.16	72.47	0.54
Ins.18		–	–	129.26	70.26	0.65	–	–	139.56	94.16	0.68
Ins.19		–	–	141.94	51.01	0.63	–	–	163.54	62.31	0.67
Ins.20		–	–	144.64	59.94	0.44	–	–	165.94	83.84	0.53
Ins.21		–	–	91.34	101.13	0.51	–	–	113.44	115.03	0.58
Ins.22		–	–	112.37	72.81	0.49	–	–	133.97	93.71	0.52
Ins.23		–	–	159.52	72.58	0.6	–	–	175.62	94.38	0.67
Ins.24		–	–	157.57	75.2	0.65	–	–	178.97	98.7	0.7
Ins.25		–	–	137.64	90.51	0.63	–	–	149.34	109.31	0.68
Ins.26		–	–	141.1	102.93	0.56	–	–	162.3	122.03	0.62
Ins.27		–	–	74.41	60.05	0.54	–	–	97.21	71.25	0.63
Ins.28		–	–	82.23	58.51	0.5	–	–	104.63	74.91	0.54
Ins.29		–	–	88.8	81.96	0.55	–	–	106.9	100.56	0.63
Ins.30		–	–	72.1	88.48	0.51	–	–	94	110.48	0.58

more prominent in the medium and large instances. Although there is also a slight superiority in the small instances, absolute superiority is not observed in this scale relative to NSGAII. Another point to be noted concerns the values of (γ) and (Δ) in instances 10–30. Since CPLEX is incapable of defining the optimal front in these instances, it is not possible to calculate the values of (γ) and (Δ).

The Pareto fronts obtained from solving the instances in different scales are presented in Figure 6. When the problem size is small, the solutions provided by CPLEX, NSGAII, and MOGWO are similar, as illustrated in Figure 6(a). This result suggests the appropriate efficiency of the algorithms in solving small-scale problems, as also described in Table 6. It can be observed in Figure 6(b), where the Pareto front obtained from solving a medium scale problem is displayed, that the solutions provided by MOGWO are closer to those of the CPLEX Pareto front. Some of the points in the NSGAII Pareto front are close to those of the CPLEX algorithm, but the NSGAII Pareto front displays a general lower quality than that of MOGWO. In Figure 6(c), only the solutions offered by NSGAII and MOGWO are presented, since CPLEX does not provide results in this scale due to the high complexity of the problem. As can be observed, some of the points in the NSGAII and MOGWO Pareto fronts are close to each other, particularly at the initial and final sections of the fronts. The reason for such similarities is quite possibly the fact that the complexity of the problem is relatively low at the initial and final points of the front, which, as illustrated in Figure 3, display a lower solution time than the middle points. All in all, the main result obtained from these comparisons is the fact that the MOGWO algorithm obtains more appropriate results than NSGAII.

We continue the analysis of the behaviour of the algorithms in solving the test instances by examining and comparing their performances in terms of solution process and solution time. One of the main factors determining the efficiency of an algorithm is the rate at which it converges into the final solution for each Pareto front member. Figure 7 presents a graphical comparison of this factor for the same members of the Pareto front in large-scale instances. It can be observed that the MOGWO algorithm converges much faster than NSGAII and obtains solutions of higher quality. In particular, the difference between the fitness of the best member and the mean fitness in each iteration is insignificant in the MOGWO algorithm, whereas it cannot be ignored in the NSGAII case. Figure 8 confirms the above conclusions. In this figure, the variance among the solutions obtained in each population illustrates the superiority of the MOGWO algorithm in finding the appropriate solution space and transferring the solutions from the exploration to the exploitation phase.

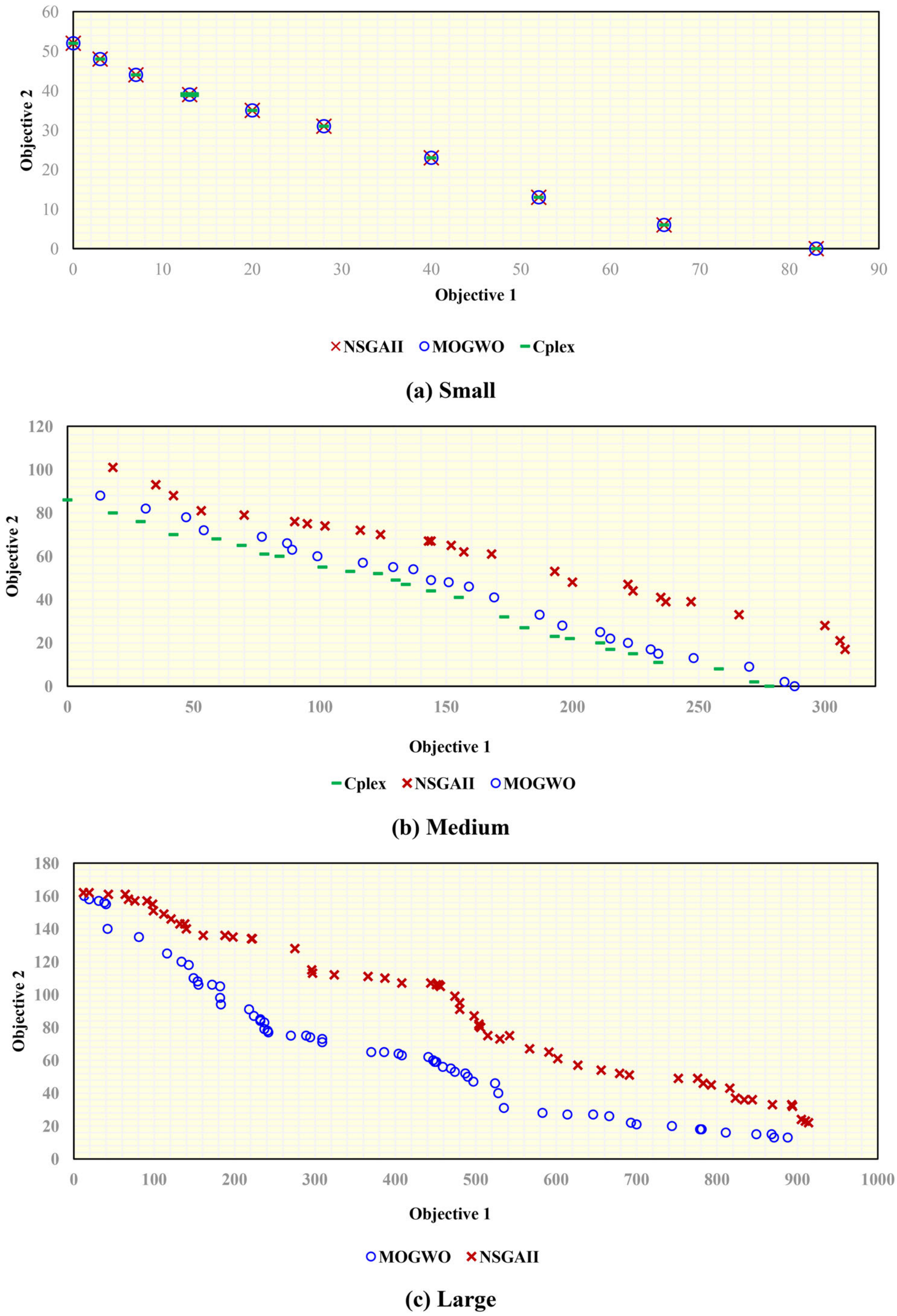


Figure 6. Dispersion of non-dominated solutions obtained by different algorithms in three scales.

We proceed now with the analysis of the case study and therefore focus on the health system in the South Khorasan Province of Iran.

6. The case study

The South Khorasan Province (SKP) is one of the eastern provinces of Iran and its capital is the county of Birjand. The area of the province equals 151,193 $(\mathrm{km})^2$, which makes it the third largest province in Iran in terms of land area. Based on the most recent information issued, is has a population of about 800,000 people. The province includes 11 counties, 25 cities, 61 rural divisions, and 28 towns. Figure 9 provides an illustrative view of the geographic location and structure of the province. Given its geographic location and shared borders with Afghanistan and the Provinces of Sistan & Baluchestan and Kerman, it has a high ethnic diversity compared to the other Iranian provinces. Its population diversity together with the dry desert climate foster the potential occurrence and prevalence of different diseases within the province. Managing its health system is particularly challenging since the tasks that must be performed require a considerable amount of coordination and discipline.

A health-monitoring plan has been periodically implemented in different areas within SKP since

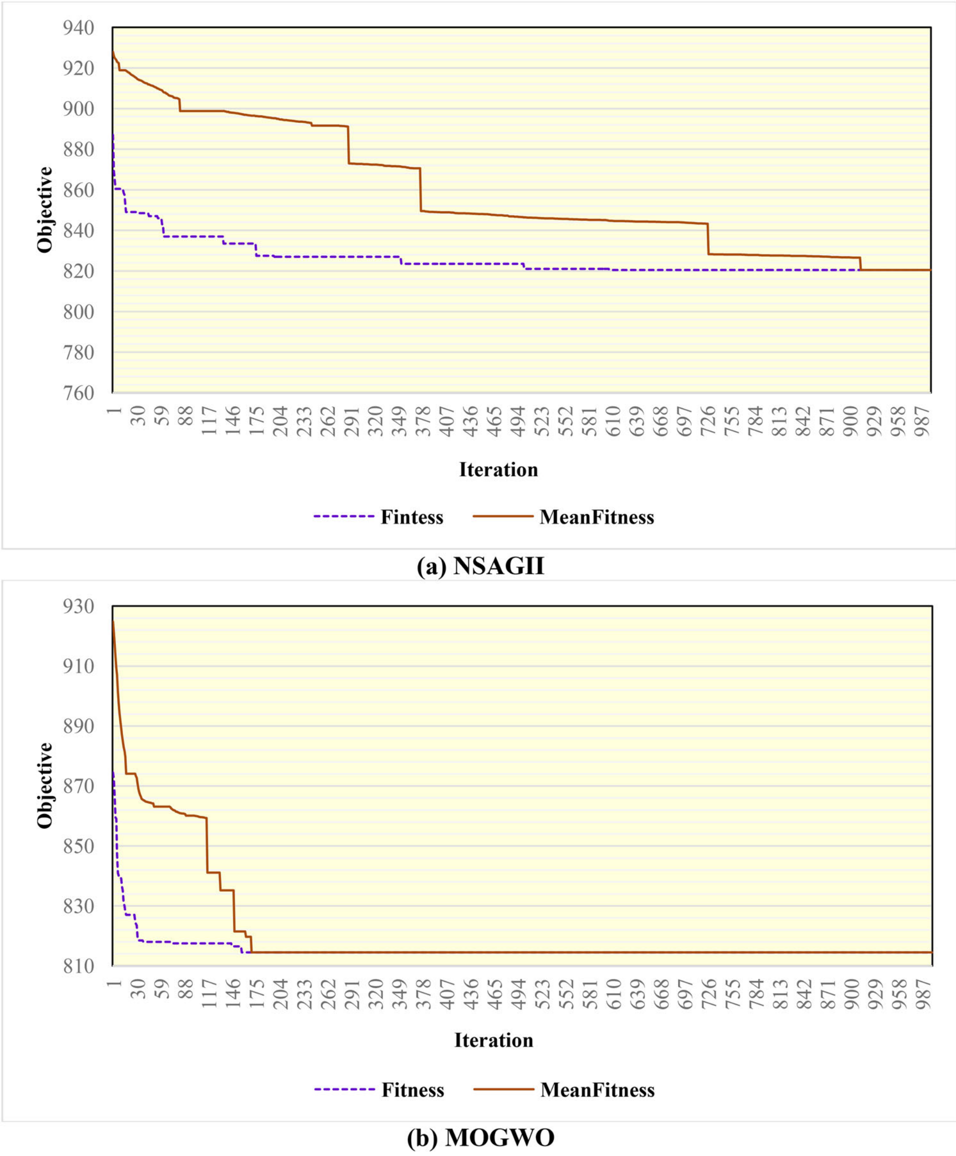


Figure 7. Comparison of the convergence of the objective function and the mean objective function for NSAGII and MOGWO.

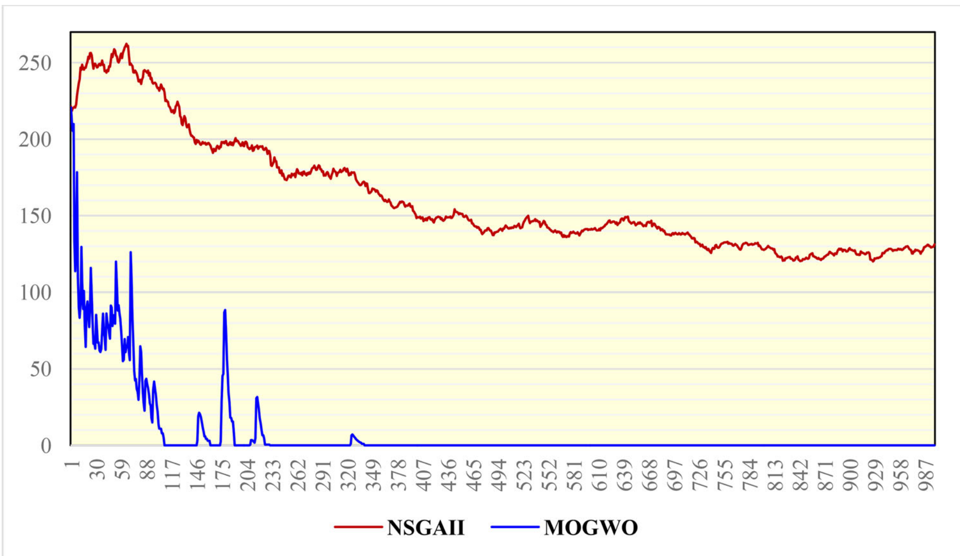


Figure 8. Comparison of diversification in the NSGAII and MOGWO algorithms.

2011. Based on the reports available, the implementation of the plan has increased the societal health level and reduced the anomalies resulting from unfamiliarity with primary health care. However, one of the main concerns of the managers has always been to increase the coverage level as far as possible and promote the quality of service provision along with reducing the costs of the system. In 2015, an ambitious plan was implemented in 82 populated areas of the province accounting for a total of 27 health services. The input information is presented in Figure 9 and Table 7. The complete information is accessible through the systems of the Statistical Centre of Iran, the Department of Roads and Urban Planning of the province, and the Deputy of Health Networks of the Birjand University of Medical Sciences.[1]

Given its superior performance relative to the NSGAII one, the MOGWO algorithm will be applied to solve the above case study. We will first present the Pareto front generated by the algorithm and then select one of its points so as to examine the resulting structure in detail.

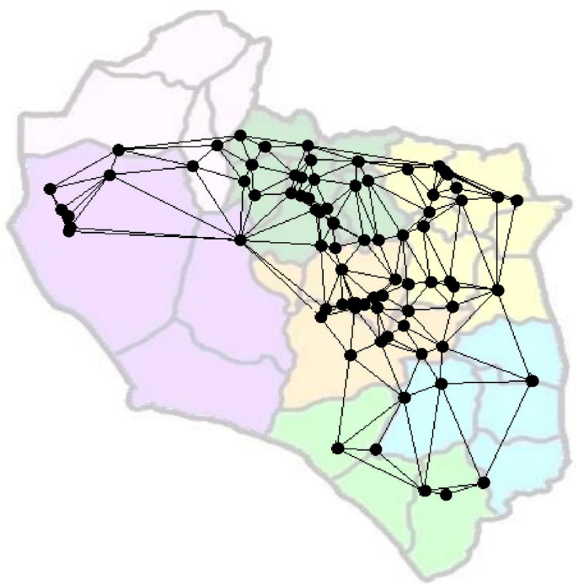


Figure 9. A view of the south Khorasan province map and the populated points connectivities based on a non-crossing edges graph.

6.1. Results and discussion

Based on the method presented to obtain non-dominated solutions, the number of Pareto front members equals the number of populated points. The run time of the algorithm for a population size of 800 and 3000 generations is 5 h 37 min. The Pareto front obtained contains 82 points, with $(obj_1, obj_2) = (7, 1051), (1127, 81)$ representing the best values relative to each objective function. According to the information available, the objective function values derived from the structure used to implement the health monitoring plan in 2015 are given by $(obj_1, obj_2) = (421, 1344)$. The corresponding structure contains a total of nine partitions. Table 8 presents the two extreme Pareto front solutions together with the one being implemented, the compromised solution selected and its improvement over the existing one.

Although specific information is required to select one solution from the members of the Pareto front, each point potentially selected by the managers can be considered as the final solution for comparison purposes with the solution being currently implemented. In this regard, point $(obj_1, obj_2) = (314, 775)$, endowed with a similar structure to that of the existing solution, is selected as the compromised solution. This solution improves the first objective function by 34.1% and the second one by 73.4% with respect to the existing solution.

Figure 10 displays the (non-dominated members composing the) Pareto front obtained when applying the MOGWO algorithm together with the existing and compromised solution selected. Note that the existing solution is dominated by all the points in the Pareto front, implying that the compromised solution is clearly superior to the existing one.

Table 8. Extreme solutions from the Pareto front, existing solution, and compromised solution.

Objective functions	f_1	f_2
Solution with the best (minimum) value of f_1	7	1051
Solution with the best (minimum) value of f_2	1127	81
Existing solution (ES)	421	1344
Compromised solution (CS)	314	775
Improvement in the compromised solution $= \frac{ES-CS}{CS} \times 100$	0.341	0.734

Table 7. Services needed by the cities.

	A1	A2	A3	A4	A5	A6	A7	…	A74	A75	A76	A77	A78	A79	A80	A81	A82
C1	1	1	0	1	1	1	1	…	1	1	0	0	1	0	0	0	0
C2	0	1	0	0	0	0	0		0	0	1	0	1	0	1	1	0
C3	1	1	1	1	0	1	1		0	0	0	1	0	1	1	0	0
C4	0	0	0	1	0	0	0		0	1	1	1	0	1	1	0	0
C5	1	1	1	1	1	0	0		1	0	1	1	0	1	0	0	1
…				…									…				
C23	1	0	1	1	0	0	0	…	1	0	1	0	1	0	0	0	0
C24	1	0	1	0	1	0	1		0	0	0	0	1	1	0	0	1
C25	1	1	0	0	1	1	1		0	0	0	1	0	1	1	0	1
C26	0	1	0	1	0	0	1		1	1	0	0	1	1	1	1	1
C27	1	0	0	1	1	1	0		1	0	1	0	0	1	1	1	1

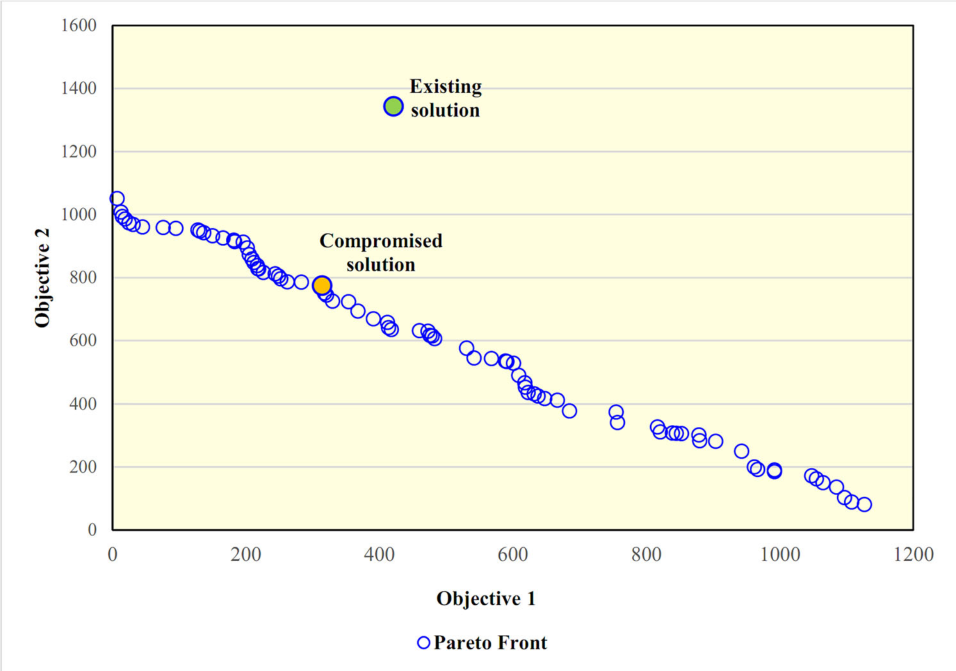


Figure 10. Pareto front along with the existing and compromised solutions.

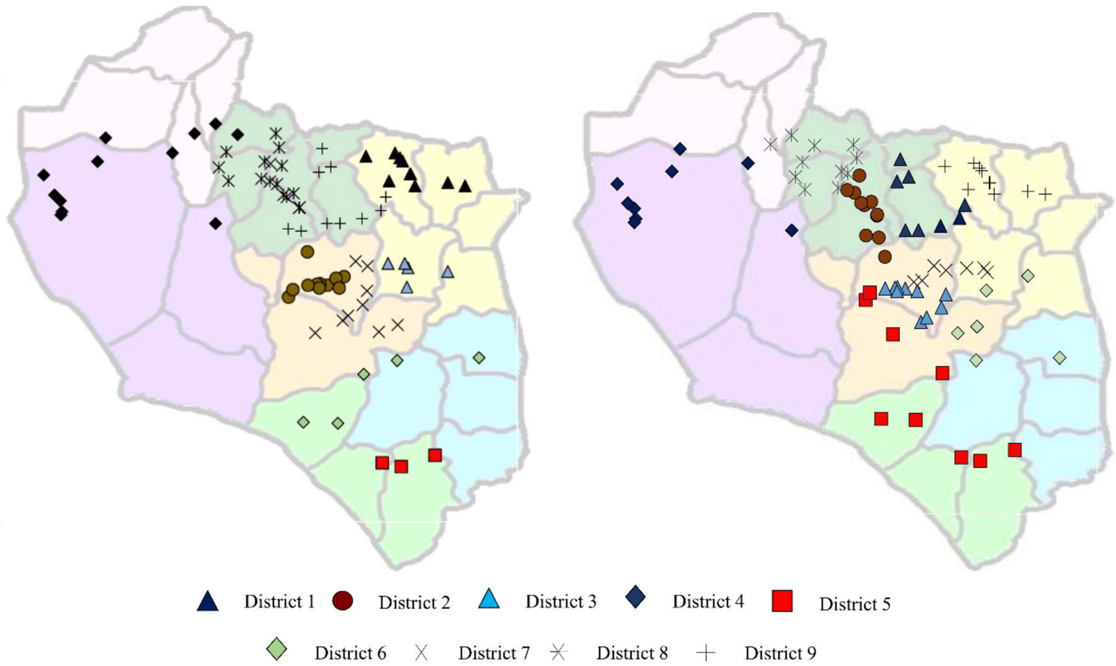


Figure 11. Existing and compromised optimum maps of the healthcare system of SKP.

Moreover, as illustrated in Figure 11, all the restrictions regarding contiguity and absence of embedded districts are properly met. In order to allow for a closer examination of the results, the structures generated by the existing solution and the compromised one are presented in Table 9 and Figure 11.

It can be observed that the districts formed in the structure of the compromised solution are more balanced in terms of number of cities per district. Though this result is not among the objectives of the problem, it describes the seemingly more appropriate partitioning created by the current model. Another important point concerns the way in which services are allocated to the different districts in both the compromised solution and the existing one, as respectively described in Tables 10 and 11. It should be emphasized that the compromised solution has been designed to minimize the number of inter-district displacements as well as the heterogeneity within districts.

Table 9. Distribution of cities among the districts.

Existing solution			Compromised solution		
Number of districts	Number of cities in a district	Total number of cities	Number of districts	Number of cities in a district	Total number of cities
1	3	3	4	9	36
2	5	10	3	11	33
1	9	9	1	6	6
2	10	20	1	7	7
2	12	24	–	–	–
1	16	16	–	–	–
9	–	82	9	–	82

Table 10. Service allocation in the compromised solution.

District 1	{3, 5, 6, 7, 8, 9, 11, 13, 15, 16, 18, 20, 21, 23, 24, 25}
District 2	{1, 4, 7, 10, 11, 14, 19, 20, 21, 22, 26}
District 3	{2, 4, 6, 9, 11, 14, 15, 21, 23, 24, 25, 26}
District 4	{1, 2, 3, 5, 6, 7, 9, 11, 12, 14, 18, 21, 23}
District 5	{5, 9, 12, 16, 18, 20, 21}
District 6	{1, 3, 4, 7, 8, 9, 10, 12, 13, 20, 21, 23, 25, 26, 27}
District 7	{3, 6, 8, 9, 11, 14, 16, 17, 21, 23, 25, 26, 27}
District 8	{1, 3, 4, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 24, 26, 27}
District 9	{2, 3, 4, 8, 9, 10, 12, 13, 15, 17, 18, 21, 22, 24}

Table 11. Services allocation in the exiting solution.

District 1	{4, 8, 9, 11, 12, 18, 20, 21, 23, 24, 27}
District 2	{1, 2, 4, 5, 6, 7, 8, 10, 10, 16, 18, 20, 21, 23, 24, 25}
District 3	{1, 2, 3, 6, 10, 11, 14, 15, 16, 17, 20, 21, 26}
District 4	{3, 5, 8, 10, 12, 16, 17, 23, 24}
District 5	{1, 1, 1, 1, 6, 7, 8, 11, 14, 16, 18, 19, 20, 22}
District 6	{2, 3, 4, 5, 9, 11, 12, 16, 17, 19, 21, 23, 26}
District 7	{1, 2, 3, 4, 5, 8, 10, 11, 13, 14, 16, 20, 23, 24, 26, 27}
District 8	{2, 4, 6, 7, 910, 13, 15, 16, 24, 26}
District 9	{1, 2, 4, 8, 9, 14, 15, 18, 21, 24, 25, 2, 6, 27}

Finally, we should note that the selection of an appropriate solution from among the Pareto front members is a matter of great importance, since the implementation of the results obtained can be regarded as a strategic-level decision. This is particularly the case when the structures defined fail to operate appropriately as a result of the population changes occurring over a long period of time. Although population changes usually take place very slowly, this problem should be considered in light of the substantial importance of keeping a balanced population within the health system. It is therefore recommended that proper scientific methods are applied to select the final solution.

7. Conclusion and future research directions

The main objectives in the design of a health system consist of increasing the coverage level and reducing its operating costs. In this regard, the design of the health infrastructures of a country is one of the most important problems in the field of macro-level decision-making. To the best of our knowledge, the current paper is the first research proposing a comprehensive mathematical model to address the partitioning problem while considering the contiguity and absence of embedded districts constraints.

The model has been designed to guarantee that all the services required by each city within a district are satisfied while minimizing the heterogeneity of districts in terms of the services required by their cities. Moreover, the distribution of the teams of experts providing services across districts has been defined to minimize their number of displacements (and subsequent operating costs). All in all, the formal structure of the model allows us to consider the problems of district generation and service provision simultaneously.

We have used the NSGAII and MOGWO algorithms to solve several numerical examples and then compared the results obtained. Since MOGWO outperforms NSGAII, it has been applied to analyse (and improve upon) the health monitoring plan implemented in the South Khorasan Province of Iran in 2015. The partitioning of the corresponding populated areas has been defined after identifying and comprehensively studying the regulations governing the health system of the country as well as investigating the plan being implemented.

From a computational viewpoint, different solution algorithms can be applied to solve the model and their behaviour in dealing with the partitioning problem should be further investigated. In addition, the definition of a comprehensive mathematical model implies that accurate solution methods such as Local Solver can be applied to the current framework, a possibility suggested when conducting future research. Finally, the model could be modified and adapted to account for the population shifts derived from migration processes, particularly when studying border and developing areas.

Acknowledgement

The authors would like to thank the anonymous reviewers and the editor for their insightful comments and suggestions. This research was partially supported by the research grant GA CR 19-13946S awarded to Dr. Tavana by the Czech Science Foundation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. We refer the reader to the corresponding websites http://nihr.tums.ac.ir, and www.amar.org.ir/.

ORCID

Madjid Tavana http://orcid.org/0000-0003-2017-1723

References

Alawadhi, S., & Mahalla, R. (2015). The political districting of Kuwait: Heuristic approaches. *Kuwait Journal of Science*, *42*(2), 160–188.

Baños, R., Gil, C., Paechter, B., & Ortega, J. (2007). A hybrid meta-heuristic for multi-objective optimization: MOSATS. *Journal of Mathematical Modelling and Algorithms*, *6*(2), 213–230. doi:10.1007/s10852-006-9041-6

Baruch, Z., Cret, O., & Pusztai, K. (1999). Genetic algorithm for circuit partitioning. Paper presented at the proceedings of fifth international conference on engineering of modern electric systems: section computer science and control systems.

Benzarti, E., Sahin, E., & Dallery, Y. (2013). Operations management applied to home care services: Analysis of the districting problem. *Decision Support Systems*, *55*(2), 587–598. doi:10.1016/j.dss.2012.10.015

Boulle, M. (2004). Compact mathematical formulation for graph partitioning. *Optimization and Engineering*, *5*(3), 315–333. doi:10.1023/B:OPTE.0000038889.84284.c7

Bozkaya, B., Erkut, E., & Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, *144*(1), 12–26. doi:10.1016/S0377-2217(01)00380-0

Brooks, S. P., & Morgan, B. J. (1995). Optimization using simulated annealing. *The Statistician*, *44*(2), 241–257. doi:10.2307/2348448

Butsch, A., Kalcsics, J., & Laporte, G. (2014). Districting for arc routing. *INFORMS Journal on Computing*, *26*(4), 809–824. doi:10.1287/ijoc.2014.0600

Camacho-Collados, M., Liberatore, F., & Angulo, J. (2015). A multi-criteria Police Districting Problem for the efficient and effective design of patrol sector. *European Journal of Operational Research*, *246*(2), 674–684. doi:10.1016/j.ejor.2015.05.023

Coello, C. A. C., Lamont, G. B., & Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems* (Vol. 5, 2nd ed.). Berlin, Germany: Springer.

Contreras, I., Fernández, E., & Reinelt, G. (2012). Minimizing the maximum travel time in a combined model of facility location and network design. *Omega*, *40*(6), 847–860. doi:10.1016/j.omega.2012.01.006

D'Amico, S. J., Wang, S.-J., Batta, R., & Rump, C. M. (2002). A simulated annealing approach to police district design. *Computers & Operations Research*, *29*(6), 667–684. doi:10.1016/S0305-0548(01)00056-9

Datta, D., Figueira, J., Gourtani, A., & Morton, A. (2013). Optimal administrative geographies: An algorithmic approach. *Socio-Economic Planning Sciences*, *47*(3), 247–257. doi:10.1016/j.seps.2013.03.002

Datta, D., & Figueira, J. R. (2011). Graph partitioning by multi-objective real-valued metaheuristics: A comparative study. *Applied Soft Computing*, *11*(5), 3976–3987. doi:10.1016/j.asoc.2011.01.044

Datta, D., & Figueira, J. R. (2012). Some convergence-based M-ary cardinal metrics for comparing performances of multi-objective optimizers. *Computers & Operations Research*, *39*(7), 1754–1762. doi:10.1016/j.cor.2011.10.013

Datta, D., Figueira, J. R., Fonseca, C. M., & Tavares-Pereira, F. (2008). Graph partitioning through a multi-objective evolutionary algorithm: A preliminary study. Paper presented at the proceedings of the 10th annual conference on genetic and evolutionary computation.

Datta, D., Fonseca, C. M., & Deb, K. (2008). A multi-objective evolutionary algorithm to exploit the similarities of resource allocation problems. *Journal of Scheduling*, *11*(6), 405–419. doi:10.1007/s10951-008-0073-9

Datta, D., Malczewski, J., & Figueira, J. R. (2012). Spatial aggregation and compactness of census areas with a multiobjective genetic algorithm: a case study in Canada. *Environment and Planning B: Planning and Design*, *39*(2), 376–392. doi:10.1068/b38078

De Assis, L. S., Franca, P. M., & Usberti, F. L. (2014). A redistricting problem applied to meter reading in power distribution networks. *Computers & Operations Research*, *41*, 65–75. doi:10.1016/j.cor.2013.08.002

De Barros Franco, D. G., & Arns Steiner, M. T. (2018). Clustering of solar energy facilities using a hybrid fuzzy c-means algorithm initialized by metaheuristics. *Journal of Cleaner Production*, *191*, 445–457. doi:10.1016/j.jclepro.2018.04.207

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. doi:10.1109/4235.996017

Galvão, L. C., Novaes, A. G., De Cursi, J. S., & Souza, J. C. (2006). A multiplicatively-weighted Voronoi diagram approach to logistics districting. *Computers & Operations Research*, *33*(1), 93–114. doi:10.1016/j.cor.2004.07.001

García-Ayala, G., González-Velarde, J. L., Ríos-Mercado, R. Z., & Fernández, E. (2016). A novel model for arc territory design: Promoting Eulerian districts. *International Transactions in Operational Research*, *23*(3), 433–458. doi:10.1111/itor.12219

Haugland, D., Ho, S. C., & Laporte, G. (2007). Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, *180*(3), 997–1010. doi:10.1016/j.ejor.2005.11.070

Inayoshi, H., & Manderick, B. (1994). The weighted graph bi-partitioning problem: A look at ga performance. In Y. Davidor, H. P. Schwefel, & R. Männer (eds.) *Parallel problem solving from nature—PPSN III* (pp. 617–625), Berlin, Heidelberg: Springer.

Jovanovic, R., Tuba, M., & Voss, S. (2016). An ant colony optimization algorithm for partitioning graphs with supply and demand. *Applied Soft Computing*, *41*(C), 317–330. doi:10.1016/j.asoc.2016.01.013

Kalcsics, J. (2015). Districting problems. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (chap. 23, pp. 595–622). Berlin, Heidelberg: Springer.

Khodaparasti, S., Maleki, H. R., Jahedi, S., Bruni, M. E., & Beraldi, P. (2017). Enhancing community based health programs in Iran: A multi-objective location-allocation model. *Health Care Management Science*, *20* (4), 485–499. doi:10.1007/s10729-016-9366-2

Kim, J., Hwang, I., Kim, Y.-H., & Moon, B.-R. (2011). Genetic approaches for graph partitioning: a survey. Paper presented at the proceedings of the 13th annual conference on genetic and evolutionary computation.

Kim, Y.-H., & Moon, B.-R. (2004). Investigation of the fitness landscapes in graph bipartitioning: An empirical study. *Journal of Heuristics*, *10*(2), 111–133. doi:10.1023/B:HEUR.0000026263.43711.44

Knight, V. A., Harper, P. R., & Smith, L. (2012). Ambulance allocation for maximal survival with heterogeneous outcome measures. *Omega*, *40*(6), 918–926. doi:10.1016/j.omega.2012.02.003

Lewis, M., Kochenberger, G., & Alidaee, B. (2008). A new modeling and solution approach for the set-partitioning problem. *Computers & Operations Research*, *35*(3), 807–813. doi:10.1016/j.cor.2006.04.002

Li, W., Church, R. L., & Goodchild, M. F. (2014). An extendable heuristic framework to solve the p-compact-regions problem for urban economic modeling. *Computers, Environment and Urban Systems*, *43*, 1–13. doi:10.1016/j.compenvurbsys.2013.10.002

Lin, R.-C., Sir, M. Y., & Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega*, *41*(5), 881–892. doi:10.1016/j.omega.2012.11.003

Liu, R., Xie, X., & Garaix, T. (2014). Hybridization of tabu search with feasible and infeasible local searches for periodic home health care logistics. *Omega*, *47*, 17–32. doi:10.1016/j.omega.2014.03.003

Maini, H., Mehrotra, K., Mohan, C., & Ranka, S. (1994). Genetic algorithms for graph partitioning and incremental graph partitioning. Paper presented at the proceedings of the 1994 ACM/IEEE conference on supercomputing.

Mirjalili, S., Saremi, S., Mirjalili, S. M., & Coelho, L. D S. (2016). Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. *Expert Systems with Applications*, *47*, 106–119. doi:10.1016/j.eswa.2015.10.039

Moradi-Lakeh, M., & Vosoogh-Moghaddam, A. (2015). Health sector evolution plan in Iran; equity and sustainability concerns. *Int J Health Policy Manag*, *4*(10), 637–640. doi:10.15171/ijhpm.2015.160

Ricca, F., Scozzari, A., & Simeone, B. (2013). Political districting: From classical models to recent approaches. *Annals of Operations Research*, *204*(1), 271–299. doi:10.1007/s10479-012-1267-2

Ríos-Mercado, R. Z., & Escalante, H. J. (2016). GRASP with path relinking for commercial districting. *Expert Systems with Applications*, *44*, 102–113. doi:10.1016/j.eswa.2015.09.019

Ríos-Mercado, R. Z., & López-Pérez, J. F. (2013). Commercial territory design planning with realignment and disjoint assignment requirements. *Omega*, *41*(3), 525–535. doi:10.1016/j.omega.2012.08.002

Ríos-Mercado, R. Z., & Salazar-Acosta, J. C. (2011). A GRASP with strategic oscillation for a commercial territory design problem with a routing budget constraint. In I. Batyrshin, G. Sidorov (eds.) *Advances in soft computing* (pp. 307–318). Berlin, Heidelberg: Springer.

Rummler, A., & Apetrei, A. (2002). Graph partitioning revised – *A multiobjective perspective. Proc. of 6th MultiConf. On Syst., Cybern. and Informatics.*

Salazar-Aguilar, M. A., Ríos-Mercado, R. Z., & Cabrera-Ríos, M. (2011). New models for commercial territory design. *Networks and Spatial Economics*, *11*(3), 487–507. doi:10.1007/s11067-010-9151-6

Salazar-Aguilar, M. A., Ríos-Mercado, R. Z., & González-Velarde, J. L. (2011). A bi-objective programming model for designing compact and balanced territories in commercial districting. *Transportation Research Part C: Emerging Technologies*, *19*(5), 885–895. doi:10.1016/j.trc.2010.09.011

Salazar-Aguilar, M. A., Ríos-Mercado, R. Z., González-Velarde, J. L., & Molina, J. (2012). Multiobjective scatter search for a commercial territory design problem. *Annals of Operations Research*, *199*(1), 343–360. doi:10.1007/s10479-011-1045-6

Shirabe, T. (2012). Prescriptive modeling with map algebra for multi-zone allocation with size constraints. *Computers, Environment and Urban Systems*, *36*(5), 456–469. doi:10.1016/j.compenvurbsys.2011.12.003

Shortell, S. M., & Kaluzny, A. D. (2000). *Health care management: Organization, design, and behavior*. Clifton Park, NY, USA: Cengage Learning.

Steiner, M. T. A., Datta, D., Neto, P. J. S., Scarpin, C. T., & Figueira, J. R. (2015). Multi-objective optimization in partitioning the healthcare system of Parana State in Brazil. *Omega*, *52*, 53–64. doi:10.1016/j.omega.2014.10.005

Steiner Neto, P. J., Datta, D., Arns Steiner, M. T., Canciglieri Júnior, O., Rui Figueira, J., Pereira Detro, S., & Tadeu Scarpin, C. (2017). A multi-objective genetic algorithm based approach for location of grain silos in Paraná State of Brazil. *Computers & Industrial Engineering*, *111*, 381–390. doi:10.1016/j.cie.2017.07.019

Tavares-Pereira, F., Figueira, J. R., Mousseau, V., & Roy, B. (2007). Multiple criteria districting problems. *Annals of Operations Research*, *154*(1), 69–92. doi:10.1007/s10479-007-0181-5

Wang, T., Wu, Z., & Mao, J. (2007). A new method for multi-objective tdma scheduling in wireless sensor networks using pareto-based pso and fuzzy comprehensive judgement. In R. Perrott, B. M. Chapman, J. Subhlok, R. F. de Mello, & L. T. Yang (Eds.), *High performance computing and communications* (pp. 144–155). Berlin, Heidelberg: Springer.

Yamada, T. (2009). A mini–max spanning forest approach to the political districting problem. *International Journal of Systems Science*, *40*(5), 471–477. doi:10.1080/00207720802645246

Yanık, S., Sürer, Ö., & Öztayşi, B. (2016). Designing sustainable energy regions using genetic algorithms and location–allocation approach. *Energy*, *97* (15), 161–172. doi:10.1016/j.energy.2015.12.116

Zhou, L., Geng, N., Jiang, Z., & Wang, X. (2018). Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega*, *81*, 220–233. doi:10.1016/j.omega.2017.11.005