

MODELING THE DISTRIBUTION OF MAMMOGRAPHY SCREENINGS USING A MULTI-OBJECTIVE APPROACH

**Yerlis Valdelamar Gonzalez, Leandro Luiz Fragoso, Vothan Salomao, Washington Alves de
Oliveira, Priscila C. Berbert Rampazzo**

Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas

R. Pedro Zaccaria, 1300, Limeira - SP, 13484-350

yerlis2499@gmail.com, lelu.fragoso@hotmail.com, vothansalomao@gmail.com,

waoliv@unicamp.br, priscila.rampazzo@fca.unicamp.br

ABSTRACT

Premature detection of breast cancer is essential to reduce the number of deaths. However, this detection depends on the adequate location and allocation of mammography equipment to meet the demands of women who must perform the exam. In this sense, this work presents a multi-objective programming approach to model the problem of locating and allocating mammography units, maximizing the demand coverage, and minimizing the distance traveled by women to access the mammography exam. Exact resolution methods are studied and implemented to capture and analyze the Pareto frontier: the weighted sum, epsilon-restricted, and goal programming, employing different variations. The results show that these methods offer alternative and satisfactory solutions to the problem and can determine a dispersed Pareto frontier, validating the conflict between the objectives.

KEYWORDS. Location-allocation problems. Classical methods . Multi-Objective optimization.

Paper topics: SA – PO na Área de Saúde, PM – Programação Matemática, ADM – Apoio à Decisão Multicritério

1. Introduction

Mammography is an imaging technique used to detect changes in the breast, including lumps or calcifications that may indicate cancer. In this sense, mammography screening is recommended for women over 40 or with a family history of breast cancer. Its adequate allocation is essential for early detection through mammography screening and diagnostic tests, increasing the chances of successful treatment and cure, so the service must be available. However, mammography equipment allocated without any criteria conduce to long waiting times and delays in diagnosis and cancer treatment. Unfortunately, in different parts of the world, including Brazil, the lack of investment makes it difficult to achieve geographic coverage of mammography equipment. One of the main reasons for this is the need for more presentations of projects that make investments in public health in Brazil feasible. Also, the “poor distribution” limits the access to the exam for numerous women Rodrigues et al. (2019). Consequently, there is an actual problem related to the equitable and appropriate distribution of mammography equipment in Brazil, as is exposed in (Amaral et al., 2023).

The Mammography Unit Location-Allocation Problem (MULAP) consists of deciding in which cities a specific quantity of mammography units must be located and the percentage of demand that must be covered for each mammography unit. The demand for mammography screening is the number of women who need to take a mammography exam. In this sense, some authors have proposed heuristics and meta-heuristics algorithms to solve the problem using data from some regions of Brazil as in (De Campos et al., 2020; Souza et al., 2019). De Campos et al. (2022) considered the Multi-Objective Mammography Unit Location-Allocation Problem (MOMULAP), simultaneously improving the covered demand and the weighted distance. The authors implemented the Strength Pareto Evolutionary Algorithm (SPEA) and Non-dominated Sorting Genetic Algorithm II (NSGA II). Similarly, De Campos et al. (2023) modeled the problem to attend to these two objectives and maximize the equity of access to mammography units employing the same solution methods. (Barbosa et al., 2023) present a multi-objective approach to solving the problem using two classic methods: weighted sum and epsilon restricted, aiming to improve the coverage of demand of mammography units and simultaneously minimizing the number of mammography units employed. They used instances from Rondonia and Espirito Santo. According to our research, it is the only work that uses exact methods to solve the MOMULAP, explicitly considering the two objective functions mentioned above.

Figure 1 represents the keywords obtained from the works used as references for our research, which were grouped using the Leiden Algorithm Traag et al. (2019). The works that are part of the scope and development of this research are mainly divided into the main clusters: [1] mammography, breast neoplasm, health service accessibilities, parameters, diagnostic screening programs, national health programs, universal health insurance, breast cancer, fatality, mortality, screening; [2] facility location, maximal covering, p-center problem, conditional covering problem, epsilon-constraint method, large scale emergency, p-median problem, partial coverage, genetic algorithm, mathematical programming, simulated annealing; [3] multi-objective, fuzzy analytic hierarchy process, infectious waste disposal centers, Pareto solution, tabu search algorithm, two-level problem; [4] meta-heuristic, mammography unit location, NSGA II, SPEA II, variable neighborhood search, mammography screening; [5] health, mammography equipment, spatial distribution; [6] location-allocation problem, health-care facility, Hong Kong. The keywords were listed and counted, forming a list of occurrences. In the network structure, the keywords are represented by a node; the edges link keywords that appear together in the same document. Each edge has a weight associated with the frequency that two keywords are cited together, considering all documents.

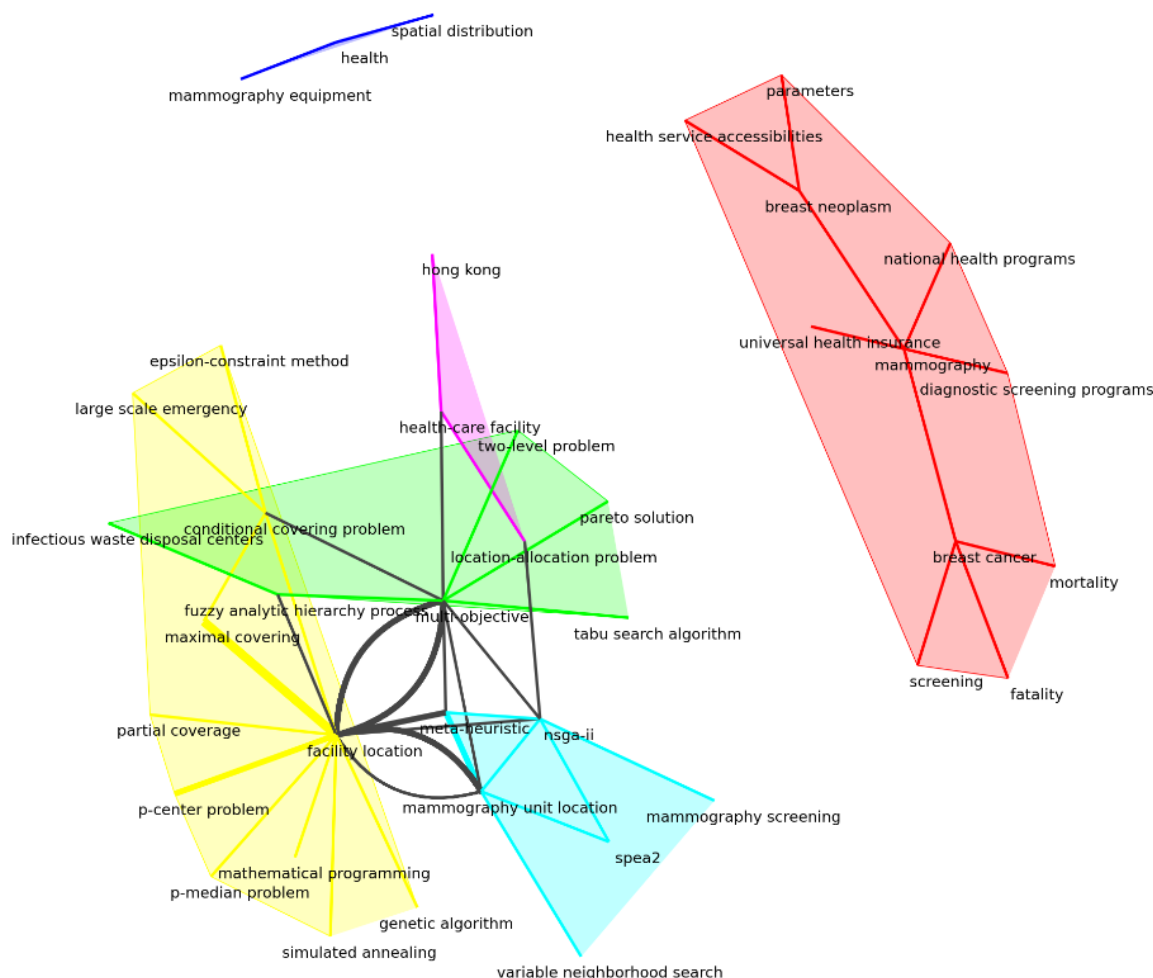


Figure 1: Bibliographic References: groups of keywords.

When it comes to the problem of mammography unit location-allocation and screening, most of the existing literature focuses on multi-objective optimization. However, these works often overlook important aspects such as the approximation of the Pareto Frontier, extreme solutions, utopian solutions, and the exploration of trade-offs with classical methods. Our work aims to fill this gap by providing a comprehensive analysis of these issues. In this sense, the main contribution of this work is to implement three multi-objective classical methods to solve MOMULAP: the weighted sum, epsilon-restricted, and goal programming methods. The idea is to represent the trade-offs between objectives, focusing on analyzing the quality of solutions they offer and the advantages and disadvantages of each method to improve the coverage of demand of mammography units while minimizing the distance traveled by women who need to take the mammography exam. A sensitive analysis of coverage radii is also performed, employing three different radii and analyzing the divergence in solutions between the regions. Furthermore, this work handles real data from the other areas of Brazil that still need to be explored.

2. Mathematical formulation

The MOMULAP formulation presented here lies in the proposal of De Campos et al. (2020) and De Campos et al. (2023) with the difference that the radius parameter R is variable between 30,60 and 90 km. The following considerations are essential for understanding the mathematical formulation.

- a) N is the set of n candidate cities to host p mammography units.
- b) Each mammography unit can perform 6,758 screenings per year. Sistema Único de Saúde (SUS) fixes this parameter.
- c) Each city j has the annual demand dem_j of women in the age group indicated to perform mammography screenings, which is 58.9%, which belongs to the age range of 50-59 and 60-69 years.
- d) It is assumed that all cities have adequate hospital infrastructure to host mammography equipment.
- e) The distance from a city to another city that hosts mammography equipment is compared with the radius R (in km); this distance cannot exceed R . The SUS fixed this parameter at 60 km, and two other additional radii were considered, 30 km and 90 km, to analyze the importance of radio coverage. Figure 2 shows the importance of the coverage radius of a mammography unit. A mammography unit will only cover the demand of cities within that radius. The variation of this parameter directly influences the assignment of mammography units for each town and the percentage of demand it will cover in other cities.

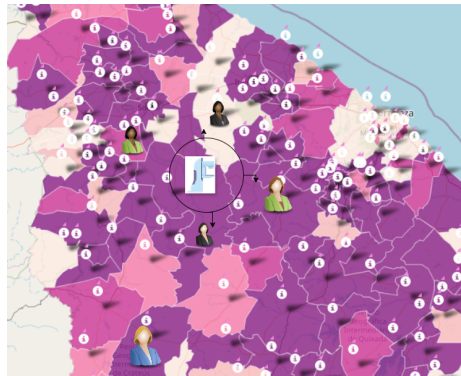


Figure 2: Coverage radius of a mammography unit

- f) The demand for mammography screenings from one city can be fully or partially covered by mammography units installed in one or more other cities.
- g) If a city hosting mammography equipment is fully covered, it can meet another city's demand for mammography screenings.

2.1. Description of parameters and decision variables

Parameters:

$N = \{1, \dots, n\}$ candidate cities to host mammography units

d_{ij} distance from city i to city j in km;

dem_j annual demand for mammography screenings in city j ;

cap annual mammography screenings capacity of a mammography unit;

p total number of mammography units to be located;

R maximum distance to a mammography unit in km;

S_i set of cities whose distance from city i is less or equal to R . $S_i = \{j \in N \mid d_{ij} \leq R\}$, $i \in N$.

Decision variables:

x_{ij} continuous variable into interval $[0, 1]$ that indicates the fraction of from city j , which is served by mammography units in the city i ;

y_i integer variable indicating the total number of mammography units located at city i .

Auxiliary decision variables:

t_i binary variable for demand coverage; $t_i = 1$ if the city i is fully covered by mammography units hosted itself, otherwise $t_i = 0$;

z_i binary variable for mammography designation; $z_i = 1$ if the city i hosts mammography units, otherwise $z_i = 0$.

Mathematical modeling

$$\text{Maximize} \quad f_1(x) = \sum_{i \in N} \sum_{j \in S_i} dem_j \cdot x_{ij} \quad (1)$$

$$\text{Minimize} \quad f_2(x) = \sum_{i \in N} \sum_{j \in S_i} x_{ij} \cdot d_{ij} \cdot dem_j \quad (2)$$

$$\text{subject to} \quad \sum_{i \in S_j} x_{ij} \leq 1, \quad \forall j \in N, \quad (3)$$

$$\sum_{i \in N} y_i = p, \quad (4)$$

$$\sum_{j \in S_i} dem_j \cdot x_{ij} \leq cap \cdot y_i, \quad \forall i \in N, \quad (5)$$

$$t_i \leq z_i, \quad \forall i \in N, \quad (6)$$

$$t_i \leq x_{ii}, \quad \forall i \in N, \quad (7)$$

$$t_i \geq dem_i \cdot x_{ii} - dem_i + 1, \quad \forall i \in N, \quad (8)$$

$$t_i \geq x_{ij}, \quad \forall i, j \in N, \quad (i \neq j), \quad (9)$$

$$y_i \leq p \cdot z_i, \quad \forall i \in N, \quad (10)$$

$$z_i \leq y_i, \quad \forall i \in N, \quad (11)$$

$$x_{ij} \in [0, 1], \quad \forall i, j \in N, \quad (12)$$

$$y_i \in \mathbb{Z}^+, \quad \forall i \in N, \quad (13)$$

$$t_i \in \{0, 1\}, \quad \forall i \in N, \quad (14)$$

$$z_i \in \{0, 1\}, \quad \forall i \in N. \quad (15)$$

The bi-objective model involves two conflicting objective functions, which (1) aims to maximize the covered demand for mammography screenings and (2) minimize the distance traveled by the users. According to (3), the demand fraction for mammography screenings of the city j can only be satisfied by mammography units located at cities within the radius R , ensuring no over-coverage because the total fraction is limited by 100%. All available mammography equipment must be located by (4). Constraints (5) are the capacity constraints of each mammography unit. According to (6)–(8), the binary variable t_i takes the value 1 if the located mammography to it fully covers the city i ; otherwise, $t_i = 0$. Constraints (9) ensure that a city i will only serve another city j if its demand has just been fully served by itself. If mammography is installed in a city i , then z_i takes the value 1 by (10) and (11); otherwise, $z_i = 0$. The sets of constraints (12)–(15) specify the non-negativity and the integrity of variables for the model.

3. Solution methods

3.1. Weighted sum (WS) method

This method is based on assigning weights to each objective function according to the level of importance. Considering the model (1)–(15) and the weight vector $v \in \Re^m$, with $v_i \geq 0$, $i = 1, \dots, m$, and $\sum_{i=1}^m v_i = 1$, the WS method is based on the resolution of the following problem.

$$\text{Minimize} \quad \sum_{i=1}^m v_i f_i(x) \quad (16)$$

$$\text{subject to} \quad x \in S. \quad (17)$$

The model considers the weight vector with $v_1 + v_2 = 1$. Thus, for the implementation, the WS method considered pairs of weights $(v_1, 1 - v_1)$ such that v_1 shifts 0.01 from 0.0 to 1.0, representing the importance given to f_1 .

3.2. Epsilon-restricted (ER) method

This method is based on optimizing the objective function of the major priority, converting the other's objectives into inequality constraints. The ER method resolves the following problem by considering the model (1)–(15) and choosing the objective function i .

$$\text{Minimize} \quad f_i(x) \quad (18)$$

$$\text{subject to} \quad f_j(x) \leq \epsilon_j, \quad i \neq j, \quad (19)$$

$$x \in S. \quad (20)$$

Motivated by Rodrigues et al. (2019) in the implementation considered here, f_1 was chosen as the objective function of significant priority, and f_2 was incorporated into the constraints set. The ER method involved creating 100 variations of epsilon values to limit f_2 , considering the extreme values of f_2 .

3.3. Goal programming (GP) method

In the context of this work, this method is applied as follows, where d_i^+ and d_i^- represent the over-achievement and under-achievement of objectives understood as deviation variables (positive and negative), respectively. Decision-makers specify their preference level g_i (target) for each objective. Considering the feasible solutions set S of the model (1)–(15), the GP method is based on resolving

the following problem.

$$\text{Minimize} \quad \sum_{i=1}^2 (d_i^- + d_i^+) \quad (21)$$

$$\text{subject to} \quad f_1(x) - d_1^- + d_1^+ = g_1, \quad (22)$$

$$f_2(x) - d_2^- + d_2^+ = g_2, \quad (23)$$

$$x \in S, \quad (24)$$

$$d_i^-, d_i^+ \geq 0. \quad (25)$$

To implement the GP method, we determine the worst value for each target as at least 50% of the demand that must be satisfied for f_1 , and at most 30% of the total sum of distances for f_2 .

4. Computational experiments

4.1. MOMULAP instances

We considered 10 different instances, divided into 5 synthetic instances and 5 real instances, with data obtained from the DATASUS website (see Table 1). Each instance comprises the number of cities N , the number of mammography units p , the capacity of each mammography cap , the demand of each city dem_j , and the maximum distance allowed R .

Table 1: Main data of instances

Instance	N.Cities	N.Units	Capacity	Demand	State
Synthetic					
SA15	15	140	2000	258220	N/A
SB33	33	250	3000	2447171	N/A
SC60	60	40	2000	276343	N/A
SD90	90	60	4000	189068	N/A
SE150	150	250	3500	3697955	N/A
Real					
RA22	22	22	6758	453517	Acre
RD62	62	72	6758	276343	Amazonas
RE102	102	78	6758	201867	São Paulo
RF184	184	325	6758	4760409	Ceará
RG295	295	403	6758	3697955	Santa Catarina

4.2. Weighted sum analysis

For weighted sum analysis, we define the level of importance of each objective function. Table 2 shows an example of the discretization of the most representative pairs of weights in the first column. It shows four different pairs and the values for objective function f_1 and objective function f_2 for a coverage radius of 60 km in the instance of 150 cities. GAP was 0.0% for all cases. Figure 2 shows the objective space between the two objective functions for the instance of 150 cities.

As is shown in Figure 3 (a)-(c), the impact generated on the value of the objective functions is considerable. When the importance of objective function f_1 decreases, its value decreases, but objective function f_2 decreases (which is desirable), and when the importance of objective function f_2 decreases, its value increases, but objective function f_1 increases (which is desirable), showing clearly that a conflict exists between these two functions. Also, the objective function f_2 value is impacted by the coverage radius, being greater or equal when the coverage radius increases. We obtained similar results from the other instances.

Table 2: WS method results for 150 cities

W1	W2	f_1	f_2	Radio
1	0	874999.99	16842465.57	30
0.96	0.04	871763.07	368672.03	30
0.77	0.23	218440.0	407931.00	30
0.65	0.35	184881.0	317145.81	30
1	0	875000.0	34047045.11	60
0.96	0.04	871763.0	7368672.03	60
0.77	0.23	218440.0	407931.00	60
0.65	0.35	184881.0	317145.81	60
1	0	875000.0	50778650.84	90
0.96	0.04	871763.0	7368672.03	90
0.77	0.23	218440.0	407931.00	90
0.65	0.35	184881.0	317145.81	90

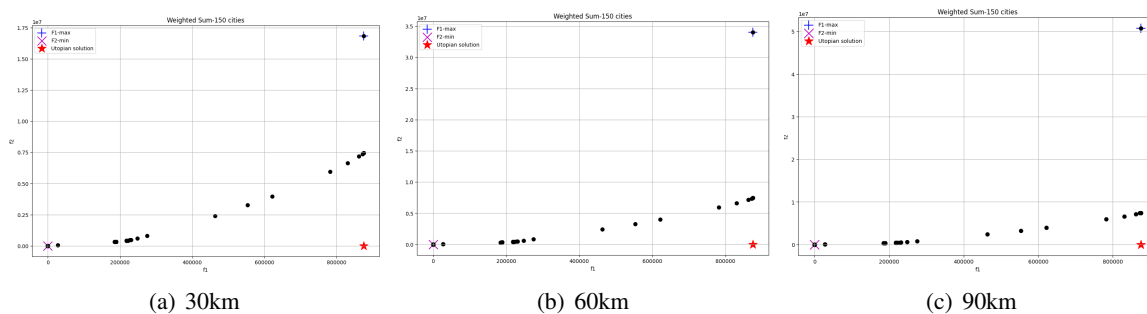


Figure 3: Weighted Sum for 150 cities

4.3. Epsilon-restricted analysis

In a similar way, for the Epsilon-restricted method, 100 different values of epsilon were generated. Table 3 shows an example containing 10 of the 100 values of epsilons and objectives function's values, for instance, of 150 cities. For this instance, the GAP was 0.0 % in all cases. Objective f_1 was established as the most important objective due to the persistent low demand coverage of women who need to undergo mammography examinations in Brazil, and objective f_2 became an additional constraint. Figure 4 (a)-(c) show the objective space of the instance of 150 cities.

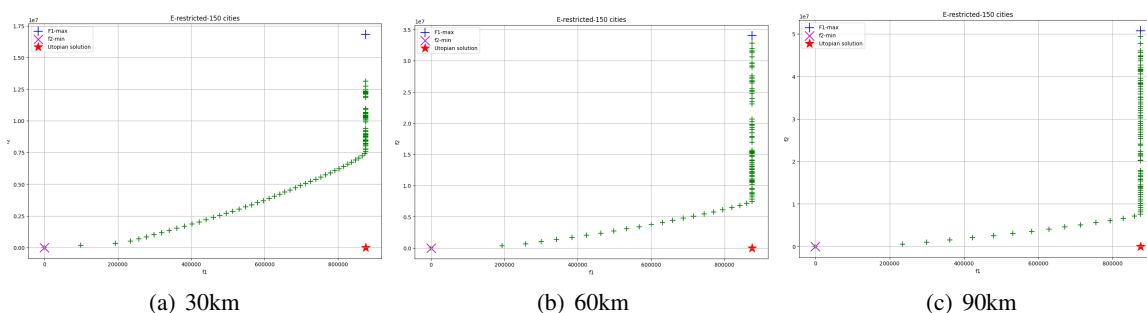


Figure 4: Epsilon Restricted for 150 cities

The results presented show that the value of objective function f_2 is greater when a radius of 90km

Table 3: ER method results for 150 cities and 60km

Epsilons	f_1	f_2
340470.45	194414.67	340470.45
680940.90	258298.72	680940.90
1021411.35	300374.43	1021411.35
2383293.15	462638.62	2383293.15
3404704.51	568095.74	3404704.51
5107056.76	715856.42	5107056.76
17363993.00	875000.0	14680729.09
13957566.50	875000.0	13957566.50
18044933.91	875000.0	12977220
23832931.57	875000.0	15364825

is considered, indicating that a greater distance has been traveled to serve the cities. The value of objective function f_1 increases and sometimes remains the same, indicating that a greater demand is satisfied when the 90km radius is considered. This method also allows us to observe the comparison between the two functions, similar to the weighted sum method, which helps us understand the model's behavior.

4.4. Goal programming analysis

Table 4 shows an example of the performance of f_1 and f_2 when the model takes different deviations for the instance of 150 cities, taking into account the coverage radius of 60 as in weighted sum and epsilon restricted method. For this instance, the GAP was 0.0% in all cases. Ten of the 100 goal values are exposed. For almost all solutions, the deviations focused on negative deviations in objective function f_1 , indicating that a considerable part of demand is not covered.

Table 4: GP method results for 150 cities and 60km

Deviation	f_1	f_2
1809671.77	57982.26	96963.48
1772464.88	113865.69	193926.97
1735257.99	169749.12	290890.46
1712393.87	211289.78	387853.95
1710807.97	231552.22	484817.44
1715337.52	245699.21	581780.93
1721687.63	258025.65	678744.42
1728309.32	270080.49	775707.90
1735007.70	282058.66	872671.39
1741856.59	293886.31	969634.88

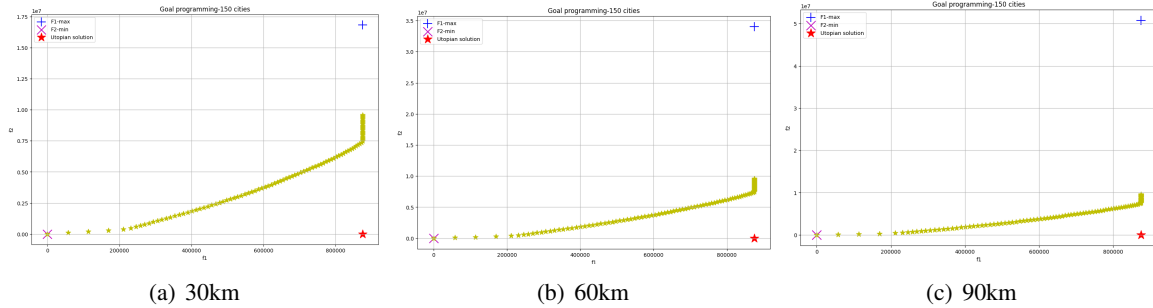


Figure 5: Goal Programming for 150 cities

The deviations were smaller when a radius of 90km was considered, as it is possible to cover a greater percentage of demand when a greater number of cities is considered. There were no significant differences between the radii 30, 60, and 90km, only in three different goal values given the nature of the instance about its instances. Even so, a significant amount of demand still needed to be met when using this method, as is shown in Figure 5 (a)-(c).

Concluding this part of the analysis, it's important to note that classical methods offer viable alternatives for solving MOMULAP, both for real-world and synthetic instances. However, it's crucial to be mindful of the model's complexity and various parameters. As the instance size grows, the solver may not always achieve optimal values in the WS method, leading to a 2.0% GAP in some instances.

4.5. Comparison of frontiers

Figure 6 (a)-(c) compares the Pareto frontiers obtained by the three classic methods for the instance of 150 cities, revealing a higher quality of solutions when the problem is addressed through the Epsilon-restricted method. This method finds a greater variety of solutions than weighted sum and goal programming, reaching better values for f_1 and f_2 . The impact of the coverage radius on the quality of the solutions is also evident, these being better when the coverage radius is 90km since it is less restricted in terms of distances, and therefore, the method finds better ways to locate and allocate the mammography units in the different cities. The information about the instances that were used to test the methods can be found at <https://github.com/YValdelamar/MamografiaData.git>.

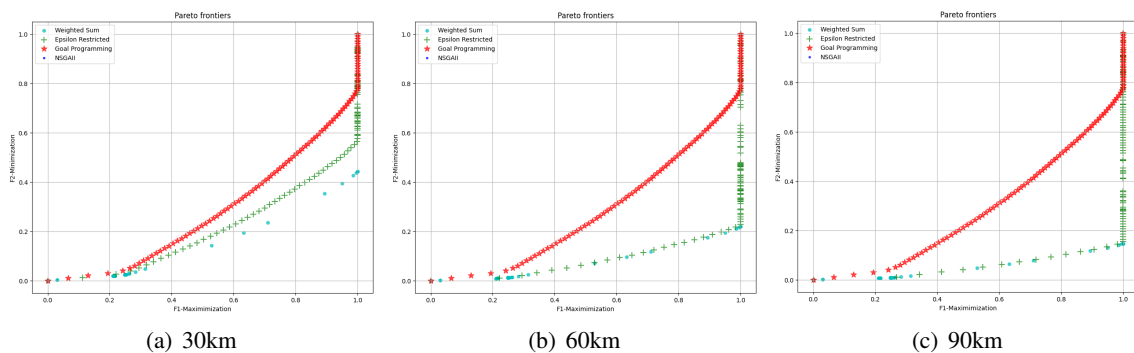


Figure 6: Comparison of frontiers for 150 cities

4.6. Hypervolume indicator

This indicator evaluates the distribution of different solutions across the Pareto frontier, and a reference point delimits it Audet et al. (2021). The reference point is 0 for objective function f_1 , and for objective function f_2 , it is calculated as in De Campos et al. (2023).

- For each city j , the furthest distance to the other cities under a specific radius is found. 30, 60, and 90 km were considered.
- Each calculated distance is multiplied by the demand of city j .
- These results are sum, and that value represents the nadir point to objective function f_2 .

In our comprehensive study, we calculated the hypervolume indicator for the three methods in the instance of 150. This calculation took into account the three coverage radii, as shown in Table 5.

Table 5: Hyper-volume Indicator. 150km

WS	ER	GP	Radius
0.244	0.249	0.248	30
0.608	0.611	0.610	60
0.980	0.983	0.980	90

For the three radii, the hypervolume indicator of the epsilon-restricted method is slightly better. Additionally, for the radius of 90km, the hypervolume indicator is much better than 30km and 60km, confirming the results of the previous individual analysis of each method. The Pareto frontiers generated by the three methods represent an acceptable distribution of solutions.

4.7. Comparison of classical methods configurations

Figure 7 (a)-(c) compares the three methods for the instance of Amazonas when a radius of 60km is considered. Simulating the role of the decision maker, we choose the three solutions that have a better value for f_1 . The percentage of coverage is higher for the solution of the WS method and smaller for the GP method. In the three solutions offered by the three methods, it is clear that there is a large percentage of demand that is not yet covered; this is because Amazonas is a region where the distances between cities are large, which means that, given the coverage radius set by the SUS of 60 km for mammography units, it is not possible for mammography units to widely reach the demand of a large number of cities.

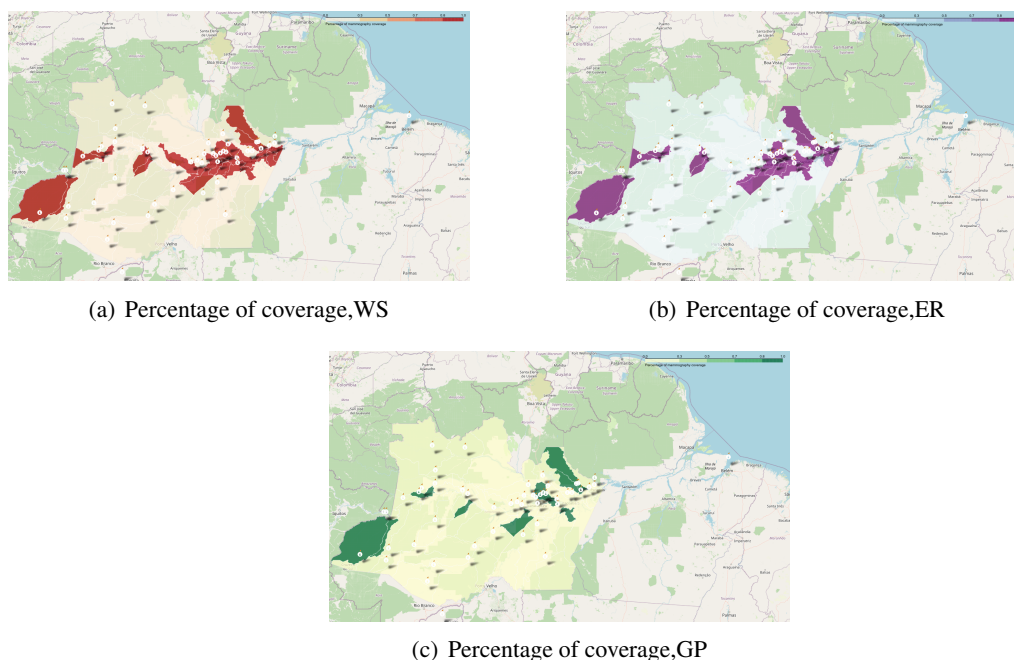


Figure 7: Comparison of coverage with the three methods. Amazonas

5. Conclusions

The MOMULAP is a problem deeply connected with a decision maker's decisions, who decides the best solution for each situation; this is how the methods employed constitute an appropriate and helpful tool to support that decision. Classical techniques offer suitable solutions for the resolution of MOMULAP, especially in small and medium instances. Through them, it is possible to represent the trade-offs between the objective functions justifying the use of multi-objective approaches and showing the complexity of the problem, despite some difficulties presented in the resolution part as the incumbent solutions encountered by the solver for some variations of the WS method and big instances. A significant percentage of demand is still uncovered; this is justified in part by the sensitive analysis of the three coverage radii, which demonstrates the importance of the coverage radius of each mammography unit because when this radius is higher, the value of f_1 is higher, too, and therefore the percentage of demand coverage is higher too.

References

- Amaral, P., Luz, L., Cardoso, F., and Freitas, R. (2023). Distribuição espacial de equipamentos de mamografia no brasil. *Revista Brasileira de Estudos Urbanos e Regionais*, 19:326–341.
- Audet, C., Bignon, J., Cartier, D., Le Digabel, S., and Salomon, L. (2021). Performance indicators in multiobjective optimization. *European journal of operational research*, 292(2):397–422.
- Barbosa, R. P., Souza, M. J. F., and de Miranda Junior, G. (2023). Mammography unit location: Reconciling maximum coverage and budgetary constraints. In *ICEIS (1)*, p. 187–194.
- De Campos, M. V. A., De Assis, R. S. L., Souza, M. J. F., De Siqueira, E. C., Silva, M. A. L., and De Souza, S. R. (2023). Multi-objective mammography unit location-allocation problem: A case study. Available at SSRN 4397830.

- De Campos, M. V. A., De Assis, R. S. L., Souza, M. J. F., Silva, M. A. L., De Siqueira, E. C., and De Souza, S. R. (2022). Algoritmos NSGA-II e SPEA2 para o problema multiobjetivo de localização de mamógrafos. *Anais do LIV Simpósio Brasileiro de Pesquisa Operacional*, 54.
- De Campos, M. V. A., De Sá, M. V. S. M., Rosa, P. M., Penna, P. H. V., De Souza, S. R., and Souza, M. J. F. (2020). A mixed linear integer programming formulation and a simulated annealing algorithm for the mammography unit location problem. In *ICEIS (I)*, p. 428–439.
- Rodrigues, D. C. N., Freitas-Junior, R., Rahal, R. M. S., da Silveira Corrêa, R., Peixoto, J. E., Ribeiro, N. V., Ferreira, N. C., and Soares, L. R. (2019). Difficult access and poor productivity: Mammography screening in brazil. *Asian Pacific Journal of Cancer Prevention: APJCP*, 20(6): 1857.
- Souza, M. J. F., Penna, P. H. V., Stilpen, M., Rosa, P. M., Monteiro, J. C., and Lisboa, M. R. (2019). Localização de mamógrafos: formulações e estudo preliminar de caso de rondônia. *LI Simpósio Brasileiro de Pesquisa Operacional*, 2.
- Traag, V., Waltman, L., and Van Eck, N. (2019). From louvain to leiden: guaranteeing well-connected communities. *sci. rep.* 9, 5233.