

A TWO PHASE ALGORITHM FOR THE DISTRIBUTION OF MAMMOGRAPHY UNITS WITH GEOGRAPHIC CLUSTER PARTITIONING

Yerlis Valdelamar Gonzalez ,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, waoliv@unicamp.br, priscila.rampazzo@fca.unicamp.br

ABSTRACT

Several studies have shown the effectiveness of mammography screening for the early detection of breast cancer, which is essential for increasing patients' chances of survival. However, it is not possible without an adequate distribution of mammography units. In this sense, this work proposes a two-phase implementation to achieve an adequate distribution of the mammography units, so that the most significant possible demand is met using both, fixed and mobile units. To this end, a mixed mathematical formulation is used to locate fixed mammography units, and a routing algorithm is proposed to distribute and generate the routes of mobile mammography units. The algorithms are tested with synthetic and real instances, proving that using these two types of mammography units can lead to better demand coverage.

KEYWORDS. Mammography screening. Vehicle Routing Problem. Constrained clustering.

Paper topics: SA – PO na Área de Saúde, L&T – Logística e Transportes

1. Introduction

Each year, thousands of women around the world are diagnosed with breast cancer, as this type of cancer is one of the most deadly. Randomized trials and incidence-based mortality studies have demonstrated that the mammography screening exam can reduce breast cancer mortality [Duffy et al., 2020]. There are direct benefits for using mammography exams for countries with population-based mammography screening programs [Cevik et al., 2018]. So it is essential that women, who are principally affected by breast cancer, have access to the mammography screening exam in order to start an effective treatment in case of an early detection of cancer. Thus, this exam is principally recommended for women over 40 years or women who present a family history of breast cancer.

However, an effective mammography distribution of mammography units faces different problems, such as a lack of resources (diagnostic machines), the cost of a trained workforce, among others Cevik et al. [2018]. In Brazil, a lack of investment exists, making it difficult to achieve a high geographic coverage of mammography screenings. Also, there exists a poor distribution of mammography units Rodrigues et al. [2019], while some states have enough mammography units to cover the demand, others do not have enough, so this generates an inappropriate distribution, limiting access to the exam [Amaral et al., 2023].

Some work has been carried out in Brazil to address this situation, seeking to model this real-life problem and find solutions that can offer an alternative to the appropriate location of mammography units. The Mammography Unit Location-Allocation Problem (MULAP) consists of deciding in which cities to locate the mammography units in and the percentage of demand that must be covered. Different methods, such as heuristics and meta-heuristics algorithms, have been implemented to solve the problem using data from some regions of Brazil, using a mono-objective and multiobjective approach, as in [De Campos et al., 2020; Souza et al., 2019]. This approach aims to provide a perspective on the allocation of fixed mammography units, considering the available number of mammography units in each state and the maximum coverage radius, among other parameters.

Similarly, with the same objective of improving that coverage of demand, the Mobile Mammography Unit Routing Problem (MMURP) was proposed by Rosa et al. [2020], where the objective is to maximize demand and minimize the distance traveled by Mobile Mammography Units (MMUs). The use of mobile mammography machines helps to partially or completely meet the demand in those cities that cannot be covered by fixed mammography machines, for reasons as a lack of infrastructure to house mammography equipment or because the city is out of the coverage radius. Da Cunha et al. [2019] researched to analyze the effectiveness of using mobile mammography units and fixed mammography units and how this could impact the improvement of the coverage of the demand served, concluding that for the state studied, by using the two types of mammography units (mobile and fixed) it could meet 100% of the demand and reduce the existing equipment.

Thus, this paper presents a two-phase algorithm for solving the problem of the distribution of mammography units. In the first phase, the mammography unit location and allocation problem is solved using a Mixed-Integer Linear (MILP) formulation, with a mono-objective approach, to maximize the demand served. So, fixed mammography units are allocated based on a set of cities, demand, coverage radius, among other parameters. The second stage consists of taking the cities with residual demand, resulting from the first phase, and solving the MMURP problem to maximize the demand covered by the MMUs and minimize the distance traveled. For this phase, a K-means constrained algorithm is developed to cluster cities with residual demand and define the depots. A nearest neighbor heuristic is implemented to create vehicle routes in each cluster. The algorithm

is tested with synthetic instances for validation of the algorithm's performance and with the instances belonging to Acre and Amazonas states. The method proposed here also helps determine the number of mobile mammography units needed to meet 100% of the demand, providing effective solutions to determine the distribution of mammography units in Acre and Amazonas states.

2. Mathematical formulation

2.1. First stage: Formulation for the MULAP

The first part of this proposal consists of solving the MULAP problem using its single-objective approach, seeking to maximize the demand resulting from the allocation of fixed mammography units in cities. The formulation presented here fits mainly with the work of De Campos et al. [2020] and also of De Campos et al. [2024]. Thus, $N = \{1, \dots, n\}$ are the candidate cities to host mammography units; d_{ij} is the distance from city i to city j in km; dem_j is the annual demand for mammography screenings in city j ; cap is the annual mammography screenings capacity of a mammography unit; p is the total number of mammography units to be located; R is the maximum distance to a mammography unit in km; S_i is the set of cities whose distance from city i is less or equal to R . $S_i = \{j \in N \mid d_{ij} \leq R\}$, $i \in N$. Unlike the work of De Campos et al. [2020] it is assumed that all cities have infrastructure, therefore all cities are candidate to host mammography unit.

The decision variables are: x_{ij} which is a 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 which is a integer variable indicating the total number of mammography units located at city i . t_i which is a 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 , a binary variable for mammography designation; $z_i = 1$ if the city i hosts mammography units, otherwise $z_i = 0$.

Mathematical modeling

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

$$\text{subject to} \quad \sum_{i \in S_j} x_{ij} \leq 1, \quad \forall j \in N, \quad (2)$$

$$\sum_{i \in N} y_i \leq p, \quad (3)$$

$$\sum_{j \in S_i} dem_j \cdot x_{ij} \leq cap \cdot y_i, \quad \forall i \in N, \quad (4)$$

$$t_i \leq z_i, \quad \forall i \in N, \quad (5)$$

$$t_i \leq x_{ii}, \quad \forall i \in N, \quad (6)$$

$$t_i \geq dem_i \cdot x_{ii} - dem_i + 1, \quad \forall i \in N, \quad (7)$$

$$t_i \geq x_{ij}, \quad \forall i, j \in N, \quad (i \neq j), \quad (8)$$

$$y_i \leq p \cdot z_i, \quad \forall i \in N, \quad (9)$$

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

$$x_{ij} \in [0, 1], \quad \forall i, j \in N, \quad (11)$$

$$y_i \in \mathbb{Z}^+, \quad \forall i \in N, \quad (12)$$

$$t_i \in \{0, 1\}, \quad \forall i \in N, \quad (13)$$

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

(1) aims to maximize the covered demand for mammography screenings. According to (2), 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%. Not all available mammography equipment must be located by (3). Constraints (4) are the capacity constraints of each mammography unit. According to (5)–(7), the binary variable t_i takes the value 1 if the located mammography fully covers the city i ; otherwise, $t_i = 0$. Constraints (8) 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 (9) and (10); otherwise, $z_i = 0$. The sets of constraints (11)–(14) specify the non-negativity and the integrity of variables for the model.

2.2. Second stage: Characterization of MMURP

The second part of this proposal consists of developing an algorithm for the MMURP. This implementation aims to minimize the distance traveled by each MMU and meet the highest possible demand for mammography screenings. To develop the MMURP model, some of the assumptions made by Rosa et al. [2020] were considered, but some differences are also highlighted: the depots are not pre-defined, they are defined during the implementation of K-means as is detailed after; the algorithm also have the flexibility to assign a free quantity of vehicles, this parameter is defined depending of the instance; each MMU has only a mamography unit inside with capacity of 6.758; a city can be visited for more than one MMU; finally, the demand can be fully or partially met by the fixed mammography units and also the mobile mammography units. The following details are essential to understand how the problem is characterized.

- There is a set of n cities with residual demand, where $\bar{N} \subset N$, all of which are candidates to be served by the mobile mammography units.
- Each city has a residual dem of mammography screenings.
- There is a set V of M mobile mammography units to visit the cities and meet the demand. The quantity of MMUs available is fixed for each experiment and some variations are of this parameter are also considered.
- Each MMU has a fixed Q capacity, set at 6.758 mammography screenings. It is assumed that each MMU has only one mammography unit.
- The MMUs depart from depots, which are defined during the algorithm execution. The depots are part of the set of cities, so its demand have to be attended too.
- There is a distance d_{ij} between the cities.
- A city can be visited more than once by different MMUs with the aim that the demand is covered as much as possible.
- The city's demand can be fully or partially met by the MMUs.
- The vehicles depart and return to the depot.

The idea here is to minimize the distance traveled by the MMUs while meeting the greatest possible demand, allowing for partial coverage by more than one mammography unit in each city.

3. Solution methods

The MULAP is solved by the CPLEX solver, which uses various advanced algorithms inside to solve the problems. For the MMURP, an algorithm is proposed that involves a restricted K-means for clustering cities with residual demands and depot definitions, and a nearest neighbor heuristic for route generation, making the MMURP a multi-depot vehicle routing problem (MVRP). Details of the methods are presented below.

3.1. K-means Clustering

K-means is a highly efficient clustering algorithm proposed by Lloyd [1982] based on the standard distance measure, calculating the average distance demand, given a set of k clusters and thus determining the centroids Chong [2021]. So, given a set of points $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ where $\mathbf{x}_i \in \mathbb{R}^d$, set of $S = \{S_1, S_2, \dots, S_k\}$ with k clusters, the objective is to minimize:

$$\arg \min_{\{S\}} \sum_{j=1}^k \sum_{\mathbf{x} \in S_j} \|\mathbf{x} - \mu_j\|^2 \quad (15)$$

where μ_j is the centroid of cluster S_j :

$$\mu_j = \frac{1}{|S_j|} \sum_{\mathbf{x} \in S_j} \mathbf{x} \quad (16)$$

Distance-Constrained Variant

For this study, as explained in Rosa et al. [2020], the cities to be served by the mobile mammography cannot be more than 500 km away from the depot, which is why a distance restriction is implemented. That is, K-means is implemented to define the clusters and centroids. After that, a verification is performed to ensure that the points are within the coverage radius of the centroid to which they were assigned. If the point does not meet the distance restriction, it is assigned to the cluster of another depot where it does meet the restriction. With maximum intra-cluster distance $D_{\max}$:

$$\|\mathbf{x}_p - \mathbf{x}_q\| \leq D_{\max} \quad \forall \mathbf{x}_p, \mathbf{x}_q \in S_j \quad (17)$$

Nearest-Neighbor Heuristic (NNH)

Considering the depots, the set of cities with demand, and the routes, the MMU must start at the depots, move toward the first customer i , and execute the route, always respecting capacity. While the candidate cities are yet to be visited, the closest feasible customer must be found, where the distance is optimal and does not exceed the vehicle capacity. The route and capacity are updated. The served customer is then removed from the list, and the MMU moves to the next node. Upon completing the route, it must return to the depot. For a depot d and city set N with demand dem_i , R route, the algorithm:

1. Start at depot: $i \leftarrow d, R \leftarrow [d], Q \leftarrow Q_{\max}$
2. While $N \neq \emptyset$:
 - Find nearest feasible customer: $i^* = \arg \min_{i \in N} \{d_{ij} | q_i \leq Q\}$
 - Update route: $R \leftarrow R \oplus i^*$

- Update capacity: $Q \leftarrow Q - q_{i^*}$
- Remove served customer: $N \leftarrow N \setminus \{u^*\}$
- Move to next node: $j \leftarrow i^*$

3. Return to depot: $R \leftarrow R \oplus d$

4. Computational experiments

The algorithms were implemented using Python 3.10 and performed on a personal computer with an Intel Core i5-1135G7 CPU 2.40GHz and 8GB of RAM running the Ubuntu 22.04 operating system. Four instances were used; the first one is to validate the proposed algorithms, the second one is data belonging to Acre state, the third one is for Amazonas state and the last one is also a synthetic instance. Thus, the MULAP presented in the first stage is developed for the group of instances containing the information below and presented in Table 1.

4.1. Description of instances

The information for the synthetic instances does not follow a specific source and for real instances is obtained from DATASUS, which also considers the projected 2020 population census of the IBGE. Each instance contains the following parameters:

- Number of cities N : this parameter corresponds to the quantity of cities by state;
- Number of mammography units p : is defined as the quantity of equipment that is available to served the demand in each state;
- Capacity of each mammography unit cap : corresponds to the number of mammography screenings that can be perform annually by a mammography unit.
- Demand of each city dem_j : it represents the demand for mamography screenings in each city;
- Maximum distance R : this value is fixed by SUS as 60km and defines the maximum coverage radius where a fixed mammography unit can be used to served the demand of some city.
- Quantity of vehicles Qty : represents the quantity of available mobile mammography units. There is not an specific method to defined this. For this work, were consider 2,3 and 4 vehicles for instances with nearest cities in terms of distance and 5,6,7 for vehicles for cities with more distance cities.
- Maximum distance D_{max} : represents the maximum coverage distance in which a mobile mamography unit can be used to meet the demand of the cities. This parameter is fixed by SUS in 500km.
- Distance matrix $d_{i,j}$: this is a symmetric matrix with size $N \times N$ containing the distance between cities.
- Coordinates of cities $Coor_i$: this is a list containing the latitude and longitude for each city.

Table 1: Data of instances

Instance	N.Cities	N.Units	Capacity	Demand	State
Inst 8	15	140	2000	24000	N/A
Inst 22	22	250	6758	453517	Acre
Inst 62	62	72	6758	276343	Amazonas
Inst 184	184	325	8000	4760409	N/A

The algorithm for MULAP is executed and then returns the cities whose demand was not 100% met by the fixed mammography units. In other words, any city with residual demand is considered a candidate for service by the MMUs. This information, such as the coordinates of the cities and the distances between them, becomes an input for the MMURP algorithm. The second stage algorithm is executed, and the results for the instances are presented in Table 2. There were performed 12 experiments varying three times the quantity of vehicle for each instance.

4.2. Analysis of results

The solution is analyzed considering the demand coverage in each stage. The information is presented in Table 2, The first column *Cities* corresponds to the number of cities in each instance, the column *TotalDem* corresponds to the total demand of the cities, the column *DemMF* is about the quantity of demand that is covered by the algorithm in the first stage and *DemMM* is about the quantity of demand that is covered by the algorithm of the second stage, the column *TotalAtt* exposed the quantity of coverage demand was fully or partially covered by the two algorithms, and finally, the last column *Qty of vehicles* represents the quantity of vehicles for each experiment.

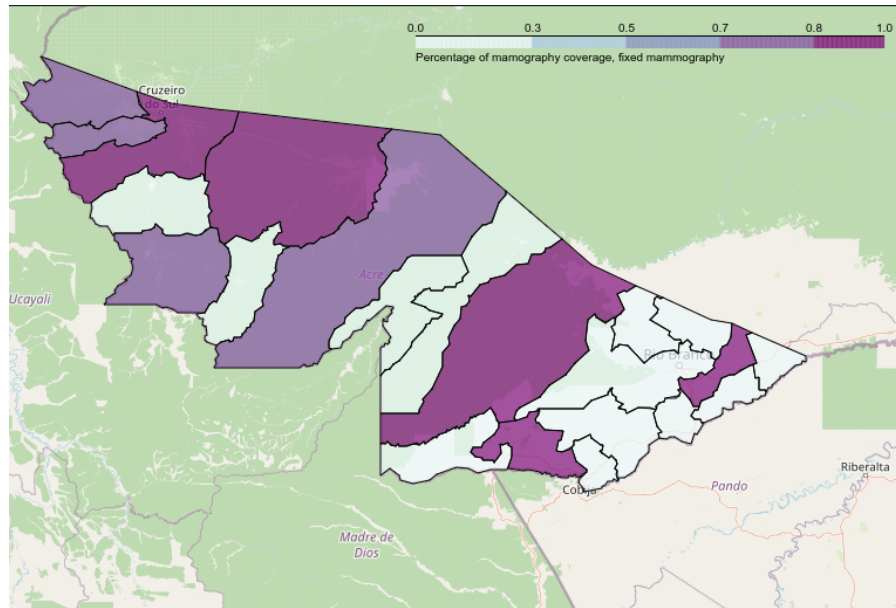
Table 2: Solutions for the algorithms for variations in quantity of vehicles

Cities	TotalDem	DemMF	DemMM	TotalAtt	TotalAtt/TotalDem	Qty of vehicles
8	24000	10000	10748	20748	0.8645	2
8	24000	10000	12758	22758	0.9482	3
8	24000	10000	14000	24000	1.0000	4
22	453517	148676	29233	177909	0.3922	5
22	453517	148676	42749	191425	0.4220	7
22	453517	148676	56014	177909	0.4513	10
62	276343	275500	843	276343	1.0000	2
62	276343	275500	843	276343	1.0000	3
62	276343	275500	843	276343	1.0000	4
184	4760409	2196350	40000	2236350	0.4697	5
184	4760409	2196350	51876	2248226	0.4722	7
184	4760409	2196350	75876	2272226	0.4773	10

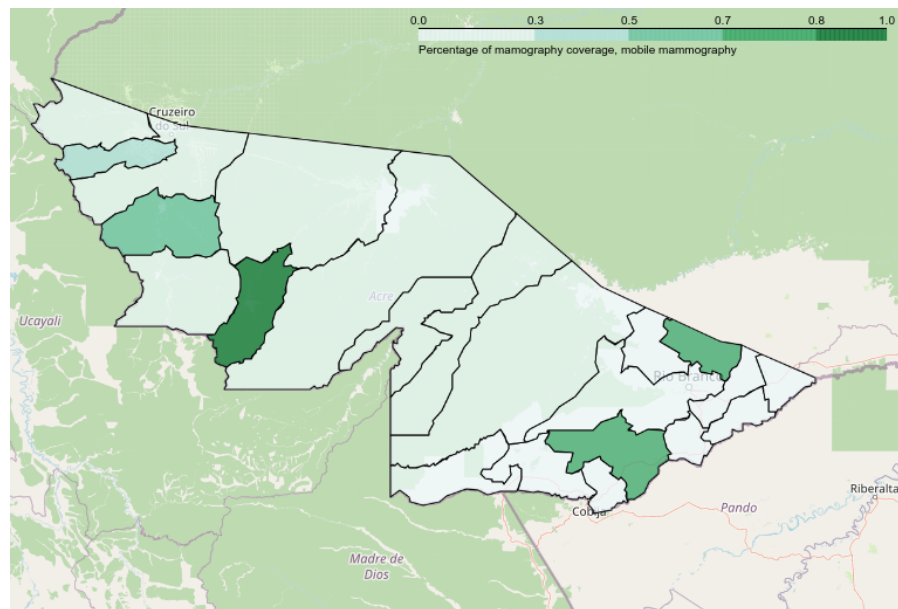
In the case of the instance of 8 cities, in the first stage, the demand of cities 3,5,6,7 are fully covered by the fixed mammography units and the city 8 is partially covered by the city 5 which have all the quantity of mammography units located. In the second stage, the quantity of mobile mammography units were varied in 2, 3 and 4 vehicles. For the 4 vehicles, the cities with residual demand were covered in a 100% for the MMUs. Here, the algorithm allocated one vehicle per cluster, which is only composed by one city.

For the case of Acre state, just 2 cities were fully covered for the fixed mammography units, 9 cities were partially covered, and 11 were not covered, so all these residual demand passed to the MMRUP algorithm. Here in order to achieve and equilibrium, were determinate 1 vehicle per cluster, so the five vehicles achieved cover an additional percentage of the demand as is shown in Figure 1, where (a) represent the percentage of coverage for fixed mammography units and (b) represents the percentage of coverage for mobile mammography. In total, 47 MMUs will be

necessary to cover all the demand, especially considering that the city of RIO BRANCO has a considerably high demand compared to the other cities in this state. As the number of vehicles available in the second phase increases, so demand covered also increases.



(a) Acre state, fixed mamography

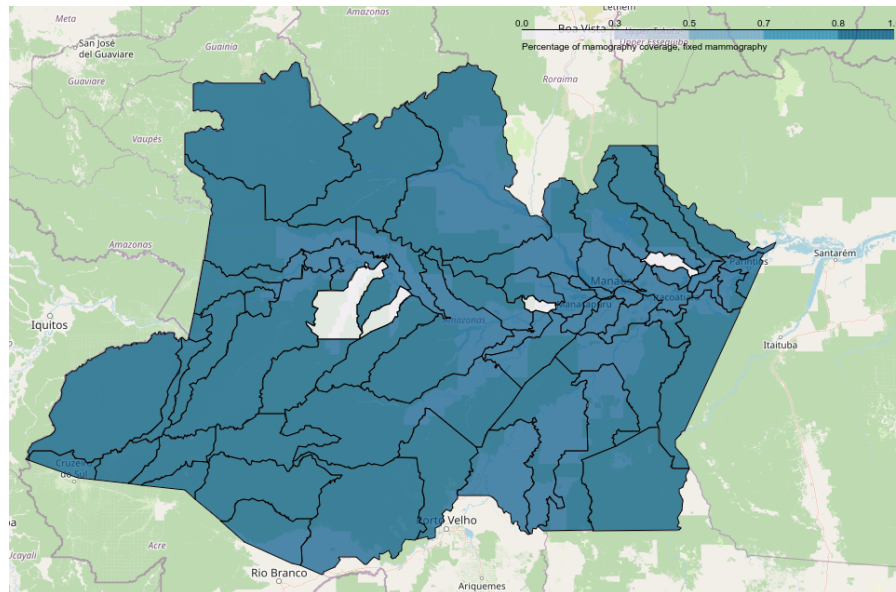


(b) Acre state, mobile mamography

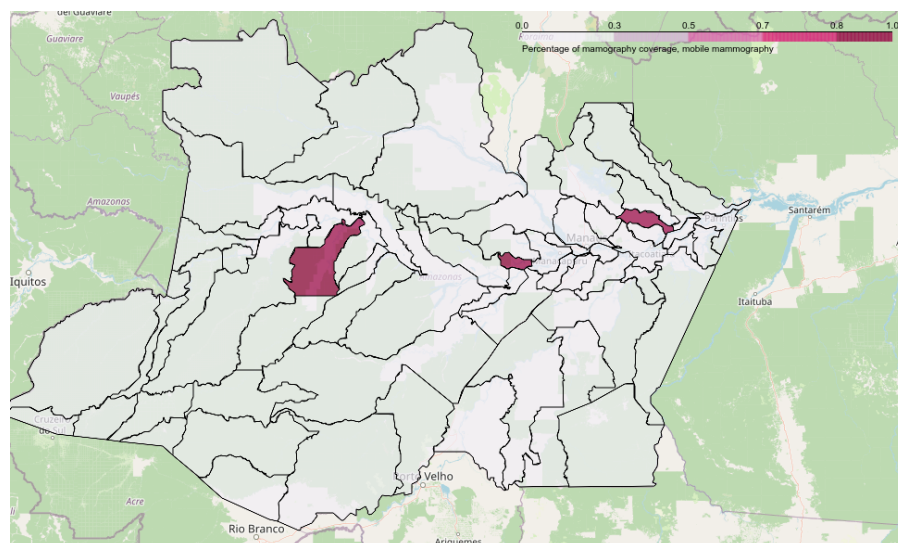
Figure 1: Percentage of demand coverage for Acre state

For the instance of 62 cities, corresponding to Amazonas, 59 cities were full covered for the mamography fixed units. Here, the MULAP allocates mamography units in most cities, allowing for full coverage in most cases due to low demand. The remaining cities have full coverage for the mobile mamography units too, achieving a 100% of demand coverage in all cases when

the different quantity of vehicles are considered as is shown in Figure 2 where (a) represent the percentage of coverage for fixed mammography units and (b) represents the percentage of coverage for mobile mammography.



(a) Amazonas state, fixed mamography



(b) Amazonas state, mobile mamography

Figure 2: Percentage of demand coverage for Amazonas state

Finally, for the synthetic instance of 184 cities, 101 cities were fully covered by the fixed mammography units and the remaining demand of the other cities were passed to the second stage, where the demand coverage was about 47% for all the variations of quantity of vehicles. Here, demand coverage is not significantly affected by the number of vehicles. Most cities were not covered by the MULAP and were passed as candidate cities for the MMRUP. However, the 500km distance restriction significantly affects the vehicles' ability to cover the cities, as well as their

limited capacity.

5. Conclusions

The demand coverage rate is one of the main problems facing the distribution of mammography units in Brazil, so alternative solutions are needed to address this situation. Currently, both fixed and mobile equipment are available in several states across the country, making it important to decide how to use them to achieve the greatest possible coverage.

Thus, the results demonstrate that mobile mammography units can help increase coverage in cities that fixed mammography units cannot reach. However, restrictions such as the coverage radius of fixed mammography units and the distance from the depot to candidate cities that a mobile mammography unit can cover reduce the possibility of meeting the demand in a larger number of cities, especially in states where the distances between cities are very large.

Some limitations of the study are based on the impossibility of comparing it with the actual situation of demand management in the states considered, since DATASUS does not have yet a record of the number of mobile mammography units currently serving the states' demand, the routes they follow and also information related to the depots of MMUs. Additionally, random values are considered for variations in the number of vehicles, for example, which also limits the comparison of the problem with reality.

As future work, it would be interesting to implement the algorithms for other states in Brazil. Likewise, it would be important to consider additional objective functions and perform multiobjective approaches.

6. Acknowledgments

The authors would like to acknowledge the support of CNPq (process 309925/2021-5) and FAPESP (process 2022/05803-3 and 2020/09838-0).

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.
- Cevik, M., Ayer, T., Alagoz, O., and Sprague, B. L. (2018). Analysis of mammography screening policies under resource constraints. *Production and Operations Management*, 27(5):949–972.
- Chong, B. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, 4(5):37–40.
- Da Cunha, G. N., Vianna, C., Mosegui, G. B. G., da Silva, M. P. R., and Jardim, F. N. (2019). Breast cancer screening: modeling improvement of access using mobile mammography units. *seguimiento del cáncer de mama: modelo de mejora del acceso con el uso de mamógrafos móviles. Revista Panamericana de Salud Publica= Pan American Journal of Public Health*, 43.
- De Campos, M. V. A., de Assis, R. d. S. L., Souza, M. J. F., de Siqueira, E. C., Silva, M. A. L., and de Souza, S. R. (2024). Multi-objective mammography unit location–allocation problem: A case study. *Operations Research for Health Care*, 41:100430.
- 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 *International Conference on Enterprise Information Systems*, p. 428–439.

- Duffy, S. W., Tabár, L., Yen, A. M., Dean, P. B., Smith, R. A., Jonsson, H., Törnberg, S., Chen, S. L., Chiu, S. Y., and Fann, J. C. (2020). Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. *Cancer*, 126(13):2971–2979.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.
- 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): 1857–1864.
- Rosa, O. A. S., Rosa, P. M., Penna, P. H. V., and Souza, M. J. F. (2020). Um algoritmo construtivo para o problema de roteamento de unidades móveis de mamografia. *Simpósio Brasileiro de Pesquisa Operacional*.
- 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.