ORIGINAL ARTICLE

Location of ecopoints through integration p-medians model with geographical information system

Localização de ecopontos através da integração modelo p-medianas com sistema de informação geográfica

Jonnathan Vinicius Lopes Silva[1], Narciso Ferreira dos Santos Neto[1]

[1]Universidade Estadual de Montes Claros – UNIMONTES, Programa de Pós-graduação em Modelagem Computacional e Sistemas – PPGMCS, Montes Claros, MG, Brasil. E-mail: eng.jonnathanlopes@gmail.com; narciso_santos1@yahoo.com.br

Abstract: The configuration of logistics networks for waste oil and fats collection services is a strategic point of great importance, since their disposal contributes to a number of environmental disturbances, problems in sewage pipes and wastewater treatment systems. Starting from this premise, this paper proposes location of ecopoints in the city of Montes Claros-MG, capable of optimizing the collection of the waste oil and fats in all urban area of the municipality. This proposal of location was constructed in two stages: in the first, the collection network was constructed by integrating a location model of facilities with a geographic information system (GIS). It was possible to model a collection network in the city of Montes Claros - MG containing 87,748 points referenced by geographic coordinates and representatives of residential points considered as potential generators of these residues. During the creation of the model, due to the high density of the network and the low quantity of facilities (p), which causes limitation in the solution, it was observed the need for thematization of the map and consequently the network partitioning in zones, which allowed to determine the initial amount of installations that were to be found in the solution. Finally, an application in the native language of Transcad, GISDK, was developed with a programming interface that was developed to automate the process of integration of the model in the GIS. The model allowed the creation of 12 networks with densities below 12000 points, allowing gain in the solution of the problem.

Keywords: Geographic information system; Facility location; OGR; P-medians.

Resumo: A configuração de redes logísticas para serviços de coleta de óleo e gordura residual (OGR) é ponto estratégico de grande importância, visto que a sua disposição de forma incorreta contribui para uma série de distúrbios ambientais, problemas em tubulações de esgoto e sistemas de tratamento de água. Partindo dessa premissa, o trabalho propõe localização de ecopontos na cidade de Montes Claros-MG, capazes de otimizar a coleta do OGR em toda área urbana do município. Essa proposta de localização foi construída em duas etapas: na primeira, deu-se a construção da rede de coleta de OGR através da integração de um modelo de localização de facilidades a um sistema de informação geográfica (SIG). Foi possível modelar uma rede de coleta na cidade de Montes Claros – MG contendo 87.748 pontos referenciados por coordenadas

geográficas e representantes de pontos residenciais considerados potenciais geradores desses resíduos. Durante a criação do modelo, devido à densidade da rede ser elevada e a quantidade de facilidades (p) ser baixa, o que causa limitação na melhoria da solução, observou-se a necessidade de tematização do mapa e por consequência o particionamento da rede em zonas, o que permitiu determinar a quantidade inicial de instalações que deveriam ser encontradas na solução. Por fim, foi desenvolvido um aplicativo na linguagem nativa da Transcad, GISDK, com uma interface gráfica cujo objetivo foi automatizar o processo de integração do modelo no SIG. O modelo possibilitou a criação de 12 redes com densidades inferiores a 12.000 pontos, permitindo ganho na otimização da solução do problema.

Palavras-chave: Sistema de informação geográfica; Localização de instalações; OGR; P-medianas.

1 Introduction

With the increase in urbanization and the creation of new points of food consumption, the generation of waste in society increases significantly. An example is ORF, oil and residual fat, obtained from frying in the preparation of food in homes, food consumption, and industrial establishments. These residues, in particular, often do not have an adequate collection and disposal system, creating the problem of viable logistics for their correct disposal, since they cannot simply be discarded in the environment due to the environmental impacts they may cause (Castellanelli, 2008).

When improperly disposed of in sewers, the oil mixes with organic material causing inconvenience for water treatment companies, increasing maintenance costs due to clogging of pipes and grease boxes. Another way to dispose of these residues is through storm drains, causing obstructions and retention of solid residues that can contribute to floods. Also, ORF can pollute the water table, since there are sewage network connections with rivers, lakes, and streams, which occur in several Brazilian municipalities. In this situation, there is a decrease in oxygenation causing the death of creatures that live in these places (Pitta et al., 2009).

In 2016, Brazilian production of vegetable oil was 8.52 million tons and the country had a consumption of 7.54 million tons, ranking third in the domestic consumption of soybean oil. Among the countries presented, Brazil is only behind China and the United States (USDA, 2016). These rankings prove that it is necessary to create strategic and viable solutions for the correct destination of these residues, minimizing impacts and creating alternatives that generate returns for society.

Seeking to solve this problem, the Basic Sanitation Company of São Paulo (SABESP) created the frying oil recycling program (PROL) in 2007, with the adhesion of 1500 residential condominiums, which were within the project acting area. This had a positive impact that was monitored through the number of pipes clearances. This project also provided income for the needy population involved in waste collection and was disseminated in partnership with SABESP to other municipalities such as Osasco, Registro, Itapetininga, Lins, Jales, Presidente Prudente and Pindamonhangaba (SABESP, 2017).

The study by Castellanelli (2008) on the feasibility of collecting and recycling frying oil in the city of Santa Maria in the Rio Grande do Sul to obtain biodiesel presents conclusions that recycling these residues brings economic, environmental and public health benefits. He also pointed out that only in the neighborhood that was the focus of his research, 3,623 liters of frying oil were released monthly into the environment, with an average monthly generation of residual oil of 0.21 liters per person and in the entire city, it was estimated that 33,371 liters of frying oil was launched into the environment daily.

Gândara (2015) in his thesis on the production and disposal of residual frying oil in the city of São Paulo, highlights the use of this residue as a raw material for the production of

resin for paints, soap, detergent, fabric softener, soap, glycerin, animal feed, lubricant for agricultural machinery and biodiesel. He also estimated that only 2.5% to 3.5% of the vegetable oil used in food preparation in the country was recycled.

According to Guabiroba & D'Agosto (2011), the use of renewable biofuel is an attraction for countries worldwide because it has a lower level of pollution than other fuels, such as diesel oil. This form of renewable energy can be obtained from different sources of raw materials, like vegetable oils from soybeans, canola, sunflower, peanuts, castor beans. However, the fact that biodiesel is also obtained from residual raw material such as animal fat and residual frying oil is highlighted here. A constraint in the use of these residues as raw material is the cost of collection, as they are not concentrated in a given location, but are dispersed throughout the urban area.

The central issue of this paper is the lack of disposal points, and consequently, the lack of an ORF collection system in the city of Montes Claros-MG since these residues cannot be collected by the same conventional urban collection system since it needs specific containers, a specific destination location, a specific vehicle for collection, besides increasing the operational time of collection due to the correct handling. In addition, there is a potential for reusing these residues due to the presence of a Petrobras biodiesel plant installed in the municipality. A collection system can contribute to the development of the region by producing biodiesel through ORF, creating a source of income for waste pickers, also minimizing the impacts caused to the environment by avoiding the disposal of this waste in the sewage network.

In this context, the objective of this research was to develop a model for locating ORF collection points in the city of Montes Claros-MG. Thus, it is expected to contribute to the feasibility of collecting these residues, and to enable their use in the production of biodiesel and other products, at an acceptable cost. A fundamental step to make this collection feasible is the correct location of the disposal points, referred to in this article as eco points.

The optimized choice of eco points allows greater coverage of disposal in generating areas, in addition to assisting the agencies or companies responsible for the collection in making the decision to choose these points. The central objective of this paper is the optimal location of eco points through a model of the location of facilities existing in the literature.

The analysis of the location of facilities aims to define the places for the installation of facilities that should serve customers in an optimized way concerning some criteria of choice. The location of facilities has a very broad approach and is of fundamental importance, as it directly influences the context that is applied. An example of facility location decision-making is when you need to define locations for vehicle patrol bases, fire stations, and outpatient clinics. The correct and optimized location of these facilities is of fundamental importance, as it can preserve lives. In the context of transport, when this location has a poor decision, it may increase unnecessary costs in the logistics processes (Daskin, 2013).

According to Oliveira (2012), the location problems of the facilities are based on each demand point at the nearest facility and then locate that facility at the center of those points. The distance between points and facilities is considered weights or costs. Equipment location problems can be discrete or continuous, and various models and heuristics for treating problems considered NP-Difficult, that is, there is no optimal solution in polynomial time for all cases.

One of the solution methods addressed in a large class of facility location problems in various contexts is described by Pereira (2005) as p-medians. This method consists of selecting p vertices of a graph representing the network to minimize the sum of the distances of each vertex to its nearest median. This method can be applied to problems involving the demand for points

as restrictions, then called as enabled p-medians and in problems that only the distances between points are considered, called un-enabled p-medians.

This paper was based on the Lagrangean / surrogate technique through the combined use of heuristic techniques of Lagrangian relaxation and optimization by subgradients applied to the p-median model and described by Senne & Lorena (2000), in which the algorithm used to find facilities is available on the author's website so that other researchers can apply in modeling their problems.

As the central issue of this paper is characterized as a transport problem, it was necessary to use geographic information systems (GIS) in its modeling. Nyerges (1990) defines GIS as an integrated system of hardware, software, people, data, and institutional tools for collecting and disseminating data in terrestrial areas. According to Fitz (2008), a GIS comprises a software package that integrates all of these resources previously mentioned in its platform to create a geographic database and also allows acquiring, store, process, manipulating, and visualizing spatial data that are referenced by geographic coordinates.

Brito (2006) brings an analysis of the use of GIS to solve the problem of transporting solid urban waste to calculate the necessary vehicle fleet and the generation of itineraries for the final disposal of this waste. Thus, it shows the economy of the computational model in solving the problem when compared to the empirical method that is intuitive and based on the particular experience of operators.

Brasileiro & Lacerda (2008) present the analysis of the use of GIS in the routing of vehicles for the collection of household solid waste. The work presents the use of the platform to solve routing problems through the arc model that aims to minimize the total extension to be traveled by the collecting vehicles. The authors show a 41% reduction in the total distance traveled and a decrease of 68% in the total travel time comparing to the real situation. The GIS provided the use of real information on distance and travel time of the analyzed road network and made it possible to change parameters for modeling the problem.

Biju (2015) presents in his dissertation on the application of GIS in the solution of problems related to civil construction solid waste. He points out the importance of geographical information systems in decision-making in integration with multi-criteria Decision Analysis (MCDA), reducing time and cost in solving the problem of indicating areas for disposal of these residues.

This paper seeks to contribute to the integration of the Geographic Information System and the model of the location of facilities in the treatment of the problem of dispersion of the points inserted in a waste collection network, in this case in particular the oil and residual fat.

2 Literature review

2.1 Geographic information system

This paper uses GIS as a modeling tool, defined as a data processing system in a graphical or alphanumeric form with a focus on the spatial analysis of these data. According to Nielson (2014), Geographic information systems become, an increasingly, a powerful tool in solving highly complex problems. This is motivated by its ability to integrate the visualization and analysis of geographic data in several application domains with the aid of computational tools with high processing capacity. The application of GIS

platforms has become increasingly common due to human needs in the search for solutions to various problems in a variety of applications.

Longley et al. (2005) cite a series of problems of geographic location that can be solved with the help of the GIS, such as deciding the location of new clinics, the decision on routes and schedules of vehicles from delivery companies, selection of routes for new highways, better installation of retail stores, better management of forests for cutting, roads and new plantings, decision-making in the agriculture sector on the application of fertilizers and pesticides in different parts of the plantation.

According to Pina & Santos (2000), GIS are computational systems that aim to assist in the process of interpreting situations that occur in geographic space. With a GIS, it is possible to achieve the integration of conventional database operations such as capture, storage, manipulation, analysis, and presentation. It should also be noted that a GIS is capable of performing statistical analysis and visualization of maps, all of which distinguish a GIS from a conventional information system.

Longley et al. (2005) define that the anatomy of a GIS is composed of six components, as can be seen in Figure 1 below. The first part described and the most fundamental is the Network, it allows fast communication as well as sharing digital information and using the Internet to access data and geographic maps. The second component is the hardware that is defined as a device with which the user will perform operations with the GIS through basic commands such as typing, clicks, notes. This device allows the return of information visually or audibly.

The third component is the Software that runs on the user's local machine and has macro scripts to optimize tasks that are performed repeatedly by the user. Examples of GIS software providers are ESRI developer of ArcGIS, MapInfo Corp, Caliper developer of Mapitude, and Transcad among many others. Some of this software has its routines to deal with transport problems such as vehicle routing problems and the creation of minimum path arrays.

The fourth component of GIS anatomy is the database that consists of the digital representation of aspects of some specific area of the earth's surface that will be used to treat problems. The GIS also has as a component the procedures, management, control, and reports keeping the activities of the GIS always within established budgets. Finally, the GIS has the user as its component, the people who will program, design, supply and interpret data according to specific skills.
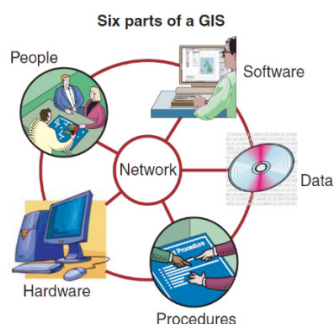

Figure 1. Six components of GIS. Source: Longley et al. (2005, p. 24).

Peuquet & Marble (1990) have a more specific approach to the anatomy of a GIS by describing that most platforms have the following main components:

- Data entry system that captures and or processes spatial data derived from existing maps;

- Data storage and retrieval system that allows organizing spatial data so that it can be retrieved quickly by the user, allowing updates and corrections of spatial databases with precision;

- Data analysis and manipulation system with several functions, such as, for example, creating rules for changing data or creating parameter estimates and restrictions to optimize space and time for simulation models;

- Data communication system between original and modified data, in addition to the possibility of generating an output called a digital map or digital cartography.

It must pay attention to the quality of data and spatial attributes that are manipulated in operations in the GIS. Spatial data have errors and uncertainties inherent in their nature and need to be corrected and not manipulated to be hidden from the user's reach. It is a mistake to attest to the quality of a final GIS product due to its appearance independent of the applied cartographic visual techniques.

Errors in spatial data can occur in several stages of the modeling process, for example, in the initial and conceptual perception, approximations of determination of geographic locations that directly depends on the ability to acquire hardware information such as (GPS and satellites), choose of map projections, in the measurement of parameters inherent to the natural variation of the phenomena or precision of the measurement and recording device made by the observer. For these variations and errors in spatial data to be eliminated or reduced to an acceptable level, it is necessary to know the errors, propagation parameters, and standards and tools for the optimization of sampling and treatment of these errors (Burrough, 1986).

Thill (2000) presents the perspective of geographic information systems in transport applications in several problem domains, for example, for the treatment of traffic flow, peak hours, road infrastructure, among other applications, it is called SIG-T. Naturally, the SIG-T Softwares have modeling, manipulation, data analysis, routines, and facilities for handling problems related to transportation, more appropriate to transportation that are not served by conventional GIS software. Basically for transporting application, three classes of GIS models are essential:

• Field models: representation of the continuous variation of a given phenomenon over space. For example, the elevation of the land.

• Discrete models: in which discrete objects such as points, lines, or polygons make up space. Highways, urbanized areas, barriers can use this model.

• Network models: used to represent linear connections fixed on the surface with continuous reference. Roads and flight airlines use this model.

The representation of networks has great importance in transport applications, as these problems often only need a network model to represent the data. This model is based on the representation by arcs and nodes and allows the representation of objects through x and y coordinates (longitude and latitude). Network models are associated with the creation of graphs that, in a one-dimensional form, can represent objects referenced on the Earth's surface in a two-dimensional or three-dimensional form. An example relevant to the use of graphs in problem-solving was made by Goldbarg & Goldbarg (2012) is on Figure 2.
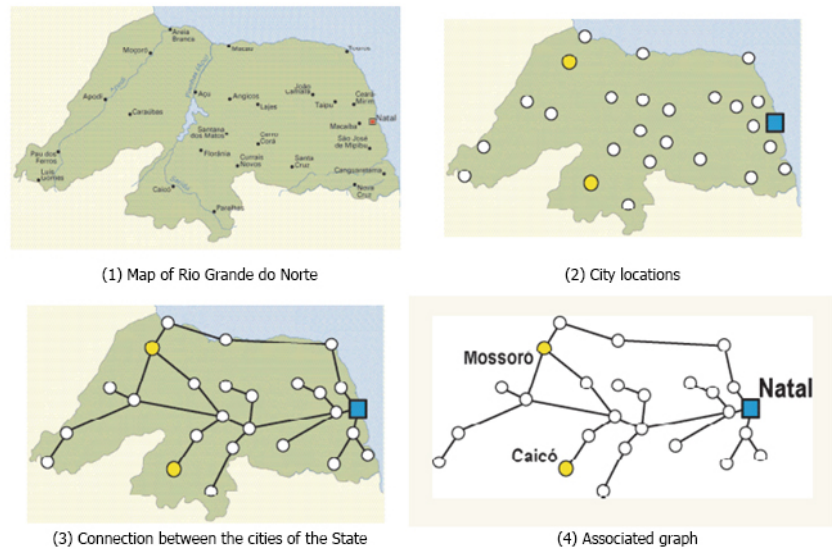
(1) Map of Rio Grande do Norte

(2) City locations

(3) Connection between the cities of the State

(4) Associated graph

Figure 2. Fork formation from a map. Source: Goldbarg & Goldbarg (2012).

It is possible to visualize that the nodes in Figure 2 represent the cities of the state of Rio Grande do Norte while the arcs that are the connections between these nodes represent the roads or paths existing between these cities. It is important to note that the representation of networks and arches does not necessarily have to have the same geographical course as real roads, as in this specific case they only represent a connection between cities. From this network, it is possible to associate a distance or minimum cost matrix that represents the minimum distance between all points that have a path between them.

2.2 Location of facilities

To determine the collection points, it was necessary, in addition to using the GIS defined in the previous topic, to use a location optimization model in the studied scenario. Decisions to locate facilities are considered strategic in distribution systems. It is possible to highlight these decisions when industrial sectors need to define places for the installation of factories, points of sale of stores, when the public sector needs to define places to install schools, hospitals, and other services for society. In each of these segments, the location of facilities has its due importance, whether in terms of production and marketing capacity or the quality and scope of the service provided.

The location models can be of the network location type, discrete location, and mixed whole programming model. Facility location problems have many approaches for different applications in which it is possible to highlight simple linear models, non-enabled and enabled models, non-linear, deterministic and probabilistic models. The objectives in location models can be of the minsum or minmax type. The minsum models aim to minimize the average distances. The minmax models aim to minimize the maximum distances.

Unqualified models are not restricted about demand allocation, however, there are situations where capacity restrictions must be obeyed and this demand must be allocated with due care. There are localization models called the single-stage, and they are focused on only one phase of the distribution model, different from multistage models that need to

obey several hierarchical levels. Static location models aim to optimize performance for a certain time, while dynamic models have data such as cost, demand, the capacity that varies over time (Klose & Drexl, 2005).

According to Sule (2001) installation location problems are usually categorized into the following classes: p-median, p-center, non-enabled, enabled, and quadratic. Highlighting the non-enabled model, which is the focus of this paper, it is possible to define it as a minsum type location problem and contains a fixed cost, and in this problem the capacity of each installation is unlimited.

2.2.1 P-median problem

The location model for integration with the GIS used in this paper for the installation of eco points is called P-medians. This model can be defined through a network of n-vertex installations where N = (1 ..., n). Assigning a symmetric matrix of distances between these vertices, called matrix $[d]_{ij}$ of cost and an allocation matrix $[x]_{ij}$, where $X_{ij}=1$ if **j** is a median and $X_{ij}=0$ otherwise. It is also possible to infer on the allocation matrix that $X_{ij}=1$ if node **i** is allocated at median j and $X_{ij}=0$, when node **i** is not allocated at median **j**. According to the definition of ReVelle & Swain (1970), the p-median problem can be defined with the following mathematical model:

$$\min \sum_{j=1}^{n} \sum_{i=1}^{n} a_i x_{ij} d_{ij}$$

The objective function presents the term a_i which denotes the demand for vertex i. When the objective function takes this term into account, the problem is defined as a qualified P-median model or PMC, as it takes into account the demand of the network's vertices. When this variable is not considered in the objective function, the problem is said to be non-enabled P-medians or PMNC, considering only the cost matrix or minimum paths between the vertices forming the network.

Three restrictions are necessary for this type of treatment, the first (Equation 1) being the guarantee that each node i can be allocated in only one median j. Thus, the problem creates a unique set of points allocated to each node determined as median and the different sets created have no allocation relationship between them, which is, each one has its allocation independently of its respective median.

$$\sum_{j=1}^{n} x_{ij} = 1, \ \forall \ i \in N \tag{1}$$

The second constraint (Equation 2) ensures that a median that has its nodes allocated is not allocated to another node. Thus, if i is allocated to j $(X_{ij}=1)$, then vertex j cannot be allocated to any other vertex except itself, with $X_{jk}=0$ and $X_{jj}=1$.

$$x_{ij} + \sum_{k=1}^{n} x_{jk} \le 1, \ \forall \ i,j \in N, \ k \ne j \tag{2}$$

The third constraint (Equation 3) is the one that defines the number of medians (p), being the sum of vertices that allocate itself and not allocated to any other vertex. Thus, if

in a problem it is necessary to find 3-medians (p = 3), the network must have 3 sets of points allocated at 3 distinct vertices.

$$\sum_{j=1}^{n} x_{jj} = p \qquad\qquad (3)$$

In this paper, an important study is related to the density of the network to be modeled for the collection of OGR. In concluding the paper of Zhao et al. (2015) on network density and the solution of the p-median problem, the authors state that the choice of network density is a crucial step for the final results of the solution p -medians. In very dense networks, the final result is impaired.

Han et al. (2013) examined solutions found in p-median location models for networks with a density variation of 500 to 70,000 points and concluded that for densities greater than 10,000 points, the solutions found had no improvement and gradually worsened .

Zhao et al. (2015) performed density variation tests to solve the p-median problem on a network of 67,020 and observed that in addition to consumption and computational effort significantly increase to densities greater than 12,000, the solutions did not show improvement or have improvement insignificant solution. The authors conclude that the choice of network density affects the final solution to be found by the p-median problem.

3 Methodological procedures

3.1 Construction of the SIG-P medianas integration model

To achieve the proposed objective of generating a model for the location of eco points for the collection of OGR in the city of Montes Claros-MG, quantitative-descriptive research was adopted, in which the work steps are integrated, as shown in Figure 3. In this figure, note the steps in the process of integrating the facility location model into the GIS to solve the problem.
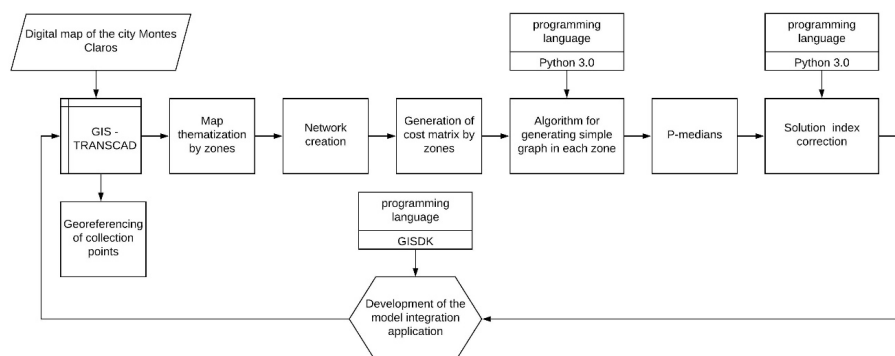

Figure 3. Process of integration of the p-median model to GIS. Source: Elaborated by the author.

The first stage presented consists of the virtualization of the city map of Montes Claros on the SIG TransCAD platform. The geographic information system chosen is

manufactured and marketed by the company Caliper and has routines aimed at transport applications that a conventional GIS does not have.

The second stage consists of georeferencing the residences on the virtualized map to create a collection network. At that time, residential points were inserted through geographic coordinates that can be considered OGR generators and are spatially dispersed in the city. The collection points comprise a database with 87,748. After the insertion of the points forming the collection network, it was necessary to carry out the thematization of the map and the creation of zones, modeled according to the regional centers found on the Montes Claros city hall website in the infrastructure and urban services section.

After the virtualization of the map and insertion of the collection points, it was possible to obtain the cost matrix responsible for providing the term d_ij shown in Equation 1. It was found that the cost matrix generated in TransCAD generates a multigraph, that is, a graph with parallel loops and arcs. This problem was solved with the development of an algorithm written in python 3.0 to generate a simple graph from the TransCAD cost matrix. The written code generates a DAT format file, which consists of a simple graph of labeled edges with minimum distances between points in the network. This file is an entry for the p-median program.

The next step in the methodological process was the application of the p-median model to the modeled collection network. After the application of the program pmedian.exe (Senne & Lorena, 2000), it was verified that the generated output provides an error in the indexes of the facilities found and their associated nodes. This error was corrected with the development of an algorithm in python 3.0, this was of fundamental importance for correctly integrating the model with the GIS TransCAD.

Finally, the integration validation consists of the development of a script in the native language of TransCAD called GISDK. This script allows you to allocate the medians on the digitized map of the collecting network and graphically present the connection of each node in the graph to its median.

3.2 Characterization of field research

Montes Claros is a municipality located in the north of the state of Minas Gerais, has a territorial area of 3,568,941 km2, according to the demographic census of IBGE (2011), O an estimated population of 400 thousand inhabitants, 95% of which live in urban areas and the rest in rural areas.

As in many Brazilian municipalities, the municipality does not have any computational tool to position and optimize OGR collection, which is then done empirically, most of the time, and due to the population growth in the urban area, it is done A policy for the collection of OGR is necessary, which must be generated in scattered points of its demographic area.

Through the database provided by the transport laboratory of the State University of Montes Claros, it was possible to model the city map on the SIG TransCAD platform as can be seen in Figure 4 below.
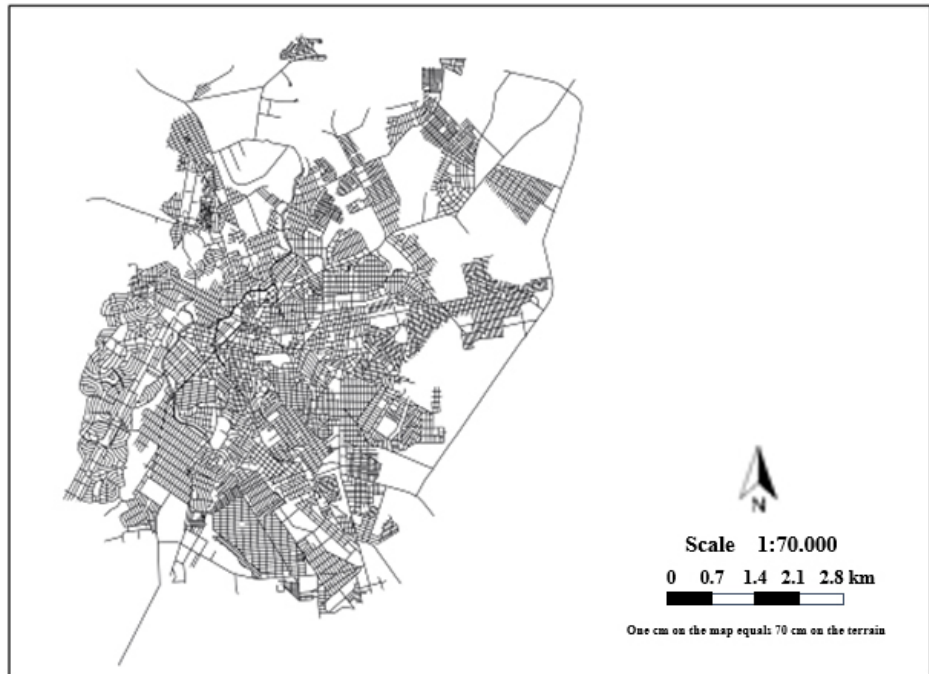
Figure 4. Digital map of the city of Montes Claros Virtualized in the GIS. Source: Unimontes Transport Laboratory.

It is possible to observe in Figure 4 that the virtualized map presents the study region with the streets represented by lines formed by points referenced by pairs of geographical coordinates. This virtualization, as approached by Nielson (2014), allows visual and spatial analysis of these data becoming a powerful tool in dealing with geographical problems that could not be solved without the aid of the use of a specific system as is the case with the problem addressed in this paper.

4 Discussion of results

Unlike the representation of networks presented by Goldbarg & Goldbarg (2012), the GIS manages with a certain degree of precision, depending on the error handling as described by Burrough (1986), the spatial data collection instruments and the cartographic digitization, to represent the route of the streets, roads and highways of a given modeled area. This allows the treatment of transport and location problems, as it is possible on the SIG platform to calculate the distance between points taking into account the actual distance from the studied scenario, in addition to the other specific analyzes and routines, such as vehicle routing.

The motivation for using GIS in this research was due to its great potential for processing geographic data. The use of GIS in location problems has proved to be efficient in several studies of the location of facilities, as in the paper of Carnasciali & Delazari (2011) that sought to solve the location problem of the banking sector by integrating ArcGIS with a specialist location system. The authors implemented the problem in the GIS through the cartographic adequacy of the city of Curitiba provided by the Curitiba Urban Research and Planning Institute.

Khan & Samadder (2014) also addressed the potential of geographic information systems in municipal solid waste management with auxiliary methods. The authors presented a GIS aid approach for landfill location and optimization of collection routes. The use of GIS reduces the total cost of collecting solid waste.

A relevant study carried out by Mapa & Lima (2012) addresses the qualities of solutions to problems of location-allocation of facilities generated by a GIS obtained after the combined use of facility location routines and transportation problems. The GIS presented efficient results in non-qualified models, as is the case of the model developed in this research and described by Sule (2001), as a location problem of the minsum type with a fixed cost, in this problem the capacity of each installation is unlimited.

Considering the current works and the gains in the treatment of geographic data that the GIS made possible in different contexts and studies, this paper approaches from the same perspective, the integration of the GIS and the model of the location of facilities to create a logistic network of collection of OGR in the municipality of Montes Claros-MG, seeking to highlight the processing capacity of the GIS in the search for optimization of solutions to complex problems.

After the virtualization of the map in the GIS as described in the methodological procedures, it was observed in the insertion of the collection points in the GIS TransCAD, the need for thematicization of the map in 12 zones corresponding to the 12 regional poles presented on the Montes Claros city hall website in the menu infrastructure and maps. The theme of the map is shown in Figure 5a, where each regional pole is differentiated by color.
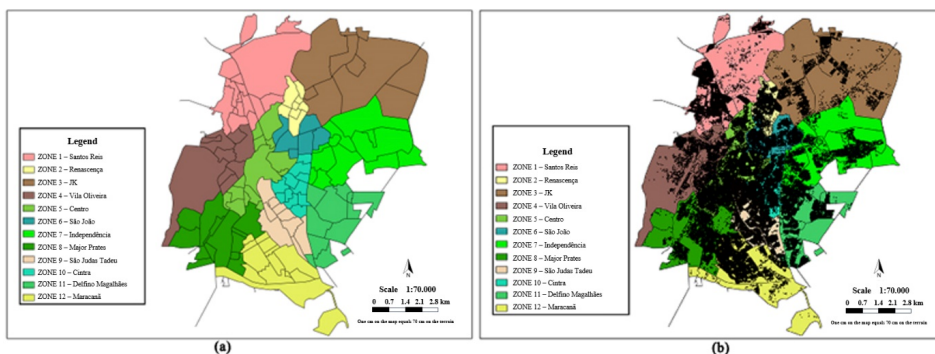

Figure 5. (a) Thematization of the map (b) insertion of points for the collection network. Source: Elaborated by the author.

This thematization allowed the creation of zones, helping to define the number of facilities that should be modeled on the problem. It also facilitated the manipulation of the database that was used to create the collection network, as it allowed the isolation of the points contained in a given zone to create a cost matrix that lists only those points.

Figure 5b shows the insertion of 87,748 points, which are represented by small black dots in each zone of the map. These points are referenced by pairs of geographic coordinates and represent potential OGR generating points. As noted by Zhao et al. (2013) in the study about the density of networks in location problems, it can be considered that the network modeled in this paper has a high density. Thus, the proposal of zoning through thematization allowed circumventing this problem and to reduce the total unity of the network, due to its partitioning by zones.

Besides, as noted by Han et al. (2013) and Zhao et al. (2015), modeling locations with a dense network harms the computational effort in addition to hampering the search for

optimal solutions. With the creation of a cost matrix from a dense network in TransCAD, there was a considerable increase in the consumption of computational space that was verified in the order of 140GB and also an increase in time in the execution of the algorithms applied in the integration of the problem to the GIS, making it impossible to solve the problem.

It is important to mention that despite the partitioning performed and the creation of several cost matrices, the problem was modeled from the original network, without affecting its characteristics of the distance between points. This network generated in the SIG TransCAD is called Network and contains all the information necessary to create the matrix of minimum paths between the points contained in the map. It was possible through this partitioning, the modeling of smaller matrices in each zone, enabling the application of the p-median model in a targeted way in each zone, which would not be possible if only a cost matrix was used for the dense network.

As described in the methodological procedures, there was a need for an algorithm for the treatment of cost matrices generated by TransCAD. The nature of these matrices is a multigraph, that is, a graph with parallel edges and loops at all vertices. The developed algorithm reads the graph generated by TransCAD, identifies loops and parallel edges, and removes those loops and parallel edges of the graph in each zone. The initial total number of edges for each cost matrix is $E = N^2$, where N is the number of points or vertices for a given zone on the map. The algorithm generates a file where the number of edges becomes equal to:

$$E = \frac{N^2}{2} - \frac{N}{2} \qquad (2)$$

Table 1 shows the sizes occupied in memory for each cost matrix for each zone generated by TransCAD and the reduction of this file after the application of the developed algorithm. The values found are following the studies of Zhao et al. (2015), as it reduces the need for storage and, consequently, reduces the time needed to solve the problem of the location of facilities. It was observed during the matrix generation process that while the GIS needed several days to generate the cost matrix of the dense network, in a few minutes it was possible to generate cost matrices for the partitioned networks.

Table 1. Results of the application of the algorithm for simple graph generation.

Zone	Cost matrix size generated by TransCAD	Size of the cost matrix after applying the Algorithm	Reduction %
1	1.59 GB	0.3564 GB	78
2	225 MB	51.1 MB	77
3	540 MB	121 MB	78
4	429 MB	110 MB	74
5	2.17 GB	0.487 GB	78
6	489 MB	110 MB	78
7	753 MB	169 MB	78
8	923 MB	206 MB	78
9	1.21 GB	0.271GB	78
10	832 MB	186 MB	78
11	960 MB	215 MB	78
12	1.45 GB	0.326 GB	78

Source: Elaborated by the Author.

It can be seen in Table 1 that the algorithm developed in the Python language allowed an improvement in the processing of data on the GIS platform, both about the storage capacity and the processing time of the p-median solver. This is important since large graph problems (dense networks) consume a lot of time to perform the optimization calculations until obtaining a satisfactory solution. It is possible to observe that the algorithm made it possible to significantly reduce the file to be manipulated in most zones in the order of 78% comparing to the original file.

5.1 TransCAD integration application

To integrate the p-median model with the GIS TransCAD, it was necessary to develop an application in the native GISDK language of TransCAD. This application, whose interface can be verified in Figure 6, allows automating all data generation and program execution to the model and allows the user to choose in which zone the p-median model will be applied, as well as the number of facilities to be located. Another objective of this application is the creation of a set of medians in the TransCAD called selection set, in which the indexes of these facilities are part of a set selected within the GIS.
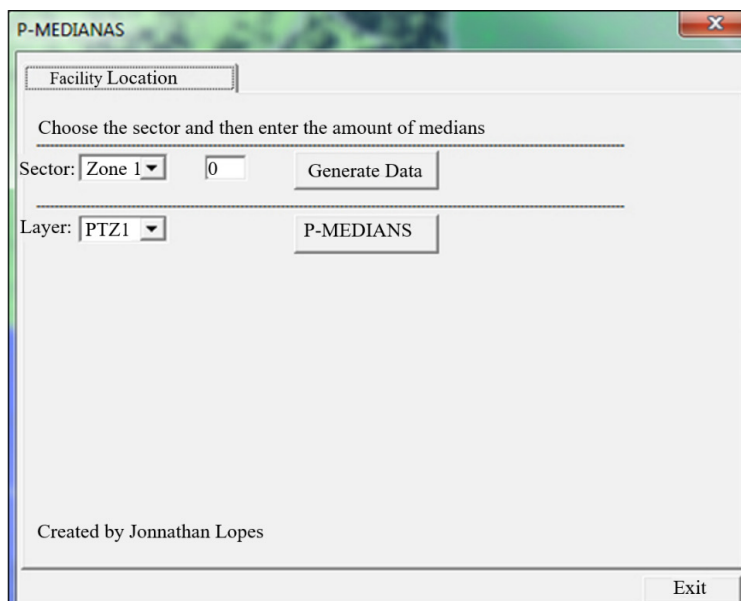

Figure 6. Integration application p-medians and TransCAD. Source: Elaborated by the author.

Figure 6 shows that the application has a box called sector for selecting the zone, a p field to inform how many points among the set of points inserted in the selected zone will be considered facilities. After filling in these fields, the user must click on generated data button, whose function is to automate the network creating process creation process, apply the creation of a simple graph algorithm and create the input files for the p-median problem solver. Besides, the user must select in which sets of points the p-median method should be applied and then click on the p-median button to perform the integration with the GIS.

The application internally runs the p-median.exe program (Senne & Lorena, 2000), which uses the Lagrangean / Surrogate relaxation technique to the p-median model, and then find the solutions and corrects the node and facility indications found in the solutions, as these differ from the indexes of the georeferenced points in TransCAD. If these indexes were not corrected, an incorrect reference would occur at the time of integration. The application reads the solution found by the program and checks which facilities are found. While doing this search, find the coordinates (latitude, longitude) of the medians on the map and connect each facility to its related node.

Figure 7 below shows the p-median model in the search for solutions in zone 4. The yellow dot is the location where the facility should be installed. The application was executed to search for *p=1* solution and search for *p=3*, where p is the number of facilities to be found by the integration.
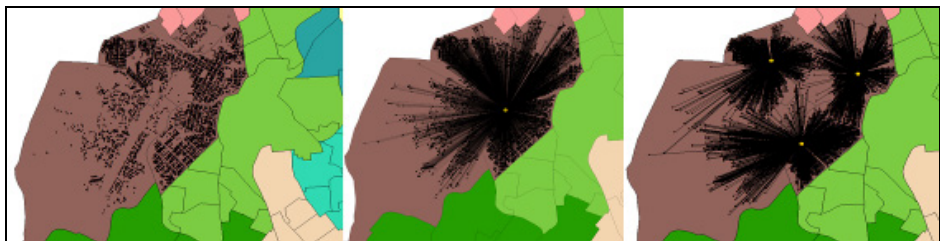

Figure 7. Application of the model in zone 4 of the Montes Claros city map. Source: Elaborated by the author.

Figure 7 shows that facilities are differentiated from other points through colored circles. Each related node is linked to its facility through lines that represent the allocation of these points in their respective facilities. Also, it is possible to note that the increase in the number of facilities allows for an improvement in the coverage of the ready-to-use areas that are on the peripheries of the zones, as occurs at *p=3*. This observation is important because in a waste collection system, the more homes that adhere to the disposal, the greater the collection. Besides, the total cost of the problem becomes even lower, reducing the total distance to be traveled by collectors in door-to-door collections.

After applying the p-median facility location model in each area of the Montes Claros city map, it was possible to find 12 optimal locations for installing OGR storage points. These points can be seen in Figure 8 below by black dots and are identified and shown in Table 2 below, in each zone the facility ID(index) and its respective coordinates (longitude, latitude) were identified.
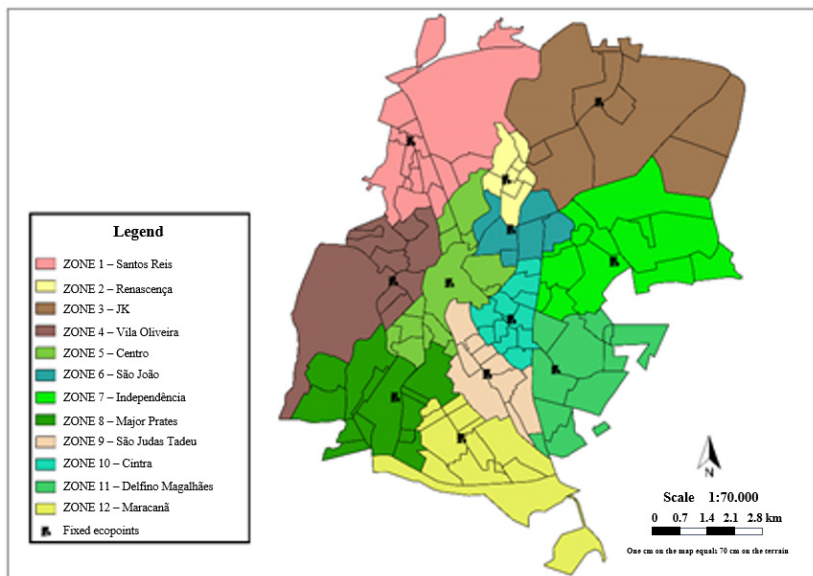
Figure 8. P-median modeling for facility location. Source: Elaborated by the author.

The number of facilities shown in Figure 8 for the installation of eco points throughout the city was defined according to the number of zones on the map, but the total cost (km) of the system tends to decrease as the number of facilities increases. This increase in facilities must be observed when planning a collection system, as it causes an increase in financial costs.

What is suggested is to combine a location strategy based on optimization to aid the decision-making process, which must be accompanied by an analysis of the need for the number of eco points in each zone following some specific criterion, such as the number of houses and waste generation capacity.

Table 2. Solutions found in the 12 zones with parameter p = 1 in each zone.

Zone	Network density	Facility ID(Index)	Coordinate (Longitude, Latitude)
1	9788	6806	-43.874911, -16.692506
2	3663	853	-43.852342,-16.701067
3	5652	2654	-43.830550,-16.683342
4	5372	2225	-43.878847,-16.724301
5	11438	5675	-43.865585,-16.724616
6	5379	2100	-43.851162,-16.712375
7	6663	3430	-43.826803,-16.719565
8	7364	2834	-43.878213,-16.750673
9	8546	1935	-43.856524,-16.745299
10	7001	3916	-43.850898,-16.732011
11	7520	3719	-43.840408,-16.744177
12	9362	5369	-43.862523,-16.759921

Source: Elaborated by the author.

Table 2 shows the result of the output created by the developed application, indicating the number of residential points (network density) that are allocated in each eco point modeled by the p-median problem in the GIS. It is possible to observe that for the 12 zones the density of the network created does not exceed the values addressed in the works of Han et al. (2013) and Zhao et al. (2015), where the loss of gain in the solution begins to be observed location problem. This proves that the proposed model minimizes the density effect of the collection network.

In addition, it is observed that zone 5, referring to the central region of the city, has a very large allocation of points in a single eco point. In this case, it is recommended studying the application of more than one eco point to minimize the dispersion effects of the points. In addition, it should be noted that the model is purely mathematical and issues such as the strategic location of eco points must be observed.

According to Nunes et al. (2016), eco points are places for waste disposal with the aim of sensitizing the population and promoting environmental education in addition to solving the problem of the correct destination of waste generated in society. Furthermore, when eco points are modern and well-planned, they allow a change in society's perception of waste collection as well as the structure that surrounds it, such as the collectors and the activity they carry out. In addition to creating all this change in people's mentality in relation to recycling and the correct disposal of waste, eco points bring the view that society can deal in a planned way with the waste generated.

This paper only aims to assist decision-making in the search for optimal locations through a purely mathematical model and issues such as the strategic location of eco points must be observed. Therefore, aspects of the feasibility of the installation in relation to infrastructure costs are not part of the proposed approach. It is recommended using the coordinates found as a starting point for the definition of eco points, looking for easy access places with a large flow of people such as schools, for example.

6 Conclusion

As discussed in the studies by Castellanelli (2008) and Guabiroba & D'Agosto (2011), about the potential use of residual oil in the production of biodiesel, the importance of this work in the study region, which has a Biodiesel plant from Petrobras with a production capacity of approximately 152.9 million liters of biodiesel per year and does not have an adequate collection system and is often discharged into the sewage system.

This paper presented a model that helps the planning of creating eco points that facilitate the supply of raw material for obtaining biofuel through reuse of OGR. The development of a systematic collection policy in the city can contribute to the social, environmental, and economic development of the region.

Thus, this paper seeks to contribute to the social and environmental sphere, as it is part of the solution to the problem of the correct destination of OGR in the region and what can create a source of income for waste pickers. It is important to note that the complementation of this research is the creation of a logistics model with the routing of vehicles in the modeled OGR collection network, seeking optimized routes between eco points and biodiesel factories, soap, and other products that can be produced from OGR as approached by Gândara (2015). The use of eco points allows the population to be sensitized and promotes education, as discussed by Nunes et al. (2016).

It is concluded in this paper that the integration of the p-median model and the geographic information system in the proposed modeling through zones, allows with low computational cost, to find points for installing facilities in the dense network. This

integration was also shown to be efficient in the researches of Carnasciali (2007), Arakaki & Lorena (2006), Khan & Samadder (2014) which proves that the use of GIS is fundamental in problems of locating facilities in various contexts.

In this research, it was observed that the integration of the GIS to the p-median model is not enough to locate installations. Based on the observation made in the studies by Zhao et al. (2015) Han et al. (2013), Zhao et al. (2015), it is necessary to study the density problem of the modeled network, since networks with densities above 12,000 points can present solutions without optimization gains or with insignificant gains, besides bringing great computational consumption making it impossible to search for solutions to the problem. This problem was overcome by the zoning and partitioning of networks where it can be seen in Table 2 networks with densities below 12,000 points

In addition to contributing to the computational effort and the search for an optimal solution, the map's zoning model can allow the planning and allocation of resources in a targeted or priority manner. Public agencies or the company responsible for collection may create prioritization standards for the installation of eco points in the zones, according to restrictions raised in each of them, such as the creation of eco points gradually in areas with a greater number of collection points, as is the case of zone 5 presented in table 2 with 11438 points, or in areas that have a greater capacity for generating waste, which requires a quantitative study of the OGR generation capacity.

The developed application allowed automating the entire modeling process and it is a tool that can be used in the planning sectors, making it important in helping decision making about location decisions. The integration of the visualization capacity, spatial analysis of the GIS, and the localization model combine in a powerful tool to support the spatial decision.

References

Arakaki, R. G. I., & Lorena, L. A. N. (2006). Uma heurística de localização-alocação (HLA) para problemas de localização de facilidades. *Production*, 16(2), 319-328. http://dx.doi.org/10.1590/S0103-65132006000200011.

Biju, B. P. (2015). *Utilização do sistema de informação geográfica (SIG) na indicação de possíveis áreas aptas à disposição de resíduos de construção e de demolição* (Dissertação de mestrado). Universidade Tecnológica Federal do Paraná, Curtiba.

Brasileiro, L. A., & Lacerda, M. G. (2008). Análise do uso de SIG no roteamento dos veículos de coleta de resíduos sólidos domiciliares. *Engenharia Sanitária e Ambiental*, 13(4), 356-360. http://dx.doi.org/10.1590/S1413-41522008000400002.

Brito, R. A. F. (2006). *Uso de sistema de informação geográfica para a análise do transporte e disposição final dos resíduos sólidos* (Dissertação de mestrado). Faculdade de Engenharia de Ilha Solteira, Ilha Solteira.

Burrough, P. A. (1986). *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford: Oxford University Press. http://dx.doi.org/10.1080/10106048609354060.

Carnasciali, A. M. S. (2007). *Integração de sistemas de informações geográficas e sistema especialista visando auxiliar a tomada de decisão locacional do setor bancário* (Dissertação de mestrado), Universidade Federal do Paraná, Curitiba.

Carnasciali, A. M. S., & Delazari, L.S., (2011). A localização geográfica como recurso organizacional: utilização de sistemas especialistas para subsidiar a tomada de decisão locacional do setor bancário. *Revista de Administração Contemporânea*, 15(1), 103-125. https://doi.org/10.1590/S1415-65552011000100007.

Castellanelli, C. (2008). *Estudo da viabilidade de produção do biodiesel, obtido através do óleo de fritura usado, na cidade de Santa Maria - RS* (Dissertação de mestrado). Universidade Federal de Santa Maria, Santa Maria.

Companhia de Saneamento Básico do Estado de São Paulo – SABESP (2017). *Programa de reciclagem de óleo de fritura da SABESP*. Retrieved in 2017, March 14, from http://site.sabesp.com.br/uploads/file/asabesp_doctos/programa_reciclagem_oleo_completo.pdf

Daskin, M. (2013). *Network and discrete location*. Hoboken: Wiley.

Fitz, P. (2008). *Geoprocessamento sem complicação*. São Paulo: Oficina de Textos.

Gândara, G. M. F. (2015). *Produção e descarte de óleo residual de fritura na cidade de São Paulo/SP: proposta para uma gestão integrada com auxílio das geotecnologias* (Tese de doutorado). Universidade de São Paulo, São Paulo.

Goldbarg, M. C., & Goldbarg, E. (2012). *Grafos: conceitos, algoritmos e aplicações*. Rio de Janeiro: Elsevier - Campus.

Guabiroba, R., & D'Agosto, M. (2011). O impacto do custo de coleta do óleo residual de fritura disperso em áreas urbanas no custo total de produção de biodiesel– estudo de caso. *Transportes*, 19(1), 68-76. http://dx.doi.org/10.14295/transportes.v19i1.372.

Han, M., Håkansson, J., & Rebreyend, P. (2013). *How do different densities in a network affect the optimal location of service centers?* (Working Paper for Transportation Research, pp. 1-17). Falun, Suécia: Dalarna University. Retrieved in 2018, November 20, from http://www.diva-portal.org/smash/get/diva2:627956/FULLTEXT01.pdf

Instituto Brasileiro de Geografia e Estatística - IBGE (2011). *Censo Demográfico 2010: características da população e dos domicílios: resultados do universo*. Rio de Janeiro: IBGE. Retrieved in 2016, January 20, from https://sidra.ibge.gov.br/pesquisa/censo-demografico/demografico-2010/inicial.

Khan, D., & Samadder, S. R. (2014). Municipal solid waste management using Geographical Information System aided methods: a mini review. *Waste Management & Research*, 32(11), 1049-1062. http://dx.doi.org/10.1177/0734242X14554644. PMid:25352293.

Klose, A., & Drexl, A. (2005). Facility location models for distribution system design. *European Journal of Operational Research*, 162(1), 4-29. http://dx.doi.org/10.1016/j.ejor.2003.10.031.

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2005). *Geographic information systems science* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Mapa, S. M. S., & Lima, R. S. (2012). Uso combinado de sistemas de informações geográficas para transportes e programação linear inteira mista em problemas de localização de instalações. *Gestão & Produção*, 19(1), 119-136. http://dx.doi.org/10.1590/S0104-530X2012000100009.

Nielson, D. (2014). *Geographic Information Systems (GIS)*. New York: Nova Science Publishers Inc.

Nunes, F. A., Fo., Osório, N.B., & Macêdo, C. F. (2016). Projeto Ecoponto na Escola, uma experiência de Educação Ambiental intergeracional em escolas públicas de Palmas – TO. *REMEA – Revista Eletrônica do Mestrado em Educação Ambiental*, 237-256. https://doi.org/10.14295/remea.v0i0.5659.

Nyerges, T. L. (1990). Locational referencing and highway segmentation in a geographic information system. *ITE Journal*, 60(3), 27-31. Retrieved in 2017, January 20, from https://pdfs.semanticscholar.org/54f6/57019f4caa168cbb3ca7ac66bc7a685c0340.pdf

Oliveira, M. (2012). *Sistema de localização de facilidades: uma abordagem para mensuração de pontos de demanda e localização de facilidades* (Dissertação de mestrado). Universidade Federal de Goiás, Goiânia.

Pereira, M. (2005). *Um método branch-and-price para problemas de localização de p-medianas* (Tese de doutorado). Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

Peuquet, D., & Marble, D. (1990). *Introductory readings in geographic information systems*. London: Taylor & Francis.

Pina, M. F., & Santos, S. M. (2000). *Conceitos básicos de sistemas de informação geográfica e cartografia aplicados à saúde.* Brasília: OPAS.

Pitta, O., Jr., Nogueira, M., No., Sacomano, J., & Lima, J. (2009). Reciclagem do óleo de cozinha usado: uma contribuição para aumentar a produtividade do processo. In *Proceedings of the 2nd International Workshop Advances in Cleaner Production* (pp. 1-10). São Paulo: ACPN. Retrieved in 2017, March 13, from http://www.advancesincleanerproduction.net/second/files/sessoes/4b/2/m.%20s.%20nogueira%20-%20resumo%20exp.pdf

ReVelle, C., & Swain, R. (1970). Central facilities location. *Geographical Analysis*, 2(1), 30-42. http://dx.doi.org/10.1111/j.1538-4632.1970.tb00142.x.

Senne, E., & Lorena, L. (2000). Lagrangean/surrogate heuristics for p-median problems. In M. Laguna & J. González Velarde. *Computing tools for modeling, optimization and simulation: interfaces in computer science and operations research* (pp. 115-130). Boston: Kluwer Academic.

Sule, D. (2001). *Logistics of facility location and allocation*. New York: Marcel Dekker.

Thill, J. (2000). *Geographic information systems in transportation research*. Bingley, UK: Emerald.

United States Department of Agriculture – USDA. (2016). *Oilseeds: world markets and trade*. USA: USDA. Retrieved in 2017, February 20, from https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf

Zhao, X., Carling, K., Dan, Z., & Håkansson, J. (2013). *Network density and the p-median solution* (Working Paper for Transportation Research, pp. 1-11). Retrieved in 2017, March 13, from https://www.researchgate.net/publication/267448802_Network_density_and_the_p-median_solution

Zhao, X., Rebreyend, P., & Håkansson, J. (2015). *Does road network density matter in optimally locating facilities?* (Working Papers in Transport, Tourism, Information Technology and Microdata Analysis, pp. 1-14). Retrieved in 2017, March 13, from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A847912&dswid=-5563