GIS-Based Planning of Rural Distribution Grid using MILP Approach

POLITECNICO

MILANO 1863

Nour Mohamed Esawe Ahmed

School of Industrial and Information Engineering

LM-28 - Electrical engineering

Politecnico di Milano

Acknowledgements

Foremost, I would like to express my gratitude to my advisor Professor Marco Merlo for the continuous support of my thesis and research, for his patience, guidance, and immense knowledge that allowed me to achieve the goals of my research.

Besides my advisor, I would like to pass my sincere thanks to my Co-advisor PhD - Student Silvia Corigliano who facilitated my research path, shared all her knowledge, and guided me all the time of my research.

I would also like to thank Enel GI&N for their cooperation and providing me the required information to complete my thesis study.

My sincere thanks also goes to the professors from the School of Industrial and Information Engineering of Politecnico di Milano for conveying the great knowledge that I have right now and that allowed me to finish this thesis.

I would like to pass my special thanks to my wife that stood behind me and provided me all the support and motivation at each single step in my career and in my life.

Last, but not least, I would like to thank my parents for giving birth to me at the first place and supporting me throughout my entire life.

Abstract

Grid extension in rural areas is a challenging problem due to the difficult terrain, and the long distance from existing grid. This makes the routing phase of any electrification planning tool to be the most critical phase. In order to enhance the competitiveness of the grid extension solution with other solutions (e.g. off-grid solutions), a robust, efficient and accurate optimization model is mandatory to obtain the optimum network topology with the minimum possible cost while satisfying the electrical constraints..

In this thesis, a MILP-Based approach is developed to effectively perform grid routing and search for optimal substation connection to feed the grid. Moreover, the model performs cable size categorization to follow the hierarchical structure of the distribution network and to minimize the cost of lines as possible. This model is designed to be integrated with GISEle which is a rural electrification planning tool. GISEle injects into the model the GIS data in the form of a discretized map that contains information about the location of loads, and terrain characteristics of the area under study.

Eventually, `pandapower`, which is an open-source power system analysis tool, is integrated with the model and is used to validate the electrical constraints by running an actual load flow on the output topology. The performance of the model is tested on a real case study based in Brazil and is compared with the heuristic approach that GISEle is currently using in the grid routing phase.

The results showed that, overall, MILP-Based approach significantly enhances the economic efficiency of the output topology compared with the heuristic approach and effectively allocates the size of cables following Main & Collateral approach.

Moreover, the flexibility of the model allows it to be initialized by heuristic algorithms to reduce the computational effort.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Despite the global increase of population with access to electricity rising from 83% in 2010 to 89% in 2017, there still remains an amount of 840 million people without electricity mainly in rural areas (International Energy Agency, 2019). This problem is further complicated because of the lack of a global definition of access to energy since a village can be considered electrified if it contains atleast 10% of electrified households (Livemint, 2018). As a result, this could cause ambiguity and an incorrect perception of the problem. By considering these results and the expected growth of global energy consumption to grow by 50% between 2018 and 2050 (U.S.-Energy-Information, 2019), it becomes imperative to search and develop new electrification strategies to cope with the energy growth and guarantee the access to electricity for all population around the world.

1.1 Rural Electrification Challenge

Rural areas are getting particular attention in power system planning for several reasons. First, these areas, especially in developing countries, are being at a huge disadvantage when it comes to access to electricity compared to urban areas. This is due to the characteristics of these areas including: low-population, remotely located with difficult terrain, and demand characterized by a low consumption profile. As a result, it becomes extremely costly to implement electricity schemes to connect these areas to the network. Additionally, the poor income in rural areas significantly impacts the affordability of any electrification solution. Moreover, the access to

electricity will promote production and enhance education and healthcare in rural areas.

By considering these aspects, any improvement in the economic efficiency of the electrification solutions is of paramount importance. This calls out for the development of robust, economically efficient, and accurate electrification planning tool to facilitate the planning process and to find the optimal electrification solution that increases the affordability in rural areas.

The goal of this thesis is to develop a MILP-based model to effectively plan the network topology with the objective to minimize the cost while respecting the electrical constraints including voltage limits, thermal limits, and energy balance of the system. The optimization model will be integrated with the GIS tools to make use of the geographical data such as the location of the populated areas, and the nature of the terrain of the area being electrified. In addition to that, the output topology will based on the typical hierarchical structure of an electrical network such that the cables are categorized to: Main feeder (carrying high power), and Collateral lines (carrying low power) with correspondece to the amount of load that they feed. Moreover, supplementary tools are proposed to enhance the operation of the MILP model including `pandapower` as an open-source power system analysis tool to validate the results by running an actual load flow. Eventually, the MILP model will be tested on a real study case to evaluate the performance of the approach.

1.2   Thesis Outline

The work of this thesis is divided into a number of chapters. The content of each chapter is summarized as the following:

- In Chapter 2, a review of the available tools and approaches (heuristic and numerical) in the field of electrification planning is presented while stating the pros and cons of each of them.

- In Chapter 3, GISEle project is introduced while stating the building-block

that will be improved by the work depicted in this thesis. Moreover, the result of the case studies that have been conducted is also represented.

- In Chapter 4, the proposed MILP-Based model is introduced along with the limitations of the model and the proposed suppelementary tools to overcome those limitations.

- In Chapter 5, the real case study of Cavalcante is presented in which the proposed model will be tested.

- In Chapter 6, the results of applying the model on the study case are presented and is compared with GISEle's current approach to point out improvements.

Chapter 2

Literature Review

In this chapter, the recent work done in the field of power system planning will be introduced while comparing between two types of approaches: Heuristic, and Numerical approach. Moreover, GIS tool will be introduced since the work of this thesis is based on the data obtained from it.

2.1 Algorithms & Models for Distribution Network Planning

Several algorithms and models are based on GIS tools to allow effective planning of the electrification system by means of giving the opportunity to perform geospatial analysis. Therefore, it is important to introduce it to allow proper understanding of the algorithms and models presented.

2.1.1 Geographical Information System (GIS)

It is difficult to define the term of GIS since it sums up different subject areas. However, the most accepted definition is the one provided by the National Centre of Geographic Information and Analysis (NCGIA):

> *A GIS is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modelling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources (NCGIA, 1990)*

The main advantage of GIS is that it integrates different types of data according to their geographical coordinates. These data are mapped on the Earth's surface to facilitate data analysis, record-keeping, and improve decision making.

Map Projection

Map projection is concerned with the representation of the earth's surface onto the flat surface of map. In order to preserve the consistency of the geographical data, they must be mapped using a common Coordinate Reference System (CRS). The role of CRS is to define how this two-dimensional projection takes place with the help of coordinates. The choice of the map projection and the CRS depends on the location of the area that is being studied. There are basically three families of map projections as shown in Figure 2.1: (a) cylindrical projection, (b) conical projection, (c) planar projection. However, every map projection has specific distortions in angular conformity, distance and area, and is hence, prone to errors. In fact, the choice of the map projection is based on a trade-off to preserve only the characteristics that are serving the analysis being conducted. For instance, the mercator projection shown in Figure 2.2 is used when angular relationships are important because it perfectly preserve the angular proportions.

CRS can be divided to Cartesian Coordinate Reference System and Geographic Coordinate Reference System (GCS). The latter system is the one that is very common. It uses latitude (with reference to the equator line) and longitude (with reference to the Greenwich Meridian) to describe the position of an object on the earth's surface. The most popular one is called WGS84. In Figure 2.3, shows the pictorial view of GCS. The lines of latitude are parallel to the equator line and the lines of longitude with the prime meridian through Greenwich which are measured in degrees unit.

Geo-Processing Data Types

There are two data types that are mainly incorporated into GIS tools: *Vector Type* and *Raster Type.*

Vector data are mainly used to present data that are discrete. It represent real
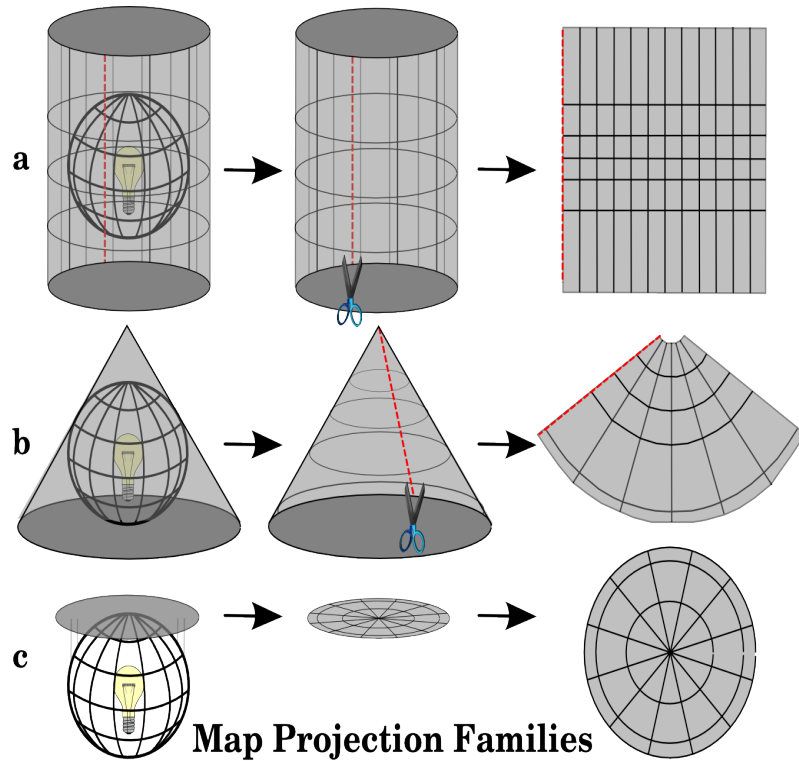
Figure 2.1: Families of Map Projections
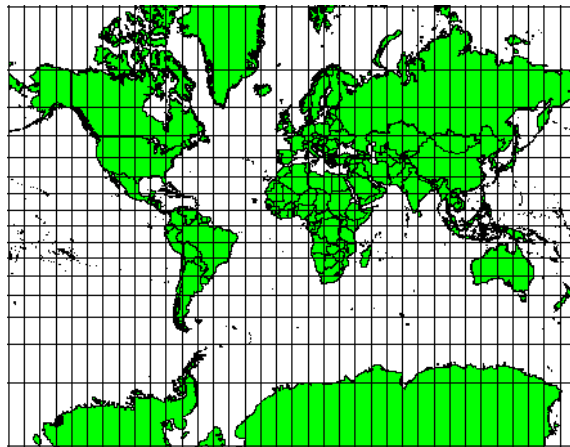
Figure 2.2: Mercator Map Projection

world features within GIS environment and those features contain attributes, text or numerical information to describe them. Moreover, each feature is represented by a geometry (point, polygons, polylines) which is made up of one or more interconnected
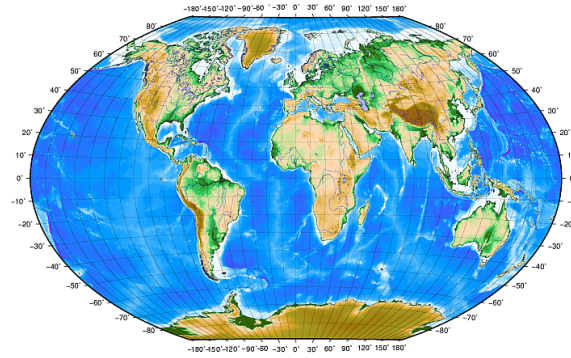
Figure 2.3: Pictorial view of GCS

vertices. The position of the vertex in the space is located using X, and Y axis. Vector data can for example be used to represent trees, roads, and building footprints

On the other hand, *Raster* data are composed by matrices of pixels called cells. Each pixel or cell contain a value that represent the nature of the area covered by that cell. Raster data are used to represent data that are changing continuously in space. For instance, the variation of the color and density of cover for grasslands is best represented by raster data. If vector data type is used, much information would be lost in the process of simplifying the features to a single polygon.

Each homogeneous collection of GIS data composes a layer. It can be composed by either a raster or vector data. In fact, a GIS map is a collection of raster and vector layers as shown in Figure 2.4.

Terrain Modeling

The main problem within this thesis is related to finding the shortest path between a given set of target points and a source node. In order to model the terrain where the path will be going through, there are two approaches: vector-based approach, and raster-based approach. It is important to examine the key difference between the two to allow proper evaluation of the routing performance.

Raster-based Approach

In this approach, a weight is allocated to each cell that represents the spatial nature of that cell. The shortest path will then be the path that takes the minimum weighted
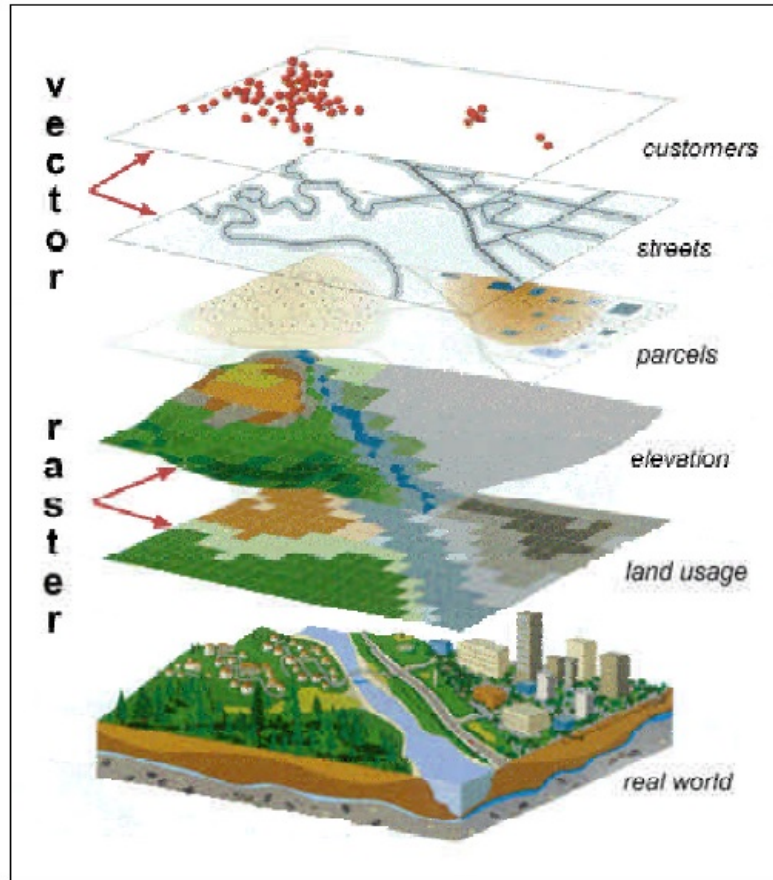
Figure 2.4: GIS map composition of raster and vector data.

cells to reach the destination. This approach is characterized by simpler and faster computation time which makes it a perfect choice to cover big areas. However, this does not come at no cost which is mainly related to the geographical distortions created due to the rasterization process. There are two types of distortion that are usually created: *elongation* and *deviation* as described by (Hekken and Oosterom, 1993). The distortion is depicted in Figure 2.5. The theoretical distance between the source and the distance is shown in figure 2.5 (a). The path followed using raster cells exhibits the so-called stair effect, and this is where the elongation error arises. The error of the entire path is equivalent to the maximum deviation shown in figure 2.5 (c) with a total value of 41% greater than the real distance.

Vector-based Approach

In this approach, homogeneous terrain areas are represented by polygons with an
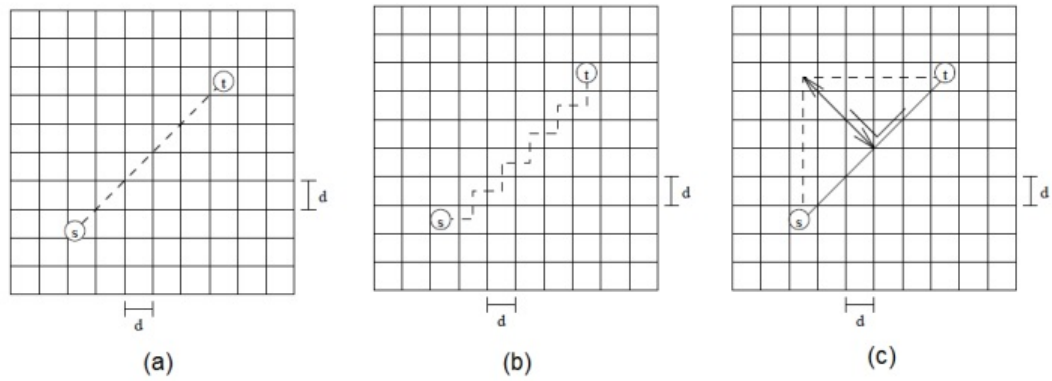
Figure 2.5: Path between two raster cells. (a) Real distance, (b) Stair Effect, (c) The maximum deviation

assigned weight. Mathematical procedures allow to compute the optimal incidence angles to find the least cost path. Therefore, it is possible to find the exact solution by tracing those points. However, this also does not come at no cost such that the complexity and computational burden on the algorithm is heavily increased.

The routing phase is the most important phase in electrification planning since the technical and economic feasibility of a grid extension solution depends on this phase. In order to facilitate the process of finding the optimal topology of the network, the electrical network is represented by a Graph which opens the opportunity of using Graph Theory. The terminal nodes (generator/loads) are represented by vertices and the lines are represented by the weighted edges. In this section, several routing algorithms and models will be presented to point out the key differences and the gaps of each of them. In order to properly convey the concepts that will be introduced, some important terminologies need to be clearly defined before going into the routing phase.

Definition 2.1.1 *Graph: A Graph G is composed of a set of finite Vertices V and a subset of Edges E connecting those vertices. The weight of each edge is evaluated based on the metric distance between the nodes which are connected by their corresponding edge.*

Definition 2.1.2 *Tree: A Tree T is any connected graph without the presence of*

any cycle such that there is always exactly one path from one vertex to another inside the Tree.

Definition 2.1.3 *Forest: A forest contains multiple Trees which are interconnected.*

Minimum Spanning Tree Problem

Initially, the routing problem is transformed to a Minimum Spanning Tree (MST) problem which is based on finding the optimal connection of **all** vertices that leads to the minimum total edge weight without any cycles. This optimization problem can be represented by the following objective function:

$$min \sum_{e \in T} w(e) \tag{2.1}$$

An example that showcases the MST problem is shown in Figure 2.6. It can be observed from the solution obtained by MST that it is incomplete and not accurate since it only gives the information about the connection between target nodes only in the form of straight lines. Thus, the output still misses the information regarding the optimal path to take to actually implement these connections. The optimal path is the path that takes into account all the spatial nature (terrain, obstacles, and etc..) underneath those straight lines obtained by the MST. This problem becomes even more significant in the case of rural areas since the probability of having forests, rivers and natural obstacles is high, and this implies that it would be economically infeasible to have a straight line connection within this type of environment especially for LV and MV networks. Therefore, this calls out for more sophisticated algorithms that overcome this limitation.

The algorithms that can overcome the limitation of MST are to be presented below. It is worthwhile mentioning some definitions that are required to properly understand the algorithms.

Definition 2.1.4 *Shortest Path: Let G = (V,E) be a connected weighted graph. The shortest path between two distinct vertices u and v of G, is the path that has a sum of weight less than or equal to any other path connecting u and v.*
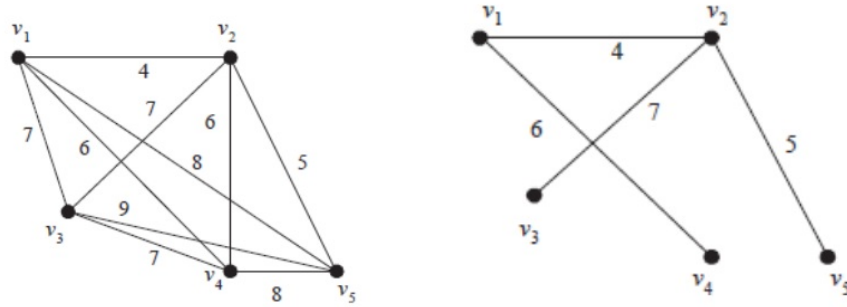
(a) Example of weighted, not ori-
ented graph

(b) MST connecting all the nodes of
the graph shown in a

Figure 2.6: Minimum Spanning Tree Example

Definition 2.1.5 *Distance: Let G = (V,E) be a weighted graph. The distance $d_{u,v}$ between two vertices u and v of G, is the sum of weights of edges of the shortest path from u to v. If no path exists between them, then $d_{u,v} = +\infty$. If u = v, then $d_{u,v}$.*

2.1.2 Heuristic Algorithms

Prim's Algorithm

This algorithm is based on a greedy approach. It was developed in 1930 and it is considered as one of the first solutions to the MST. The algorithm can be articulated in the following steps:

1. Starting by selecting an arbitrary point from the graph to initialize the Tree T = (V',E') where V' = v_0 and E' = ϕ

2. As long as V' ≠ V, perform the following:

   - Find all the possible edges e(u,v) that connect u ∈ V to v ∉ V'. Not to mention that e ⊂ E.

   - Select the minimum-weighted edge e and add it to E' and v to V'.

   - Break the loop is there is no possible edge.

3. T will be the minimum MST only in the case if V' = V. Ohterwise, no MST is found in the graph G.

The procedure of Prim's algorithm is articulated in Figure 2.7. Regarding the com-

putional time, it depends on the size of the graph in terms of the amount of vertices and the edges.
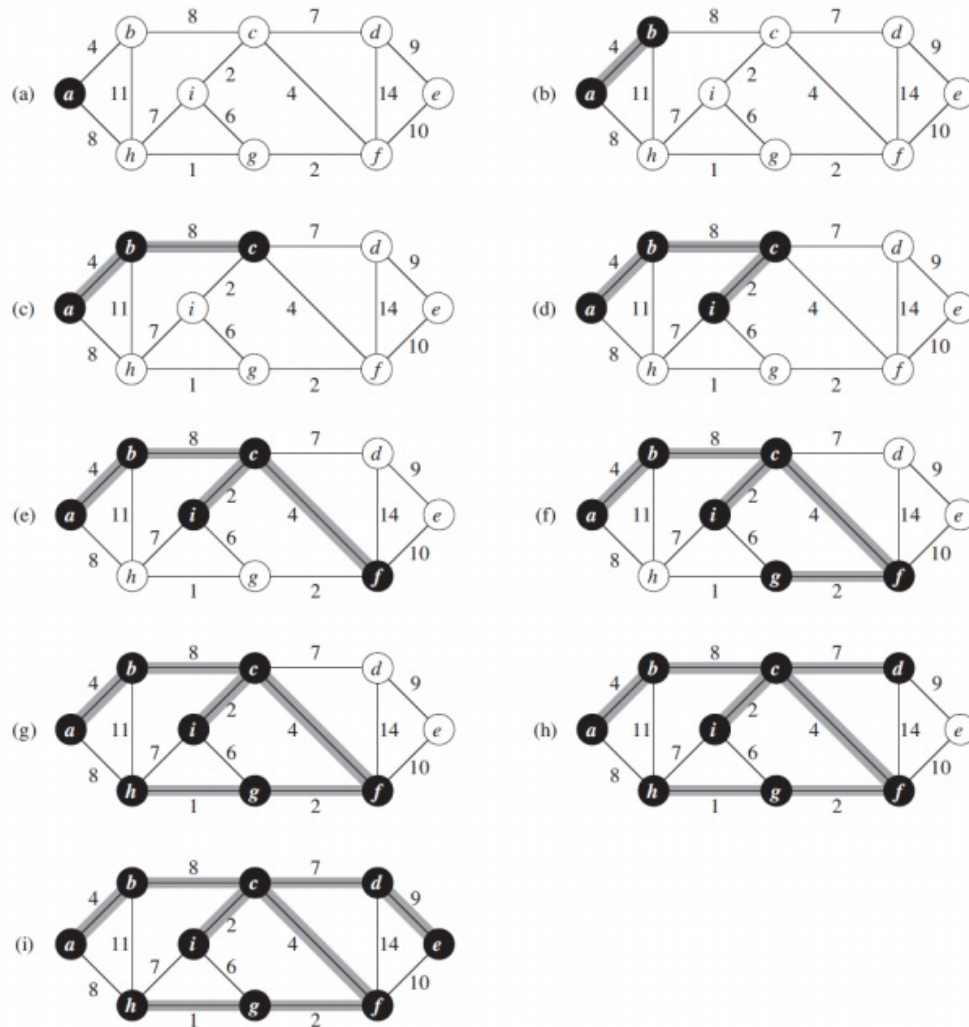
Figure 2.7: Procedure of Prim's Algorithm to find the MST of the graph

Kruskal's Algorithm Kruskal's algorithm is another greedy-based algorithm. It is considered a variation of Prim's Algorithm. At the initial step, the edges are sorted in an ascending order based on their respective weight. The algorithm starts removing one edge at a time from the set and all edges will be completely evaluated when the set is empty. This process can be demonstrated more thoroughly in Figure 2.8.

These greedy algorithms are considered heuristic algorithms due to the fact that the
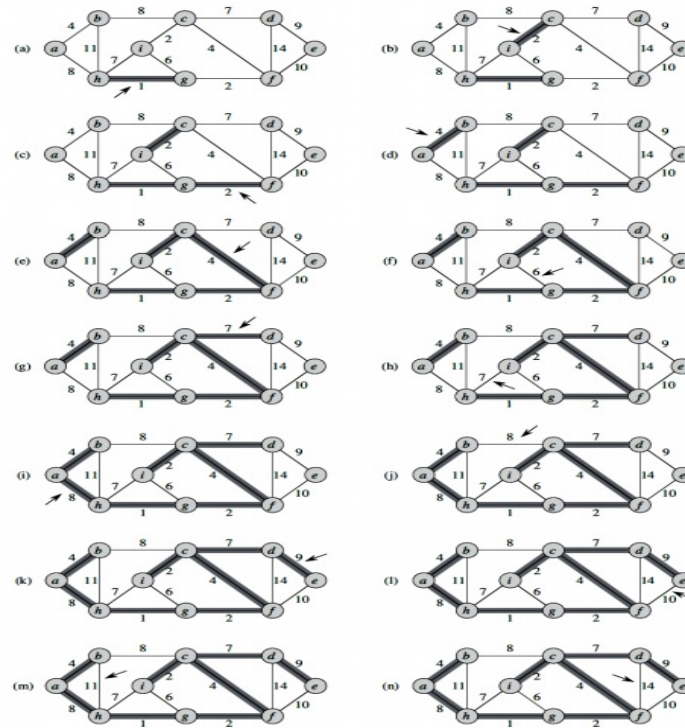
Figure 2.8: Procedure of Kruskal's Algorithm to find the MST of the graph

decision is taken in a step-wise fashion and if any incorrect decision is taken at the beginning of the process, this will result in a final solution which is far from being optimal. Moreover, the output of the greedy algorithm is MST oriented topology. This means that the nodes are connected only using straight lines without giving the possibility of having intermediate connections. Due to these drawbacks, they are not suitable to be applied for the case of electrification networks. However, they can be used to provide preliminary sub-optimal results due to their low computational time.

Dijkstra's Algorithm

This algorithm was named after its creator Edsger W. Dijkstra in 1956. The objective of the algorithm is to solve the problem of finding the shortest path between two given nodes of a graph. As an example, consider two nodes: s (source), and t (terminal/load), in order to find the shortest path using Dijkstra's algorithm, the procedure can be broken down into the following steps:

1. A minimum path tree T = (V',E') is created which initially contains only the source node.

2. The distance between the source node and all the other nodes are computed. Therefore, composing the so-called Distance Matrix. Initially, the distances D_{sv} are $+\infty$ except D_{ss} is equal to zero.

3. While a terminal t $\notin$ V': Find the minimum-cost edge e connecting u to v where u $\in$ V'and v $\notin$ V'.

4. The new minimum-cost edge e is then added to E' and v is added to V'.

The output of the algorithm is the shortest path between the source node and the target node. It does not require all the vertices to be connected. However, since the source node is connected to target node using the minimum path found, it does not necessarily compose a spanning tree of G. An example of Dijkstra'a Algorithm that showcases the steps to find the shortest path is shown in Figure 2.9.

The major drawback of Dijkstra's algorithm is that it connects one source node to one target node. However, if the problem involves several terminal nodes that exist in graph composed by a pool of several nodes, the problem becomes too complex using this algorithm.

Steiner Tree Problem

Moving into another type of problem in Graph Theory, which is the Steiner Tree problem. It can be ummarized as:

> Consider an undirected connected weighted graph G = (V,E) and a subset of vertices S ⊂ V called terminal nodes. Find a tree T=(V',E') where V' ⊂ S and the weight w(T) are minimized
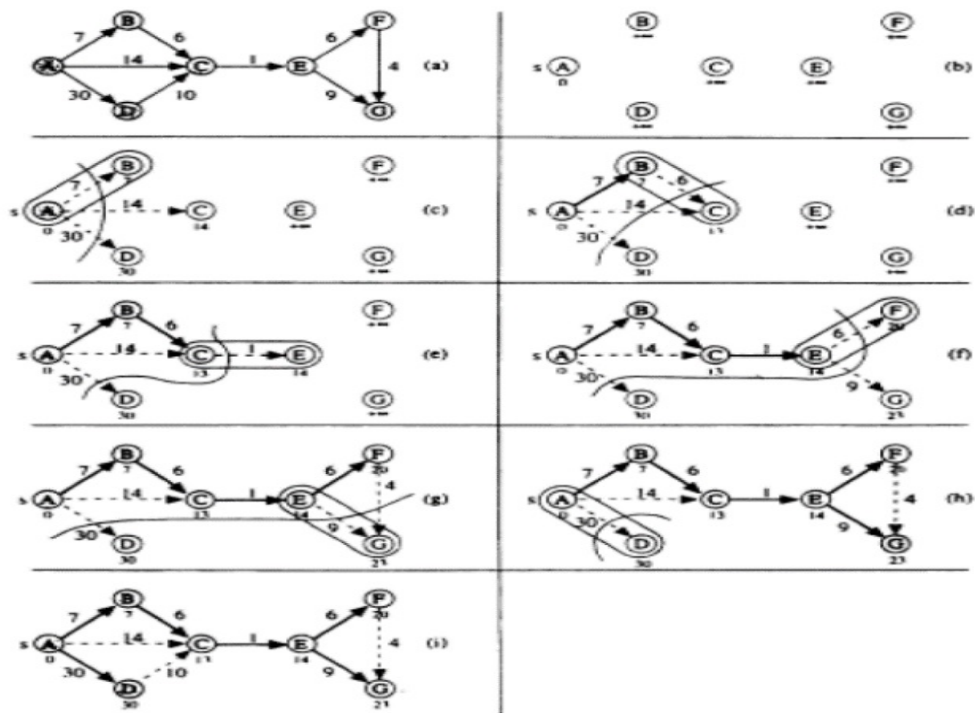
Figure 2.9: Example of Dijkstra Algorithm

In other words, the Steiner Tree problem is based on finding the optimal connection for the terminal nodes that exploits the minimum-weighted edges. Unlike MST problem, instead of spanning across all vertices, it spans across a given subset of vertices. The aim is the same as Dijkstra's algorithm, but here it involves multiple target nodes instead of one which proves to be more suitable in the case of routing of electrification network.

In Steiner Tree problem, the vertices can be divided to two groups: *terminals* and *non-terminal* nodes. The *terminal* nodes, which represent populated/feeding points, must be included in the solution. Meanwhile, the *non-terminal* nodes represent the selected territory where the line passes through to form the grid. The total cost will then be the sum of the weight of all the selected edges. An example of Steiner Tree solution is shown in Figure 2.10. The minimum path found for the connection of terminal nodes (a to f) exploiting the road network is the solution the Steiner Tree problem.
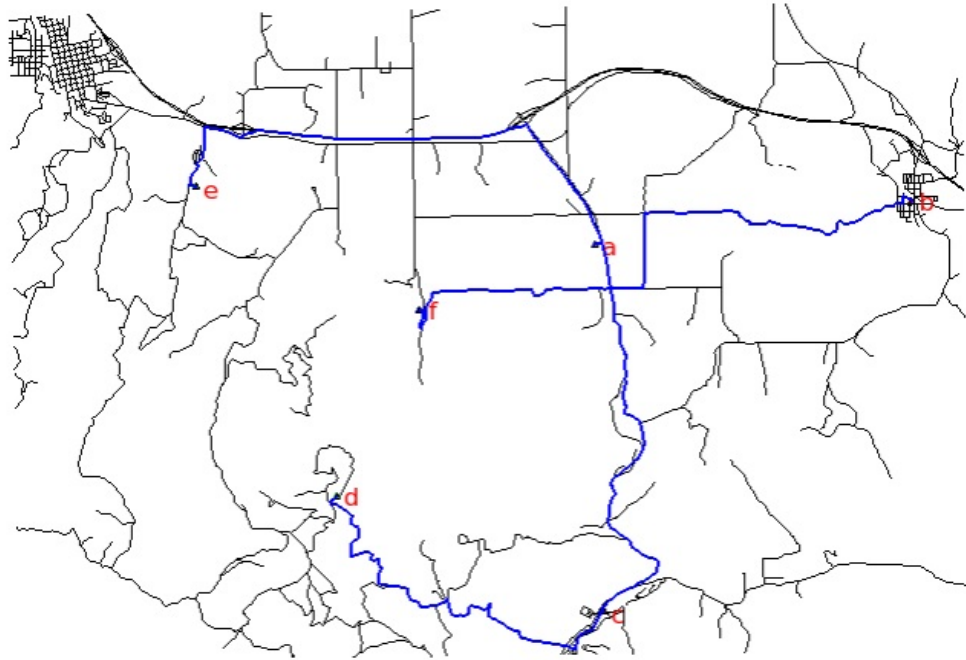
Figure 2.10: Example of Steiner Tree solution

Steiner Tree concept despite being a simple approach, it is part of a complex type of problems called NP-Complete. The issue is that it is not possible to determine how much time is required to reach the solution. This makes it difficult to be solvable by algorithms that take polynomial amount of time to reach a solution. In order to overcome the high computational time issue, the authors (Carnovali and Edme, 2019) have made use of MST and Dijkstra together to have an approximate solution of the Steiner Tree problem. The problem is broken down to small sub-problems, instead of having to process all the terminal nodes inside the cluster, each sub-problem will only be concerned about finding the connection between two terminals only. In this way, the computational effort is greatly reduced. The resulting grid is shown in Figure 2.11. The populated points (Yellow color) are connected using MV line (blue) following the non-terminal nodes (grid of points) to take into account the cost of the terrain. Moreover, in the grid connection phase, Dijkstra algorithm was found to be suitable in this case since it only needs to define the distance between one point inside the cluster and its nearest substation available. It performs this
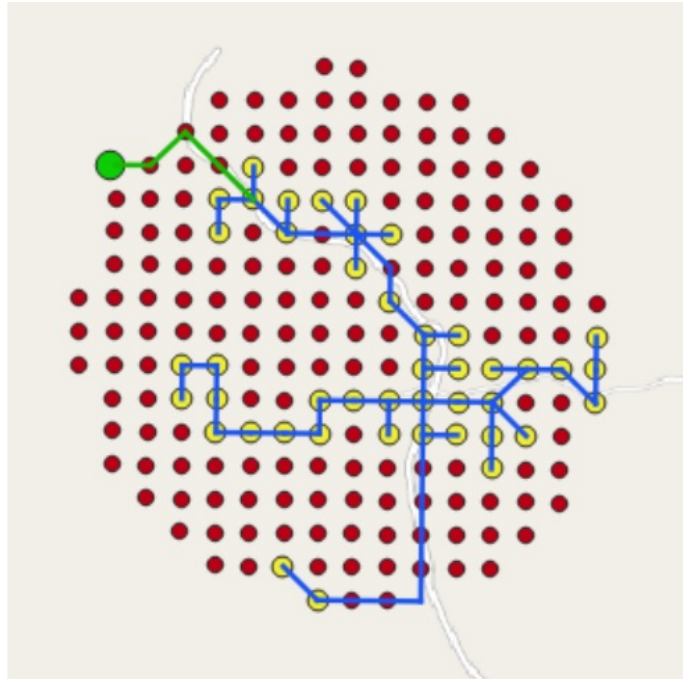
Figure 2.11: Example of the resulting grid using Steiner Tree approximated solution

operation for all the nodes inside the cluster forming the distance matrix, and the minimum cell is chosen among them.

When the previous algorithms and Steiner Tree concept are deployed in the field of electrification networks, a big gap is encountered which must be filled. The gap is in the output topology such that there are no information about whether that output topology is respecting the electrical constraints (voltage and thermal limits) of the lines or not. The optimization is only considering the spatial aspects without paying attention to the electrical constraints. In the case that the output topology violates the electrical constraints, the output will no longer be meaningful or useful due to infeasibility from an electrical perspective. This massively motivates the need for another more sophisticated approach that fill this gap.

Another gap that needs to be filled is related to the choice of the substation to feed the cluster. In the previous work, the choice of the substation is only based on the distance such that the nearest substation to the cluster is chosen (Carnovali and Edme, 2019). However, this is not enough to decide on the economic feasibility of the chosen substation. For instance, if the cluster is featured by a low demand, and

the nearest substation is coming from HV line, it would be economically infeasible to implement a HV/MV transformer just to feed a small amount of load. Hence, the choice of the substation must depend on the amount of power required by the cluster, and the distance to the cluster which is being fed. Moreover, the algorithm should provide the possibility of having more than one cluster sharing the same substation if it can accommodate the total amount of power required by the clusters interconnected.

An improvement has been developed to fill the gap related to the optimization of substation connection (Vinicius, 2020). The developed algorithm is articulated in Figure 2.12. The algorithm finds a point inside the cluster and a grid connection point. As was mentioned above, Dijkstra is preferred over Steiner tree to find the connection of this particular part since it is a *shortest path* problem. Dijkstra is featured by a much faster computational time than Steiner Tree. An example of the resulting grid output after importing this improvement is shown in Figure 2.13. For the big cluster (blue), the load is high and therefore, the algorithm chooses the far HV substation (Black square) instead of the closer MV substation (red square) since it is sufficient to supply the load. However, a flaw appears here, since the algorithm did not consider the possibility of having the green cluster to share the same substation with the blue cluster.

A further improvement was done to allow the possibility of permitting the clusters to share the same substation if the power limit allows it (Vinicius, 2020). The improved algorithm is shown in Figure 2.14.
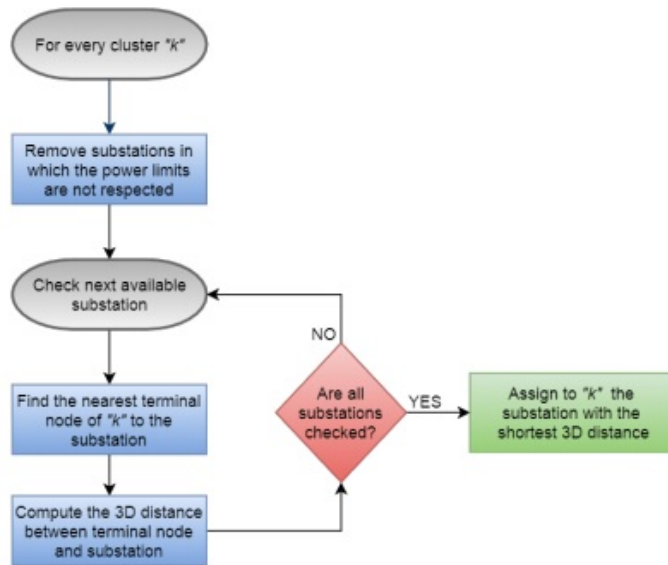
Figure 2.12: Optimization Algorithm of Substation Connection
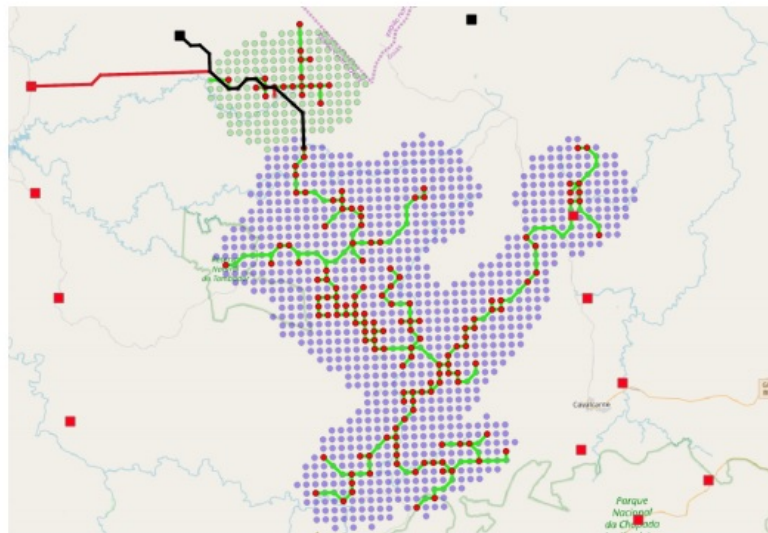
Figure 2.13: Resulting Grid with the Optimization of Grid Connection
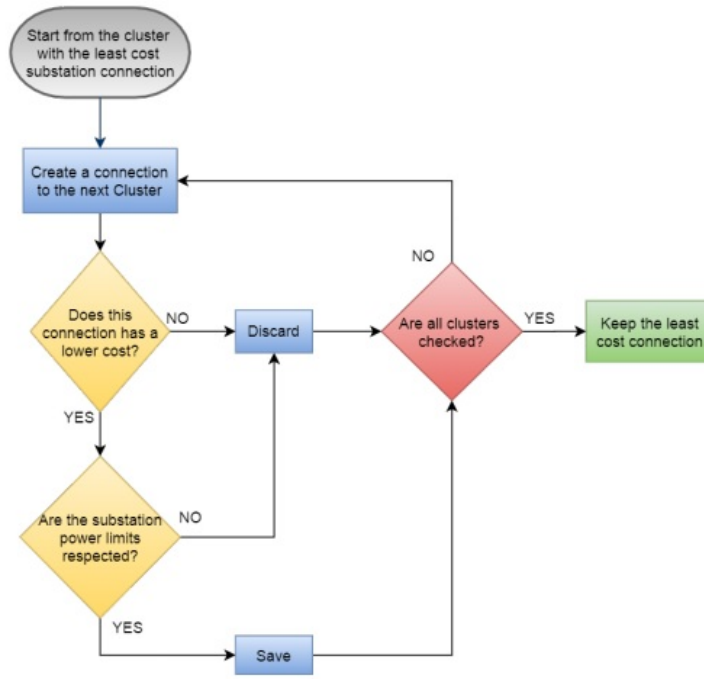
Figure 2.14: Improved Optimization Algorithm of Substation Connection to include cluster merging

2.1.3 Numerical Algorithms

The main advantage of numerical algorithms (e.g. MILP which is the focus of this thesis) over the heuristic approaches is the possibility of constraining the optimization to fit the environment of the problem. In case of power systems, it is possible to restrict the optimization of the topology to meet the power system stability requirements.

Generally, a Mixed-Integer Linear Programming (MILP) model consists of an objective function which is mainly to minimize the cost of the network, and a set of constraints that can be expressed as linear equalities or inequalities, and a set of decision variables. The decision variables can be exploited to decide on the optimal path to take. For instance, if the line is part of the network, the decision variable is 1, otherwise, 0. The equality constraint can be exploited to set the energy balance of the network, and the inequality constraint can be used to set the capacity limit of the lines.

In (Shu et al., 2012), A GIS-Raster-based MILP optimization model has been developed for spatial power system planning. The objective function is designed to minimize the line investment cost. The feasibility of the network routing is evaluated based on the candidate line routes and the hourly network evaluation for peak power loads to ensure the stability of the power system.

The input area is first rasterized into $N_R * N_C$ square cells as shown in figure 2.15 (a) where (b) is the rasterized enviromental map, and (c) is the altitude map. The darker the area, the higher is the altitude. The resolution is defined as W/N_R where N_R is number of cells in each row; N_C is the number of cells in each column; $C_{i,j}$ represent the cell in row i and column j.

(a)   (b)   (c)
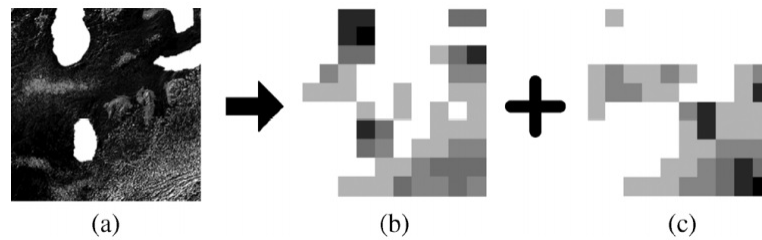
Figure 2.15: Rasterized map

The cost of the terrain is only related one spatial aspect which is the elevation. To follow the grid of points, the line can only reach its 8 neighboring nodes from 8 directions as shown in figure 2.16, where $S_{hr} = \sqrt{2}$ for odd directions, and 1 otherwise. Hence, the line length is equal to $\sqrt{S_{hr}^2 + S_{vt}^2}$ where S_{hr}^2 and S_{vt}^2 are the horizontal and vertial differences between the centers of the two cells, respectively.

(a)   (b)   (c)
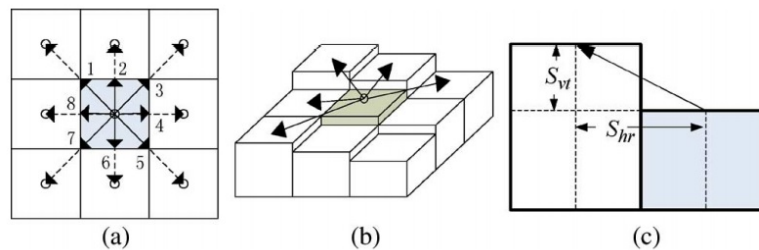
Figure 2.16: Routing principle based on grid of points

Where each direction of the 8 neighboring nodes has a different elevation as shown in figure 2.16 which is translated into the objective function.

The result of testing the model on a study case, while taking into account the cost of the terrain, is shown in figure 2.17.
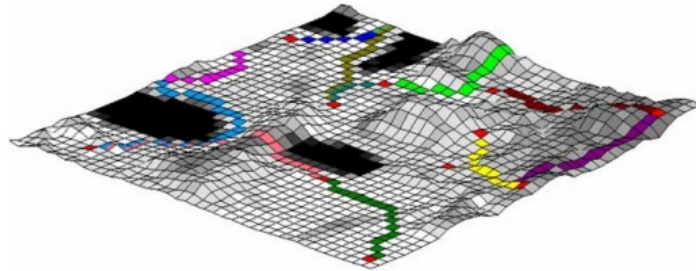
Figure 2.17: Three-dimensional optimal routing of the lines

Regarding the technical constraints, (Shu et al., 2012) relied on constraining the optimization using DC power flow. To recall, DC power flow results are based on the assumption that the voltage profile is flat (voltage is 1 p.u for all nodes). Therefore, in the context of distribution system, it is a critical drawback such that it is not possible to maintain the voltage within the acceptable limits. Moreover, the model does not check for the optimal substation to feed the network, and also does not check about the capability of the substation to feed the network.

Another numerical algorithm developed by (Samui, Samantaray and Panda, 2012) which is based on a Direct Solution. The algorithm searches for all possible radial paths with the root of each path being the substation. It also exploits graph theory to represent the network. It starts by creating a graph of all possible paths as shown in Figure 2.18. Then, the path with the minimum cost is the optimal solution, with voltage, capacity and radiality constraints. In order to reduce the number of possible paths, it applies the principle of optimality theorem which based on dynamic programming optimization technique. For instance, if node i is involved in the optimal path to node j, then there is no need to check the optimal path to node i. In this way, the computational load is heavily reduced.
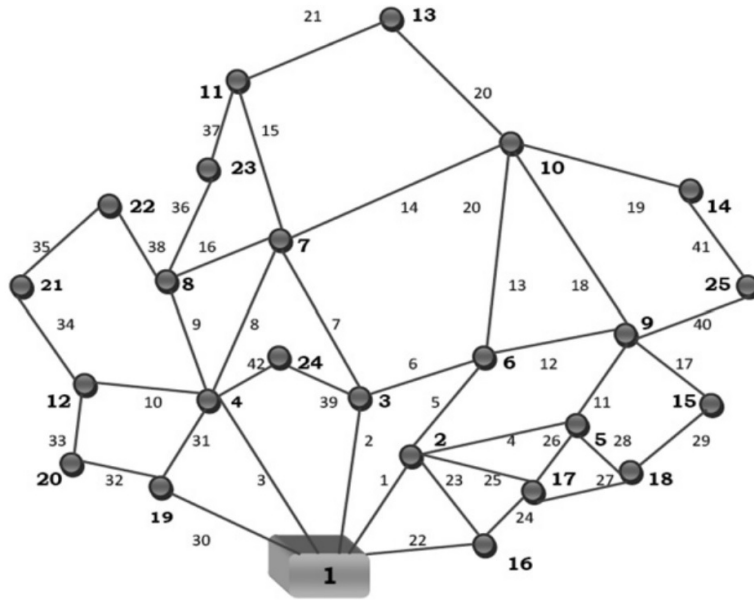
Figure 2.18: Graph of possible paths for 24 nodes

The optimal number of substations is evaluated on the basis of cost per kVA which includes the cost of the substation, cost of feeders, and the cost of energy loss. An example of the result is shown in Figure 2.19.
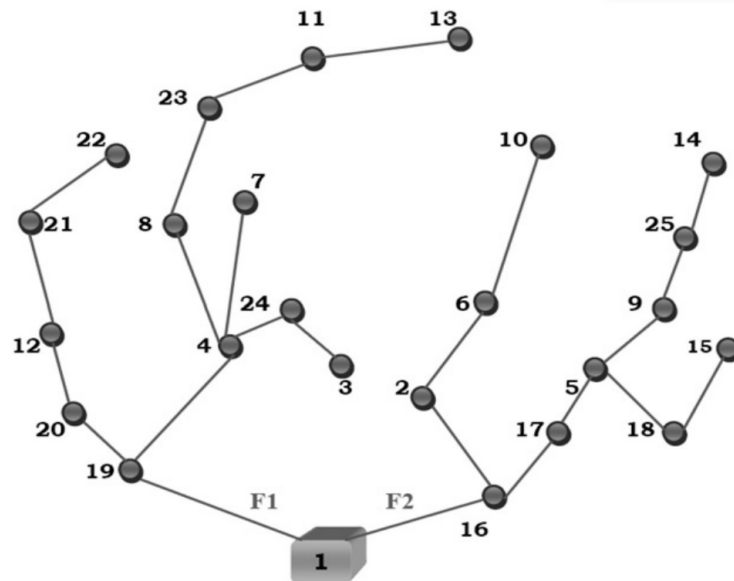
Figure 2.19: Optimal connection of the radial distribution network consisting of 24 nodes

However, as it can be observed, the cost of terrain is not taken into account such that it depends on direction connection without intermediate connection points. Moreover, despite minimizing the computational burden, it is not suitable to be applied for large areas with thousands of nodes as the computational burden will be extremely high.

In (Lotero and Contreras, 2011), a MILP-based approach was developed which is oriented towards the minimization of investment cost, operation and maintenance, and load shedding cost (translating reliability cost). The model gives a pool of solutions with different reliability percentages expressed in the common indices used in distribution system (e.g. SAIFI, SAIDI) and it allows the decision maker (The user) to analyze and choose from the pool of solutions. The main advantage is that the model is greatly focused on the reliability constraints such that the user can customize the system reliability level with respect to the availability of the budget. However, the routing of the loaded points does not consider any spatial aspects which makes the output topology to be too simplistic. In addition to that, it assumes the the loaded points are aggregation of demand which simplifies the network.

In (Paiva et al., 2005), the scope MILP optimization model is to effectively plan, not only the primary MV grid, but also the secondary LV grid. In other words, it considers the hierarchical structure of the distribution network as depicted in Figure 2.20. Most of the research efforts have ignored the LV grid despite the fact that the cost of the secondary grid may be comparable with the cost of the primary grid. Moreover, the structure of the secondary grid may affect the optimality of the primary grid topology. Therefore, integral planning of both grids is fundamental to achieve a global optimal network topology.

In (Ganjavi, 2001), it considers both MV and LV networks, but in a two-layer sequential procedure by adopting a dynamic programming algorithm. In other words, the optimization of primary and secondary grid is performed as separate problems. However, in (Paiva et al., 2005), the optimization is done simultaneously which guarantees the global optimal solution for the integral distribution system. Moreover, it also considers the possibility of having primary grid and secondary grid sharing some
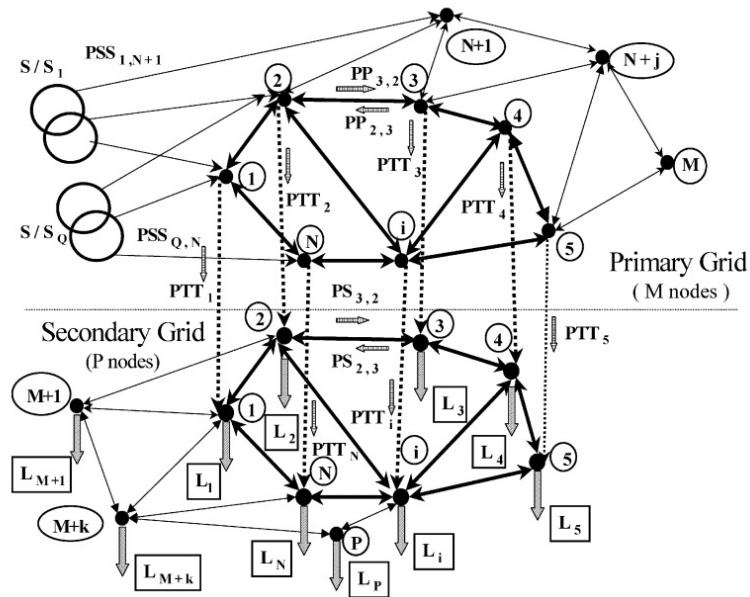
Figure 2.20: The hierarchical structure of the distribution system.

parts of the line where applicable to minimize the cost. An example of the resulting grid is shown in Figure 2.21.

The drawback of the previous models is the suitability to be applied on a large-scale network. In (Paiva et al., 2005), the model has been tested on a test case that is composed of 37 nodes, and the computation time is 29305 seconds with 47 729 971 iterations which is significant with respect to the size of the problem. In addition to that, it does not consider any geospatial data which is a significant computational effort to be added to the problem.

2.2 Rural Electrification Planning Tools

Several electrification planning tools have been developed with the objective to efficiently plan the electrification network in rural areas with the best economic use of local resources such as existing grid (possible extension), and renewable resources.

In (Ciller et al., 2019), it is a compact electrification planning tool which is called *Reference Electrification Model* (REM). It has been tested on several areas including
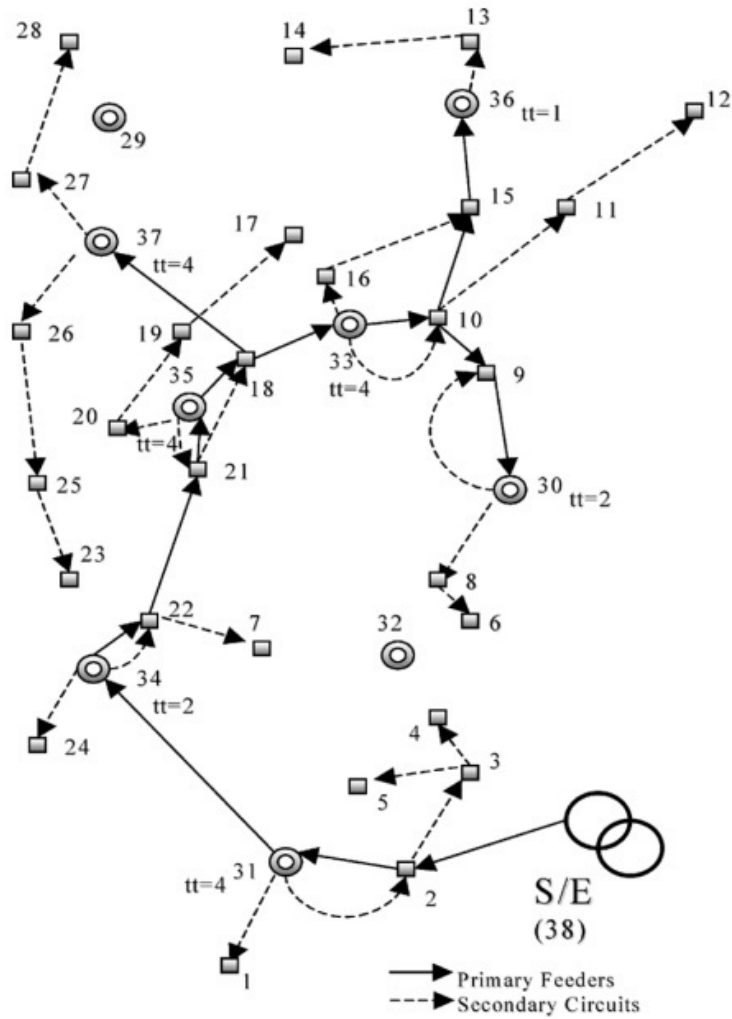
Figure 2.21: Example of the resulting integral distribution network (Paiva et al., 2005)

Uganda, Colombia, Kenya, and more. This tool is characterized by the following features:

- It is a compact tool such that it does not focus on specific aspect at a time, but it combines different planning works in one tool. It searches for the most economic electrification mode between off-grid,and grid extension solutions for each area.

- Moreover, it works with geo-referenced data and performs clustering technique to group customers in the form of independent energy communities as shown in Figure 2.22. Then, they are then analyzed individually and the optimal

electrification solution for each of them is identified between off-grid systems or grid extension solution.
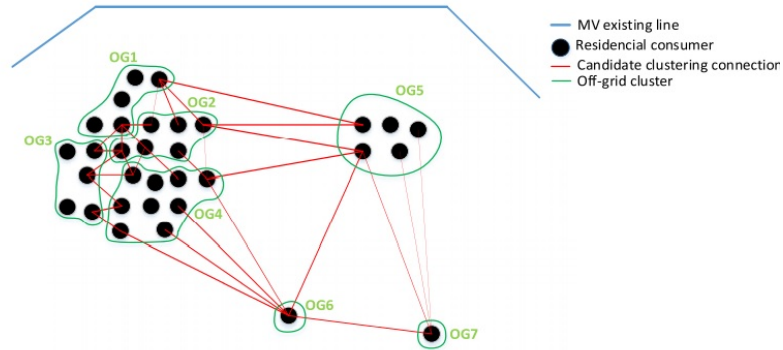
Figure 2.22: Example of Clustering Process in REM

- It gives the possibility of not supplying the entire demand if it leads to over-costly solutions.

An example of the final solution is shown in Figure 2.23. As it can be observed, some of the demand is not electrified due to leading to being expensive.
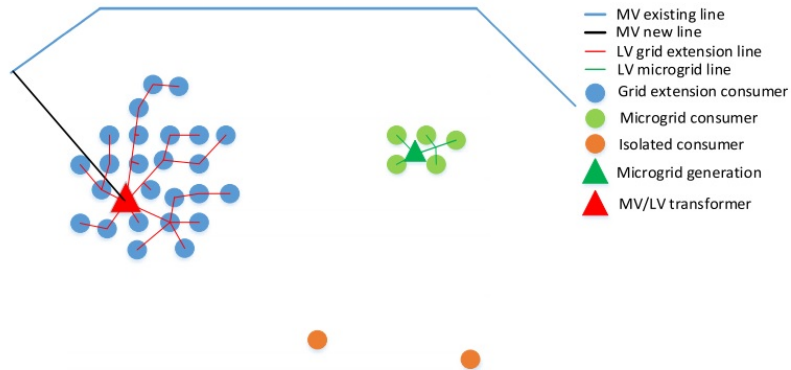
Figure 2.23: Example of the final solution using REM. (Ciller et al., 2019)

The main gap found in this tool is related to the pool of the proposed energy resources. The only considered energy resources are solar photovoltaic (PV), and diesel-based energy system. This can significantly affect the results since the inclu-

sion of wind and hydroelectric energy within the possible solutions could potentially lead to a totally different solution.

Furthermore, another weak point spotted is related to the grid routing. It relies on an external tool which is called *Reference Network Model* (RNM). It designs the routing of the network based on the GPS coordinates and the demand of each consumer and the generated energy by distributed energy sources (DER) (Mateo Domingo et al., 2011). It takes into account the electrical constraints including the voltage limits, capacity constraints, and continuity of supply. However, in the spatial aspects, it takes into account the position of lakes, natural reserves, and roads without paying attention to the impact of rivers, elevation, and land cover on the routing of the network.

Another famous electrification tool developed which is OnSSET (Open Source Spatial Electrification Tool, n.d.) and it is based on a bottom up optimization approach. It performs estimation, analysis and visualization of the most cost-effective electrification strategy. The most important advantage is being an open-source tool which is developed in Python. It performs the same functions as *REM* and evaluates the optimal electrification strategy between grid extension and Microgrid solutions. However, it does not exploit the population density layer in its full extent such that it does not efficiently deal with densely populated areas. Each populated cell is treated in as an independent entity. Moreover, another drawback spotted is that the the energy demand is estimated by multiplying the amount of population of each cell by the desired *ESMAP Multi-Tier Framework for Measuring Energy Access* in a 2030 horizon. As a result, it is not based on a real optimization since the load profile is not realistic. This leads to a sizing output with a high degree of approximation, and does not give possibility for energy mixes. In turn, this leads to inaccurate evaluation of the LCOEs.

In (Corigliano et al., 2020), a geospatial-based electrification tool has been developed which is called GIS for Electrfication (GISEle). It processes the geographical region and identifies the populated areas which are then grouped into a number of clusters in a strategic way. Moreover, the tool starts connecting these clusters using an iterative

minimum-path algorithm which based on an approximated Minimum Steiner Tree. The routing algorithm takes into account the cost of the terrain which affects the cost of line deployment in difficult areas.

The output is the internal connection of all the populated points in each cluster. Eventually, the tool starts an evaluation process for electrification solution between a stand-alone mini-grid solution, and grid extension to feed the clusters. The least-cost solution is chosen based on the one with the lowest Levelized-Cost-of-Electricity.

The main advantage is that it is an even more compact tool, it combines the different planning works in one tool: territory analysis, load estimation, generation sizing (with all possible energy resources), and electric network routing. Moreover, the electrification solution is chosen among the different technologies of Microgrid and Grid Extension.

By comparing GISEle with the available tools, the result is that GISEle is more efficient in terms of considering all aspects simultaneously (e.g. algorithms for grid routing and microgrid planning), detailing the spatial analysis, and achieving a result that is more practical without being too generic or too specific.

On the other hand, as a first version, the grid routing algorithm in GISEle does not consider the electrical constraints while forming the network topology. Moreover, the reliability assessment of the grid is not considered.

Chapter 3

GISELe: GIS-Based Electrification Tool

(Corigliano et al., 2020) developed a GIS based open-source rural electrification planning tool that is called GIS for electrification (GISEle). The aim of the tool is to find the optimum electrification solution (e.g. off-grid, grid extension) to electrify a rural area. In this chapter, the concept behind GISELe and its building blocks will be introduced. In addition to that, the new update to the project will be described which applies Main & Collateral approach.

3.1   Introduction to GISELe Project

The work depicted in this Thesis is considered an improvement of GISELe Project which is a GIS based open source tool for the design of the optimal electrification strategy in developing countries. (Corigliano et al., 2020). It exploits the spatial information such as Population Density, Elevation, Roads, and processes them to estimate the loads of the area, the potential cost for building electric lines and the renewable resources potential. Moreover, the nodes are then clustered in a strategic way using a density-based clustering algorithm (DBSCAN) to identify densely populated areas. The aim of the clustering phase is to reduce the overall cost per capita for the connection to the grid, and to reduce the computational effort as it will be shown in the following chapters how significant this issue could be. Furthermore, the network topology in each cluster is then created using Steiner Tree methodology. The algorithm is implemented in Python that is fed by GIS data as inputs. The

procedure will give indications related to the levelized cost of energy for building microgrids or connecting to the electrical network.

The entire procedure of GISELe project is described in Figure 3.1. This is the initial concept developed for GISEle which can be articulated in the following steps:
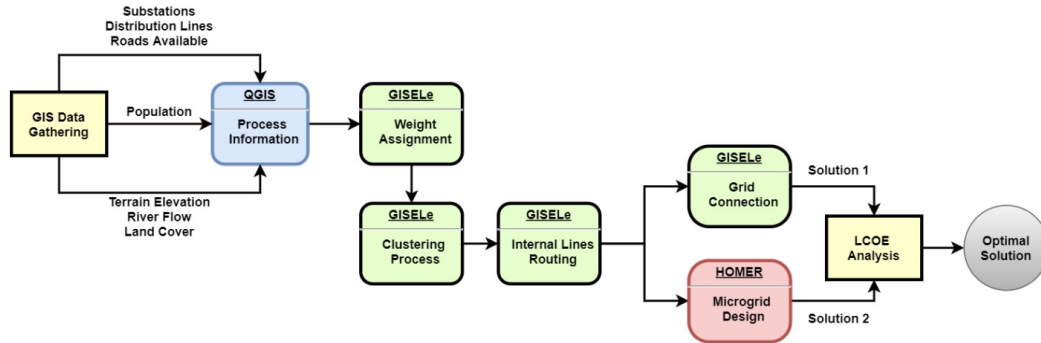
Figure 3.1: Flowchart of GISELe's initial concept

- The procedure starts by gathering the GIS data (population, terrain, and etc) of the area that needs to be electrified. The spatial data are then processed and prepared to be inserted in to GISEle tool.

- GISEle exploits the GIS data and starts assigning weight to each node in the vector layer obtained from the GIS application. The value of the weight translates the difficulty of the terrain such that each spatial aspect has a contribution factor to the weight of each node which acts a penalty factor that increases the cost of connecting that node based on the spatial nature around that node.The result is the grid of points as shown in Figure 3.2 that forms the basis for all the following analysis.

- In the clustering phase, GISEle starting clustering the area to sort the densely populated in small groups in a strategic way with aim to reduce area of the cluster while covering as much as population as possible. An example is shown in Figure 3.3.

- The next step is the routing phase in which the optimal routing of the internal lines of each cluster is computed. The routing algorithm exploits the Graph
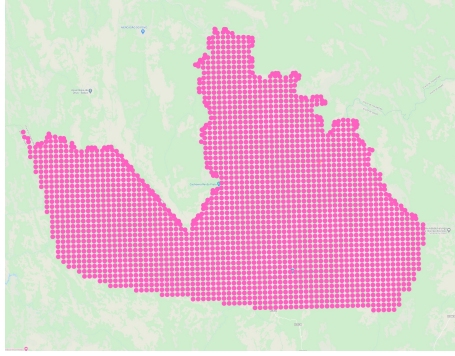
Figure 3.2: Area discretization to create grid of points
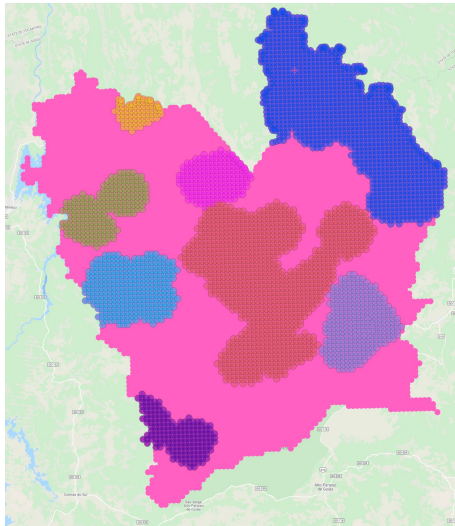
Figure 3.3: An example of the result of clustering process of GISEle

Theory in which the power system network is represented by a graph. In GISLEe's current approach, the routing algorithm relies on solving a Steiner Tree problem that was described in the previous chapter. Then, search for optimal grid connection. The aim of this thesis is improve this specific block of the entire algorithm.

- In GISELe's last step, it makes a comparison between the solution of Grid Extension and Microgrid solutions that could potentially be installed. The comparison is evaluated on the basis of LCOE analysis. The evaluation of Microgrid solution is done by the incorporation of Homer Software with GISELe.

3.2 Main & Collateral Approach with GISELe

The approach that is adopted in the previous work appears to be simplistic since in a real case, the scenario is different. As shown in Figure 3.4, it is evident that in a real case, the grid does not simply follow a shortest path solution. The reason behind this goes back to the original structure of the power system which is based on the famous hierarchical structure that is composed of HV transmission network, MV distribution network, and LV distribution network. The HV transmission is used to connect areas that are featured by heavy loads, and then the transmission network is further branched to create the so-called collaterals to feed ligher loads. This mechanism allows to efficiently categorize the size of the lines with correspondence to the nature of the load being supplied. As a result, the cables will be optimally sized which increases the economic efficiency of the topology.
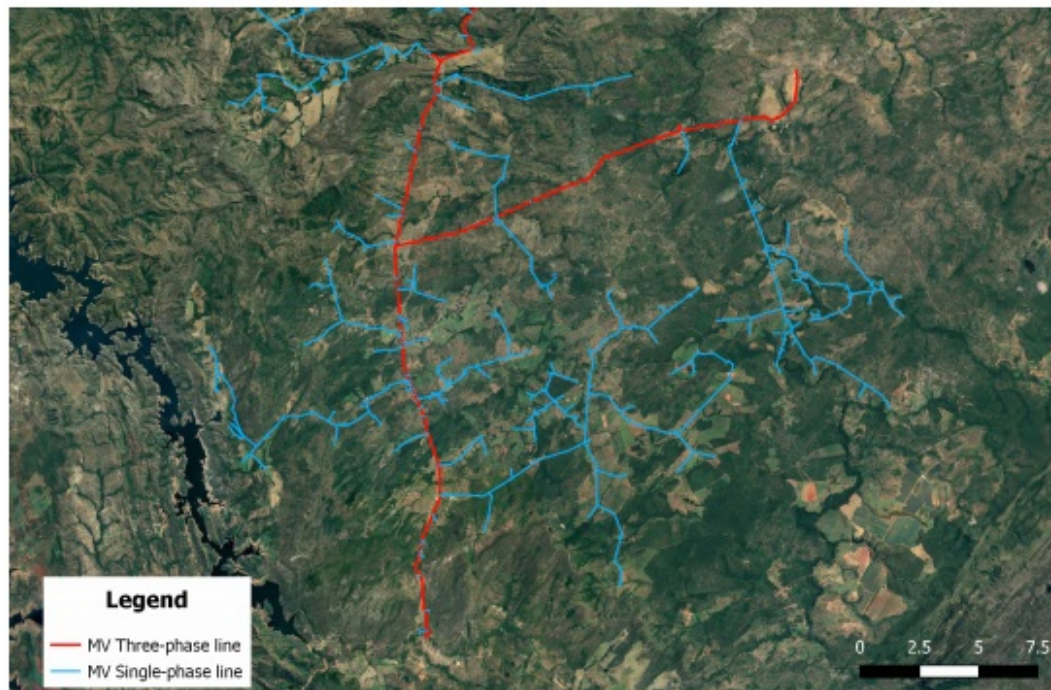
Figure 3.4: Part of MV Distribution Grid in Brazil. (Enel GI&N)

An addition in the routing algorithm that was developed to take into account this

hierarchical nature of the power system network (Vinicius, 2020). The lines are classified to:

- **Main Feeder**: It correlates the role of a HV Transmission lines which feeds heavy loads.

- **Collateral**: It represents the derivations from the main feeder to feed light loads.

In this way, each cable type will be optimized to match their respective loads. This variation in the routing algorithm is depicted in the flowchart shown in Figure 3.5. In order to route the main feeder, the resolution of the grid of point is reduced to allow the identification of densely populated points with a population threshold of 30. Then, the routing algorithm is applied to find the connection of those densely populated areas. Furthermore, the resolution is increased again to start connecting the lightly populated areas to form the collateral lines. However, in order to connect the collateral lines, the model needs to be initialized with the route already obtained for the main feeder which will be the starting point for all the collateral lines. The method of initialization is based on weight re-assignment such that weight of all the points that belong to the main feeder will be re-assigned to be zero. In this way, the model will instantly compose the main feeder before connecting the collateral lines. An example of the resulting grid is shown in Figure 3.6.
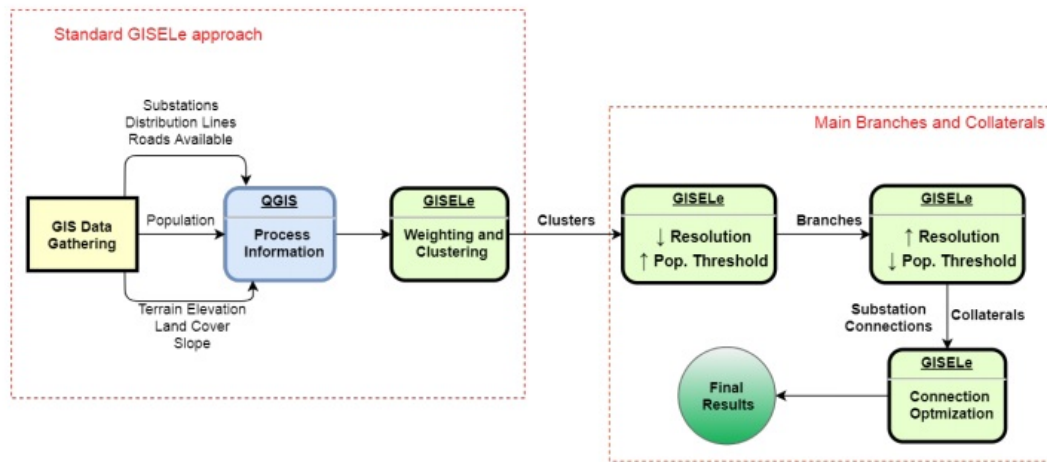
Figure 3.5: Modification of GISELe algorithm to apply Main & Collateral Approach
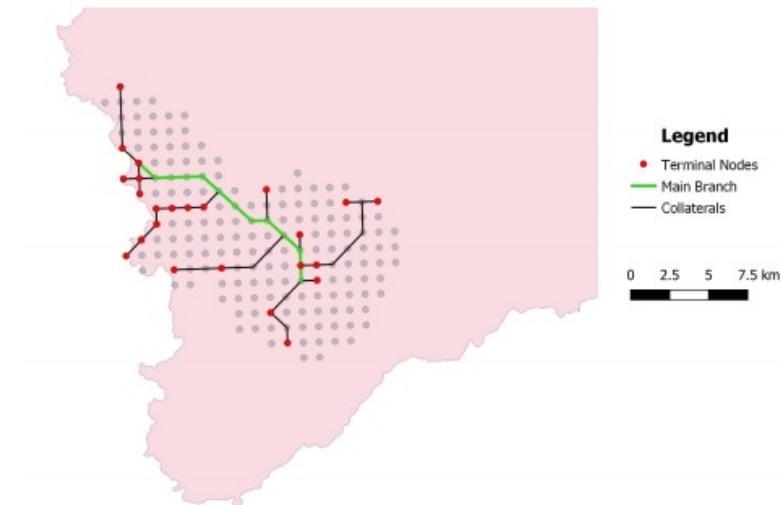
Figure 3.6: Resulting grid using Main & Collateral Approach

3.3 Case Studies

GISEle has been tested on two case studies. The result of each case study is represented in this section.

3.3.1 Namanjavira

GISELe has been tested on the area of Namanjavira which is located in the province of Zambezia in Mozambique. This area is characterized by a total area of 100 km^2 and has a total population of 5 million inhabitants. The lack of electricity in that area is translated into the percentage of people living below the poverty line which is about 70.05 % of the population. The result of running GISEle on that area is shown in Figure 3.7 which is conducted within the thesis work of (Carnovali and Edme, 2019).

In figure 3.7(a), it shows the result of grid routing to connect the entire area of Namanjavira. Meanwhile, in 3.7(b) it is the result of connecting the most densely areas only. The grid created around the most densely areas achieved 60% of population being connected. The result of GISEle approach in this case, it found that it is economically feasible to connect densely populated areas while the remaining sparse populated areas are electrified using Microgrid solutions. Hence, it made the best of use of network expansion and the allocation of energy resources of the country.

(a) (b)
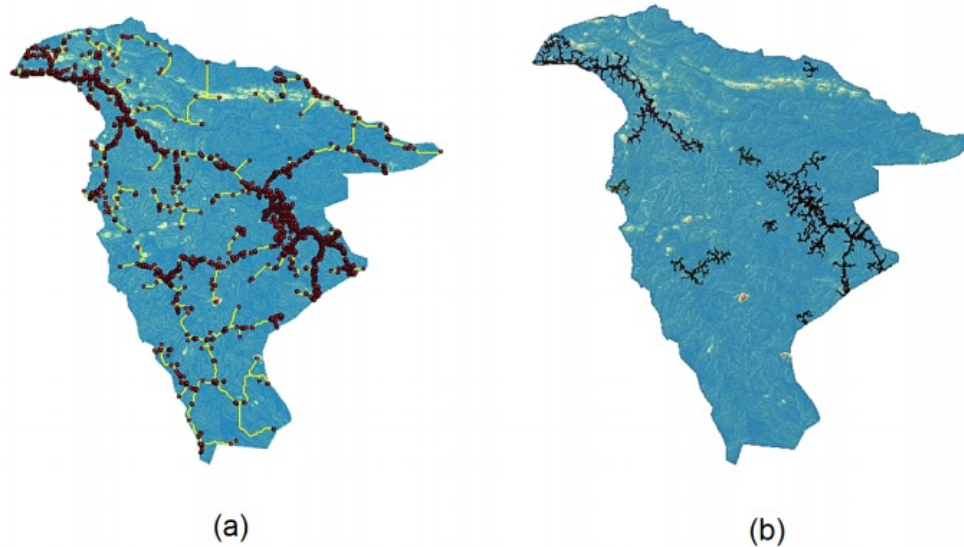
Figure 3.7: GISEle's resulting grid on the area of Namanjavira

3.3.2 Cavalcante

After the new addition in GISEle's algorithm developed by (Vinicius, 2020) to include the categorization of the line with respect to the size of the cable in order to have a more realistic result, it has been tested on another area which is Cavalcante that is located in Brazil. The existing grid in Cavalcante before the expansion is shown in figure 3.8 and after the expansion using GISEle is shown in figure 3.9. It can be observed how the result is now more realistic after categorizing the lines with respect to the size of the cables to have a Main Feeder, and its Collateral lines.
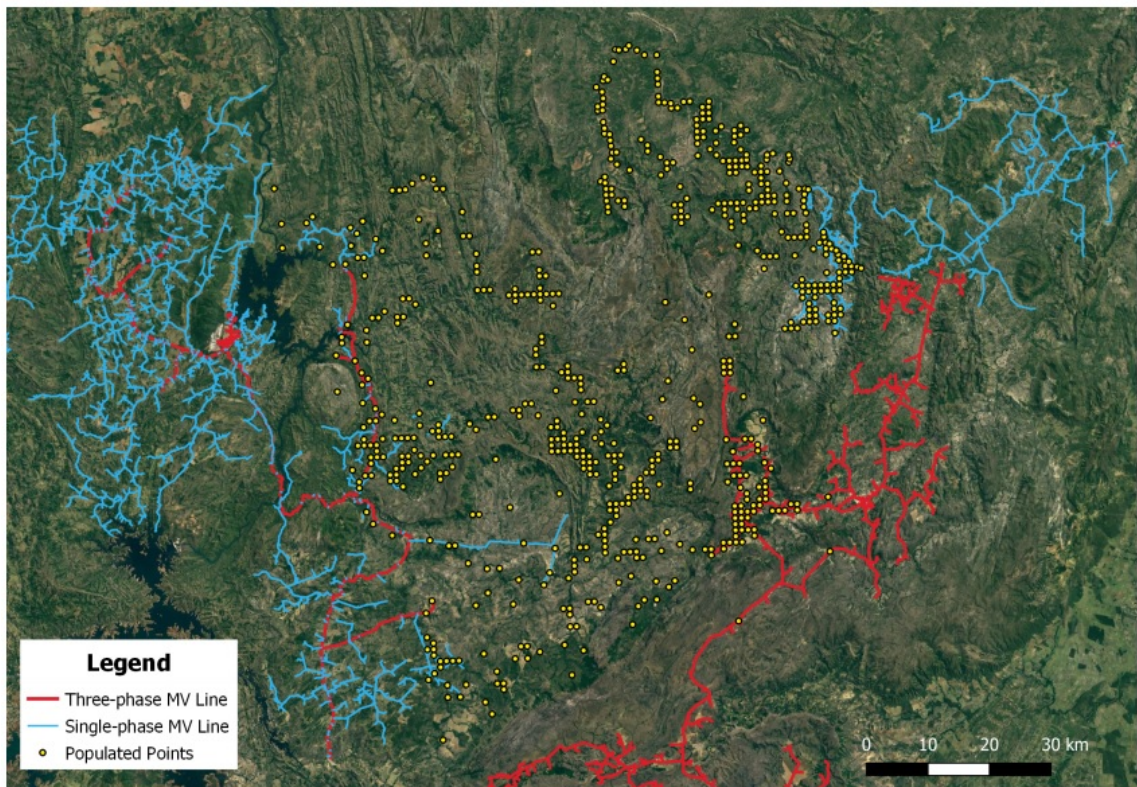
Figure 3.8: The existing grid in Cavalcance before the expansion

The gap related to the electrical aspects is yet to be filled. This in turn motivates the work of this thesis. Several scenarios could happen which could potentially affect the structure of either the Main Feeder or the collateral lines. The details related to the voltage constraints need to be sufficiently addressed to allow a realistic assessment of the grid topology. The work developed in this thesis using MILP approach is also
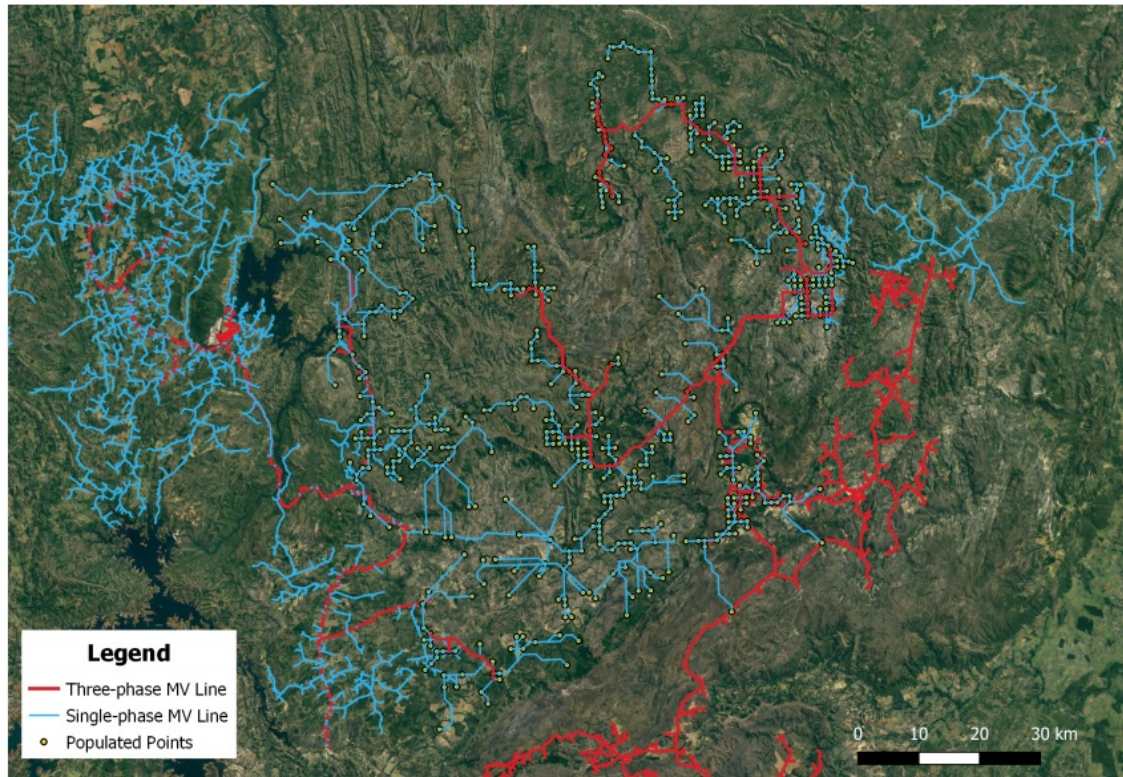
Figure 3.9: The grid after the expansion using GISEle's result

performed on the same area to allow the comparison between the two outputs, and to point out the differences between the two approaches.

Chapter 4

Proposed MILP-Based Approach

In this chapter, a new approach for power system planning will be proposed with the objective of filling the gaps that exist in GISELe's current approach, and to improve the economic efficiency of the solution.

4.1 Proposed Model

The new approach is based on MILP which provides the flexibility needed to set the electrical constraints on the routing process. MILP based optimization has been used extensively in research and literature. Not to mention, the wide range of MILP solvers available (Gurobi, CPLEX, XPress, and more). Throughout this chapter, the evolution of the MILP-Based model will be thoroughly presented starting from the base model, and then the development process of the base model. The model is implemented using Pyomo which is a Python based open source software package that supports a wide range of types of optimization problems (William E Hart, Watson and Woodruff, 2011). In fact, the modeling objects are fully embedded in Python, and this is what makes it unique over the likes of AMPL and GAMS. This advantage opens up the opportunity of exploiting the different set of supporting libraries provided by Python which is a full-featured high level programming language.

In Pyomo, there are two types of mathematical models: *Abstract Model*, and *Concrete Model* (William E. Hart et al., 2017). In the first model, it provides a context for initializing the parameters of the model with the data values only when a solution is to be obtained. In other words, the parameters of the model are kept in their

symbolic form till being initialized in a later step with the data, and this is why it is also called *Symbolic Model*. In the latter model (*Concrete Model*), the model has to be directly defined with the data values at the time of model definition. In the case study of this thesis, the *Abstract Model* is the chosen mathematical model due to its flexibility and to facilitate the manipulation of data obtained from QGIS.

Model M0: Linear Connection

The objective function of the model is the minimization of the total investment costs, given by the summation of costs for building electric lines and substations. This model is the starting point such that nodes are connected based on a linear connection without any intermediate points (Not following grid of points). The model M0 can be articulated as follows:

Sets

N	set of nodes (vertices of the graph)
A	set of arcs of the graph
links	Possible connections in the graph
PS	Node corresponding to primary substation

Scalars

V_{ref}	Reference Voltage [kV]
A_{ref}	Reference Apparent Power [MVA]
R_{ref}	Per-uni Kilometric line Resistance [ohm/km]
X_{ref}	Per-uni Kilometric line Resistance [ohm/km]
P_{max}	Maximum Allowed Power Flow [MVA]
C_f	Fixed cable cost [\$/km]
E_{min}	Minimum Voltage Level [p.u]
E_{max}	Maximum Voltage Level [p.u]
M_{max}	Maximum Allowed Value for z [p.u]
M_{min}	Minimum Allowed Value for z [p.u]

Parameters

l_{ij}	$(i,j) \in A$	Real length of connection (i,j) in km
P_{sub_i}	$i \in N$	Local demand of node i in MW

Rl_{ij}	$(i,j) \in A$	Resistance of connection in pu

$$\text{where } Rl_{ij} = R_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$$

Xl_{ij}	$(i,j) \in A$	Resistance of connection in pu

$$\text{where } Xl_{ij} = X_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$$

Variables

$x_{ij} \in (0,1)$	$(i,j) \in A$	1 if connection (i,j) exists, otherwise 0
$P_{ij} \geq 0$	$(i,j) \in A$	Power flow of connection (i,j)
$E_i \geq 0$	$i \in N$	Voltage at node i in p.u
z_{ij}	$(i,j) \in A$	Variable to linearize the problem

Objective Function

$$\min \left\{ \sum_{(i,j) \in A} C_f * l_{ij} * x_{ij} \right\} \tag{4.1}$$

Constraints

$$x_{ij} + x_{ji} \leq 1 \qquad\qquad \forall i, j \in N \qquad (4.2)$$

$$\sum_{j:(j,j)\in A} \sum_{i:(j,j)\in A} x_{ij} = length(N) - 1 \qquad \forall i, j \in N(i \neq j) \qquad (4.3)$$

$$\sum_{j:(j,j)\in A} P_{ji} - \sum_{j:(j,j)\in A} P_{ij} = P_{sub_i} \qquad \forall i \in N(i \neq j) \qquad (4.4)$$

$$P_{ij} \leq P_{max} x_{ij} \qquad\qquad (i,j) \in A(i \neq j) \qquad (4.5)$$

$$z_{ij} = e = Rl_{ij} P_{ij}^{pu} + Xl_{ij} 0.5 P_{ij}^{pu} \qquad (i,j) \in A(i \neq j) \qquad (4.6)$$

$$M_{min} * x_{ij} \leq z_{ij} \leq M_{max} * x_{ij} \qquad\qquad (4.7)$$

$$(E_i - E_j) - (1 - x_{ij})M_{max} \leq z_{ij} \qquad\qquad (4.8)$$

$$(E_i - E_j) - (1 - x_{ij})M_{min} \geq z_{ij} \qquad\qquad (4.9)$$

$$z_{ij} \leq (E_i - E_j) + (1 - x_{ij})M_{max} \qquad\qquad (4.10)$$

$$X_{ij} = 0 \qquad\qquad (i,j) \notin links \qquad (4.11)$$

$$E_{min} \leq E_i \leq E_{max} \qquad\qquad \forall i \in N \qquad (4.12)$$

$$E_i = 1 \qquad\qquad \forall i \in PS \qquad (4.13)$$

In the objective function given by (4.1) is to minimize the cost of network connection. Constraint 4.2 ensures that are no loops created. Constraint 4.3 is created to preserve the radiality of the network. Constraint 4.4 is the power conservation rule for each node such that the difference between the incoming power and the outgoing power is equal to the difference between power consumed and power generated by that node. Constraint 4.5 bounds the capacity of the line. Constraint 4.6 is the approximated formula of the voltage drop. Constraints from 4.7 to 4.10 are used to linearize the problem. Constraint 4.11 ensures that if a connection is to be used, it must exist within the pool of possible connections. Constraint 4.12 set the minimum and maximum voltage limits of each node. Constraint 4.13 set the voltage of the feeding point to be 1.

The output of this model is considered too simplistic since it only considers linear connection without intermediate points to take into account the cost terrain. Each node can be connected to the next node only using the possible links which are only

composed of target points (either load or feeding point) without intermediate points. An example of the result is shown in Figure 4.1.
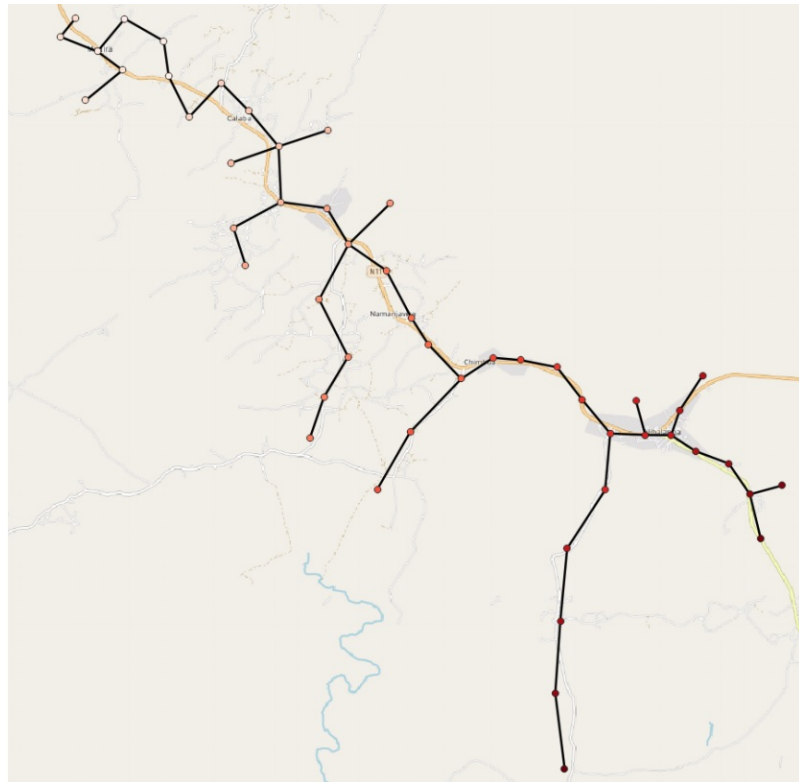
Figure 4.1: Example of grid result using model M0

Model M1: Grid of Points with Grid Connection Optimization

In this model, the objective is to allow the model to start routing following the grid of points to facilitate the consideration of the terrain cost such that the routing of the lines should deviate as much as required to avoid difficult terrain. Moreover, the objective function is modified to allow the model to search for the optimal substation to feed the network. The model is presented as shown below:

Sets

N	set of nodes (vertices of the graph)
A	set of arcs of the graph
$links$	Possible connections in the graph

Scalars

V_{ref}	Reference Voltage [kV]
A_{ref}	Reference Apparent Power [MVA]
R_{ref}	Per-uni Kilometric line Resistance [ohm/km]
X_{ref}	Per-uni Kilometric line Resistance [ohm/km]
P_{max}	Maximum Allowed Power Flow [MVA]
PPS_{max}	Maximum Power output of Substation [MVA]
PPS_{min}	Minimum Power output of Substation [MVA]
C_f	Fixed cable cost [$/km]
C_{PS}	Transformer Installation cost [$]
E_{min}	Minimum Voltage Level [p.u]
E_{max}	Maximum Voltage Level [p.u]
M_{max}	Maximum Allowed Value for z [p.u]
M_{min}	Minimum Allowed Value for z [p.u]
NF	Number of feeding points

Parameters

l_{ij}	$(i,j) \in A$	Real length of connection (i,j) in km
$l_{ij_{weighted}}$	$(i,j) \in A$	Scaled length of connection (i,j) in km
$PS_i \in (0,1)$	$i \in N$	1 if node i is an available substation, otherwise 0
$V_i \geq 0$	$i \in N$	Set-point Voltage of the feeding point (Substation)
$TX_i \geq 0$	$i \in N$	Scaling Factor of the cost of transformer installation
PSS_i	$i \in N$	Local demand of node i (Secondary Substation)
Rl_{ij}	$(i,j) \in A$	Resistance of connection in pu

where $Rl_{ij} = R_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$

Xl_{ij}	$(i,j) \in A$	Resistance of connection in pu

where $Xl_{ij} = X_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$

Variables

$x_{ij} \in (0,1)$	$(i,j) \in A$	1 if connection (i,j) exists, otherwise 0
$k_i \in (0,1)$	$i \in N$	1 if node i is chosen as a substation, otherwise 0
$P_{ij} \geq 0$	$(i,j) \in A$	Power flow of connection (i,j)
$PPS_i \geq 0$	$i \in N$	Substation Power Output
$E_i \geq 0$	$i \in N$	Voltage at node i in p.u
z_{ij}	$(i,j) \in A$	Variable to linearize the problem

Objective Function

$$\min \left\{ \sum_{(i,j) \in A} C_f * l_{ij_{weighted}} * x_{ij} + \sum_{i \in N} k_i * TX_i * C_{PS} \right\} \qquad (4.14)$$

Constraints

$$x_{ij} + x_{ji} \leq 1 \qquad\qquad \forall i, j \in N \quad (4.15)$$

$$\sum_{j:(j,j)\in A} \sum_{i:(j,j)\in A} x_{ij} = length(N) - \sum_{i\in N} k_i \qquad\qquad \forall i, j \in N(i \neq j) \quad (4.16)$$

$$\sum_{j:(j,j)\in A} P_{ji} - \sum_{j:(j,j)\in A} P_{ij} = PSS_i - PPS_i * k_i * PS_i \qquad\qquad \forall i \in N(i \neq j) \quad (4.17)$$

$$P_{ij} \leq P_{max} x_{ij} \qquad\qquad (i, j) \in A(i \neq j) \quad (4.18)$$

$$z_{ij} = e = Rl_{ij} P_{ij}^{pu} + Xl_{ij} 0.5 P_{ij}^{pu} \qquad\qquad (i, j) \in A(i \neq j) \quad (4.19)$$

$$M_{min} * x_{ij} \leq z_{ij} \leq M_{max} * x_{ij} \qquad\qquad (4.20)$$

$$(E_i - E_j) - (1 - x_{ij})M_{max} \leq z_{ij} \qquad\qquad (4.21)$$

$$(E_i - E_j) - (1 - x_{ij})M_{min} \geq z_{ij} \qquad\qquad (4.22)$$

$$z_{ij} \leq (E_i - E_j) + (1 - x_{ij})M_{max} \qquad\qquad (4.23)$$

$$X_{ij} = 0 \qquad\qquad (i, j) \notin links \quad (4.24)$$

$$E_{min} - E_{min} * k_i + k_i * PS_i * V_i \leq E_i \qquad\qquad \forall i \in N \quad (4.25)$$

$$E_{max} - E_{max} * k_i + k_i * PS_i * V_i \geq E_i \qquad\qquad \forall i \in N \quad (4.26)$$

$$PPS_i \leq k_i * PS_i * PPS_{max} \qquad\qquad \forall i \in N \quad (4.27)$$

$$PPS_i \geq k_i * PS_i * PPS_{min} \qquad\qquad \forall i \in N \quad (4.28)$$

$$\sum_{i\in N} PPS_i - PSS_i = 0 \qquad\qquad (4.29)$$

In the objective function given by (4.14), the objective is to minimize the cost of network connection and grid connection. The first part of the objective function is related to the network connection, the cost is directly related to the distance while scaling it based on the weight of the connecting nodes as shown in the following expression:

$$l_{ij_{weighed}} = l_{ij} * \frac{weight[i] + weight[j]}{2} \qquad\qquad (4.30)$$

Thus, it acts as a penalty factor to translate the cost of the terrain and spatial nature in each territory as explained in the previous chapters. However, the distance parameter ($l_{ij_{weighted}}$) is not the same parameter used to calculate the voltage drop

since after scaling it, it will no longer be the real distance between the two nodes. Hence, it will no longer represent the actual voltage drop. The distance parameter ($l_{ij_{weighted}}$) is only used to translate the cost of terrain and it must not by any means affect the actual voltage drop. Therefore, a different parameter is created (l_{ij}) to calculate the actual voltage drop using the real distances. Eventually, in order to follow the grid of points, each node has only 8 possible connections which are the neighbourhood nodes as depicted in Figure 2.16, and these node does not necessarily need to be one of the target nodes.

For the second part of the objective function, the model chooses the optimal substation as a feeding point based on its distance by using the decision k_i. As was mentioned in the previous chapters, it is not sufficient to choose the optimal substation only based on the distance. The type of the substation affects the capacity, and in turn, affects the cost since the apparent power (VA) of the transformer and other associated equipment (protection, and etc) is directly related to the cost. In order to convey this aspect to the model, another paramter is multiplied by the cost of the substation (TX). This parameter acts as a scaling factor, such that the higher the capacity of the substation, the higher is the value of TX. Moreover, if there is a possibility of extending an existing line, the TX in that case will be zero since no new substation equipment will have to be installed.

Constraint 4.15 ensures that are no loops created. For Constraint 4.16 it is only usable in case if the model will be used to connect only the populated nodes (no intermediate connection) because it would be possible in that case only to determine the number of lines before the execution of the model.

Constraints 4.25 and 4.26 set the minimum and maximum voltage limits of each node, and if the node is substation, it will be set to 1 p.u. The parameter V_i plays the role of the set-point voltage of the substation. In case the feeding point is not an actual substation, but an extension from an existing line, this parameter will represent the actual voltage at the point of line extension. Constraints 4.27 and 4.28 set the minimum and maximum power of the substation. The control variable k_i is used by the model to choose among the possible set of substations available, and the

model knows which are the available substation by means of inserting the parameter PS_i. They are both binary, but k_i cannot be set true if PS_i is false (translating the logic of AND). The only difference between the two is that PS_i is given as input (part of the data), and k_i is an output variable. Lastly, the constraint 4.29 sets the power balance constraint such that the output power of all substation must be equal to the sum of all the load. An example of the resulting grid is shown in Figure 4.2.
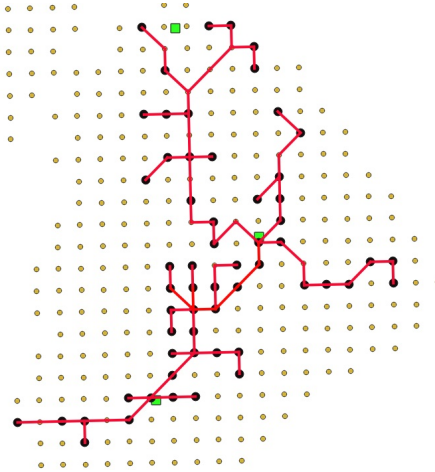
Figure 4.2: Example of resulting grid using model M1

4.2 Main & Collateral Approach with MILP

As described in the previous chapter, distribution grid in rural areas are usually composed by a main branch and collaterals. This aspect has been introduced and modelled within this thesis project, with 3 different methods.

Model M2: Post-optimization

The first method is a post-optimization method, which is to exploit the approximated load flow to allow the categorization of the lines. This is performed after the optimization process. Given the maximum power of the collateral lines, if the power flow is higher than this threshold, then the line is part of the main feeder, otherwise, it is part of the collateral lines. An example the result of using this method is shown in Figure 4.3.
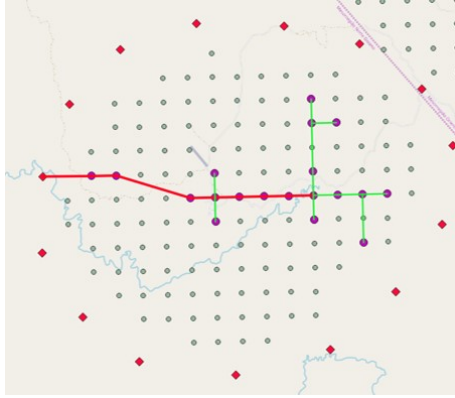
Figure 4.3: Example of applying post-optimization categorization of lines

The main feeder is shown in red, and the collateral is in green. The advantage of this method is that it is simple to be applied. However, it requires the model to design for the entire topology (Main and Collateral) simultaneously and this puts a heavy computational load on the optimization model. It is suitable to be applied for small areas or clusters, however, in case of big clusters, it may not be possible to obtain a solution. Moreover, considering the fact that the categorization is done after the optimization process, the optimality of the result may be negatively affected.

Model M3: Optimization-based Categorization

The most accurate way to design the main branch and collaterals is to include the cable type decision variables within the MILP model. This requires the modification of the optimization model as shown in red below:

Sets

N	set of nodes (vertices of the graph)
A	set of arcs of the graph
links	Possible connections in the graph
cables	Types of cables (One for Main feeder, and one for Collateral line)

Scalars

V_{ref}	Reference Voltage [kV]
A_{ref}	Reference Apparent Power [MVA]
R_{ref}	Per-uni Kilometric line Resistance [ohm/km]
X_{ref}	Per-uni Kilometric line Resistance [ohm/km]
PPS_{max}	Maximum Power output of Substation [MVA]
PPS_{min}	Minimum Power output of Substation [MVA]
C_{PS}	Transformer Installation cost [\$]
E_{min}	Minimum Voltage Level [p.u]
E_{max}	Maximum Voltage Level [p.u]
M_{max}	Maximum Allowed Value for z [p.u]
M_{min}	Minimum Allowed Value for z [p.u]
NF	Number of feeding points

Parameters

l_{ij}	$(i,j) \in A$	Real length of connection (i,j) in km
$l_{ij_{weighted}}$	$(i,j) \in A$	Scaled length of connection (i,j) in km
$PS_i \in (0,1)$	$i \in N$	1 if node i is an available substation, otherwise 0
$V_i \geq 0$	$i \in N$	Set-point Voltage of the feeding point (Substation)
$TX_i \geq 0$	$i \in N$	Scaling Factor of the cost of transformer installation
PSS_i	$i \in N$	Local demand of node i (Secondary Substation)
P_{max_C}	$C \in cables$	Maximum Allowed Power Flow [MVA]
C_{f_C}	$C \in cables$	Fixed cable cost [\$/km]
$Rl_{ij,C}$	$(i,j) \in A$	Resistance of connection in pu
	$C \in cables$	where $Rl_{ij} = R_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$
$Xl_{ij,C}$	$(i,j) \in A$	Resistance of connection in pu
	$C \in cables$	where $Xl_{ij} = X_{ref} * l_{ij} * \frac{A_{ref}}{V_{ref}^2}/1000$

Variables

$x_{ij,C} \in (0,1)$	$(i,j) \in A, C \in cables$	1 if connection (i,j) exists, otherwise 0
$k_i \in (0,1)$	$i \in N$	1 if node i is chosen as a substation, otherwise 0
$P_{ij,C} \geq 0$	$(i,j) \in A, C \in cables$	Power flow of connection (i,j)
$PPS_i \geq 0$	$i \in N$	Substation Power Output
$E_i \geq 0$	$i \in N$	Voltage at node i in p.u
$z_{ij,C}$	$(i,j) \in A, C \in cables$	Variable to linearize the problem

Objective Function

$$\min \left\{ \sum_{(i,j) \in A} \sum_{C \in cables} C_{f_C} * l_{ij_{weighted}} * x_{ij,C} + \sum_{i \in N} k_i * TX_i * C_{PS} \right\} \qquad (4.31)$$

Constraints

$$x_{ij,C} + x_{ij,C} \leq 1 \qquad \qquad \forall i,j \in N \quad (4.32)$$

$$\sum_{j:(j,j) \in A} \sum_{C \in cables} x_{ij,C} \leq 1 \qquad \qquad \forall i,j \in N \quad (4.33)$$

$$\sum_{j:(j,j) \in A} \sum_{i:(j,j) \in A} x_{ij} = length(N) - \sum_{i \in N} k_i \qquad \forall i,j \in N (i \neq j) \quad (4.34)$$

$$\sum_{j:(j,j) \in A} P_{ji,C} - \sum_{j:(j,j) \in A} P_{ij,C} = PSS_i - PPS_i * k_i * PS_i \qquad \forall i \in N (i \neq j) \quad (4.35)$$

$$P_{ij} \leq P_{max_C} x_{ij,C} \qquad \qquad (i,j) \in A (i \neq j) \quad (4.36)$$

$$z_{ij,C} = e = Rl_{ij,C} P_{ij}^{pu} + Xl_{ij,C} 0.5 P_{ij}^{pu} \qquad \qquad (i,j) \in A (i \neq j) \quad (4.37)$$

$$M_{min} * x_{ij,C} \leq z_{ij,C} \leq M_{max} * x_{ij,C} \qquad \qquad (4.38)$$

$$(E_i - E_j) - (1 - x_{ij,C}) M_{max} \leq z_{ij,C} \qquad \qquad (4.39)$$

$$(E_i - E_j) - (1 - x_{ij,C}) M_{min} \geq z_{ij,C} \qquad \qquad (4.40)$$

$$z_{ij,C} \leq (E_i - E_j) + (1 - x_{ij}) M_{max} \qquad \qquad (4.41)$$

$$X_{ij} = 0 \qquad \qquad (i,j) \notin links \quad (4.42)$$

$$E_{min} - E_{min} * k_i + k_i * PS_i * V_i \leq E_i \qquad \qquad \forall i \in N \quad (4.43)$$

$$E_{max} - E_{max} * k_i + k_i * PS_i * V_i \geq E_i \qquad \qquad \forall i \in N \quad (4.44)$$

$$PPS_i \leq k_i * PS_i * PPS_{max} \qquad \qquad \forall i \in N \quad (4.45)$$

$$PPS_i \geq k_i * PS_i * PPS_{min} \qquad \qquad \forall i \in N \quad (4.46)$$

$$\sum_{i \in N} PPS_i - PSS_i = 0 \qquad \qquad (4.47)$$

The difference between this model and Model M1 presented is the new set of indices that is added in order to categorize cables: either main branch and or collaterals. Each variable and parameter related to the links is then indexed by the cables set. The set of cables belong to the types of cables which refer to Main and Collateral line types. The main advantage of this method is that the line categorization and line routing are simultaneously optimized, leading to the most effective cost minimization. An example of the resulting grid using this method is shown in Figure 4.4. The objective of categorizing the lines efficiently is met such that the most expensive cable is deployed where needed based on the load flow result, otherwise, the cheaper (smaller) cable is used.
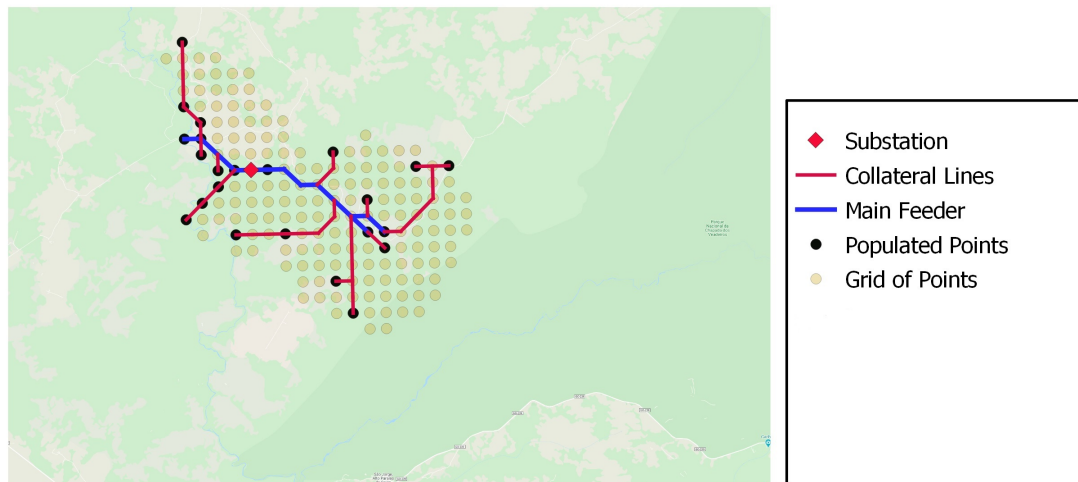
Figure 4.4: Example of grid result using Optimization-based line categorization

The increased accuracy of the algorithm comes at the expense of higher complexity and computational burden. The high computational time and convergence issues make the model feasible to be applied only in very small areas with limited number of nodes as it will be shown later in the results section.

However, this method could be used as a verification tool to check the economic and technical optimality of other simplified categorization methods. In order to use it as a verification tool, the model can be initialized with the solution of GISEle, then it will verify and modify it as needed to meet the electrical constraints.

Model M4: Sequential Optimization

This method was inspired by (Vinicius, 2020). It is based on the manipulation of the resolution of grid of points. The concept behind the main feeder is to connect concentrated loads (heavy populated territories), and the lightly loaded points are connected using the collateral lines. The algorithm will be divided to a two-step process, and this is where the name Sequential comes from. First, to connect the main feeder, the resolution of the grid of points is reduced to 4 km spaced nodes which is enough to detect heavy populated points. An example of reducing the resolution is shown in Figure 4.5. By reducing the resolution, the area of each square will be higher. The centroid of each square represent the vertices of the graph under study. This method facilitates the detection of concentrated population.
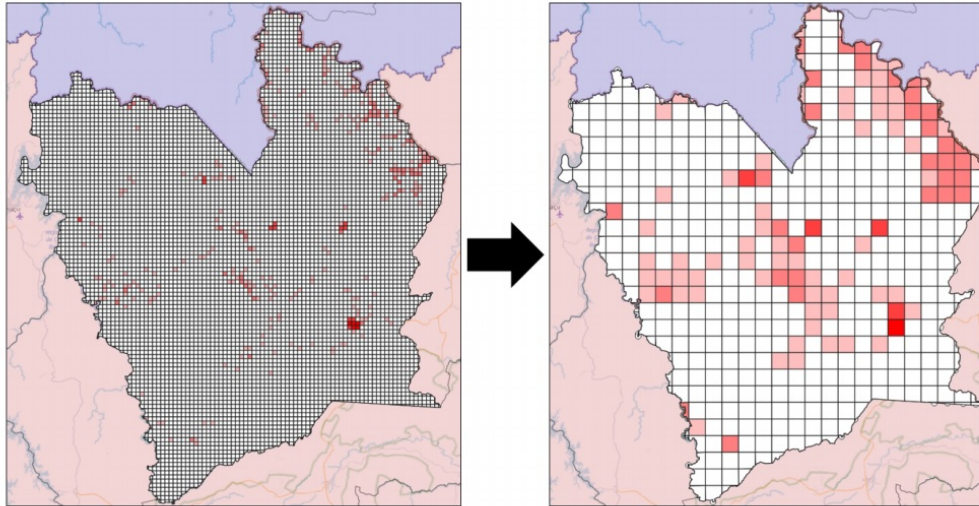
Figure 4.5: Example of reducing the resolution from 1 km to 4 km

A population threshold is set to connect only populated points that have *population* $\geq$ 30. The MILP model will then find the optimal connection of those nodes to form the topology of the main feeder. The second step is to increase the resolution back to 1 km spaced grid of points. Then, for the final time, the MILP is run again to find the optimal topology for the collateral lines. The parameter PS_i will have a value of 1 for all nodes that belong to the main feeder. In this way, the MILP model recognizes the main feeder points as potential feeding points for collateral lines. As

a result, the collateral will be derived from the main feeder, which is the objective of this approach. The following figures demonstrate the execution steps of this optimization method which is divided to two steps: Main Feeder Creation and Collateral Line Creation as shown in Figures 4.6 and 4.7.
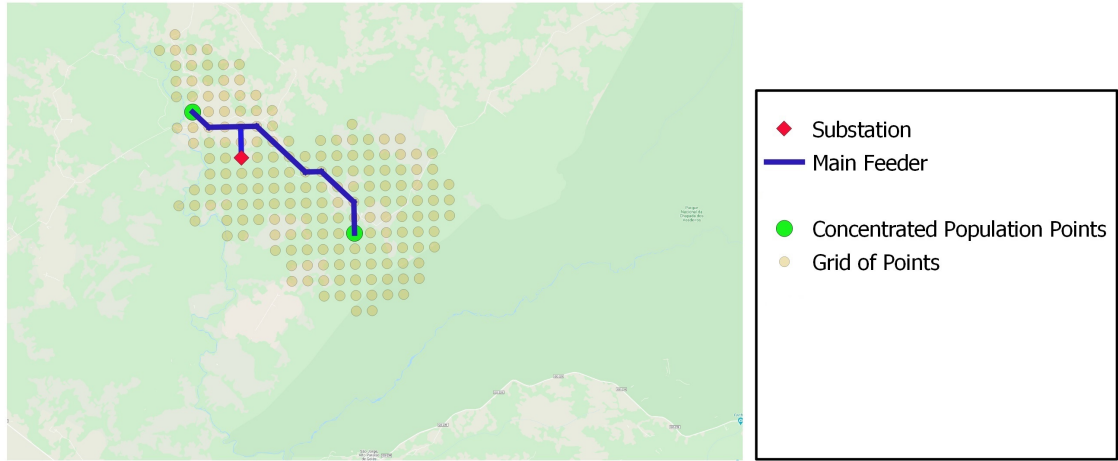
Figure 4.6: Step 1 of Sequential Optimization: Main Feeder Topology Creation
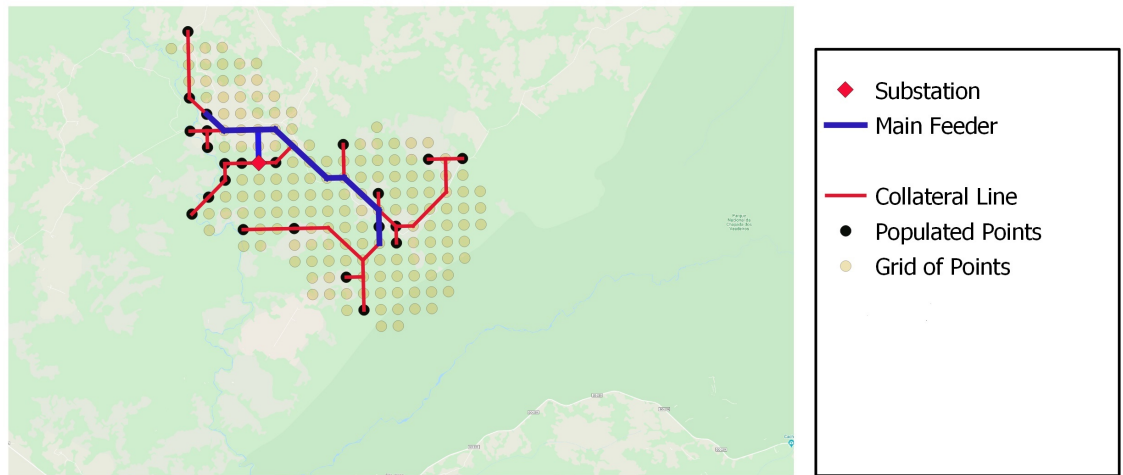
Figure 4.7: Step 2 of Sequential Optimization: Collateral Line Topology Creation

It can be observed from previous figures that the topology of the main feeder is changed from the one obtained using model M3. This is due to relying on connecting concentrated population points (population > 30 as a threshold).

However, this method also does not come at no cost. Due to the fact that the categorization is performed in two steps using MILP model by manipulating the resolution and depending on the idea of connecting densely populated points to form the Main Feeder. This makes the optimally of the solution prone to errors such that the load flow is broken down to two steps. For instance, the cable size used for the collateral line had to be increased in order to allow the model to converge. This implies that there is a loss of optimally at this point because there is no possibility for the model to extend the main feeder while working on connecting the collateral line points. In other words, since the optimization is divided to two steps, the model cannot change what already has been created in the previous step.

On the other hand, the main advantage of this method is that the optimization problem is broken down to two sub-problems to be solved separately. This significantly minimizes the computational load on the model, which is a very important aspect to be considered.

A summary of the comparison between the different models presented above are shown in table 4.1.

Table 4.1: Comparison of the different MILP-Based models developed

Model Index	Connection Methodology	Cable Size Categorization	Computational Effort
M0	Linear Connection	Not applied	Low
M1	Grid of Points	Not applied	High
M2	Grid of Points	Post-Optimization	High
M3	Grid of Points	Optimization-Based	Very High
M4	Grid of Points	Sequential-Optimization	Low

In the study case, the mainly used models are M3 and M4 for being the most efficient ones.

4.3 Limitations of MILP Model

Despite filling the gaps of the previously mentioned approaches, the model has some limitations to be overcome.

Computation Time and Convergence

The first limitation is the computation time of the model which is extremely high specially on big clusters (It could take 100,000 seconds or more). This limitation only persists when the cost of terrain is taken into account. However, when the topology is based on connecting populated points only without intermediate connection, this limitation is no longer significant.

The computation time depends on two factors: the number of nodes included in the grid of points, and the number of nodes to be connected (target points). Moreover, in the grid connection phase, if the substation is too far (Involving greater number of nodes), the computational time will be too high to the extent that the RAM limits are exceeded.

Simplistic Load Flow

Another limitation of the MILP model, the load flow is not real, but rather a very simplified (linearized) load flow which cannot be considered as accurate as an actual load flow. The power (generated/consumed) of each node is evaluated only on the basis of the difference between incoming and outgoing power flow. Moreover, only the magnitude of the voltage is considered, and the voltage drop is linearized based on an approximated version of the actual voltage drop.

The actual voltage drop is expressed as:

$$\Delta U = \frac{(RP + XQ) + j(XP - RQ)}{U_1} \tag{4.48}$$

However, the approximated voltage drop formula that is used in the model is:

$$\Delta U = \frac{RP + XQ}{U_1} \tag{4.49}$$

This implies that only the real part of the voltage drop formula is used based on the assumption that the active and reactive power are decoupled. The effect of this approximation becomes even more significant in case of MV networks due to the fact that X/R ratio is higher than HV networks. Thus, in that case, the higher the reactive power contribution, the higher will be the error. Therefore, in order to have accurate results, both parts (real and imaginary) of the voltage drop formula need to be considered, together with the actual load flow equations.

The limitations of MILP that are mentioned above call for further *supplementary tools* to enhance the operation of MILP model to allow to have a reliable, compact, and accurate planning tool for electrification system.

4.4   Supplementary Tools

In order to enhance the operation of MILP model, the following tools are proposed to overcome the issues previously mentioned.

4.4.1   Cluster Fragmentation

As discussed in the previous sections, a major limitation of the MILP model is the high computation time which becomes infeasible in case of big areas. This issue motivates a further supplementary process to allow MILP to be used efficiently in big areas.

The proposed solution is based on fragmenting big clusters to smaller sub-clusters. In other words, the problem is broken down to smaller sub-problems to be solved by MILP. This method is applicable only on Main Feeder since it connects concentrated population points which exist on the reduced resolution (4km). As a result, it facilitates the fragmentation process. However, for the collateral line, the topology will be based on populated points only, since the collateral lines are characterized by a high density which are connected based on 1km resolution. This difference is demonstrated in Figure 4.8, it can be observed how concentrated-load points (red points) are sparse and easy to be fragmented. Meanwhile, the high-resolution popu-

lation points (green points) that belong to collateral lines are too close to each other and more dense. Thus, fragmenting collateral lines would cause significant loss of optimality.
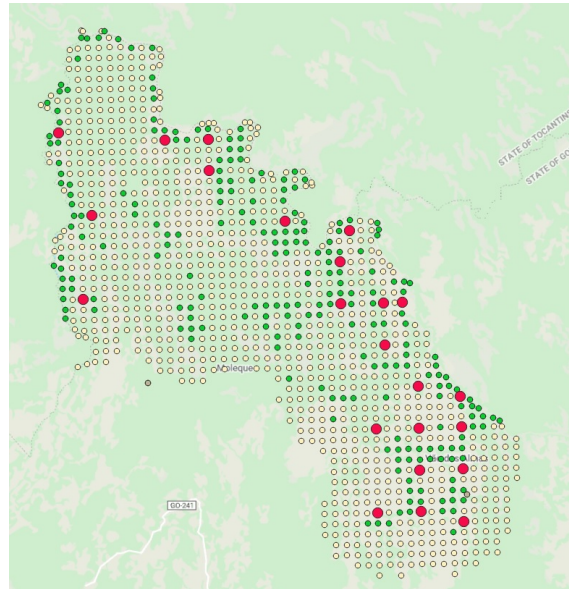
Figure 4.8: The difference between concentrated population points (low resolution) and normal population points (high resolution)

Before starting with the algorithm, there are important aspects that must be respected in order to allow this solution to be feasible:

- The voltage profile must be respected throughout the entire network. In other words, the fragmentation must not change the actual voltage of the nodes.

- A feasible starting point must be defined to start the procedure.

Due to the sparsity of the points that belong to the main feeder, density-based clustering tools (such as DBSCAN) is not suitable in this case. Therefore, the tool used to fragment the cluster is KMeans clustering tool due to its simplicity relative to the other tools.

By taking the previous aspects into account, the fragmentation process can be articulated in the following steps:

1. Apply KMeans clustering tool to start fragmenting the cluster to small sub-

clusters. The criteria will be based on the two factors that affect the computation time of the model: Number of points involved (size of grid of points), and the number of target points to be connected. Hence, the selected criteria is that each fragment contains a maximum number of points of 160 points (it could take more points, but this number is chosen to speed up the process) and a minimum of 3 target nodes (Concentrated High resolution populated points).

2. Run MILP model for each fragment separately to find the optimal internal topology of each of them.

3. Store the optimal solutions for each fragment.

4. Eventually, MILP is run on the entire full cluster while being initialized with the solutions of all the fragments that it contains. In this way, MILP will focus on finding the appropriate starting point and on choosing the optimal substation to feed the cluster.

The fragmentation algorithm can be summarized in the flowchart presented in Figure 4.9.
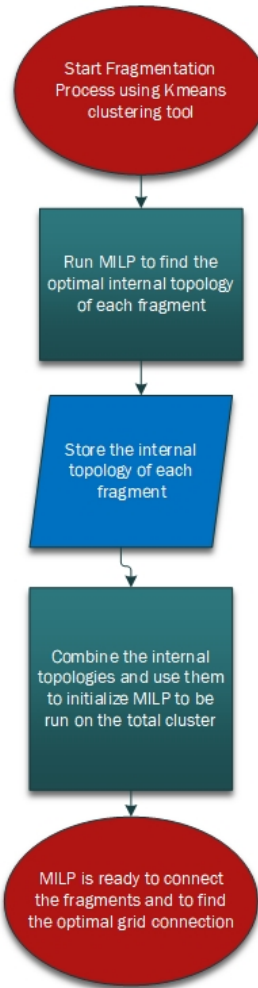
Figure 4.9: Flowchart of cluster fragmentation

4.4.2   `pandapower` - Power System Analysis Tool

`pandapower` is an open source Python-based, BSD-licensed power system analysis tool (Thurner et al., 2018). It provides power flow, optimal power flow, state-estimation, and short-circuit calculations based on **IEC 60909**. It includes a power flow solver based on PYPOWER that operates using Newton-Raphson method. Due to being python-based, it exploits the variety of supporting python libraries such as `pandas` which is used to define networks using a tabular data structure. Moreover, it can receive geographical data from QGIS such that it can be used with the interactive built-in plotting libraries in python such as `plotly` and `matplotlib`.

An example of a MV plotted network is shown in Figure 4.10

(a) Example of simple plot

(b) Example of plot on Map
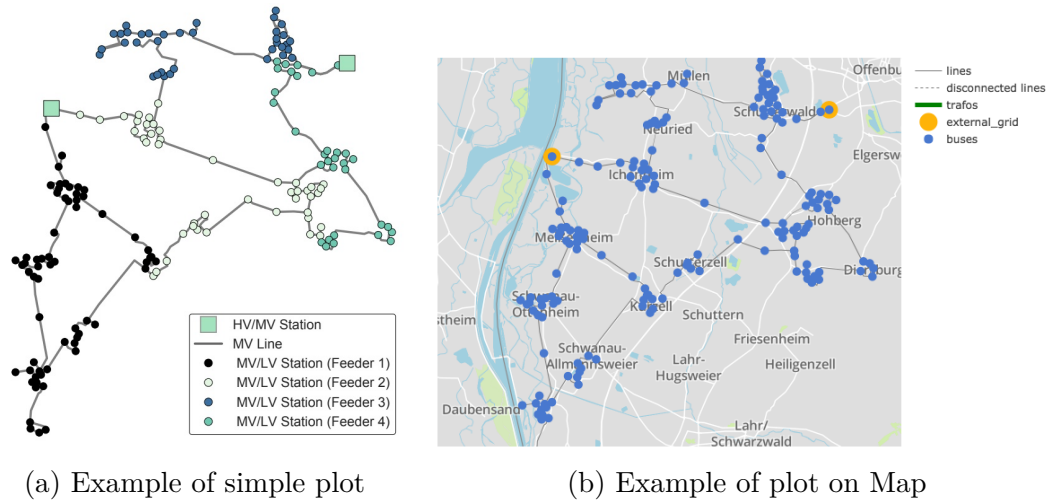
Figure 4.10: Plotting features using pandapower

Another example of the same MV network shown in Figure 4.11 which is the result of power flow analysis plotted on Map.
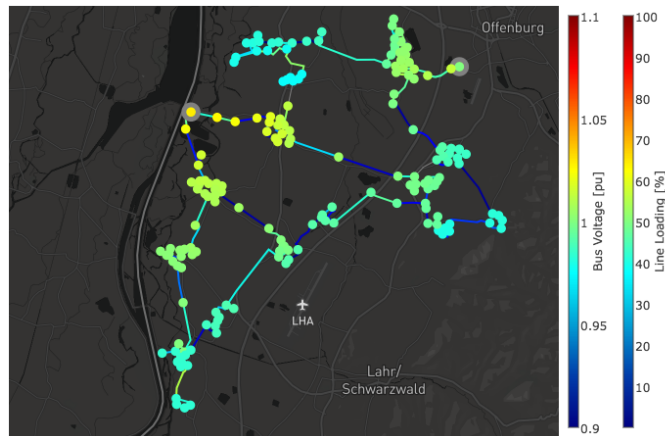
Figure 4.11: Example of Power Flow result on Map

Therefore, by considering the flexibility and tools provided by `pandapower`, it has been chosen as a suitable tool to run an actual load flow of the output topology from MILP to validate all the electrical analysis that have been approximated.

Chapter 5

Case Study

The model presented in the previous chapter has been tested on the area of Cavalcante as a real study case. It is a rural area located in Brazil, proposed by Enel, where grid expansion for unattended clients is desirable. This area has been chosen as the study case for the work of this thesis for two reasons. First, it is located in the state of Goiás in which Enel is the legal DSO. As a result, information about the existing distribution network and the location of the substations is available to allow the work of this thesis to be tested. In addition to that, the work developed by (Vinicius, 2020) has also been tested on that area. Therefore, it is good opportunity to compare the performances of the two different approaches (MILP vs GISELe approach).

5.1 Study Case: Cavalcante - Brazil

State of Goiás is located in the center-west of Brazil. It has a total area of 340.096 km^2 and population of 6 million people which makes it the most populated state. It is divided to 16 micro regions in which Cavalcante falls into one of those regions.

Cavalcante has a total population of 9709 inhabitants by 2019 as was estimated by the Brazilian Institute of Geography and Statistics. It has a total area of 6.953 km^2. It implies that the population density is considered significantly low to be around 1.4 people/km^2. The transmission system network of the state of Goiás is presented in Figure 5.1 where the Municipality of Cavalcante is highlighted in green. It can be observed that the Municipality of Cavalcante is located at the north of Goiás, on

the border with the state of Tocantins. Morever, there are two hydro power plants
(*Cana Brava* 450 MW and *Serra da Mesa* 1275 MW) located on the right side of
Cavalcante, and they are key components in the transmission network of Brazil.

Figure 5.1: Transmission network of the State of Goiás. (Cevallos-Sierra and Ramos-
Martin, 2012)

In Figure 5.2 the transmission and distribution networks of the state of Goiás is
presented in detail. The area of Cavalcante is highlighted by the yellow border.

Figure 5.2: Transmission and Distribution Network of Cavalcante

The symbols shown in the figure are representing the following:

- Red Square: *Cana Brava* with 450 MW capacity.

- Black Square: *Serra da Mesa* with 1275 MW capacity.

- Black Line: 500 kV Transmission Network.

- Red Line: Three-phase MV Line

- Green Line: Single-phase MV Line

By observing the current state of the electrical network. It becomes clear why Cavalcante was chosen as the study case such that there are several areas that could be electrified. There is a good possibility of extending the existing network to feed more areas located in Cavalcante. The objective of the proposed model is to design the optimal distribution grid, able to connect non electrified populated areas to the existing grid. The results of the developed mode, which takes into account electrical parameters, are then compared to the original GISEle approach, which is purely base of topographical analyses.

5.2 Data Analysis

The analysis of data related to area of Cavalcante has been already done by (Vinicius, 2020) using GISELe. Hence, before presenting the output of MILP, it is worthwhile to sum up the analysis of the input data that will be injected into MILP model.

5.2.1 Population

The most important geographical data to be analysed is the location of the populated points in the area under study. In Figure 5.3, the population distribution is presented which is provided by the Brazilian Institute (IGBE, 2010). Each square polygon is sized at 1 km^2. It can be observed from the distribution that the population distribution is characterized by a high sparsity. Most of the square polygons contain 10 people or none. There is one densely populated polygon located in the south which is a village that contains almost half of the total population in the entire area. The total population (summation of all squares) is 9288.
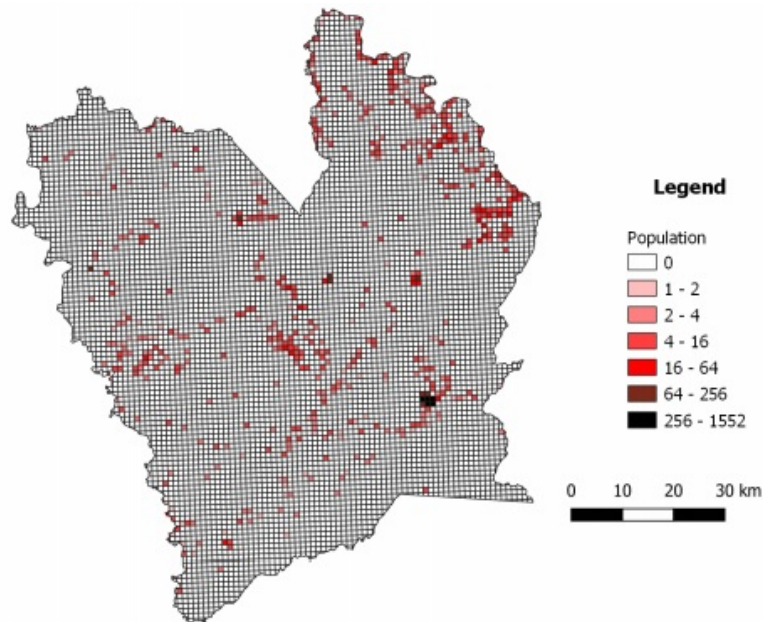

Figure 5.3: The distribution of populated points in the area of Cavalcante (IGBE, 2010)

5.2.2   Weight Assignment

As described in the previous chapters, the weight is assigned based on the spatial nature around each node. To recall, the weight acts as a penalty factor that sums up the spatial characteristics that contribute to the difficulty of deploying distribution lines. The spatial aspects taken into account are: Road Distance, Land Cover, and Elevation & Slope. By combining these aspects, the weight distribution is created as shown in Figure 5.4. It can be observed how the points that belong to the road have the lowest weight which translates the ease of line deployment along those points.
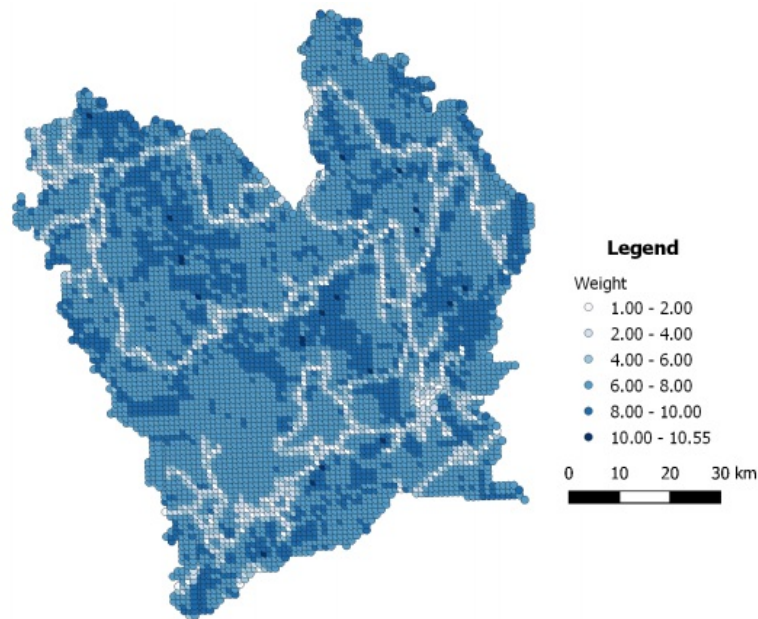
Figure 5.4: Distribution of weight assignment on Cavalcante (Vinicius, 2020)

5.2.3   Cluster Analysis

As mentioned in the previous chapters, the clustering algorithm used by GISELe is DBSCAN algorithm. Sensitivity analysis have been performed by (Vinicius, 2020) to define the two parameters: *eps* (Radius of neighborhood) and *MinPts* (Minimum population) with the objective of increasing the percentage of clustered people and reducing the total clustered area with the possibility of merging clusters if the distance between the clusters is too small. The result of the analysis defined the two

parameters to be: $eps = 3100$ and $MinPts = 20$. Figure 5.5 shows the result of the cluster analysis. There are a total number of 8 clusters with 9002 as the total number of clustered people. This implies that 96.9 % of total population is clustered, while 286 people remain unclustered.
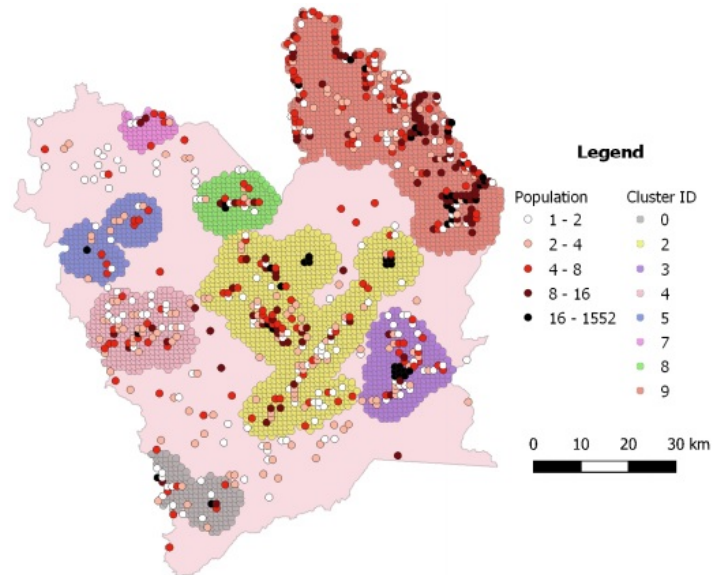
Figure 5.5: Result of the cluster analysis performed by (Vinicius, 2020)

5.2.4 Substation Availability

By exploiting the local distribution grid data provided by the DSO (Enel), and the transmission grid topology is obtained from the national transmission system operator (ONS) in Brazil, it is possible to identify the hypothetical position of the substations (Vinicius, 2020). The distance between each MV substation is estimated to be 10 km, and 20-30 km for HV substations. Figure 5.6 is the result of the estimation of substation locations.

The connection of each cluster to the substation depends on the following parameters that are related to the technical limits of the existing network:

- Single phase MV line: has a limit of connection of 300 kW.

- Three phase MV line: has a limit of connection of 3000 kW.

- HV transmission line: has no limit of connection. However, as instructed by *Enel Global Infrastructure and Networks S.r.l*, the transmission system operator of Brazil requries a power factor of 0.98 to be able to connect to the transmission network. As a result, it would be extremely costly to increase the power factor of a long MV line to 0.98. Thus, this solution is considered economically infeasible. In that case, the TX parameter of HV substation can be set to be a very high value.

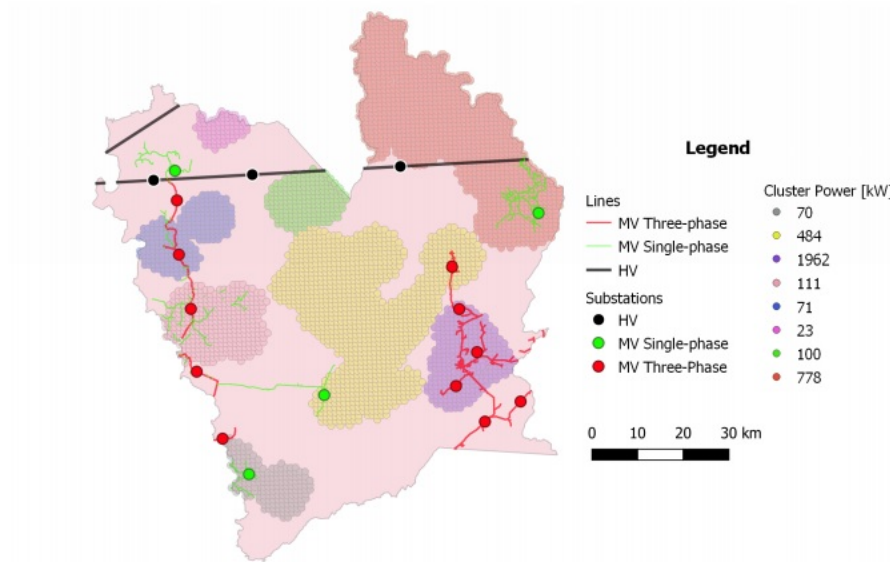Figure 5.6: Estimated Substation location as estimated using existing network (Vinicius, 2020)

Chapter 6

Results & Discussion

As described before, the objective of this thesis is improve the grid routing phase of GISELe in order to guarantee an output topology that does not violate any electrical constraints. In this section, a different usages of MILP model will be presented.

6.1 Full Network

First, the full network is presented in Figure 6.1. It is important to note that the MILP model has been applied on each cluster using the different optimization models (M3,M4) described previously to point out the performance of each one, and to compare it with GISEle model (MG) developed by (Vinicius, 2020). The electrical data adopted to perform the analysis are shown in Table 6.1. Since this data is the same used by (Vinicius, 2020), the result of the two approaches are compared with each other to point out the difference in performance. The type of cable used for the main feeder is the same as the collateral, but a 3-phase line used for the main feeder, and single-phase used for the collateral.

Table 6.1: Electrical parameters

Load per capita	400 Watt
Transformer	13.8 / 0.38 kV 45 kVA
MV Line	Alumoweld 4 AWG - 145 A 13.8kV kVA

In the next section, the map will be zoomed in to verify the performance improvement with MILP in each cluster.
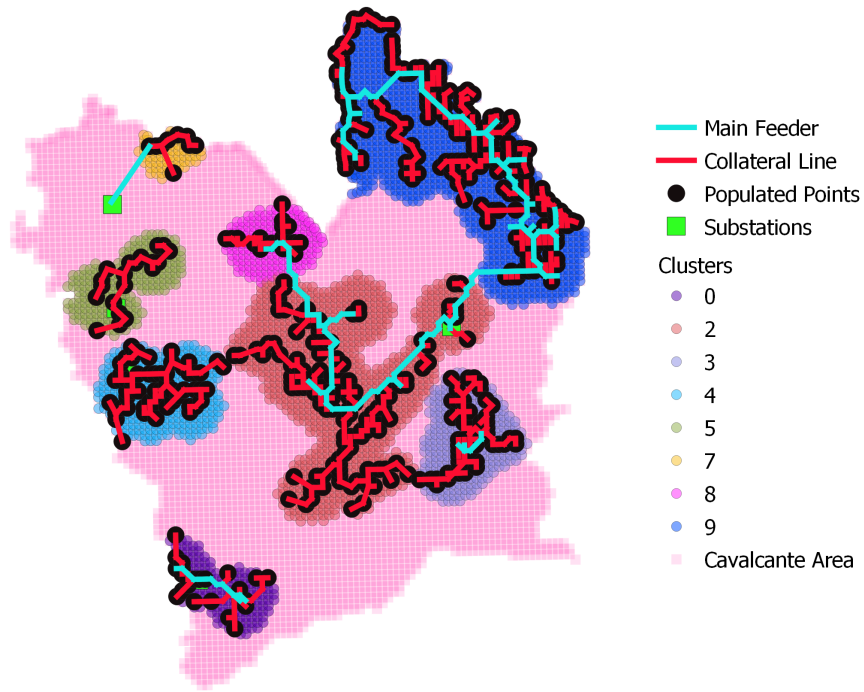
Figure 6.1: Fully Integrated MILP's Output Network Topology

In table 6.2, a summary of the statistical results of the clusters are presented while the cost of the lines have been computed using the data reported in table 6.3.

Table 6.2: Summary of the results of the Full Output Network

Cluster	Total N. of points	Covered Area $[km^2]$	Total Population	Main Feeder Length [km]	Collateral Length [km]	Main Feeder Cost [k €]	Collateral Cost [k €]	Total Cost [k €]
0	179	184	175	16	40	844	530	1374
2	1109	1225	1210	65	182	3428.75	2411.5	5840
3	303	345	4950	7	71	369.25	940.75	1310
4	347	358	291	-	84	0	1113	1113
5	224	236	179	-	38	0	503.5	503.5
7	75	70	59	12	15	633	198.75	832
8	181	193	250	18	30	949.5	397.5	1347
9	1121	1250	1945	131	216	6910.25	2862	9772
TOTAL	**3539**	**3861**	**9059**	**231**	**676**	**13130.75**	**8957**	**21142**

Table 6.3: Cost of Lines

Line Type	Average Cost [€/km]
Main Feeder	52750
Collateral	13250

The cost can then be summed up in one factor which is the cost per capita and is compared with GISEle's result in Table 6.4.

Table 6.4: Comparison of cost: MILP vs GISEle

Cluster	MILP Average Cost [€/person]	GISEle Average Cost [€/person]
0	7851	8121
2	4826	4693
3	265	267
4	3843	4267
5	2813	3048
7	14102	16639
8	5388	4975
9	5024	5128

Before the comparison, it is important to note that GISEle used a cable size slightly smaller than the main feeder for merging the clusters (cost cheaper than main feeder by 4350 €). However, both main feeder and cluster merging lines are sized the same in MILP approach. By analyzing the tables depicted, the following points have been observed.

- For clusters 4,5, and 7, it is only composed of the collateral network (Main feeder capacity is not needed) which is implemented without using grid of points (deviation to avoid difficult terrain is not considered). Therefore, the cost may get underestimated by up to 22.5%.

- For cluster 9, there are economic improvement in the topology despite having the electrical constraints.

- For clusters 0, the sizing of cables is different between GISEle and MILP such that there were some parts of the network that needed a bigger size of cable while GISEle used the size of collateral for those lines and vice versa. However, the efficient allocation of cable sizes using model M3 has leaded to an overall cost lower than GISEle's output.

In terms of computation time, GISEle definitely outperforms MILP in this aspect. The average computation time for clusters sized between 184-240 km^2 of covered area is 100-167 minutes, meanwhile, the computation time of GISEle for clusters of the same size is between 5-10 minutes. Moreover, for clusters such as cluster 9 covering an area of 1250 km^2, the computation time could exceed 1667 minutes, while with GISEle it does not exceed 120 minutes. However, with cluster fragmentation tool, the computation time of MILP can be reduced from 1667 to 167 minutes or even less.

If MILP is initialized with the solution of GISEle, the computation time can be significantly reduced to a few minutes even for clusters as big as cluster 9 and 2.

6.2   Model M3 vs GISELe

By using model M3, it means that all aspects are taken into account by MILP: terrain costs and main and collateral approach, all in one step process. This implies that the method used for line categorization is the optimization-based method (Non-Sequential). In order to allow this comparison, a small cluster from the network presented in Figure 6.1 has been chosen to allow the full operation of MILP due to the limitation of computation time.

In Figure 6.2 it presents the output of the two different approaches. It has been observed that MILP outperforms GISELe's Steiner Tree approach in both: economic and technical aspects when MILP runs in full operation mode. For instance, by looking at the connection highlighted in blue and red in GISELe's output, it can be noticed there are double connections to the main feeder which are unnecessary. Moreover, by checking the weight of the points that contribute to the path taken

by the connection, it was found that MILP also outperforms GISELe in finding the path with the minimum cost.

On the other hand, from a technical point of view, the difference in the main feeder topology between the two approaches imply that some parts of the lines in GISELe's output required a bigger size of cable and vice versa. This is due to the fact that the output obtained by MILP was based on a load flow.
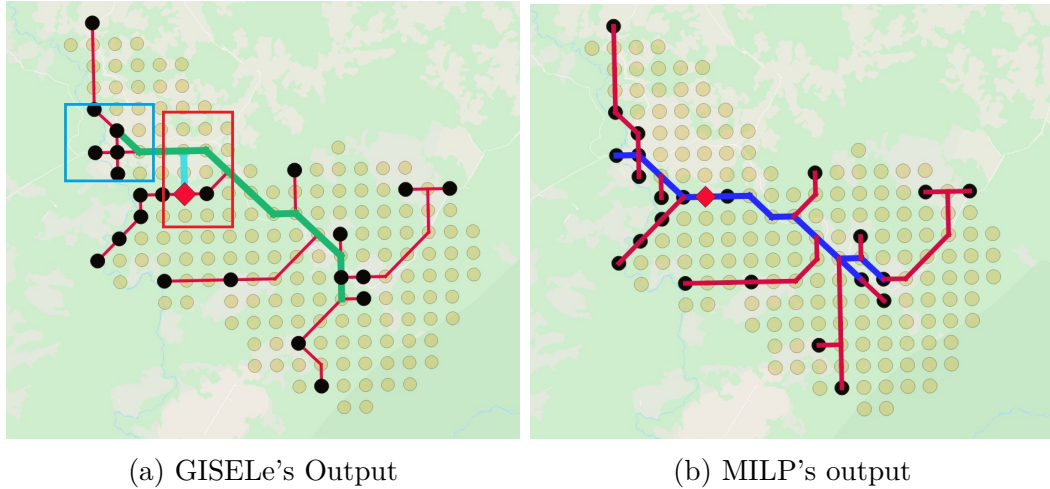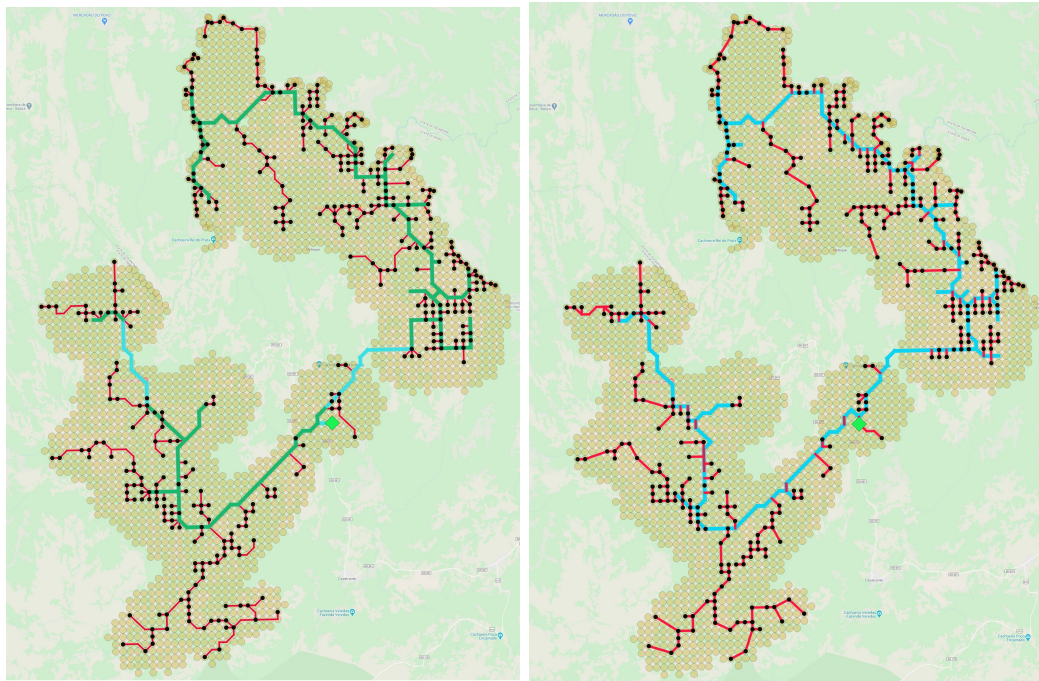

(a) GISELe's Output   (b) MILP's output

Figure 6.2: Comparison between MILP and GISELe on cluster 0

6.3   Model M4 vs GISELe

In model M4, as described previously, the optimization is broken down to two steps to find the topology of the main feeder and collateral line in two separate steps. Moreover, in order to overcome the limitation of computational effort for big clusters, *Cluster Fragmentation* methodology has also been adopted. This method has been applied on the two clusters 9 and 2 shown in Figure 6.1.

To showcase the performance of this method, it has been compared with GISELe's approach as shown in Figure 6.3. First, in the main feeder topology, MILP outperforms GISELe in terms of the optimal path to take that leads to the minimum cost. However, in the collateral topology, MILP is not able to take into account the cost of the terrain due to computational limit. Thus, MILP relied on connecting only the populated points. Although the cost of collaterals may be underestimated (up to

22.5 %) due to not considering terrain cost, MILP still outperforms GISEle in terms of cost minimization.

(a) GISELe's Output | (b) MILP's output

Figure 6.3: Comparison between MILP and GISELe on clusters 9 and 2

One flaw has been detected which is in common for both MILP and GISELe due to finding the topology for main feeder and collateral in two separated steps. This flaw can be demonstrated in Figure 6.4 as highlighted in red. It can be observed that there are unnecessary extension and deviation in the main feeder. This is due to connecting the main feeder on a separate step based on concentrated population points. However, this flaw is impossible to happen when MILP runs on full operation mode as seen in the previous comparison.

6.4 MILP initialization by GISELe's

Another possible usage of MILP is to be initialized by GISELe's solution. The role of MILP in this case is to verify the electrical constraints of GISELe's output topology, and MILP will modify it if there are any violations. In Figure 6.5, it can be observed how MILP has fixed the topology of the main feeder obtained by GISELe to meet
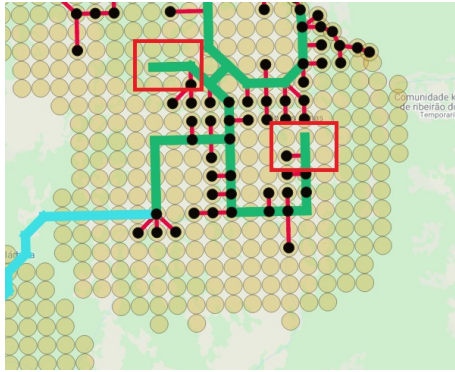
Figure 6.4: Unnecessary extensions and deviations in the Main Feeder

the power flow requirement, and then, collateral lines are modified in correspondence
to that change.
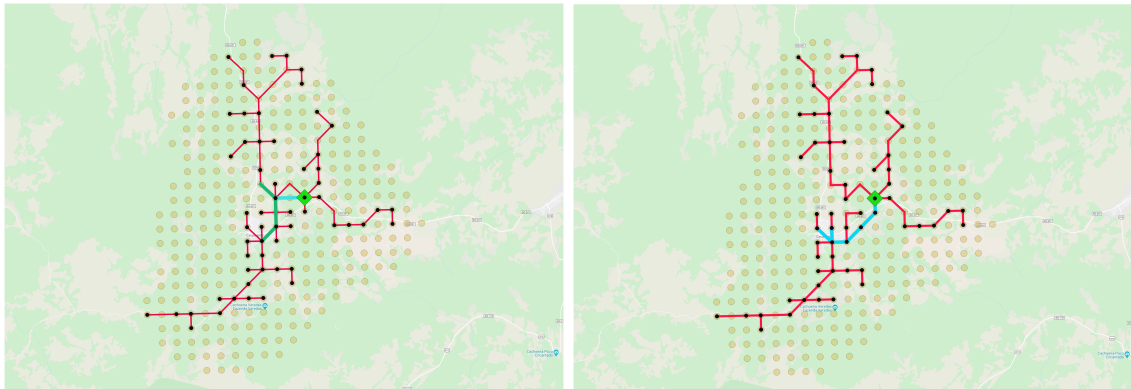
(a) GISELe's Output

(b) Initialized-MILP's output

Figure 6.5: Comparison between Initialized-MILP and GISELe

6.5 `Pandapower`: Load Flow Results

To recall, the objective of using `Pandapower` is to validate the output topology ob-
tained by GISELe by running an actual load flow (not an approximated one). Two
scenarios have been tested:

- Scenario 1: The same parameters used by (Vinicius, 2020) which are represen-
  ted in table 6.1:

- Scenario 2: A different set of parameters are chosen as shown in table 6.5.

Table 6.5: Scenario 2 electrical parameters

Load per Capita	1200 W			
Transformer	25 MVA 110/10 kV			
	Line Type	R per km	X per km	i_{max} in A
Collateral Line	34-AL1/6-ST1A 10.0	0.8342	0.36	170
Main Feeder & Interconnection Line	94-AL1/15-ST1A 10.0	0.306	0.35	350

Result of running Load Flow using Scenario 1

The load flow did not converge with the parameters chosen due to overload on both: transformers and lines, and primarily the ones feeding the clusters 9 and 2 shown in figure 6.3 and the village which is the cluster located blow them shown in figure 6.5. In this configuration, the two clusters (9 and 2) cannot share the same primary substation even after connecting two parallel transformers with increased size. However, for the other small clusters, no issues found.

This result highlights the importance of using power analysis tool such that these issues cannot be detected either by GISELe or MILP.

Result of running Load Flow using Scenario 2

On the other hand, the load flow converges with the parameters chosen in scenario two. The result of the load flow is shown in Figure 6.6. It can be observed the difference of the voltage profile of clusters 9 and 2. This is due to the difference in the total population contained in each of them. By adjusting the set-point of the substation, it was possible to satisfy the voltage limits of the two clusters.

Moreover, it can also be observed from the voltage profile that there is an abrupt decrease in one of the clusters. This abrupt change in the voltage is coming from the village that was discussed recently which is located in cluster 3 (shown in figure 6.5. This abrupt voltage change translates the amount of population contained in that village which has almost half the population (4950 people) of the entire area of Cavalcante.

For the choice of line capacity, it can be observed from the loading percentage map
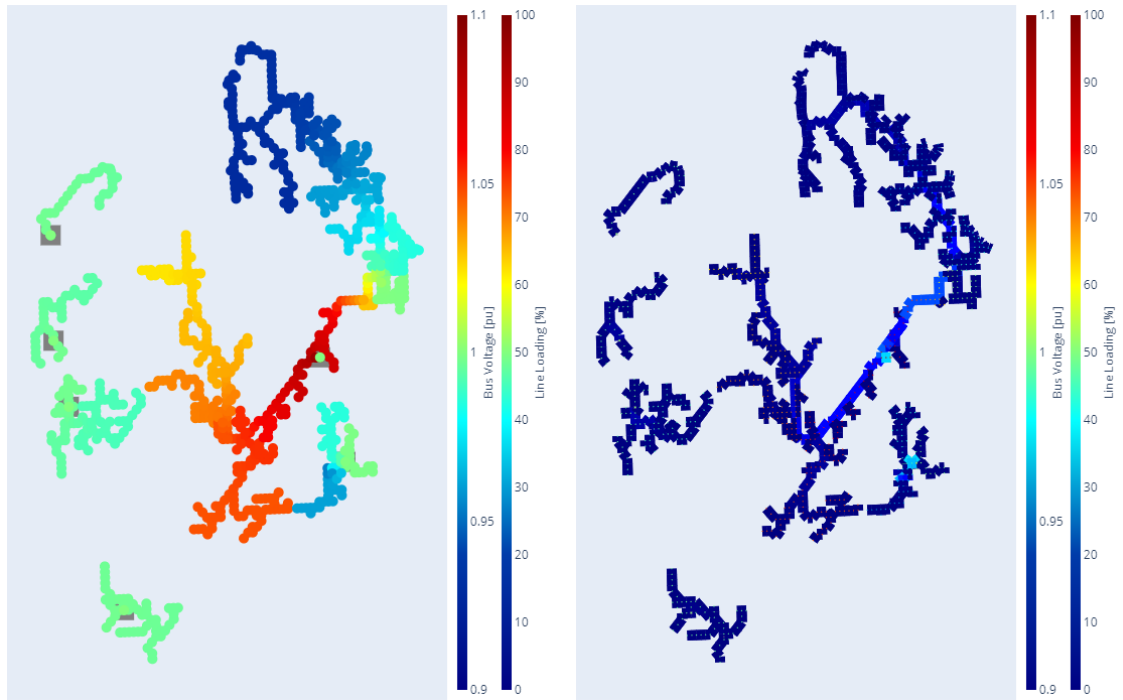
(a) Nodal Voltage Map

(b) Loading % Map
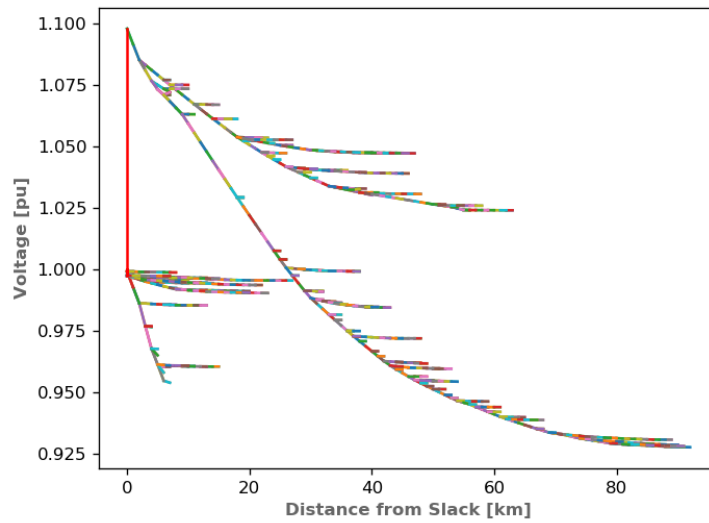
(c) Voltage Profile

Figure 6.6: Load Flow Result of Scenario 2

that both collateral and main feeder capacities could be minimized more such that the average load percentage for all clusters is around 20-30%.

Chapter 7

Conclusions and Future Work

It has been seen how challenging it is to provide electrification services in rural areas with the difficult characteristics of the terrain. Spatial analysis with the use of GIS tools is a fundamental component in electrification planning tools to reach to an optimal decision on the type of electrification solution to use. Moreover, the most important phase of any electrification planning tool is the routing phase. The grid extension solution is expensive by nature when it comes to rural areas due to the harsh environment. In order to improve the competitiveness of grid extension solution with other solutions (e.g. Microgrid), it is of great importance to adopt the most accurate optimization tools to find the optimal configuration of the network.

Recent research efforts, related to grid routing, are focused on developing between heuristic algorithms, and mathematical ones. They are either not considering electrical constraints, focusing on one aspect (e.g. Reliability), or focusing on the hierarchical structure of the distribution network without considering spatial aspects.

The work of this thesis was aimed at improving the grid routing algorithm that is adopted in GISEle electrification tool to enhance the competitiveness of grid extensions solution and to allow effective comparison with other solutions (e.g. Microgrod solutions). In addition to that, to take into account the electrical constraints which is missing from the current version of GISEle.

Different MILP-based models have been developed with different computational efficiency to allow flexible usage of MILP approach that suits more than one case. The most compact model developed is M3, which sums up several features including:

Line Size Categorization, Optimal Grid Connection, and consideration of terrain cost. The main challenge faced is the computation time required by model, such that it is significantly high for clusters with coverage area around 1250 km^2. However, the model was designed to allow initialization by other heuristic algorithms to overcome the limitation of computational effort which makes it a good candidate to be integrated with GISEle tool.

Eventually, the model has been tested on the same study case that was used by (Vinicius, 2020) to point out improvements and the difference between the current version of GISEle and the adoption of MILP-Based approach. Overall, there is no doubt that the economic improvement are noticed when MILP is fully operated without using linear connection. Moreover, the sizing of lines is efficiently allocated by using M3 model even though it may lead to higher costs. Moreover, `pandapower` has been used to electrically validate the electrical parameters which has been obtained in a very approximated way. By using `pandapower`, it was possible to run an actual load flow. In addition to that, it was found that `pandapower` can be efficiently integrated with GISEle as an effective power analysis tool with very low computational effort.

Despite having reached these results, MILP-based approach still needs improvement to be fully compact. First, the computation time has to be further optimized to allow it to run independently without any initialization from heuristic algorithms. Moreover, reliability assessment is an important aspect to be included together with the evaluation of the Structure Criticality Parameter (PCR) of the output topology. Therefore, these aspects are considered as future work to be developed.

Bibliography

Carnovali, T. and D. Edme (2019). 'GISEle: an innovative GIS-based approach for electric networks routing'. In: *Ph.D Thesis* 1 (cit. on pp. 16, 17, 36).

Cevallos-Sierra, J. and J. Ramos-Martin (2012). 'Spatial assessment of the potential of renewable energy: The case of Ecuador, Renewable and Sustainable Energy Reviews,' in: 3.3, pp. 1154–1165. DOI: 10.1016/j.rser.2017.08.015 (cit. on p. 63).

Ciller, P. et al. (2019). 'Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM)'. In: *Proceedings of the IEEE* 107.9, pp. 1872–1905 (cit. on pp. 25, 27).

Corigliano, Silvia et al. (2020). 'Holistic geospatial data-based procedure for electric network design and least-cost energy strategy'. In: *Energy for Sustainable Development* 58, pp. 1–15. ISSN: 0973-0826. DOI: https://doi.org/10.1016/j.esd.2020.06.008. URL: http://www.sciencedirect.com/science/article/pii/S0973082620302611 (cit. on pp. 28, 30).

Ganjavi, M. R. (2001). '"Coordinated comprehensive design of distribution systems at medium and low voltage levels'. In: *n Proc. 16th Int. Conf. Exhib. Elect. Distrib* 5.29.482, pp. 1–6 (cit. on p. 24).

Hart, William E. et al. (2017). *Pyomo–optimization modeling in python.* Second. Vol. 67. Springer Science & Business Media (cit. on p. 39).

Hart, William E, Jean-Paul Watson and David L Woodruff (2011). 'Pyomo: modeling and solving mathematical programs in Python'. In: *Mathematical Programming Computation* 3.3, pp. 219–260 (cit. on p. 39).

Hekken, Bemmelen J. W. Quak M. and P. Oosterom (1993). *Vector vs . raster-based algorithms for cross country movement planning* (cit. on p. 8).

IGBE (2010). 'Censo Demográfico 2010 - Características da População e dos Domicílios., Tech. rep., Instituto Brasileiro de Geografia e Estatistica.' In: (cit. on p. 65).

International Energy Agency International Renewable Energy Agency, United Nations Statistics Division (2019). *The Energy Progress Report 2019. World Bank.* URL: https://trackingsdg7.esmap.org/data/files/download-documents/2019-Tracking%5C%20SDG7-Full%5C%20Report.pdf (cit. on p. 1).

Livemint (2018). *The problem of lack of rural electricity demand.* URL: https://www.livemint.com/Opinion/nrgFBVsILzEtbJf7pwfH3H/The-problem-of-lack-of-rural-electricity-demand.html (cit. on p. 1).

Lotero, R. C. and J. Contreras (2011). 'Distribution System Planning With Reliability'. In: *IEEE Transactions on Power Delivery* 26.4, pp. 2552–2562 (cit. on p. 24).

Mateo Domingo, C. et al. (2011). 'A Reference Network Model for Large-Scale Distribution Planning With Automatic Street Map Generation'. In: *IEEE Transactions on Power Systems* 26.1, pp. 190–197 (cit. on p. 28).

NCGIA (1990). *Geographical Information System.* URL: http://www.ncgia.ucsb.edu/ (cit. on p. 4).

Open Source Spatial Electrification Tool (n.d.). URL: http://www.onsset.org/ (cit. on p. 28).

Paiva, P. C. et al. (2005). 'Integral planning of primary-secondary distribution systems using mixed integer linear programming'. In: *IEEE Transactions on Power Systems* 20.2, pp. 1134–1143 (cit. on pp. 24–26).

Samui, A., S. R. Samantaray and G. Panda (2012). 'Distribution system planning considering reliable feeder routing'. In: *IET Generation, Transmission Distribution* 6.6, pp. 503–514 (cit. on p. 22).

Shu, J. et al. (2012). 'A New Method for Spatial Power Network Planning in Complicated Environments'. In: *IEEE Transactions on Power Systems* 27.1, pp. 381–389 (cit. on pp. 21, 22).

Thurner, L. et al. (2018). 'pandapower - an Open Source Python Tool for Convenient Modeling, Analysis and Optimization of Electric Power Systems'. In: *IEEE Trans-*

actions on Power Systems. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2018.2829021. URL: https://arxiv.org/abs/1709.06743 (cit. on p. 60).

U.S.-Energy-Information, Administration (2019). *International Energy Outlook.* URL: https://www.eia.gov/outlooks/ieo/pdf/ieo2019.pdf (cit. on p. 1).

Vinicius, Gadelha T. Filho (2020). 'GIS-Based Strategy For Distribution Grid Planning in Rural Areas'. In: *Master Thesis* (cit. on pp. 37, 53, 62, 65–69, 75, 79).

Appendix A

Parameters Data

A.1   Data for Model M3

param: V_ref := 20;

param: A_ref := 1;

param: Cable_Types := 2;   *# Number of cable types*

param: R_ref :=   *#Resistance of each cable type [Ohm/km]*

1   0.42

2   0.52;

param: X_ref :=   *#Reactance of each cable type [Ohm/km]*

1   0.39

2   0.49;

param: P_max :=   *#Maximum power flow of each cable type [MW]*

1   3

2   0.3;

param: cf :=   *#Cost of lines [/km]*

1   52750

2   13250;

param: E_min := 0.9;

param: E_max := 1.1;

param: PPS_max := 100;

param: PPS_min := 0;

param: cPS := 2000;

A.2 Data for Model M4

param: V_ref := 20;

param: A_ref := 1;

param: R_ref := 0.642;

param: X_ref := 0.083;

param: P_max :=5;

param: cf := 52750;

param: E_min := 0.9;

param: E_max := 1.1;

param: PPS_max := 100;

param: PPS_min := 0;

param: cPS := 2000;