POLITECNICO DI MILANO
Doctoral Programme in Electrical Engineering
Department of Energy

Geospatial based methodology for rural electrification planning

Doctoral Dissertation of:
Silvia Corigliano

Supervisor:
Prof. Marco Merlo

Tutor:
Prof. Alberto Berizzi

The Chair of the Doctoral Program:
Prof. Marco Mussetta

2022-XXXIV Cycle

A chi è andato in un bel posto

Acknowledgments

The present thesis work is the outcome of a cooperation between many people who worked and believed in the project.
Thanks to my supervisor, Professor Marco Merlo, who guided me during my research with kindness and expertise.
Thanks to all the "Gisele" colleagues, Aleksandar, Carla, Darlain, Marina, Nour, Tommaso, Vinicius whose works became fundamental bricks of this manuscript.
Thanks to my tutor, professor Alberto Berizzi, and to the other researchers of Politecnico who communicate passion and knowledge in their subjects.
Thanks to all my office mates, to their laughters, friendship and support.
Thanks to Lacor Hospital, Fondazione Corti, Jacopo and all the special people I met in Uganda, who taught me what it means to fight and to dream.
Thanks to all the people I love and could not hear me saying one more time "non posso, devo scrivere la tesi".

Abstract

The research work presented in this manuscript has the general goal of addressing the problem of energy access in rural areas of the Global South, proposing effective solutions that could foster the electrification process considering both on and off-grid technologies. More specifically, the present thesis aims to give a significant contribution to the literature models for rural electrification planning by creating a new open source and open access modeling framework, usable by different stakeholders, from university researchers, to Non Governmental Organization (NGO)s and private companies. A novel procedure, named Gisele (Geographic Information System (GIS) for Electrification) has been developed, coded and validated over the real-life case studies. The proposed approach is a comprehensive model, subdivided into four blocks, able to estimate the energy needs of an area, size the optimal generation portfolio for off-grid systems, identify the optimal electrification solution and design the distribution network. With respect to other literature tools its main strengths reside in the ability of including a detailed sizing of hybrid microgrid and routing of the distribution grid, the reliance on optimization algorithms for the optimal electrification solution identification and the integration of a multi-objective optimization including environmental and social dimensions. The procedure has been tested for designing the rural electrification plan in a area in Mozambique and Lesotho. In both cases, that differ in terms of load distribution and geographical characteristics it performed well being able to plan an electrification strategy considering both grid extension and off-grid systems.

Acronyms

AC Alternate Current

AMC Antecedent Moisture Conditions

ARI Adjusted Rand Index

BESS Battery Energy Storage System

CAPEX Capital Expenditure

COE Cost of Energy

DC Direct Current

DEM Digital Elevation Model

DER Distributed Energy Resource

DG Diesel Generator

DSO Distribution System Operator

ESMAP Energy Sector Management Assistance Program

GA Genetic Algorithm

GDP Gross Domestic Product

GIS Geographic Information System

GUI Graphical User Interface

HDI Human Development Index

HT Hydro Turbine

HV High Voltage

IEA International Energy Agency

LCA Life Cycle Assessment

LCOE Levelized Cost of Electricity

LEC Lesotho Electricity Company

LP Linear Programming

LV Low Voltage

MILNP Mixed Integer Non Linear Programming

MILP Mixed Integer Linear Programming

MST Minimum Spanning Tree

MTF Multi Tier Framework

MV Medium Voltage

NGO Non Governmental Organization

NLP Non Linear Programming

NPC Net Present Cost

O&M Operation and Maintenance

OPEX Operational Expenditure

OSM Open Street Map

PFR Primary Frequency Regulation

PV Photovoltaic

RES Renewable Energy Sources

RI Rand Index

RISE Regulatory Indicators for Sustainable Energy

SA Stand-Alone Systems

SAIDI System Average Interruption Duration Index

SAIFI System Average Interruption Frequency Index

SDG Sustainable Development Goals

SHS Solar Home System

UNDP United Nations for Development

WT Wind Turbine

Contents

Chapter 1

Introduction and motivations

The research work presented in this manuscript has the goal of providing a small contribution to the demanding challenge of universal energy access. Sustainable energy for all by 2030 is one of the United Nations for Development (UNDP) Sustainable Development Goals (SDG), globally recognized as one of the major challenges of the millennium for which worldwide stakeholders are called to take action and provide effective solutions. Among the topics related to energy access, the research focuses on the issue of rural electrification that in 2019 still concerns 800 million people in the world without access to electricity, mainly concentrated in Sub-Saharan Africa.

The manuscript is divided into three parts, that describe the undergone research process and the results achieved.

Part I

The first part of the work is dedicated to the comprehensive study of the electricity access problem, from the diffusion of electrification in the world (Chapter 2), to the different dimensions of the rural electrification problem (Chapter 3) to the literature modeling framework for performing electrification planning studies.
Objective: analyse literature related to electricity access to understand which are the available tools and solutions for electrification.
Research questions:

1. Which are the possible strategies for electrification? What influences their adoption?

2. Which type of instruments would help stakeholders performing rural electrification planning?

3. Which are the literature gaps in models for rural electrification planning?

Part II

The second part of the manuscript describes the new modeling framework proposed by the author. It is a holistic geospatial data based procedure that could be used by various stakeholders for performing pre-feasibility studies of the electrification planning of rural areas. The procedure, named Gisele (GIS for Electrification), is composed by several modules, to cover different aspects of the problem.
Objective: create a new open-source and open-access modeling framework, usable by different stakeholders to tackle the issue of rural electrification planning.
Research questions:

1. How could open-source geospatial data be used for electrification planning?
2. How to identify the optimal electrification solution choosing among different technological options?
3. Can optimization consider also non economical aspects, such as technical and environmental dimension?

Part III

The third part of the research is rooted on real life case studies where the modeling framework has been tested to assess its performance and limitations. In particular Chapter 10 analyses the rural area of Namanjavira, in Mozambique and Chapter 11 the region of Butha-Buthe in Lesotho.
Objective: test the procedure on real life case studies and assess its performance and limitations.
Research questions:

1. How does the proposed procedure perform in two different contexts?
2. How are solutions influenced by different models' assumptions?
3. Which are the possible areas of improvement?

Part I

Overview on sustainable access to electricity

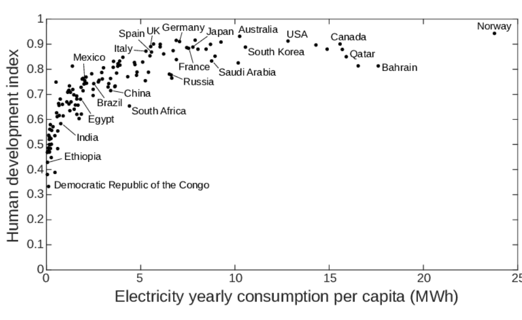
Chapter 2

Electricity access in the world

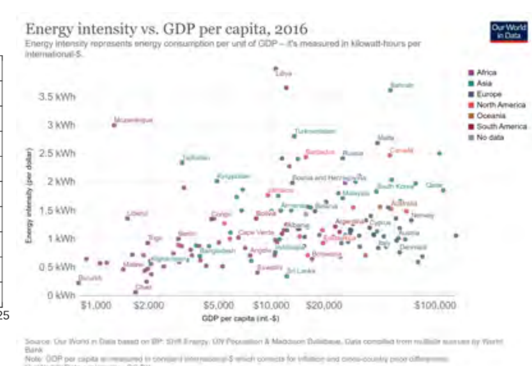
2.1 Introduction

This work starts with a simple yet fundamental statement: energy and development are strictly related. Energy is necessary for many fundamental aspects of modern society: cooking, heating, sanitation, telecommunication services, industries, transportation and more, rely of some form of energy. Graphs of figure 2.1a and 2.1b show the correlation between per capita energy consumption, Human Development Index (HDI) and Gross Domestic Product (GDP) of different countries across the world: it is evident that a high energy consumption is related to higher levels of human and economic development.

Not surprisingly, the United Nations put energy at the heart of many of the SDGs, the 17 targets that nations in the world are called to achieve by 2030.



(a) HDI versus energy consumption [1]



(b) Energy consumption versus GDP [2]

Among them, the 7th Goal, Affordable and Clean Energy, is the one most explicitly referring to energy and it is in turn composed by five interrelated targets [3]:

1. ensure universal access to affordable, reliable and modern energy services;
2. increase substantially the share of renewable energy in the global energy mix;
3. double the global rate of improvement in energy efficiency;
4. enhance international cooperation to facilitate access to clean energy research and technology, including renewable energy, energy efficiency and advanced and cleaner fossil-fuel technology, and promote investment in energy infrastructure and clean energy technology;
5. expand infrastructure and upgrade technology for supplying modern and sustainable energy services for all in developing countries, in particular least developed countries, small island developing States, and landlocked developing countries.

The first subgoal (SDG 7.1) is divided into the topic of access to electricity and the one related to access to clean cooking solutions. The present work focuses on the first of the two, aiming to propose solutions that could facilitate its achievement. Electricity access is defined as follows:

Definition 1 *Electricity access entails a household having initial access to sufficient electricity to power a basic bundle of energy services – at a minimum, several lightbulbs, phone charging, a radio and potentially a fan or television – with the level of service capable of growing over time. "pico solar" products, mainly solar lanterns which may include mobile phone chargers, are considered to be below the minimum threshold to count as having access. [4]*

The International Energy Agency (IEA), given its role as a custodian agency, is monitoring yearly the worldwide progress in the first three targets of SDG 7, providing reports, data and scenarios of evolution.
With respect to SDG 7.1, energy access, it has been found that in 2019 the number of people without electricity access had dropped to 770 million [5]. However, progress is not even across the world, South-East Asia and South America have almost reached 100% electricity access in 2019 while 75% of the population without access now lives in Sub-Saharan Africa, a share that has risen over recent years (figure 2.2). The IEA's Stated Policies Scenario,

that is the scenario considering the current framework of regulations, policies and plans, projects that in 2030 some 660 million people will still lack access to electricity. Given the expected population growth, about 940 million people will have to be connected by 2030 to reach universal access. Moreover, the COVID-19 crisis threatens progress in some parts of the world. In Sub-Saharan Africa, the number of people without access to electricity most likely grew in 2020 [6]. This means the access rate will have to more than triple between now and 2030. In Sub-Saharan Africa alone, this would mean connecting around 85 million people each year through 2030.

In South-East Asia, with an electricity access rate of 96% in 2019, Middle

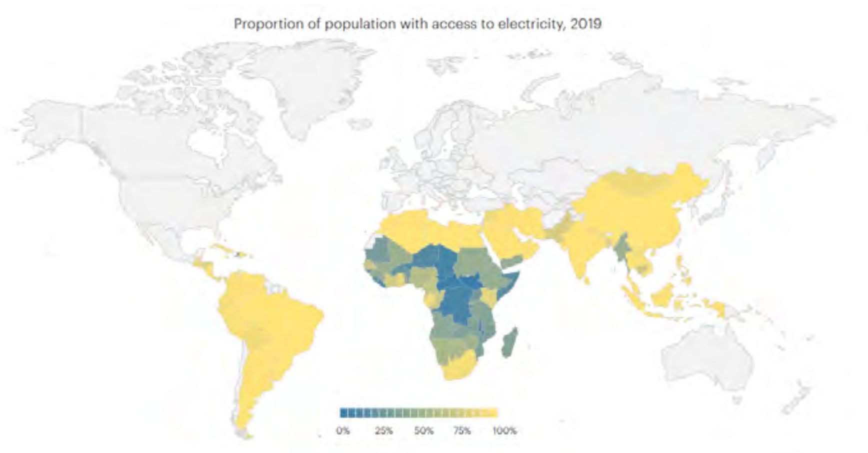


Figure 2.2: Proportion of population with access to electricity 2019 [7]

East of 92 %, and Central and South-America 97%, the main problem to be faced when dealing with access to electricity is the so called last-mile electrification [8], resulting in a struggle to reach the last, usually extremely isolated and under-served rural communities that still lack access to electricity. The extension of power grids in those areas is rarely economically justifiable and the choice of the electrification solutions adopted are often based on political rather than techno-economic motivations. On the contrary, Sub-Saharan Africa, with an electricity access rate of just 48% in 2019, still requires huge planning efforts to identify the optimal way forward. The following section is dedicated to analysing more in detail the issue of electricity access in Sub-Saharan Africa, highlighting virtuous countries that managed to significantly improve their condition during the past years.

2.2 Electricity access in Sub-Saharan Africa

Sub-Saharan Africa is the poorest region of the world with an average GDP per capita of approximately 1500 US$, not showing trend of growth since 2014. One of the issues that hinders economic growth is surely the lack of access to reliable and affordable electricity, which does not allow the proper development of the industrial sector [9].
Energy Sector Management Assistance Program (ESMAP), the global knowledge and technical assistance program administered by the World Bank, developed the Regulatory Indicators for Sustainable Energy (RISE) platform, a set of indicators intended for use in comparing the policy and regulatory frameworks that countries have put in place to support the achievement of SDG7 on universal access to clean and modern energy. Indicators are divided into the three pillars of energy efficiency, energy access and renewable energies [10]. Each country gets a score from 0 to 100 according to its average performance in the different fields. Data, reported in 2.3 demonstrate how countries in Sub-Saharan Africa still have room for improvement in building robust policies that could effectively support the energy sector. Kenya and South Africa are the only countries with a score comparable to the one of industrialized world, even though the Sub-Saharan region is seeing fast improvements during the last years. Kenya, Tanzania, and Chad had large improvements in their regulatory framework since 2017 with their RISE scores increasing of 9 points per year.

Figure 2.4 shows how the increase of electrification rate changed over the years in the Sub-Saharan countries. Negative values are given by years that showed a decrease in electrification rate. Kenya experienced the widest growth during the last 4 years going from 40 to more than 80 % electrification, followed by Rwanda with a 26% growth. Equatorial Guinea, on the opposite, started from a good electrification level in 2000, but remained stuck to a value of 65 %.

According to the IEA Stated Policy Scenario, the share of people with access to electricity will rise from 45 % to 65% in 2030, with a total number of people without access of 530 million [9].The countries that will be able to reach full access by 2030 are Kenya, Ethiopia, Rwanda and Senegal. Those countries have developed comprehensive electrification plans which entail a combination of different electrification strategies.

The Ethiopian government announced plans in its 2019 National Electrification Plan to connect 100% of households by 2025 by connecting to the

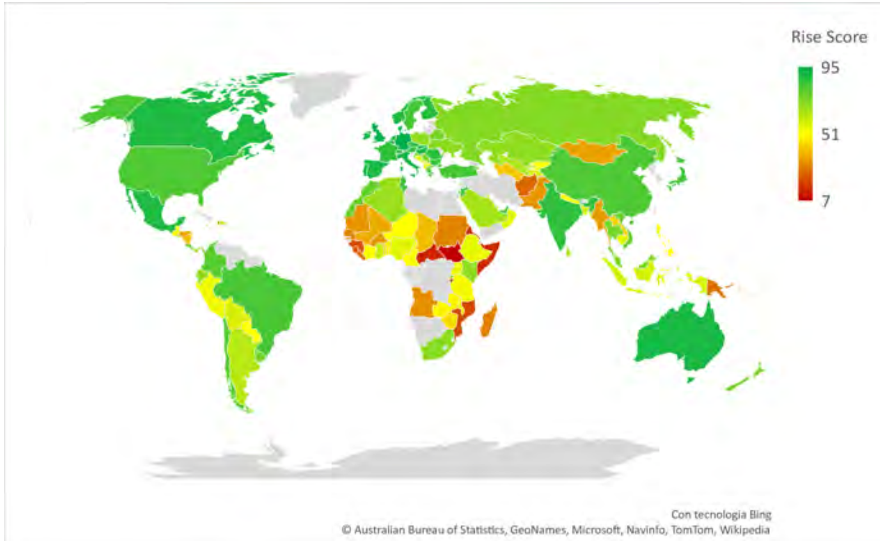


Figure 2.3: RISE scores in 2019. Author's elaboration from data of [10]

grid those 65% of households located less than 2.5 km from the existing network and putting in place decentralised solutions for the remaining 35%. As a second step, to be reached by 2030, the government plans to extend the grid to reach households located between 2.5 km and 25 km from the existing grid progressively substituting off-grid solutions with grid connectivity. By 2030, 96 % of access will be provided by the grid, 9.7 million people reached through low-cost grid densification and intensification, and about 7.8 million through the grid extension, with only a small fraction remaining with off-grid service. Connection priority will be given to social institutions (schools and clinics) and to areas with high potential for economic growth [11].
As for Rwanda, the Energy Sector Strategic Plan launched in 2018 aims to bring electricity access by 2024 to all public infrastructure, schools, health facilities, small businesses and administrative offices, in addition to households. Moreover, new roads will be provided with street lightning and reliability of electricity supply will be improved [12]. According to the Rural Electrification Strategy of 2016 both centralized and decentralized solutions will be needed to reach full electrification: 48% of households will have access to electricity by 2024 through off-grid systems and the remaining 52% will be connected to the grid following the Energy Access Roll-out Program (EARP) [13].

The Senegal National Rural Electrification Program (PNER), launched in

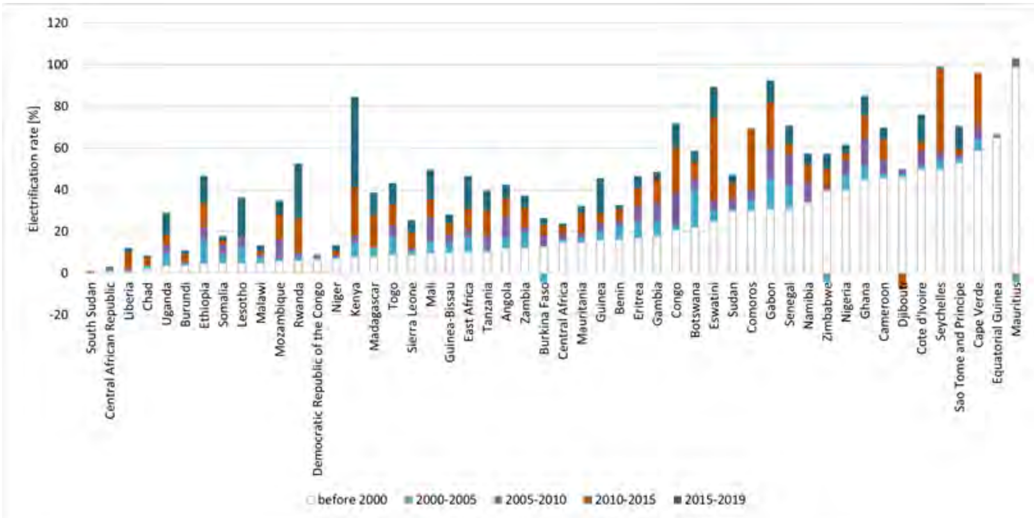


Figure 2.4: Electricity access rate in different countries of Sub-Saharan Africa

2015 was composed by a first Urgency Program , targetting 60% of electrification rate by 2017 and a second phase aimed to reach full electricity access by 2025. According to the Senegal action Agenda and electrification prospectus, by 2025, 14.234 rural villages and more than one million rural clients should be electrified mostly through grid extension (12.556 villages and 95% of rural clients), but also through solar only or solar-diesel hybrid mini-grids (1.215 villages and 4% of rural clients) [14]. Rural electrification - which will represent an additional generation capacity requirement of 180 MW by 2025 – will come from the "ramification" of the planned Medium Voltage backbone and from an ambitious mini-grid program in the Eastern part of the country [15]. Individual Solar Home Systems will electrify only 464 settlements representing less than 1% of rural clients.

Kenya case is particularly interesting because of the huge progresses it made during the past few years. A strong push to electrification rate was given by the Last Mile Connectivity Program. It started from a study on barriers to electrification in rural Kenya in 2014 [16] that underlined how electrification rates where surprisingly low despite previous significant investments in grid infrastructure. The study identified that half of the unconnected households were located in close proximity to the grid, or clustered within just 200 m of a low-voltage power line, where connecting to the grid should be relatively low-cost. However, connection cost was not affordable by the majority of residents in low income areas since it constituted the 40 % of their annual income. Within the Last Mile Connectivity Program, the connection fee was reduced to one third for a selection of communities located within 600 m from the grid in a first phase and 2km in a second phase. With respect to decentralized systems, the Kenya Off-grid Solar Access Project aims to distribute 250 000 solar home systems to power households, schools, health facilities and agriculture by 2030. According to the Kenya National Electrification Strategy, launched in 2018 universal access to electricity can be reached by 2022 by 269,000 connections to the grid through grid expansion, 2.77 million connections to the grid through grid intensification and densification (including 100,000 connections through intensification of existing mini-grids), 35,000 connections through 121 new mini-grids to serve housing clusters too distant from the network or too small to be connected to the national grid, 1.96 million connections through stand-alone solar home systems [17].

Those four countries show the relevance of developing strategies that include both decentralized and centralized systems for reaching universal access to electricity in few years. Speaking more broadly, around 15 million people are now connected to minigrids in Africa [18], while the number of people gaining access through solar home systems in sub-Saharan Africa increased from two million in 2016 (IEA, 2017a) to almost five million in 2018 (IEA analysis based on sales data provided by the Global Off-grid Lighting Association). There is no solution that fits all, and each countries' and regions peculiarities should be carefully evaluated to promote effective electrification strategies. In the next chapters, the different aspects that influence the rural electrification planning problem and the approaches adopted in literature are presented and discussed.

Chapter 3

Overview of rural electrification planning

3.1 Introduction

What clearly emerges from the analysis of the previous chapter, is the importance of reaching global electrification by creating hybrid energy systems, composed of multiple actors and operators (actually this is the trend also of industrialized countries where energy communities and distributed generators are gaining space and importance aside centralized systems). The two main strategies for electrification are [19], [20]:

- Centralized grid expansion: the traditional electrification strategy is the extension of bulk power systems by means of transmission and distribution grids. It is usually a governments' duty; it can ensure low specific energy costs but implies huge capital expenditure. Big investments could be necessary to guarantee the development of the grid in vast urban or rural settlements.

- Off grid systems: they usually rely on locally available renewable resources which could result considerably expensive in the mid-long term scenario. They can be in turn subdivided into:

  - Stand-Alone Systems (SA): those are small systems, ranging from few watts of power, able to supply just lightbulbs and phone chargers (e.g. solar kits or Solar Home System (SHS)), to some kW, to supply single households or small enterprises, typically with diesel generators or roof-mounted Photovoltaic (PV) modules.

- Microgrids: they differ from SA because they include also the distribution grid and can supply bigger loads, entire communities or big industries. According to the CIGRE WG6.22 definition [21]:

  Definition 2 *Microgrids are electricity distribution systems containing loads and distributed energy resources (such as distributed generators, storage devices, or controllable loads,) that can be operated in a controlled, coordinated way either while connected to the main power network or while islanded.*

  Size of microgrids could range from few kW to some MW of installed power.

While those two solutions are traditionally seen as complementary, they should be considered as processes that need to be properly coordinated to manage both the rural electrification need in a short time perspective and the energy growth of vast regions in the long period [20]. There is no optimal solution fitting all the contexts as exemplified in figure 3.1: the Levelized Cost of Electricity (LCOE) of different technological solutions and their convenience with respect to the others is a function of communities' characteristics, such as population, distance from the grid, complexity of terrain, etc. Nowadays in the world, 47 million people are connected to 19000 minigrids, mostly hydro and diesel powered, while 7500 additional minigrids for 27 million people more are planned, mainly in Africa and based on solar hybrid systems [18]. First estimates state that from a least cost analysis, the total number of minigrids to be installed to reach universal access to electricity by 2030, is 210.000, for a total of half a billion people served.

National electrification plans based on quantitative analysis are the most effective instrument used by governments to set effective electrification roadmaps. A proper electrification plan should consider several complementary aspects to propose the optimal context-specific solution as visually represented in 3.2. In the next sections the different dimensions are described in detail according to the following classification, proposed also in [23]: techno-economical framework, related to the state of the art technologies and costs, social dimension, related to energy demand satisfaction and socio-economic development indicators, geographic dimension, for the assessment of local resources and territory peculiarities, environmental impact and regulatory framework.

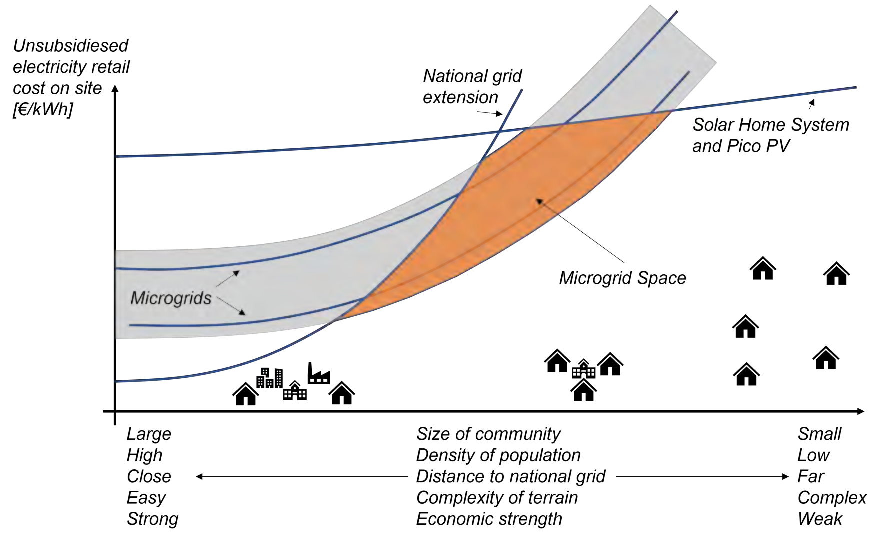


Figure 3.1: Convenience of different electrification solutions with respect to communities' characteristics [22].

3.2 Techno-economical framework

Electrification plans must cope with the costs and maturity level of different technologies. With respect to grid expansion, aside conventional three phase and single phase systems, there are other low cost technologies that could be used specifically for rural electrification, as for instance:

- Single wire earth return (SWER): it is a particular case of single-phase systems that uses only one energized conductor and the earth as a return ground wire [24]. It has been used to supply power to rural loads worldwide due to its reduced investment and maintenance costs, as well as a fast and simple deployment of lines. Different variations have been proposed and used in many countries, some of them presenting an isolating transformer between the SWER line and the three-phase system. In the direct connection variation (without isolating transformer), each single-phase load is connected to a single conductor of the main branch of the three-phase medium voltage grid. In this case the current flowing through the ground will be reflected to the distribution system, which may cause imbalances, reliability loss and safety issues if not properly planned. This issue is aggravated by the fact that in the direct connection SWER the sensitive earth protection cannot be implemented,

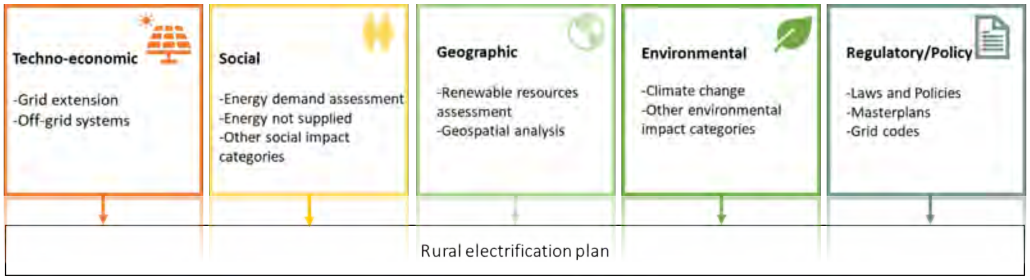


Figure 3.2: Visual representation of dimensions related to rural electrification planning

since there is a permanent nominal current flowing through the ground. For this reason the grounding system, the soil resistivity and humidity will strongly influence the applicability and efficiency of this type of configuration.

- Shielded Wire Systems (SWS): it consists on insulating the shielded wires from the towers of high voltage transmission lines for a medium voltage operation, and energize these conductors to supply the loads. In the most common configuration, it requires two shielded wires that will be energized at medium voltage and a ground wire as the third conductor. Benefit of using a SWS is to make a better use of an existent high voltage transmission line, using its shielded wires instead of building a complete new medium voltage grid. By having this dual use of the shield wire, the cost of subtransmission towers, conductors, grounding mats can be completely avoided. It was first proposed in the 1980s to be deployed in Ghana, with a SWS system that had about 526km of 161kV lines, serving up to ten thousand households and it has been in commercial operation for over 15 years [25]. The unique characteristics of this type of electrification, which requires specifically located HV lines, limit its use to very specific situations.

When speaking about Renewable Energy Sources (RES) based off-grid systems, there has been a sharp decreasing trend in prices during the last 10 years that is making them competitive with traditional sources even without the help of subsidies (e.g. residential PV modules' cost decreased up to 80% from 2010 to 2019 [26]). Analysis and long term scenarios should be updated to account for this fast changing sector.

Also, microgrids could be designed according to different criteria:

- Grid connected or isolated: If grid connected, the microgrid will be a part of the distribution systems which means that the reliability, quality of service and security of the microgrid will have the same standards as the conventional system. However it does not provide much of cost reduction, since the grid has to be deployed nonetheless and microgrids requires special attention at their protection systems to avoid damages that can be caused by reasons such as internal faults or unwanted islanding. This results in better protection equipment that can concur in more costs. Isolated microgrids can offer cheap energy where the grid connection might be impossible or expensive. The main challenge is to create a reliable system that supplies the local loads with good quality of service, since all the regulation, protection, and demand control will be provided by the microgrid itself.
- With or without a DC bus available: Given the diffusion of Direct Current (DC) based generators, such as PV modules and Battery Energy Storage System (BESS) and DC loads, in recent studies, more and more microgrids are being equipped with a DC bus available for connection or in some cases without a AC bus at all [27], [28]. DC microgrids could be potentially more beneficial than AC since they simplify the need to synchronize generators, they reduce the use of converters, they facilitate the connection of various types of Distributed Energy Resource (DER) and loads to the microgrid common bus with simplified interfaces, and reduce losses associated with the AC-DC energy conversion [29]. However, immature power protection system, higher investment cost and low compatibility with AC motors, make their usage in most cases limited.

In addition to costs of different technological options, also their reliability, the facility of maintenance and operation, the local availability of replacement pieces could discriminate one solution with respect to others. Interrelated to the social dimension is the quality of energy supplied and the quantity of energy not supplied. National grids in Sub-Saharan Africa suffer from instability problems, frequent interruptions and blackouts, thus in some cases grid reinforcements and densification should be preferred to grid extension [9].

3.3 Social dimension

Electrification plans should answer to the needs of the population living in the beneficiary area, represented in first instance by the energy demand, but also by other socio-economical indicators of development, such as opportunities for job creation or equitable distribution of income between genders and poverty classes [30]. Actually electricity access and rural development are interlinked by bidirectional casual relations characterized by complex dynamics feedbacks and behaviours (e.g. electricity access would lead to the increase of production activities that would lead to an increase in energy demand) [31]. Only when interventions are carefully planned to take into account all dynamics and social aspects of communities they can lead to the virtuous cycle of development.

The energy needs depend on many socio-economical factors, size of communities, their level of growth, their proximity to other communities and the types of activities developed. When dealing with greenfield planning, load forecasting faces the difficulty of estimating the latent energy demand, that is the load consumption once electricity becomes available, not constant but dynamically changing over time. Willis in [32] divides the process of electric demand growth into four steps:

- Immediate jump: some customers connect immediately as soon as the grid arrives, acquiring some electric appliances; richer customers or social infrastructures may already own their generators and will continue use their electric appliances connecting to the grid;
- Rapid growth: during time, new customers may connect and new appliances and equipment could be bought by the population;
- Economic growth: thanks to electrification, new born economic activities may rise requiring further electricity; increased income levels will give the possibility to costumers to buy new appliances and consume more electricity;
- in-migration: the increased attractiveness of the area will lead to an increase of population coming from the neighbouring non electrified communities.

It becomes evident that electric demand estimation is a challenging task due to many uncertainties related to present and future scenarios. A correctly sized energy system should be able to power the three main classes of users: households, social infrastructures and productive uses [33]. The literature concerning the nexus between energy and rural development indicates that access to energy, when it is supported by complementary activities – e.g. educational activities, capacity building and awareness campaigns –,

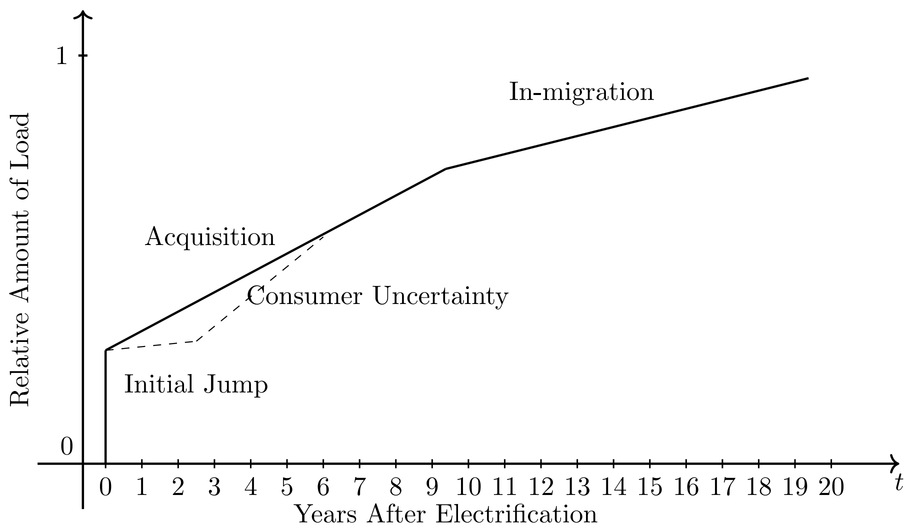


Figure 3.3: Generalized load growth behaviour of an area after electrification [32]

can be a pivotal driver in developing new business [34], with a consequent increase in the industrial energy demand. It must be recognized that not all the electrification solutions provide the same potential for demand expansion. SA have a minimal capability of guaranteeing development possibilities to the beneficiaries, since the technology is usually sufficient to power very small appliances such as lightbulbs and phone chargers for a single household, completely leaving out the possibility for more complex appliances or productive uses of electricity, fundamental for unleashing development in rural areas [34]. Lee et al. in [35] assess how Kenyan households that benefit from electricity access through SHS, and hence to the basic end uses of electricity, actually would aspire to own additional appliances with energy consumption that could not be satisfied by the SHS, resulting in a limitation of their potential development. As widely discussed by Amartya Sen in [36], limited access to electricity facilities would result in limited capability for households to start new productive activities and improving their economic and social status, making the technology itself a limit to the potential of development of the beneficiaries. Even microgrids, when not properly sized, could limit the expansion of communities and would require continuous investments to keep pace with increasing energy needs. Centralized grid extension is usually the most robust technology able to face different degrees of development. However, low energy quality, frequent interruptions and blackouts may hinder the development of big businesses and industries [20]. The amount of energy not supplied is a widespread indicator of performance of different energy systems. Other important social aspects are the opportunities for job creation, the forecasted acceptability of the technologies by the communities, affordability of different solutions, the improvement of health (that could be for example given by the limitation of local dangerous emissions) and education conditions [30, 37, 38, 39].

3.4 Geographical context

The optimal electrification solution depends on the peculiarity of the territory where it should be deployed.
Firstly, RES availability differs around the world and can drive different types of development. Noticeable is the example of Indonesia, where thanks to the availability of hydro resource, since 1990 more than 1300 micro-hydropower projects have been built, benefiting the same number of rural villages [40]. The convenience of one technology with respect to other depends roughly on the capacity factor (cf), the ratio between the energy produced by a plant during one year (E_p), depending on the availability of primary resources, and the maximum producible energy:

$$cf = E_p/P_{nom}/8760 \tag{3.1}$$

From this it derives the necessity of correctly estimating the renewable resources' potential and of having the ability to forecast the power production in a short and long term perspective. This allows for instance to evaluate if there are weak aspects in the existent electrical grid that need to be reinforced, or which would be the optimal size of new generation facilities. The renewable resource potential, could be distinguished into [41]:

- Theoretical Potential: it is the maximum theoretical renewable potential that arrives to the Earth, i.e. the Earth energy flow;
- Geographical Potential: it is the renewable potential related to a specific location;
- Technical Potential: it is the power extractable with the available technologies;
- Economical Potential: it is the technical potential that can be realized at a given financial cost level.

When speaking about energy system planning, the most important indicator that could help evaluating the upper bound for future energy supply is the technical potential. In particular, it is important to properly model power production from renewable potential in terms of time series data, representative of one year of power production. Uncertainties related to RES availability forecast may also influence sizing of systems and move to solutions that could be more or less conservative according to the degree of confidence (e.g. backup diesel generators could be needed in case of wind-based minigrid systems).

Secondly, terrain morphology, the land coverage, the difficulty in reaching isolated communities could favour either centralized or decentralized solutions. Philippines, given their peculiar territory, extending over a multitude of islands, are among the countries with highest number of installed micro-grids (896) and total capacity installed (397 MW) in the world [18]. Several Caribbean islands are also nowadays powered by off-grid microgrids.

All these types of geographical information could be gathered and processed by means of GIS:

Definition 3 *A GIS is a computers-based system to aid in the collection, maintenance, storage, analysis, output, and distribution of spatial data and information [42].*

Speaking in simple terms, geospatial data are the modern way of mapping. They provide information over vast areas in forms of vector or raster layers, that could be overlapped to derive complete information related to physical and political aspects of territories (e.g. elevation, land cover, solar irradiation) [42]. Raster GIS layers are usually created starting from remote sensing images, taken from satellites and aerial photography, while vector data creation is helped by the use of GPS systems. Given the improvement of those technologies in the recent years, high resolution geospatial data are becoming widespread on online open-source databases. The use of geospatial analysis is nowadays recognized as one of the important pillars of rural electrification planning [18].

3.5 Environmental dimension

An optimal solution for electrification should consider also its environmental footprint. The 7th SDG not only talks about access to energy but more specifically relates to access to clean energy. What is clean and what not could be measured according to different impact categories [30, 43, 38]:

- Climate change impact: this is usually measured through the greenhouse gas emissions or the equivalent CO2 emissions in atmosphere. They are subdivided into direct emissions, only related to the electricity production process, and indirect emissions, related to the life cycle of the technologies (comprising production and transport of the materials);
- Wastes production: the disposal of some technologies, especially the ones with limited lifetime, may constitute a huge problem for the environment. This is why research efforts are now focusing on end-of-life management of different components [44];
- Land use: the land surface required to produce energy with a specific source could be critical when planning extensive fields (according to literature data PV modules could occupy an area ranging between 4.5 and 7 m^2/kW against only 2.35 m^2/unit for diesel generators [45, 39]);
- Resources depletion: both traditional and RES based technologies imply the use of primary resources. In the first case the major impact is related to the non renewable fossil fuels extraction, while wind, PV and BESS make use of a critical amount of minerals and rare earths (see fig. 3.4)[46].

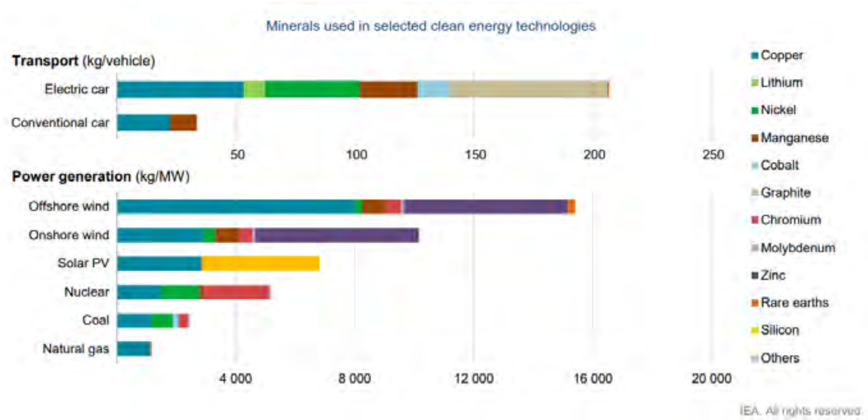


Figure 3.4: Minerals used in selected clean energy technologies [46]

New approaches are interestingly relating energy planning with water and food consumption in what is called the Water-Energy-Food nexus [47]. As stated in [48] energy needs water, water needs energy: the dependencies in both directions are set to intensify rapidly. In turn, the availability of water affects the viability of energy projects and must be considered when deciding on energy options. And the dependence of water services on the availability of energy will impact the ability to provide clean drinking water and sanitation services. Food is also interrelated to water and energy: it is enough to think about water and energy needs for agricultural production and energy production from agricultural residues. New approaches consider integrated planning at, for instance, the river basin level to guarantee the sustainability of projects.

3.6 Policy and Regulatory Framework

Rural electrification plans are inserted and have to cope with the political and regulatory framework of the countries where they should be adopted. Convenience and risks of investments change according to the in place regulations that could, to different extents, benefit some technologies with respect to others. The RISE platform by ESMAP identifies several indicators of positive enabling frameworks for grid extension, microgrids and stand-alone systems, which are here summarized in table 3.1. The scores associated to each of those indicators for countries in Sub-Saharan Africa are shown in figure 3.5, with values ranging from 0 (lowest performance), to 100 (best performance).

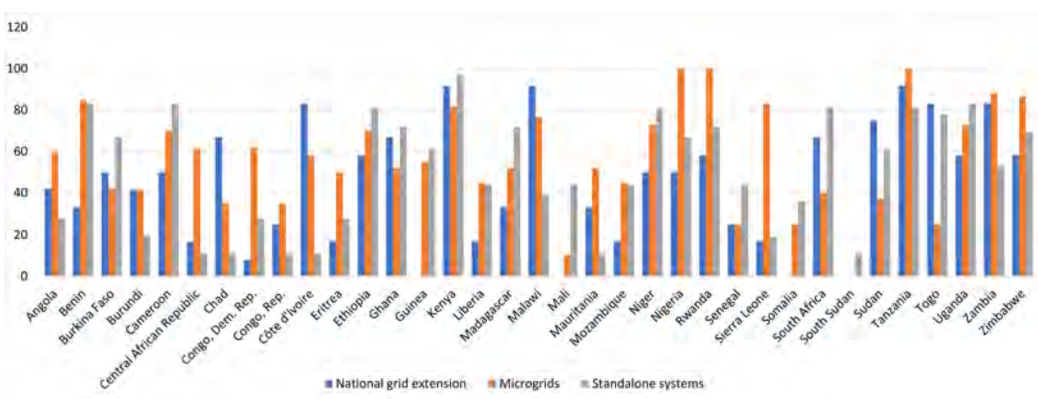


Figure 3.5: RISE scores in Sub-Saharan Africa countries. Author's elaboration from 2.3

It is interesting noticing that from this study countries seem most prepared to electrify through microgrids, that has the highest average score of 57 rather than standalone systems and grid extension. Nigeria, Rwanda and

Table 3.1: RISE indicators related to enabling regulatory framework of grid electrification, minigrids and standalone systems

Framework for centralized grid electrification
Dedicated government funding line or budget for electrification (e.g., funded national program,
budget item, rural electrification fund to finance grid extension);
Capital subsidies paid to the utilities to provide distribution systems to rural areas/villages;
Existence of consumer financing mechanisms (i.e. utility loans, on bill financing, micro-loans etc.);
and/or direct subsidies available to support the payment of connection fees by consumers;
Existence of grid quality standards (e.g. number of guaranteed hours per day, duration of the electricity, frequency of outages, etc.);
Framework for minigrids
Legal allowance of microgrid operation;
Dedicated programs for microgrid development;
Clear regulations related to the interaction between in place microgrids and main grid arrival;
Private ownership allowance;
Clear licencing procedures for microgrid operators and consumers;
Legal allowance to minigrid operators to charge a cost-reflective tariff, even more expensive than national tariff;
Publicly funded mechanisms to secure viability gap funding for operators;
Duty exemptions and/or capital subsidies for minigrid systems and/or individual components;
Specific financing facilities (access to credit etc.) available to support operators;
Existence of technical standards for minigrid connection to the grid;
Existence of safety standards for microgrids(e.g. overcurrent protection, system control, etc.)
Framework for stand-alone systems
Dedicated programs for stand-alone systems development
Duty exemptions and/or subsidies to support stand-alone home systems
Specific financing facilities available to support operators/consumers to develop/ purchase stand-alone home systems
Adoption of international quality standards for stand-alone systems
Environmental regulations on the disposal of solar devices and stand-alone home system products or components

Tanzania are the ones with the maximum score for microgrid framework. At the time of IRENA report [49] (2018), they were able to regulate the most important aspect of the microgrids sector. They had defined capacity thresholds for mini-grids (e.g., smaller than 100 kilowatt-peak, kWp) that are either exempted from licenses or undergo simplified processes, they deregulated tariffs for energy selling and promoted regulations for the operation of microgrids at the arrival of the main grid, allowing mini-grid operators to relocate assets, sell parts of their assets to the utility or become a small power producer selling electricity to the main grid at a fixed renewable feed-in tariff and/or become a distributor of electricity purchased from the main grid.

Aside energy laws, policies and masterplans, one of the fundamental pillars regulating the power sector is the grid code. Grid codes collect the set of rules that regulate power systems and energy markets operation and have the goal of ensuring operational stability, security of supply and well-functioning of wholesale markets, when present. A set of grid codes can include, for instance, connection codes, operating codes, planning codes, market codes [50]. Grid codes play a fundamental role also for regulating the integration of DER and microgrids into the national grids. The European set of grid codes, developed by the European Network of Transmission System Operators (ENTSO-E), recently included connection requirements fo distributed generators, in terms of voltage and frequency provision as well as minimum security standards. Sub-Saharan African countries still have poorly developed grid codes, that are often not even respected due to lack of control, and that lead to the instability and reliability issues explained before in the text. Electricity market in Sub-Saharan Africa is aggregated into the Eastern Africa Power Pool (EAPP), the Western Africa Power Pool (WAPP) and the Southern Africa Power Pool (SAPP). Those are cooperatives of electricity companies aimed to create common power grids and facilitate commerce, as well as to promulgate regional grid codes in a similar fashion to ENTSO-E. Those codes could set the technological requirements of different components, thus moving electrification investments plans towards specific directions. They could oblige inverter-based components to participate to frequency and voltage regulations but they could even guarantee revenues for the services provided. Finally, and this is something not yet developed even in industrialized countries' codes, they could regulate microgrids' connection to the grid ensuring that their interaction could be mutually beneficial. In this regard, the author, in collaboration with a PhD student working for WAPP, performed a study, published in [51], related to the possibility of integrating BESS in the WAPP power systems as providers of primary frequency control regulation. Different control strategies where proposed to effectively manage the BESS in a context with high amplitude frequency fluctuations and over-frequency bias given by too stringent load shedding strategies.

Chapter 4

Modeling framework in literature

4.1 Introduction

Scientific community is, since several years, active in studying solutions and providing instruments that could help stakeholders in the complex and multivariate task of rural electrification [52], [53]. The complete technical design of the new electrification strategy in rural areas should be composed by the following four points [54]:

1. Identification of optimal electrification solution, choosing between on- and off-grid solutions;
2. Sizing of generation portfolio for off-grid systems;
3. Design of electric network;
4. Eventual upstream reinforcements of the electric network and generation portfolio.

Literature is rich in tools and methodologies able to address one or more of the above aspects. They can be subdivided into different classes, according to their objective and scope. Above all, three major categories can be identified. On the one hand there are the tools whose goal is the optimization of energy systems from the generation side, defining the optimal mix of energy sources to be used to satisfy a certain demand. Those use energy models, energy flows and balances and could consider various energy vectors (e.g. heat and electricity) to represent the system under analysis. On the other hand, there are tools which use an electrical approach to design and optimize the electric power network, by means of laws typical of electrical engineering (e.g. power flow, short circuit studies). The third category is represented by tools aimed to the comprehensive electrification planning, that try to optimize at the same time both the generation and the distribution infrastructure (those will be called here on comprehensive models).

Within this chapter, the goal of the review is to describe the scope of the models, their input and output variables and ability to address the different dimensions of rural electrification planning, rather than provide details on the mathematical approaches adopted. Details on algorithms and optimization strategies will instead be provided in part II.

4.2 Review of energy models

Local perspective: off-grid systems sizing

In the literature, there is a vast range of models that are specifically designed to find the optimal combination of generation units, usually composed by RES, storage systems and diesel generators, for supplying villages or communities: HOMER [55], Microgrids.py [56], iHOGA [57], DER-CAM [58], are examples among others. All of them are focused on small systems and use as input forecasted load profiles and resources availability of the analysed area, which allow to optimize the size of hybrid microgrids, in terms of the combination of generation sources. The models for the design of off-grid systems differentiate in terms of algorithms used, technologies addressed, accuracy in the modeling of technological components, and ability to address peculiar features of rural electrification, such as uncertainties and long term planning [59],[60].

The most well-known and diffuse tool for hybrid microgrid design is the software HOMER, developed by NREL lab. It is a commercial software, able to size a hybrid microgrid composed by several types of generation sources and eventually evaluate the connection to the in place grid; it is often used as a benchmark for new developments in literature. Environmental and social aspects are included as constraints in terms of minimum fraction of energy produced by RES or maximum amount of energy not provided [55].

The Distributed Energy Resources Customer Adoption Model (DER-CAM) is a decision support tool for investment and planning of decentralized energy resources in buildings or microgrids. It was developed at the Lawrence

Berkeley National Laboratory, USA in 2000. It is a pure optimization tool whose objective function to be minimized is the total energy cost and/or CO2 emissions, such that energy balance is preserved, technologies operate within physical boundaries and financial constraints are verified. It considers both electrical, heating and cooling needs of costumers for one typical year, considering a cost for load curtailment. It allows in addition to model the distribution grid and assess the voltage and current constraints [58].

Improved Hybrid Optimization Genetic Algorithm(iHOGA) is a hybrid system optimization software tool developed by the University of Zaragoza, Spain. It can perform multi-year and multiobjective optimization, including aside the costs dimension, the equivalent CO2 emissions minimization, the unmet load, the HDI and jobs creation. In addition to electricity flows it considers water consumption and hydrogen vector [57].

Micro-Grids py is a free library of tools for the simulation and optimization of micro-grids, developed by University Of Liege, Energy Systems Research unit, Belgium together with University Mayor of San Simon, Energy Center, Bolivia; it was designed to find the optimal sizing of Lion-Ion batteries, diesel generators and PV panels in order to supply a demand with the lowest cost possible but thanks to its flexible design it can include also other resources [56].

Generally speaking, this category of tools provides a detailed modeling of different energy production technologies and is able to account for load growth scenarios and availability of RES. These tools are also able to include the climate-change related environmental dimension in form of constraints or multiobjective optimization. It is clear that they solve however only one part of the rural electrification problem, national grid is just considered as an additional energy supply source that could be used in case of connection, and should be complemented by other instruments to create a comprehensive electrification plan.

Multi-scale perspective: energy system models

A second category of tools includes modeling frameworks which perform the optimization of energy systems at different scales, from small communities to entire nations. Many of these tools are open source and freely available online and are collected under the OpenMod initiative (`https://openmod-initiative.org/`). Those tools are not specifically designed for rural electrification projects, but, thanks to their flexibility, can be easily adapted to different case studies. Calliope [61], Osemosys [62], Oemof [63], are among the most recognized models. They include the geospatial dimension of the system, which allows to have a simplified nodal representation of the electric network; multi energy flows are simulated in order to optimize the size and location of energy generators.

OSeMOSYS is an open source modelling system for long-run integrated assessment and energy planning. It has been employed to develop energy systems models from the scale of continents down to the scale of countries, regions and villages in more than 100 literature works [64]. The first version of OSeMOSYS was made available in 2008 by the KTH Royal Institute of Technology, Sweden. OSeMOSYS is typically used for the analysis of energy systems looking over the medium (10-15 years) and long (50-100 years) term. As an example, with OSeMOSYS the Temba (The Electricity Model Base for Africa) model was developed, representing each continental African country electricity supply system and transmission links between them. The detail of input data is low, it considers annual capacity factors for RES and does not allow to distinguish between on and off-grid systems. With respect to rural electrification, Osemosys as been linked to other tools to have a complete vision of the problem: in [34] Osemosys is coupled with a tool for bottom-up modeling of energy demand while in [65] with Onsset, the large scale planning tool that will be described in the next section.

The Open Energy Modelling Framework (oemof) is a modular, free and open source tool for modelling and analysing multi-scale energy systems developed by the Reiner Lemoin Institute (RLI), Germany. It includes and links the heat, power and mobility sector and describes energy systems with a graph based approach, composed of nodes with generation sources and loads and edges representing electric lines or thermal energy flows. The typical optimization is the dispatch optimization but it also provides a combined dispatch and investment optimization. The solver optimises for minimal costs that can be economic, environmental, technical or any other type of cost thanks to the flexible approach. It has been applied to several purposes, from the evaluation of RES and storage in power system to the complete sizing of microgrids [66], [67]. In this last case, a new model for hybrid microgrid optimization, micrOgridS [68], was developed, that however results to be much more simplified than the models previously described, specifically designed for off-grid systems sizing.

Calliope is a free and open-source tool developed by the ETH Zürich. With a graph approach similar to the other tools, it allows to build energy system models at scales ranging from urban districts to entire continents. A Calliope model consists of a collection of text files that fully define a model, with details on technologies, locations, resource potentials. The number and types of technologies and resources depend on the user settings. The investment planning mode identifies also which power lines should be installed and which areas should remain off-grid. This could allow to have a preliminary result of a hybrid electrification plan, however, to the authors' knowledge, in literature no specific application for rural electrification planning is available [61].

To summarize, multi-scale energy models are characterized by a high flexibility, and the types of simulations depend strongly on the user's capability of setting input parameters. They could be used for off-grid system sizing, but in this case losing the technical accuracy of models specifically designed for this purpose. They could also be used for a rough estimation of grid expansion, but in this case they are more suited for transmission grid capacity expansion rather than distribution grids in rural areas. Social and environmental impacts are generally not considered and neither are geospatial terrain characteristics.

Large scale perspective: energy modeling of countries

Finally, there exist tools that allow to have a holistic perspective on the problem of electrification and are suited for large scale pre-feasibility studies. Relying on geospatial data analysis and extremely simplified assumptions, they provide an overview of the optimal electrification solution within entire continents. Onsset [69], IntiGIS [70], RE2nAF [71], divide the area of interest into cells or regular size, or clusters in the last version of Onsset, with a desired resolution (e.g. 1km per 1km). For each cell, data related to estimated energy consumption and renewable resources availability, collected from online open-source GIS databases, are combined to define the most convenient electrification solution. Raster maps of nations or continents are provided as output, where to each cell is associated the most convenient energy solution, choosing among grid connection, SA or microgrids; grid connection cost is usually proportional to the distance to the in place grid. Those tools do not perform optimization or sizing of systems but simply compute the least cost solution, in terms of LCOE, according to techno-economical input parameters.

Onsset is the most well known tool in this category, it was developed by KTH and is being extensively used by academia and international organizations. In [9], the IEA report related to energy in the African country, Onsset has been extensively used to have a geospatial estimation of the least-cost pathway to universal access to electricity, providing detailed analysis of 44 countries. The Global Energy platform by ESMAP shows the results of Onsset over 58 countries around the world.

With regard to the four points related to rural electrification planning described at the beginning of the chapter, this category of tools is designed to address the identification of the optimal electrification solution and one of its major points of strength is the analysis of the geographical context thanks to the extensive use and combination of GIS layers.

4.3 Review of electric models

All energy models are focused on the energy supply aspect of the electrification and therefore they usually have a very simplified representation of the local distribution grid, often underestimating its costs. Electrical models, on the other hand, are more prone to find the best electric grid design, as a complement to the optimization of generation resources. Usually, the electric grid design is a Distribution System Operator (DSO)s' task and it fits an already defined energy plan. In literature, different approaches have been developed in order to solve the grid design problem that is complex and composed by several different interconnecting aspects, that can be summarized as [72] [73] [74]: (i) Secondary substations siting and sizing; (ii) Low Voltage lines routing; (iii) Primary substations siting and sizing; (iv) Medium voltage lines routing.

Strategies and models used differ according to the phase of development planners have to face. In most of the cases, especially in the industrialized world, grid does not have to be planned from scratch (what is called greenfield planning), but new investments are directed towards the reinforcement of an existing grid already in place (brownfield planning). Given the peculiarity of rural electrification problem, that involves the extension of the grid to non electrified areas, the focus of this section will be related to algorithms and models for greenfield planning. Broadly speaking, the goal of these models is to minimize the total investment and operational costs of the network, taking into account voltage and current constraints, and eventually reliability requirements. Only the peak load demand and the location of users are used as input, environmental issues are only seldom considered and social impact is considered by reliability indicators. Literature is rich in works devoted to propose innovative models that use different optimization algorithms [75, 76, 77] to solve specific aspects of distribution network planning. Models can be subdivided into static (that look only at a specific instant of time) or dynamic or multi-stage, that plan the grid during different future time intervals or according to their ability of including DER planning. The accuracy of these models comes at the expense of the limited dimension of the problem to be addressed; moreover, the majority of these works focus on the grid expansion of local urban areas in industrialized countries rather than of the electrification of wide rural areas of entire countries [54].

An interesting tool is the the Reference Network Model (RNM), a very large-scale planning tool, designed by the Instituto de Investigation Tecnologica of Comillas University [78]. It designs the high, medium and low voltage networks, planning both substations and feeders. In doing that it makes considerations and assumption about technical aspects such as voltage limits, capacity constraints, continuity of supply and geographical constraints such as forbidden ways through (such as lakes or nature reserves) and a street map.

4.4 Review of comprehensive models

Energy and electrical modeling approaches are traditionally seen as complementary, with the optimal design of electric network subordinated to an already defined energy plan. However, recently, some works are starting to merge the two approaches, to have a realistic representation of power systems: the Reference Electrification Model (REM) [79], the open source Network Planner [80], the commercial software GeoSim [81] and LAPER [82] are tools which simultaneously design the energy system and the electrical network. The authors in [83] although not developing a software tool, also present a complete procedure for rural electrification planning. In this case, the works presented constitute, according to the author's knowledge, the state of the art of this models' category and no other relevant works are present in literature.

REM is a computer model developed by the MIT and Comillas Universal

Energy Access Laboratory. Its purpose is to support large-scale electrification planning and local electrification projects (LREM). The inputs for the model are information about building locations, solar irradiance, topography, grid extent and reliability, expected consumer demand, fuel costs and infrastructure costs. After running a series of clustering and optimization algorithms specifically designed for electrification planning, REM produces lowest-cost system designs. Individual consumers are grouped into electrification clusters so that total system costs (actual and social costs) are minimized. These clusters may denote groups of customers to be connected to separate mini-grid systems, groups to be connected to the existing grid, or clusters of single customers to be supplied with stand-alone systems. The RNM previously described is used in REM to design the distribution grids. REM provides as output: (i) the optimal groupings of individual consumers into electrification clusters so that total system costs are minimized; (ii) the optimal generation mix and network layout for each of the off-grid mini-grids, selecting the size of diesel generators, PV modules and BESS;(iii) the optimal network layout for each cluster that will be connected to the grid. Social cost of energy not supplied is included in the cost function while RES usage and maximum emissions could be inserted as constraints.

The GEOSIM platform is a modular commercial software tool based on GIS. Its main innovation consists on the optimisation of energy services covering a given territory, within a given time horizon, with a view to improving the economic and social impact of rural electrification. Consequently, as it is based on the logic of land-use management, GEOSIM is initially used to select and hierarchically arrange the localities according to their own dynamism and impact on neighbouring localities (introduction of the concepts of Development Poles and hinterlands). Next, once demand forecasting has been completed within the planning horizon, the various electricity supply options (including connection to the power grid and decentralised solution such as hydroelectricity, biomass PV modules and diesel generators) are obtained with a technical-economic optimisation. REM and Geosim represent the state-of-art of rural electrification planning and they have been used as support tools for the redaction of national electrification plans in Subsaharan Africa ([84], [85]).

Logiciel d'Aide à la Planification d'Électrification Rurale (LAPER) determines the best electrification mode for a set of villages or settlements minimizing total investment and operation expenses. It uses all available geographical data, creates villages-types in order to be able to use standardised electrical equipment for electrification, and compares the costs of all possible solutions of electrification (i.e. mini-grids, diesel gensets, solar panels, small hydro- or wind generators). LAPER determines the villages which would economically benefit from being connected to the power grid and those for which a decentralised method of electrification is preferable. After this computation of the "target" solution, LAPER determines the master plan based on given annual budgets and various non-technical criteria (political, environmental...) which influence the order in which villages will be electrified. The planner may then draw what he considers as the most suitable power grid to which the maximum number of villages can be connected. The user may choose between several kinds of lines, since the main line is automatically set as a three phased one, and the secondary lines as single phased ones. The software then carries out an electric check of the designed grid.

Network Planner is an online tool developed by the Columbia university for planning grid, mini-grid, and off-grid electricity from the community scale to national scale. Network planner takes a host of inputs, including geospatial population distribution, costs of energy technologies, electricity demand and population grow-fluidity, and existing grid network, and output the least-cost solution. The Network Planner is a decision support tool for exploring costs of different electrification technology options in non electrified communities. The model combines data on electricity demands and costs with population and other socio-economic data to compute detailed demand estimates for all communities in a dataset. Then, the model computes cost projections of three electrification options and proposes the most cost-effective option for electrifying communities within a specified time horizon. These three potential electrification options include: (i) off grid: PV for households supplemented by a diesel generator for productive use; (ii) Mini-grid defined supplied by diesel generator with Low Voltage (LV) distribution for all demands type; (iii) Grid connection, considering only linear distances.

Finally, the work of Blechinger et al. [83] applies a geospatial-based procedure for the rural electrification planning of five Nigerian Federal States. It performs a multi-stage planning considering grid connection, hybrid micro-grids (composed of PV modules, diesel and batteries), and SHS. Grid routing is based on geospatial criteria but no electrical simulation is performed.

To conclude, comprehensive models are able to address three out of the four steps of rural electrification planning: identification of optimal electrification solution, optimal sizing of off-grid systems and design of electric network. None of the analyzed tools addresses the problem of eventual upstream reinforcements. Their combination of electric and energy modeling make them complete and promising for addressing the issue of energy access even though the large amount of input data required and the complex modeling frameworks could make their usage limited.

4.5 Literature outcomes

Table 4.1 and figure 4.1 summarize the main concepts and considerations drawn during the literature review. The spider diagrams show the ability of

Table 4.1: Summary of rural electrification models categories

Category	**Objective**	**Input**	**Output**	**Limitations**	**References**
			Energy Models		
Local	Off-grid systems sizing	Load and RES estimated profiles; costs.	Optimal generation mix able to supply the load.	Low scalability; no grid topology.	[55];[57]; [68]; [56]
Multi-scale	Energy systems design	Load and RES estimated profiles for each node of the system; costs.	Optimal generation mix and grid connections.	No electrical analysis and simplified grid representation.	[62],[61],[63]
Large scale	Pre-feasibility studies of large areas	Geospatial average data of load and resources; costs.	Optimal electrification solution for each cell of the grid	Simplified assumptions; no optimization	[70]; [69]; [86]
			Electrical Models		
	Optimal design of electrical grids	Location and estimated peak power of loads; costs.	Optimal grid topology (feeders/substations)	Low scalability; no generators sizing.	[75]; [76]; [87], [77], [73], [74]
			Comprehensive Models		
	Energy system design in terms of generation units and electric network	Location of loads, estimated load profile, RES availability; costs.	Design of the optimal solution for electrification. Grid topology and generators sizing.	High input data requirement	[79]; [80]; [81], [83]

the different tools (subdivided by category) of assessing the four dimensions of rural electrification problem: 1 techno-economical, 2 social, 3 geographical, 4 environmental. Given the difficulty of inserting the regulatory framework in optimization tools, this dimension is not here considered. Scores are given according to qualitative criteria summarized in table 4.2. Arrows indicate the ability of each category of tools of addressing the different steps of a rural electrification plan, according to the categories reported at the beginning of the chapter taken from [54].

Comprehensive models are the most complete ones, being able to address the majority of problems and dimensions. However, only few works are present in literature and they all present some limitations:

- Most rural electrification planning tools focus solely on economic parameters when designing the best rural electrification plan for a given region. Some of these indirectly consider other factors into the global cost, such as a penalty for NSE. However, there are no tools that perform multiobjective optimization (for instance, considering cost, emissions, or NSE) allowing them to analyze their trade-offs

- The majority of the tools consider limited options for hybrid microgrids, mainly composed by diesel generators, PV modules and batteries;

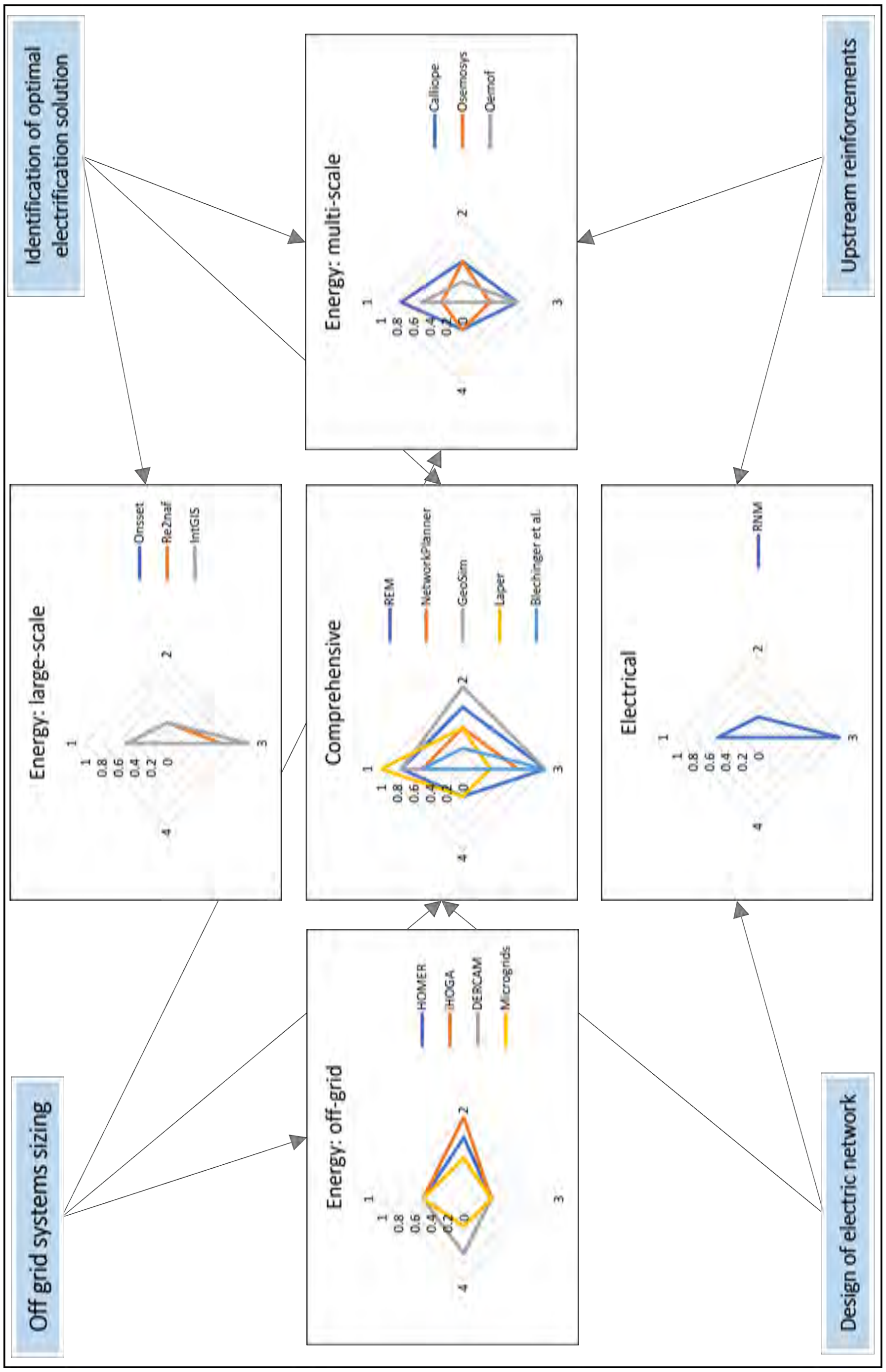
- Electric network design is simplified or neglected in the majority of the works. An exception is represented by the RNM used in REM. However, this tool was created for developed countries and not specifically adapted to the rural electrification planning problem;

- The majority of the tools are proprietary softwares and a complete open-source modeling framework for rural electrification planning has not yet been developed.

Table 4.2: Scores associated to rural electrification planning dimensions

Dimension	Score	Dimension	Score
Technological		**Geographical**	
Off-grid options	2	RES availability	1
Grid extension	2	Other	2
Tot	4	Tot	3
Social		**Environmental**	
Energy demand	1	Climate change impact	2
Demand evolution	1	Other	1
Social impact	2	Tot	3
Tot	4		

According to the findings of the literature analysis, the author decided to focus her research activities on the development of a new comprehensive procedure for rural electrification planning, aimed to push a small step forward energy-access related literature and to provide an open-source software shared with the community. The following two parts of the manuscript are devoted to describe in detail the procedure and provide examples of application.



Dimensions: 1: techno-economical 2: social 3: geographical 4: environmental

Figure 4.1: Visual representation of the tools

Part II

Proposed modeling framework for rural electrification planning

Chapter 5

Case studies

A major difficulty when performing studies for energy access and rural electrification planning is related to the data gathering process. Data are often incomplete and not reliable; few metering systems, able to provide quantitative information, are in place and only scattered information related to historical trends is available.
During the years of her doctoral studies, the author had the opportunity to work in collaboration with other students, companies and NGOs and address specific aspects of rural electrification and energy access in Sub-Saharan Africa and South America. Through these collaborations it was possible to gather data and develop and validate different procedures on realistic case studies.Those works had as outcome peer-reviewed papers, Msc thesis works and project deliverables; they are described in the following paragraphs and summarized in table 5.1:

1. Network model in Tanzania: the author was co-supervisor of a MSc thesis work aimed to model the transmission network of Tanzania and analyze the hosting capacity for distributed generators in the city of Dar el Salaam. The thesis student spent a period in Tanzania gathering data related to in place and planned transmission lines and to the distribution grid in the city.

2. Rural electrification planning in Mozambique: the author has been co-supervisor of a MSc thesis work whose goal was the rural electrification planning of the non-electrified area of Namanjavira in Mozambique; one of the students was working on the field collaborating with the NGO COSV and was able to gather data about load demand and costs of different technologies. The study lead to the publication of [88].

3. Network model in Colombia: the author was co-supervisor of a MSc thesis work aimed to study Colombia power network. Within this work data related to the Colombia power system were gathered and some analysis on the power flows and possible grid extensions were performed.

4. Rural electrification planning in Brazil: the author was co-supervisor of a MSc thesis work in cooperation with Enel infrastructure and networks, the DSO of the Cavalcante region. The collaboration allowed to gather data related to the in place distribution grid and to costs and technical characteristics of new lines expansion and to perform an analysis on the optimal grid extension topology in the area of Cavalcante; in addition to the thesis manuscript, a peer-reviewed paper was published: [89].

5. Rural electrification planning in Bolivia: within a collaboration with a NGO working in the country, a study aimed to assess how to electrify households in the rural area of Omereque, discriminating between grid connection and SA was performed. The study was published in the paper [90]. The NGO provided data related to the inplace distribution grid and the coordinates of the non-electrified households.

6. WAPP grid code: the author collaborated with the Italian company Cesi within a project for the drafting of the WAPP grid code. The collaboration focused on the identification of possible requirements for the integration of DER and micrograms in the grid.

7. Rural electrification planning in Lesotho: the author worked in cooperation with a MSc student living in Lesotho and working for the local DSO aiming to perform a study for the distribution grid expansion in the area of Butha-Buthe. The study resulted very interesting for the amount of data gathered relative to the in place grid topology, costs and technical characteristics of grid components, population and location of villages.

8. Microgrids pre-feasibility studies in Sudan: the author is participating, together with the Cesi company, to a consultancy activity for a company in Sudan. The goal is to gather data related to several villages in the country and to size the optimal hybrid microgrids that could supply the loads. Wind, solar and biomass resources are taken into account. Through local surveys, accurate data related to villages composition and availability of resources and possible load demand are gathered.

9. BESS in WAPP region: through a collaboration with a PhD student working for the WAPP system, a study on the possible performance

of BESS for primary frequency regulation in the system has been performed. Data related to measured frequency trends in the region and to the connected generators where gathered. The output of the study has been published in [51].

Table 5.1: Summary of case studies

Topic	Country	Type of cooperation	Year	Outcome
Hosting capacity and network model	Tanzania	Msc student	2018	Msc thesis
Rural electrification planning	Mozambique	Msc students/NGO COSV	2019	[88]
Network model and grid routing	Colombia	Msc students	2019	Msc thesis
Rural electrification planning	Brazil	Msc students/Enel	2019	[89]
Rural electrification planning	Bolivia	NGO Luces Nuevas	2020	[90]
Regulatory frameworks	West Africa	Cesi	2021	Section of WAPP grid code
Rural electrification planning	Lesotho	Msc students/DSO	2021	Msc thesis
Microgrids	Sudan	Cesi	2021	Proj. deliverables
BESS for Primary Frequency Regulation (PFR)	WAPP	Phd student	2021	[51]

The case studies of the rural areas of Namanjavira in Mozambique, Butha-Buthe in Lesotho, Cavalcante in Brazil and Omereque in Bolivia are the ones that were deepened the most during the years and that were used specifically to develop a new comprehensive procedure for rural electrification planning. The procedure has the goal of design the optimal electrification solution in rural areas, choosing between off-grid and grid connected systems. In the following section an insight over the selected case studies is provided.

5.1 Mozambique-Namanjavira

Country overview

Mozambique is a Sub-Saharan country with a surface of 801.590 km^2 and a population of around 30 millions of inhabitants, a growth rate of 2.5 % per year and an average population density of 36 inhabitants per km^2; 65 % of the population lives in rural areas. To date, Mozambique is one of the poorest countries in the world, with about 54.7 % of the population living below the poverty line, as shown by the low value of the country's Human Development Index (0.418 in 2016) which put it in the 181st place out of 188 nations. National access to electricity was 35 % in 2019, significantly increased from the 6% of 2007 but still far from SDGs' targets and with high disparities between rural areas (with electrification rate of 22%) and urban areas (with electrification rate of 57%) [9].

The main actor of the energy sector is MIREME (Ministerio dos Recursos Minerais e Energia), which is responsible for national energy planning and policy formulation as well as for overseeing operation and development of the sectos. The energy regulatory authority (ARENE) is the governmental consultative which works as a regulatory instrument concerning generation, transmission and sale of electricity. In 1997, the Government established the FUNAE (Fundo de Energia), an administratively and financially autonomous public institution with the role of supporting and developing the management of energy resources, being responsible for the off-grid electrification

field. It is complementary to EDM (Energia de Mocambique), the National Grid operator, a vertically-integrated, government-owned electric utility responsible for generation, transmission and distribution of electricity in the national grid.

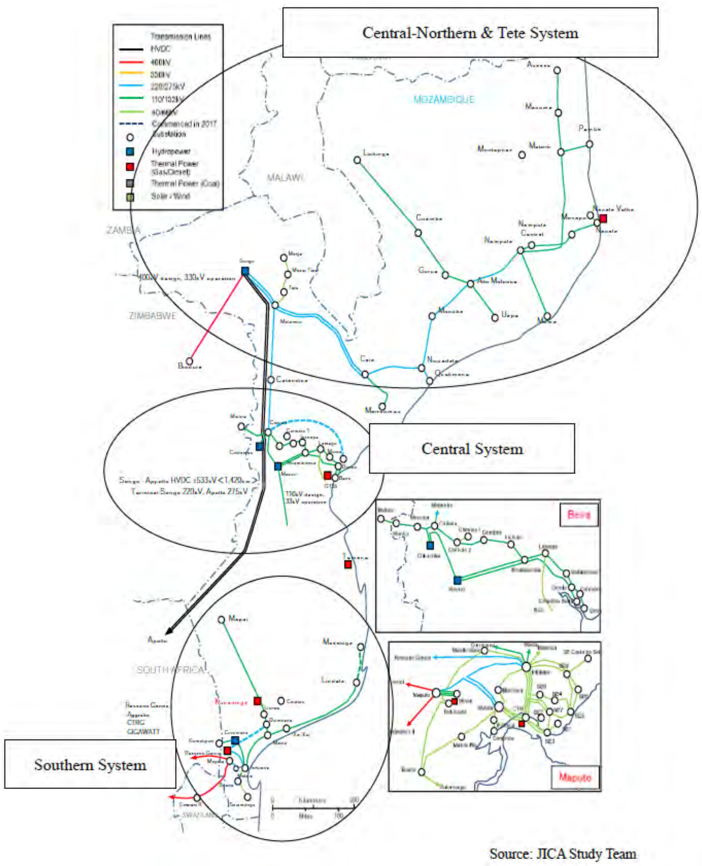


Figure 5.1: Power System of Mozambique

The structure of Mozambican power system, composed by a northern, central and southern part, is shown in 5.1. Currently, the southern transmission network is not connected to the central and northern parts that are instead connected among them through 220 kV lines. The energy regulatory framework in Mozambique is composed by several laws and policies. The Integrated Power Sector Master Plan from EDM is the main policy for Mozambique. It has a set goal to increase installed capacity to 6,001 MW by 2030 and 20% integration of renewable energy in the grid. The institutional, financial and technical approach to reach these goals is defined by the National Electrification Strategy. Within the document, the target of complete electrification by 2030, task to be accomplished by FUNAE, is also set. Although there is a strong will on the part of the Mozambican government to electrify the rural areas of its national territory, the prohibitive costs estimated for the extension of the rural electric network clearly show the importance of investing, especially in the short term, in off-grid systems, based on the abundant renewable energy resources present in the country. Investments still struggle to be implemented mainly due to a legislation that is unclear and at times adverse to the private sector involvement in the energy sector: Mozambique is the African state with the lowest investments in renewable energies, being 2.2 million dollars between 2009 and 2014. Consequently, FUNAE has so far exercised the role of implementing agency, rather than supporting third-part's rural electrification initiatives. For instance, when internal funds allowed it, it built several off-grid micro-grids. In the first time, diesel generators were the preferred technology: due to their low investment costs they allowed to electrify a relatively high number of communities. The systems were strongly subsidized, energy was available for few-hours a day and consumers had to pay a fixed monthly fee far from being close to its cost-reflective value. In the long run, subsidies were removed and due to the unsustainability for consumers even in only purchasing the fuel, most of these systems failed. Consequently, due to critics from the communities, FUNAE abandoned diesel generators and switched to renewable energy and, in particular, to solar-PV technology. It started building some small microgrids with a capacity generally around 5 to 10kWp feeding a few dozen homes and public functions. High investment costs obviously reduced the replicability of the systems, but low annual costs allowed these systems to be financially self-sustaining when considering the investment as non-repayable. FUNAE drafted the "Atlas of Renewable Energies" and the "National Portfolio of renewable water and solar energy projects" with the aim of creating the conditions for the advent of foreign investments. They are the outcome of a two-year long country-wide mapping initiative, in which renewable energy resources and electrification priorities in terms of socio-economic strategic value were mapped, and consist of a long list of pre feasibility assessments of potential off-grid projects (microgrids, and SA systems), with locations, suggested technologies, estimated available powers and strategic values. Even with the existence of the Atlas, foreign investments did not have an appreciable increase and the cause can mainly be identified in the energy sector regulatory framework. To date the Mozambican Government, as a result of repeated interlocutions with international stakeholders decided to launch a vast program of reforms in the sector, resumed in the "Nova Lei da Energia" (New Electricity Law or NEL), a new law that is expected to disrupt the energy sector and hopefully to drive Mozambique towards an energy-driven sustainable development.

Area under study

The area analysed is the Posto Administrativo of Namanjavira, in the province of Zambezia. The Zambezia Province concentrates many of the country's criticalities in terms of development and energy. It has an area of 103.478 km^2, it is the second most populated and it is one of the most densely populated provinces of Mozambique (48.7 inhabitants/km^2). Its population growth rate is also particularly high considering a 35 % increase over the last 10 years. The population pyramid shows that 64 % of the population is under 24 years old. Zambezia is the most relevant province both from political consensus and development goals perspectives since it hosts one fifth of the country population, the 93% of which is living in rural areas, characterized by a very low resilience in the face of climate change and external shocks and where 70.5% of the population lives below the poverty line. Energy access rate is also among the lowest in the country, together with the provinces of Cabo Delgado and Niassa. Therefore, most of the government strategic efforts are concentrated on it.
A researcher of the group worked for two years on an electrification project in Namanjavira area (the Italian Cooperation for Development Agency (AICS)'s ILUMINA Program implemented by the Italian NGO COSV), so a collaboration has been set up with the implementing agency, to gather data and test and validate the different procedures for rural electrification planning, described in the following chapter. Different characteristics made Namanjavira a suitable case study for the developed model:

- the area is not electrified and it is hence suitable for greenfield planning; moreover it is a rural region in Sub-Saharan Africa, area where efforts for rural electrification planning should be concentrated;

- even though data related to the in place distribution grid in the country are not available, a transmission grid crossing the area is planned for construction. This provides a useful indication of the possible position of primary substations, from where the new planned distribution grid could depart;
- Namanjavira area has a contained dimension, which allows to test different procedures with increasing degrees of complexity, testing their suitability in terms of computational error and memory requirements;
- thanks to the collaboration with COSV, reliable on-field data related to communities' characteristics, possible load demand and different technologies' costs were gathered.

Figure 5.2: Namanjavira district in Mozambique

5.2 Lesotho-Butha Buthe district

Country overview

Lesotho is an African enclaved country, entirely surrounded by South Africa. It has an estimated population of around 2 million people with a total area of just over 30 000 km^2. In 2020 the 70% of the population was living in rural areas, however with a decreasing trend of 5% every 10 years. According to IEA data on energy access, in 2019 only 36% of population had access to electricity, with a discrepancy between rural, with 26 % of population having access and urban areas, with 63 % [9].

The main authority of the energy sector is the Ministry of Energy and Meteorology and in particular its the Department of Energy. In 2004 it was established the Rural Electrification Unit as part of the Department of Energy, with the goal of addressing electrification issues within the country. Until now its main effort has been on grid extension and generally very little on off-grid solutions. The Lesotho Electricity Company (LEC) is a government-owned electricity company responsible for the electricity networks and the electricity customer interface (connections, billing and payment). The regulator of the energy sector is the Lesotho Electricity and Water Authority, with the main duties of licencing, tariff approval, monitoring and conflicts resolution.

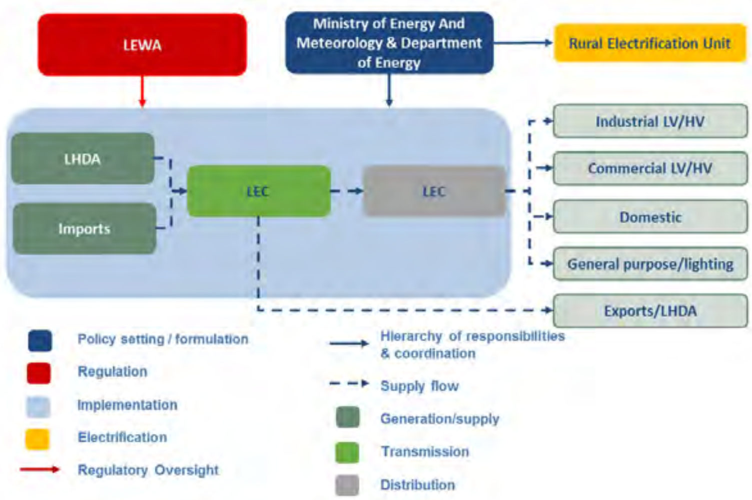


Figure 5.3: Structure of Lesotho power sector

The Lesotho Electricity Company Limited (LEC) is responsible for the transmission and distribution of electricity in Lesotho, operating under the direction of the National Rural Electrification Fund (NREF). The company has a portfolio of assets located around the country, composed of transmission and distribution lines operating at 132, 88, 66, 33 kV and 11kV, aside some places where they operate at 33kV as well as 45 High Voltage (HV)/Medium Voltage (MV) substations (see figure 5.4).

In 2018, the Government of Lesotho published a revised electrification plan, named the Lesotho Electrification Master Plan (EMP) which was drafted taking into account the country's development objectives contained in the *Na-*

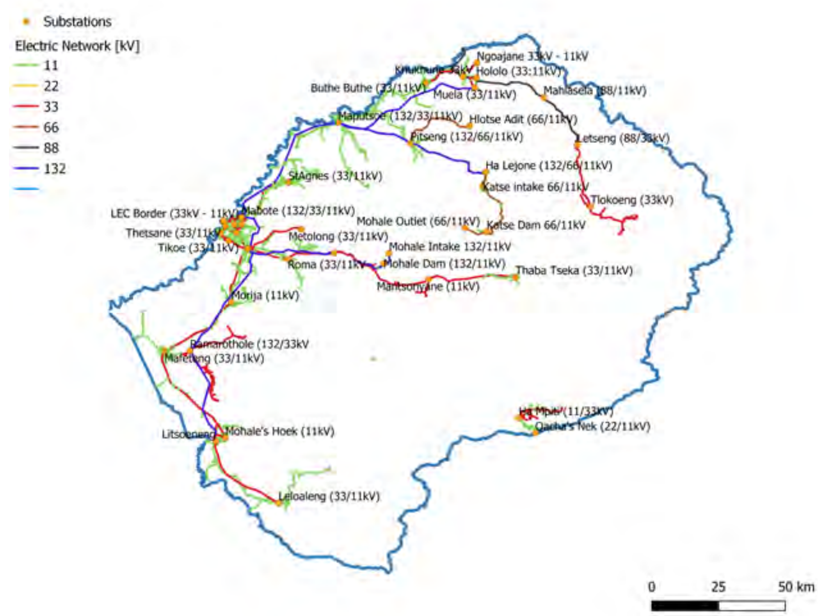


Figure 5.4: LEC power grid

tional Strategic Development Plan 2012/13-2016/17, *Lesotho Energy Policy 2015-2025* and *Lesotho Vision 2020* [91]. The primary aim of the plan is to improve electricity access in the country. The EMP found that grid extension would continue to play an important role in achieving the access target as it is the least-cost supply solution for roughly 64% of the total population, while off-grid solutions (primarily microgrids) would be least-cost for 36% of Lesotho's population. A portion of the plan therefore focuses on grid extension, but there is also a large component which focuses on the establishment of microgrids. Geographic zones in Lesotho have been classified into three groups: (i) to be electrified through grid extension; (ii) to be electrified through microgrids; (iii) to be electrified through SHS. The EMP also includes an off-grid development plan focused primarily on rural electrification, particularly those areas of Lesotho that are not easily reached by the national power grid. In terms of the annual electrification budget committed by the Government, 80% is allocated to grid electrification while only the remaining 20% will be allocated to off-grid electrification. Based on this budget, the Government estimated in its Off-Grid Master Plan Report that it would able to connect about 10,600 households to off-grid energy solutions (mainly solar lanterns and small SHS solutions) and 300 households to minigrids each year.

With the LEC's current plans and budget for electrification, as outlined in the EMP, focused on prioritising least-cost grid connections in high-density urban communities, rural access is likely to remain a challenge.

Area under study

The area analysed is the district of Butha-Buthe, one of the 10 districts of the country. Butha-Buthe has a surface of 1788 km^2 and a total population, from the census of 2016 of 118,242 inhabitants. The region is crossed by a transmission line at 132 kV and one at 88kV, however the distribution grid is still not completely developed and it mainly supplies the capital city Butha-Buthe and the surrounding villages. According to the EMP, only the village of Motete is suited for a microgrid installation, whose feasibility study is reported in the previous masterplan of 2007. A hydro turbine of 524 kW, able to supply not only Motete but also the nearby settlements of Kao and Liqhobong could be installed. The rest of the budget would be deployed for SHS and solar kits.
A MSc student working in Lesotho for the local DSO, contacted the research group in Politecnico to perform a distribution grid expansion study to reach 10 non-electrified villages in the area, considered priority villages in the EMP. He provided accurate techno-economical data related to the in place grid and population as well as indications related to the best practices of the DSO. Currently in the area more than 100 non electrified villages are present and within this manuscript the study is extended to explore also off-grid electrification options aside grid extension.
The area has been selected as a proper case study for testing the developed procedure for rural electrification planning due to several characteristics:

- it represents a further example, besides Mozambique, of a non-electrified rural area in Sub-Saharan Africa but with different geographical characteristics;
- availability of data, not publicly available, related to the in place distribution and transmission network topology, with lines and substations location, voltage levels and lines capacity;
- availability of data, not publicly available, related to villages' population and coordinates;
- accurate mapping of the area on open-source databases that provide an additional source of information.

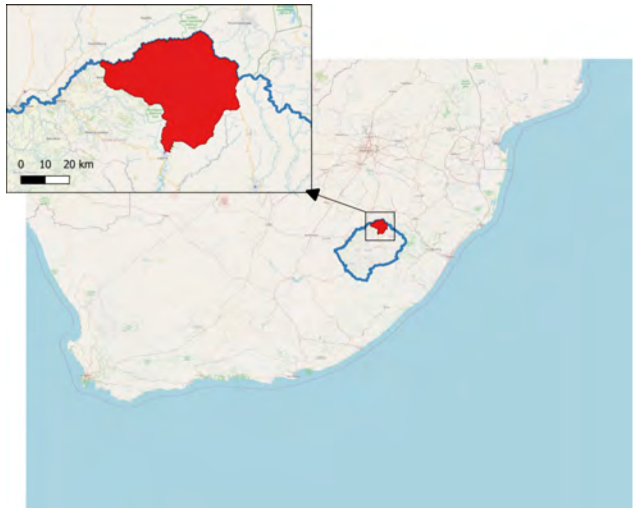


Figure 5.5: Butha-Buthe district in Lesotho

5.3 Brazil-Cavalcante

Country overview

Brazil is the world's fifth-largest country by area, covering 47.3% of the whole South America continent, and the sixth most populous with over 210 million people. 84.72% of the population lives in urban areas while 15.28% lives in rural areas according to the Brazilian Institute of Geography and Statistics [92]. The current country constitution, since 1988, divides it in 26 states from 5 different regions. Brazil is the largest national economy of South America and the world's ninth largest economy by nominal GDP and the eighth largest by Purchasing Power Parity (PPP) in 2019. Brazil is peculiar when it comes to the use of energy sources. The variety and quantity of natural resources allowed the country to have an energy mix based on RES since its conception, with the first hydro power plant built in 1883. In 2018 the yearly electricity production was 636.4 TWh, from which 83.3% coming from RES. The majority of Brazil's electricity production comes from hydro power (66.6%) with big power plants such as the Itaipu dam, built in 1982 to be the world biggest hydro power plant with installed capacity of 14 GW. The construction of these large dams cause a huge environmental impact, specially due to the reservoir that can ruin river ecosystems, cover large areas of land causing green house gas emissions from underwater rotting vegetation and displace thousands of people affecting their livelihood. For this reason the country has been experimenting a boom of alternative renewable energy in the last decades, mainly wind and photovoltaic generation. The wind power specifically had a growth in yearly energy production from 662 GWh in 2007 to 48,475 GWh in 2018 [93]. Due to reliable RES-based energy mix, Brazil has one of the lowest CO2 emissions associated with electricity production, having an yearly emission rate of 2 tCO2 per inhabitant in comparison to 6.2 tCO2 in the European Union and 15 tCO2 in USA.

Being a country of continental dimensions, the electrification process in Brazil was a hard and long process. Coming off a 21 years dictatorship that lasted until the end of the 80's, the financial crisis that the country was living started the privatization process of the energy sector at the 90's. In 1996 the unbundling of the whole energy sector happened, from generation to distribution, allowing competition in the generation sector while maintaining the natural monopolies in transmission and distribution. It was also the birth of the Brazilian Electricity Regulatory Agency (ANEEL) that is an autarchy of the government and linked to the Ministry of Mines and Energy (MME). One of main goals of the agency is to promote a free market of commercialization and to regulate the National System Operator (ONS) which is a non-profit entity responsible for coordinating the generation and transmission of the electric power system. However after the liberalization of energy sector and creation of the electricity market, the power generation did not follow the increase in the electricity demand. From 1990 to 2000, the Brazil's installed capacity increased by 33% while the electricity demand of 49% [94]. Part of this demand increase is due to the country's rural electrification process that was also ongoing during this period, which ended up almost tripling the amount of household with access to electricity from 1980 to 2000 [95].

The first rural electrification programs charged for the installation of infrastructures (poles, cables, wires, power boxes, transformers), restricting access to the service provision and excluding families that did not have the economic resources to purchase materials and services [95]. This led to a social stratification process, a fundamental aspect to understand the Brazilian rural electrification context, affecting precisely the most impoverished rural families. In 1999 the federal government started the until-then biggest rural electrification program, with the objective to extend electricity access to more than 1 million households countrywide. This, together with the constant economic development of the country, generated the mismatch between installed capacity and electricity demand causing blackouts and the country's biggest electricity crisis in 2001. In 1999, the Brazilian Interconnected System (SIN) was created, it is a meshed transmission network which today supplies 25

out of 26 states through its more than 140 thousand kilometres of electric lines. The SIN is divided into transmission and sub-transmission networks. The transmission network is composed of HV and HVDC lines of 230 kV or higher, and is responsible for supplying the electricity demand of the biggest urban centres. The sub-transmission is connected to the former and is responsible for supplying smaller urban centres and large industrial consumers, with a voltage level of electric lines from 69 kV to 138 kV. The distribution system operates at, what is defined by ANEEL, medium voltage level from 1 kV to 69 kV and at low voltage for residential and smaller consumers at 110 V or 220V.
A new rural electrification program, the "Light for All" program, was issued in 2003, with the goal to provide electricity access to 2 million households by the end of 2008. From the rural families benefited by it, 90% had a total monthly wage less than 3 times the country minimum. In 2009 the initial target set by the program was met, however during this period the number of known households without electricity access rose to more than 3 million, extending the initial duration. The difficulties and the complexity of extending the electric grid to remote areas, as well as the high investment costs, caused the program to be constantly renewed to ensure that even the isolated systems were connected to the SIN. According to the Brazilian MME, by the end of April 2018 more than 16 million people in rural areas gained access to electricity through the policy, generating around 510 thousands jobs. Brazil went from 87.5% of the total population with access to electricity in 1990 to near 100% in 2017 [6]. Even though in percentage of population the vast majority has access to electricity, it is estimated that 207 thousand families are still to be reached [95]. For this reason the "Light for All" program is expected to last at least until 2022. One of the procedures to achieve that goal, is for the federal government to work with local DSO's to find solutions for rural electrification. As an example, with the issue of the normative resolution 493 the regulator authority allows DSO's to deploy microgrids and small isolated systems to provide electricity to remote communities . It was within this context that the company Enel S.p.A, being the DSO of four states in Brazil, reached Politecnico di Milano requesting a study for strategies of rural electrification, resulting in a master thesis work.

Area under study

Enel is the DSO of the state of Goiás, located in the Center-west region of Brazil, where the country's federal district and its capital Brasilia is situated.

The state has a total area of 340 km^2 and population of 6 million people, being the most populated state of its region. It is divided in 18 micro regions, each of them composed of many municipalities. One of them is the municipality of Cavalcante, the case study area, located in the northern part of the state of Goiás which has an estimated total population of 9709 inhabitants. It covers an area of 6953 km^2, resulting in a extremely low population density

Figure 5.6: Cavalcante municipality in Brazil

of approximately 1.4 people/km^2. Part of Cavalcante municipality is already connected to the distribution grid, either through a three-phase MV line or a single-phase MV line. However, even though the DSO is able to supply the biggest urban center in the area (located at Southeast), there are still remote places that could be electrified and families to be reached out by the distribution grid. This condition together with the incentives that Enel receives to improve rural electrification, motivated the request for collaboration. The case study was set up to test different algorithms for distribution expansion planning. However, the full procedure, that discriminates between on and off-grid technologies was not tested in the area since, given the problem of last mile electrification, microgrids are not considered a possible solution by the DSO. The specific features, that made it in any case a useful case study are:

- data related to the in place distribution grid topology, with details about three phase and single phase lines were provided by Enel;
- population in the area is very scattered and electric load requirements

are higher than the cases in Sub-Saharan Africa. This allowed to test different algorithms for grid expansion;

- the low voltage distribution grid does not have to be designed, since almost each agglomerate of households is reached by MV lines.

5.4 Bolivia-Omereque

Bolivia has an area of 1,098,581 km^2 and a population of 11,513,100 people, with an annual growth rate of 1.41% . In 2019, 31% of the population lived and worked in rural areas. Bolivia is one of the poorest countries of Latin America. While urban areas such as La Paz and Santa Cruz are modern cities with a relatively good supply of modern energy services, most Bolivia's rural areas are still experiencing a lack of most basic services, including reliable and affordable access to electricity and improved biomass cook stoves. The electrification ratio of Bolivia in 2019 was 93%, being 99% in urban and 79% in rural areas [6].

Electricity is nearly exclusively generated by private companies from hydropower (36,3%) and thermal power plants mainly based on gas (59,7%). The total installed capacity is 1645 MW, the installed capacity connected to the National Grid System (SIN) in 2011 was 1.31 GW and the contribution of renewable sources other than hydropower is almost negligible. 85% of the electricity was produced in the Sistema Interconectado Nacional (SIN - National Grid System), while 15% was produced in isolated systems (mainly diesel-driven generators) . The Bolivian electricity market is strictly unbundled into three fields: generation, transmission, distribution. One company is not allowed to work in more than one of this fields. However, there is an exception for off-grid systems. Most of the electricity companies have been nationalized. All mayor cities -except Tarija and Trinidad- are connected to the national grid. The transmission lines operate at 250, 115 and 69 kV while the MV lines in Bolivia operate at 44, 24.9 and 19.9 kV and the LV lines at 220 V. The typical transformer capacity is between 25 and 15 kVA [96].

Currently, there are two transmission companies in the SIN, Transportadora de Electricidad (TDE) and ISA Bolivia which runs 53% of the transmission network in Bolivia. In Bolivia, the seven existing distribution companies enjoy a geographic monopoly in their concession areas. The largest company (in terms of kWh sold) is Electropaz, majority-owned by Spain's Iberdrola. The second place is occupied by the Empresa de Luz y Fuerza Eléctrica Cochabamba (ELFEC), which was owned by the American PPL

Global until 2007; followed by the Rural Electrification Cooperative (CRE), which operates in the Department of Santa Cruz.

In some cases, especially in the high plateau, cooperatives and community organizations access the distribution companies' network and sell electricity to small rural communities. Sometimes, those are organized enterprises that provide the service to middle-size towns, but in most cases, they are small organisations that serve family communities. The Bolivian government's efforts to improve delivery of energy services to the poor have been quite intensive in recent years. First, the broad energy sector reform programme that comprised among others the privatisation of state utilities, was implemented in the mid-1990s. The reform improved the overall performance of the electricity sector and achieved important coverage gains in urban areas, connecting and providing access to the grid for about 98% of the urban population. The access rate in rural areas, however, has grown from 13.7% in 1997 to 46,6% in 2010 [90]. In 2002, the government of Bolivia developed an ambitious rural electrification plan (PLABER – Plan Bolivia de Electrificación Rural) to increase access to electricity in rural areas from 25% to 45% within five years. However, implementation of the plan has been slow due to the ongoing political and economic crisis.

Area under study

The study area is the municipality of Omereque, in Campero province, part of the district of Cochabamba, in central Bolivia. The municipality, distributed over an area of 800 km^2, consists of 11 communities without access to electricity, 137 households in total. The organization Luces Nuevas, a Spanish NGO working for reaching the objectives of SDG7 in the rural areas of Latin America, is present in the area and is conducting studies to assess strategies for granting access to electricity to the entire population of the municipality. The municipality is crossed by MV distribution lines. Given the characteristics of the area (proximity to the national grid and low density of population), the NGO Luces Nuevas considers only two options for guaranteeing access to electricity to the households: either connection to the national grid or electrification with SA. Within the research group, a study for identifying which households to interconnect and which to provide with SHS has been performed. In addition to this, the data provided resulted particularly interesting to test some specific modules of the rural electrification planning procedure. The favourable data characteristics included:

- Availability of the precise location of each non electrified household: this allowed to avoid assumptions based on open source population datasets;

- Topology of in place MV and LV distribution grids crossing the area;
- Costs of distribution lines and substations.

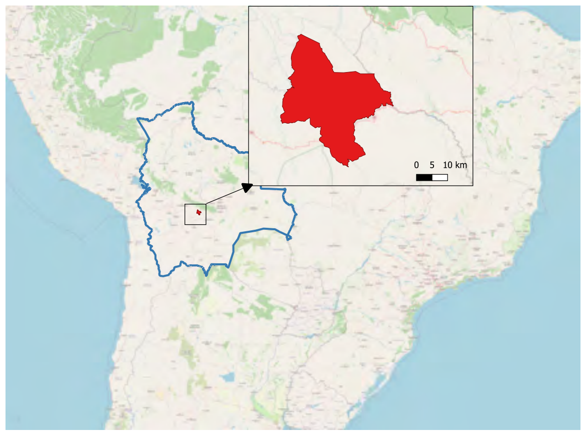


Figure 5.7: Omereque municipality in Bolivia

Chapter 6

Gisele: GIS for electrification

The present thesis aims to give a significant contribution to the literature models for rural electrification planning by creating a new open source and open access modelling framework, usable by different stakeholders, from university researchers, to NGOs and private companies. A novel procedure, named Gisele (GIS for Electrification) has been developed, coded and validated over the real-life case studies introduced in the previous chapter. The proposed approach has the following main characteristics:

- it is a comprehensive model, able to cover three out of the four steps necessary for rural electrification planning: (i) identification of the optimal electrification solution, choosing between on- and off-grid solutions; (ii) sizing of generation portfolio for off-grid systems; (iii) design of electric network;
- it covers all the different dimensions on which rural electrification depends:
    1. Techno-economical: the procedure models and compares different technologies for electrification: hybrid microgrids based on PV modules, wind turbines, hydro turbines, BESS and diesel generators and three-phase MV distribution grid.
    2. Social: electric needs of the rural communities are identified by a combination of bottom-up procedures, geospatial data and local surveys data; microgrid optimization allows to include multi-year planning and constraints on maximum energy not supplied; moreover, a multi-objective model including the minimization of total energy not supplied is developed.

3. Geographical: the whole procedure is based on the extensive use of geospatial data, in the form of raster and vector layers that provide information on the availability of RES, on the morphological characteristics of the territory, on distribution of the population and on the location of infrastructures.
4. Environmental: constraints on the minimum RES usage for off-grid systems as well as multi-objective optimization with an objective function related to emissions' minimization are included.

With respect to the state of the art of comprehensive tools, described in chapter 4, Gisele represents a valid alternative, with different functionalities and some improvements:

- the discrimination between on and off-grid solutions is based on optimization algorithms, rather than on greedy or heuristic considerations;
- it performs a detailed sizing of hybrid microgrids, considering multiple sources and multi-year planning;
- it performs a detailed routing and cost estimation of the distribution grid, able to consider the optimal path according to the geographic context;
- a multiobjective optimization is integrated, to assess some trade-off between environmental, technical and economical constraints;
- it is based on open source data and models; it is written in Python and freely available on GitHub; a first version of the tool is also provided with a Graphical User Interface (GUI) that makes it user friendly and usable also by non programmers;

The procedure is subdivided into different modules, each one devoted to solve a specific problem. Those modules can be categorized into four main blocks, which give the titles to the following sections:

1. Block 1-Energy demand assessment: starting from the boundaries of a non-electrified rural area, the block aims to subdivide the population into communities, each one to be electrified with the same strategy. In the figure 6.1, a visual representation of what is considered the *area* and what a *community* is shown: area is the non electrified region under study, while communities are the single energy systems that will be electrified in an homogeneous manner. In a second step, the electricity needs of each community are estimated.

2. Block 2-Off-grid system sizing: given the boundaries and load demand of the communities, and after the estimation of RES potential in the area, the optimal sizes and generation portfolio of hybrid microgrids able to supply the load in each community are identified.

3. Block 3-Internal grids design: the MV distribution grids able to interconnect users' within the communities are designed. The block designs one electric grid for each community.

4. Block 4-Integrated area optimization: this last block gathers all the outputs from the previous ones to perform an optimization on the whole area, design the grid interconnecting communities to the in place national grid and identify the communities to be supplied with off-grid systems.

A schematic diagram of Gisele modules and blocks, with their interrelations is reported in figure 6.2.

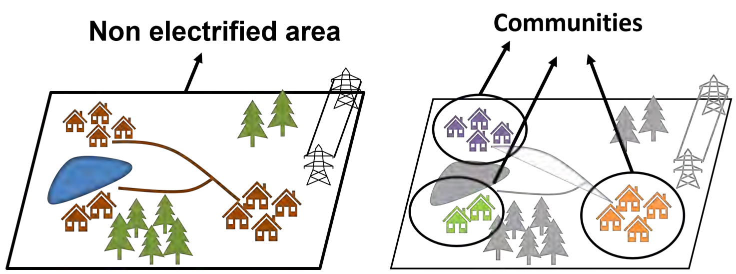


Figure 6.1: Schematic representation of non-electrified area and communities

The case studies reported in chapter 5 have been used to test the different procedures as specified in table 6.1. Namanjavira and Butha-Buthe regions have been used to test the whole procedure, as reported in chapters 7 and 8. Cavalcante case was useful to test the grid routing procedure, while Omereque for the secondary substation siting procedure.
In the following sections, all the four blocks of the proposed procedure are described in detail; grey insight boxes throughout the text are applicative examples that help demonstrating and clarifying some of the aspects of the models.

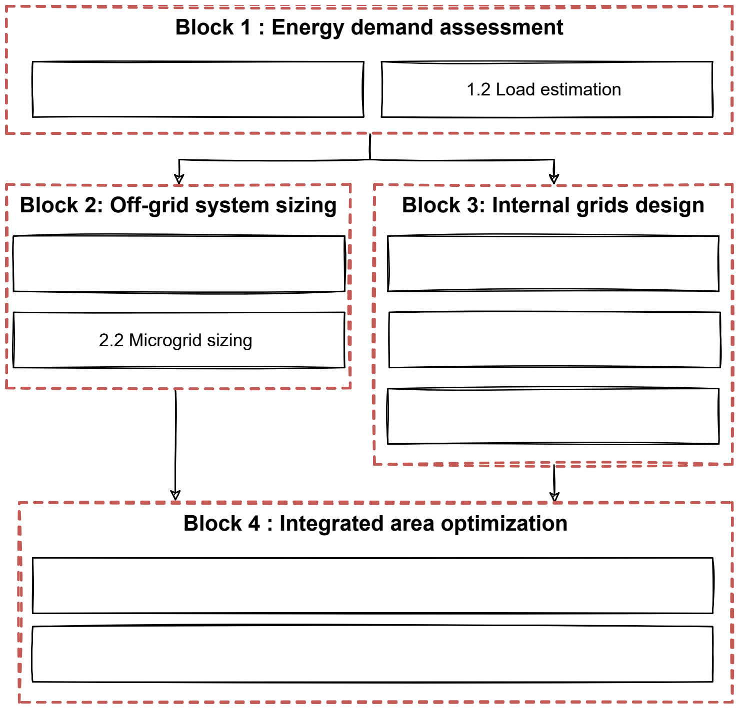


Figure 6.2: Structure of the proposed modeling framework

6.1 Block 1: Energy demand assessment

6.1.1 Introduction

The assessment of energy needs of communities is one of the fundamental aspects that should be considered to propose an effective electrification strategy, as widely discussed in chapter 3. Given a rural area to be electrified, the proposed approach aims to compute for each community the optimal electrification strategy, deciding whether to install an off-grid microgrid or to connect it to the national grid. For this reason, the starting point of the proposed electrification planning procedure is the identification of rural communities, i.e. clusters of users that will be electrified with the same strategy, and of their load demand. Accordingly, the energy demand assessment block (Block1) is divided into two modules:

1. Identification of communities to be electrified in an homogeneous manner, either -on or off-grid;
2. Estimation of energy demand for each community.

Table 6.1: Application of the different case studies

Case study	Blocks	Modules
Mozambique - Namanjavira	All blocks	All modules
Lesotho - Butha Buthe	All blocks	All modules
Brazil - Cavalcante	1,3	1.1, 3.2, 3.3
Bolivia - Omereque	3	3.1

The main steps of the first block of the Gisele procedure are shown in the flow-chart of figure 6.3 and described in the following sections.

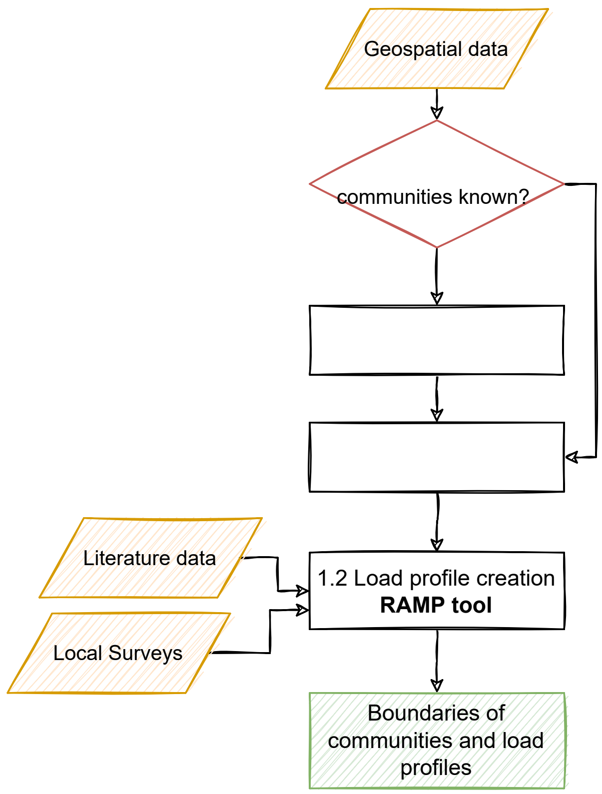


Figure 6.3: Structure of the Block 1: Energy demand assessment

6.1.2 Module 1.1: Identification of communities

To characterize the area under study and define the extent of interconnected energy systems, the type of useful information that could be found can be one of the following:

- boundaries of existing villages and communities, in the form of polygon vector layer or similar: in this case each village could be considered one separate energy system;
- raster file with population density: the distribution of population is an indicator of the existing communities and load centers and can be used to extrapolate the boundaries of possible energy systems.

While raster files with population densities are available for the whole world from different sources of data [97], [98], this is not the case of precise boundaries of communities, that in disperse and rural areas are often not mapped. In case the boundaries of communities are not well defined and geospatial data are not available, the first step for identifying communities is the clustering of populated points, to find dense areas. A preprocessed dataset that already provides some clusters of populated areas, is the one developed by the KTH and used as input to the Onsset procedure [99]. It can be used as an interesting input, however, due to the procedure followed, that merges adjacent populated cells, clusters usually are very small, even much smaller than existing communities, so that the problem can become unmanageable.

Literature Review

Definition 4 *Cluster analysis, or clustering, is the process of partitioning a set of data objects(or observations) into subsets. Each subset is a cluster, such that objects in a cluster are similar to one another, yet dissimilar to objects in other clusters [100].*

Different clustering techniques have been developed during time; following the classification of [100], they can be organized in the following categories:

- Partitioning methods: they divide the set of n observations into k groups, where $k <= n$, usually on the basis of distance criterion. K-means, K-medoids are representative methods of this class [101];
- Hierarchical methods: they create a hierarchical decomposition of the group of observations. They can be either agglomerative, if observations are grouped into clusters of growing size, or divisive, if an initial big cluster is subdivided gradually. BIRCH and Chamaleon are examples of algorithms in this category;

- Density-Based methods: they consider a density criterion rather than distance to identify clusters. DBSCAN [102], OPTICS [103], DENCLUE [104] are some of the most well known methods of this class;
- Grid-based methods: they quantize the object space into a finite number of cells that form a grid structure. All the clustering operations are performed on the grid structure (i.e., on the quantized space). STING and CLIQUE represent this category of models.

Clustering related literature is continuously expanding and modern and more sophisticated methods are available. The methods differ in terms of time complexity, required input parameters, management of outliers and clusters' shape identification.

Clustering based approaches for communities identification are followed by some of the literature works for comprehensive rural electrification planning ([83] identifies population clusters, without however specifying the followed procedure, while REM adopts a bottom-up clustering procedure developed specifically for the purpose).

Modeling approach

Density based clustering has been found to be the most suitable class for selecting the communities for the following reasons: a) it allows the creation of clusters with non convex shapes, so closer to real aspect of communities b) it considers the presence of outliers, that are points that do not belong to any cluster c) it is suited also for large datasets.

DBSCAN and HDBSCAN [105] have been identified as the most suitable algorithms given their adaptability to big datasets, limited memory requirements and computational complexity, scalability to different problems, relative simple parameter tuning and results interpretation.

DBSCAN requires in input two parameters, ϵ, a distance measure which describes the maximum radius to consider, and $MinPts$, the minimum number of points to form a cluster: a point p is a core point if at least $MinPts$ are found within its *eps*-neighbourhood.

In the proposed methodology, as first step, the population raster is transformed into a grid of points, with defined coordinates and a value of population associated. Those points represent the n observations to be clustered. Points are weighted according to the population associated and $MinPts$ give an indication of the minimum size of villages to be considered for electrification. The pseudocode of the procedure is reported in the algorithm 1 [100].

eps and $MinPts$ could be defined so that the cluster density is equal to the average population density ρ of rural communities in the area:

$$\epsilon = \sqrt{\frac{MinPts}{\rho \cdot \pi}} \tag{6.1}$$

Algorithm 1: DBSCAN

```
Data: D: dataset containing all the populated n points, their coordinates and
      associated population
Data: eps: the radius parameter
Data: MinPts: the minimum population threshold
Result: A set of density-based clusters
mark all objects as unvisited;
for All unvisited objects p do
    mark p as visited;
    if the eps-neighborhood of p has at least MinPts population then
        create a new cluster C;
        add p to C;
        let N be the set of objects in the eps-neighborhood of p ;
        for each point p' in N do
            if p' is unvisited then
                mark p' as visited;
                if the eps-neighborhood of p' has at least MinPts points then
                    add those points to N;
                end
            end
            if p' not yet a member of any cluster then
                add p' to C;
            end
            output C;
        end
    else
        mark p as noise
    end
end
```

This is a parameter that is not available easily. Moreover, not all villages, even in the same administrative division of a country have the same structure, some of them may be more packed and with an urbanized structure while others may be scattered and less dense.

HDBSCAN has been developed more recently by the same authors of DB-SCAN and is a combination of a hierarchical algorithm and a density based one. It overcomes the limit of homogeneous density clusters and requires as input only the $MinPts$ parameter.

There is also the possibility to combine DBSCAN and HDBSCAN to avoid problems related to the definition of a multitude of small clusters close to each other in densely populated areas as proposed in [106].

Obviously, it is not straightforward to identify which is the algorithm and the combination of parameters that bring to the optimal solution. Some qualitative and quantitative criteria could be used:

- Outlier identification: outliers are the points that should be excluded by the process of electrification. They are so sparse that do not justify even a microgrid with a distribution grid, but should be electrified with SA;
- Villages identification: a good clustering procedure should be able to identify the boundaries of recognised communities.

The accuracy of the tools in identifying communities can be validated only in areas where the boundaries are known, as shown in the grey box below.

Insight

Lesotho-Butha Buthe The different results, strengths and limits of the clustering algorithms are exemplified in the following graphs and pictures, reporting as a case study the area of Butha-Buthe district in Lesotho. The area is particularly interesting because of the availability of Open Street Map (OSM) data representing the boundaries of existing residential areas (fig. 6.4). Population data are collected from the High Resolution Settlement Layer of Columbia University [97], with a resolution of 30 meters.

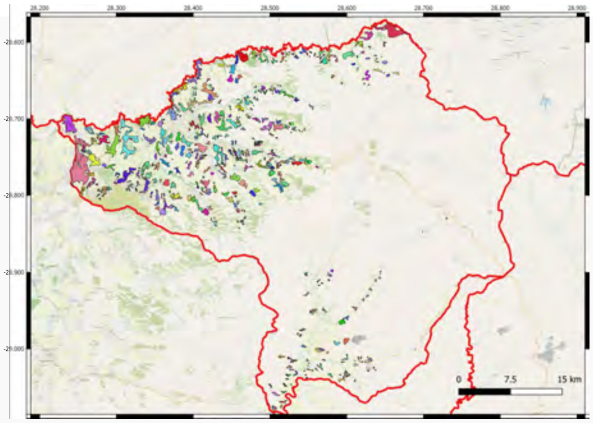


Figure 6.4: Map of villages

The probability distribution of population and population density in each of the 359 residential areas are reported in figure 6.5. The $MinPts$ parameter is defined as the minimum size of clusters (communities) and chosen equal to 40, to exclude only the 10 % of smallest areas. The density threshold, necessary to define a reasonable value of ϵ to run DBSCAN, is also set so to exclude the 10 % less dense areas, at value of $650 people/km^2$.

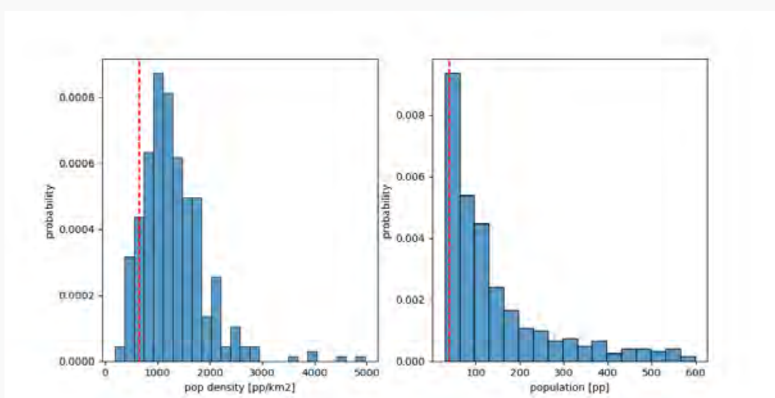


Figure 6.5: Characteristics of villages

The core radius ϵ is found from equation 6.1 to be equal to $140m$. The results are reported in table 6.2.

Table 6.2: Results of comparison between DBSCAN and HDBSCAN

Cluster strategy	**DBSCAN**	**HDBSCAN**
ϵ	140 m	-
$MinPts$	40	40
N cluster	430	352
N outlier	3234, 11% tot. pop.	2466, 8% tot. pop.
N outliers that are real outliers	1811	1813
Adjusted Rand Index (ARI)	0.45	0.35
Rand Index (RI)	0.94	0.96

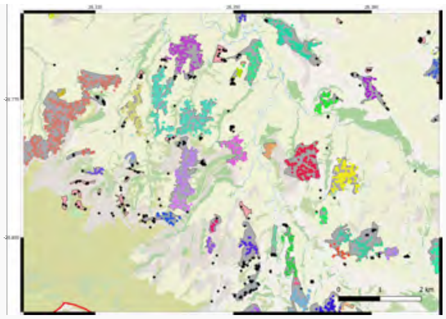
The RI and the ARI are used to evaluate the accuracy of the clustering algorithms in detecting the correct shape of villages. RI can be shortly defined as a measure of the percentage of correct decisions made by the algorithm:

$$RI = \frac{TP + TN}{TP + FP + FB + TN} \tag{6.2}$$

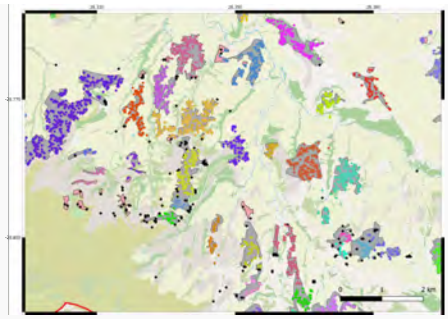
Where TP is the number of true positives, TN is the number of true negatives, FP are the false positives and FN are the false negatives. The value of RI is comprised between 0 and 1, with 0 indicating total disagreement between clustering and reality and 1 perfect classification. ARI corrects the RI normalizing it with respect to the RI that would be obtained with random classification (expected(RI)). ARI

$$ARI = \frac{RI - expected(RI)}{max(RI) - expected(RI)} \tag{6.3}$$

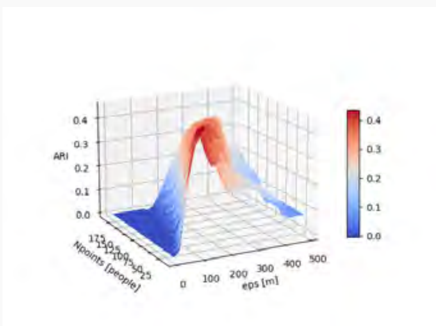
The total number or real outliers, points not belonging to any community of at least 40 inhabitants, is of 6021 which is 3 times larger than the outliers detected by the algorithms.
Graphs 6.7a and 6.7b provide a sensitivity analysis of the adjusted Rand index varying the parameters of the clustering algorithms. They show a strong dependence of DBSCAN output on the ϵ parameter, with ARI varying from a minimum of 0 to a maximum of 0.45 while the $MinPts$ does not influence much the behaviour. HDBSCAN is instead more unstable when varying the $MinPts$ parameter and ARI varies from a minimum of 0.15 to a maximum of 0.47. As a general outcome, the two clustering algorithms show a comparable performance and the choice could be done on a case by case basis.



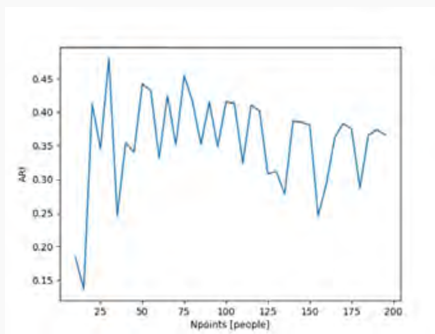
(a) DBSCAN output with $\epsilon = 140$ and $MinPts = 40$



(b) HDBSCAN output with $MinPts = 40$



(a) Variation of Adjusted Rand Index, running DBSCAN with different ϵ and $MinPts$ values



(b) Variation of Adjusted Rand Index, running HDBSCAN with different $MinPts$ values

The two algorithms show an equivalent performance. HDBSCAN is more robust with respect to the parameters' choice, however with DBSCAN it is possible to have more control on the characteristics of the identified clusters and there is no risk of clustering wide sparse areas. Since DBSCAN is also one of the algorithms mostly recognized and tested in literature it is generally considered as the best choice for Gisele.

The procedures can be run iteratively, changing the parameters in input, choosing the required combination of clusters given a certain minimum electrification rate in the area. For this reason, the user version of Gisele is provided with a sensitivity analysis, that allows the planner to try different combinations of parameters and to categorize them according to the percentage of people and the percentage of area electrified, so to choose in a qualitative way the best combination.

6.1.3 Module 1.2: Demand Assessment

Literature Review

Once communities are identified, it is necessary to make a reliable estimation of the electric needs of each area, to implement the most effective electrification strategy. Electric grids are typically designed to sustain the peak load power, so rough estimations of power per capita multiplied by coincidence factors are enough to size the lines and estimate their cost [72]. Optimal sizing of hybrid minigrids, on the other hand, requires the estimation of the load profile, to define the dispatching logics of the different energy sources.

The techniques to model electric demand are subdivided in literature into two main classes, top-down and bottom-up models which differ in terms of assumptions and input data used [31, 107, 108, 109, 110]. Other approaches are also found in literature such as system dynamics [34], input-output analysis or scenario approaches.

The top-down approach uses macroscopic data and econometric models to model electricity demand trend based on economic theories. Those models make extensive use of historical data related to energy use of entire sectors (e.g. residential) to extrapolate future trends on the basis of economic and technological variables (price or appliance ownership). They rely on aggregate data and on historical trends which are able to correctly drive the model whenever there are no big discontinuities in the systems. Their usage for estimating load demand in unelectrified rural areas of the Global South is however challenged by the lack of accurate data and by their difficulties in properly detecting differences between rural and urban contexts. Bottom up approaches are based on a very specific and detailed data collection that allows to model energy services and uses thus leading to realistic load projections [9]. Estimated load consumption is computed for each single user and then aggregated up to the system level. [111] and [112] propose two tools (namely LoadProGen and RAMP) for the stochastic estimation of load demand starting from the collection of data related to user categories and appliances functioning windows and time. Given their level of detail, if enough data on the estimated future use of appliances and number of customers is available, also load growth scenarios can be evaluated. The disadvantage of this approach is that usually it requires an extensive on-field data collection campaign based on local surveys to gather all the necessary information.

System dynamics approaches are useful methods for assessing all the four steps of energy demand development, described by [32], since they are based on the creation of virtuous dynamic cycles of growth. However their complexity makes their application suited only for specific and localized sites [34].

Modeling approach

The load profile of each of the communities identified with the previously described module 1.1 is created by means of a bottom-up procedure. This is preferred to a top-down approach given the low availability of historical trends and data from which to infer possible load profiles, in particular in Sub-Saharan Africa. The tool used to the scope is RAMP, a python-based tool able to create synthetic load profiles starting from classes of users, appliances and usage habits [112]. The main input parameters required are described in the table 6.3.

Table 6.3: Main RAMP input parameters

Parameter	**Description**
User_j	Name of each user class
N_j	Number of users within each user class
Appliance_{ij}	Name of appliance associated with j class of user
n_{ij}	Number of appliances i within class j
P_{ij}	nominal power absorbed by appliance ij
ft_{ij}	functioning time: total time appliance is on during the day
fc_{ij}	functioning cycle: minimum time appliance ij is on after switch-on
fw_{ij}	functioning window: period during the day when each appliance ij can be on
Rfc_{ij}	random variation of functioning cycle
Rfw_{ij}	random variation of functioning window

The load profile of each user of a specific user class is provided by the combination of the usage pattern of each appliance ij, computed by defining, in a stochastic manner, the times t_{ij} the appliance ij is switched on within the day. Those times must be selected within the identified functioning windows fw_{ij}. Once the appliance is on, it must remain on for at least the previously defined functioning cycle fc_{ij}. The overall daily load profile results from the aggregation of the user classes' profiles j. Each profile j will be different from the others of the same user class category because of the stochastic variability provided by the input parameters and by the randomized selection of appliances switch on-times.

Since computations are time intensive, the tool is run in advance to create a pool of possible profiles that are then combined inside Gisele simulations to recreate realistic profiles for each community. Notice that these preliminary analyses have been performed for Sub-Saharan rural communities; for other areas in the world, different assumptions should be adopted and RAMP tool should be run again.

To make reasonable assumptions with respect to the social structure of Sub-Saharan communities, a combination of literature works, data from the Global Survey on Energy Access program launched by WorldBank and real measurements are used. The WorldBank survey campaign on energy access, launched with support from the ESMAP aims to collect on field data following the Multi Tier Framework (MTF). Those data are now available for 8 countries (Ethiopia, Kenya, Liberia, Malawi, Niger, Nigeria, Rwanda, Zambia) and to the author's knowledge are the most complete source of data related to electricity consumption at household level [113, 114, 115, 116]. Questionnaires were distributed to statistical samples of households living in villages spread across the countries. Questions ranged from energy consumption habits, type of electricity access, satisfaction and affordability of energy resource as well as number and type of electric appliances owned. In Ethiopia and Rwanda, moreover, questionnaires at community level where also created, and they constitute a useful source of information related to communities' composition in terms of available categories of users and electric consumption habits.

RAMP creates a different profile for each type of user, which has specific energy consumption habits. For this reason, the first step of the proposed procedure is the identification of possible user categories forming typical communities. The types of users are distinguished into three main groups: households, business activities and social facilities, classification used in the majority of the studies related to load forecasting [117, 107]. Each of the categories has different types of users, that are distinct in terms of energy needs. Those are chosen to be as general and representative as possible and are listed and described as follows:

Households
Households are distinguished into the 5 energy access tiers identified by the WorldBank in [118] and described in 6.4. To each of them a different pool of appliances and energy consumption is associated, whose detail is reported in annex A, table A.1.

Table 6.4: Electric appliances belonging to different households categories

Category	Electric Appliances	Load level
Tier 1	Lights, phone chargers, radio	Very low load (3-50 W)
Tier 2	..., TV, PC, fan	Low load (50-200W)
Tier 3	..., refrigerator, food processor, rice cooker	Medium load (200-800W)
Tier 4	..., washing machine, iron, hair dryer, toaster, microwave	High load (800-2000W)
Tier 5	..., air conditioner, space heater, vacuum cleaner, water heater, electric cookstove	Very high load (2000W or more)

Social facilities

Those are the social infrastructures, typically public, that are generally present in rural communities. They are subdivided into the three main categories of health centers, schools and worship centers. The type and number of electric appliances typically owned is obtained from the MTF data of Ethiopia [113], since in the country WorldBank questionnaires where also proposed to those facilities. Figure 6.8, 6.9, 6.10 show in different colours the number of appliances (each row is a different appliance) owned by each infrastructure (each column), within electrified rural and urban communities resulting from the questionnaires. White colour corresponds to no appliances, yellow to a high number and blue to low number. Facilities in urban communities are more developed, they use a wider number of appliances with a higher power consumption. For this reason,inside Gisele procedure, for each category of user, two profiles are created, one typical of rural and the other typical of urban communities. The details are reported in table 6.5 and in annex A, table A.1.

Table 6.5: Electric appliances belonging to social facilities

Infrastructure	Electric Appliances	Load level
Health Center rural	Lights, phone chargers, sterilizers, fridge	High load (>1000W)
Health Center urban	Lights, phone chargers, sterilizers, fridge, microscope, centrifuge, blood monitor	Very high load (> 5000 W)
School rural	Lights, PC	Low load(100-200W)
School urban	Lights, PC, TV	Medium load(500-1000W)
Worship rural	Lights	Very low load (50-100W)
Worship urban	Lights, TV, PC	Low load (500-700W)

Business activities

The importance of productive uses of electricity for the real development

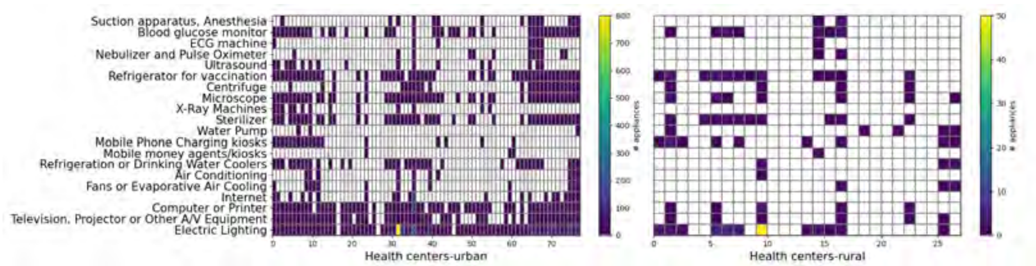


Figure 6.8: Number of appliances belonging to each urban (left) and rural (right) health center in Ethiopia, author elaboration from MTF data

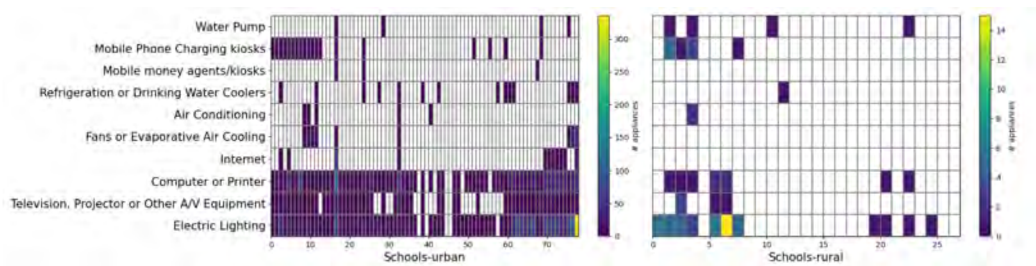


Figure 6.9: Number of appliances belonging to each urban (left) and rural (right) education facility in Ethiopia, author elaboration from MTF data

of communities is becoming widely recognized in literature [119], [120], [121]. Business activities are however manifold and not easy to categorize. Since a bottom-up approach defining each single appliance could be complex to adopt, given that no data is available on wide scale, the authors in [117] use a baseline profile for commercial and productive activities scaled according to some proxy indicators (such as number of households or road density). In case no specifications on the case study are provided, this approach could also be replicated in the proposed procedure, using the same data, given the difficulty in recreating realistic profiles without a reasonable amount of measurements and surveys. In the case studies reported in 5, related to Mozambique and Lesotho, some standard load profiles, created with values from literature and values retrieved from on-field surveys, are created in advance and combined for typical users. The on-field data collected in Namanjavira, Mozambique, are used as reference:

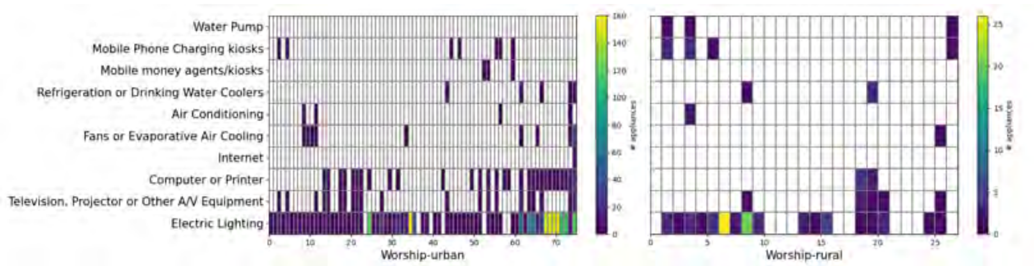


Figure 6.10: Number of appliances belonging to each urban (left) and rural (right) worship place in Ethiopia, author elaboration from MTF data

Table 6.6: Examples of appliances belonging to businesses

Infrastructure	**Electric Appliances**	**Load level**
Shop	Lights, phone chargers, radio (refrigerator)	Low/Medium load (100-1000W)
Carpenter	Lights, phone chargers, utensils, radio, fan	High load (> 1000 W)
Hairdresser	Lights, phone chargers, radio	Low load(100-200W)
Office	Lights, PC, TV, Fan, radio	Medium load(500-1000W)
Tailor	Lights, phone chargers, radio, fan	Low load (100-500W)

For assessing the number of users of each category within communities (i.e the parameter N_j), and generalize it to be replicable inside Gisele procedure, an analysis of the MTF data in Ethiopia and Rwanda is performed. The percentage of each category of households in the community with respect to the total number of households in the two countries is shown in fig. 6.11. From those data, an average value for urban and rural communities has been retrieved. The average number of social facilities and their probability of existence in the communities of the two countries is shown in fig. 6.12. Data related to the number of infrastructure facilities have also been compared to the number of households in the villages and to their extension to find some useful proxies; however, since no specific trends correlating number of facilities to population or surface area have been highlighted, average values are used in the procedure.

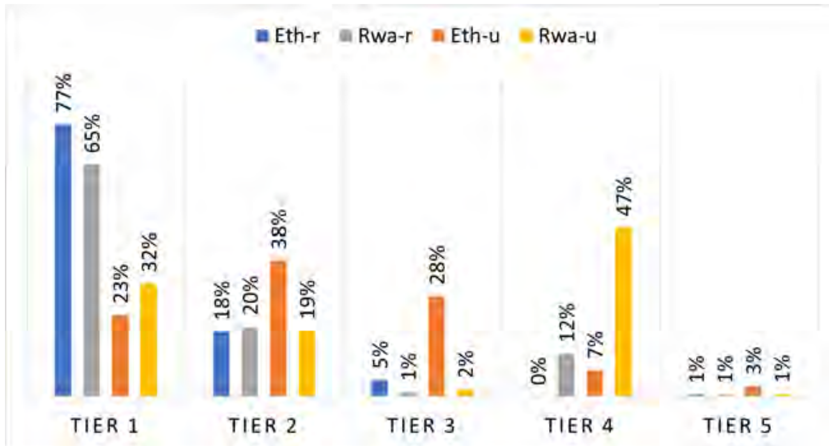
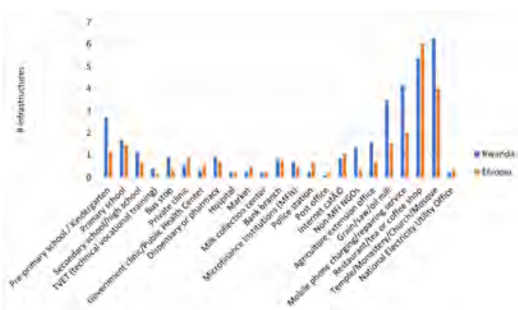
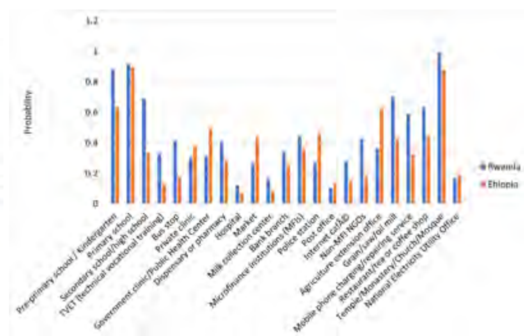


Figure 6.11: Percentage of grid-connected households belonging to each tier of energy use



(a) Average number of infrastructures in communities of Rwanda and Ethiopia



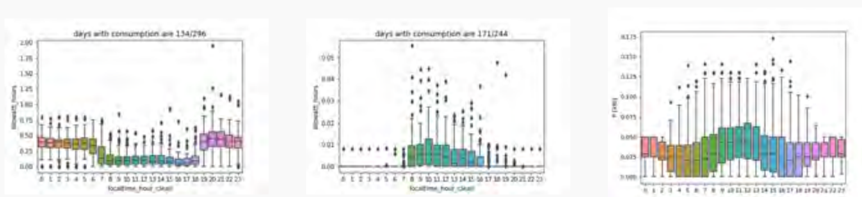
(b) Probability of the presence of at least one infrastructure in the communities of Rwanda and Ethiopia

Figure 6.12: Infrastructures in Rwanda and Ethiopia communities

To create the load profile of the community, given its characteristics coming from geospatial data (geographic extension, number of households, type (urban or rural) and type of economy), single user profiles (whose input data are resumed in annex A) are assembled. There are some lucky cases in which some infrastructures are mapped and data are publicly available. In this case the aggregated profile can be more accurate. The following Insight box provides an example of the application of the load demand creation procedure in some villages where the actual consumption from microgrid is monitored.

Insight

The research group is collaborating with a company working in the minigrids sectors, which provided the measured energy consumption of 10 minigrids installed in Sub-Saharan Africa. For privacy reasons, the name and locality of those sites as well as the company name must remain anonymous. The data provided is disaggregated at the meter level, usually corresponding to few households, shops or schools. Running RAMP to simulate single users profiles (more specifically schools), the results reported in the following third sub figure are found. It is hard to evaluate the quality of the simulated profiles, and compare them to measured ones (first two sub figures) since specificities of each case are hardly captured by algorithms.

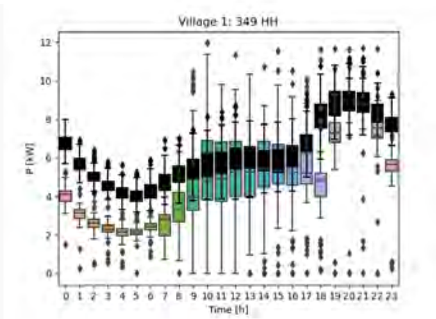


(a) Measured school profile, village1

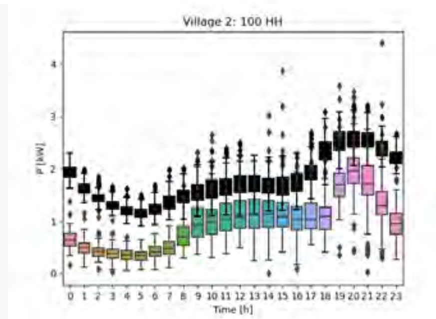
(b) Measured school profile, village 2

(c) Simulated rural school profile

On the other hand, categorizing profiles according to the type of users, an interesting comparison between measurements and simulations can be performed. 6.14a and 6.14b show the comparison between simulated residential load profiles, considering Ethiopian statistics on households' tiers' percentages (77% households in Tier 1, 18% in Tier 2 and 5% in Tier 3) and the aggregated measured ones.

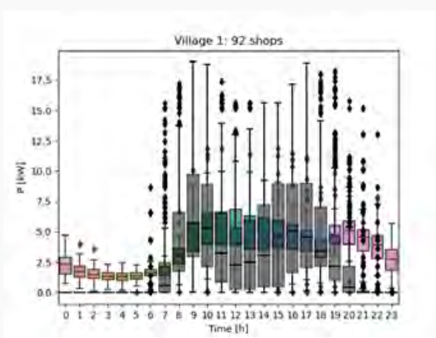


(a) Comparison between measured (colored) and simulated profiles (black), residential load village 1

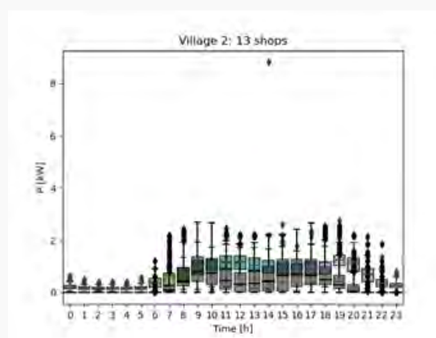


(b) Comparison between measured (colored) and simulated profiles (black), residential load village 2

Evening peak is well represented by the created profiles (in terms of peak power values and timing), however the nocturnal consumption is overestimated. Real profiles have a greater variability caused by all the non deterministic patterns of load consumption (e.g. non availability of money to pay for the connection or failures) that cause more frequent peaks and lags. The same comparison is done for the category of commercial activities (shops), since it is the second most represented in communities an aggregated analysis makes sense (see fig. 6.15a and fig. 6.15b). For building the shop profile, on-field data collected in the region of Namanjavira, Mozambique and reported in annex A are used.

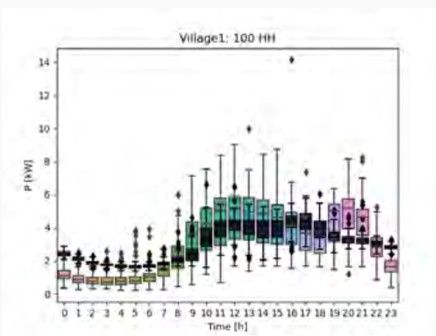


(a) Comparison between measured (colored) and simulated profiles (black), commercial load village 1

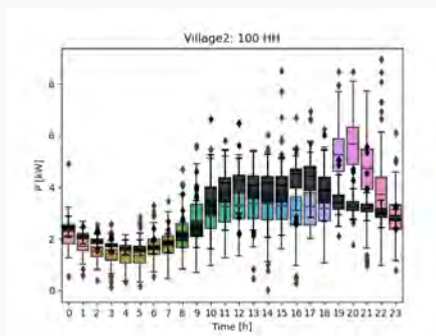


(b) Comparison between measured (colored) and simulated profiles (black), commercial load village 2

In this case, even though the order of magnitude of the power request is similar (peaks around 10kW in the first village and 2kW in the second one), the shape of the profiles is quite different. Finally, the consumption of the whole communities is plot in fig. 6.16a and 6.16b. The simulated profiles are obtained using the three categories of users, social facilities, households and business activities described in the chapter. The number of users belonging to each category in the community was known while their profiles were created starting from data retrieved from MTF and detailed in annex A. The two villages have the same number of households, however they have slightly different profiles: the second one sees a higher energy consumption and higher evening peak. In any case the simulated profile follows with good approximation the measured ones both in terms of energy content and power peak. The average daily energy has an error of 1% in the first case and of 2% in the second case while the peak power has an error of around 20% and 25% in the two cases.



(a) Comparison between measured (colored) and simulated profiles (black), total load village 1



(b) Comparison between measured (colored) and simulated profiles (black), total load village 2

The simulations demonstrate how the bottom-up procedure for formulating load profiles is effective even using generic data collected from MTF in Rwanda and Ethiopia. Thanks to the common features that most of villages in rural SubSaharan Africa share it is hence possible to generalize the procedure even in absence of on-field data and use it as input to the rest of Gisele modules.

6.1.4 Contributions and limitations

The proposed approach constitutes a valid procedure to assess electric needs on wide areas and to identify boundaries of communities. It is based on a density-based clustering approach and on the integration of a bottom-up procedure for load assessment. The availability of precomputed load profiles based on extended local surveys brings to a high scalability of the approach. After the first identification of communities, some further steps could be required to make the problem treatable.

- Subdivide communities/clusters that are too big to be electrified with a MV grid. This is because the whole procedure focuses on the design of the distribution grid and no considerations related to the positioning of new primary substations is made;
- Group communities that will likely be connected together and with the same electrification strategy because of their proximity. This allows to simplify the problem avoiding to include in the optimization non necessary variables.

Those could both be faced by a further clustering process; the author tested and used agglomerative clustering to address both of the problems, but no in depth investigation has been performed resulting in a topic candidate for future research.

6.2 Block 2: Off-grid system sizing

6.2.1 Introduction

The second block of the proposed rural electrification planning procedure is devoted to the definition of the optimal configuration of hybrid microgrids, able to satisfy the energy demand of the identified communities. After defining the RES potential in each of the communities, a Mixed Integer Linear Programming (MILP) model is used to identify the optimal portfolio of generators that could supply the load on a defined time frame. The net present cost of the identified off-grid solution for each community will, in the third block of Gisele, be compared to the grid extension, to select the optimal solution for the whole area under analysis. The main steps of this block are shown in the flow-chart of figure 6.17.

6.2.2 Module 2.1 Energy resources assessment

RES assessment focuses on wind, solar and hydro resources. Biomass resource has not been considered due to the difficulty in the estimation of its distributed potential. Biomass potential depends in fact on the availability

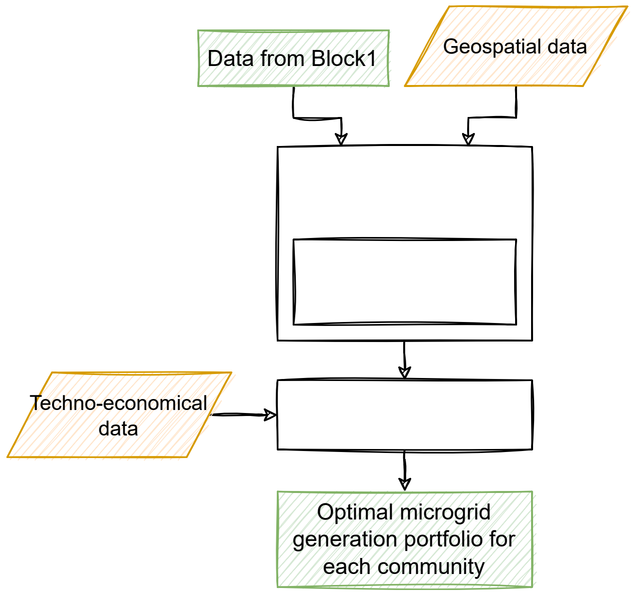


Figure 6.17: Structure of the Block 2: off-grid system sizing

of different resources such as crops residuals, manure, forest residuals or energy crops. However, freely available geospatial data related to crop types have low resolution and it is hard to derive useful indicators at community level [122]. Geothermal resource instead has been neglected since for micro-grid applications is not yet diffused, given the large capital expenditures and expensive preliminary studies.

Literature Review

Solar and wind resources

Numerous studies have been performed with the goal of providing estimates of RES potential in different geographic regions and with various time resolutions. Two methods have been extensively used to provide global data related to RES availability: meteorological reanalysis and satellite images processing [123]. Meteorological reanalysis, that is the process of integrating model results with observations distributed irregularly in space and time into a spatially complete gridded meteorological dataset, has emerged as an important data source for renewable energy modeling studies. Those data are particularly important for several reasons: reanalysis data are available globally, they provide several decades of coverage and they are usually freely available. A major advantage is that reanalyses can provide data for locations or timesteps where no direct observations are available through their integration of measurements and numerical models. Commonly used global reanalyses of the most recent generation include NASA's, MERRA, MERRA-2 [123] and the ERA5 dataset from the the European Centre for Medium-Range Weather Forecasts (ECMWF) [124]. Meteorological reanalysis has the advantage of global coverage that may come however at the expenses of accuracy. When speaking about solar irradiation data, a valid alternative approach is the use of satellite images as the freely available hourly dataset SARAH, which allows to take into account the cloud coverage and accurately estimate atmospheric conditions relevant for surface irradiance. In [125] and [126] the methodologies for deriving hourly time series data of PV and wind energy production in Europe starting from meteorological reanalysis and satellite data are explained. The databases are available and downloadable online for free from the website `https://www.renewables.ninja/`. The Global Wind Atlas, developed by a partnership among the World Bank and DTU university, is a free, web-based application which provides average wind speed and power output data across the world with a 250 m resolution. Data were derived from meteorological reanalysis of the ERA5 database. The PVGIS database uses instead satellite data to provide freely available web-accessible hourly data of PV power production across Europe and Africa, with 1 km resolution [127].

Hydro Resource

The estimation of the hydro power potential as a distributed energy source is not so straightforward as for solar and wind. Within big hydro power projects, as for instance the building of a dam, measurement campaigns are performed on-site to estimate the river flow rate and the available prevalence. When speaking about smaller projects for rural electrification however, it could be relevant to have a perspective on the available hydro resource on a wide areas, in terms of power extractable from run-of-river mini hydro turbines. The available geospatial data covering wide areas only provide the average flow rate of rivers along the year [128], [129]. However, river regime may have a very high seasonality, even with dry months, especially when speaking about small rivers, heavily influenced by rain regime. In this case an alternative solution to the on-site data collection is a pre-feasibility study via hydrologic modeling. Combining geospatial data related to altitude, land cover, soil and rain and temperature patterns it is possible to estimate the amount of water absorbed by the terrain and the one flowing on the surface, thus estimating the river flow rate of streams. One of the mostly recognized models in literature is the SWAT (Soil and Water Assessment tool) model ([130, 131, 132]). It is a physically based hydrological model based on water balances coming from a simplification of the hydrogeologic cycle (see fig. 6.18):

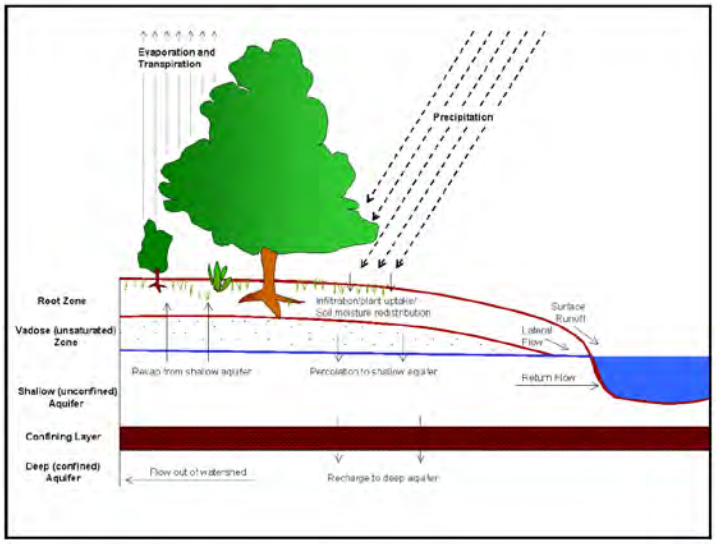


Figure 6.18: Water balance considered in the SWAT model [130]

$$SW_t = SW_0 + \sum_{i=1}^{t} \left(R_{dqy} - Q_{surf} - E_a - w_{\text{seep}} - Q_{gw} \right) \tag{6.4}$$

where SW_t is the final soil water content $(\mathrm{mmH_2O})$, SW_0 is the initial soil water content $(\mathrm{mmH_2O})$, t is the time (days), R_{day} is the amount of precipitation on day i (mm $\mathrm{H_2O}$) , Q_{surf} is the amount of surface runoff on day i(mm $\mathrm{H_2O}$) , E_a is the amount of evapotranspiration on day i $(\mathrm{mmH_2O})$, w_{seep} is the amount of percolation and bypass flow exiting the soil profile bottom on day i ($\mathrm{mmH_2O}$), and Q_{gw} is the amount of return flow on day i (mm $\mathrm{H_2O}$) . The surface runoff is the amount of water neither absorbed by the terrain or by vegetation nor evaporated in atmosphere which, therefore, flows in surface. It is computed through the Curve Number (CN) method, developed by the USDA-Soil Conservation Service (SCS) that became Natural Resources Conservation Service (NRCS) [133], for predicting direct runoff or infiltration from rainfall excess in water resources management.

The runoff CN is the result of an empirical analysis of runoff from small catchments and hill slope plots monitored by the USDA (United States Department of Agriculture). In hydrology, CN is used to determine value of rainfall which infiltrates into soil or an aquifer and, consequently, how much rainfall becomes surface runoff: a high CN means high runoff and low infiltration (urban areas), instead the lower the curve number, the more permeable the soil is (sand). Combining this parameter with a detailed morphological analysis of the target region it is possible to forecast how surface water moves through the territory and how much water funnels in each point.
CN is a function of land-use/land-cover and hydrological soil group. To each combination of this two features corresponds an empirical value of CN_{II}, obtainable by means of tables based on the ones provided by USDA technical release number 55 (TR-55) [134]. Those values are referred to normal Antecedent Moisture Conditions (AMC), i.e. the starting condition when the rainfall begins. The AMC are classified as dry, normal or moist conditions according to the ratio of P (precipitation) and PET (potential evapotranspiration):

Dry Conditions are measured for $\frac{P}{PET} < 0.8$

$$CN = CN_I = \frac{4.2 \cdot CN_{II}}{10 - 0.058 \cdot CN_{II}} \tag{6.5}$$

Normal Conditions when $0.8 \leq \frac{P}{PET} < 0.9$

$$CN = CN_{II} \tag{6.6}$$

Moist Conditions when $\frac{P}{PET} \geq 0.9$

$$CN = CN_{III} = \frac{23 \cdot CN_{II}}{10 + 0.13 \cdot CN_{II}} \tag{6.7}$$

There are many methods available in literature to compute PET, compared in [135].SWAT model includes as user-defined options the Priestley-Taylor, the Penman/Monteith and the Hargreaves. Once the value of CN is determined, the set of equations which lead to compute the runoff in the desired area are defined.

$$Q = \begin{cases} 0 & \text{for } P \leq 0.05S \\ \frac{(P-0.05S_{0.05})^2}{(P+0.95S_{0.05})} & \text{for } P > 0.05S \end{cases} \tag{6.8}$$

Where Q is the water runoff [mm/month], P is the Rainfall[mm/month], and S is the potential maximum soil moisture retention after runoff begins.

The value of S is linked to CN through the formula:

$$S = \frac{25400}{CN} - 254 \tag{6.9}$$

And

$$S_{0.05} = 1.33 * (S_{0.20})^{1.15} \tag{6.10}$$

The SWAT model, after computing the water runoff in each point of the region, through an hydrologic analysis that starts from the elevation layer (also called Digital Elevation Model (DEM))provides as result useful information about the hydrology of the region like flow accumulation, watershed basins and rivers' paths.

Modeling approach

The hourly per unit power profiles of wind and pv resources are automatically downloaded for each community through an API from the website `https://www.renewables.ninja/`, considering the coordinates of the communities. Those are yearly profiles, that will be later reshaped to consider only few typical days a year more easily utilizable by the MILP model for microgrid sizing.

Hydro power resource is evaluated according to the flow chart depicted in figure 6.19. The first step is the selection of one or more a hydrological basins in the area considered: in this step, there is a part of manual evaluation of the basin, which should be wide enough to have few or possibly none inlet points (points from where water coming from other basins enters) and small enough to avoid computational issues. If there are inlet points, it is necessary to estimate at first the flow rate entering in the inlet (through an iterative procedure that analyses other basins) and sum it up to the water accumulating inside the basin. The evaluation of monthly river discharge is done through the QGIS plug-in QSWAT, using as input the data reported in table 6.7

Weather data (temperature, wind speed, relative humidity, solar irradiation, precipitation), come from Climate Forecast System Reanalisys of the National Centers for Environmental Prediction [139]: they are high resolution data related to different stations. After running the SWAT model, which provides as output the average monthly discharge of the rivers for all the years for which climate data is available (ranging usually from the '80s to modern times), the model should be calibrated. This is typically done considering some real measurements from gauges in the river basin and adapting

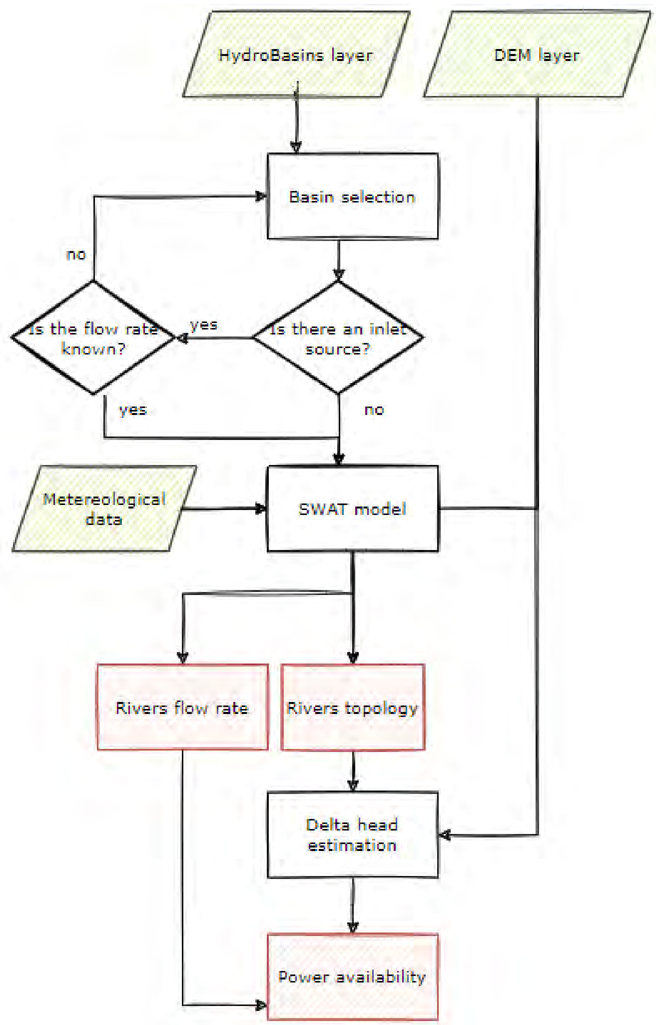


Figure 6.19: Flow chart of hydro-power potential assessment

the model parameters to fit those data [140]. Data related to the rivers' not surveilled are then obtained from the calibrated model. The African Database for Hydrometric Indices (ADHI) [141] contains different hydrometric indices, describing runoff characteristics, seasonality, floods and low flows, for about 1500 rivers covering all regions of Africa between 1950 and 2018. It includes 1466 stations with at least 10 years of data coming from on the collection of stations from the Global Runoff Data Center (GRDC) and the SIEREM database [142], [143]. It was created with the scope of calibrating and validat-

Table 6.7: Geospatial layers and data used as input to SWAT procedure

Data	Type	Resolution	Source
Elevation	Raster	30m	NASA SRTM Digital Elevation [136]
Basin area	Vector		HydroBasin [128]
Rivers	Vector		HydroRivers [128]
Soil	Raster	7km	Digital Soil Map- FAO [137]
Landuse	Raster	400m	USGS GLCC [138]
Weather data		text files	CFSR [132]

ing hydrological models in the Africa region. It is the most complete dataset for this kind of study but given the fact that data are old and hard to verify and many basins do not have station based data, the choice is to avoid using those data not to include further uncertainties. As a first comparison, the order of magnitude of the annual river flow rate is compared to that of the Hydroriver dataset ([128]), to check if macro errors occur.
The average head along a river bed is computed using the DEM layer. Along each branch of the river with a unique flow rate (without inlet or outlet flows), the elevation is sampled each 200m . This is considered a reasonable length for the derivation channel of run-of-river turbines. Given errors coming from the discretization of the surface, the average head along the river is computed with a linear regression of the elevation of the points sampled, as show in figure 6.20.

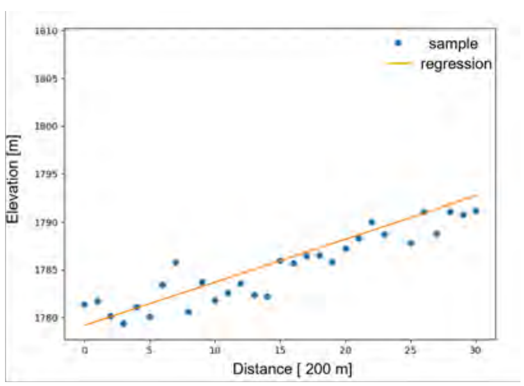


Figure 6.20: Evaluation of available head

The monthly nominal power of the turbine is then computed with the formula:

$$P(t) = Q(t) \cdot \rho \cdot g \cdot h \tag{6.11}$$

Being Q(t) the average river flow rate on a monthly basis $[m^3/s]$, g is the gravitational constant $[m/s^2]$, ρ is the water density $[kg/m^3]$and h is the available head [m].

The hydro resource associated to each community is computed considering the closest rivers with an average power availability above a certain threshold. The output is an average monthly power profile for each of the rivers and an associated possible size for the hydro turbine, chosen among a set of possible sizes with costs varying according to economies of scale.

The steps of the procedure related to a basin in Zambezia province, Mozambique are exemplified in figure 6.21. At first the hydrological basin is chosen (6.21a), in this case it has only an outlet point, where water from all the rivers converges, so a single simulation is enough. Secondly the area is subdivided by the SWAT model into sub basins (6.21b) and for each of them the river flow is computed (6.21c), finally the average power potential is computed (6.21d).

6.2.3 Microgrid sizing

The RES potential estimated with module 2.1, together with the load profile estimated within Block1, are used as input to the next procedure, devoted to define the optimal size of hybrid microgrids for each of the communities.

Literature Review

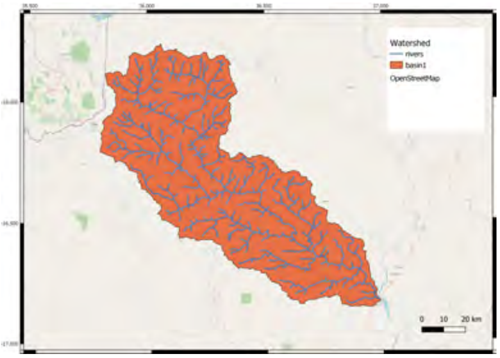
Modeling approaches for defining the optimal hybrid microgrids generation portfolio in literature use different optimization techniques:

- Numerical programming models: use Linear Programming (LP), MILP, Non Linear Programming (NLP) techniques to find the optimal size and dispatching strategy of several generation sources [56], [144];

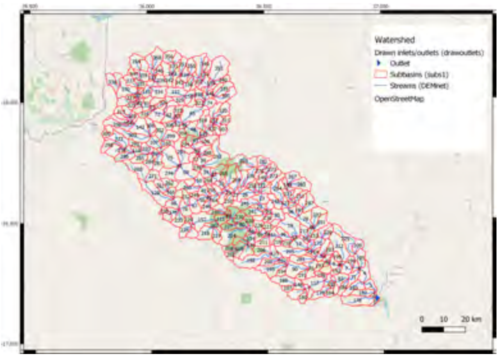
- Heuristic programming: heuristic optimization algorithms such as genetic or particle swarm optimization perform the optimal sizing of the system [145], [146], subject to dispatching set with predefined rules;

- Two-stage programming: in this case heuristics algorithm perform sizing, finding possible combinations of generation portfolio and are coupled in a loop with LP or MILP that optimize dispatching [147].

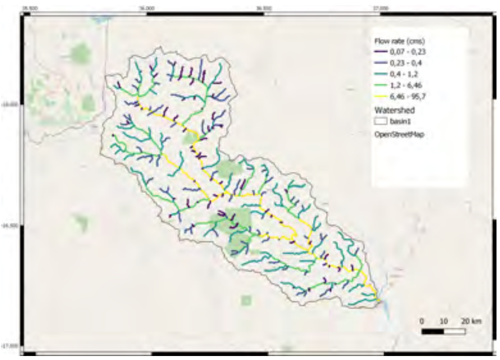
As for the tools for off-grid systems sizing presented in chapter 4, they are all based on different approaches: iHOGA uses Genetic Algorithm (GA),



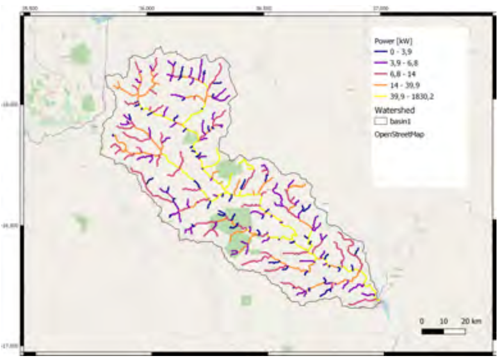
(a) Main hydrological basin



(b) Rivers subbasins



(c) Rivers average flow rate



(d) Rivers average power potential

Figure 6.21: Main steps of the procedure for hydropower potential assessment

HOMER heuristic algorithms, but specific details are not provided, Microgrids.py is a LP model and DER-CAM is based on MILP.

Modeling approach

The model used for microgrid sizing is a readaptation in Python and *pyomo* ([148, 149])of the MILP model presented in [144] and [150] developed within the research group in Politecnico di Milano. The choice of the model stands in the possibility of designing hybrid microgrids with several generation sources (Diesel Generator (DG),PV, Wind Turbine (WT), Hydro Turbine (HT) and BESS), being able to include also integer variables, such as the number of diesel generators.

The model is composed of the following sets, variables and constraints:

Sets:

i	Available technologies	$i \in \{g, p, wt, b, ht\}$
g	DG type	
p	PV type	
wt	WT type	
ht	HT type	
b	BESS type	
h	Hour of the day	
y	Year	

Scalars:

F	Cost of fuel	[$/l]
$\overline{Y}$	project lifetime	[years]
N_d	number of typical days	[day/year]
$\overline{ENS}$	maximum energy not supplied	[%]
$\underline{RES}$	minimum energy produced by RES	[%]
A	cost coefficient DG	[l/h]
B	cost coefficient DG	[l/h/kW]
γ_d	load forecast error	[0-1]
γ_{pv}	PV forecast error	[0-1]
γ_{wt}	WT forecast error	[0-1]
r	project discount factor	[0-1]
M	big constant	-

Parameters:

CC_i	capital cost of technology i	[$/unit]
$M_{i\setminus\{g\}}$	Operation and Maintenance (O&M) yearly cost of one unit of i	[$/unit/y]
M_g	O&M hourly cost of one unit of DG	[$/unit/hour]
H_g^{life}	Lifetime of g	[hours]
H_b^{life}	Lifetime of b	[kWh]
$Y_{i\setminus\{g,b\}}^{life}$	Lifetime of technology i	[years]
η_b	BESS efficiency	[-]
η_{ht}	HT efficiency	[-]
D_h	load demand	[kW]
$P_{h,p}^{pv}$	per unit power available from PV at time h	[kW/unit]
$P_{h,w}^{wt}$	per unit power available from WT at time h	[kW/unit]
$P_{h,ht}^{riv}$	power available from river at time h	[kW]
$\underline{P_g}$	min power of DG	[0-1]
$\underline{P_{ht}}$	min power of HT	[0-1]
$\overline{PQ_b}$	maximum BESS power-to-energy ratio	[kW/kwh]
C_b	BESS capacity	[kWh/unit]
C_g	DG capacity	[kW/unit]
C_{ht}	HT capacity	[kW/unit]
$\overline{DOD_b}$	maximum BESS depth of discharge	[0-1]

d_h	discount factor for hour h	[-]
d_y	discount factor for year y	[-]
w_h	weight of each hour	[-]

Variables:

IC_i	Investment cost of technology i	[$]
$O\&M_i$	O&M cost of technology i	[$]
RC_i	replacement cost of technology i	[$]
SV_i	salvage value of technology i	[$]
N_i	number of installed units of technology i	[-]
$U_{h,g} \in (0,1)$	number of DG units of type g active at hour h	[-]
$U_{h,ht} \in (0,1)$	number of HT units of type ht active at hour h	[-]
$FC_{h,g}$	fuel consumption of g at time h	[l]
$P^{dch}_{h,b} \geq 0$	discharging power of BESS b at time h	[kW]
$P^{ch}_{h,b} \geq 0$	charging power of BESS b at time h	[kW]
P^{ren}_h	sum of PV and WT power injected into the system at hour h	[kW]
$P^{dg}_{h,g}$	power produced by DG at time h	[kW]
$P^{ht}_{h,ht}$	power injected by HT at time h	[kW]
D^u_h	unmet demand	[kW]
$w^{dch}_{h,b} \in (0,1)$	1 if BESS is in discharging mode at time h, 0 if charging	-
$Q_{h,b}$	BESS energy level	[kWh]
$R^{dg}_{h,g}$	reserve to be provided by DG of type g	[kW]
$R^{sb}_{h,b}$	reserve to be provided by BESS of type b	[kW]

Objective function: NPC minimization

Objective function: NPC minimization

$$\min NPC = \sum_i (IC_i + O\&M_i + RC_i - SV_i) \tag{6.12}$$

Investment cost

$$IC_i = N_i \cdot CC_i \tag{6.13}$$

O&M costs

$$O\&M_{i\setminus\{g\}} = N_i \cdot M_i \cdot \sum_{y=1}^{\overline{Y}} d_y \tag{6.14}$$

$$O\&M_g = \sum_{h=1}^{\overline{H}} d_h \cdot (M_g \cdot U_{h,g} + F \cdot FC_{h,g}) \tag{6.15}$$

Replacement cost

$$RC_g = \frac{CC_g}{H^{life}_g} \cdot \sum_{h=1}^{\overline{H}} d_h \cdot U_{h,g} \tag{6.16}$$

$$RC_b = N_b \cdot CC_b / H_b^{life} \cdot \sum_{h=1}^{\overline{H}} d_h \cdot P_{h,b}^{dch} \tag{6.17}$$

Salvage value

$$SV_{i \setminus \{g,b\}} = d_{\overline{Y}} \cdot N_i \cdot CC_i \frac{Y_i^{life} - \overline{Y}}{Y_i^{life}} \tag{6.18}$$

$$\tag{6.19}$$

Constraints:

Power balance

$$\sum_b \left(P_{h,b}^{dch} \cdot \eta_b - \frac{P_{h,b}^{ch}}{\eta_b} \right) + \tag{6.20}$$

$$+ P_h^{ren} + \sum_g P_{h,g}^{dg} + \sum_{ht} P_{h,ht}^{ht} + D_h^u = D_h$$

Energy not supplied

$$\sum_{h=1}^{\overline{H}} D_{(y-1)\cdot(24\cdot N_d)+h}^u \leq \sum_{h=1}^{\overline{H}} D_{(y-1)\cdot(24\cdot N_d)+h} \cdot \overline{ENS} \tag{6.21}$$

Renewable production

$$P_h^{ren} \leq \sum_p N_p \cdot P_{h,p}^{pv} + \sum_w N_w \cdot P_{h,w}^{wt} \tag{6.22}$$

Minimum renewable production

$$\sum_{h=1}^{\overline{H}} (P_h^{ren} + \sum_{ht} P_{h,ht}^{ht}) \geq \sum_{h=1}^{\overline{H}} D_h \cdot \underline{RES} \tag{6.23}$$

Diesel generator constraints

$$FC_{h,g} = A \cdot U_{h,g} + B \cdot P_{h,g}^{dg} \tag{6.24}$$

$$P_{h,g}^{dg} + R_{h,g}^{dg} \leq C_g \cdot U_{h,g} \tag{6.25}$$

$$P_{h,g}^{dg} \geq \underline{P_g} \cdot U_{h,g} \cdot C_g \tag{6.26}$$

$$U_{h,g} \leq N_g \tag{6.27}$$

Hydro turbine constraints

$$P_{h,ht}^{ht} \geq= C_{ht} \cdot U_{h,ht} \cdot \underline{P_{ht}} \tag{6.28}$$

$$P_{h,ht}^{ht} \leq P_{h,ht}^{riv} \cdot \eta_{ht} \tag{6.29}$$

$$P^{ht}_{h,ht} \leq C_{ht} \cdot U_{h,ht} \tag{6.30}$$

$$U_{h,ht} \leq N_{ht} \tag{6.31}$$

BESS constraints

$$Q_{h,b} = Q_{h-1,b} + (P^{ch}_{h,b} - P^{dch}_{h,b}) \cdot \Delta h \tag{6.32}$$

$$Q_{h,b} \geq N_b \cdot C_b \cdot (1 - DOD_b) + R^{sb}_{h,b} \tag{6.33}$$

$$Q_{h,b} \leq N_b \cdot C_b \tag{6.34}$$

$$P^{dch}_{h,b} \leq N_b \cdot C_b \cdot \overline{PQ}_b \tag{6.35}$$

$$P^{ch}_{h,b} \leq N_b \cdot C_b \cdot \overline{PQ}_b \tag{6.36}$$

$$P^{dch}_{h,b} \leq w^{dch}_{h,b} \cdot M \tag{6.37}$$

$$P^{ch}_{h,b} \leq (1 - w^{dch}_{h,b}) \cdot M \tag{6.38}$$

Reserve requirements

$$R_h = \gamma_d \cdot D_h + \gamma_{pv} \cdot \sum_p N_p \cdot P^{pv}_{h,p} + \gamma_{wt} \cdot \sum_w N_w \cdot P^{wt}_{h,w} \tag{6.39}$$

$$R_h \leq \sum_g R^{dg}_{h,g} + \sum_b R^{sb}_{h,b} \cdot \eta_b. \tag{6.40}$$

$$\tag{6.41}$$

The model requires in input, in addition to several techno-economic parameters, the RES production and the load demand. Simulations are run along all the years of the project duration, but to make the problem treatable, only a reduced number of typical days for each year is selected N_d. The annual RES and load profiles are hence resampled according to N_d: average values are obtained for RES, while for the load typical daily profiles are extracted randomly from the ones obtained with the procedure of Block1, described in section 6.1 and increase during time to account for possible scenarios of growth. The choice of the number of days influences the weight that each simulated hour has on the total project lifetime. In particular, the parameter w_h and computed as $8760/(N_d \cdot 24)$ is used to correctly rescale all the hourly variables (e.g. fuel and maintenance costs for DG)

The function to be minimized is the Net Present Cost (NPC) formulated as in (6.12), comprising the initial investment of components, the operation and maintenance costs, the replacement cost and the salvage value. For each technology multiple types of generators, with different costs and technical characteristics can be defined with the sets (g,p,wt,ht,bess) (e.g. diesel generator with different sizes). Obviously, the wider those sets, the higher the computational effort and the risk of not convergence of the model. The O&M costs are defined in (6.14) for PV, WT and BESS as a fixed amount per year y, supposed to be encountered at the last hour of the year. The O&M costs of DG, detailed in (6.15), depend also on the fuel consumption. The costs are actualized with the yearly d_y and hourly d_h discount factors computed as: $d_y = 1/(1+r)^y$ and $d_h = 1/(1+r)^{h/(N_d \cdot 24)}$, being $\overline{H}$ the total lifetime in hours.

The replacement cost of DG and BESS is computed considering that their lifetime depends instead on their usages: the one of DG is expressed in (6.16) and depends on the working hours of the generator, assuming that H_g^{life} are the total working hours before replacement; the one of BESS, expressed in (6.17) depends instead on the number of cycles performed by the battery. The related cost becomes in this way distributed along the corresponding lifetime, instead of being concentrated at the time of the actual replacement. The other components are assumed to have a lifetime Y_i^{life} longer or equal than the microgrid project lifetime $\overline{Y}$, hence their replacement cost is not considered. For these components,namely PV, WT and HT, the salvage value is computed in (6.18). Since the DG and BESS replacement cost is addressed as a distributed cost, there is no need to consider its salvage value.

The power balance constraint at the AC busbar is reported in (6.20), where the load demand must be balanced by batteries charging and discharging, power produced by RES , by DG and by the HT. To avoid the over-sizing of the system, load shedding is typically admitted in these contexts. In particular, in the proposed formulation, the constraint is enforced to be below a given threshold of the yearly demand (see (6.21)), so that significant mismatches of unmet demand along the project lifetime are avoided through the $\overline{ENS}$ factor.

The sum of PV and WT power injected into the system (P_h^{ren}) and computed in (6.22) is at most equal to its availability, where $P_{h,p}^{pv}$ is the generation available from the PV generator of type p, and $P_{h,w}^{wt}$ is the generation available from the WT generator of type w.

The following block of constraints is devoted to defining the behaviour of DG. In particular, (6.24) describes a linear fuel consumption curve according to coefficients A and B, suitable for small size DG; (6.25) and (6.26) limit the working area of the units within the generator capacity and the minimum power $\underline{P_g}$ and consider the reserve $R_{h,g}^{dg}$ to be provided; (6.27) limits the total number of active generators.

HT behaviour is regulated by the subsequent block of equations. The maximum power producible at each time interval cannot be neither higher than the power generated by the river multiplied by the turbine efficiency eq. (6.29) nor higher than the maximum plant capacity (6.30).

The behaviour of BESS is ruled by the last groups of equations where (6.32) defines the energy level $Q_{h,b}$, limited by (6.33) and (6.34); the discharging and charging power are capped in (6.35) and (6.36) by the maximum power-to-energy ratio $\overline{PQ}_b$; (6.37) and (6.38) aim at avoiding that the batteries discharge and charge during the same time interval.

To account for the unpredictability related to real-time dispatching of the system, a reserve requirement R_h to be provided by DG and BESS is established in (6.39), proportional to the unpredictability of load (γ_d) and availability of renewables (γ_{pv} and γ_{wt}).

6.2.4 Contributions and limitations

The procedure described is particularly efficient and suited to perform pre-feasibility studies of microgrid potential in large areas. Once the hydro potential is assessed, it can be run iteratively on all the communities of a non electrified area to analyse their different suitability for microgrid investments. The integration of accurate microgrid design is not often included in large scale planning models where smaller amount of possible resources are considered (e.g. in REM and NetworkPlanner) or no sizing is performed (Onsset). Also, the procedure allows to perform multi-year planning and to include different constraints on the minimum amount of RES production.

The are however some limitations and aspects that could be improved: (i) there is no possibility of choosing between Alternate Current (AC) or DC microgrids; (ii) no feedback loop, evaluating the optimal size of the community to be supplied by a microgrid is present; (iii) biomass resources or improved BESS modeling including degradation and variable efficiencies could be included. These topics could be the focus of future research activities.

6.3 Block 3: Internal grids design

6.3.1 Introduction

The goal of the third Block of the proposed procedure is the design of the distribution network interconnecting costumers within communities. The design is based on a geospatial topological approach, which does not have the claim of performing an accurate electrical analysis, but rather to give a reliable estimation of the costs and possible paths and location of feeders and substations.

The block is divided into three modules:

1. Secondary Substations siting: the first step is the identification of the possible location of MV/LV transformers that could supply all the loads in the communities;

2. Cost surface creation: the goal of this module is to provide a realistic representation of costs for building electrical lines. With a combination of vector and raster geospatial data, the morphological characteristics of the terrain as well as the presence of infrastructures and obstacles are modelled. A penalty factor for building electric lines in different areas is computed;

3. Electric grid routing: the least cost topology of radial distribution grids interconnecting costumers is identified. The designed grid is assumed to represent the MV grid, while the LV grid is not designed for sake of simplicity.

The main steps of this block of Gisele are shown in the flow-chart of figure 6.22.

6.3.2 Module 3.1: Secondary Substations siting

The goal of this module is the identification of the optimal number and location of secondary substations. The MV grid designed in the following steps in facts does not reach in most of the cases the single households but supplies secondary substations with MV/LV that in turn supply the LV grid. Substations normally consist simply in pole mounted MV/LV transformers, so the two terms will be used in the text interchangeably.

Literature Review

The problem of optimal location and number of secondary substations (i.e. MV/LV transformers) is complex and multivariate. Distribution feeders and substations should respect geographical and technical constraints, i.e., voltage drops and loading of the conductors, while connecting costumers with adequate value of security and reliability [96].

Significant research has been carried out on methods for planning substation locations and LV networks, with interest arising as computation instruments began to be diffused in electrical design [151, 152, 153, 154, 155, 156]. These methods typically introduce a simplified structure of the problem that

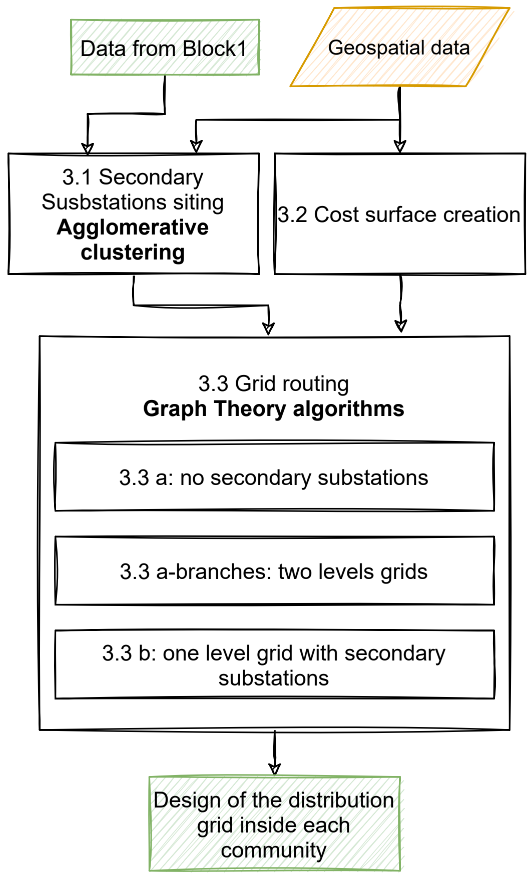


Figure 6.22: Structure of the Block3: optimal design of clusters' grid

is then used for formulating a tractable mathematical optimization problem or offer rule-based approaches. Some older methods require candidate substation locations as input.

Currently proposed planning models to site secondary substations can be divided into three main categories [157]:

- Mathematical or numerical methods (e.g., integer or mixed-integer programming, branch and bound and, network-flow programming algorithm): the models allow to reach a global optimum of the solution but often incur in convergence problems when the size and complexity of the system grows substantially [158, 159].
- Heuristic and metaheuristic algorithms: several techniques have been developed to solve problems with different sizes and characteristics. They

do not guarantee the optimality of the solution and the convergence and depend on a variety of different input parameters, but when properly set could allow to solve complex problems with good accuracy. Genetic Algorithms [160], Imperialist Competitive Algorithm [161], Particle Swarm Optimization [162], are some examples of the wide variety of algorithms and methods present in the literature.

- Unsupervised learning techniques: data analysis techniques, and in particular clustering techniques, could be an effective way to find the optimal location of distribution substations on the basis of little input data, usually just the population density [163]. The most commonly used algorithm of this category is K-means. In [157], the authors use a combination of K-means and Dijkstra's algorithm to design the LV grid. In [164] a GIS-based and Semi-Supervised Learning Algorithm based on K-means is developed, in [165] K-means with post processing techniques are used for large scale planning.

Some literature works also mix different approaches within iterative procedures in order to increase the accuracy of the results and speed up the computational process [166, 167].

Most of literature works, however, fail in addressing some of the aspects typical of rural electrification planning like the low population density, with population dispersed over wide territories and low load per capita, and the requirement of simplified approaches flexible enough to cope with low budgets and input uncertainties.

Modeling approach

Clustering approach

The approach followed to site secondary substations is based on population clustering due to several reasons. It is a scalable approach, able to find solution of problems of different sizes, usually with reasonable computational effort. There is only a little data and few parameters to be set in input and population data is available on online free databases. Electrical modeling, which could have several unknowns due to the greenfield approach, is performed in a second step, not to overload the optimal substation siting problem. Moreover, those algorithms do not choose the substation location on the basis of a predetermined set of solutions, as other techniques usually do, but are able to locate them on the basis of geometric criteria.

Following some preliminary analysis published in [96], where the author compared different clustering algorithms, specifically K-means, agglomerative and the algorithm reported in [168], agglomerative clustering algorithm has been selected to be the one most suited to the scope of secondary substations siting. It outperformed the other algorithms given its ability to include distance constraints (used as proxies for maximum LV feeders length), its reduced computational complexity and its easily to set input parameters (actually just the parameter of maximum distance is required).

Agglomerative clustering belongs to the family of hierarchical clustering algorithms [169]. It is a bottom-up approach that, starting from several clusters, one for each observation, i.e., populated point, aggregates the closest clusters until reaching a maximum number of clusters or a certain distance threshold, while minimizing clusters dissimilarities. It is necessary to measure dissimilarities among observations in order to decide which clusters to combine first, choosing a metric and a linkage criterion. In the proposed procedure, the Euclidean distance metric, i.e., the linear distance among two points, is used to compute pairwise distances among observations. The selected linkage criterion, i.e., the measure of the distance between to clusters, is the complete-linkage-clustering. This method computes the distance between two clusters (C_1 and C_2) as the distance among the two points (x,y) (each one belonging to one clusters) which are farthest away from each other:

$$D(C_1, C_2) = max(d(x, y)) \quad x \in (C_1), y \in (C_2) \tag{6.42}$$

The procedure, described in the algorithm 2 is stopped when the selected distance threshold, given in input to the algorithm, is reached.

Agglomerative clustering procedure is run for each of the communities identified in Block 1, as explained in chap. 6.1. The output clusters represent the area of influence of each secondary substation, which is then sited by computing the centroid of each cluster. The possibility of imposing a distance threshold on the created clusters avoids the risk of creating too extended clusters, incurring in high voltage drops. The distance threshold is set as twice the maximum LV line length, that is taken either from local regulations or, when not present, from literature. Whether only one cluster is found, the community will be electrified through LV lines and supplied by a single secondary substation.

Notice that only radial distances from households to the grid are considered. It is a good proxy, given that the goal of this work is the routing of the medium voltage grid and in depth analyses of the low voltage grid are not

Algorithm 2: Agglomerative clustering

```
Data: X: dataset containing all n the populated points p and their coordinates
Data: d_max: the distance threshold
Result: A set of clusters
while size (C) >1 do
    construct a cluster C_i for each object p in X; C = C_1, ...C_n;
    compute the distance matrix;
    find x,y so that D(C_x, C_y) <= D(C_i, C_j) for C_i, C_j in C;
    if D(C_x, C_y) < d_max then
        remove C_x and C_y from C;
        merge C_x and C_y into cluster C_xy;
        compute the distance matrix;
    else
        return
    end
end
```

required.

MILP based approach

Within the work published in [90], an alternative approach for secondary substations siting has been adopted. The proposed model is a MILP model with the objective of finding the location and number of transformers, so to minimize the total cost for substations and low voltage feeders while respecting the maximum LV feeder length constraint, given by local regulations and caused by technical limits.

A set of points t of the possible locations for the transformers is identified within the target area. Those points lie in a buffer of limited distance around the existing houses (given the maximum length d_{max} of LV feeders enforced by local regulations).

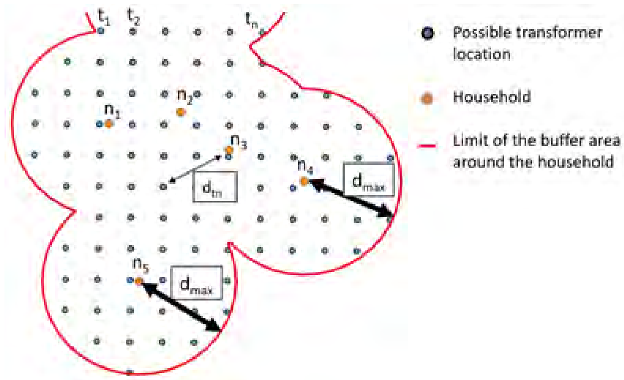


Figure 6.23: Graphical schematization of secondary substations transformer positioning according to the local regulatory framework on maximum feeder length

The MILP algorithm selects in which of those points to install the transformers, according to a least-cost criterion. In particular, the total cost C_{LV} of the LV network is given by equation (6.43), where ct is the cost of one substation t, x_t is a binary variable equal to 1 if the point is selected for the installation of the transformer, cl is the cost per meter of LV feeders and d_{tn} is the distance between transformer t and node n. tn is a bi-dimensional set containing all the possible combinations of transformers and nodes that respect the distance constraint, and are hence within the voltage reach.

$$\min C_{LV} = ct \cdot \sum_{t} x_t + cl \cdot \sum_{t} d_{tn} \cdot x_t \tag{6.43}$$

Equation 6.44 is defined for every node n and guarantees the connection of every household to a transformer, by stating that each node n must have at least one transformer t within its distance buffer radius.

$$\forall n \sum_{t} d_{tn} \cdot x_t \geq 1 \tag{6.44}$$

The following Insight box provides a detail of the application of the two approaches in the area of Omereque, Bolivia, described in chap. 5. It underlines how agglomerative approach performs much better in terms of computational effort and reduced complexity, without significantly losing in accuracy. It is hence the algorithm of choice for the complete Gisele procedure.

Insight

The algorithms for secondary substation siting were applied to the rural area of Omereque in Bolivia, where the location of unelectrified households was provided as input by the local NGO. In the area regulations required a maximum LV feeders length of 600m.
Figure 6.24 shows the input data for the MILP algorithm and the output results of the model. The final number of identified transformers depends on their relative cost with respect to the LV line cost.

Figure 6.24: Results of MILP for MV/LV transformers siting

(a) Possible location of transformers within 600m buffer radius

(b) Optimal transformers placement

Agglomerative algorithm is also run on the same area to compare the two approaches and the results are reported in 6.10

Table 6.10: Comparison between MILP model and agglomerative clustering for secondary substations siting

Parameter	MILP	Agglomerative
N cluster	51	41
Households at distance ¿ 600m	0	4%
Computational time [s]	0.29	0.002
LVlines cost	176.9	216
Trafo cost [k$]	178.5	143.5
Tot cost [k$]	355.4	359.5

> The number of transformers selected by the MILP model is higher than the ones selected by the agglomerative clustering; this allows to have all the houses within the maximum distance of 600m from the substation. Agglomerative clustering instead causes the 4 % of the houses being further away from the substations. The total cost, computed as in equation (6.43), is lower for the MILP since it is its objective function, however, the percentage difference is just of 1%.
> The reasonable errors, the much lower computational time and the easier input preparation, made the agglomerative algorithm being chosen to be integrated inside Gisele procedure.

6.3.3 Module 3.2: Cost surface creation

The goal of this module is to aid in the design of a realistic grid topology, able to consider all the geographical and morphological characteristics of the territory as well as the presence of infrastructures. Through the creation of a cost surface, electric lines, designed in the third module, will not follow the shortest path (i.e. shortest linear distance between substations) but will instead follow low cost corridors, typical along roads.

Literature review

The cost of installation of new electric feeders depends on the characteristics of the terrain on which they are built. Preferably, new medium voltage lines are built along roads, given the ease of access and the reduction of costs related to private property rights. The cost of lines deployment, given by human labour and equipment for installation, as well as operation and maintenance costs, depends on the accessibility of the area, increase as the distance from roads increases and in presence of difficult terrains (e.g. high slopes, presence of forests). Finally, there are areas where it would be impossible or forbidden to install lines such as lakes, large rivers or national parks and protected areas. A way to include all this aspects into the electric grid routing problem, is to create a cost surface, where a different cost (or weight), is associated to each type of terrain [170].

In literature, two main approaches to model the real surface are followed: vector models and raster models [171].
Vector models subdivide the area into homogeneous polygonal regions, associating to each of them a different unitary cost. Edges between different

regions (that could for instance represent roads), can also have costs associated. The least cost path between two points in the region is calculated using the law of light refraction, which defines the optimal angle of a line crossing two regions [172, 173]. The main advantages of this type of approach are the possibility of detecting the exact optimal paths of the lines, and the relatively low amount of data that needs to be stored. On the other hand, however, the discretization of the surface, although useful to represent elements with defined boundaries (e.g. lakes or protected areas), leads to more approximations when representing continuous characteristics (e.g. variation of the elevation). Finally, this approach requires complex and time consuming algorithms that may not be suited for the analysis of wide areas. Vector models could be greatly simplified by modeling only the road graph, thus representing the surface with a network of edges, that are the only allowable path. The cost of lines depends in this case only on their length. This approach is fast and simple and typically used for routing lines in urban areas, where feeders cannot pass over buildings [174]. In rural areas, however, feeders do not always follow the roads path, and sometimes complete data related to the full road graphs are not even available.
The complementary approach to vectorial representation is the use of raster models. In this case, surface is represented by a matrix of pixels with a defined resolution and a cost is associated to each of them. The total cost for connecting two points is given by the sum of the costs of the pixels the line is crossing. The least cost path can be found by graph theory algorithms such as Dijsktra or A* that will later be described [175]. This approach is simple and quite efficient but could lead to some distortions in shapes representation given by the discretization of spaces: this could be evident for instance in the representation of roads or rivers. To increase the accuracy of the model, resolution must be increased, thus leading however to and increase in memory requirements and computational effort. A solution to this issue is described in [170], where the authors proposed to model the raster surface with an irregular grid of points. A degree of importance is associated to each pixel, and only the ones with higher values would be used to actually represent the surface, thus avoiding having a dense grid of pixels with homogeneous values. Different strategies could be employed to identify the most important points of a mesh. The cited paper proposed the Very Important Point algorithm: the value of importance is computed by computing the distance of the pixel value (its costs) to the one of the surrounding pixels, moreover, only the pixels with higher importance, hence significantly different from the surrounding are used to represent the terrain. What is less straightforward in the method

is the way in which to compute the shortest path between points that could be subject to many approximations. An alternative to increase accuracy of raster models is the hybridization with vector models, that is the representation of some vectorial features (typically roads) on top of a continuous raster surface [176].

Modeling approach

The composition of raster and vector models is the approach used in this work, since it is a good compromise between accuracy and simplicity. More specifically, the final cost surface provided in input to the grid routing procedure is given by a point vector layer composed of:

- regular grid of points, that is equivalent to a raster surface, since each point represents the centroid of a pixel of a raster layer, with a penalty factor associated to it (*GridPts*);
- additional points, sampled along the roads lines network, with unitary cost associated (*RoadPts*);

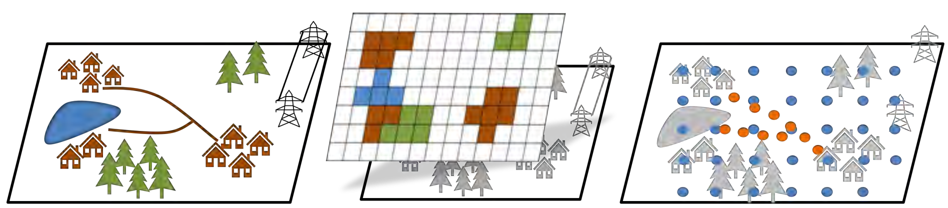


Figure 6.25: Schematization of the cost surface creation. In blue the regular grid of points and in orange the additional road points

The penalty factor associated to each *GridPts* is the result of the combination of different characteristics, gathered by sampling information from different geospatial layers.

- Distance from roads: the weight increases linearly up to a certain threshold when it is at is maximum value; the distance is computed as the euclidean distance of each point from a roads vector layer;
- Slope: the weight increases exponentially with the slope; slope is computed from the DEM raster layer and sampled at each point;

- Protected Areas: weight is much higher than the others to avoid crossing of those areas; protected areas are usually found in the form of polygon vector layers;
- Lakes: weight is much higher than the others to avoid crossing of lakes;
- Land Cover: weight differs according to the type of land cover (e.g. forests have a higher weight than agriculture land); land cover values in each point are sampled from land cover raster layers.

Also, to realistically associate a penalty factor to the cost of the lines deployed, their total cost is split into the voices of materials, works and permission. Each of them is affected differently by different types of surface. To give and example, cost of works, that is e.g. man hours of work and fuel for transportation, is significantly reduced by the presence of roads, while the cost of conductors could increase greatly when insulation for underwater cables is needed. Table 6.11 shows in detail the coefficients considered. Their numerical values have been derived from a combination of data from real-life grid-expansion projects, mainly related to the repartition of costs into the different categories and literature works and authors' assumptions for the weighting coefficients.
The penalty factor pf associated to each point when considering only Capital Expenditure (CAPEX) of lines deployment is given by:

$$pf = \sum_{i=1}^{5} (inv_i \cdot (1 + w_i)) \tag{6.45}$$

where inv_i is the percentage of investment cost associated to each of the five categories i (Material conductor, material poles, material additional, works, permission) and w_i is the weighting coefficient given by the combination of all the territory characteristics (land cover, roads, etc.) If also Operational Expenditure (OPEX) needs to be considered, to compute the total NPC of lines deployment, the penalty factor associated to each point is given by the following equation, where r is the average project discount rate:

$$pf = \sum_{i=1}^{5} (inv_i \cdot (1 + w_i)) + \sum_{y=1}^{\overline{Y}} (O\&M_y \cdot (1 + w_{O\&M}))/(1 + r)^y \tag{6.46}$$

The additional points, sampled along the road network ($RoadPts$) have a weighting coefficient of 0 for each of the cost categories.

Table 6.11: Coefficients used for the cost surface creation

	Material -cond	Material-poles	Material_add	Works	Permission	O&M
Tot cost	0.18	0.36	0.06	0.22	0.18	0.02
LandCover						
1 Trees cover areas	0	0	0	4	-0.5	4
2 Shrubs cover areas	0	0	0	1	0	1
3 Grassland	0	0	0	0	0	0
4 Cropland	0	0	0	0	4	0
5 Vegetation aquatic	0	2	0.1	4	0	4
6 Sparse vegetation	0	0	0	1	0	1
7 Bare areas	0	0	0	0	0	0
8 Built up areas	0	0	0	0	1	0
9 Snow and/or Ice	0.2	2	0.1	2	0	2
10 Open water	10	0	10	6	0	6
Distance from roads [m]						
<100	0	0	0	0.1	0	0.1
<500	0.2	0	0.2			
600	0.7	0	0.7	0.1	0	0.1
Slope						
<2°	0	0	0	0	0	0
>2°and <5°	0	0.1	0	0.1	0	0.1
>5°and <10°	0	0.5	0	0.5	0	0.5
>10°and <20°	0	2	0	2	0	2
>20°	0	10	0	10	0	10
Protected areas [yes]						
100	0	10	0	10	0	10
River [cms]						
>100	0	10	2	10	0	10

6.3.4 Module 3.3: Grid routing

After the preliminary steps useful to identify candidate positions of secondary substations and the cost for deploying electric lines, the grid routing procedure is adopted. In this module, only the grid interconnecting users within the communities is designed, so the procedure is run iteratively on each of the communities and the output is a number of radial grids equal to the number of communities. The definitive grid topology, also interconnecting the communities to the in place grid and between each other will be designed in the last part (Block4) of Gisele procedure. Splitting the grid routing into two steps allows to reduce the computational burden and treat larger areas.

Literature Review

Grid routing is the task of identifying the optimal path of new electric lines able to connect new costumers in a region. The objective is generally the minimization of costs while respecting electrical constraints such as the voltage drops and current limits. The optimization problem is commonly solved in literature using one of the three approaches [73]:

- Graph Theory based algorithms: they can be used to find least cost paths and trees within weighted graphs not allowing, however, the possibility to include further constraints. They have the advantage of being simple and requiring small computational effort [177];

- Heuristic and methaeuristic algorithms: several algorithms belong to this class (Genetic Algorithms, Particle Swarm Optimization, Tabu Search). They do not provide the exact solution of the problem and have the risk of incurring into local optima, but they have the advantage of modeling also non linear constraints typical of load flow computations[74];
- Numerical exact methods such as mixed integer MILP, NLP: they have the advantage of finding the global optimal solution, at the expense of the approximation of the problem due to linearization (in case of MILP) or of high computational complexity and convergence issues (NLP)

When trying to adapt literature algorithms to the problem of greenfield rural electrification planning, the following gaps are identified:

1. algorithms are typically designed and tested for urban areas optimization;
2. number of variables and constraints are so high that models are tested only on small areas, with few number of nodes;
3. few works consider the real terrain topology to design electric grids;

Among the tools specific for rural electrification planning and described in the first chapter, Network Planner uses a simple graph theory algorithm (the minimum spanning tree), to find the minimum length of feeders' lines, [83] uses an approach based on graph theory algorithm and cost surface, while REM uses RNM, which is based on heuristic algorithms.

Modeling approach

For designing the grid inside the clusters, a graph-theory based approach is used. Within this phase, priority is given to the simplicity and replicability of the algorithm, which has to be applied multiple times in areas composed also by several loads without incurring in solver issues. Moreover, inside each community, distances are relatively short and likely no voltage or current problems could arise. This approach can provide a good approximation of the cost for deploying a grid within communities without the necessity of long and heavy computations. Electrical consistency of the grid is evaluated a posteriori, leading to a manual redefinition of communities' boundaries whether some constraints cannot be satisfied with standard cables and voltage levels.

The goal of the proposed procedure (module 3.3) is the creation of a radial grid, connecting all the loads at the minimum cost. The model starts from the conceptualization of the grid as weighted a graph with nodes (loads) and edges. To identify the load points, during the research activities two approaches have been investigated:

- 3.3-a: within the first works, to each point of the regular grid of points forming the cost surface, was also associated the value of the population living in the surrounding area (sampled from raster data). In this case load nodes are all $GridPts$ with a population overcoming a certain minimum threshold $MinPop$[88];
- 3.3-b: as an improvement, the procedure described in section 6.3.2 for the optimal siting of secondary substations was developed. Load nodes are in this case all the secondary substations points.

The simplest approach that can be applied to design a radial grid is the computation of the Minimum Spanning Tree (MST), defined as:

Definition 5 *A Minimum Spanning Tree is a subset of the edges of a connected, edge-weighted undirected graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight.*

In the problem under consideration, the weights of the edges could be represented by their linear or 3D distance, the parameter to be minimized to reduce costs, and nodes are the loads. Two well known heuristic algorithms (Prim's and Kruskal's) allow to solve the problem, arriving at the optimal solution in a time dependent on the number of nodes and vertices [178], [179].

This approach comes naturally with many simplification and results to be too simplistic to be applied to the problem under consideration. In particular, nodes are connected with straight lines, that do not consider the characteristics of the terrain and the cost surface associated. Two nodes would be connected if they are close enough even in presence of physical obstacles (such as lakes) between them.

To solve this issue, the cost surface described in the previous section is used. The cost surface describing the territory, is transformed into a weighted *graph* $G = (V, E)$ where V are the vertexes (the pixels' centroids) and E are the edges connecting a couple of nodes (i.e. all the possible connections). The edges composing the graph could be: (i) the ones that connect each vertex to its eight neighbours as shown in figure 6.26; (ii) edges connecting two of the additional points sampled along the roads *RoadPts*; (iii) edges connecting one of the additional road points with a point on the regular grid if their distance is lower than a certain threshold.

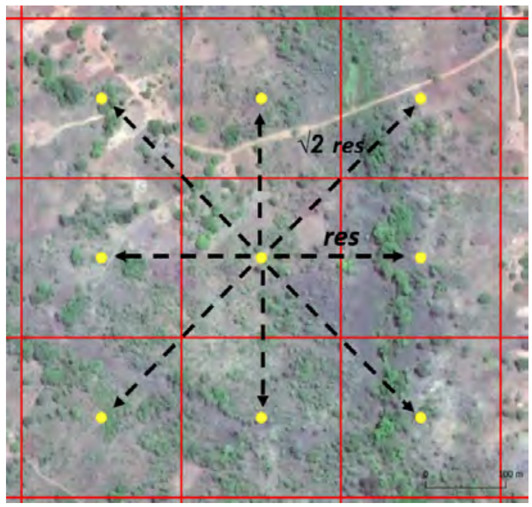


Figure 6.26: Allowed grid connections

The weight of a edge connecting nodes i and j represents the cost of the electric line connecting the two vertices:

$$W_{ij} = \begin{cases} L_{ij} \cdot UC \cdot \frac{pf_i + pf_j}{2} & \text{if } L_{ij} < \sqrt{2} \cdot \text{res or i,j} \in RoadPts \\ \infty & \text{if } L_{ij} > \sqrt{2} \cdot \text{res} \end{cases} \tag{6.47}$$

Which considers the distance between the extremes L_{ij}, the unitary capital cost of electric line UC $\$/km$ and a mean *"penalty factor"* between those related to the extremes i and j. *res* is the resolution of the regular grid of points and $\sqrt{2}\cdot$ *res* is the length of the diagonal (i.e. the max allowed distance). Building edges with a longer distance would not allow to have an accurate representation of the line cost, because obstacles or morphology changes in the middle would not be taken into account.

Given the model of the terrain described previously, and two nodes to be connected, their least cost distance is provided by the following definition:

Definition 6 *The least cost distance between two nodes is the single path through the space between a given source location and a destination location that has the least total accumulated cost.*

The shortest path problem on top of raster maps is widely addressed in GIS related literature [170, 180, 181] with the well-known Dijsktra algorithm [182] being widely used.

However, the solution of the shortest or minimum weight path between a source and target node, is not enough to solve the problem of optimal grid routing, not considering the creation of a minimum cost spanning tree. The problem under consideration could be actually better described by what is called the Steiner Tree problem

Definition 7 *Given an undirected graph with non-negative edge weights and a subset of vertices, usually referred to as terminals, the Steiner tree problem in graphs requires a tree of minimum weight that contains all terminals (but may include additional vertices).*

A generic example of Steiner tree is shown in figure 6.27.

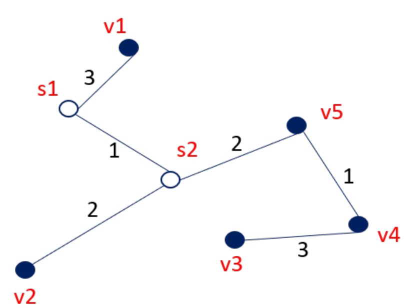


Figure 6.27: Steiner tree with Steiner nodes (S) and terminal nodes (V)

The nodes, divided in *terminals* and *non-terminals* represent respectively the nodes that must be included in the solution, and the ones that could be part or not of the final tree, according to their convenience. Terminal nodes, are the load points, found with one of the two strategies previously described (3.3-a or 3.3-b). Non terminal nodes are instead constituted by all the cost-surface points, both the *RoadPts* and the *GridPts*. This problem is classified as NP-complete and the solution can only be approximated to have viable computational effort. In literature, several algorithms are proposed to address the problem, nevertheless those approaches require a consistent amount of memory, therefore are suitable only for small graphs involving a limited amount of terminal nodes [183, 184, 185]. Two of the most known algorithms are the graph Iterated 1-Steiner approach [186], with a time complexity of $O(V \cdot L + L^4 logL)$ and the MST approximation, computed starting from the metric closure of the graph, with a time complexity of $O(V \cdot L^2)$, where V is the total number of vertexes and L the number of terminal nodes [187].

Definition 8 *The metric closure of a graph G is the complete graph in which each edge is weighted by the shortest path distance between the nodes in G.*

The MST approximation procedure, here on also called M1, is described by the pseudocode 3.

Algorithm 3: M1 Steiner tree-MST approximation

Data: $G = (V, E, w)$: a weighted graph
Data: L: a terminal set, $L \subset V$
Result: A Steiner tree T
1 Construct the metric closure G_L on the terminal set L;
2 Find a MST T_L of G_L;
3 $T \leftarrow \emptyset$;
4 **for** *edge* $e = (u, v) \in E(T_L)$ *in a depth-first-search order of* T_L **do**
5 Find a shortest path P from u to v on G;
6 **if** P *contains less than two vertices in* T **then**
7 Add P to T;
8 **else**
9 Let p_i and p_j be the first and the last vertices already in T;
10 Add subpaths from u to p_i and from p_j to v to T;
11 **end**
12 **end**
13 output T;

To construct the metric closure and the shortest path between two points (step 1 and 5), Dijsktra algorithm is used.

Steiner tree approximation algorithms inspired many other literature works but none of them, according to the authors' knowledge, applied the procedures to least cost raster surfaces, characterized by high number of vertices and edges. To overcome computational limits and be able to provide a solution even for large clusters, the author developed an iterative procedure which combines the potentialities of Kruskal and Dijkstra algorithms, following an approach similar to the interesting work of researchers at Reiner Lemoine Institute (RLI), where the cost surface is used to build the optimum connection pathways of a radial grid [83, 188].

The new greedy approach (aslo called M2) is then applied as an alternative to the MST approximation, resulting the preferred solution in case of large graphs or grids that require multiple branches. In the proposed approach, an initial, rough model of the electric connections is initially traced by computing the MST of the overall set of cluster's populated points. An analysis of the MST solution is then performed, making a distinction between *ShortLines* connecting neighbouring points ($L_{ij} < \sqrt{2}\cdot\text{res}$) and *LongLines*. *ShortLines* are already well defined since they connect neighbouring points and their cost can be estimated by equation (6.47). Therefore they become part of the final tree. *LongLines* instead need to be fragmented into a sequence of adjacent elementary edges in order to analyze their cost. To do so, the terminals i and j are respectively set as source and target for the Dijkstra algorithm which returns the *least cost path* connecting them. During each iteration the algorithm is designed to continuously store the newly created connections updating the corresponding edges cost to a value of zero: this makes it capable of recognizing already built paths which, if crossed, would lead to no additional costs. Once all *long lines* have been fragmented and re-structured the algorithm stops: the final solution is a tree, connecting all the terminal nodes. This tree will only be made of elementary edges, each one with an associated cost, which algebraically summed gives the total network cost. A logical scheme of the procedure is reported in the pseudocode 4 and a graphical exemplification is shown in figure 6.28.

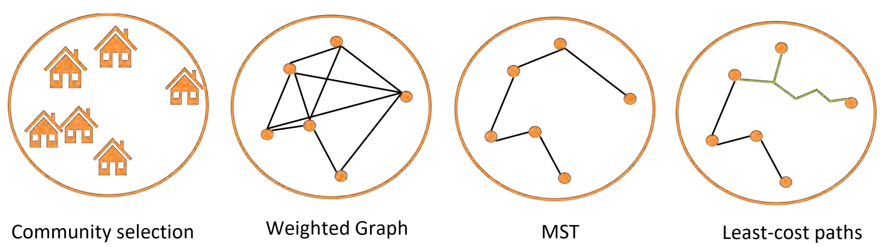


Figure 6.28: Schematic representation of the steps of M2

For each cluster, if the number of nodes allows it and does not create memory issues, the MST approximation (M1) and the new algorithm proposed (M2) are compared to find the least cost Steiner tree. Otherwise, only the MST + Dijsktra algorithm is used.
As shown in the example of figure 6.29, the convenience of one algorithm with respect to the other depends on the case study. The two sub figures are related to two different case studies, with a different disposition of terminal

Algorithm 4: M2 MST + Dijsktra

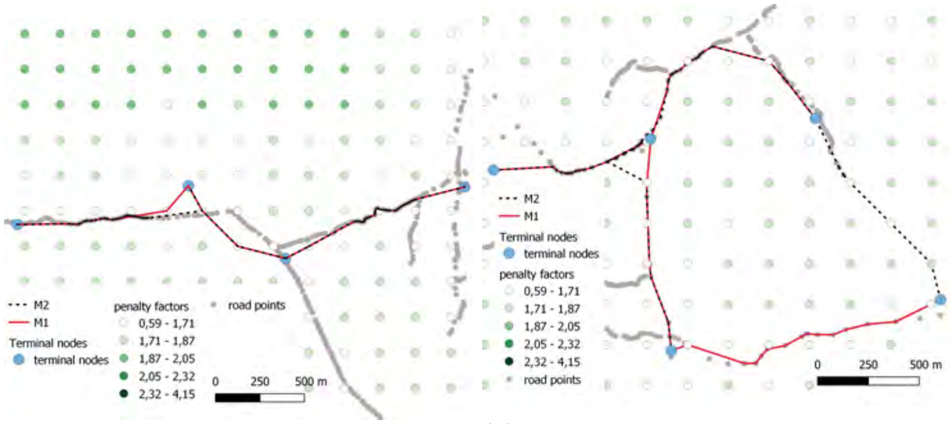
Data: $G = (V, E, w)$: a weighted graph
Data: L: a terminal set, $L \subset V$
Result: A Steiner tree T

```
T ← ∅;
Construct the MST T_L of G on the terminal set L;
for edge e = (u, v) ∈ E(T_L) do
    if len(e) <= √2 · res then
        e = ShortLine;
        update G: w(e)=0;
        add e to T;
    else
        e = LongLine
    end
end
while #(LongLines) in T_L > 0 do
    pick shortest long line e′ = (u′, v′) ;
    find shortest path P from u′ to v′ on G;
    add P to T;
    add edges of P to G with w=0;
    remove e′ from T_L;
end
```

nodes (blue points) in two areas of the territory where $RoadPts$ are shown in grey and $GridPts$ are coloured with different shades of green according to their penalty factor. In the case 1, M2 is able to create a grid with lowest cost, branching the feeders in a non terminal node, while in case 2 it is M1 with a lowest cost since it exploits the road at the bottom of the figure. To summarize, the Steiner MST +Dijsktra (M2) algorithm has the disadvantage of being iterative, losing the opportunity of globally optimizing the graph, but it allows to create ramifications from graph edges when they are convenient. Also, it allows to significantly reduce the computational time.

3.3-a-branches

With respect to method 3.3-a, that considers as terminal nodes all the points of the regular grid having a total population that overcomes a certain threshold, a two steps procedure has also been investigated. The goal is to design a grid with two different types of cable, and with a hierarchic structure: a main feeder constituting the backbone of the grid and the collateral cables, originating from the main feeder and connecting all the users.



(a) Case 1: MST + Dijkstra better cost

(b) Case 2: Steiner MST approximation better cost

Figure 6.29: Example of comparison between M1 and M2 for optimal grid routing

The first step to create the main branch is to lower the resolution of the input regular grid, therefore reducing the number of points. At a lower resolution each point covers a larger area and consequently a larger population. A new higher population threshold is defined to identify a new set of terminal nodes (L2). This must be done carefully, since an over-reduction of the resolution would create a too-simplistic grid, in which clusters could end up having a single or no terminal nodes. On the other hand, if the resolution is not reduced enough, the main branch grid would end up connecting most of the terminal nodes making the creation of collaterals irrelevant. Also, the choice for the population threshold used, which determines the terminal nodes of the main branch to be connected, adds another layer of complexity to this trade-off. In order to keep a sufficient level of detail,resolution is not reduced over one third of the area of the smallest cluster. In this way it is assured that every cluster would be delimited by at least three internal nodes, keeping a minimum amount of detail for the analysis.
After selecting the terminal nodes in the low resolution grid, the Steiner tree is computed, with one of the two methods (M1 and M2) previously described. It could happen that clusters localized in sparse areas, even lowering the resolution, might not have sufficient people to require a high power line. In these cases, no main branch is created, and the internal cluster grid will be a standard Steiner tree composed of a single collateral. Fig. 6.30 shows an example of the routing algorithm for creating the main branches. There are two grid of points of different resolutions, one lower (light blue points) and one higher

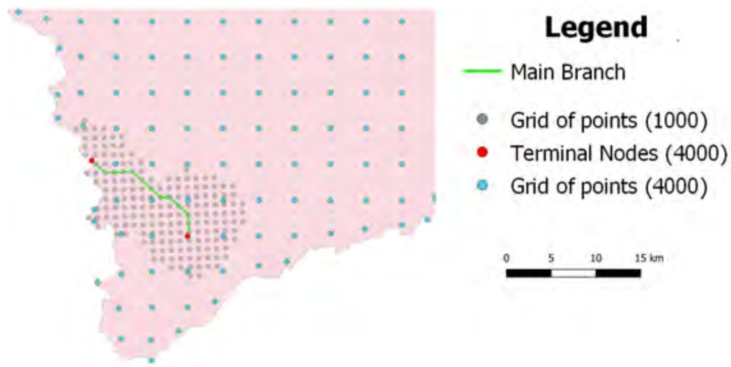


Figure 6.30: Example of main branch creation using resolutions of 1000 and 4000 meters

delimiting the clustered area (grey points) also called the cost surface. The lower resolution grid of points is used to find the densely populated areas, or the terminal nodes (red points), that the main branch will connect. Once defined, the terminal nodes are correspondent to the nearest point in the cost surface. Even in the creation of the main branch, the high resolution cost surface is used for the Steiner tree computation since it permits the creation of a more detailed and accurate least-cost path. Once the algorithm finishes, a main branch will be created connecting all the densely populated areas with the least-cost path based on the cost surface.

In the second step, the Steiner tree algorithm is run again, but assigning to every edge e of G that also belongs to a main branch, a weight $w(e)$ close to zero. In this way the grid routing algorithm will create a Steiner tree of the terminal nodes exploiting the already in place main branch, creating the collaterals. For clusters that do not have a main branch, the grid routing algorithm will have no changes with respect to the standard approach, and will compute a Steiner tree connecting all the populated points. Figure 6.31 exemplifies the topology of the grid that can be obtained with the one step approach (3.3a) on the left and with the branches approach (3.3a-branches).

This two steps method (denominated 3.3-a-branches), is particularly interesting when the grid to be designed must arrive to the final users, without intermediate substations. This was the case in Cavalcante region of Brazil, described in the following example box and in the journal paper [89]. In general, method 3.3-b, with the localization of secondary substations, is the one used and proposed in the updated version of Gisele.

Algorithm 5: 3.3-a-branches

```
Data: G = (V, E, w): a graph
Data: L ⊂ V: a terminal set
Data: MinPop: the minimum population threshold
Result: Two overlapping Steiner trees T1, T2
Create P: a low resolution grid of populated points;
for p in P do
    if population p > MinPop then
        add p to the new terminal set L1;
    end
end
create T1 applying procedure M1 or M2 to (G = (V, E, w),L1);
add edges of T1 to G with w=0;
create T2 applying procedure M1 or M2 to (G = (V, E, w),L)
```

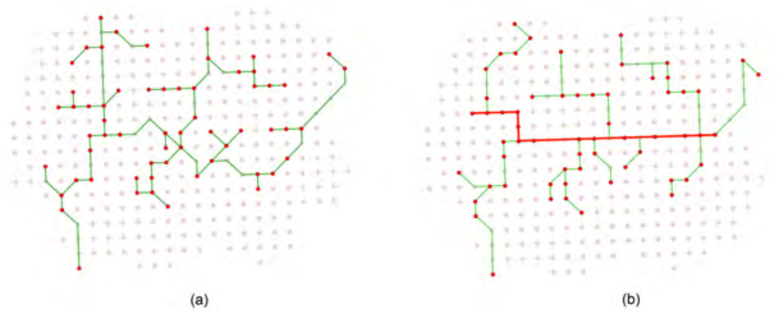


Figure 6.31: Example of (a) a simple least-cost topology and (b) a least-cost topology considering a hierarchical structure composed of a main branch and collaterals

Insight

Brazil-Cavalcante The procedure 3.3-a-branches has been developed in collaboration with Enel Infrastructure and Networks and tested on the area of Calcante in Brazil. In that area, the standard electrification approach does not include LV lines since the MV distribution grid directly supplies small agglomerates of households. There is however a hierarchical structure of the grid with main three phase branches covering long distances and single-phase collaterals reaching single costumers.

High resolution cost surface was created with 1000 m resolution, the highest available detail of the population density. Population was then resampled at 4000m resolution (see fig. 6.32) and main branches where created considering as terminal nodes all the points with a minimum population of 30 people. Collaterals were instead designed with a population threshold of 1, to connect all the people in the areas.

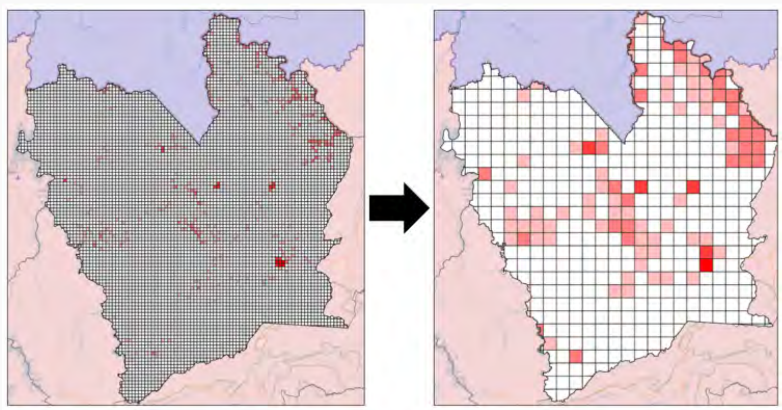
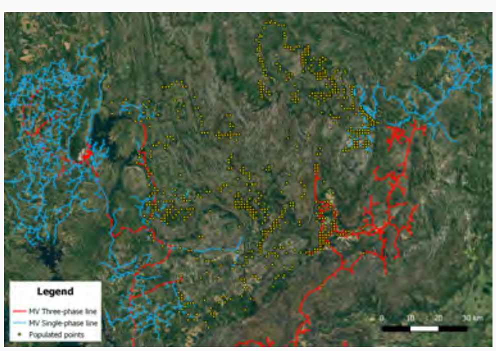


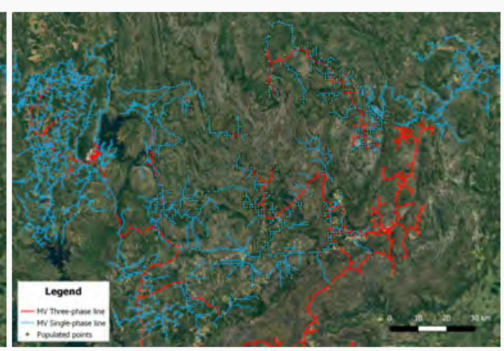
Figure 6.32: Creation of a low resolution populated grid of points

The procedure allowed to have a grid expansion design similar to the one already in place (see fig. 6.33a), where collaterals represent the single-phase lines and the main branches are the three phase systems.

Figure 6.33: Distribution grid expansion in Cavalcante region



(a) Distribution grid before the expansion



(b) Distribution grid after the expansion

Table 6.12 shows a comparison of the results obtained with the one-step procedure 3.3-a and the two step procedure 3.3-a-branches. Given the possibility of differentiating between the cost of collaterals (3000 $/km) and main branches (10000 $/km), the second procedure leads to a more realistic cost estimation, reducing total cost with respect to 3.3-a (where cost 10000 $/km for all the lines) despite increasing total network length.

Table 6.12: Comparison between grid routing procedure with and without main branches

Cluster	3.3-a		3.3-a-branches	
	Length [km]	Cost [k$]	Length [km]	Cost [k$]
1	51.87	2531.23	73.24	1283.67
2	233.01	12720.90	310.47	6887.64
3	75.65	2360.60	84.73	888.27
4	93.73	5221.42	93.73	1740.3
5	41.17	2302.76	41.17	767.51
6	32.6	2026.41	32.6	1193.69
7	36.21	1881.47	49.45	1444.74
8	312.12	17900.74	472.7	10864.8
Total	**876.36**	**46945.55**	**1158.09**	**25070.6**

6.3.5 Contributions and limitations

The Block3 of the proposed procedure allows to design least-cost distribution grids based on topological approaches. It has the following advantages:

- It allows to perform large scale analysis.
- It requires a minimum amount of input data; mainly open-source geospatial data easily and automatically downloadable.
- It considers the geographical characteristics of the territory, feeders' routing exploits the presence of roads and avoids obstacles.
- It provides a reliable cost estimation of distribution grids inside communities.

The procedure has two main limitations: (i) it does not design LV grids; (ii) it does not consider electrical constraints in the design phase, neither for substations nor for feeders' routing. Further studies already going on in the research group are focusing on the improvement of the design of the low voltage grid and on consequent optimization of secondary substation siting and sizing.

6.4 Block 4: Integrated area optimization

6.4.1 Introduction

The last block of the proposed modeling framework has the goal of identifying the optimal electrification strategy for each community choosing between the connection to the in place grid or the electrification through a hybrid microgrid. While the graph-based approach was effective for internal clusters grid routing because it is simple and straightforward, for the integrated area optimization this type of approach is not well suited, since it would not allow to choose between different electrification solutions and include additional constraints. For this reason different strategies have been investigated, and a MILP and GA based models have been developed and compared. In addition, a MILP based multi-objective optimization algorithm has been developed to consider, aside the minimization of costs, the environmental and technical dimension.

The main steps of this block of the procedure are shown in the flow-chart of figure 6.34.

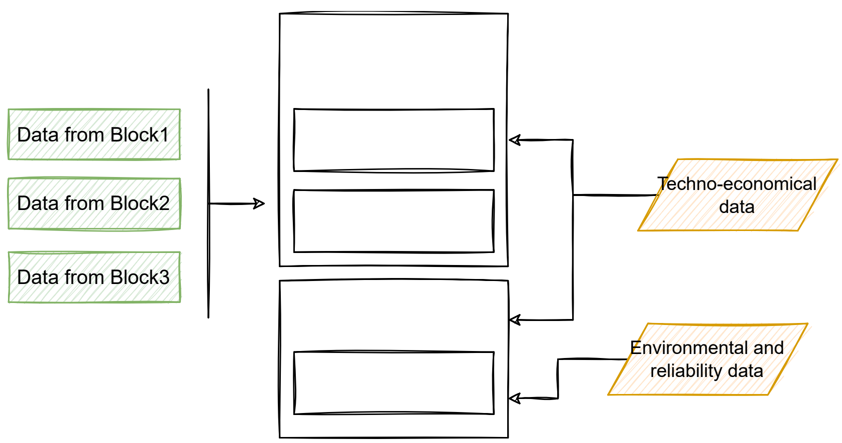


Figure 6.34: Structure of the Block4: integrated area optimization

The integrated optimization is able to design the new distribution grid topology while at the same time selecting the optimal electrification strategy. Figure 6.35 shows a simplified representation of the concept behind the block: starting from already identified communities (colored houses), a weighted graph is created and through the optimization algorithms only the convenient connections are kept: isolated communities will be electrified through off-grid systems.

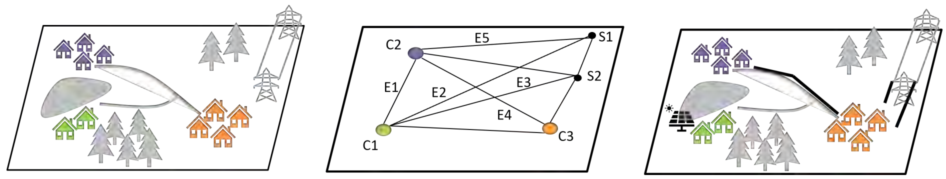


Figure 6.35: Schematization of integrated area optimization

6.4.2 Module 4.1: Single-objective optimization

Literature Review

The goal of this module is to perform an integrated optimization, comprising the design of the electrical grid. Some of the different literature modeling algorithms for grid routing reported in chapter 6.3 have been applied also for the discrimination between on and off-grid solutions: the authors in [189] propose a Mixed Integer Non Linear Programming (MILNP) model, while in [190] heuristic models are used.

Comprehensive tools described in chapter 4 adopt different approaches for the choice of the optimal electrification solutions. [83] makes a cost comparison of mini-grid electrification and grid extension for each community and distinguishes the electrification process in three phases where different criteria of choice for the solution are adopted. REM performs at first a bottom-up greedy procedure for clustering costumers and decides iteratively which to connect to the grid and which to electrify with off-grid systems. In a second step, through the use of RNM, the electric grid is designed. Network Planner also performs a cluster by cluster analysis, using the heuristic algorithm based on MST described in [190] and [191], to iteratively interconnect non electrified settlements with a radial grid. GEOSIM bases its analysis on geographical distances, selecting areas where densification of the grid should be put in place, areas for grid extension and potential areas for off-grid systems. None of literature approaches, according to the author's knowledge, performs an integrated optimization, where all the solutions are evaluated contemporaneously.

Input data preprocessing

As for the internal grid routing procedure, the area is represented as a weighted graph, with nodes and edges:

- Nodes: the communities (represented by their centroids) and the points of connection to the existing grid; those could be existing primary substations (HV/MV), MV/MV conversion substations or direct connections to inplace MV feeders;

- Edges: the links between communities and with the existing grid.

To simplify the optimization procedures, the created graph is not complete, i.e. only some of the possible connections are preselected as optional edges. This is done with the help of the Delaunay triangulation:

Definition 9 *A Delaunay triangulation for a given set P of discrete points in a general position is a triangulation DT(P) such that no point in P is inside the circumcircle of any triangle in DT(P). Delaunay triangulations maximize the minimum angle of all the angles of the triangles in the triangulation.*

The steps followed to identify the graph edges can be listed as:

1. Delaunay triangulation: Delaunay triangulation is performed to connect all the communities (see fig. 6.36);

2. Removal of long lines: Delaunay edges that overcome a maximum distance threshold are discarded;

3. Cost of connections: Dijsktra algorithm is applied to each couple of nodes (communities), connected by a Delaunay edge to compute the real connection cost. The cost surface extensively described in sec. 6.3 is used at the scope;

4. Removal of expensive lines: edges that overcome a certain cost threshold are discarded;

5. Connection to inplace grid: each community is also connected through a least-cost path to the three (or less if not available) closest connection points;

The created edges, with their length and cost associated, are the binary decision variables of the problem. The models will select only some of the edges, to interconnect clusters and grid at the minimum cost.

Input to the model are also the distribution lines electric characteristics, the internal grids created in sec. 6.3, the peak power associated to the secondary substations, the costs of off-grid systems and the cost of energy supplied by the national grid.

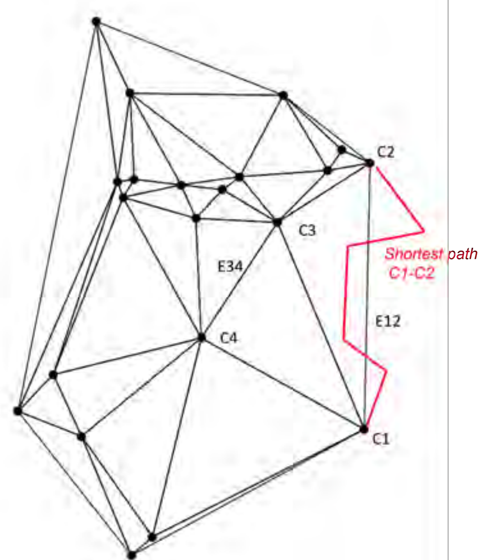


Figure 6.36: Schematization of Delaunay triangulation and computation of edges shortest path cost

Module 4.1.1: GA-based optimization

Heuristic and metaheuristic algorithms, as previously described, have the advantage of being able to simulate non linear problems, since the penalty function can have any shape and are therefore well suited to model power flow constraints in electric network design. Among the different heuristic algorithms developed in literature for grid planning, genetic algorithms are one of the most promising and scalable to different problems, given their low dependency on initial solutions and parameters tuning [74]. They are also well suited for dealing with mixed integer problems, as demonstrated in [192] and [193], which is the type of problem that has to be faced in this module, given the objective of selecting which edges to keep and which to drop. Those are the main reasons why an approach based on genetic algorithms has been developed and investigated.
The algorithm has been coded in Python with the help of available libraries *pandapower* [194] and *geneticalgorithm*.

The main structure is an electric grid modeled and solved adapting the pandapower package, here in after defined a *pandapower network* where each community is represented by its whole internal grid: the set of secondary substations form the load buses (P-Q buses) and the radial grid previously designed constitutes the network lines. The possible points of connection to the existing grid are modeled as external grid buses (that are slack buses). The decision variables are the connections between communities and with the grid connection points. At each iteration of the GA, some of those connections will be activated and the corresponding line in the network will be created. At this point a loadflow is run to check if electrical constraints are respected. If some communities are not connected to any external grid buses, they are considered unsupplied buses and the load flow there is not computed. Figure 6.37 shows on the left the pandapower network built starting from the internal grids designed in Block3. Each separated grid is one community. The yellow square is the possible point of connection to the in place grid. On the right the results after a first genetic-algorithm iteration are shown. Load flow is computed on the new grid that interconnects some communities between each other and some with the in place grid. The unsupplied buses are grey since load flow was not computed.

Chromosome coding:

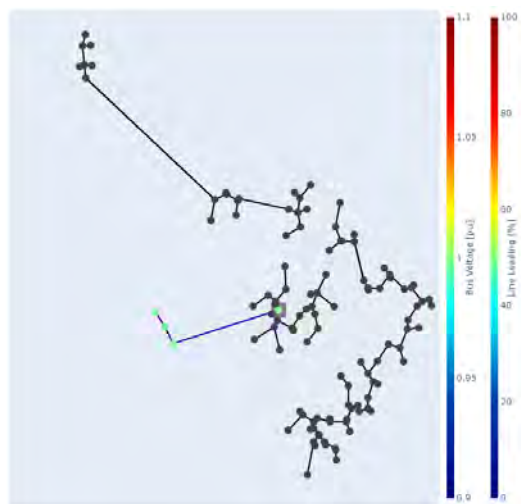
Each chromosome has a length equal to the number of communities N. Each element must be an integer number bounded between 1 and the total possible number of links coming out from the node plus 1, which is the case in which the node is off-grid. This strategy for chromosome coding allows to create only radial grids, avoiding loops that are not wanted in the type of topology to be designed. The total NPC is the fitness function to be minimized:

$$\sum NPC_{link} + \sum NPC_{mg} + \sum NPC_S + \sum E_{grid}\, coe \tag{6.48}$$

It is given by the sum of lines NPC ($\sum NPC_{link}$), (precomputed as explained in the Input data preprocessing paragraph), the one of the primary substations or interconnection nodes ($\sum NPC_S$), the one of microgrids installed $\sum NPC_{mg}$ (precomputed in Block2) and the energy bought from the grid $\sum E_{grid}$ multiplied by the wholesale price of energy *coe*. Constraints have been defined as an additional cost in the penalty function, whose value is given by the following equation:

Figure 6.37: Pandapower network model and genetic algorithm first iteration

(a) Starting pandapower network. Blue are the buses, yellow is the external grid connection and in grey the internal grids



(b) Network after first genetic iteration and load flow computation. Unsupplied buses in grey

$$\begin{cases} f_1 = NPC; \\ f_2 = C_{max} + g_E + g_I + g_P \\ f_3 = C_{max} * 2 + \sum_{i=1}^{S} \overline{P_{S,i}} - P_{load,i} \end{cases} \tag{6.49}$$

The cost function is such that f_1 is activated if no constraints are violated, and it represents the real cost for the given configuration. f_2 is activated in case voltage E is below minimum voltage level $\underline{E}$ in any of the nodes, current is above limits $\overline{I}$ in some of the lines, or the power requested from the substations overcomes the maximum power available $\overline{P_S}$. g_E counts the number of nodes i where $E_i < \underline{E}$. $g_I$counts the number of edges i,j for which $|I_{i,j}| > \overline{I}$ and g_P counts the number of substations where $P_i > \overline{P_S}$. f_3 is activated when, even before computing the load flow, it is clear that the total power requested to the substations $P_{load,i}$ overcomes their maximum power providable by the substations. C_{max} is the highest possible value of a feasible solution.

Module 4.1.2: MILP based optimization

As an alternative to the genetic approach, a MILP model has been developed, with the goal of obtaining accurate results in shorter timeframe. Those models are more rarely found in literature for distribution planning, since they allow to model only linear constraints. When applied, usually voltage constraints are neglected [75]. The proposed model considers power balances and voltage constraints, computing voltage drops on the lines starting from the following equation, resulting to be a well known linear approximation of the voltage drop problem over a radial network:

$$\Delta E = \frac{RP + XQ}{E_1} \tag{6.50}$$

and further simplifying it assuming $E_1 = E_{nom}$. E is the end voltage, R and X are respectively the line resistance and reactance, P and Q are the active and reactive power flows on the line and E_1 is the starting voltage. This is a linearized formula for voltage drop calculations extensively used in literature when dealing with distribution grids and microgrids planning: [167], [195], [196].

Power losses are neglected since given the usual small value of currents involved, their influence on power balance and line fluxes would be minor.

The model is composed by the following sets, parameters, variables and constraints:

Sets:

C	Communities
S	Substations
N	All nodes of the graph: Communities and Substations
A	Allowed connections in the graph $\in (NxN)$

Scalars:

coe	wholesale cost of electricity	[$/kWh]
V_{ref}	reference voltage	[kV]
A_{ref}	reference apparent power	[MVA]
R_{ref}	specific kilometric line resistance	[ohm/km]
X_{ref}	specific kilometric line reactance	[ohm/km]
$\overline{P_A}$	maximum allowed power flow over the lines	[MW]
$\underline{E}$	minimum voltage level	[p.u.]
$\overline{E}$	max voltage level	[p.u.]

Parameters:

$D_{C,c}$	Energy demand of community c during microgrid's lifetime	$c \in C$	[MWh]
$\underline{P_c}$	Peak demand of community c	$c \in C$	[MW]
$\overline{P_s}$	Maximum power providable by substation s	$s \in S$	[MW]
NPC_c	Net Present Cost of microgrid in community c	$c \in C$	[$]
NPC_s	Net Present Cost of substation s	$s \in S$	[$]
$NPC_{A,ij}$	Net Present Cost of connection *(i,j)*	$((i,j) \in A)$	[$]
$L_{A,ij}$	length of connection *(i,j)*	$((i,j) \in A)$	[km]
$L_{N,c}$	length of node c	$c \in C$	[km]
$Rl_{A,ij}$	resistance of connection *(i,j)*	$((i,j) \in A)$	[p.u]
$Xl_{A,ij}$	reactance of connection *(i,j)*	$((i,j) \in A)$	[p.u.]
E_s	per unit voltage of substation s	$s \in S$	[p.u.]

Variables:

$x_{ij} \in (0,1)$	1 if connection (i,j) exists , 0 otherwise	$(i,j) \in A$	
$y_c \in (0,1)$	1 if microgrid is installed in node (c), 0 otherwise	$c \in C$	
$z_s \in (0,1)$	1 if substation in node (s) is used, 0 otherwise	$s \in S$	
$P_{A,ij}$	power flow of connection (i,j)	$(i,j) \in links$	[MW]
$P_s \geq 0$	power provided by substation s	$s \in S$	[MW]
$E_{N,i} \geq 0$	actual value of the voltage at node N	$i \in N$	[p.u.]

Constraints:

Objective function: NPC minimization

$$\min \sum_{(i,j)\in links} NPC_{A,i,j}\; x_{ij} + \sum_{(c)\in C} NPC_c\; y_c + \tag{6.51}$$

$$+ \sum_{(s)\in S} NPC_s\; z_s + \sum_{(c)\in C} (1 - y_c)\; D_{C,c}\; coe$$

Radiality rule

$$\sum_{(i,j)\in A} x_{ij} = card(C) - \sum_{(s)\in S} y_s \tag{6.52}$$

Power flow conservation on each substation

$$\sum_{j:(s,j)\in A} P_{A,js} - \sum_{j:(j,s)\in A} P_{A,sj} = -P_s \qquad s \in S(s \neq j) \tag{6.53}$$

Power flow conservation on each community

$$\sum_{j:(c,j)\in A} P_{A,jc} - \sum_{j:(j,c)\in A} P_{A,cj} = P_c\; (1 - y_c) \qquad c \in C(c \neq j) \tag{6.54}$$

Feeders max loading

$$-\overline{P_A}\; x_{ij} \leq P_{A,ij} \leq \overline{P_A}\; x_{ij} \qquad (i,j) \in A(i \neq j) \tag{6.55}$$

Substation max loading

$$P_s \leq \overline{P_s}\; z_s \qquad s \in S \tag{6.56}$$

Linearized voltage drop

$$(E_{N,i} - E_{N,j}) - (1 - x_{ij}) \leq Rl_{ij}\; P^{pu}_{A,ij} + Xl_{ij}\; cos\phi P^{pu}_{A,ij} \tag{6.57}$$

$$(E_{N,i} - E_{N,j}) + (1 - x_{ij}) \geq Rl_{ij}\; P^{pu}_{A,ij} + Xl_{ij}\; cos\phi P^{pu}_{A,ij} \tag{6.58}$$

Voltage limits

$$\underline{E} < E_{N,i} < \overline{E} \qquad i \in N \tag{6.59}$$

$$E_s = 1 \qquad s \in S \tag{6.60}$$

Line parameters

$$Rl_{ij} = R_{ref} \cdot (L_{A,ij} + L_{N,j}) \cdot \frac{A_{ref}}{V^2_{ref}} / 1000 \tag{6.61}$$

$$Xl_{ij} = X_{ref} \cdot (L_{A,ij} + L_{N,j})) \cdot \frac{A_{ref}}{V^2_{ref}} / 1000 \tag{6.62}$$

$$\tag{6.63}$$

The objective function is the minimization of the total NPC, computed with the equation (6.51) as the sum of the lines NPC, the one of the substations, the microgrids installed and the energy bought from the grid. Each of the terms is multiplied by a binary variable, that is activated only when the technology is used. The radiality rule (6.52) ensures that the resulting grid is radial, with communities either supplied by a microgrid or connected to the grid. Equations (6.53) and (6.54) guarantee that the power balances is respected both in the susbstations' nodes and in the communities' nodes while (6.55) and (6.56) constraint the maximum power flowing on feeders and provided by susbstations. The voltage difference between two points, given by the linearized equations (6.57) and (6.58) is active only if the line connecting them is created. It depends on the line impedance computed through equations (6.61) and (6.62), where the total length of the connection is given by length of the line and the length of the node L_N. Since each community is represented only by one node, this last parameter is useful to avoid totally neglecting the voltage drop inside the community. It is set as equal to half of the community total extension. In case the node is a substation, this parameter is set to 0.

Over the tests performed, the inclusion of voltage drops in the model drives to convergence issues in case the problem becomes very large. In this case, instead of voltage drops limits, a constraint of the maximum length of the feeders could be added. In this case, instead of voltage $E_{N,i}$, the variable $d_{N,i}$ related to the distance of nodes from the national grid is used. This is constrained to be lower than a maximum length of the feeders, $\overline{d}$, given as input to the model and dependent on the voltage and size of the cables used. Since the distance constraint does not depend on the power flowing on the feeders, equations are more simple with respect to the model previously described and allow to treat bigger problems. The linearization of the distance constraint is performed to an additional variable kij and a big constant M.

Constraints:

Linearized distance

$$k_{ij} = e = (-L_{ij} - L_{N,i}) \cdot x_{ij} \qquad (i,j) \in A(i \neq j) \qquad (6.64)$$

$$-M \cdot x_{ij} \leq k_{ij} \leq M \cdot x_{ij} \qquad (6.65)$$

$$(d_{N,i} - d_{N,j}) - (1 - x_{ij})\, M \leq k_{ij} \leq (d_{N,i} - d_{N,j}) + (1 - x_{ij})\, M \qquad (6.66)$$

$$k_{ij} \leq (d_i - d_j) + (1 - x_{ij})\, M \qquad (6.67)$$

Distance limits

$$d_{N,i} < \overline{d} \qquad i \in N \qquad (6.68)$$

$$d_{N,i} = 0 \qquad i \in S \qquad (6.69)$$

(6.70)

6.4.3 Module 4.2: Multi-objective optimization

Finally, one further research effort has been spent in order to develop a multiobjective integrated model, able to discriminate different electrification solutions not only from an economic perspective but also in terms of their environmental and social impact. This is particularly important when the analyses are run by policymakers in charge of providing rural electrification plans, that should consider as much as possible all the different dimensions of development.

Literature review

The approach commonly found in literature to address the multi-dimensionality of rural electrification planning problem, is the multi-criteria decision making. This technique entails a two phase modeling of the system. At first, several solutions for electrification are generated by means of optimization processes. Secondly, the different electrification strategies are evaluated and ranked on the basis of some quantitative and qualitative indicators. Those works are mainly comparing different off-grid electrification strategies, evaluating which type of generator (e.g. PV modules, wind, diesel), or hybrid microgrid, performs better according to chosen criteria [30, 37, 38, 197].
Other works insert different aspects inside the cost function to be minimized as forms of monetary penalties: typically the energy not supplied can be represented by the cost for adopting alternative energy sources [79]. In this way, the optimization is able to find a unique solution, that balances several dimensions. The critical aspect of this type of approach is the difficulty of determining the cost of non monetary aspects as this would theoretically require extensive and well-designed surveys or experiments of the customers involved.

Multiobjective optimization offers a possible alternative optimizing simultaneously more than one objective function, without losing the possibility of analyzing several trade-offs.

Definition 10 *A general multi-objective optimization problem is defined as the minimization (or maximization) of the objective function set* $F(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))$ *subject to inequality constraints* $g_i(\boldsymbol{x}) \leq 0, i = \{1, \ldots, m\}$, *and equality constraints* $h_j(\boldsymbol{x}) = 0, j = \{1, \ldots, p\}$. *The solution of a multi-objective problem minimizes (or maximizes) the components of a vector* $F(\boldsymbol{x})$

where $\boldsymbol{x}$ *is a n-dimensional decision variable vector* $x = (x_1, \dots, x_n)$ *from some universe* Ω.

Generally, different objective functions are conflicting and hence there is no solution that can simultaneously optimize each single objective. Multi-objective problems are characterized by a set of good compromise solutions which constitute the different possible trade-offs. The concept of Pareto optimal solutions and Pareto frontier become fundamental to represent the space of solutions:

Definition 11 *Pareto optimal solutions are solutions which, when evaluated, produce vectors whose performance* f_i *cannot be improved without adversely affecting another* f_j*, with* $i \neq j$*. In other words,* x_i^* *is Pareto optimal only if there exists no other* x_i *that would improve* f_i *while maintaining all other* $f_{j\neq i}$ *constant.*

The Pareto front or Pareto frontier is the set of all the optimal solutions. Chiandussi et. al in [198] provide a review of multi-objective optimization techniques for engineering applications:

- Global criterion: the global criteria aims is to minimize a function $f(\boldsymbol{x})$ (global criterion) which is a measure of how close the decision maker can get to the ideal vector. The most common formulation is:

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} \left(\frac{f_i^0 - f_i(\boldsymbol{x})}{f_i^0} \right)^p$$

  where $f_i(\boldsymbol{x})$ is one objective functions i and f_i^0 is its minimum value. p is a coefficient and k the total number of objective functions.

- Weighted sum method: the objective function is a weighted sum of the value of the single objective functions. Varying the different weights associated to the functions they gain more or less importance and the Pareto solutions are found.

- ϵ - constrained method: with this approach, no tentative is made to aggregate different criteria, but only one of the original objectives is minimized while the others are transformed to constraints fixing threshold values ϵ_i. Pareto solutions are found changing the values of ϵ_i.

Considering the different categories of tools described in chapter 4, multiobjective optimization has been used for optimal off-grid systems sizing

[57, 199, 39], to add environmental and social objectives to the economic one. However, to the author's knowledge there is no tool available in literature that performs multi-objective optimization for rural electrification planning [54].

When speaking about works related to optimal grid topology planning, usually multidimensionality is considered by taking into account reliability factors (typically System Average Interruption Duration Index (SAIDI) and System Average Interruption Frequency Index (SAIFI)) [200], [201], [202].
SAIDI is the mean duration of interruptions in minutes, computed as follows:

$$\text{SAIDI} = \frac{\sum D_i N_i}{N_T}. \tag{6.71}$$

Where D_i is the annual outage time at location i, N_i is the number of costumers in the location and N_T is the total number of costumers.
SAIFI is the average frequency of interruption, computed as:

$$\text{SAIFI} = \frac{\sum \lambda_i N_i}{\sum N_i} \tag{6.72}$$

Where λ_i is the failure rate at location i. From the distribution grid planner perspective, environmental dimension is not as relevant as the technical one, that directly impacts on maintenance costs quality of service. Reliability and energy not supplied are often included in single objectives models as additional costs or constraints but some works also make use of multi-objective optimization ([202], [203])

Table 6.13 summarizes the main literature works and approaches previously described and the different electrification dimensions they address. The techno-economical framework (described in chapter 3) is separated into economical dimension (ec.) that is the minimization of total cost, and the technical dimension (tec.), that could include indicators such as reliability or failure rate of components. The environmental dimension (env.), typically covers the minimization of emissions and eventually other aspects such as land-use, the social dimension (soc.) is related, among others, to affordability and acceptability of technologies and the institutional dimension (inst.) to barriers for development.

Modeling approach

The proposed model starts from the single-objective MILP model previously described and adds to the least cost objective f_{ec} two objective functions,

Table 6.13: Summary of research works about multi-dimensional rural electrification planning

Model	Electrification options	Dimensions	Reference
Multicriteria	On grid, Off grid (wind, PV, diesel)	Ec., tec., soc., inst., env.	[30]
Multicriteria	Off grid (diesel, hydro, PV, biomass)	Ec., tec., soc., inst., env.	[37]
Multicriteria	Off grid (diesel, hydro, PV, biomass,wind)	Ec., tec.,soc., inst., env.	[38]
Multicriteria	Off grid ()	Ec., tec., soc., env., inst.	[197]
Multiobjective	Off grid	Ec., env.	[199]
Multiobjective	Grid extension	Ec., tec.	[202]
Single cost function	Grid extension	Ec., tec.	[201]
Single cost function	On grid, Off grid (PV, diesel)	Ec., tec.	[79]

for the social f_{soc} and environmental f_{env} dimension. Using a combination of the global criterion and of the weighted sum method, the functions are then normalized and summed, to obtain a unique objective function to be minimized (see eq. 6.73).

$$f_{MO} = \alpha \cdot (\frac{\overline{f_{ec}} - f_{ec}}{\overline{f_{ec}} - \underline{f_{ec}}}) + \beta \cdot (\frac{\overline{f_{env}} - f_{env}}{\overline{f_{env}} - \underline{f_{env}}}) + \gamma \cdot (\frac{\overline{f_{soc}} - f_{soc}}{\overline{f_{soc}} - \underline{f_{soc}}}) \tag{6.73}$$

α, β, γ are the weights, comprised between 0 and 1 associated to each of the objectives. $\overline{f_i}$ and $\underline{f_i}$ are the maximum and minimum possible values of each of the objective functions obtained running the model three times, each time with just one function activated.
The decision variables are the same of the single-objective with the addition, for each cluster, of choosing between three different types of microgrids, with different minimum RES percentage. The algorithm for hybrid microgrid sizing (described in section 6.2) is run three times, not to bias the solution:

1. Microgrid with minimum NPC;
2. Microgrid with at least 50% of energy produced by RES;
3. Microgrid with 100% of energy produced by RES.

A new set k is introduced to consider the three different types of microgrid. To each of them, different costs and emissions are associated. In order to provide a procedure that could be as general as possible, one of the main issues to be faced in this analysis is the availability of public data. Those, as will be explained later, influenced many of the assumptions and of the modeling choices.

Economic dimension

The objective function is the same as the single objective optimization, that is the minimization of the NPC, with the only difference being the NPC of the microgrids, a variable dependent on the set k, related to the type of microgrid.

$$\begin{aligned} f_{ec} = \ & \min \sum_{(i,j)\in A} NPC_{A,i,j}\ x_{ij} + \sum_{(i,k)\in C} NPC_{C,i,k}\ y_{ik} + \\ + & \sum_{(i)\in S} NPC_{S,i}\ z_i + \sum_{(i,k)\in C} (1 - y_{i,k})\ D_{C,i}\ coe \end{aligned} \tag{6.74}$$

Environmental dimension

The environmental impact of the proposed solution is evaluated by starting from the assumption that the energy mix of the country is not influenced by the electrification strategy identified and that it will remain constant during the lifetime of the project. The considered indicator is the CO2 emissions released in the atmosphere during the project lifetime. Both the possibility of considering direct emissions (only related to electricity production phase) and total emissions have been investigated. To compute indirect emissions, precomputed literature Life Cycle Assessment (LCA) values for each component were used [204]. LCA is a methodology for computing the environmental footprint of products or processes along their life cycle. It is based on energy and mass balances across defined boundaries (e.g. cradle to grave) dependent on the geographic localization of the processes (producing a PV module in China has a different impact with respect to producing it in Europe due to different processes and energy mix of the countries). This is why, to provide an accurate analysis, values from literature should be avoided and a LCA should be performed for each specific case study. However being in this work the focus mostly on the methodology, recognized values from literature are used and the integration of LCA databases and computations is proposed as a further improvement. Input-Output approach was also investigated and resulted potentially interesting for European countries [205, 206]. The advantage is the possibility of using country-related emissions disaggregated into different industrial sectors, considering also imports and exports. However given that there is no availability of finely disaggregated matrices of data for Sub-Saharan Africa, the approach was discarded.

$$f_{env} = \min \sum_{(i,j)\in A} em_{A,i,j}\ x_{ij} + \sum_{(i,k)\in C} em_{C,i,k}\ y_{ik} + \sum_{(i)\in S} em_{S,i}\ z_i + \\ + \sum_{(i,k)\in C} (1 - y_{i,k})\ E_{C,i}\ em_{nat}\ \ [kgCO_2] \tag{6.75}$$

The specific national emissions are computed knowing the energy mix of the country as follows:

$$em_{nat} = \frac{\sum_i (em_i * EP_i)}{\sum_i (EP_i)} \tag{6.76}$$

Where i is the index representing the different types of energy sources, em_i are the specific emission related to that source and EP is the average, over the last five years, of the energy produced by each source, taken from Energy Information Administration (EIA) International Data Browser. $em_{A,i,j}$ and $em_{S,i}$, the emissions relative to the construction of electric cables and substations, would be active only in case total emission are considered. However, also in this case they are considered negligible with respect to the other components, and their value is set close to 0.

Social dimension

Among the several social aspects a new electrification project can touch, attention has been focused on the hours of available energy during the year for each customer. This affects the social sphere of projects since non reliable electricity is an issue that can hinder development of communities, increase costs of energy provision (in case it is necessary to provide it from alternative sources, e.g. diesel generator) and prevent the sustainability of business activities. Unavailability of electricity also depends on the technical sphere, that is the reliability and failure rate of components, so some works in literature (see the literature review section) categorize it as technical.

National grids in less developed countries often provide a low quality of service, with frequent and long lasting interruptions of service and microgrids could be an effective solution to provide a more stable and higher quality electricity provision.

The WorlBank commissioned two different studies where among other aspects, the reliability of national grids in the world was assessed: Enterprise Surveys and Doing Business [207]. Those data are publicly available and can hence be used in the proposed model as a proxy for the estimation of hours of unavailability of national electricity. The first one is a firm level survey of a representative sample of the private sector in different economies. Among the various topics covered in the survey, the reliability and availability of infrastructure is assessed with several indicators; number of electrical outages in a typical month and the average duration of electrical outages are the ones that allow to estimate the average hours of unavailability during the year.

Doing Business is a complementary survey focusing on measuring the complexity of business regulations and quantifying the ease of doing business across countries via indicator sets and rankings. SAIDI and SAIFI indicators are directly provided by utilities of the largest business cities. Comparing the value of indicators provided by the two different surveys, it appears that Doing Business reports results are much more optimistic, with values of SAIDI that are on average just the 15% of the ones registered with Enterprise surveys [208]. Given the likely bias in the utilities self-reporting their performance, Enterprise Survey data are used in the present work to compute the average yearly hours on unavailability of national grids in each country (un_{nat}) [h].

The objective function is given by the minimization of the total energy not supplied:

$$f_{soc} = \min \sum_{(i)\in C} un_{line,i}\ D^h_{C,i} + \sum_{(i)\in C} un_{mg,i,k}\ D^h_{C,i}\ y_{ik} + \\ + \sum_{(i,k)\in C} (1 - y_{ik})\ D^h_{C,i}\ un_{nat}\ \ [MWh] \tag{6.77}$$

The average yearly hours of unavailability of the national grid and of the microgrids $un_{mg,i,k}$ are multiplied by the average hourly energy consumption of each community $D^h_{C,i}$ to compute the total energy not supplied.

In addition, a term related to the possible outages from electric lines ($un_{line,i}$) is added. This term is computed considering an average failure rate per km of lines (λ_{line}) taken from literature multiplied by the average repair time (t_r) and by the length of the line connecting communities to the national grid d_i. The distance is computed according to the constraints described in the single-objective MILP model.

$$un_{line,i} = d_i \cdot \lambda_{line} \cdot t_r \tag{6.78}$$

This term results small with respect to the total reliability of the national grid, so it does not move the equilibrium with respect to the technology choice. However, it helps to build grids that prioritize important loads, tending to connect bigger loads with shorter lines to reduce the total energy not supplied.

The most difficult term to be estimated is the unavailability of the microgrid $un_{mg,i,k}$ and in particular the repair time of different components, term that is really case specific, depending not only on the site and technologies but also on the company in charge of maintenance of the plant. Within the WorldBank questionnaires of the MTF, used in the analysis of the load profiles, households were asked also to make some estimation of the outages duration, both in case their main source of supply are microgrids and in case it is the national grid. However, data collected where not enough to make some reasonable assumption since most of the households did not properly answer. Ferrall in [209] tried to map reliability patterns of minigrids and national grids in East Africa, with a collaboration with a company installing several sensors in 10 minigrids across Tanzania and Kenya finding an average SAIDI index of 379 but with a strong variation between the different cases. In the proposed procedure, some literature values related to the failure rate of different microgrid components are used to provide a first estimation of their yearly unavailability.

6.4.4 Contributions and limitations

The proposed procedure represents a novel methodology to perform an integrated design of on and off-grid technologies. The development of three models, gives the flexibility of performing the analysis most suited to the area under consideration. In particular:

- Genetic algorithm: it can include non linear constraints and is able to run load flows during optimization; it has the disadvantage of not reaching global optimal solution but the capacity of providing feasible solutions; it can be run without the use of commercial softwares (e.g. Gurobi) needed for MILP;
- Single-objective MILP: it reaches the global optimal solution and can include some linearized electrical constraints; it could have convergence issues when applied to large areas;
- Multi-objective MILP: it is useful to perform a comprehensive evaluation of the electrification problem, considering environmental and social dimension aside the economical aspects.

The proposed procedure, is innovative and effective in performing rural electrification planning, designing both the optimal grid topology and identifying off-grid solutions. However, it still presents some limitations. The MILP model, that within the first tests appeared to be more promising with respect to genetic, considers communities as singles nodes rather than modeling the whole internal grids: this could lead to great approximations in case the size of communities becomes very large. For this reason, this is a topic that has already being addressed within the research group and will constitute the material for the next publication. In addition, further efforts will be put on the inclusion of different dimensions in the multi-objective optimizations and on a comprehensive data collection campaign that could help obtaining more realistic results.

Part III

Applications

Chapter 7

Mozambique-Namanjavira

The goal of the analysis is the study of the optimal electrification strategy of the rural area of Namanjavira, in the province of Zambezia, in Mozambique. This is to test and validate the algorithms developed in the PhD thesis framework. The area, described in detail in chapter 5 has been selected because of a collaboration with the NGO COSV, working on electrification projects for rural communities, that allowed to collect on-field data useful for the analysis. Simulations have been run with a processor Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz 3.00 GHz, with 128 Gb of RAM. In the following sections the data gathered and the results obtained are described.

7.1 Case study set-up

The objective of the case study is to test and demonstrate the applicability of the developed procedures in an area that concentrates all the typical features of non-electrified rural areas in Sub-Saharan Africa: scattered population, mainly distributed along roads, availability of renewable resources, in particular solar energy, type of environment mainly composed by trees, forest and croplands. Some challenges, especially the lack of detailed data related to villages population and distribution grid, made the use of geospatial analysis particularly important. The data and assumptions used for the various blocks of the procedure are summarized in the following paragraphs, following the same subdivision used in the modeling part (chapter 6).

Block1: Energy demand assessment

Data related to boundaries of existing communities in this area are not available on open source databases , neither is information regarding rural villages' population. Data related to population density are retrieved from the High Resolution Settlement Layer [97].
The load profiles of the communities are created with the bottom-up tool RAMP described in chapter 6.1, starting from data collected from local surveys. Data were collected only in the village of Namanjavira (with 706 households and a surface of 7 km^2) but they were extended to the other communities, scaling the profile according to the population. Details on the input parameters for RAMP procedure are reported in annex A and table 7.1.

Table 7.1: User classes and number of users in Namanjavira

User	**N**	**User**	**N**
Administration	1	Office	1
Barber	3	Police	1
Health Center	1	Primary School	1
Households T1	654	Public Lights	1
Households T2	52	Secondary School	1
Merchant	21	Tailor	7
Worship	4		

Block2: Off-grid system sizing

All the input parameters required by the MILP procedure are detailed in annex B. They derive from literature data. For each river, the hydro turbine with the size closer to the average power production is chosen. Rivers with flow rate higher than 20 m^3/s or average power production higher than 1 MW are excluded from the analysis because they would be suited for large power plants and not exploitable by small scale microgrids.

Block3: Internal grids design

The geospatial data used in the procedure for the creation of the cost surface are listed in the table 7.2. They are a mix of raster and vector layers which provide information on the elevation, from which also the slope layer is then computed, average rivers' flow rate, to penalize crossing of large rivers, landcover, with 10 different landcover types provided at high resolutions and roads path. There are no protected areas in Namanjavira. Roads' layer from

OSM is filtered to consider only the most significant roads and remove the small pedestrian paths. The weighting procedure for the cost surface creation follows the criteria and the numerical coefficients detailed in section 6.3.4.

Table 7.2: Number of users of each category

Data	Type	Source
Elevation	Raster-30m	NASA SRTM Digital Elevation [136]
Rivers	Vector-lines	HydroRivers [128]
Landcover	Raster-30m	Esa Landcover CCI [210]
Protected Areas	not present	
Roads	Vector-lines	OSM[211]

Block4: Integrated area optimization

No accurate data related to the in place distribution grid was available. The planned expansion of the transmission grid in the area is used as a reference to make assumptions on the possible location of primary substations. Their construction cost is assumed to be a sunk cost, and just an additional cost for interconnection... is included in the optimization. Since the voltage levels of distribution grid in the country is either 11 kV or 33kV, both levels are considered as different options is the analysis. The electrical parameters assumed for the distribution feeders are reported in table 7.3. The investment cost for MV lines was retrieved by on-field interviews with the local grid operators. The wholesale cost of energy in the country is the MV electricity tariff in Mozambique, as published in the website of EDM `https://www.edm.co.mz/en/website/page/electricity-tariffs`.

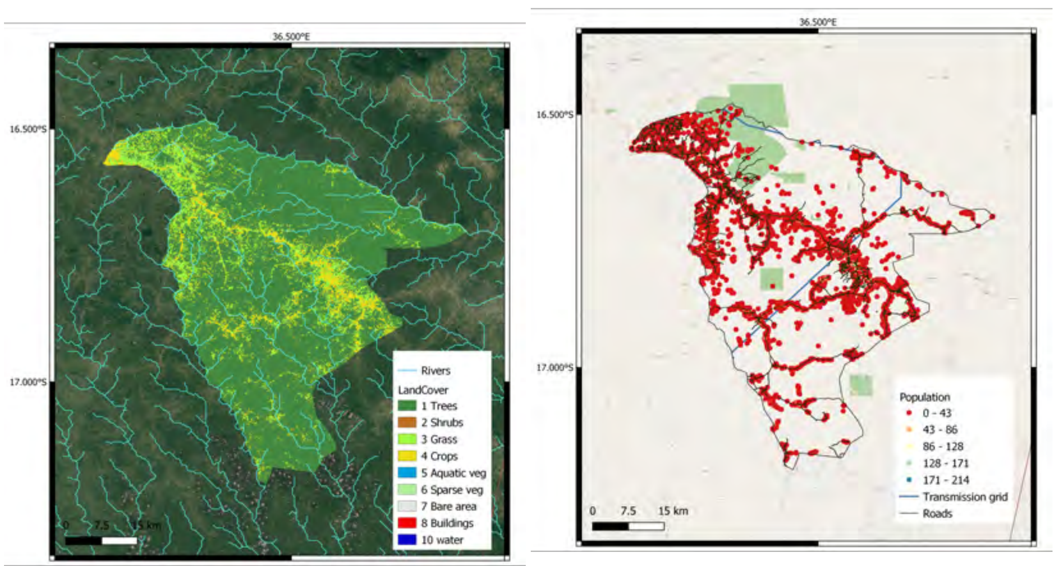
Table 7.3: Electrical parameters

Data	Value	u.m.
Investment cost	20	k$/km
Lifetime	40	years
r	0.45	ohm/km
x	0.36	ohm/km
c	9.7	nF/km
I	0.26	kA
coe	0.06	$/kWh

7.2 Results and discussion

The main geospatial data gathered for the area under consideration are shown in 7.1. The area is prevalently covered with trees (80% of the total surface). The 11 % is covered by grassland and the 9 % is dedicated to cropland.

Figure 7.1: Geospatial representation of Namanjavira



(a) Landcover and rivers in Namanjavira

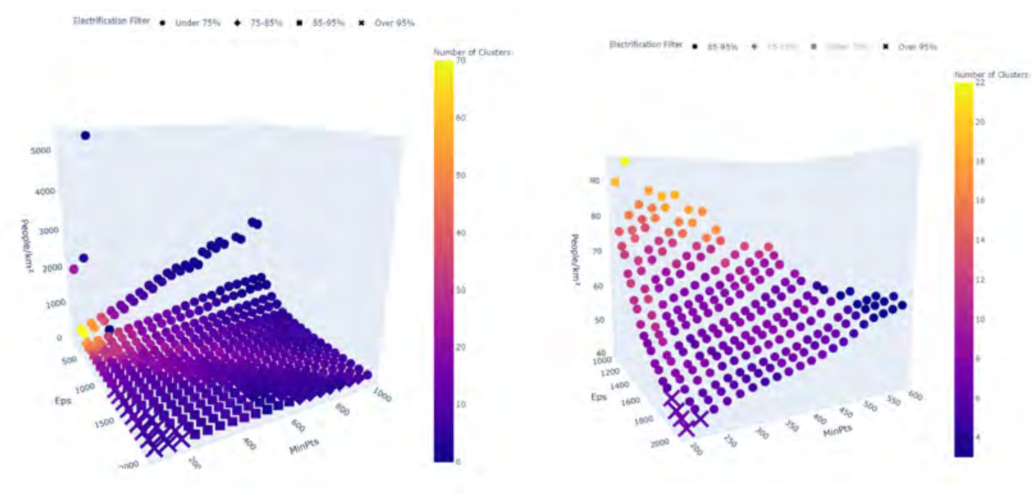
(b) Population, transmission grid and roads in Namanjavira

The area under study, according to the administrative subdivision, has a surface of $2673km^2$; the total population in the area is 64 135 people, with a population density of $24pp/km^2$, less than the average country value.

Block1: Energy demand assessment

Since data related to boundaries of existing communities in this area are not available, the density based clustering algorithm DBSCAN is run as the first step of the procedure to subdivide the population into densely inhabited clusters. No data related to the average population density of Mozambican villages was available, hence a sensitivity analysis on the DBSCAN parameters has been run, to identify the optimal combination of ϵ and $MinPts$: ϵ has been varied between 500m and 2000m and $MinPts$ between 100 and 1000 people. The criteria of choice are the electrification of at least the 85%

Figure 7.2: Sensitivity analysis on DBSCAN parameters



of the population, a minimum village population of 100 inhabitants and a population density as high as possible, since it is the value that mostly affects the convenience of investments. Preliminary results of the analysis are shown in figure 7.2. The images show how the average density of clusters changes according to $MinPts$ and ϵ. Points are coloured according to the identified number of clusters and the shapes represent the percentage of population clustered and that will hence be supplied by electricity with the proposed strategy. The second image has a bigger level of detail. As expected, the higher the $MinPts$ value and the lower the ϵ, the higher the number of outliers and hence the lower the percentage of future electrification rate. Number of clusters increase up to 70 with $\epsilon = 500$ and $MinPts = 100$ The values of $MinPts = 200$ and $\epsilon = 1000m$ have been chosen as a good compromise between all the necessities. The 87 % of the population is included in the communities that have an average population density of $24pp/km^2$. 19 clusters are found (fig. 7.3).

The load profiles of the communities, as stated previously, have been created with the RAMP tool, starting from on-field data collection, considering 13 user categories in each of the communities. The number of users of each category is computed rescaling the number in the reference community according to the number of households and surface area. The average load profile of cluster 5, disaggregated into the various user categories, is shown in figure 7.4. A 2% annual increment of the load is considered, set equal to the average rate of growth of rural population in Mozambique.

Figure 7.3: Clusters polygons

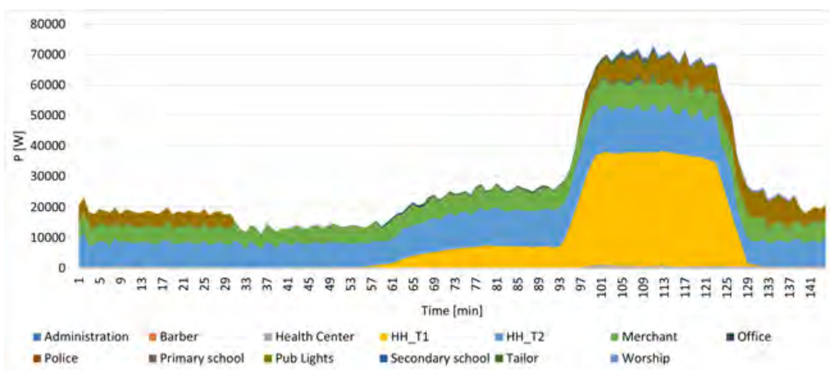
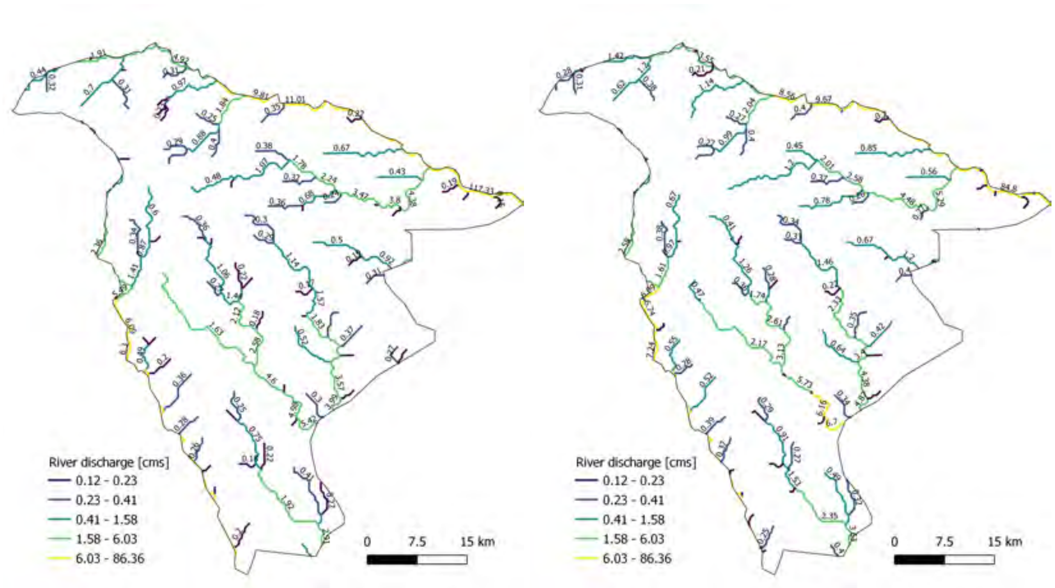


Figure 7.4: Disaggregated daily load profile of cluster 5

Block 2: Off-grid system sizing

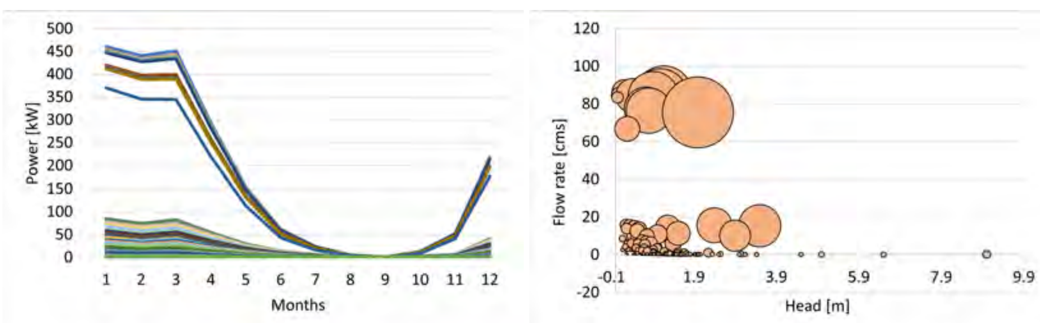
Before sizing the hybrid microgrid for each of the communities, the hydro power potential in the area has been estimated, as part of the assessment of the local resources. The monthly average river flow rate is estimated according to the procedure explained in chapter 6.2.

Figure 7.5 shows a comparison of the average annual discharge estimated through the SWAT procedure and the one reported in the HydroAtlas layer.



(a) HydroRivers (b) SWAT procedure

Figure 7.5: Average annual river discharge [cms] from literature data and SWAT procedure implemented



(a) HydroRivers (b) SWAT procedure

Figure 7.6: Average annual river discharge [cms] from literature data and SWAT procedure implemented

The order of magnitude of all the river streams is comparable, with flow rates going from 0.15 to 118 m^3/s in the Hydrorivers case and from 0.11 to 86 m^3/s in the simulated scenario.

Figure 7.6 shows the computed monthly power output of each of the river streams (each line is a river) and their characteristics in terms of flow rate and head (each 200m); the size of the orange bubbles is proportional to the average power availability. Given the absence of significant high grounds, the available head of the rivers is low, never overcoming 4m each 200m in rivers with significant flow rate, corresponding to an average slope of 2% (2 degrees). Rivers with higher power availability are the main streams with high flow rate, that gather water from multiple sub basins.

SWAT procedure and the CN method assumptions rely on the rainfall regime; for this reason, during the dry season power production is close to zero in all the rivers. Actually, this is a demonstrated trend of rivers in the region, as shown by figure 7.7, where the measured flow rate of the Rio Licuare, situated in Zambezia, over different years from 1968 to 1976 is plot: data are taken by the ADHI database.

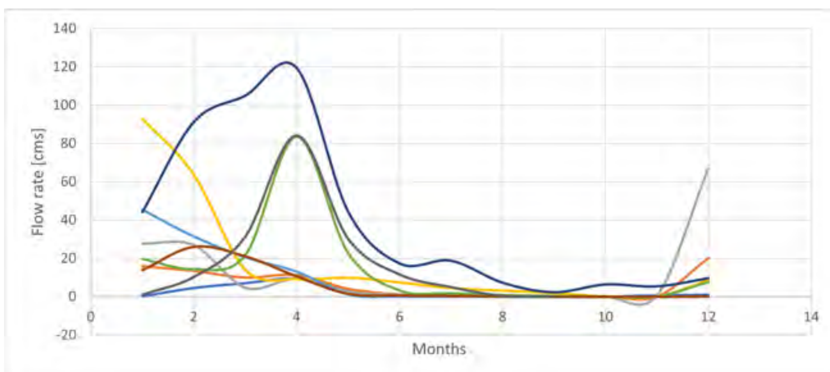


Figure 7.7: Measured flow rate of Rio Licuare in Mozambique, author elaboration from [141]

After the evaluation of the hydro-power potential, the MILP procedure for microgrid sizing has been run, considering also solar and wind resources availability, using data from `www.renewables.ninja`. In terms of hydro resource, to each community are associated the three rivers with highest average power output in a radius of maximum 5km. For each year of the simulation, 12 typical days (1 per month) are considered, averaging the resources profiles.

The results of the optimal generation portolio of microgrids resulting from the MILP optimization, are reported in figure 7.8. Cluster 7, given the much larger dimension, is shown on a different scale. Hydro resource is exploited when possible. It is interesting to notice that the least cost solution is a mix of different resources and that PV and BESS, with the assumptions adopted are competitive with diesel generators.

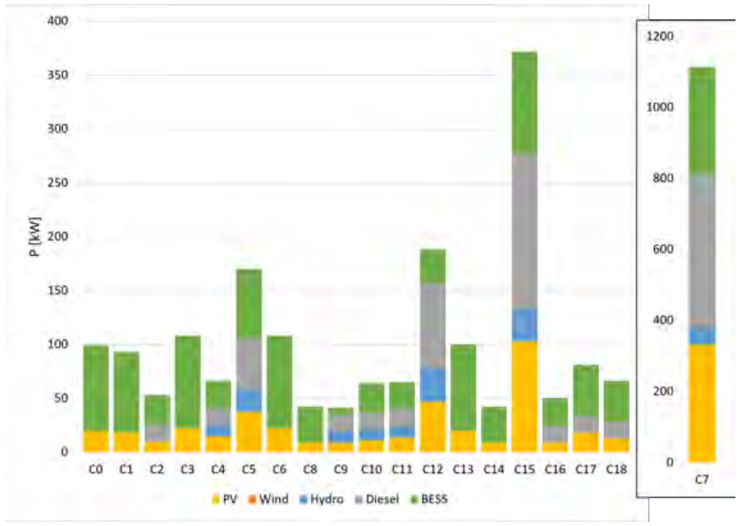


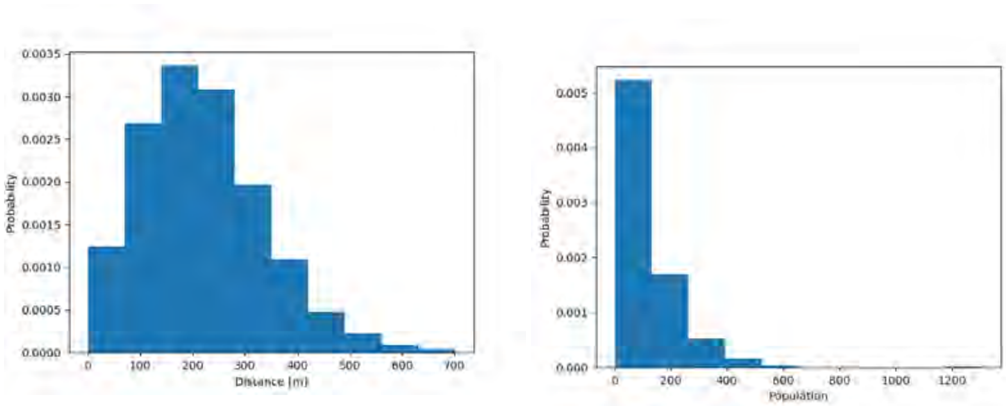
Figure 7.8: Optimal microgrids' generation portfolio, minimum NPC, Namanjavira

Table 7.4: Results of hybrid microgrid sizing

Cluster	PV [kW]	Hydro [kW]	Diesel [kW]	BESS [kWh]	Inverter [kW]	NPC [$]	En. Demand [MWh]	En. Produced [MWh]	LCOE [$/kWh]	CO2 [ton]	CO2 spec. [kg/pp]
C0	20	0	0	79	31.1	87	338	331	0.33	0	0
C1	18	0	0	75	28.3	82	315	310	0.34	0	0
C2	10	0	16	27	14.1	94	372	357	0.33	158	326
C3	22	0	0	86	33.9	95	372	364	0.33	0	0
C4	15	10	16	25	25.5	195	888	848	0.29	262	237
C5	38	20	48	64	52.8	509	2826	2695	0.24	887	233
C6	22	0	0	86	34	95	371	363	0.33	0	0
C7	333	50	432	299	390.7	4018	20288	19317	0.26	10099	363
C8	9	0	0	33	13.5	37	154	150	0.31	0	0
C9	9	10	16	6	12.3	164	799	761	0.27	230	229
C10	11	10	16	27	18.8	150	688	658	0.29	122	141
C11	14	10	16	25	24	179	851	814	0.28	213	209
C12	47	30	80	31	54.7	721	3926	3734	0.24	1397	271
C13	20	0	0	80	31	88	342	334	0.33	0	0
C14	9	0	0	33	13.6	37	154	150	0.32	0	0
C15	103	30	144	95	121.9	1347	6755	6431	0.27	3110	337
C16	9	0	16	25	12.9	93	369	355	0.33	163	343
C17	18	0	16	47	25.4	153	688	660	0.29	288	340
C18	13	0	16	37	22.6	108	435	419	0.33	161	288

Internal grids design

For the design of clusters' grids, the algorithms of procedure 3.3 b are applied, so at first the procedure for secondary substations placement (module 3.1) is run. Agglomerative clustering is run for each of the 19 clustered communities, with a distance threshold of 1000 m, considered as twice the maximum length of distribution grid (maximum of 500m). Substations are set in the



(a) Probability distribution of points-centers distances

(b) Probability distribution of population

Figure 7.9: Results of agglomerative clustering for secondary substations siting

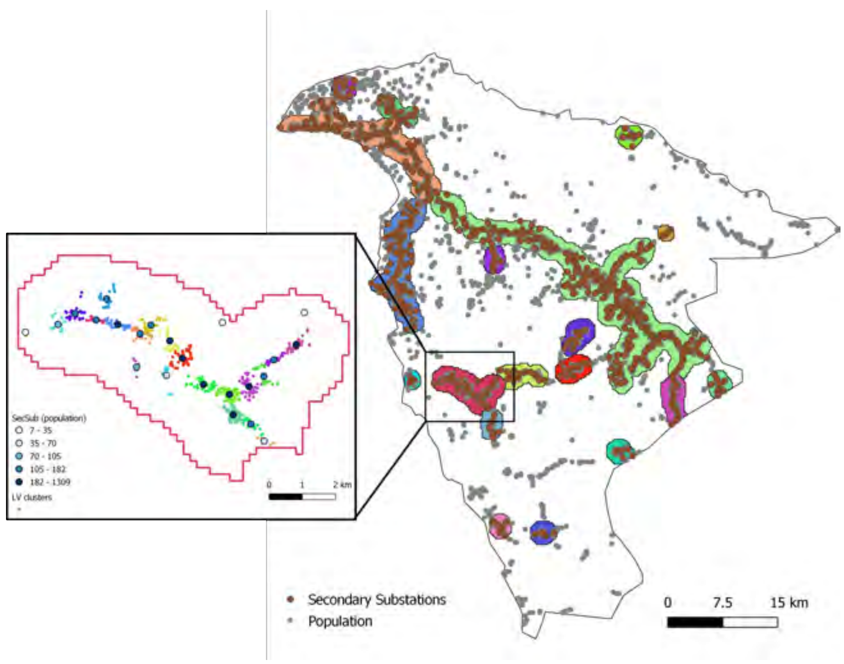


Figure 7.10: Results of the two steps clustering procedure

centroids of each low voltage cluster identified. Figure 7.9 shows the probability distribution of the distances between populated points and secondary substations (left image) and the probability distribution of the number of people supplied by one susbstation. The constraint of maximum 500m distance is satisfied in the 98% of the cases. Figure 7.10 graphically show the results of the two steps clustering process, which starting from population data, arrives to the identification of MV and LV clusters. Numerical results are reported in table 7.5 that shows the characteristics of each community in terms of population, peak load estimation and number of secondary substation. A rough estimation of the investment cost for secondary substations installation, considering they just consist on pole mounted transformers at the cost of 5 k$ each and of the LV grid installation (at a cost of 10 k$/km) are also reported; costs have been collected through on-field surveys. The LV grid is not routed in the proposed procedure so the estimation of its length is provided by the summation of the distances between households and transformers.

Table 7.5: Results of clustering procedures

Community	Population [pp]	Peak Load [kW]	Load per hh [W/pp]	Area [km2]	N SecSub	SecSub cost [k$]	LV Length [km]	LV Cost [k$]
0	421	10.87	129.04	9.6	4	20	14	142
1	393	10.4	132.37	8.4	6	30	12	121
2	483	12.04	124.67	9.76	5	25	18	184
3	490	12.19	124.37	10.8	7	35	14	141
4	1104	27.22	123.26	18.52	12	60	35	352
5	3803	85.19	112	39.72	23	115	118	1182
6	490	12.08	123.31	9.24	8	40	14	140
7	27854	608.27	109.19	249.12	191	955	876	8762
8	221	5.21	117.98	4.84	3	15	7	72
9	1008	24.88	123.41	18.48	14	70	29	288
10	863	21.56	124.89	15.32	10	50	31	309
11	1021	26.15	128.05	16.28	9	45	29	289
12	5155	119.06	115.48	69.84	61	305	175	1754
13	428	11.23	131.19	9.52	6	30	13	134
14	221	5.32	120.27	4	3	15	5	54
15	9227	203.32	110.18	85.68	77	385	291	2910
16	476	12.06	126.67	9.68	6	30	17	174
17	849	21.29	125.39	15.44	16	80	31	305
18	559	13.75	122.99	9.88	11	55	17	166

The graphical representation of the cost surface (created following the module 3.2, described in chapter 6) and the results of the graph theory based routing procedure for designing the medium voltage grid inside the communities (module 3.3) are shown in 7.11. The resolution adopted for the creation of the cost surface is of 200m, with a resulting cost-surface grid composed of a total of 66902 $GridPts$. The number of additional points, sampled along the roads $RoadPts$ is 43417. Points are colored according to their penalty factor, computed considering both OPEX and CAPEX over the grid lifetime of 40 years. It is evident how the cost of building lines is influenced by the presence of roads, with a minimum penalty factor of 1.37, arriving to a maximum weight of 15.31 in difficult areas, mainly forests and steep terrains.

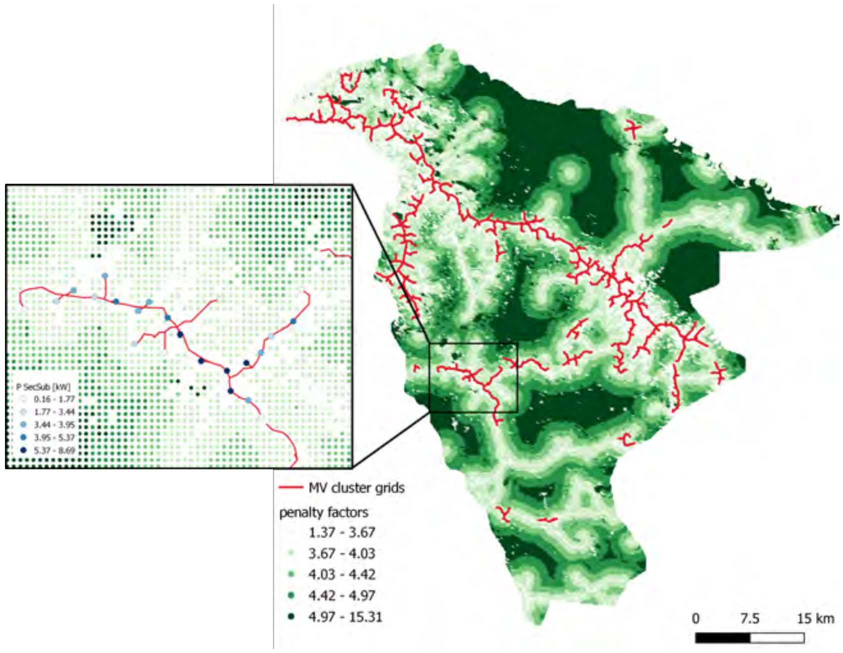


Figure 7.11: Results of the internal grids routing procedure

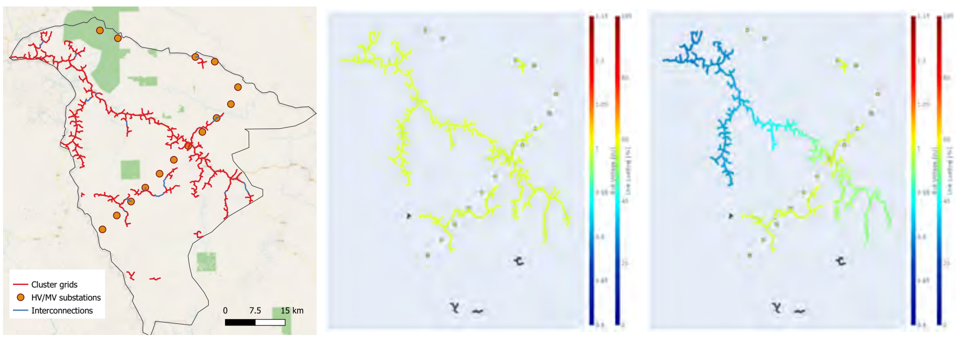
Numerical results of the grid routing procedure are reported in table 7.6. For each cluster both algorithms M1 (Steiner MST procedure) and M2 (MST +Dijsktra procedure) have been run to find the cheapest solution. NPC_{10} is computed along the project lifetime (that is 10 years, same lifetime of micro-grids), hence the specific cost is in most cases lower than the line investment cost, due to the salvage value. The minimum value of NPC, when line is built along roads is of 11.35 k$/km, however it is not constant among the clusters because of the penalty factors associated to the terrain (see section 6.3). NPC_{40} is instead computed over the grid lifetime, to have an estimation of the total cost associated to electric lines deployment. In this case, along the roads the NPC would be 26.4 k$/km. In only one case (cluster 2) M1 finds a solution with la cost lower than M2, with a negligible cost difference of 7 %. In terms of computational effort, in all cases M2 outperforms M1, with a time of the order of magnitude of seconds againts time approaching one hour for M1 (e.g in cluster 18). In the bigger clusters (C5, C7, C12 and C15), only M2 was run to avoid memory issues.

Table 7.6: Results of grid routing procedure

Community	Grid Length [km]	NPC [k$]	Spec. cost [k$/km]	Algorithm	Time1 [s]	Time2 [s]	Cost diff [%]
10 years							
0	2.877	47	16	M2	65.6	1.7	4
1	4.279	119	28	M2	60.9	2.3	4
2	4.522	77	17	M1	25.2	2.3	7
3	6.366	113	18	M2	294.9	2.8	3
4	9.837	166	17	M2	574.8	5.6	8
5	20.539	369	18	M2-only	0	11.4	0
6	5.95	110	19	M2	215.6	3.7	13
7	163.317	3028	19	M2-only	0	108.4	0
8	1.438	26	18	M2	0.8	0.9	17
9	10.613	218	21	M2	358.1	6	7
10	7.588	135	18	M2	437.6	4.6	7
11	6.168	116	19	M2	485.1	3.8	10
12	56.905	1055	19	M2-only	0	36	0
13	4.179	81	19	M2	177.1	2.7	2
14	1.473	29	20	M1-M2	4.3	1	0
15	68.331	1165	17	M2-only	0	46.8	0
16	5.725	94	16	M2	58.2	2.6	11
17	13.523	229	17	M2	1923.1	8.5	10
18	9.238	168	18	M2	2654.7	9	3
40 years							
0	2.82	86	30	M1-M2	78.8	2.1	0
1	4.427	232	54	M1-M2	73.9	2.7	0
2	4.512	140	31	M1	28.1	2.6	11
3	5.674	208	33	M1-M2	311.4	3	0
4	8.548	297	30	M2	642.6	6	2
5	19.686	702	34	M2-only	0	12.7	0
6	6.079	206	35	M2	250.8	4.4	7
7	151.998	5434	33	M2-only	0	122.5	0
8	1.438	46	32	M2	0.8	1	13
9	9.891	405	38	M2	362.2	6.4	1
10	7.677	242	32	M2	458.5	4.9	3
11	6.168	208	34	M2	514.4	4.1	5
12	54.275	1966	35	M2-only	0	38.7	0
13	4.082	143	34	M2	187.3	2.9	0
14	1.475	47	32	M1-M2	5.6	1.1	0
15	64.383	2074	30	M2-only	0	51.5	0
16	5.517	181	32	M2	65.9	2.8	5
17	12.867	421	31	M2	1965.2	9.1	1
18	9.311	298	32	M1-M2	2630.6	9.8	0

Block4: Integrated area optimization

The results of the integrated area optimization, found with the MILP model and a voltage level of 33kV for the distribution grid are shown in 7.12. Voltage constraints are not active, since the low power values and high voltage level do not create issues on the lines, as shown by the load flow results in the second subfigure. However, if the area were electrified at a voltage level of 11kV, this configuration would lead to voltage problems, as shown in the third subfigure, with load flow results showing a minimum of 0.89 p.u.. Loadflows have been run with the *pandapower* library of Python.



(a) Optimal topology from MILP and 33kV (b) Load flow on optimal topology at 33kV (c) Load flow on optimal topology at 11kV

Figure 7.12: Results of integrated area optimization, milp 33kV

The procedure has been run again lowering voltage level to 11kV, where voltage constraints in the algorithms become useful to identify feasible solutions.The MILP model leads to the result of figure 7.13. It is a feasible solution, with voltage drops constraints respected. Table 7.7 shows the error computed by the MILP model when estimating the minimum voltage level in each cluster. Since cluster 7 has a wide extension (40km approximately), error reaches high values (up to 7 %). To improve accuracy, the procedure should be run in smaller clusters where internal drops are negligible. In any case, the estimated voltage is always lower than the real one, so the procedure leads to a conservative result. The genetic algorithm allows to avoid voltage approximations, given the opportunity of including non linear constraints and running the power flow at each iteration. However it could converge to a different solution at each run, not guaranteeing the optimality, as shown in the zoom of figure 7.14. Genetic algorithm should be run several to obtain a pool of feasible solutions among which to choose the optimal.

Both with genetic and MILP algorithm and with both voltage levels, the clusters that are selected to be electrified with an off-grid system are C0,C1,C2,C8. Their characteristics, together with the ones of the other clusters, are summarized in table 7.8. What off-grids clusters have in common

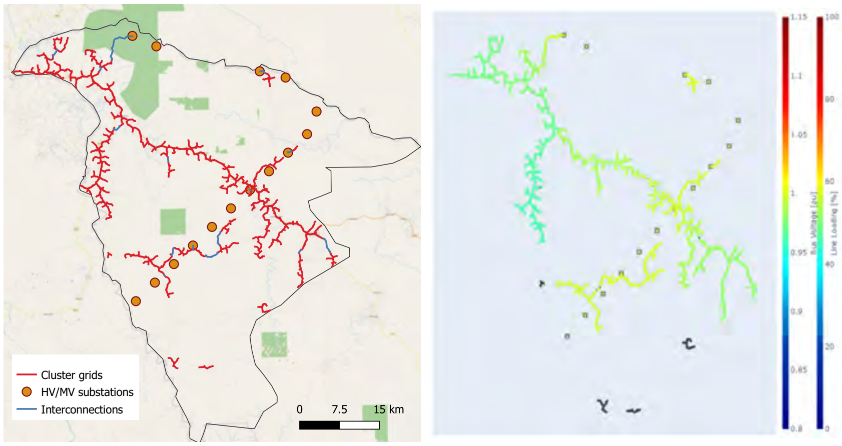


Figure 7.13: Results of integrated area optimization, milp 11kV

Table 7.7: Errors in the MILP nodal voltage computation

	33kV			11kV		
Cluster	E milp [p.u.]	E load flow [p.u.]	error [%]	E milp [p.u.]	E load flow [p.u.]	error [%]
C3	1	1	0	0.991	0.994	0
C4	0.98	0.996	2	0.915	0.978	6
C5	1	1	0	0.991	0.994	0
C6	0.98	0.996	2	0.915	0.979	7
C7	0.98	0.996	2	0.915	0.978	6
C9	1	1	0	0.994	0.998	0
C10	1	1	0	0.993	0.998	1
C11	1	1	0	0.992	0.997	0
C12	0.976	0.995	2	0.934	0.958	2
C13	0.98	0.996	2	0.915	0.984	7
C14	1	1	0	1	1	0
C15	0.977	0.995	2	0.941	0.97	3
C16	1	1	0	1	1	0
C17	0.977	0.995	2	0.974	0.98	1
C18	0.977	1	2	0.941	0.97	3

is their small population (less than 500 people), low population density (less than 50 pp/km2) and high distance from the in place grid (above 10km). However, those are not the only clusters with these characteristics: C6,C13 are also small, with low density and far from the national grid. Given their proximity to cluster 7, however, it is more convenient to interconnect them rather than supply with an off-grid system. As a further improvement, to

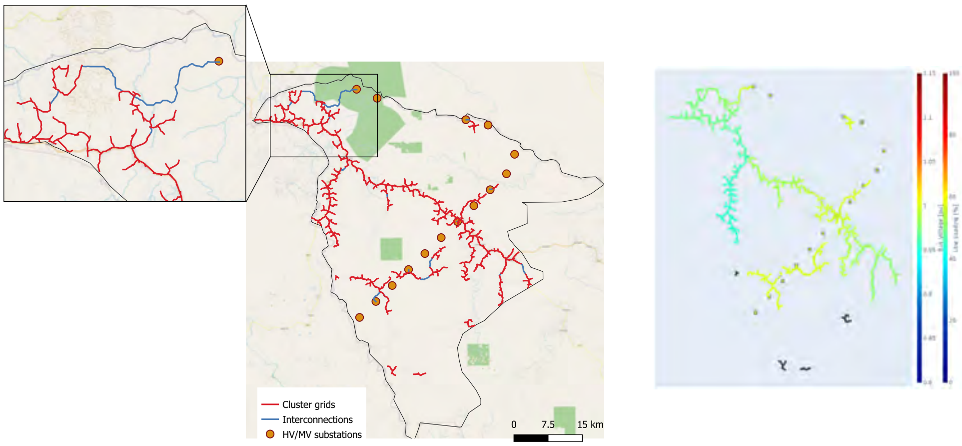


Figure 7.14: Results of integrated area optimization, genetic algorithm 11kV

evaluate the accuracy of the solution, Gisele could be run other times, splitting the wider clusters, to understand where it would be reasonable to split the systems.

The last part of the analysis is related to the multiobjective optimization. In this case, not to include too many variables in the problem, a voltage level of 33kV has been set. Specific direct emissions related to the national grid energy mix are 90 kgCO2/MWh , while the estimated hours of unavailability per year are 104. The model has been run at first only considering direct emissions, with different combinations of weights, to create a set of Pareto solutions, shown in 7.15 and 7.16a. The maps show in grey the clusters interconnected to the grid, and with different colors the type of microgrid chosen: 0 is the lowest cost, 1 is the one with at least 50% of RES production and 2 is with 100% RES, i.e. null emissions. Both the social and environmental objectives push toward solutions with only microgrids, so when the cost is not considered, all the clusters are off-grid. When all the three dimensions are considered almost at the same level, balance between on and off-grid systems is created. The least cost microgrids are preferred to the low emission ones only when the environmental dimension is not active. When the social dimension is considered, that is the minimization of energy not supplied, all the connections to C7 are removed, given their long distance. When economical and evironmental dimension are considered, C14 is electrified with a microgrid given its low dimension that does not justifies a new interconnection to

Table 7.8: Summary of optimal electrification solution for each cluster

Cluster	Population [pp]	Area [km2]	Pop density [pp/km2]	Closest grid [km]	Closest cluster [km]	Electr.
0	421	9.6	43.9	17	4	mg
1	393	8.4	46.8	12	4	mg
2	483	9.76	49.5	17	7	mg
3	490	10.8	45.4	2	0.3	grid
4	1104	18.52	59.6	15	0.5	grid
5	3803	39.72	95.7	1	0.3	grid
6	490	9.24	53	18	2	grid
7	27854	249.12	111.8	0	0.3	grid
8	221	4.84	45.7	10	2	mg
9	1008	18.48	54.5	0.6	1	grid
10	863	15.32	56.3	3	2	grid
11	1021	16.28	62.7	2	2	grid
12	5155	69.84	73.8	15	1	grid
13	428	9.52	45	15	0.8	grid
14	221	4	55.3	0.5	2	grid
15	9227	85.68	107.7	11	0.3	grid
16	476	9.68	49.2	1	16	grid
17	849	15.44	55	5	0.5	grid
18	559	9.88	56.6	12	1	grid

the grid.
The graph 7.16a shows the value of the single objective functions in the different simulations performed. Dots are colored according to the values of the multiobjective function: given by the weighted sum of the normalized single objectives (see eq. (6.73)) , it can range from 1, its optimal value, to 0. The value gets closer to 1 when the optimization gets closer to a single-objective, i.e. the weights of two functions are 0 and the remaining is 1. In the other three subfigures of figure 7.16 it can be seen the interdependence of one objective function with respect to the three weights. The value of weights of the three different dimensions are shown in the x-axis, with different colours and different point sizes. The cost function is at its maximum level either when emissions or when energy unavailability have a high weight (light coloured dots for high environmental weight and large size of dots for high social weight); that value is related to the full electrification with microgrids. It then tends to decrease increasing its weight, on the right of the x-axis. The social dimension has also a strong decreasing trend, its value halves when the weight passes from 0 to 1. The coloured dots show clearly how the increase of environmental weight has a positive influence on the reduction of the social

function, opposite to the one of the cost (small dots have a low weight of the cost function).

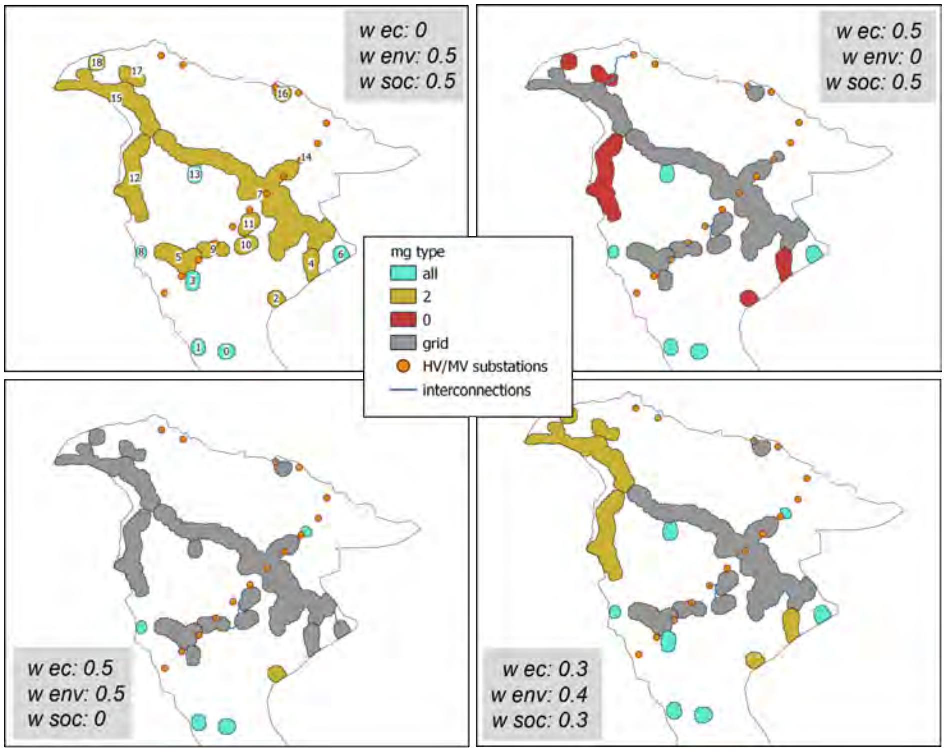
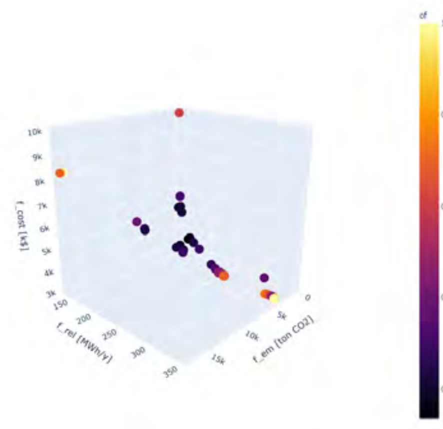
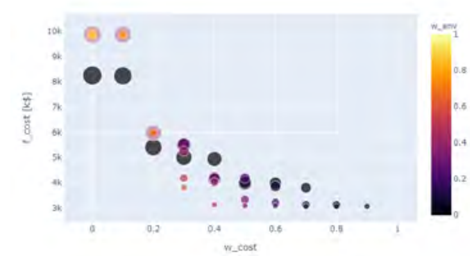


Figure 7.15: Results of multi-objective optimization in Namanjavira

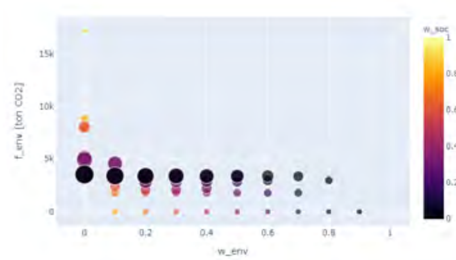
The variation of single objectives when simultaneously changing their weights can be also appreciated in the graphs of figure 7.17. In these graphs also the comparison with the optimization results considering indirect emissions from an LCA perspective is shown. The top three graphs consider the environmental dimension with only direct emissions, the bottom three also include indirect emissions. They are coloured according to the values of each of the three functions, blue when it is minimum and red when is maximum, and the axes show the weights of each dimension. Direct emissions have a maximum value of 17kton corresponding to 312 kg/CO2 per person over the project lifetime while indirect emissions have their maximum at 19 kton corresponding to 352 kg/CO2 per capita. It can be noticed however that, when moving from direct to indirect emissions, the interrelations between the objective functions change only slightly: the area with higher cost increases because the algorithm tends to choose a solution with only microgrids even



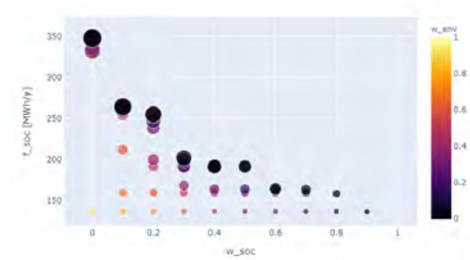
(a) Pareto solutions in Namanjavira



(b) Values of f_{cost} function varying its weight



(c) Values of f_{env} function varying its weight

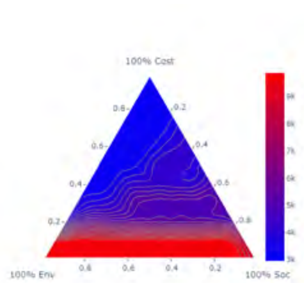


(d) Values of f_{soc} function varying its weight

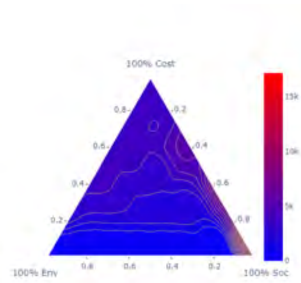
Figure 7.16: Pareto solutions of multi-objective optimization

at lower weights of the environmental function.

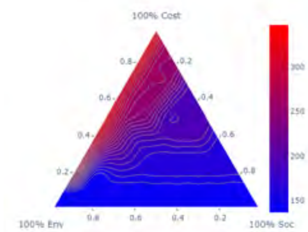
Numerical results from the different optimization strategies in the area are summarized in table 7.9. The lowest NPC from the integrated area optimization (in the table NPC tot) is achieved from the MILP algorithm with 33kV voltage level (i.e. without the activation of voltage constraints). At 11kV the MILP and the genetic lead to the same results even though the genetic algorithm had to be run several times so that the minimum cost solution could be selected. The time for one single run has been of around half an hour. The column *NPC tot* in the table does not include the cost of building electric lines inside communities that are shown in the column *NPC int grids*. Actually this is the largest component of cost, counting as more than double the sum of other costs. In the last line it is shown the largest NPC, obtained when in the multi-objective optimization the cost function has weight=0 and



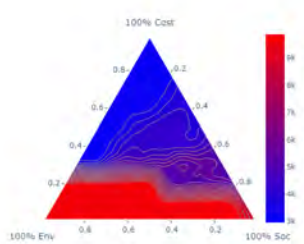
(a) NPC [k$] (dir. em.)



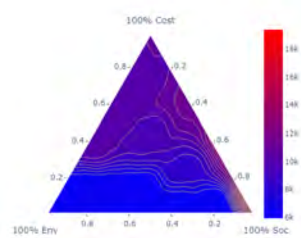
(b) CO2 [ton] (dir. em.)



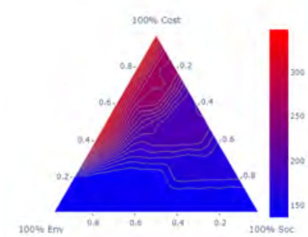
(c) Unavailability [MWh] (dir. em.)



(d) NPC [k$](indir. em.)



(e) CO2 [ton](indir. em.)



(f) Unavailability [MWh](indir. em.)

Figure 7.17: results of multi-objective optimization with different weights

all the communities are electrified with microgrids. In this case, the cost for microgrids generation sources (NPC mg) and the ones for internal grids are comparable, to demonstrate the importance of including the grid design even in tools for off-grid systems sizing.

Finally, table 7.10 summarizes the computational time required from the different modules of the Gisele procedure. Those numbers do not consider all the pre analyses that are required to identify the inputs combinations (e.g. clustering sensitivity) bust just report the time for running each of the algorithms. The energy resources assessment process is still partially manual for the creation of hydropower resource layer; moreover, the retrieval of wind and solar data with API from Renewables.ninja has some limitations on the hourly quantity of data so it could cause some delays when the request is iterated for several communities. The most time consuming steps are the grid routing procedure, mainly due to the algorithm M1 and the input data processing for the final MILP optimization, where Dijsktra algorithm has to run several times on long distances. The time for the whole procedure is almost 4 hours.

Table 7.9: Summary of optimization strategies

Opt strategy	Voltage	NPC tot	mg	NPC mg	NPC interc	NPC grid en	NPC int grids	time [s]
milp	33	2919.794272	C0,C1,C2,C8	300.1765	336.438	2283.179772	7345.955	2
milp	11	3066.039272	C0,C1,C2,C8	300.1765	482.683	2283.179772	7345.955	10
genetic (min)	11	3066.039272	C0,C1,C2,C8	300.1765	482.683	2283.179772	7345.955	1800
milp mo	33	9879	all	9879	0	0	7345.955	10

Table 7.10: Computational time

Block	**Module**	**Time**
Block1	1.1 Identification of communities	16 seconds
Block1	1.2 Load estimation %	14 minutes
Block2	2.1 Energy resource assessment	partially manual
Block2	2.2 Microgrid sizing	31 minutes
Block3	3.1 Secondary Substations siting	1 minute
Block3	3.2 Cost surface creation	15 minutes
Block3	3.3 Grid routing	2 hours
Block4	Input data processing	1 hour 45 minutes
Block4	4.1 Single-objective optimization	2 seconds
Block4	4.2 Multi-objective optimization	10 seconds

Chapter 8

Lesotho-Butha Buthe

The goal of the analysis is the study of the optimal electrification strategy of the of Butha-Buthe region, in Lesotho. The area has been selected thanks to a collaboration with the local DSO, which allowed to collect data useful for the analysis. Simulations have been run with a processor Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz 3.00 GHz, with 128 Gb of RAM. In the following sections, the data and the results obtained applying Gisele procedure are described.

8.1 Case study set-up

The specific case study allows to test the developed procedure in a context different to the Namanjavira case. First of all, the area is in prevalence mountainous, and the population is distributed along the valleys. Secondly, Lesotho has a higher GDP per capita, population has higher living standards that lead to higher requirements in terms of energy needs. Finally, available data are much more complete and accurate, and they provide indication on the in place distribution grid, communities characteristics and costs of components. In this case, the main challenge is related to the tractability of the problem, since in the area there is a high number of small communities, that could cause convergence problems during the process of integrated area optimization. Nevertheless, the wide number of communities offers the opportunity of drawing conclusions on the applicability of the algorithms and generalize the trends of optimal electrification strategy identification.

Block1: Energy demand assessment

Residential areas of Butha-Buthe region are mapped on OSM, and the map of the distribution grid serving the area was provided by the contacts in Lesotho. Those data were used as starting point for identyifing communities to be electrified without the need of population clustering. Load profiles were generated starting from MTF data and adapting them to the energy and peak load values considered in Lesotho masterplan for electrification (see table 8.1). The categories of users and the associated RAMP input parameters are reported in annex A. The number of households in each community are used as a proxy of the number of users of the other categories.

Table 8.1: User classes parameters extracted from Lesotho masterplan

User	Peak Power [kW]	Energy [MWh/year]
Households	0.4	2.8
Schools	0.8	12.6
Health Centers	1.5	15.4
Churches	0.5	3.7
Other Businesses	0.8	5.9

Block2: Off-grid system sizing

All the input parameters required by the MILP procedure for microgrid sizing are detailed in annex B. They derive from literature data. For each river, the turbine with the size closer to the average power production is chosen. Rivers with flow rate higher than 20 m^3/s or average power production higher than 1 MW are excluded from the analysis because they would be suited for large power plants and not exploitable by small scale microgrids.

Block3: Internal grids design

The maximum length for low voltage grid is set at 500m, data coming from discussion with local DSO. The geospatial data used in the procedure come from the same sources of the Mozambique case study (see tab. 7.2). The weighting procedure for the cost surface creation follows the criteria and the numerical coefficients detailed in section 6.3.4.

Block4: Integrated area optimization

Geospatial data related to the in place transmission and distribution network in the country was available, together with the position and voltage level of the substations. Following indications from Lesotho Masterplan, voltage level for grid expansion is set at 11kV. The possible points of connection to the in place grid are both the 33/11 kV substations and the existing 11kV feeders. In the last case, the author made assumptions on the possible position of these (found according to the roads and communities' position) and on their voltage level, lower than unitary per unit value, to consider losses along the line. The costs for connecting new lines in the two possible configurations have been provided by the contact in Lesotho and are reported in table 8.2 together with the electrical parameters assumed for the distribution feeders. Lesotho electricity demand is fulfilled by the 72 MW hydroelectric power plant installed in Muela and by the imports from the surrounding country of South Africa and from Mozambique. The electricity price depends on the tariffs of the bilateral agreements Lesotho Electricity Company (LEC) stipulated with the different importing companies and vary according to the seasonality and to the daily hours. The average price is around 0.06$/kWh, but it can reach up to 0.2$/kWh in the peak hours of the peak season. Simulations have hence been run with different values of Cost of Energy (COE) to analyze the effect of the price on the optimal electrification strategy. As concerns the emissions related to the production of energy from the national grid, the direct emissions related to Lesotho internal power production are null, since the only production plant is a hydroelectric plant. However, around 40% of the total Lesotho electricity consumption comes form imports from Mozambique and South Africa, hence the national emissions are computed as a weighted average of the emissions derived from the energy mix of these

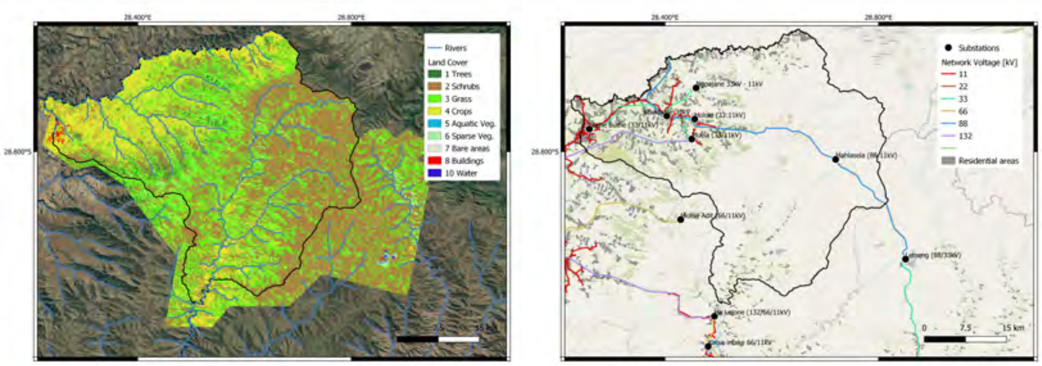
Table 8.2: Electrical parameters in Lesotho

Data	Value	u.m.
Investment cost	12.4	k$/km
Lifetime	40	years
r	0.45	ohm/km
x	0.36	ohm/km
c	9.7	nF/km
I	0.26	kA
coe	0.06	$/kWh
Connect. 11kV feeder	0.55	k$
Connect. 33/11kV sub	1.25	k$

countries.

8.2 Results and discussion

The main geospatial data gathered for the area under consideration are shown in figure 8.1. Butha Buthe is a mountain area, covered by 44% of shrubs, 33% by grassland and 17% by cropland.



(a) Landcover and rivers (b) Population, transmission grid and roads

Figure 8.1: Geospatial data of Butha-Buthe region

Block1: Energy demand assessment

The region of Butha-Buthe has a surface of $1788 km^2$ and a total population of 118,242 inhabitants with a population density of 66 pp/km^2. Geospatial data of the distribution grid in place (red and green lines) and of the residential areas (grey polygons) have been exploited to identify communities to be electrified as shown in figure 8.2. A buffer of 1000 m has been created around the 11kV and 33kV grids (the striped area), supposing that the communities within the buffer already have access to electricity, or will be connected in a short time. As for the other communities, a buffer around residential areas, with a radius of 100m has been drawn, to unify into single clusters close communities. The result is a total of 101 clustered communities (the coloured polygons), with a population ranging from 100 people (blue polygons) since smaller communities are discarded, to 3550 people (yellow polygons). Figure 8.3 shows the number ID associated to each of the communities.

Load profiles are created using as benchmark the power values from the Lesotho masterplan (see table 8.1). Those energy consumption and peak

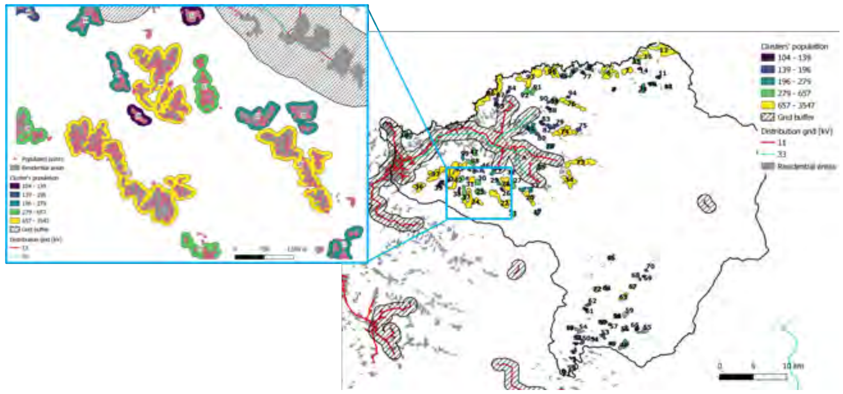


Figure 8.2: Identified communities in Butha-Buthe

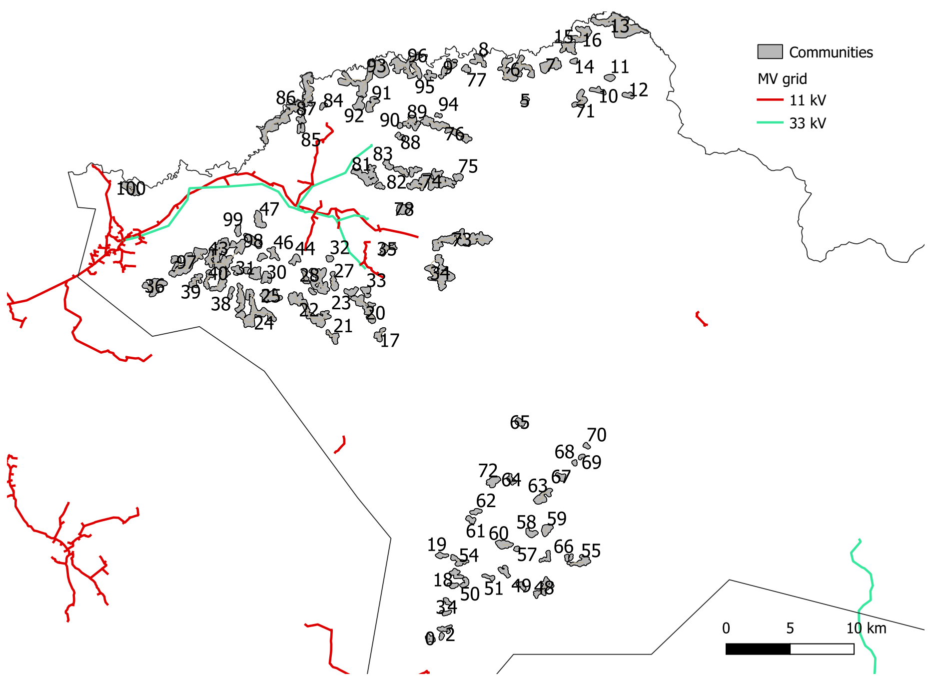
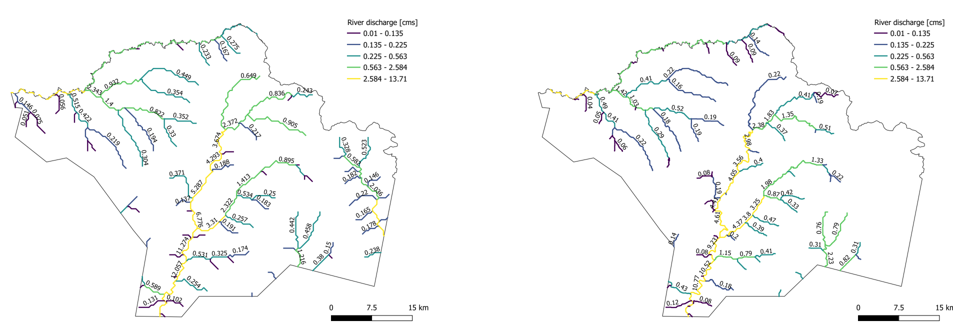


Figure 8.3: Identified communities in Butha-Buthe

power values are set as target load values to be reached at the end of project lifetime (after 10 years). From year 0 to years 10 a linear load growth trend of 2% has been considered.

Block2: Off-grid system sizing

The hydro resource availability is obtained by running the SWAT procedure on two hydro basins that cover the area. The computed average annual river flow rate was scaled and multiplied by a factor of 2 to arrive at the results of figure 8.4, comparable to the ones of the HydroRivers database: the order of magnitude of river flow rate is the same in the two cases, with values ranging from 0.05 to 25 m^3/s in the Hydrorivers database and from 0.01 to 30 m^3/s in the computed scenario. In the sub figure 8.4a, showing the HydroRivers vector layer, also rivers belonging to other basins, not simulated with SWAT since they are outside the populated areas are visible. The two



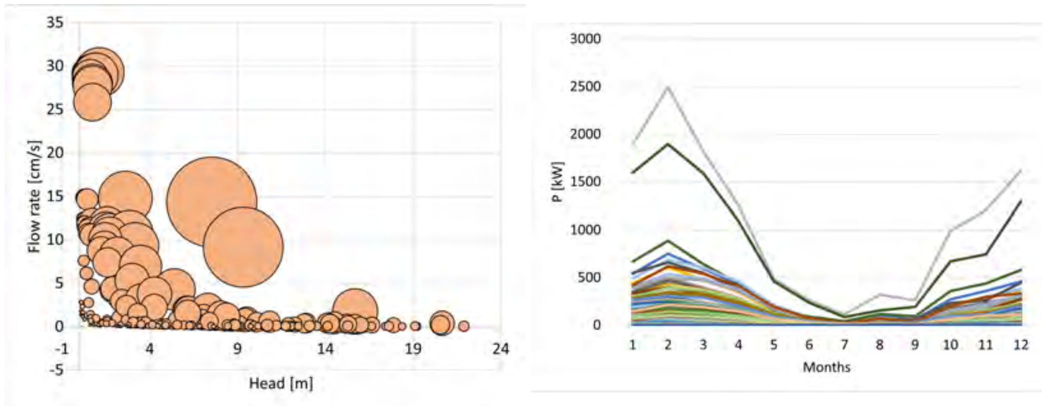
(a) HydroRivers

(b) SWAT procedure

Figure 8.4: Average annual river discharge [cms] from literature data and SWAT procedure implemented

images of figure 8.5 resume the main characteristics of the rivers in the area, computed with the proposed procedure. Subfigure 8.5a shows the average annual flow rate and the average head each 200m along the rivers. The bigger the dots, the higher the available average power availability. The two bigger dots, corresponding to the rivers with higher potential, are easily spotted also in subfigure 8.5a, as the two upper curves. The subfigure in fact shows the monthly average rivers' power output, also considering deviation channels' length of 200m. Given the dependence of the rivers' flow rate on the rainfall regime, during the central months of the years the power potential is significantly reduced, up to 95% reduction from February to July. This shows how important it is to consider the seasonality of flow regimes to proper size RES based systems.

In a second step, the procedure for microgrid sizing has been run for each of the communities, with three different constraints of minimum RES pene-



(a) Computed rivers flow rate vs head

(b) Computed rivers' monthly power

Figure 8.5: Butha-Buthe rivers' characteristics

tration (0%, 50% and 100%); for sake of simplicity, only the results related to the unconstrained simulation, with minimum RES of 0% are reported in table 8.3 since the others are only used by the multiobjective optimization module. PV, BESS and Diesel generators are installed in all the communities while wind resource does not result to be convenient. Communities 66 and 69 represent an exception, where the small size and the availability of hydro resource allow to have full RES based microgrid with few PV and BESS installed. It is noticeable how the presence of hydropower resource allows to decrease the microgrids' LCOE from around 0.25 $/kWh to 0.17 $/kWh in the best cases. Hydroturbines are installed in 49 out of the 101 communities with sizes ranging from 10 to 50 kW. It is worth mentioning that the procedure considers is a distributed hydropower potential, where it is assumed that several hydropower turbines could be installed along the same river to supply all the microgrids. This simplification allows to achieve a first indication of costs for microgrids deployment; actually, further analysis should evaluate the possibility of supplying several communities with a bigger power plant and consider constraints that do not allow to install multiple turbines at short distance along the same rivers. BESS capacity to PV ratio ranges from 0.2 to 3 hours when the Diesel capacity installed is smaller and the BESS need to supply power for longer hours.

ID	PV [kW]	Hydro [kW]	Diesel [kW]	BESS [kWh]	Inverter [kW]	NPC [k$]	En. Prod. [MWh]	LCOE [$/kWh]	CO2 [ton]	CO2 spec. [kg/pp]
C0	10	20	16	9	12	234	1886.28	0.17	119	572
C1	20	10	16	56	29	187	1303.85	0.19	60	552

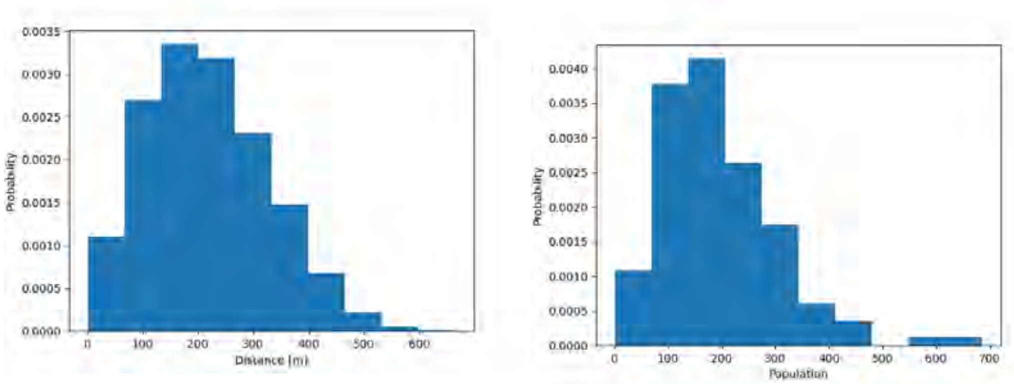
ID	PV [kW]	Hydro [kW]	Diesel [kW]	BESS [kWh]	Inverter [kW]	NPC [k$]	En. Prod. [MWh]	LCOE [$/kWh]	CO2 [ton]	CO2 spec. [kg/pp]
C2	13	20	16	33	19	252	1880.95	0.18	55	264
C3	13	20	16	17	20	223	1476.52	0.2	40	288
C4	5	20	16	2	6	173	1387.45	0.17	6	52
C5	30	0	16	26	40	236	1296.41	0.25	602	5540
C6	185	20	96	98	220	1539	8730.37	0.24	3716	2757
C7	116	10	64	99	164	973	5475.47	0.24	2285	2841
C8	54	10	16	98	72	390	2193.57	0.24	541	2110
C9	59	10	16	125	80	411	2327.25	0.24	563	2022
C10	35	0	16	28	41	260	1457.81	0.24	671	4982
C11	36	0	16	32	44	265	1479.84	0.24	671	4820
C12	38	0	16	43	46	280	1561.41	0.24	682	4483
C13	197	10	96	137	219	1630	9183.9	0.24	4136	2909
C14	23	10	16	20	27	231	1269.81	0.25	322	3090
C15	94	10	48	81	111	799	4597.46	0.24	1884	2869
C16	110	10	64	49	125	933	5265.4	0.24	2325	3021
C17	61	10	16	133	84	411	2332.83	0.24	543	1948
C18	9	20	16	4	10	184	1386.12	0.18	41	331
C19	15	20	16	15	21	221	1509.15	0.2	42	289
C20	153	10	64	93	227	1027	5610.56	0.25	2331	2824
C21	57	0	32	25	66	453	2505.13	0.25	1242	4012
C22	220	10	112	109	236	1822	10225.64	0.24	4801	3009
C23	56	0	16	116	74	359	1958.98	0.25	664	3065
C24	218	0	112	97	231	1727	9773.8	0.24	4858	3191
C25	104	0	48	100	120	807	4522.56	0.24	2083	3231
C26	35	0	16	14	42	251	1404.18	0.24	675	5295
C27	79	0	64	33	87	707	3830.64	0.25	1966	3701
C28	145	0	80	50	152	1146	6455.82	0.24	3250	3356
C29	44	0	32	51	53	344	1885.8	0.25	830	4018
C30	106	0	48	104	122	813	4554.85	0.24	2082	3204
C31	113	0	48	143	137	854	4780.5	0.24	2085	3038
C32	34	0	16	29	39	263	1481.74	0.24	685	4930
C33	53	0	32	83	66	404	2165.78	0.25	899	3558
C34	218	40	128	51	232	2007	11499.68	0.24	4865	2687
C35	56	10	32	52	65	487	2746.48	0.24	993	2831
C36	124	0	64	64	133	964	5411.61	0.24	2654	3345
C37	54	0	32	56	64	436	2378.98	0.25	1093	3787
C38	39	0	16	15	46	255	1413.73	0.25	670	5222
C39	40	0	16	10	47	242	1313.88	0.25	627	5585
C40	37	0	16	9	41	247	1368.45	0.24	665	5527
C41	38	0	16	36	45	273	1517.96	0.24	676	4682
C42	52	0	32	53	60	429	2322.25	0.25	1084	3886
C43	505	0	304	132	528	4438	24247.88	0.25	12240	3450
C44	46	0	32	57	57	352	1913.11	0.25	828	3927
C45	46	0	32	58	58	336	1800.98	0.25	759	3965
C46	40	0	16	53	49	291	1626.57	0.24	686	4217
C47	98	0	48	63	111	767	4316.89	0.24	2080	3404
C48	20	30	16	12	25	330	2499.85	0.18	251	812
C49	19	20	16	12	23	288	2030.13	0.19	256	1101
C50	7	20	16	6	10	230	1817.51	0.17	108	551
C51	19	20	16	27	25	244	1823.48	0.18	103	525
C52	15	20	16	19	20	213	1604.07	0.18	54	340
C53	9	20	16	6	12	226	1575.08	0.19	97	625
C54	18	30	16	14	21	373	2721.93	0.19	234	678
C55	57	50	16	87	76	572	4557.71	0.17	262	403
C56	61	0	16	145	84	374	2053.4	0.25	631	2718
C57	34	0	16	28	39	263	1480.17	0.24	689	4947

ID	PV [kW]	Hydro [kW]	Diesel [kW]	BESS [kWh]	Inverter [kW]	NPC [k$]	En. Prod. [MWh]	LCOE [$/kWh]	CO2 [ton]	CO2 spec. [kg/pp]
C58	51	0	16	86	66	328	1799.07	0.25	664	3477
C59	78	20	32	86	93	673	4058.88	0.23	1193	2103
C60	51	0	32	56	60	397	2162.08	0.25	970	3837
C61	37	0	16	9	43	245	1356.98	0.25	659	5553
C62	43	0	16	87	58	300	1642.39	0.25	609	3688
C63	111	20	64	73	124	1016	6095.15	0.23	2336	2573
C64	32	20	16	23	41	288	1940.2	0.2	293	1353
C65	11	20	16	22	18	209	1916.91	0.15	46	217
C66	6	20	0	2	6	163	1320.83	0.17	0	0
C67	28	30	16	22	36	374	2931.42	0.17	292	766
C68	11	20	16	10	15	197	1608.1	0.17	54	337
C69	17	20	0	15	20	182	1385.03	0.18	0	0
C70	18	20	16	22	24	231	1796.25	0.17	72	375
C71	59	0	16	127	80	369	2042.5	0.25	671	2914
C72	52	20	16	91	68	420	2665.04	0.21	469	1397
C73	264	50	144	75	293	2378	13708.81	0.24	5748	2641
C74	113	20	64	59	127	1038	5817.86	0.24	2404	2793
C75	38	0	16	40	50	274	1533.1	0.24	667	4512
C76	105	10	48	105	124	860	4915	0.24	2005	2830
C77	24	10	16	21	29	234	1300.25	0.24	356	3275
C78	48	10	16	77	61	366	2102.29	0.24	573	2362
C79	36	0	16	34	45	268	1504.71	0.24	677	4720
C80	34	10	16	34	42	276	1535.87	0.24	405	2737
C81	53	10	16	103	71	406	2275.59	0.24	578	2144
C82	38	0	16	42	45	277	1558.39	0.24	680	4467
C83	30	0	16	39	42	240	1302.64	0.25	571	5252
C84	30	0	16	12	38	231	1268.54	0.25	624	5979
C85	35	0	16	19	43	256	1429.99	0.24	681	5171
C86	276	10	144	111	292	2276	12798.43	0.24	6161	3041
C87	35	0	16	34	41	268	1505.32	0.24	682	4748
C88	26	10	16	31	36	231	1268.69	0.25	306	2928
C89	72	10	32	63	96	532	2995.17	0.24	1099	2809
C90	29	10	16	24	34	267	1507.05	0.24	463	3227
C91	61	10	16	132	84	427	2354.89	0.25	556	1969
C92	49	10	32	11	51	474	2574.33	0.25	1072	3332
C93	256	20	144	66	274	2205	12478.33	0.24	5866	2972
C94	26	10	16	24	32	238	1347.92	0.24	374	3184
C95	41	10	16	57	50	332	1851.8	0.24	534	2672
C96	156	20	80	67	202	1330	7460.65	0.24	3175	2798
C97	309	0	160	126	326	2458	13889.86	0.24	6956	3150
C98	85	0	48	51	92	676	3769.08	0.24	1834	3524
C99	54	0	16	100	69	344	1884.68	0.25	669	3273
C100	85	0	48	51	95	683	3826.98	0.24	1866	3517

Table 8.3: Results of microgrids sizing procedure

Block3: Internal grids design

For the design of internal grids, module 3.3 b (see fig. 6.22) was used, hence the first step has been the siting of the MV/LV transformers with agglomerative clustering algorithm. Figure 8.6 shows the probability distribution of the distances between populated points and the corresponding substations. It can be seen that just in the 1% of cases the predefined threshold of 500m is not respected. The LV clusters supply an average of 195 people, ranging between a minimum of 12 to a maximum of 686. 57 communities don't need the design of the internal grid because, according to the assumptions adopted, they are small enough to be supplied with LV grid through a single MV/LV transformer.



(a) Probability distribution of points-centers distances

(b) Probability distribution of population

Figure 8.6: Results of secondary substations siting in Butha-Buthe

The second step of the internal grids' routing procedure has been the creation of cost surface whose result is shown in figure 8.7. Here it is evident the influence of the terrain's slope: flat valleys have a penalty factor around 1 while mountains' sides have values up to 20. In the image the penalty factor (the coefficient to be multiplied by the lines base cost to obtain the total NPC) is computed considering the NPC over the grid lifetime (i.e. 40 years). As for the internal grids' creation, model M2 (MST+Dijsktra) leads to the same or even better (lower cost) results than model M1 in all the communities. Detailed results related to each communities' main characteristics obtained from Block1 (fig. 6.3) and Block3 (fig. 6.22) procedure are reported in table 8.4. The total low voltage length is estimated as the sum of the distances between each household and the MV/LV transformer, so it is slightly overestimated, while the MV grid length is more accurate being designed with the grid routing procedure.

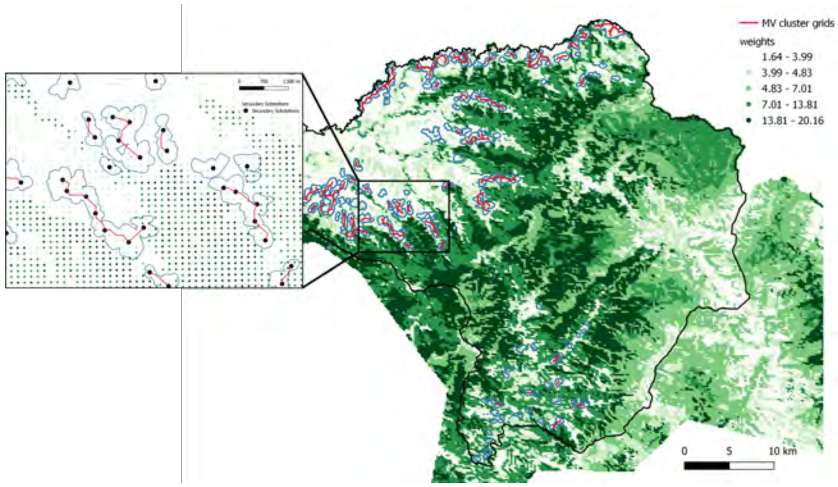


Figure 8.7: Results of the internal grids routing procedure

ID	Pop.	Load [kW]	Area [km2]	Load/cap [kW/pp]	Pop. dens. [pp/m2]	MV/LV [num]	Length LV [km]	Length MV [km]	NPC [k$]	Alg.
0	208	33	0.41	0.16	512.5	1	8	0	0	
1	108	22.4	0.21	0.21	522.6	1	2.8	0	0	
2	206	33.7	0.49	0.16	424.2	1	9.1	0	0	
3	139	25.5	0.3	0.18	459.1	1	4	0	0	
4	124	24.4	0.27	0.2	457	1	4.1	0	0	
5	109	21.8	0.31	0.2	350.3	1	3.3	0	0	
6	1348	141.2	3.17	0.1	424.8	7	76.1	5.4	62.6	M2
7	804	88	1.39	0.11	577.6	3	39.7	1.2	13.9	M2
8	257	35.1	0.79	0.14	323.9	2	14.2	0.6	7.2	M2
9	278	37.7	1	0.14	278.9	2	13.3	0.9	13.5	M2
10	135	25.3	0.5	0.19	266.9	1	8	0	0	
11	139	27.8	0.41	0.2	340.3	1	5	0	0	
12	152	27.3	0.4	0.18	382.8	1	6	0	0	
13	1422	148.2	4.31	0.1	329.6	8	71	5.8	68.9	M2
14	104	21.3	0.25	0.2	410.7	1	2.5	0	0	
15	657	74	0.96	0.11	683.2	3	32	1.5	18.9	M2
16	770	84.5	1.45	0.11	531.2	4	36.7	1.8	22.1	M2
17	279	38.3	0.57	0.14	491.3	2	9.3	0.7	5.3	M2
18	124	24.2	0.33	0.2	370.7	1	4	0	0	
19	144	26.1	0.37	0.18	385.9	1	5.3	0	0	
20	825	90	1.86	0.11	443.6	5	29.8	2.6	31.5	M2
21	310	41.7	0.61	0.13	503.8	2	8.6	0.8	11.7	M2
22	1596	165.6	3.17	0.1	503	7	77	4.3	52.7	M2
23	217	35.4	0.58	0.16	373.5	1	8.6	0	0	
24	1522	158.7	3.3	0.1	460.6	7	73.8	5.4	68.8	M2
25	645	72.1	1.34	0.11	481.8	3	26.6	1.1	12.1	M2
26	127	23.7	0.23	0.19	545	1	3.7	0	0	
27	531	62.8	0.72	0.12	733.8	2	21.3	0.8	8.9	M2
28	968	104.1	2.23	0.11	434.3	4	57.9	2.1	23.8	M2
29	207	33.4	0.43	0.16	485.9	2	6.7	0.7	6.2	M2
30	650	73.1	0.68	0.11	951.5	1	27.7	0	0	
31	686	76.8	0.83	0.11	821.9	1	36.3	0	0	

ID	Pop.	Load [kW]	Area [km2]	Load/cap [kW/pp]	Pop. dens. [pp/m2]	MV/LV [num]	Length LV [km]	Length MV [km]	NPC [k$]	Alg.
32	139	25.5	0.28	0.18	494.9	1	3.6	0	0	
33	253	34.9	0.55	0.14	461.9	1	10.4	0	0	
34	1811	186	3.1	0.1	584	7	76.3	4.3	60.2	M2
35	351	45.5	0.72	0.13	484.2	2	16.1	0.8	7.7	M2
36	794	87.1	1.68	0.11	471.5	4	48.1	1.7	22.2	M2
37	289	38.8	1.18	0.13	244.9	3	16.4	1.7	21	M2
38	128	24.8	0.36	0.19	355.6	1	7.9	0	0	
39	112	22.6	0.78	0.2	143.6	2	5.9	0.8	10.6	M2
40	120	23.6	0.62	0.2	194.7	2	4.9	1.1	14.5	M2
41	144	25.9	0.21	0.18	686.7	1	2.8	0	0	
42	279	38.1	0.47	0.14	598.6	1	12.6	0	0	
43	3547	394.6	5.84	0.11	607.8	13	205.3	8.4	95.6	M2
44	211	35.2	0.53	0.17	397.9	1	8.9	0	0	
45	191	32.1	0.36	0.17	536.5	1	6.7	0	0	
46	163	28.9	0.6	0.18	269.7	2	5.4	0.4	4.3	M2
47	611	69.5	0.89	0.11	683.3	2	32.5	0.7	8.2	M2
48	310	40.9	1.18	0.13	261.6	2	16.3	1.1	9.1	M2
49	232	33.6	0.35	0.14	654.5	1	7.7	0	0	
50	196	31.9	0.41	0.16	480.1	1	9.6	0	0	
51	196	32.3	0.33	0.16	595.1	1	9.2	0	0	
52	160	28.5	0.59	0.18	272.2	1	7.6	0	0	
53	155	27.5	0.56	0.18	275.9	1	8.2	0	0	
54	346	45.3	0.67	0.13	513.6	2	15.8	0.6	6.6	M2
55	650	73	0.92	0.11	707	2	33.6	1	7.3	M2
56	232	32.8	0.47	0.14	498.5	1	13	0	0	
57	139	25.6	0.19	0.18	742.3	1	2.5	0	0	
58	191	32.7	0.5	0.17	384.3	1	6.7	0	0	
59	567	65.2	0.65	0.11	868.3	2	18.4	0.5	3.6	M2
60	253	35.1	0.77	0.14	327	2	9.3	0.4	3.2	M2
61	119	23.4	0.38	0.2	316	1	5.1	0	0	
62	165	29	0.36	0.18	454.9	1	6.7	0	0	
63	908	98.6	1.03	0.11	883.1	3	32.5	1.2	13.1	M2
64	217	34.3	0.38	0.16	564.7	1	8.5	0	0	
65	211	34.3	0.41	0.16	517.9	1	7.7	0	0	
66	113	22.7	0.33	0.2	340.7	1	4.7	0	0	
67	382	48.5	0.37	0.13	1023.7	1	12.1	0	0	
68	160	28.1	0.18	0.18	904.7	1	2.6	0	0	
69	124	23.8	0.18	0.19	686.2	1	2.7	0	0	
70	191	32	0.21	0.17	892	1	4.2	0	0	
71	230	32.5	1.02	0.14	224.9	2	14.3	0.9	8.8	M2
72	335	44	0.71	0.13	471	1	19.6	0	0	
73	2177	221.3	3.89	0.1	559.4	8	101	5.2	65.9	M2
74	861	93.9	3.73	0.11	230.5	7	49.6	4.1	56.2	M2
75	148	26.1	0.44	0.18	338.4	1	6.8	0	0	
76	709	79.2	2.47	0.11	286.5	5	39.7	3.8	55.5	M2
77	109	21.8	0.37	0.2	293.6	1	4	0	0	
78	242	34.1	0.92	0.14	263.6	1	13.4	0	0	
79	143	25.5	0.54	0.18	267.8	1	9.4	0	0	
80	148	26.7	0.33	0.18	450.5	1	5.1	0	0	
81	270	37.2	2	0.14	134.7	1	15.6	0	0	
82	152	27.1	0.59	0.18	256.6	2	5.7	0.6	9.8	M2
83	109	22.5	0.45	0.21	240.1	1	5.9	0	0	
84	104	21.1	0.26	0.2	398.6	1	3.3	0	0	
85	132	24.3	0.43	0.18	302.8	1	7	0	0	
86	2026	207.8	5.36	0.1	378.1	13	94.9	8.2	86.7	M2
87	144	26.4	0.37	0.18	384.5	1	6.3	0	0	

ID	Pop.	Load [kW]	Area [km2]	Load/cap [kW/pp]	Pop. dens. [pp/m2]	MV/LV [num]	Length LV [km]	Length MV [km]	NPC [k$]	Alg.
88	104	20.9	0.38	0.2	273	1	4.5	0	0	
89	391	49.3	1.07	0.13	365	3	20.9	1.2	14.6	M2
90	143	26.2	0.19	0.18	745	1	3.1	0	0	
91	283	37.9	0.87	0.13	326.3	2	12.5	0.8	12	M2
92	322	42.1	0.75	0.13	430.8	1	21.5	0	0	
93	1974	202.9	4.38	0.1	450.9	10	94.6	7	87.8	M2
94	117	22.8	0.19	0.19	618.8	1	4.2	0	0	
95	200	32.7	0.53	0.16	380	1	12.7	0	0	
96	1135	119.7	2.75	0.11	413	7	51.4	4.5	57.6	M2
97	2209	224.5	3.41	0.1	648.5	8	126.4	4.2	50.8	M2
98	521	61.1	0.88	0.12	592.3	2	23.9	0.7	7.4	M2
99	204	33.5	0.41	0.16	503.8	1	10.4	0	0	
100	530	61.6	1.08	0.12	492.1	3	38.3	1.2	17.1	M2

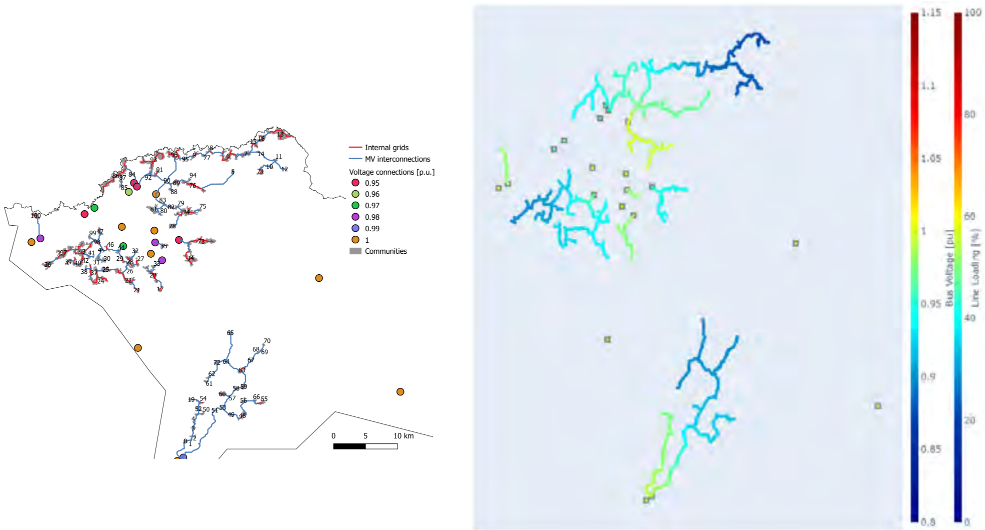
Table 8.4: Results of internal grid routing procedure

Block4: Integrated area optimization

As a first step, the possible connections between communities are identified using Delaunay triangulation and secondly, their cost is evaluated with the shortest path (Dijsktra) calculation. The total number of possible connections is 453 which significantly reduces the size of the problem with respect to the modeling of a complete graph, that would have more than 10 thousand connections when considering all the combinations between communities. The result is a MILP problem with 1024 integer variables and 1024 continuous variables. Figure 8.8 shows the results of the MILP algorithm run with a COE of the national grid of 0.06 $/kWh. The possible points of connection to the inplace grid have different per unit voltage values, as shown by the coloured dots in the picture, to account for the voltage drops along the medium voltage lines.

Both considering or not hydropower resource, all the communities are connected to the national grid and no microgrids are installed. It can be seen how only few communities are directly connected to the grid, while the majority exploit connections to the neighbouring communities. This demonstrates the importance of an integrated optimization, where the planning is done for all the communities in the same moment.

The inclusion of voltage drop constraints leads to the creation of a grid that is not the one with lowest cost, to avoid that voltage in any node decreases below 0.9 p.u.. The percentage difference between the NPC of the two solutions is just of the 0.5%, actually negligible if considering the set solver gap of 1%. However, due to the approximations of the linearized formula



(a) Load flow results

Figure 8.8: Results of integrated area optimization, milp algorithm

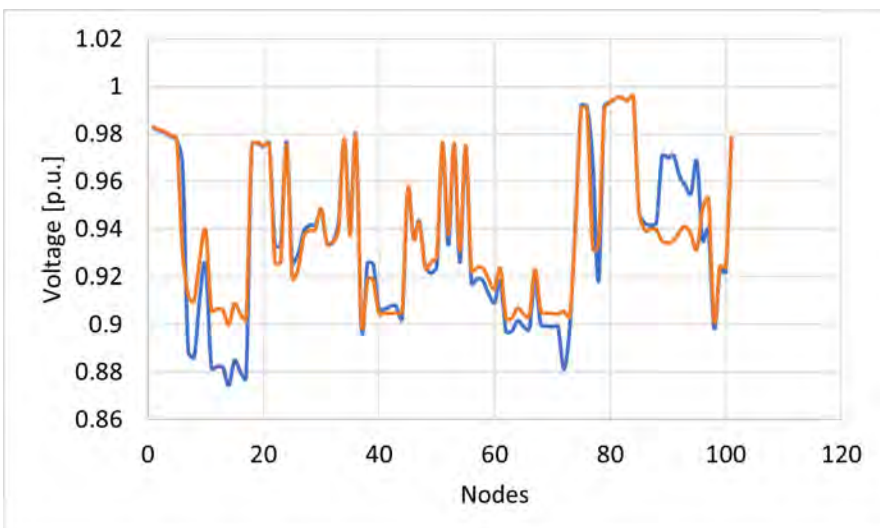
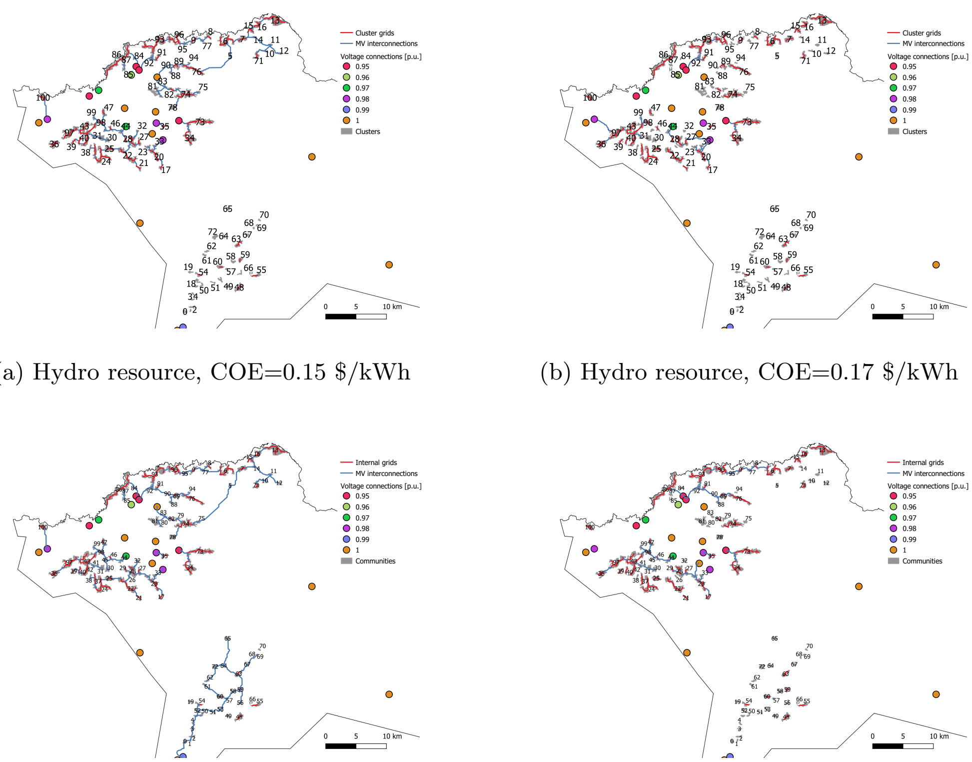


Figure 8.9: Comparison between real voltage and MILP estimation in the grid nodes

for voltage drops included in the MILP algorithm and to the neglection of power losses (resulting equal to the 5% of total load demand) the power flow computed on the resulting grid shows voltage levels lower than the minimum admissible, reaching 0.88, as shown in figure 8.8a. The highest error in the p.u. voltage estimation by the MILP is of 4%, while the average is of 1%. It is lower than in the case of Mozambique since the size of clusters is lower and hence their approximation with just one central node is closer to reality.

To have a more complete view on the possible electrification solutions, a sensitivity analysis on the grid COE has been run and both the cases with and without hydropower resource have been considered (fig. 8.10).



(a) Hydro resource, COE=0.15 $/kWh

(b) Hydro resource, COE=0.17 $/kWh

(c) No hydro resource, COE=0.15 $/kWh

(d) No hydro resource, COE=0.15 $/kWh

Figure 8.10: Sensitivity analysis on integrated area optimization

When no hydropower resource for off-grid systems is considered, below 15$ /kWh the solution remains unchanged and all the communities are connected to the grid. Overcoming this value, some communities start being electrified with microgrids. The first communities to be disconnected from the main grid are the south-eastern communities n. 19, 49, 48, 55, 66, 70. With a COE of 0.17 $/kWh the number of installed microgrids reaches 65, totally disconnecting the south-eastern area. When hydropower resource is considered, the threshold for microgrids installation is lower and community n. 65

becomes disconnected already at a COE of 0.10 $/kWh; the southeastern is electrified with off-grid systems at a COE of 0.15 $/kWh and subsequently the most remote communities of the northwestern part are disconnected.

Actually, the southestern area was identified in the Lesotho electrification master plan (EMP) as a possible site for off-grid system deployment, as shown in figure 8.11, where the communities polygons with their IDs are superimposed to the masterplan map. In particular, the deployment of a mini-hydro turbine of around 500kW on the Motete river was proposed as a solution to provide energy to the three villages of Motete, Kao, and Liqhobong (see also chapter 5), corresponding approximately to the communities from numbers 55 to 70. The analysis considered an average river flow rate of 2 m^3/s, in line with what estimated by the SWAT procedure and a head of around 30m. The approach followed in the masterplan, with a single bigger turbine exploiting a bigger jump, possibly also with a small dam, differentiates from the distributed resource estimation here proposed; however, as shown they can lead to similar results useful in a preliminary phase.

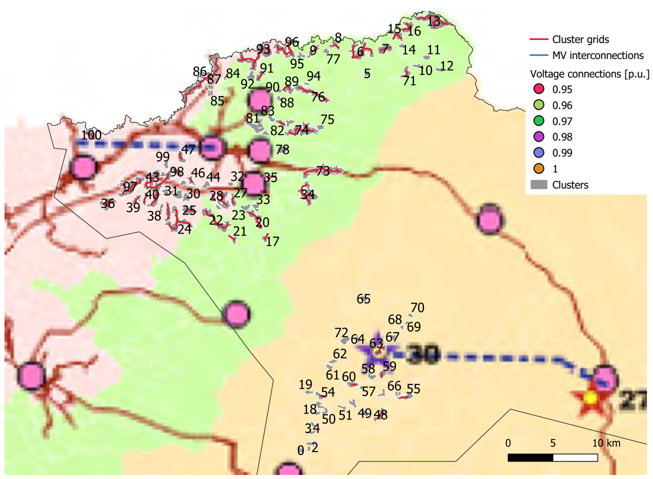


Figure 8.11: Lesotho electrification masterplan: purple stars are proposed off-grid sites, pink dots are substations, red lines are the inplace network, dotted lines are the proposed new high 88kV lines

Table 8.5 reports the numerical values of the listed scenarios, with the decomposition of the total NPC between the cost of interconnecting communities, the cost of buying energy from the grid, the cost of supplying energy through microgrids and the cost of installing MV grid inside communities.

Table 8.5: MILP optimization results changing grid cost of energy and hydropower consideration

COE	hydro	npc	N mg	npc (mg)	npc (connections)	npc energy from grid	npc (int grid)	time [s]
0.06	no	24413	0	0	2066	21101	1246	8
0.06	yes	24413	0	0	2066	21101	1246	8
0.15	no	55875	6	1983	1951	50696	1246	3
0.15	yes	55264	32	10573	1355	42090	1246	1
0.17	no	609969	65	27249	680	31822	1246	1
0.17	yes	60851	81	30659	414	28532	1246	1

The cost repartition is different from the Mozambique case where the cost of internal grids was up to 20 times the cost of intra-clusters connections. The creation of many more smaller communities in Lesotho case in fact, reduces the cost of interconnecting users within communities and moves the burden to intra-clusters interconnections: the two costs are in this case almost equal. This leads to a solution that is more biased towards off-grid systems that have a lower expenditure for grid creation. Once more, there is no solution that fits all, and an evaluation of the optimal size of communities from an economical standpoint could be helpful in providing the most promising scenario.

Subsequently, the genetic algorithm is run. It provides feasible solutions from an electrical standpoint, however it struggles in finding the real optimum with such a high number of possible connections.

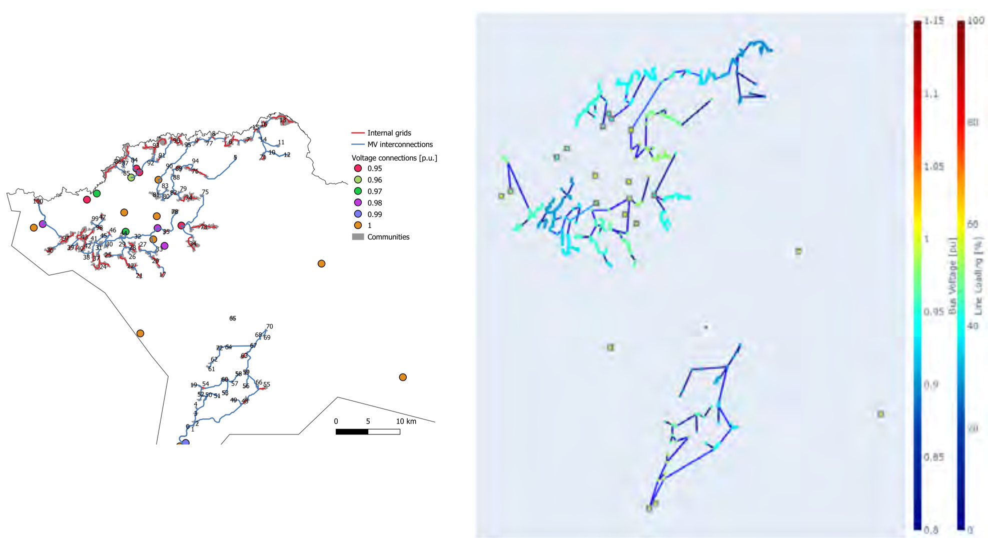


Figure 8.12: Results of integrated area optimization, genetic algorithm

Figure 8.12 shows the result of genetic algorithm run with a population of 40 individuals and a total of 2000 iterations; the computational time has been of 12 hours. Considering the COE from the national grid of 0.06, only community 65 is electrified with an off-grid system and the others are interconnected: this is a remote community, one of the farthest from the existing grid (15 km) and from other communities (5km), with few inhabitants (211 people) and small surface area (0.41 km^2). As in the Mozambique case, genetic algorithm is hence effective in identifying which communities to electrify with grid tied and off-grid systems, but it is hardly capable of designing the optimal grid topology. Intra-clusters connection are exemplified with straight lines in the second subfigure, to put in evidence also the line loading, that otherwise would not be visible and that is perfectly inside the limits. This is actually a common characteristic of rural electrification planning, where the real constraint is on voltage rather than on current limits.

As last step, multi-objective optimization was run in the area. The three complementary objective functions of minimization of costs, emissions and energy not supplied are the ones reported in section 6.4.3. The cost of energy bought from the grid was set at 0.06 \$/kWh and the hydropower resource was considered. Specific emissions associated to grid energy are 140 kgCO2/MWh and 416 kgCO2/MWh direct and indirect emissions respectively. Estimated yearly hours of unavailability are instead 174. Given the high amount of communities, failure rate related to lines' length was neglected and the total energy not supplied was computed as the sum of energy not provided by the grid and by the microgrids.

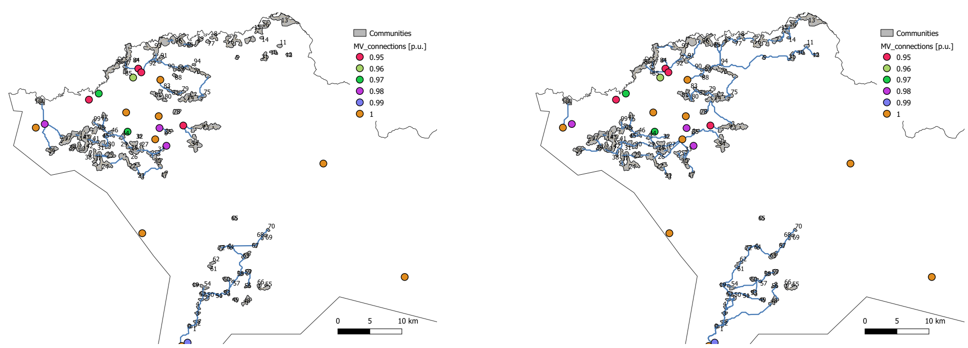
Table 8.6 and 8.13 show some of the most significant results of simulations run with different weights associated to the objective functions and with the inclusion of only direct emissions. When the weight of the cost function is lower than 0.5, all the clusters are electrified with microgrids, more convenient from an environmental and social perspective, since their reliability, according to the assumptions utilized, is higher.

When the weight of the cost is further increased the amount of microgrids suddenly decreases to below 10 given their higher net present cost. In the case where cost and social criteria share the same weight of 0.5, all communities are electrified with microgrids but the choice is on the microgrid portfolio with lowest cost (type 0). On the contrary, when the weight of cost decreases, the choice goes to microgrids with higher share of renewables (type 2).

The ternary plots of figure 8.14 show the value of the different objective functions when changing their weights: blue colour corresponds to lower values and reds to higher, less optimal values. The three plots above show results

Table 8.6: Multiobjective optimization results changing weights

w_{cost}	w_{env}	w_{soc}	f_{cost}	f_{env}	f_{soc}	f_{obj}	mg
0.5	0.5	0	25305	50530	716	0.77	$C2_2$,$C55_2$,$C65_2$,$C66_2$,$C69_2$
0.5	0	0.5	64629	114576	38	0.60	all
0.7	0.1	0.2	25322	52459	710	0.75	$C55_0$,$C65_0$,$C66_0$
0.7	0.2	0.1	24779	51317	724	0.8	$C65_0$
0.9	0	0.1	26017	54476	713	0.88	$C15_0$ $C65_0$
0.9	0.1	0	25251	51892	712	0.94	$C65_1$



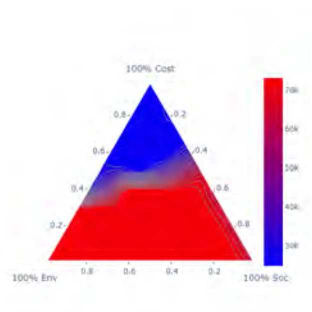
(a) Multi-objective results with $w_{cost} = 0.5, w_{env} = 0, w_{soc} = 0.5$

(b) Multi-objective results with $w_{cost} = 0.5, w_{env} = 0.5, w_{soc} = 0$

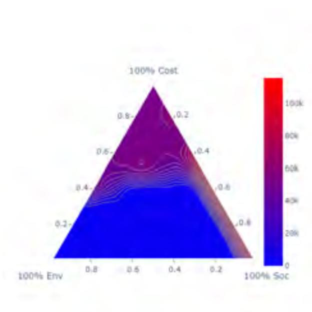
Figure 8.13: Butha-Buthe multi-objective optimization results

of simulations that consider direct emissions, while the three plots below correspond to simulations with indirect emissions. It can be clearly seen how the social and environmental objectives push toward similar solutions, that is the electrification with microgrids. As in Mozambique case, a smoother trend is observed in the direct emissions case given the preference, in the case with indirect emissions, towards solutions with more microgrids installed even at low environmental weights. Specific emissions are higher than in Mozambique given the higher energy consumption of the communities: they have a maximum value of 2443 kg CO2/capita in the case of direct emissions and 3246 kg CO2/capita for indirect emissions.

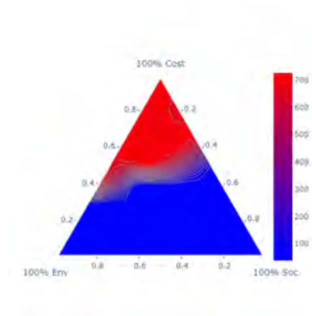
Table 8.7 summarizes the computational time required by each module of the procedure. Grid routing, when only model M2 is used is a fast procedure of only one minute, increasing to more than 2 hours when both M1 and M2 are tested. The iterative procedures for microgrid sizing and input data processing within Block4 are particularly time consuming giving the large



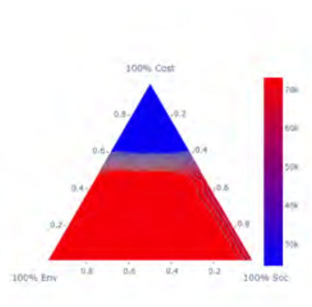
(a) NPC [k$] (dir. em.)



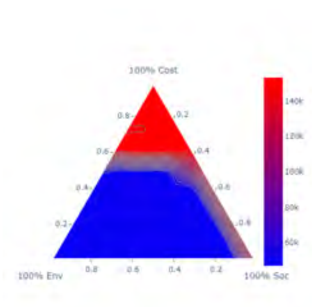
(b) CO2 [ton] (dir. em.)



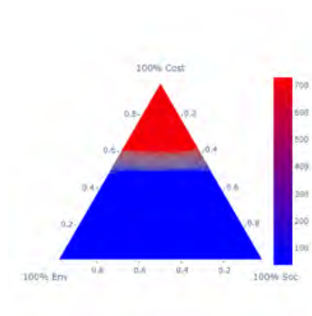
(c) Unavailability [MWh] (dir. em.)



(d) NPC [k$](indir. em.)



(e) CO2 [ton](indir. em.)



(f) Unavailability [MWh](indir. em.)

Figure 8.14: results of multi-objective optimization with different weights

Table 8.7: Computational time

Block	Module	Time
Block1	1.1 Identification of communities	34 seconds
Block1	1.2 Load estimation %	8 seconds
Block2	2.1 Energy resource assessment	partially manual
Block2	2.2 Microgrid sizing	1 hour, 10 minutes
Block3	3.1 Secondary Substations siting	1 minute
Block3	3.2 Cost surface creation	26 minutes
Block3	3.3 Grid routing	1 minute/2 hours
Block4	Input data processing	1 hour 46 minutes
Block4	4.1 Single-objective optimization	1 second
Block4	4.2 Multi-objective optimization	10 seconds

amount of communities. The procedure for load estimation is fast in this case because of the utilization of precomputed load profiles. The time for running RAMP procedure and creating standard profiles from MTF data is here not reported. The Gisele procedure takes globally from 3 hours and 30 minutes to 5 hours and 30 minutes according to the grid routing algorithm used.

Chapter 9

Discussion

The methodological framework presented in this work has the advantage of being comprehensive and able of managing several aspect of the rural electrification planning problem. On the other hand, its subdivision into four blocks, each of them entailing methodological and data based assumptions could drive to a sub-optimal solution and impact on the robustness of the approach. A complete sensitivity analysis on all the input data, cross-linking all the dimensions of the problem would be extremely time consuming and computational intensive and outside from the scope of this work. A critical discussion on the parameters and methodological assumptions is however helpful to increase the confidence of users and of future developers in the utilization of the model and to raise awareness on the points of attention. For this reason, the present chapter aims at discussing the impact that different assumptions can have on the obtained results and at proposing some preliminary sensitivity analyses on the parameters that are considered most critical. The numerical input parameters used in input to the four blocks of the procedure are qualitatively categorized according to their level of expected uncertainty into the following classes:

- Low uncertainty parameters: all the input data that are supposed to be gathered, with low difficulties by the user, belong to this class. The accuracy of those values depends mainly on the availability of data, but no further uncertainty related to non-verifiable assumptions is added. To this category belong all the data related to costs of microgrids components, technology characteristics, and the cost and electrical characteristics of power system components.
- Medium uncertainty parameters: all the geospatial data gathered from online databases as well as data related to renewable resources avail-

ability have an accuracy that is not directly controlled by the user. It is hardly imaginable that the user is able to verify all the information with on-field collected data, so in this case the uncertainty is exogenous. Luckily, geospatial databases are the object to continuous research and improvement, hence accuracy of these data is supposed to improve during time.

- High uncertainty parameters: in this category are inserted those parameters that are very difficult to be estimated with high accuracy. They could be collected by the user from different sources (e.g. on-field surveys), but hardly the user will get comprehensive information, so he should probably rely on experience and assumptions to estimate them

- Modelling assumptions: those are parameters that do not directly represent physical measurable quantities, but are needed to model and simplify the system to obtain significant results. Clustering parameters, the resolution of the grid of points, the number of typical days (exploited in the off-grid system sizing procedure) belong to this category. Given the absence of a straightforward right or wrong value for these data, some qualitative considerations may be needed to guide the user to their choice.

The four following sections report the detail of the parameters that, according to this categorization, belong to each of the classes. Although each input data would deserve an appropriate sensitivity analysis and quantification of its impact on the final result, this could drive to an infeasible approach due to the huge number of variables to be processed, consequently in this chapter the focus has been limited to the third and fourth categories of parameters which are the most critical, for which some preliminary numerical analysis are performed. The last section of the chapter proposes instead a discussion on the methodological assumptions at the basis of the blocks of the procedure, to highlight which type of impact they could have on the solution and which could be the possible developments to put in place to improve the accuracy of the results and their robustness. The focus will be given to future improvements of the work and possible research paths.

9.1 Low uncertainty parameters

In this section the parameters that, according to the author analysis are considered to have a low uncertainty, are listed. As previously said, the main reason for categorizing them in this way is that there could be a possibility that a fully-informed user is in possession of the correct (or at least reliable) value for these parameters, which are objective and retrievable from on-field sources and literature. In this case, uncertainty may be mostly related to future changes of technologies and costs hence a scenario analysis may be useful to test the optimal electrification strategy with different assumptions of future paths. Low uncertainties parameters are reported in table 9.1, classified in their main categories, with their description and an indication of the blocks of the procedure that could be influenced by their values.

Table 9.1: Low uncertainty parameters

Parameter	Description	Influence on
Parameters of table B.1	Techno-economic parameters for hybrid microgrid sizing	Block2, Block4
Maximum LV length	Maximum length of LV lines, used to site secondary transformers	Block3, Block4
Parameters of table 7.3	Techno economic parameters of the national grid	Block4
National grid emissions	Specific emissions related to energy production from the national energy mix	Block4 (multi obj)

All of the techno-economic parameters related to hybrid microgrid sizing, and required hence in the Module 2.2 are considered as low uncertainty parameters. Although it may not be straightforward to estimate the trajectory of new technologies in the future, in terms of performance and costs, if the investment is planned in a short time frame, those parameters could be easily derived by reports, companies quotations and already active projects. The costs of microgrid components influence the economic convenience of microgrids with respect to national grid connection, that is the final output of the whole proposed procedure. In addition, the differential cost of one component with respect to the other may impact the result of the MILP procedure for microgrid sizing and the optimal portfolio of the generation units. The project lifetime $\overline{Y}$, RES forecast errors, the maximum energy not supplied and the minimum energy produced by RES, are decisional parameters, for which a right or wrong solution does not exist, but their value depends on the type of scenario the user wants to simulate. The parameter number of typical days (N_d), also from table B.1, is instead categorized as a critical parameter since it is rather a modeling assumption with no physical basis than a scenario setting.

Maximum low voltage line length is a parameter that should be retrieved from local regulations or from DSOs best practices. Also in this case, a fully-informed user may have access to this information, or could at least provide a realistic estimation. This parameter could have a higher uncertainty with respect to the others since in some circumstances, the DSOs may use different types of proxies for designing low voltage grid. In the case of Mozambique, for instance, this data was not provided by grid operators in the country but was derived from international best practices.

The techno economic parameters related to medium voltage network lines are also parameters could also be retrieved by the user from different sources. The cost of feeders needs to be an average cost, considering the possible type of cables that could be deployed. Resistance and reactance also should be average values on the basis of what is typically deployed in the area. Average wholesale cost of energy is also usually available from electric utilities websites.

9.2 Medium uncertainty parameters

To this category belong those externally preprocessed data, downloaded from online geospatial databases, that provide geographically distributed information. The user, even though not having the possibility of validating all the data, could look for layers with high accuracy and spatial resolution, information which is often provided alongside the datasets.

Table 9.2: Medium uncertainty parameters

Parameter	Description	Influence on
Data of table 6.7	Input layers for SWAT procedure	Block2, Block4
Layer of roads	line vector layer with location of roads	Block3, Block4
Elevation layer	raster layer with elevation data	Block3, Block4
Landcover layer	raster layer with landuse categories	Block3, Block4
Inplace grid	vector layers with inplace distribution and transmission grids	All blocks
Population layer	raster layer with population density	all blocks
RES potential data	average profiles of wind and solar power potential	Block2, Block4

Among these parameters, the most impactful are the population and grid layers. Population data is the cornerstone of the whole procedure and a high error in its estimation, both in terms of geographical distribution and

absolute value could lead to meaningless results. The layer used in the case studies is the High Resolution Settlement Layer, from Columbia university, with a resolution of 30m that, according to the paper published in [212] has a building identification with an average precision and recall of 0.95 and 0.91, respectively. New population data are continuously being updated, so the accuracy in the population estimates is expected to increase in the future. In-place grid layers are instead not easily found online, and even when available the data provided is not very reliable and often incomplete, so in this case the best option for users is to gather this information from local stakeholders.
As for roads, landcover and elevation layers, the uncertainty in these data brings uncertainty to the evaluation of the cost surface and consequently to the optimal path of lines. The elevation layer used in the case studies, from the NASA SRTM, has a really good accuracy, with a high resolution of 30m and an error in the estimation of the absolute height lower than 6 meters for 90% of the data in Africa [213]. The landcover layer used in the case studies has a high spatial resolution (20m) but a lower accuracy as reported in the deliverable [214] tested for the countries of Gabon, Kenya, Ivory Coast and South Africa (where accuracy was varying from 40% to 90% of correct estimates). OpenStreetMap, mainly based on volunteering mapping, is used for the assessment of roads' position. In [215], the positional accuracy of the OSM road and building data is studied in the South African context. The results show that the accuracy is heterogeneous across the country. On average, the percentages of roads that are within 10 m of the Chief Directorate: National GeoSpatial Information (CD: NGI) roads are in the range from 65% to 94% for the nine state provinces with low urban density areas that tend to have a lower positional accuracy.

Among the resources, for wind and solar atlases the are well established databases tested in several works. As for the *renewableninjas* database, in [216], the author assessed the accuracy of the PV and Wind power time series in Norway. For what concerns PV, it is shown that yearly capacities are comparable with similar works for other regions, with an average yearly correlation of 0.844. Moreover, Mean Absolute Error (MAE) was below 8% while the maximum Root Mean Square Error (RMSE) was 14%. As for capacity factor, the maximum yearly difference was below 3%. Furthermore, no significant geographical differences accuracy were noticeable, even considering varying altitudes. For wind power, initial analyses showed noticeably low and varying correlation in wind production and the low resolution of the MERRA-2 data set. Authors developed and tested two different local wind speed adjustments in an attempt to improve accuracy, not achieved in all the studied locations. New wind models with high resolution (e.g. [217]) are under development which will allow to overcome such inaccuracy.

As future developments of the present thesis work, all this information could be analyzed together to perform an integrated geospatial uncertainty and sensitivity analysis assessment, as suggested in literature in the work of Crosetto [218].

9.3 High uncertainty parameters

This section reports the parameters for which, according to the author's experience, it is hard to retrieve reliable information. As extensively discussed in the previous chapters, load demand estimation in contexts where energy has not yet arrived, is particularly critical, both in terms of actual latent demand and future load growth. Also the costs of building lines in different environments, i.e. the weighting coefficients of table 6.11 could be not immediately retrievable from distributors and the user may need to include some assumptions.

Table 9.3: High uncertainty parameters

Parameter	Description	Influence on
Parameters of table 6.3	RAMP input parameters	Block1, Block2, Block4
Load growth	Load growth coefficient	Block1, Block2, Block4
Grid weights of table 6.11	Coefficients for the cost surface creation	Block3, Block4
Indirect emissions	LCOE emissions of the project	Block4 (multi-obj)
Reliability parameters	Parameters for determining the multi-objective social dimension	Block4 (multi-obj)

In this case, the suggested option would be the scenario-based analysis, changing the values of the parameters to assess their impact on the results. The ones that are expected to have a higher impact are the parameters related to the load estimation, which influence almost all the blocks of the procedure. High load demand and high load growth rate will lead to solutions biased towards grid interconnection rather then off-grid systems.

9.4 Modelling assumptions parameters

In this section are reported those input data required by the different algorithms (e.g. clustering, routing) used in the procedure. Those parameters are mathematical instruments not related to physical quantities and are used to create a treatable model. Those data, reported in table 9.4 are non verifiable assumptions, for which a right or wrong value does not exist, they rather influence the mathematical formulation and can drive the solution in different directions.

Table 9.4: Parameters related to modeling assumptions

Parameter	Description	Influence on
ϵ and $MinPts$	DBSCAN input parameters	all blocks
Grid Resolution	resolution of the grid of points composing the cost surface	Block3, Block4
N_d	Number of typical days	Block2, Block4
Input data preprocessing parameters in section 6.4.2	Parameters for choosing the number of possible connections	Block4

These data are critical, since there is no obvious optimal solution. The user should set them once some confidence with the model is acquired. The increase of grid resolution and of the number of typical days improve the model accuracy while increasing its computational time. Also the choice of the maximum number of connections per clusters may be done according to a compromise between time and accuracy. On the contrary, DBSCAN parameters are related to the size of communities to be electrified and the percentage of people to be supplied with stand alone systems, in this case the choice should mainly be based on strategic considerations.

9.5 Preliminary Sensitivity analysis

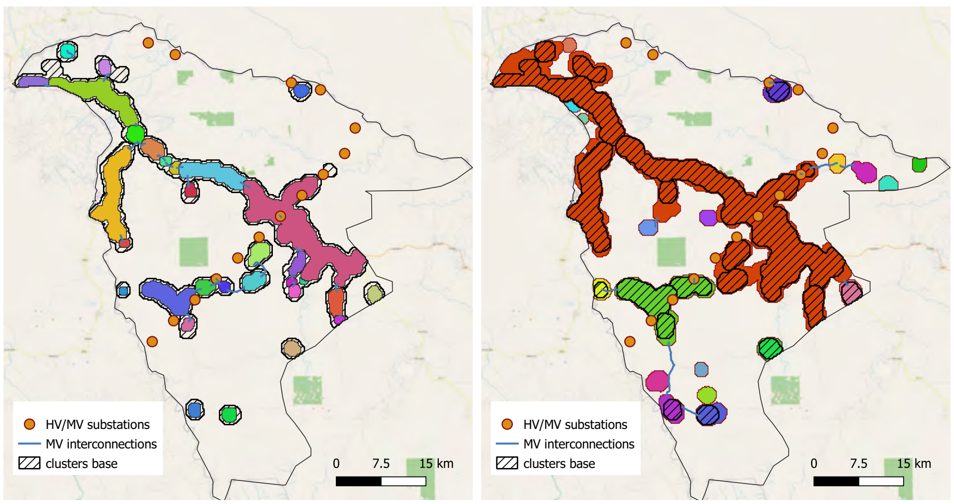
In this section the results of a preliminary sensitivity analysis on some of the most critical parameters (belonging to the last two categories previously listed and choosing the ones that affect a high number of Blocks) are reported. Sensitivity analysis has been run on the case study of Mozambique, selected as an effective case study and with a smaller computational cost (i.e. suited for a numerical sensitivity study devoted to clarify the proposed approach). Results are reported in table 9.5. The first parameter assessed has been the yearly load growth coefficient, that has been increased to 5% and 10%. In this case Block2 (module 2.2) and Block4 (module 4.1.2) of the procedure were

Table 9.5: Sensitivity analysis

Case	Load Growth %	ϵ m	$MinPts$	Resolution m	NPC k$	N microgrids
Base	2%	1000	200	200	2919	4
1	5 %	1000	200	200	3269	4
2	10 %	1000	200	200	3769	3
3	2%	800	200	200	2883	5
4	2%	1000	100	200	3288	10
5	2%	1000	200	600	2923	4

run again to evaluate the impact on the results. The total peak power of the communities increased in the first case of 23% and in the second case of 61%. In the Case 1 the number of communities electrified with off-grid systems does not change while in the Case 2 it decreases of one, becoming more convenient the interconnection to the grid for community number 8 (C8 as shown in the figure 7.3) when energy consumption is expected to grow significantly. This was in fact, among the communities that were not interconnected in the base case, the one closest to other clusters and to the national grid. The value of NPC increases in the two cases of 12 % and 30% respectively due to an increase in the cost for energy consumed, both from the one supplied from the grid and supplied from off-grid systems.

The second categories of parameters assessed are the ones related to the clustering procedure for the identification of communities. In this case all the blocks of the Gisele procedure are affected and were run again. The change in the values of ϵ and $MinPts$ (Case 3 and Case 4) has above all an impact on the number of communities: a 20% decrease in the value of ϵ leads to a number of communities increasing from 19 to 30, reason why the sensitivity has not been pushed further, not to increase too much the computational complexity, while a 100 % increase in the value of $MinPts$ adds just one community. As shown in the figure 9.1, clusters become smaller when decreasing ϵ and bigger when decreasing $MinPts$. The total percentage of electrified people goes from 85% in the base case to 82% in the case with decreased ϵ and 95% with decreased $MinPts$. It is interesting how this is reflected in the resulting optimal electrification solution: in the first case, the number of off-grid systems identified almost corresponds to the one of the initial case, with just one additional off-grid system, to supply a community that is smaller than the ones identified in the base case (C6) and hence more conveniently electrified with off-grid systems. In the case with decreased $MinPts$ the number of off-grid systems increases significantly up to 10. Those new microgrids are mainly designed in areas that in the base case were not even considered for electrification, given their sparse population density. This validates the approach and gives the interesting indication of the importance of establishing a threshold for the number of people to be electrified in the area.



(a) Case 3: $\epsilon = 800$ m (b) Case 4: $MinPts = 1000$

Figure 9.1: Communities identification changing values of ϵ and $MinPts$

Finally, the level of resolution of the regular grid of points for the cost surface creation ($GridPts$) has been changed and Block 3 and Block 4 were run again. This parameter is the one that less affected the results: this has been increased of three times and no significant change in the solution was perceived, with the same number of communities electrified with off-grid systems as in the base case and an increase in the total NPC of just the 0.1%. Differently from the others, this parameter actually affects just Block3 and Block4 so it has a lower impact on the overall procedure while helping to reduce the computational burden. In this case, a qualitative analysis could be performed by the user to address the regularity of the territory: in the case of montane area with steep valleys and peaks, a good resolution may help capturing the characteristics of the terrain; on the contrary, in a rather flat area like the Mozambique case, resolution could be worsened to reduce the computational effort without affecting the results.

9.6 Methodological assumptions

This final section reports, for each of the blocks, which are the strongest methodological assumptions in the algorithms identification and usage and which could be the possible paths for improvement.

Block1: energy demand assessment

- Population does not expand geographically: all the analysis considers a static distribution of the population, with load increasing homogeneously only where it is present at year 0. The shape of the communities and the grid routing procedure are based on the current distribution of population. In reality, communities tend to enlarge during time, as thoroughly explained in the book [32]. A multi-year planning that considers possible trends in the future population distribution (e.g. higher probability of expansion of communities along roads) could be an interesting improvement for the tool.

- Communities are identified through density based clustering: after the identification of communities as densely inhabited clusters, no check is performed to assess if those are also the optimal sizes for creating homogeneous energy communities. A techno-economic analysis to iterate and improve communities boundaries could be useful (eg. two smaller LV microgrids may be preferable to a big MV system).

- The load growth model is a constant percentage annual increase, homogeneous for all the clusters. Actually more urbanized areas may have faster trends of growth with respect to others, and this trend may decrease during time. Constructing a database that could allow to categorize the communities and associate them the most suitable load growth trend could a demanding yet very useful improvement.

- Uniform type of demand in all clusters: all the clusters are supposed to have the same types of users and the profiles are scaled according to the proxy indicators of population and area of communities. As said before, an analysis able to categorize communities and associate them different types of users' composition and appliance usage trends may be useful to improve this step. In the research group of Politecnico, some researchers and MsC students are studying exactly this topic and some possible solutions.

Block2: off-grid system sizing

- Hydro resource is "inexhaustible": there is no limitation on the number of communities that could install mini hydro turbines on the same river. In reality, it would not be feasible to build several small plants close to each other, so there may the need to improved the model by aggregating communities that could be supplied by the same plant.
- Typical days are computed by averaging the values of several days: other approaches, such as using clustering techniques for their identification may be more effective in detecting the hourly variability of the resources without using important patterns.

Block3: internal grids design

- Secondary Substations siting is based on a maximum length for low voltage lines, while no power constraints are included. This is reasonable in rural areas with scattered population but may be more critical in urban areas. Adding some parameters related to the size of transformer and their associated costs may be useful for a better evaluation of their optimal location.
- No electrical constraints are considered: this is not problematic when the sizes of communities are smalls and violations are not expected (e.g. Lesotho case), however sometimes electrical issues may arise when performing the final optimization. In this case now the option would be to ex-post split the bigger clusters into two and perform again simulations.

Block4: Integrated area optimization

- Communities are approximated with one single node: this approximation does not allow to evaluate voltage drops and losses within communities, hence losing accuracy, however it increases the treatability of the problem. New studies in the research group have focused on overcoming this limitation and they will be discussed and published in new works.
- Only one type of line is considered: all the grid is designed as a three-phase medium voltage line of one specific type with given electrical characteristics. The inclusion of multiple sizes of lines in the design process resulted, in preliminary analyses performed, very time consuming and caused several convergence issues. For this reason, the optimal design of all the line branches is left to the user in a post-processing phase.

- Power losses are neglected to keep the problem linear. Even though usually loads are small and so the value of losses is not really impactful (they were estimated as 5% of the load demand in the cases studied), it may be interesting to analyze literature works related to grid planning where they have been linearized and evaluate their possible inclusion in the model.
- Hybrid microgrid electrification and grid extension are complementary electrification solutions: it is assumed that microgrids should be deployed where grid does not arrive and viceversa. Actually, it may be interesting to investigate also cases where microgrids could support the grid for improving the total reliability and reducing hence the energy not supplied. This may be relevant in case in the community are present non detachable loads such as hospitals.

Chapter 10

Conclusions

The research work presented in this manuscript has the general goal of addressing the problem of energy access in rural areas of the Global South, proposing effective solutions that could foster the electrification process considering both on and off-grid technologies. More specifically, the research has three main objectives, listed in the introductory part, to be achieved through the answer to correspondent six research questions. In the following paragraphs, following the same subdivision, the work performed is summarized and the main research outcomes are provided in terms of answer to the research questions. Answering the last question, considerations about limitations and possible future paths are provided.

Objective 1*: analyse the literature related to electricity access to understand which are the available tools and solutions for electrification.*
The objective was achieved through the analysis of more than 200 peer reviewed journal articles and books on the topic of rural electrification and energy access, as well as documents and data from international agencies. The analysis has been divided into three parts, the first one being devoted to the analysis of the problem of access to electricity throughout the world, to put in evidence virtuous cases and adopted solutions. Sub-Saharan Africa results to be the most critical region, with still less than 50% of the population without access to electricity in 2019. The second part focused on the electrification planning problem, to understand which are the different aspects that should be taken into consideration to design an effective electrification solution. Thirdly, an analysis of the state of the art of tools and procedures available in literature to help stakeholders in different phases of rural electrification planning projects has been performed.

Research questions:

1. Which are the possible strategies for electrification? What influences their adoption?

2. Which type of instruments would help stakeholders performing rural electrification planning?

3. Which are the literature gaps in models for rural electrification planning?

Research answers:

1. The three main strategies for electrification are the interconnection to the national grid, the installation of microgrids and the provision of small SA. In order to identify the optimal solution and promote the most effective electrification plan, several aspects must be taken into considerations. They can be categorized into 5 main classes: techno-economic, social, geographical, environmental and regulatory. An effective electrification plan should hence be able to consider all the possible technological options properly evaluating their costs and technical characteristics, the geographic context with the resources availability, the distribution of population as well as its energy needs and be able to propose solutions that have low environmental impact and that are feasible from a regulatory standpoint.

2. According to the aspect of rural electrification planning that need to be addressed, different type of software tools could be adopted. They have been categorized into energy models, electric models and comprehensive models. The first ones are helpful for identifying the optimal portfolio of generation sources, the second ones are used to design the grid topology while the third category, which is the most complete, performs both the design of the new distribution grids and of the new generation sources to be deployed.

3. Literature is still short of models for comprehensive rural electrification planning, able to design both on and off-grid solutions. Among the ones available some shortcomings have been highlighted: (i) simplified electric network design; (ii) consideration of limited options for microgrid sizing; (iii) non consideration of multi-objective optimization with environmental aside cost dimension (iv) most of them are proprietary softwares, not publicly available.

Objective 2: *create a new open-source and open-access modeling framework, usable by different stakeholders.*
A new procedure has been developed, coded in Python and validated on four different real life study-cases. The procedure is composed by four main blocks, devoted to solve different aspects of the rural electrification planning process. Given a non-electrified rural area, the procedure assesses the energy needs of the different communities, sizes hybrid microgrids that could supply the loads according to the availability of RES, designs the distribution grid and identifies which communities to electrify with off-grid systems and which to interconnect to the in place grid.

Research questions:

1. How could open-source geospatial data be used for electrification planning?
2. How to identify the optimal electrification solution choosing among different technological options?
3. Can optimization consider also non economical aspects, such as technical and environmental dimension?

Research answers:

1. Open source geospatial data can provide almost all the necessary information for performing pre-feasibility studies for rural electrification planning. In particular, in the proposed procedure, raster data related to population are the input to identify, through clustering approaches densely inhabited rural communities and define the optimal siting of secondary transformers. Raster data related to elevation, land cover, soil type, irradiance and wind speed availability are used to assess the RES potential availability and define low cost corridors for deploying electric lines. Finally, vector layers with information on road networks, protected areas, electric networks location and other infrastructures are used to complement the analysis and have an accurate representation of the territory. The availability of geospatial data is improving over time, new layers with better accuracy are being developed and released on a monthly or weekly basis, so the developed procedure will have the opportunity keep the pace and continuously grow and improve the level of detail of the analyses performed.
2. In order to identify the optimal electrification solution, several steps are performed. First of all, communities are identified and outliers, i.e.

sparse and remote households are excluded from the analysis and assumed to be electrified with SA. For the selected communities, their energy needs are assessed and the minimum cost for electrifying them with hybrid microgrids, able to exploit all the available resources is computed. These preliminary steps are necessary to arrive at the final integrated optimization where the cost of building electric lines to interconnect communities among each other and with the inplace national grid is compared to the cost of building microgrids. The solution is not on a community by community basis but it is identified for the whole interest area, since the optimal solution for one community could influence the others.

3. A multi-objective optimization model has been developed to address this question. The model includes the environmental and social dimension aside the economic in terms of minimization of CO2 emissions and of energy not supplied. The approach is flexible enough to include further dimensions and constraints.

Objective 3: *test the procedure on real life case studies and assess its performance and limitations.*
During the years of doctoral studies, several collaborations with companies, NGOs, and researchers have been activated with the goal of gathering useful data and validate different aspects of the procedure. Among the different works performed related to analyses in 9 areas in Sub-Saharan Africa and South America, four of them have been deepened and used to validate parts of the procedure. Out of these four, the rural area of Namanjavira in Mozambique and Butha-Buthe in Lesotho, have been selected to test and validate the full procedure for rural electrification planning.

Research questions:

1. How does the proposed procedure perform in two different contexts?
2. How are solutions influenced by different models' assumptions?
3. Which are the possible areas of improvement?

Research answers:

1. The procedure developed in the PhD project performed well in both the two cases analysed and was able to design the electrification plan taking into account on and off-grid technologies. The two cases, although similar in terms of the extension of the non-electrified areas, differed for

other aspects such as the distribution of the population, the morphology of the territory, the energy needs, the voltage level of the grid and the data availability. In Mozambique, where the population is distributed along roads without clear distinction of communities' boundaries, the main criticality resided in the identification of the optimal size of communities. The choice of a small number of large communities lead to strong approximations in the voltage drops calculations. In Lesotho, with a large number of small communities, the critical aspect has been the computational effort in the integrated area optimization.

2. One of the strongest assumptions that later influences all the analysis is the size of the communities. Selecting large communities that cover most of the area, in fact, increases the cost of interconnecting consumers within communities and decreases the cost of interconnecting communities with the in place grid: in this case micrograds are hence penalized. Also, the consideration of hydropower resource, given its economic convenience, was demonstrated to move the equilibrium towards solutions oriented towards hybrid microgrids rather than grid interconnection. Finally, assumptions on communities' energy needs could influence the final results: as shown in the case studies, the communities electrified with off-grid systems are the smallest communities, far from the in place grid and from other communities. Overestimating or underestimating their real energy needs could make grid connections respectively more or less convenient.

3. According to the research performed and the results obtained, some areas of improvement can be identified and suggested:

    - Improvement of the energy needs assessment: the process of data gathering for automatizing and generalizing load demand assessment starting from geospatial data is now at the beginning. Further research could be needed to collect further data and be able to automatize the communities' load profile creation removing the need of on-field data collection.
    - Improvement in the definition of communities' size: develop an iterative clustering process to select the size that could minimize the microgrids cost, considering both LV and MV microgrids;
    - Inclusion of the internal grids modeling in the integrated MILP optimization: this will lead to a significant reduction of voltage drops related errors, but in depth analyses will need to be performed to avoid computational issues;

- Improvement of the multi-objective optimization to include further dimensions and improve the estimation of microgrids and grid reliability;
- Inclusion of the biomass resource as possible source of energy in the hybrid microgrid optimization process.

Most of these points are already being addressed within the research group by MSc and PhD students. Moreover, since the procedure is open source and available online, the wish is that new cooperations will arise and *Gisele* will continue to grow and improve during time.

Appendix A

Input parameters for load profiles

A.1 Generic-MTF

Table A.1: *RAMP* configuration parameters for generic user categories

Appliance_{ij}	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
School-urban							
External Lights	4	25	17-06	0	1	12	0
Internal Lights	18	20	7-17	0	0.5	4	0
PC	13	50	7-17	0	0.5	4	0
TV	3	60	7-17	0	0.5	2	0
School-rural							
External Lights	2	25	17-06	0	1	12	0
Internal Lights	4	20	7-17	0	0.5	4	0
Worship-urban							
External Lights	5	25	17-06	0	1	12	0
Internal Lights	20	25	18-22	0	1	4	0
TV	1	100	16-21	0	1	4	0.2
PC	3	50	16-21	0	0.5	3	0.2
Worship-rural							
External Lights	2	25	17-06	0	1	12	0
Internal Lights	4	25	18-22	0	1	4	0
Health-urban							
Internal Lights	36	20	8-12; 14-24	20	3	12	20

Table A.1: *RAMP* configuration parameters for generic user categories

$Appliance_{ij}$	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
External Lights	15	25	16-24	20	13	13	20
Phone charger	10	5	0-24	20	0.5	5	20
Sterilizer	2	1500	6 22	20	0.5	1	20
TV	3	60	7 17	20	0.5	2	20
PC	10	50	8-12; 17-24	20	0.1	5	20
Fridge	4	250	0-24	20	0.5		20
Fridge2	2	500	0-24	20	0.5		20
Microscope	3	200	7 17	20	0.5	3	20
Centrifuge	3	200	7 17	20	0.5	5	20
Monitor	3	50	7 17	20	0.5	5	20
Health-rural							
Internal Lights	7	20	8-12; 14-24	20	3	12	20
External Lights	3	25	16-24	20	13	13	20
Phone charger	10	5	0-24	20	0.5	5	20
Sterilizer	2	1500	6 22	20	0.5	1	20
Fridge	2	250	0-24	20	0.5		20
Households Tier1							
Internal Lights	3	5	18-03	0.2	0.5	5	0.2
Phone charger	2	5	0-10;13-16;18-24	0.2	0.5	3	0.2
Radio	1	5	6-10;17-24	0.2	0.5	5	0.2
Households Tier2							
Internal Lights	4	5	18-03	20	0.5	5	20
Phone charger	2	5	0-10;13-16;18-24	20	0.5	3	20
Radio	1	5	6-10;17-24	20	0.5	5	20
External Lights	1	7	18-7	20	1	12	20
TV	1	90	8-15; 17-24	20	0.1	4	20
PC	1	60	8 24	20	0.1	4	20
Fan	1	60	4 24	20	0.1	6	20
Households Tier3							
Internal Lights	8	5	18-03	20	0.5	5	20
Phone charger	4	5	0-10;13-16;18-24	20	0.5	3	20
Radio	1	5	6-10;17-24	20	0.5	5	20
External Lights	1	7	18-7	20	1	12	20
TV	1	90	8-15; 17-24	20	0.15	4	20
PC	2	60	8 24	20	0.15	4	20

Table A.1: *RAMP* configuration parameters for generic user categories

Appliance$_{ij}$	n$_{ij}$	P$_{ij}$ [W]	fw$_{ij}$ [h]	Rfw$_{ij}$ %	fc$_{ij}$ [h]	ft$_{ij}$ [h]	Rft$_{ij}$ %
Fan	2	60	4 24	20	0.15	6	20
Fridge	1	200	0-24	20	0.5		20
Food processor	2	350	18-20	20	0.15	0.5	20
Water pump	2	60	12 01	20	0.15	3	20
Rice cooker	2	60	12-15; 20-01	20	0.15	1	20
Households Tier4							
Internal Lights	16	7	18-03	20	0.5	5	20
Phone charger	4	5	0-10;13-16;18-24	20	0.5	3	20
Radio	1	5	6-10;17-24	20	0.5	5	20
External Lights	2	7	18-7	20	1	12	20
TV	1	150	8-15; 17-24	20	0.15	4	20
PC	2	60	8 24	20	0.15	4	20
Fan	2	60	4 24	20	0.15	6	20
Fridge	2	300	0-24	20	0.5		20
Food processor	2	350	18-20	20	0.5	1	20
Iron	1	1000	6 20	20	0.5	1	20
Ari dryer	1	1000	17-24	20	0.15	0.5	20
Toaster	1	1000	6-9; 18-21	20	0.15	0.5	20
Microwave	1	700	6-9; 11-14;18-21	20	0.15	1	20

A.2 Mozambique-Namanjavira

Table A.2: *RAMP* configuration parameters for Namanjavira user categories

Appliance$_{ij}$	n$_{ij}$	P$_{ij}$ [W]	fw$_{ij}$ [h]	Rfw$_{ij}$ %	fc$_{ij}$ [h]	ft$_{ij}$ [h]	Rft$_{ij}$ %
Primary School							
External Lights	7	25	16-21	0	5	5	0
Internal Lights	15	20	16-21	0	5	5	0
Fridge	1	60	8-18	0	10	10	0
Secondary School							
External Lights	5	25	16-21	0	5	5	0

Table A.2: *RAMP* configuration parameters for Namanjavira user categories

Appliance_{ij}	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
Internal Lights	25	20	16-21	0	5	5	0
Worship							
External Lights	3	25	16-21	0	7	7	0
Internal Lights	8	20	16-21	0	7	7	0
Sound System	1	20	16-21	0	2	6	30
Phone Charger	1	5	16-21	0	1	3	90
Administrative Headquarters							
External Lights	2	25	16-18	0	2	2	0
Internal Lights	6	20	16-18	0	2	2	0
Computer	2	50	8-18	0	2	6	20
Phone Charger	4	5	8-18	0	1	3	90
Police Station							
External Lights	1	20	16-5	0	12	12	0
Internal Lights	4	15	16-5	0	4	6	0
Phone Charger	1	5	18-5	0	1	3	30
Office							
External Lights	5	20	16-18	0	1	1	30
Internal Lights	9	15	16-18	0	1	1	30
Fridge	2	70	8-18	0	1	3	0
Phone Charger	4	5	8-18	0	1	5	20
Electronics	5	100	8-18	0	1	5	20
Public Lights							
Street Lights	30	30	16-5	0	13	13	0
Community Lights	30	25	16-22	0	7	7	0
Merchants							
External Lights	2	25	16-22	0	6	6	0
Internal Lights	3	20	16-22	10	6	6	0
Freezer	1	300	10-18	0	8	8	0
Sound System	1	20	10-22	0	2	6	30
Phone Charger	1	5	10-22	0	1	3	90
Fan	1	80	10-22	0	3	6	10
Tailors							
External Lights	1	25	16-19	0	3	3	0
Internal Lights	3	20	16-19	0	4	6	10
Phone Charger	1	5	10-19	0	1	3	90

Table A.2: *RAMP* configuration parameters for Namanjavira user categories

$Appliance_{ij}$	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
Radio	1	5	10-19	0	3	6	25
Fan	1	60	10-19	0	3	6	30
Barbers							
External Lights	1	20	16-19	0	7	7	0
Internal Lights	3	15	16-19	0	4	6	10
Radio	1	5	10-19	0	1	5	25
Phone Charger	1	5	10-19	0	1	3	90
Households Tier1							
External Lights	1	15	16-21	0	7	7	0
Internal Lights	3	15	16-21	0	4	6	10
Radio	1	5	10-21	0	1	5	25
Phone Charger	1	5	10-21	0	1	3	90
Fan	1	25	10-21	0	1	4	30
Households Tier2							
External Lights	1	20	16-21	0	5	5	0
Internal Lights	3	15	16-21	0	3	4	20
TV	1	100	10-21	0	2	7	20
Fridge	1	70	10-21	0	11	11	0
Radio	1	5	10-21	0	1	5	25
Phone Charger	1	5	10-21	0	1	3	90

A.3 Lesotho-Butha-Buthe

Table A.3: *RAMP* configuration parameters for Butha-Buthe user categories

$Appliance_{ij}$	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
School							
External Lights	4	25	17-06	0	5	5	0
Internal Lights	18	20	7-17	0	5	5	0
PC	13	50	7-17	0	10	10	0
TV	3	60	7-17	0	10	10	0

Table A.3: *RAMP* configuration parameters for Butha-Buthe user categories

Appliance_{ij}	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
Worship							
External Lights	3	25	16-21	0	7	7	0
Internal Lights	8	20	16-21	0	7	7	0
TV	1	20	16-21	0	2	6	30
PC	1	5	16-21	0	1	3	90
Merchants							
External Lights	2	25	16-22	0	6	6	0
Internal Lights	3	20	16-22	10	6	6	0
Freezer	1	300	10-18	0	8	8	0
Sound System	1	20	10-22	0	2	6	30
Phone Charger	1	5	10-22	0	1	3	90
Fan	1	80	10-22	0	3	6	10
Tailors							
External Lights	1	25	16-19	0	3	3	0
Internal Lights	3	20	16-19	0	4	6	10
Phone Charger	1	5	10-19	0	1	3	90
Radio	1	5	10-19	0	3	6	25
Fan	1	60	10-19	0	3	6	30
Barbers							
External Lights	1	20	16-19	0	7	7	0
Internal Lights	3	15	16-19	0	4	6	10
Radio	1	5	10-19	0	1	5	25
Phone Charger	1	5	10-19	0	1	3	90
Households Tier3							
Internal Lights	8	5	18-03	20	0.5	5	20
Phone charger	4	5	0-10;13-16;18-24	20	0.5	3	20
Radio	1	5	6-10;17-24	20	0.5	5	20
External Lights	1	7	18-7	20	1	12	20
TV	1	90	8-15; 17-24	20	0.15	4	20
PC	2	60	8 24	20	0.15	4	20
Fan	2	60	4 24	20	0.15	6	20
Fridge	1	200	0-24	20	0.5		20
Food processor	2	350	18-20	20	0.15	0.5	20
Water pump	2	60	12 01	20	0.15	3	20
Rice cooker	2	60	12-15; 20-01	20	0.15	1	20

Table A.3: *RAMP* configuration parameters for Butha-Buthe user categories

Appliance_{ij}	n_{ij}	P_{ij} [W]	fw_{ij} [h]	Rfw_{ij} %	fc_{ij} [h]	ft_{ij} [h]	Rft_{ij} %
Households Tier4							
Internal Lights	16	7	18-03	20	0.5	5	20
Phone charger	4	5	0-10;13-16;18-24	20	0.5	3	20
Radio	1	5	6-10;17-24	20	0.5	5	20
External Lights	2	7	18-7	20	1	12	20
TV	1	150	8-15; 17-24	20	0.15	4	20
PC	2	60	8 24	20	0.15	4	20
Fan	2	60	4 24	20	0.15	6	20
Fridge	2	300	0-24	20	0.5		20
Food processor	2	350	18-20	20	0.5	1	20
Iron	1	1000	6 20	20	0.5	1	20
Ari dryer	1	1000	17-24	20	0.15	0.5	20
Toaster	1	1000	6-9; 18-21	20	0.15	0.5	20
Microwave	1	700	6-9; 11-14;18-21	20	0.15	1	20

Appendix B

Input parameters for microgrid sizing

Table B.1: Parameters of hybrid microgrid sizing model

Symbol	Parameter	Value	U.M.
F	Cost of fuel	0.75	[$/l]
$\overline{Y}$	project lifetime	10	[years]
N_d	number of typical days	12	[day/year]
$\overline{ENS}$	maximum energy not supplied	5	[%]
$\underline{RES}$	minimum energy produced by RES	0	[%]
γ_d	load forecast error	0.1	[0-1]
$\gamma_p v$	PV forecast error	0.1	[0-1]
$\gamma_w t$	WT forecast error	0.25	[0-1]
ir	Interest Rate	0.06	[0-1]
PV modules			
C_p	Unitary capacity	1	[kW]
CC_p	Capital cost	1400	[$/unit]
M_p	O&M yearly cost	10	[$/unit/y]
Y_p^{life}	Lifetime	20	[years]
Wind turbine			
C_w	Unitary capacity	10	[kW]
CC_w	Capital cost	27000	[$/unit]
M_w	O&M yearly cost	540	[$/unit/y]
Y_w^{life}	Lifetime	20	[years]
Diesel generator			

Table B.1: Parameters of hybrid microgrid sizing model

Symbol	Parameter	Value	U.M.
C_g	Unitary capacity	16	[kW]
CC_g	Capital cost	11000	[$/unit]
M_g	O&M hourly cost	0.2	[$/unit/h]
H_g^{life}	Lifetime	15000	[hours]
A	cost coefficient	0.4672	[l/h]
B	cost coefficient	0.3	[l/h/kW]
$\underline{P_g}$	Min Power of DG	0.3	[0-1]
	BESS		
C_b	Unitary capacity	1	[kWh]
CC_b	Capital cost	400	[$/unit]
M_b	O&M yearly cost	10	[$/unit/y]
H_b^{life}	Lifetime	3000	[kWh]
Y_b^{life}	Lifetime	15	[years]
η_b	efficiency	0.975	[-]
$\overline{PQ_b}$	maximum BESS power-to-energy ratio	1	[kW/kwh]
$\overline{DOD_b}$	maximum BESS depth of discharge	0.9	[0-1]

Table B.2: Parameters of hydro turbines

Turbine	**Capacity** [kW]	**Capital Cost** [k$]	**O&M Cost** [k$/y]	**Lifetime** [years]	**P min** [kW]	**Efficiency** -
1	1	10	0.2	25	0.05	0.8
2	10	80	1.6	25	0.5	0.8
3	100	400	8	25	5	0.8
4	1000	2000	40	25	50	0.8
5	10000	15000	300	25	500	0.8

Bibliography

[1] Benjamin Pillot, Marc Muselli, Philippe Poggi, and João Batista Dias. On the impact of the global energy policy framework on the development and sustainability of renewable power systems in sub-saharan africa: the case of solar pv. *arXiv preprint arXiv:1704.01480*, 2017.

[2] Our World in Data. Energy intensity vs. gdp per capita, 2016. *Our World in Data*, 2020. https://ourworldindata.org/grapher/energy-use-per-capita-vs-gdp-per-capita.

[3] United Nations Development Programme. Sustainable Development Goals- Goal 7: Affordable and Clean Energy. http://www.undp.org, 2020.

[4] IEA. Defining energy access: 2020 methodology, 2020.

[5] IEA, IRENA, UNSD, World Bank, and WHO. Tracking SDG 7: The Energy Progress Report. 2021.

[6] International Energy Agency. World Energy Outlook 2020 - Event - IEA. *World Energy Outlook 2020 - Event - IEA*, pages 1–25, 2020.

[7] IEA. SDG7: Data and Projections, 2020.

[8] Julia Tomei, Jennifer Cronin, Héctor David Agudelo Arias, Samir Córdoba Machado, Maycol Francisco Mena Palacios, Yenny Marcela Toro Ortiz, Yemilson Espidio Borja Cuesta, Reiner Palomino Lemus, William Murillo López, and Gabrial Anandarajah. Forgotten spaces: How reliability, affordability and engagement shape the outcomes of last-mile electrification in Chocó, Colombia. *Energy Research and Social Science*, 59(March 2019), 2020.

[9] International Energy Agency. Africa Energy Outlook 2019 Africa Energy Outlook 2019. 2019.

[10] The World Bank and ESMAP. *REGULATORY INDICATORS FOR SUSTAINABLE ENERGY: SUSTAINING THE MOMENTUM.* 2020.

[11] Ministry of Water irrigation Federal Democratic Republic of Ethiopia and Electricity. National Electrification Program 2 .0; Integrated Planning for Universal Access. 2019.

[12] Republic of Rwanda Ministry of infrastructure. Energy Sector Stategic Plan. Technical Report March 2015, 2018.

[13] Republic of Rwanda Ministry of infrastructure. Rural Electrification Strategy. (September):1–23, 2016.

[14] ESMAP and WBG. Senegal's SE4ALL. Rural Electrification. Action Agenda and Investment Prospectus 2018. 2018.

[15] PYRAMIDE ENVIRONNEMENTAL CONSULTANTS. Projet d'Electrification Rurale par Mini-réseaux solaires de 1000 villages répartis dans sept (07) régions du Sénégal Rapport d'Evaluation Environnementale Stratégique et Sociale. 2020.

[16] Kenneth Lee, Eric Brewer, Carson Christiano, Francis Meyo, Edward Miguel, Matthew Podolsky, Javier Rosa, and Catherine Wolfram. Barriers to electrification for" under grid" households in rural kenya. Technical report, National Bureau of Economic Research, 2014.

[17] Kenya Ministry of Energy. Kenya National Electrification Strategy: Key Highlights 2018. page 36, 2018.

[18] ESMAP. Mini Grids for Half a Billion People: Market Outlook and Handbook for Decision Makers. Executive Summary. Energy Sector Management Assistance Program (ESMAP) Technical Report 014/19. 2019.

[19] Emanuela Colombo, Stefano Bologna, and Diego Masera. *Renewable energy for unleashing sustainable development.* 2013.

[20] Alberto Berizzi, Maurizio Delfanti, Davide Falabretti, Stefano Mandelli, and Marco Merlo. Electrification Processes in Developing Countries: Grid Expansion, Microgrids, and Regulatory Framework. *Proceedings of the IEEE*, 107(9):1981–1994, 2019.

[21] Chris Marnay, Spyros Chatzivasileiadis, Chad Abbey, Reza Iravani, Geza Joos, Pio Lombardi, Pierluigi Mancarella, and Jan Von Appen.

Microgrid evolution roadmap. *Proceedings - 2015 International Symposium on Smart Electric Distribution Systems and Technologies, EDST 2015*, pages 139–144, 2015.

[22] EU Energy Initiative Partnership Dialogue Facility. Mini-grid Policy toolkit. 2014.

[23] Magda Moner-Girona, Daniel Puig, Yacob Mulugetta, Ioannis Kougias, Jafaru AbdulRahman, and Sándor Szabó. Next generation interactive tool as a backbone for universal access to electricity. *Wiley Interdisciplinary Reviews: Energy and Environment*, 7(6), 2018.

[24] Humberto Calil Bertollo. Contribuições ao estudo dos aterramentos de sistemas monofilares com retorno pelo terra. 2008.

[25] Francesco Iliceto. *Rural Electrification with the Shield Wire Scheme in Low-Income Countries: Design, Construction, and Operation*. World Bank, 2016.

[26] International Renewable Energy Agency. *Renewable Power Generation Costs in 2019*. 2020.

[27] Parimalram Achintya Madduri, Jason Poon, Javier Rosa, Matthew Podolsky, Eric A Brewer, and Seth R Sanders. Scalable dc microgrids for rural electrification in emerging regions. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 4(4):1195–1205, 2016.

[28] Pascal Hategekimana, Adria Junyent Ferre, Etienne Ntagwirumugara, and Joan Marc Rodriguez Bernuz. Assessment of feasible dc microgrid network topologies for rural electrification in rwanda: Studying the kagoma village. In *2020 International Conference on Smart Grids and Energy Systems (SGES)*, pages 854–859. IEEE, 2020.

[29] Hossein Lotfi and Amin Khodaei. Ac versus dc microgrid planning. *IEEE Transactions on Smart Grid*, 8(1):296–304, 2015.

[30] M. Juanpera, P. Blechinger, L. Ferrer-Martí, M. M. Hoffmann, and R. Pastor. Multicriteria-based methodology for the design of rural electrification systems. A case study in Nigeria. *Renewable and Sustainable Energy Reviews*, 133(April):110243, 2020.

[31] Fabio Riva, Annalisa Tognollo, Francesco Gardumi, and Emanuela Colombo. Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies

and insights from a modelling perspective. *Energy Strategy Reviews*, 20(February):71–89, 2018.

[32] H.L. Willis. *Spatial Electric Load Forecasting [Book Review]*, volume 10. 2005.

[33] World Bank Group. The World Bank. http://www.worldbank.org.

[34] Fabio Riva, Francesco Davide Sanvito, Francesco Tonini Tonini, Emanuela Colombo, and Fabrizio Colombelli. Modelling long-term electricity load demand for rural electrification planning. *2019 IEEE Milan PowerTech, PowerTech 2019*, pages 1–6, 2019.

[35] Kenneth Lee, Edward Miguel, and Catherine Wolfram. Appliance ownership and aspirations among electric grid and home solar households in rural kenya. *American Economic Review*, 106(5):89–94, 2016.

[36] Amartya Sen. Equality of what? In *The Tanner Lecture on Human Values*, number 1, pages 197–220. Cambridge University Press, 1980.

[37] Francesco Fuso Nerini, Mark Howells, Morgan Bazilian, and Maria F. Gomez. Rural electrification options in the Brazilian Amazon. A multi-criteria analysis. *Energy for Sustainable Development*, 20(1):36–48, 2014.

[38] Md Mizanur Rahman, Jukka V. Paatero, and Risto Lahdelma. Evaluation of choices for sustainable rural electrification in developing countries: A multicriteria approach. *Energy Policy*, 59:589–599, 2013.

[39] Marina Petrelli, Davide Fioriti, Alberto Berizzi, Cristian Bovo, and Davide Poli. A novel multi-objective method with online Pareto pruning for multi-year optimization of rural microgrids. *Applied Energy*, 299(February):117283, 2021.

[40] Bernard Tenenbaum, Chris Greacen, and Dipti Vaghela. Mini-Grids and Arrival of the Main Grid. *Mini-Grids and Arrival of the Main Grid*, 2018.

[41] Nima Izadyar, Hwai Chyuan Ong, W. T. Chong, and K. Y. Leong. Resource assessment of the renewable energy potential for a remote area: A review. *Renewable and Sustainable Energy Reviews*, 62:908–923, 2016.

[42] Paul Bolstad. *GIS Fundamentals : A First Text on Geographic Information System5 th Edition.* 2016.

[43] Subhes C. Bhattacharyya. Review of alternative methodologies for analysing off-grid electricity supply. *Renewable and Sustainable Energy Reviews*, 16(1):677–694, 2012.

[44] Hengky K. Salim, Rodney A. Stewart, Oz Sahin, and Michael Dudley. Drivers, barriers and enablers to end-of-life management of solar photovoltaic and battery energy storage systems: A systematic literature review. *Journal of Cleaner Production*, 211:537–554, 2019.

[45] Nuria Martín-Chivelet. Photovoltaic potential and land-use estimation methodology. *Energy*, 94:233–242, 2016.

[46] International Energy Agency. The Role of Critical Minerals in Clean Energy Transitions. *IEA Publications*, page 283, 2021.

[47] Philip Andrews-Speed and Sufang Zhang. The Water-Energy-Food Nexus. *China as a Global Clean Energy Champion*, pages 215–243, 2019.

[48] IEA. Water Energy Nexus- Excerpt from the World Energy Outlook 2016. *Iea*, page 60, 2016.

[49] International Renewable Energy Agency (IRENA). *Policies and regulations for renewable mini-grids*. 2018.

[50] Thomas Ackermann, Nis Martensen, Tom Brown, Peter-Philipp Schierhorn, Francisco Gafaro Boshell, and Maria Ayuso. Scaling Up Variable Renewable Power: The Role of Grid Codes. page 106, 2016.

[51] J Abayateye, S Corigliano, M Merlo, and D Zimmerle. BESS Primary Frequency Control Strategies for the West Africa Power Pool. 2022.

[52] Francis Kemausuor, Morkporkpor Delight Sedzro, and Isaac Osei. Decentralised Energy Systems in Africa: Coordination and Integration of Off-Grid and Grid Power Systems—Review of Planning Tools to Identify Renewable Energy Deployment Options for Rural Electrification in Africa. *Current Sustainable/Renewable Energy Reports*, 5(4):214–223, 2018.

[53] James Morrissey. Achieving universal electricity access at the lowest cost: A comparison of published model results. *Energy for Sustainable Development*, 53:81–96, 2019.

[54] Pedro Ciller and Sara Lumbreras. Electricity for all: The contribution of large-scale planning tools to the energy-access problem. *Renewable and Sustainable Energy Reviews*, 120(November 2019):109624, 2020.

[55] HOMER Energy LLC. HOMER Pro - Microgrid Software for Designing Optimized Hybrid Microgrids. http://www.homerenergy.com/.

[56] Sergio Balderrama, Walter Canedo, Miguel Fernandez, Vincent Lemort, and Sylvain Quoilin. Techno-economic optimization of isolate microgrids including PV and Li-Ion Batteries in the Bolivian context. *The 29th International Conference on Efficiency, Cost, Optimization and Environmental Impact of Energy Systems*, (January):1–12, 2016.

[57] Rodolfo Dufo-López, Juan M. Lujano-Rojas, and José L. Bernal-Agustín. Comparison of different lead-acid battery lifetime prediction models for use in simulation of stand-alone photovoltaic systems. *Applied Energy*, 115:242–253, 2014.

[58] Salman Mashayekh, Michael Stadler, Goncalo Cardoso, and Miguel Heleno. A Mixed Integer Linear Programming Approach for Planning and Design of Networked Multi-Energy Microgrids. *Applied Energy*, 187(November):154–168, 2017.

[59] R. Luna-Rubio, M. Trejo-Perea, D. Vargas-Vázquez, and G. J. Ríos-Moreno. Optimal sizing of renewable hybrids energy systems: A review of methodologies. *Solar Energy*, 86(4):1077–1088, 2012.

[60] Sunanda Sinha and S.S. Chandel. Review of software tools for hybrid renewable energy systems. *Renewable and Sustainable Energy Reviews*, 32:192–205, apr 2014.

[61] Stefan Pfenninger and Bryn Pickering. Calliope: a multi-scale energy systems modelling framework. *Journal of Open Source Software*, 3(29):825, sep 2018.

[62] Mark Howells, Holger Rogner, Neil Strachan, Charles Heaps, Hillard Huntington, Socrates Kypreos, Alison Hughes, Semida Silveira, Joe DeCarolis, Morgan Bazillian, and Alexander Roehrl. OSeMOSYS: The Open Source Energy Modeling System. An introduction to its ethos, structure and development. *Energy Policy*, 39(10):5850–5870, 2011.

[63] S. Hilpert, C. Kaldemeyer, U. Krien, S. Günther, C. Wingenbach, and G. Plessmann. The Open Energy Modelling Framework (oemof) - A

new approach to facilitate open science in energy system modelling. *Energy Strategy Reviews*, 22(July):16–25, 2018.

[64] T. Niet, A. Shivakumar, F. Gardumi, W. Usher, E. Williams, and M. Howells. Developing a community of practice around an open source energy modelling tool. *Energy Strategy Reviews*, 35(February):100650, 2021.

[65] Nandi Moksnes, Alexandros Korkovelos, Dimitrios Mentis, and Mark Howells. Electrification pathways for Kenya-linking spatial electrification analysis and medium to long term energy planning. *Environmental Research Letters*, 12(9), 2017.

[66] Simon Hilpert, Franziska Dettner, and Ahmed Al-Salaymeh. Analysis of cost-optimal renewable energy expansion for the near-term jordanian electricity system. *Sustainability (Switzerland)*, 12(22):1–21, 2020.

[67] Md Nasimul Islam Maruf. A novel method for analyzing highly renewable and sector-coupled subnational energy systems-case study of schleswig-holstein. *Sustainability (Switzerland)*, 13(7), 2021.

[68] Sarah Berendes, Paul Bertheau, and Philipp Blechinger. Sizing and Optimization of Hybrid Mini-Grids with micrOgridS-an Open-Source Modelling Tool. *3rd International Hybrid Power Systems Workshop*, (May), 2018.

[69] Dimitrios Mentis, Mark Howells, Holger Rogner, Alexandros Korkovelos, Christopher Arderne, Eduardo Zepeda, Shahid Siyal, Costantinos Taliotis, Morgan Bazilian, Ad De Roo, Yann Tanvez, Alexandre Oudalov, and Ernst Scholtz. Lighting the World: the first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa. *Environmental Research Letters*, 12(8), 2017.

[70] Irene Pinedo and Javier Domínguez. INTIGIS: Evaluación de Alternativas de Electrificación Rural basada en Sistemas de Información Geográfica. *XXII Congreso de Geógrafos Españoles - AGE*, I(January 2011):12, 2011.

[71] Magda Moner-Girona, Daniel Puig, Yacob Mulugetta, Ioannis Kougias, Jafaru AbdulRahman, and Sándor Szabó. Next generation interactive tool as a backbone for universal access to electricity. *Wiley Interdisciplinary Reviews: Energy and Environment*, 7(6), 2018.

[72] H. Lee Willis. *Power Distribution Planning Reference Book*. 2004.

[73] Pavlos S. Georgilakis and Nikos D. Hatziargyriou. A review of power distribution planning in the modern power systems era: Models, methods and future research. *Electric Power Systems Research*, 121:89–100, 2015.

[74] Mahdi Sedghi, Ali Ahmadian, and Masoud Aliakbar-Golkar. Assessment of optimization algorithms capability in distribution network planning: Review, comparison and modification techniques. *Renewable and Sustainable Energy Reviews*, 66:415–434, 2016.

[75] Alessandro Bosisio, Edoardo Amaldi, Alberto Berizzi, Cristian Bovo, and Stefano Fratti. A MILP approach to plan an electric urban distribution network with an H-shaped layout. *2015 IEEE Eindhoven PowerTech, PowerTech 2015*, 2015.

[76] Edelmiro Miguez, José Cidrás, Eloy Díaz-Dorado, and José Luis García-Dornelas. An improved branch-exchange algorithm for large-scale distribution network planning. *IEEE Transactions on Power Systems*, 17(4):931–936, 2002.

[77] Alex Valenzuela, Iván Montalvo, and Esteban Inga. A decision-making tool for electric distribution network planning based on heuristics and georeferenced data. *Energies*, 12(21), 2019.

[78] Carlos Mateo Domingo, Tomás Gómez San Román, Álvaro Sánchez-Miralles, Jesús Pascual Peco González, and Antonio Candela Martínez. A reference network model for large-scale distribution planning with automatic street map generation. *IEEE Transactions on Power Systems*, 26(1):190–197, 2011.

[79] Pedro Ciller, Douglas Ellman, Claudio Vergara, Andres Gonzalez-Garcia, Stephen J. Lee, Cailinn Drouin, Matthew Brusnahan, Yael Borofsky, Carlos Mateo, Reja Amatya, Rafael Palacios, Robert Stoner, Fernando De Cuadra, and Ignacio Perez-Arriaga. Optimal Electrification Planning Incorporating On- And Off-Grid Technologies- And Reference Electrification Model (REM). *Proceedings of the IEEE*, 107(9):1872–1905, 2019.

[80] Francis Kemausuor, Edwin Adkins, Isaac Adu-Poku, Abeeku Brew-Hammond, and Vijay Modi. Electrification planning using Network

Planner tool: The case of Ghana. *Energy for Sustainable Development*, 19(1):92–101, 2014.

[81] IED. GEOSIM [online]. Date Accessed: 2019-09-10, 2018.

[82] R. Fronius. Rural electrication planning software (LAPER). pages v5–20–v5–20, 2005.

[83] Philipp Blechinger, Catherina Cader, and Paul Bertheau. Least-Cost Electrification Modeling and Planning - A Case Study for Five Nigerian Federal States. *Proceedings of the IEEE*, 107(9):1923–1940, 2019.

[84] Rwanda Energy Group. TheNational Electrification Plan: Report on definition of technologies (On-grid and off-grid) at village level. (June), 2019.

[85] IED. Cambodia-rural electrification plan. Technical report, 2011.

[86] IEA. Southeast Asia Energy Outlook 2019 – Analysis. *Iea*, (October), 2019.

[87] Thai Phuong Do and Marie Cecile Alvarez-Herault. Distribution network planning tool for rural areas. *2019 IEEE Milan PowerTech, PowerTech 2019*, pages 1–6, 2019.

[88] Silvia Corigliano, Tommaso Carnovali, Darlain Edeme, and Marco Merlo. Holistic geospatial data-based procedure for electric network design and least-cost energy strategy. *Energy for Sustainable Development*, 2020.

[89] T Gadelha Filho, Corigliano Silvia, Dimovski Aleksandar, Bolognesi Massimo, and Merlo Marco. Rural electrification planning based on graph theory and geospatial data: A realistic topology oriented approach. *Sustainable Energy, Grids and Networks*, 28:100525, 2021.

[90] Nicolò Stevanato, Silvia Corigliano, Marina Petrelli, Francesco Tonini, Marco Merlo, and Emanuela Colombo. Rural areas electrification strategies through shadow costs analysis - Bolivian Highlands case study. *Energy for Sustainable Development*, 65:162–174, 2021.

[91] AETS. FORMULATION OF THE LESOTHO ELECTRIFICATION MASTERPLAN. 2017.

[92] IBGE. Censo demográfico. 2010.

[93] EPE. Balanço energético nacional: Ano base 2018. *EPE - Empresa de Pesquisa Energética*, page 67, 2019.

[94] Engenharia El and Engenharia El. Álvaro Rocha Albuquerque Fluxo de Caixa em Risco : Uma Nova Abordagem para o Setor de Distribuição de Energia Elétrica Dissertação de Mestrado Álvaro Rocha Albuquerque Fluxo de Caixa em Risco : Uma Nova Abordagem para o Setor de Distribuição de Energia El. 2008.

[95] Alexandre Cosme José Jeronymo and Sinclair Mallet-Guy Guerra. Caracterizando a evolução da eletrificação rural brasileira. *Redes*, 23(1):133, 2018.

[96] Silvia Corigliano, Federico Rosato, Carla Ortiz Dominguez, and Marco Merlo. Clustering techniques for secondary substations siting. *Energies*, 14(4):1028, 2021.

[97] Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. High Resolution Settlement Layer (HRSL) [online]. Date accessed: 2019-09-20, 2016.

[98] Andrew J Tatem. Worldpop, open data for spatial demography. *Scientific data*, 4(1):1–4, 2017.

[99] Babak Khavari, Alexandros Korkovelos, Andreas Sahlberg, Mark Howells, and Francesco Fuso Nerini. Population cluster data to assess the urban-rural split and electrification in Sub-Saharan Africa. *Scientific Data*, 8(1):1–11, 2021.

[100] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques.* 2012.

[101] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[102] Martin Ester, Jörg Sander, Hans-Peter Kriegel, and Xiaowei Xu. A Desity-Based Algorithm for Discovering Clusters. *Data Mining and Knowledge Discovery*, 2(2):169–194, 1998.

[103] Mihael Ankerst, Markus M. Breunig, Hans peter Kriegel, and Jörg Sander. Optics: Ordering points to identify the clustering structure. pages 49–60. ACM Press, 1999.

[104] Hajar Rehioui, Abdellah Idrissi, Manar Abourezq, and Faouzia Zegrari. Denclue-im: A new approach for big data clustering. *Procedia Computer Science*, 83:560–567, 2016. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.

[105] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. Density-based clustering based on hierarchical density estimates. In Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[106] Claudia Malzer and Marcus Baum. A Hybrid Approach to Hierarchical Density-based Cluster Selection. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2020-Septe:223–228, 2020.

[107] Luca Lorenzoni, Paolo Cherubini, Davide Fioriti, Davide Poli, Andrea Micangeli, and Romano Giglioli. Classification and modeling of load profiles of isolated mini-grids in developing countries: A data-driven approach. *Energy for Sustainable Development*, 59:208–225, 2020.

[108] Fabio Riva, Helene Ahlborg, Elias Hartvigsson, Shonali Pachauri, and Emanuela Colombo. Electricity access and rural development: Review of complex socio-economic dynamics and causal diagrams for more appropriate energy modelling. *Energy for Sustainable Development*, 43:203–223, 2018.

[109] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, 2009.

[110] Subhes C. Bhattacharyya and Govinda R. Timilsina. Modelling energy demand of developing countries: Are the specific features adequately captured? *Energy Policy*, 38(4):1979–1990, 2010.

[111] Stefano Mandelli, Marco Merlo, and Emanuela Colombo. Novel procedure to formulate load profiles for off-grid rural areas. *Energy for Sustainable Development*, 31:130–142, 2016.

[112] Francesco , Sergio Balderrama, Sylvain Quoilin, and Emanuela Colombo. Generating high-resolution multi-energy load profiles for re-

mote areas with an open-source stochastic model. *Energy*, 177:433–444, 2019.

[113] Gouthami Padam, Dana Rysankova, Elisa Portale, Bryan Bonsuk Koo, Sandra Keller, and Gina Fleurantin. *Ethiopia–Beyond Connections: Energy Access Diagnostic Report Based on the Multi-Tier Framework*. World Bank, 2018.

[114] Lucia Luzi, Yunhui Lin, Brian Bonsuk Koo, Dana Rysankova, and Elisa Portale. Zambia–beyond connections: Energy access diagnostic report based on the multi-tier framework, 2019.

[115] Bryan Bonsuk Koo, Dana Rysankova, Elisa Portale, Niki Angelou, Sandra Keller, and Gouthami Padam. Rwanda–beyond connections. 2018.

[116] Sunita Dubey, Ehui Adovor, Dana Rysankova, and Bonsuk Koo. Kenya-beyond connections. 2020.

[117] Giacomo Falchetta, Nicolo Stevanato, Magda Moner-Girona, Davide Mazzoni, Emanuela Colombo, and Manfred Hafner. The M-LED platform: Advancing electricity demand assessment for communities living in energy poverty. *Environmental Research Letters*, 16(7), 2021.

[118] Mikul Bhatia and Nicolina Angelou. Beyond Connections Energy Access Redefined. World Bank. Energy Sector Management Assistance Program (ESMAP). *The World Bank*, pages 1–224, 2015.

[119] Harald Olk and Juliane Mundt. *Photovoltaics for productive use applications*. GIZ, 2016.

[120] Lucius Mayer-Trasch, Mohua Mukherjee, and Kilian Reiche. *Productive use of energy-PRODUSE: measuring impacts of electrification on small and micro-enterprises in sub-Saharan Africa*. GIZ, 2013.

[121] Lighting Global. The market opportunity for productive use leveraging solar energy (pulse) in sub-saharan africa, 2019.

[122] IFPRI. Global spatially-disaggregated crop production statistics data for 2010 version 2.0. 2019.

[123] Michele M Rienecker, Max J Suarez, Ronald Gelaro, Ricardo Todling, Julio Bacmeister, Emily Liu, Michael G Bosilovich, Siegfried D Schubert, Lawrence Takacs, Gi-Kong Kim, et al. Merra: Nasa's modern-era retrospective analysis for research and applications. *Journal of climate*, 24(14):3624–3648, 2011.

[124] Hans Hersbach, B Bell, P Berrisford, AJMS Horányi, J Muñoz Sabater, J Nicolas, R Radu, D Schepers, A Simmons, C Soci, et al. Global reanalysis: goodbye era-interim, hello era5. *ECMWF newsletter*, 159:17–24, 2019.

[125] Stefan Pfenninger and Iain Staffell. Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data. *Energy*, 114:1251–1265, 2016.

[126] Iain Staffell and Stefan Pfenninger. Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, 114:1224–1239, 2016.

[127] Marcel Šúri, Thomas A Huld, Ewan D Dunlop, and Heinz A Ossenbrink. Potential of solar electricity generation in the european union member states and candidate countries. *Solar energy*, 81(10):1295–1305, 2007.

[128] Bernhard Lehner and Günther Grill. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15):2171–2186, 2013.

[129] Alexandros Korkovelos, Dimitrios Mentis, Shahid Hussain Siyal, Christopher Arderne, Holger Rogner, Morgan Bazilian, Mark Howells, Hylke Beck, and Ad De Roo. A geospatial assessment of small-scale hydropower potential in sub-saharan Africa. *Energies*, 11(11), 2018.

[130] Philip W Gassman, Manuel R Reyes, Colleen H Green, and Jeffrey G Arnold. The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASABE*, 50(4):1211–1250, 2007.

[131] Vincenzo Sammartano, Lorena Liuzzo, and Gabriele Freni. Identification of potential locations for run-of-river hydropower plants using a GIS-based procedure. *Energies*, 12(18):1–20, 2019.

[132] Yihun Taddele Dile and Raghavan Srinivasan. Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: An application in the blue nile river basin. *Journal of the American Water Resources Association*, 50(5):1226–1241, 2014.

[133] Surendra Kumar. Mishra and V. P. (Vijay P.) Singh. *Soil Conservation Service Curve Number (SCS-CN) Methodology*. Springer Netherlands, 2003.

[134] R. G. Cronshey, R. T. Roberts, and N. Miller. Urban Hydrology for Small Watersheds (Tr-55 Rev.). *Soil Conservation*, (Technical Release 55 (TR-55)):1268–1273, 1985.

[135] CGIAR CSI. Global Geospatial Potential EvapoTranspiration & Aridity Index Methodology and Dataset Description Global. (1948):1–16, 2005.

[136] Tom G Farr, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. The shuttle radar topography mission. *Reviews of geophysics*, 45(2), 2007.

[137] FAO and UNESCO. FAO-UNESCO soil map of the world, 1:5000000. Africa. *Fao Soil Bulletin*, VI(1):346, 1977.

[138] T.R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, J Zhu, L. Yang, and J.W Merchant. Globlal Land Cover Characteristics Data Base Version 2.0. *International Journal of Remote Sensing*, 21(6/7):1303–1330, 2000.

[139] Suranjana Saha, Shrinivas Moorthi, Hua Lu Pan, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, Robert Kistler, John Woollen, David Behringer, Haixia Liu, Diane Stokes, Robert Grumbine, George Gayno, Jun Wang, Yu Tai Hou, Hui Ya Chuang, Hann Ming H. Juang, Joe Sela, Mark Iredell, Russ Treadon, Daryl Kleist, Paul Van Delst, Dennis Keyser, John Derber, Michael Ek, Jesse Meng, Helin Wei, Rongqian Yang, Stephen Lord, Huug Van Den Dool, Arun Kumar, Wanqiu Wang, Craig Long, Muthuvel Chelliah, Yan Xue, Boyin Huang, Jae Kyung Schemm, Wesley Ebisuzaki, Roger Lin, Pingping Xie, Mingyue Chen, Shuntai Zhou, Wayne Higgins, Cheng Zhi Zou, Quanhua Liu, Yong Chen, Yong Han, Lidia Cucurull, Richard W. Reynolds, Glenn Rutledge, and Mitch Goldberg. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1057, 2010.

[140] Boini Narsimlu, Ashvin K. Gosain, Baghu R. Chahar, Sudhir Kumar Singh, and Prashant K. Srivastava. SWAT Model Calibration and Uncertainty Analysis for Streamflow Prediction in the Kunwari River Basin, India, Using Sequential Uncertainty Fitting. *Environmental Processes*, 2(1):79–95, 2015.

[141] Yves Tramblay, Nathalie Rouché, Jean-Emmanuel Paturel, Gil Mahé, Jean-François Boyer, Ernest Amoussou, Ansoumana Bodian, Honoré

Dacosta, Hamouda Dakhlaoui, Alain Dezetter, et al. Adhi: the african database of hydrometric indices (1950–2018). *Earth System Science Data*, 13(4):1547–1560, 2021.

[142] JF Boyer, C Dieulin, N Rouche, A Cres, E Servat, and JE Paturel. Sierem: an environmental information system for water. *Climate Variability and Change–hydrological Impacts*, (308):19, 2006.

[143] Claudine Dieulin, Gil Mahé, Jean-Emmanuel Paturel, Soundouss Ejjiyar, Yves Tramblay, Nathalie Rouché, and Bouabid EL Mansouri. A new 60-year 1940/1999 monthly-gridded rainfall data set for africa. *Water*, 11(2), 2019.

[144] Marina Petrelli, Davide Fioriti, Alberto Berizzi, and Davide Poli. Multiyear planning of a rural microgrid considering storage degradation. *IEEE Transactions on Power Systems*, 36(2):1459–1469, 2020.

[145] Claudio Brivio, Matteo Moncecchi, Stefano Mandelli, and Marco Merlo. A novel software package for the robust design of off-grid power systems. *Journal of Cleaner Production*, 166:668–679, 2017.

[146] Bo Zhao, Xuesong Zhang, Peng Li, Ke Wang, Meidong Xue, and Caisheng Wang. Optimal sizing, operating strategy and operational experience of a stand-alone microgrid on dongfushan island. *Applied Energy*, 113:1656–1666, 2014.

[147] Bei Li, Robin Roche, and Abdellatif Miraoui. Microgrid sizing with combined evolutionary algorithm and MILP unit commitment. *Applied Energy*, 188:547–562, 2017.

[148] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. *Pyomo–optimization modeling in python*, volume 67. Springer Science & Business Media, third edition, 2021.

[149] William E Hart, Jean-Paul Watson, and David L Woodruff. Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3(3):219–260, 2011.

[150] Godfrey Gladson Moshi, Cristian Bovo, Alberto Berizzi, and Leonardo Taccari. Optimization of integrated design and operation of microgrids under uncertainty. *19th Power Systems Computation Conference, PSCC 2016*, 2016.

[151] Dai Hongwei, Yu Yixin, Huang Chunhua, Wang Chengshan, Ge Shaoyun, Xiao Jim, Zhou Yi, and Xin Rui. Optimal planning of distribution substation locations and sizes—model and algorithm. *International Journal of Electrical Power & Energy Systems*, 18(6):353–357, 1996.

[152] D I Sun, D R Farris, P J Cote, R R Shoults, and M S Chen. Perform Fixed Char e Network Anal. *IEEE Transactions On Power Apparatus And Systems*, 101(3):602–609, 1982.

[153] MA El-Kady. Computer-aided planning of distribution substation and primary feeders. *IEEE transactions on power apparatus and systems*, (6):1183–1189, 1984.

[154] WM Lin, MT Tsay, and SW Wu. Application of geographic information system for substation and feeder planning. *International Journal of Electrical Power & Energy Systems*, 18(3):175–183, 1996.

[155] Dale M Crawford. AREAS Dale. (2):2–7, 1975.

[156] Eloy Díaz-Dorado, Edelmiro Miguez, and José Cidrás. Design of large rural low-voltage networks using dynamic programming optimization. *IEEE Transactions on Power Systems*, 16(4):898–903, 2001.

[157] Gabriela C Cabrera-Celi, Edison G Novoa-Guaman, and Paul F Vasquez-Miranda. Design of secondary circuits of distribution networks using clustering and shortest path algorithms. In *2017 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*, pages 1–6. IEEE, 2017.

[158] Antonio Marcos Cossi, Rubén Romero, and José Roberto Sanches Mantovani. Planning and projects of secondary electric power distribution systems. *IEEE Transactions on power systems*, 24(3):1599–1608, 2009.

[159] Mostafa Esmaeeli, Ahad Kazemi, Heydar-Ali Shayanfar, and Mahmood-Reza Haghifam. Sizing and placement of distribution substations considering optimal loading of transformers. *International Transactions on Electrical Energy Systems*, 25(11):2897–2908, 2015.

[160] Jorge E. Mendoza, Miguel E. López, Héctor E. Peña, and David A. Labra. Low voltage distribution optimization: Site, quantity and size of distribution transformers. *Electric Power Systems Research*, 91:52–60, 2012.

[161] S Najafi and R Gholizadeh. On optimal sizing, siting and timing of distribution substations. In *18th Electric Power Distribution Conference*, pages 1–6. IEEE, 2013.

[162] Ihsan Jabbar Hasan, Chin Kim Gan, Meysam Shamshiri, Mohd Ruddin Ab Ghani, and Rosli Bin Omar. Optimum feeder routing and distribution substation placement and sizing using pso and mst. *Indian Journal of Science and Technology*, 7(10):1682–1689, 2014.

[163] Jiagang Xu, Rutooj D. Deshpande, Jie Pan, Yang-Tse Cheng, and Vincent S. Battaglia. Electrode Side Reactions, Capacity Loss and Mechanical Degradation in Lithium-Ion Batteries. *Journal of The Electrochemical Society*, 162(10):A2026–A2035, 2015.

[164] Li Yu, Di Shi, Xiaobin Guo, Zhen Jiang, Guangyue Xu, Ganyang Jian, Jinyong Lei, and Chaoyang Jing. An efficient substation placement and sizing strategy based on gis using semi-supervised learning. *CSEE journal of power and energy systems*, 4(3):371–379, 2018.

[165] Luis González-Sotres, Carlos Mateo Domingo, Álvaro Sánchez-Miralles, and Manuel Alvar Miró. Large-scale mv/lv transformer substation planning considering network costs and flexible area decomposition. *IEEE transactions on power delivery*, 28(4):2245–2253, 2013.

[166] Alejandro Navarro and Hugh Rudnick. Large-scale distribution planning - Part I: Simultaneous network and transformer optimization. *IEEE Transactions on Power Systems*, 24(2):744–751, 2009.

[167] Zhaoyu Wang, Bokan Chen, Jianhui Wang, Jinho Kim, and Miroslav M. Begovic. Robust optimization based optimal DG placement in microgrids. *IEEE Transactions on Smart Grid*, 5(5):2173–2182, 2014.

[168] Joseph A. Lukes. Efficient algorithm for the partitioning of trees. *IBM Journal of Research and Development*, 18(3):217–224, 1974.

[169] K Vinothkumar and MP Selvan. Hierarchical agglomerative clustering algorithm method for distributed generation planning. *International Journal of Electrical Power & Energy Systems*, 56:259–269, 2014.

[170] Thomas R. Etherington. Least-Cost Modelling and Landscape Ecology: Concepts, Applications, and Opportunities. *Current Landscape Ecology Reports*, 1(1):40–53, 2016.

[171] Joost van Bemmelen, Wilko Quak, Marcel van Hekken, and Peter van Oosterom. Vector vs . Raster-based Algorithms for Cross Country Movement Planning. *Proceedings Auto-Carto*, 11:pages 304–317, 1993.

[172] Joseph S.B. Mitchell. Geometric Shortest Paths and Network Optimization. *Handbook of Computational Geometry*, pages 633–701, 2000.

[173] Joseph Mitchell and Christos Papadimitiou. The Weighted Regions Problem: finding shortest paths through a weighted planar subdivision.

[174] Federico Rosato. Heuristic graph partitioning with preferred cluster sizes and application to the generation of realistic distribution grid topologies. In *2021 IEEE 15th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, pages 1–7. IEEE, 2021.

[175] Stefano Bagli, Davide Geneletti, and Francesco Orsi. Routeing of power lines through least-cost path analysis and multicriteria evaluation to minimise environmental impacts. *Environmental Impact Assessment Review*, 31(3):234–239, 2011.

[176] Yosoon Choi, Jeong-Gi Um, and Myong-Ho Park. Finding least-cost paths across a continuous raster surface with discrete vector networks. *Cartography and Geographic Information Science*, 41(1):75–85, 2014.

[177] Uwe Deichmann, Craig Meisner, Siobhan Murray, and David Wheeler. The economics of renewable energy expansion in rural Sub-Saharan Africa. *Energy Policy*, 39(1):215–227, 2011.

[178] Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50, 1956.

[179] Robert Clay Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957.

[180] Insu Hong and Alan T. Murray. Assessing Raster GIS Approximation for Euclidean Shortest Path Routing. *Transactions in GIS*, 20(4):570–584, 2016.

[181] Joram Schito, Ulrike Wissen Hayek, and Martin Raubal. Enhanced multi criteria decision analysis for planning power transmission lines. *Leibniz International Proceedings in Informatics, LIPIcs*, 114, 2018.

[182] Camil Demetrescu, Irene Finocchi, and Giuseppe F. Italiano. *Algoritmi e Strutture Dati*. Mc Graw Hill, 2 edition, 2008.

[183] Bang Ye Wu and Kun Mao Chao. *Spanning trees and optimization problems*. 2004.

[184] Gabriel Robins and Alexander Zelikovsky. Minimum Steiner Tree Construction*. *Handbook of Algorithms for Physical Design Automation*, 2008.

[185] Stefan Hougardy, Jannik Silvanus, and Jens Vygen. Dijkstra meets Steiner: a fast exact goal-oriented Steiner tree algorithm. *Mathematical Programming Computation*, 9(2):135–202, 2017.

[186] Andrew B. Kahng and Gabriel Robins. A New Class of Iterative Steiner Tree Heuristics with Good Performance. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1992.

[187] Bang Ye Wu and Kun-Mao Chao. *Spanning trees and optimization problems*. Chapman and Hall/CRC, 2004.

[188] Paul Bertheau and Catherina Cader. Electricity sector planning for the Philippine islands: Considering centralized and decentralized supply options. *Applied Energy*, 2019.

[189] Yakubu Abdul-Salam and Euan Phimister. How effective are heuristic solutions for electricity planning in developing countries. *Socio-Economic Planning Sciences*, 55:14–24, 2016.

[190] Yakubu Abdul-Salam and Euan Phimister. The politico-economics of electricity planning in developing countries: A case study of Ghana. *Energy Policy*, 88:299–309, 2016.

[191] Lily Parshall, Dana Pillai, Shashank Mohan, Aly Sanoh, and Vijay Modi. National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. *Energy Policy*, 37(6):2395–2410, 2009.

[192] Takao Yokota, Mitsuo Gen, and Yin-Xiu Li. Genetic algorithm for non-linear mixed integer programming problems and its applications. *Computers & Industrial Engineering*, 30(4):905–917, 1996.

[193] Kusum Deep, Krishna Pratap Singh, Mitthan Lal Kansal, and C Mohan. A real coded genetic algorithm for solving integer and mixed

integer optimization problems. *Applied Mathematics and Computation*, 212(2):505–518, 2009.

[194] L. Thurner, A. Scheidler, F. Schafer, J. H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun. pandapower - an open source python tool for convenient modeling, analysis and optimization of electric power systems. *IEEE Transactions on Power Systems*, 2018.

[195] T. H.M. El-Fouly, H. H. Zeineldin, E. F. El-Saadany, and M. M.A. Salama. A new optimization model for distribution substation siting, sizing, and timing. *International Journal of Electrical Power and Energy Systems*, 30(5):308–315, 2008.

[196] M. A. Farrag, M. M. El-Metwally, and M. S. El-Bages. A new model for distribution system planning. *International Journal of Electrical Power and Energy Systems*, 21(7):523–531, 1999.

[197] A. López-González, B. Domenech, and L. Ferrer-Martí. Formative evaluation of sustainability in rural electrification programs from a management perspective: A case study from Venezuela. *Renewable and Sustainable Energy Reviews*, 95(July):95–109, 2018.

[198] G Chiandussi, M Codegone, S Ferrero, and F E Varesio. Comparison of multi-objective optimization methodologies for engineering applications. *Computers and Mathematics with Applications*, 63(5):912–942, 2012.

[199] Bo Zhao, Xuesong Zhang, Peng Li, Ke Wang, Meidong Xue, and Caisheng Wang. Optimal sizing, operating strategy and operational experience of a stand-alone microgrid on Dongfushan Island. *Applied Energy*, 113:1656–1666, 2014.

[200] Mohammad Jooshaki, Ali Abbaspour, Mahmud Fotuhi-Firuzabad, Hossein Farzin, Moein Moeini-Aghtaie, and Matti Lehtonen. A milp model for incorporating reliability indices in distribution system expansion planning. *IEEE Transactions on Power Systems*, 34(3):2453–2456, 2019.

[201] Gregorio Munoz-Delgado, Javier Contreras, and Jose M. Arroyo. Distribution Network Expansion Planning with an Explicit Formulation for Reliability Assessment. *IEEE Transactions on Power Systems*, 33(3):2583–2596, 2018.

[202] Seyed Mahdi Mazhari, Hassan Monsef, and Ruben Romero. A Multi-Objective Distribution System Expansion Planning Incorporating Customer Choices on Reliability. *IEEE Transactions on Power Systems*, 31(2):1330–1340, 2016.

[203] S. Ganguly, N. C. Sahoo, and D. Das. Multi-objective planning of electrical distribution systems using dynamic programming. *International Journal of Electrical Power and Energy Systems*, 46(1):65–78, 2013.

[204] Ottmar Edenhofer, Ramón Pichs-Madruga, Youba Sokona, Kristin Seyboth, Susanne Kadner, Timm Zwickel, Patrick Eickemeier, Gerrit Hansen, Steffen Schlömer, Christoph von Stechow, et al. *Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change*. Cambridge University Press, 2011.

[205] Matteo V. Rocco, Alberto Di Lucchio, and Emanuela Colombo. Exergy Life Cycle Assessment of electricity production from Waste-to-Energy technology: A Hybrid Input-Output approach. *Applied Energy*, 194:832–844, 2017.

[206] Matteo V. Rocco and Emanuela Colombo. Exergy Life Cycle Assessment of a Waste-to-Energy Plant. *Energy Procedia*, 104:562–567, 2016.

[207] Doing Business, Doing Business, and Doing Business. Getting electricity: Measuring reliability, prices and transparency. pages 70–77, 2015.

[208] Jay Taneja. Measuring Electricity Reliability in Kenya. page 6, 2016.

[209] Isa Ferrall, Arthur Jacquiau-chamski, and Daniel M Kammen. A collaboration between RAEL and Sparkmeter Electricity reliability patterns in grids and minigrids in East Africa A Minigrid focus in California and in Energy Access. (February):1–21, 2020.

[210] Wei Li, Natasha MacBean, Philippe Ciais, Pierre Defourny, Céline Lamarche, Sophie Bontemps, Richard A Houghton, and Shushi Peng. Gross and net land cover changes in the main plant functional types derived from the annual esa cci land cover maps (1992–2015). *Earth System Science Data*, 10(1):219–234, 2018.

[211] OpenStreetMap contributors. OpenStreetMap [online]. Date Accessed: 2019-09-20.

[212] Tobias G. Tiecke, Xianming Liu, Amy Zhang, Andreas Gros, Nan Li, Gregory Yetman, Talip Kilic, Siobhan Murray, Brian Blankespoor, Espen B. Prydz, and Hai-Anh H. Dang. Mapping the World Population One Building at a Time. *Mapping the World Population One Building at a Time*, 2017.

[213] E Rodriguez, Cs Morris, and Je Belz. An assessment of the SRTM topographic products. *Photogrammetric Engineering and Remote Sensing*, 72(3):249–260, 2006.

[214] CrowdVal Project. Accuracy assessment of the 20 M Land Cover Map of Africa. *LACO-Wiki*.

[215] Lindy-Anne Siebritz. Assessing the Accuracy of OpenStreetMap Data in South Africa for the Purpose of Integrating it with Authoritative Data. *Master of Science in Engineering dissertation, University of Cape Town*, 2014.

[216] Miguel Muñoz Ortiz, Lisa Kvalbein, and Lars Hellemo. Evaluation of open photovoltaic and wind production time series for Norwegian locations. *Energy*, 236, 2021.

[217] Christopher Jung and Dirk Schindler. Integration of small-scale surface properties in a new high resolution global wind speed model. *Energy Conversion and Management*, 210(March):112733, 2020.

[218] M. Crosetto and S. Tarantola. Uncertainty and sensitivity analysis: Tools for GIS-based model implementation. *International Journal of Geographical Information Science*, 15(5):415–437, 2001.