Programa de Doctorado en Energía Eléctrica

THE RURAL ELECTRIFICATION PLANNING PROBLEM: STRATEGIES AND SOLUTIONS

Autor:
Pedro Ciller Cutillas

Directores:
Dra. Sara Lumbreras Sancho
Dr. José Ignacio Pérez Arriaga

Instituto de Investigación Tecnológica
Escuela Técnica Superior de Ingeniería ICAI
Universidad Pontificia Comillas

Madrid, 2020

A mi familia

"Sometimes science is more art than science, Morty. A lot of people don't get that." Rick Sánchez

Agradecimientos

A mi familia, que siempre se ha preocupado por mi formación y me ha proporcionado todos los medios necesarios.

A Sara, mi directora once sobre diez. Tu ayuda y tu apoyo han sido clave para que esta tesis haya llegado a buen puerto. Gracias por estar siempre ahí y por todas las horas que me has dedicado. No podría haber tenido una directora mejor.

A Ignacio, por darme la oportunidad de trabajar en los problemas que se abordan en esta tesis. Gracias por enseñarme que el rigor y la precisión también son importantes en ingeniería. He aprendido mucho de nuestras interacciones.

A Fernando, por proporcionar ideas que han sido fundamentales para el desarrollo de esta tesis y del modelo REM. Espero que vuelvas a jugar al ajedrez pronto.

A Andrés, por todas las horas de trabajo que hemos compartido. Gracias por animarme en los momentos complicados y espero leer tu tesis pronto. ¡Mucho ánimo!, ya solo te falta el último empujón.

A Clara, por los buenos ratos que hemos pasado trabajando juntos. ¡Que no panda el cúnico! :)

A Santos, espero que te salgan genial los exámenes y trabajos de fin de máster.

A mis compañeros del IIT, ha sido un placer compartir con todos vosotros estos años. Alberto, Salva, Fernando, Guillermo, Lorenzo, Timo, Nicolás, Leandro, Deniz, Mauricio, Sotres, Andrea, Mercedes, Jose Pablo, Celia, Peña, Leo, Rodrigo, Isaac, Diego, David, Rafa, Marta, Jesús David, Fran, Santi, Miguel, Peyman, Portela, Otaola, Nacho, Nenad y Quanyu. Sé que me dejo muchas personas sin nombrar, pero todo el IIT ha sido parte de esta tesis.

A mis compañeros del UEA Lab que han participado en esta aventura: Carlos, Rafa, Reja, Olamide, Stephen, Matthew, Pablo, Marc, Turner, Cailinn, Vivian, Patricia, Douglas y Claudio.

A los profesores y otras personas que, de un modo u otro, han contribuido a mi formación. Parte de esta tesis también es vuestra.

"If the mind is to emerge unscathed from this relentless struggle with the unforeseen, two qualities are indispensable: first, an intellect that, even in the darkest hour, retains some glimmerings of the inner light which leads to truth; and second, the courage to follow this faint light wherever it may lead." Carl von Clausewitz

"You have to have the fighting spirit. You have to force moves and take chances." Bobby Fisher

"A veces pierdes y a veces ganas, pero lo importante es la lucha." Ana Matnadze

Abstract

Universal Access to Energy is one of the most significant challenges of our time, and energy is an enabling factor that fosters development in several fields such as education and healthcare. The United Nations' seventh Sustainable Development Goal (SDG7) acknowledges the importance of energy access, and it establishes the target of achieving universal access to modern forms of energy that are affordable, reliable, and sustainable by 2030. Significant efforts are imperative to meet this deadline as there are approximately 840 million people that currently do not have access to electricity.

Establishing an electrification agenda is a complex task that depends on many socio-political factors. A suitable electrification plan should rely on solid hypotheses, rigorous analysis, and accurate data.

Computer-based models have recently gained momentum in electrification planning, as they can identify the lowest-cost designs that provide desired levels of electricity access in large-scale areas. The automated calculation of the designs can help optimize the allocation of resources devoted to universal electricity access, expediting development.

In this thesis, we focus on one electrification planning tool: the Reference Electrification Model (REM). REM determines the least-cost electrification mode for each consumer (i.e., a standalone system, a mini-grid, or an extension of the power grid). REM calculates detailed technical designs at the building level, optimizing the generation of off-grid systems and the networks of mini-grids and grid extensions.

REM is the result of ongoing teamwork. The first prototype of REM was presented in the master thesis of Douglas Ellman, which was defended at MIT, Cambridge, Massachusets, USA, in 2015. This first prototype is the starting point of this thesis.

The first prototype of REM provided inconsistent results, and substantial efforts were devoted to scrutinizing and improving its algorithms. The first part of this thesis describes several upgrades implemented into the first prototype of REM, which resulted in robust performance after the upgrades.

The second part of this thesis focuses on the development of new algorithms in REM. We present a novel method that quickly estimates the network cost of any potential low-voltage mini-grid that could appear in the solution of a large-scale planning case. We also present two clustering algorithms. The first clustering algorithm groups the consumers into mini-grids, and the second one determines which consumers should be electrified with extensions of the power grid. The new algorithms provide more optimal results than the original algorithms of REM or present other advantages.

Resumen

El acceso universal a la energía es uno de los mayores desafíos de nuestro tiempo y la energía es un factor empoderante que acelera el desarrollo en campos diversos como educación y sanidad. El séptimo objetivo de desarrollo sostenible de las naciones unidas reconoce la importancia del acceso a la energía, y aspira a alcanzar un acceso universal a formas de energía moderas que sean asequibles, fiables y sostenibles en el 2030. Es imperativo realizar esfuerzos significativos para cumplir este plazo ya que actualmente más de 840 millones de personas no tienen acceso a electricidad.

Establecer la agenda de electrificación de una región subdesarrollada es una tarea compleja que depende de multitud de factores sociopolíticos. Además, cualquier plan debe basarse en hipótesis sólidas, un análisis riguroso y datos precisos.

Recientemente, los modelos computacionales han ganado importancia, ya que pueden identificar los diseños de menor coste que satisfacen los niveles de acceso a la electricidad deseados en áreas de gran tamaño. El cálculo automatizado de los diseños puede ayudar a optimizar la distribución de recursos dedicados al acceso universal a la electricidad, acelerando su desarrollo.

En esta tesis nos enfocamos en un modelo computacional de planificación: el Modelo de Electrificación de Referencia (REM, por sus siglas en inglés). REM determina el modo de electrificación de menor coste para cada consumidor (es decir, un sistema aislado, una mini-red o una extensión de la red), y propone diseños técnicos detallados a nivel de edificio, optimizando la generación de los sistemas aislados y mini-redes y la red de las mini-redes y extensiones de red.

REM es el resultado de un gran trabajo en equipo. El primer prototipo de REM se presentó en la tesis de máster de Douglas Ellman, que fue defendida en el MIT, Cambridge, Massachusets, USA, en 2015. Este primer prototipo es el punto de partida de esta tesis doctoral.

El primer prototipo de REM proporcionaba resultados inconsistentes, y se dedicó un esfuerzo sustancial a analizar y mejorar sus algoritmos. La primera parte de esta tesis describe varias mejoras que se implementaron en el primer prototipo de REM, que mostró un comportamiento robusto tras las mejoras.

La segunda parte de esta tesis se centra en el desarrollo de nuevos algoritmos en REM. Presentamos un nuevo método que estima rápidamente el coste de red de cualquier mini-red potencial que podría ser parte de la solución de un caso de planificación a gran escala. También presentamos dos algoritmos de clustering. El primer algoritmo de clustering agrupa los consumidores en mini-redes y el segundo determina qué consumidores deberían ser electrificados con extensiones de red. Los nuevos algoritmos proporcionan resultados más óptimos que los algoritmos originales de REM o presentan otras ventajas.

Contents

List of Figures

List of Tables

Abbreviations

AC	Alternating Current
CAPEX	Capital Expenditures
CIEMAT	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas
CLARA	Clustering for LARge Applications
CLARANS	Clustering LARge Applications based on RANdomized Search
CNSE	Cost of Non-Served Energy
DC	Direct Current
DER-CAM	Distributed Energy Resources - Customer Adoption Model
DL	Distribution Layout
DP	Development Pole
EDCL	Energy Development Corporation Limited
EF	Electric Feasibility
EFOM	Energy Flow Optimization Model
GIS	Geographic Information Systems
HOMER	Hybrid Optimization Model for Multiple Energy Resources
HRSL	High Resolution Settlement Layer
HV	High Voltage
IED	Innovation Energie Développement
iHOGA	improved Hybrid Optimization by Genetic Algorithms
IIT	Institute for Research in Technology
KTH	Royal Institute of Technology
LAPER	Logiciel d'Aide à la Planification d'Électrification Rurale

LCOE	Levelized Cost of Electricity
LREM	Local REM
LV	Low Voltage
MAR	Minimal Area Rectangle
MARKAL	MARKet Allocation
MIT	Massachusetts Institute of Technology
MILP	Mixed Integer Linear Programming
MINLP	Mixed Integer Non-Linear Programming
MIP	Mixed Integer Programming
MST	Minimum Spanning Tree
MV	Medium Voltage
NL	Network Layout
NPAM	Network Performance Assessment Model
NSE	Non-Served Energy
OnSSET	Open Source Spatial Electrification Tool
OPEX	Operational Expenditures
OSeMOSYS	Open Source Energy Modelling System
PAM	Partitioning Around Medoids
PCA	Principal Component Analysis
REM	Reference Electrification Model
RLI	Renier Lemoine Institut
RNM	Reference Network Model
SDG7	Seventh Sustainable Development Goal
T	Topography

TE	Transmission Expansion
TIMES	The Integrated MARKAL-EFOM System
UEA Lab	Universal Energy Access Laboratory
UR	Upstream Reinforcements
WB	World Bank

> "Perhaps I could best describe my experience of doing mathematics in terms of entering a dark mansion. You go into the first room and it's dark, completely dark. You stumble around, bumping into the furniture. Gradually, you learn where each piece of furniture is. And finally, after six months or so, you find the light switch and turn it on. Suddenly, it's all illuminated and you can see exactly where you were. Then you enter the next dark room..." *Andrew Wiles.*

1 INTRODUCTION

Access to electricity remains a crucial challenge in many parts of the world. The complexity of electrification planning in an underserved region is partially exemplified by the technical alternatives available: centralized grid extension has been the status quo for over a century, but off-grid mini-grid and individual standalone systems have recently gained popularity.

Public governments, private investors, and entrepreneurs (among others) would benefit from the knowledge of the least-cost electrification modes and system designs over their territories of interest, to be used as a basis upon which to add further considerations. The magnitude of the universal energy access challenge, the amount of information involved, and the diversity of options for intervention compel the use of computer-based planning models.

In this thesis, we describe one of these models — the Reference Electrification Model (REM) — which is a state of the art tool. The first prototype of REM was developed at the Massachusets Institute of Technology (MIT), but it provided inconsistent results. The first part of this thesis (chapter 3 and chapter 4) focuses on scrutinizing and improving the previously developed algorithms of REM so that the model could be applied in actual planning projects. The second part of this thesis (chapter 5 and chapter 6) focuses on the development of novel algorithms in REM.

This chapter introduces the electrification planning problem from the techno-economic perspective. We also present the motivation of this thesis, the problem that the thesis is addressing, and the structure of the thesis.

Part of this chapter has been published in the following papers:

Ciller, P., Lumbreras, S., 2020. Electricity for all: The contribution of large-scale planning tools to the energy-access problem. Renewable and Sustainable Energy Reviews 120, 109624. https://doi.org/10.1016/j.rser.2019.109624

Ciller, P., Ellman, D., Vergara, C., Gonzalez-Garcia, A., Lee, S.J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., Palacios, R., Stoner, R., de Cuadra, F., Perez-Arriaga, I., 2019. Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM). Proceedings of the IEEE 107, 1872–1905.

https://doi.org/10.1109/JPROC.2019.2922543

1.1. The path to Universal Access to energy

Access to electricity is a critical enabling factor of human development, as it allows improvements in different areas such as education and healthcare. The importance of energy access is acknowledged in the seventh Sustainable Development Goal (SDG7), which comprises several targets and indicators to measure progress. Target 7.1 aims at achieving universal access to forms of energy that are reliable, affordable, and modern by 2030. Substantial progress will be necessary to meet this deadline as there are around 840 million people without access to electricity, and projections show that 650 million people could lack access to electricity in 2030, being 585 million located in sub-Saharan Africa (International Energy Agency et al., 2019).

Nevertheless, there are reasons for optimism. The population without access to electricity has dropped since 1990, and the electrification rates have grown substantially in recent years. People with access to electricity raised from 83% to 89% between 2010 and 2017, and 153 million people achieved access to electricity annually between 2015 and 2017. Considerable progress was made in South Asia (especially in India and Bangladesh), where electricity access raised from 75% to 91% in the period 2010-2017 (International Energy Agency et al., 2019).

Regretfully, the electrification rates are staggering in sub-Saharan Africa, where 600 million people do not have access to electricity (International Energy Agency, 2019). Kenia and Ethiopia have made significant progress, but the countries with the lowest electrification rates are located in Africa (i.e., Burundi, Chad, Malawi and the Democratic Republic of Congo are the only countries with an electrification rate lower than 20% (World Bank, 2020)). The challenge is even more intricate because the population of Africa is expected to increase to 2.4 billion people by 2050 (World Population Review, 2020). Figure 1-1 shows the status of electricity access in 2017.

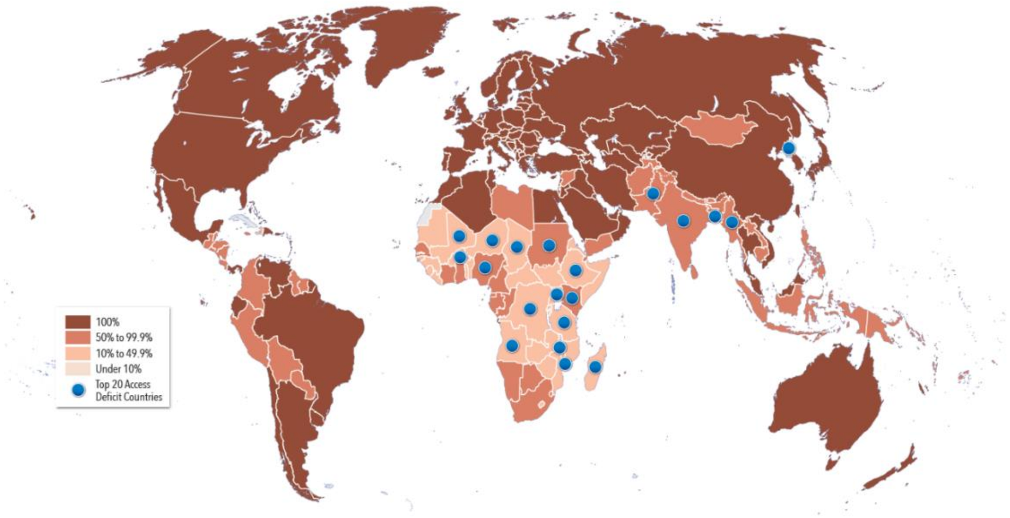


Figure 1-1: Population percentage with access to electricity by country in 2017. Source: (International Energy Agency et al., 2019).

There are also approximately 2.9 billion people that do not have access to clean cooking facilities, with 1.3 billion being located in India and China. The use of polluting elements to cook such as burning biomass is producing around 4 million deaths each year as well as contributing to deforestation and climate change. Projections estimate that 2.2 billion people will not have access to clean cooking technologies by 2030.

Finally, there is not a commonly agreed and well-established definition of electricity access. For instance, in some countries a village could be considered electrified even if only 10% of its consumers have access to electricity, which causes an erroneous perception of the magnitude of the universal energy access challenge. Moreover, simple metrics such as the number of people that have access to electricity or clean cooking facilities do not include information regarding relevant aspects such as availability of the service (i.e., the frequency and duration of interruptions) or quality of service (i.e., voltage stability), among other considerations.

1.2. The electrification planning process

Electrification planning of an underserved region is a complex task that involves social, financial, political, and regulatory aspects, and the need for electrification is mainly in rural areas but we cannot exclude peri-urban or urban zones. A plan backed-up by rigorous, reliable analysis has a higher chance of being successfully implemented. To that end, several computer models and planning methodologies have been developed by different institutions (Moner-Girona et al., 2018). These tools provide valuable assistance in finding the best techno-economic electrification plan for a region, which usually involves a combination of standalone systems, mini-grids, and grid extension designs.

Figure 1-2 shows the different phases of the electrification planning process, which is represented as a set of sequential steps. In practice, it turns out to be more of an iterative process that could last many years and needs to adapt to the new information available or a changing political or regulatory framework.

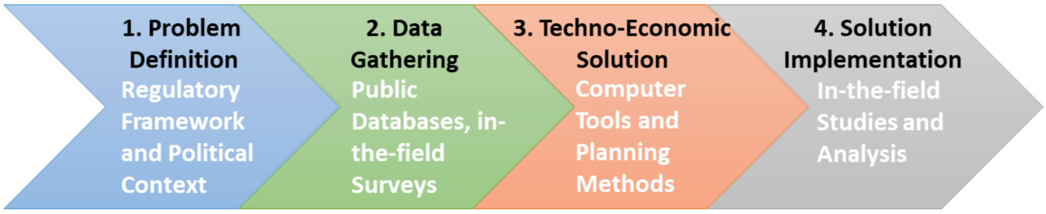


Figure 1-2: The electrification planning process. Source: adapted from (Ciller and Lumbreras, 2020).

The goal of the first step is to obtain a clear definition of the problem (i.e., what is the region to be electrified, what is the electrification target and the budget available), and the regulatory framework. There may be, for example, penalties for using polluting technologies or subsidies for the use of clean energy. Grid-connections could be mandatory for consumers located close to the power grid.

The second step aims at obtaining all the input data needed for the planning process. It typically includes catalogs of generation and network components, the location of the consumers and their demand profiles, the layout and techno-economic characteristics of the power grid, some techno-economic parameters such as discount rates and the number of years to consider for the project, the topography of the region, and information related to off-grid generation such as solar irradiance or wind speed.

The preferences of the consumers should be collected as part of the data-gathering process as they could constraint the electrification solution (Santos Pérez, 2015). For example, certain consumers could prefer to avoid diesel generation because it requires specialized maintenance, safety concerns, may lead to corruption, logistic problems in the supply of fuel, or cost, among other reasons.

Electrification planning tools usually require a considerable amount of data. The process of obtaining the necessary information is difficult in developing countries, where there is generally a scarcity of reliable data (Cader et al., 2018), but is critical for sound planning. For example, detailed demand profiles are difficult to obtain (Blodgett et al., 2017; Louie and Dauenhauer, 2016), but they have a substantial impact on the electrification plan.

The goal of the third phase is to obtain the best techno-economic solution for the problem that was defined in the first phase, and the input data obtained in the second phase. Computer tools play a prominent role in this phase. An electrification plan should determine the electrification mode for each consumer (i.e., standalone system, mini-grid, or grid extension) as well as generation designs for mini-grids and standalone systems, and network designs for mini-grid and the extensions of the power grid. It should also include a reasonable estimation of the total cost of the plan and electric designs for all systems in the solution.

The fourth phase aims at the successful implementation of the solution obtained. Local teams should analyze the solution obtained in the third phase, adapt it to the particular conditions of their locality, and prepare detailed engineering designs ready for construction.

The next section gives an overview of the third phase, which is the focus of this thesis.

1.3. The techno-economic planning problem

The techno-economic planning problem aims at determining the optimal combination of isolated systems, mini-grids, and grid extensions that electrify a given set of consumers and their corresponding demand profiles. The solution of the problem should also include generation designs for the isolated systems and mini-grids and network designs for the mini-grids and grid extensions.

The objective function that is considered most frequently is the total investment and operation cost (which should include the cost associated with lack of supply), so most models address this problem by performing a cost minimization. There are also relevant constraints that should not be neglected, such as the reliability of the systems, the total emissions, the total budget, or the existence of protected areas.

The optimization of grid extensions has been thoroughly studied in problems where it is the only electrification alternative (Georgilakis and Hatziargyriou, 2015). However, the inclusion of mini-grids and isolated systems as viable electrification solutions makes the electrification planning problem particularly challenging. The number of decision variables needed to characterize an electrification plan is very high. The most important ones are those defining:

- The grouping of consumers into clusters, and the determination of the electrification mode of each cluster (a combination of standalone systems, a mini-grid or an extension of the power grid).
- The generation design for each off-grid system (generation and storage technologies, capacities, and locations), and its dispatch strategy (i.e., which technologies are used to satisfy demand at each specific hour).
- The network design of each mini-grid and grid extension: voltage, type, and layout of distribution lines, as well as the location and the capacity of any new transformers and substations.
- Reinforcements or additional generation resources that might be needed upstream.
- The timeline of the project, which is usually divided into several phases.

A brute force approach that would guarantee the global optimum is provided in Algorithm 1-1[1]. However, the computational resources needed to apply Algorithm 1-1 successfully are far beyond what is presently available. Even the optimization of the network and layout of a single mini-grid is a process that requires significant computational resources, and there are generally thousands of possible mini-grids in an electrification planning project.

[1] *Algorithm 1-1 assumes a static approach (i.e., it optimizes the solution for a specific year, but it does not consider temporal implications regarding the implementation of the solution or the evolution of the systems). A dynamic approach is out of the question with the current computational resources, and all large-scale planning models approach the large-scale planning problem in a static way.*

Algorithm 1-1 Brute Force

```
procedure Brute Force
        for each possible clustering configuration C
            for each cluster c ∈ C
                calculate the cost of mini-grid
                calculate the cost of grid-extension
                calculate the cost of isolated systems
                select the least-cost option
            end for
            calculate the total cost of configuration C
            if C is the least-cost configuration evaluated so far then
                S = C
            end if
        end for
        return S
end procedure
```

The models and methods that address this problem necessarily need to apply smart heuristic algorithms or modeling simplifications. The higher is the level of modeling complexity, and the deeper is the exploration of the space of candidate solutions, the more realistic will be the planning recommendations; but they come at the expense of a high computational burden.

1.4. The role of computer models: the Reference Electrification Model (REM)

Several electrification planning models have been developed to deal with the techno-economic electrification planning problem. They exploit the advantages of Geographic Information Systems (GIS), providing instant access to public databases and the least-cost electrification solution for a large-scale region (Moner-Girona et al., 2018). Most of these tools provide first-pass information based on quick estimations. This is useful, but the estimations lack the level of detail needed for an implementable electrification plan. These tools apply certain simplifications, such as grouping the consumers into villages or cells beforehand (although the best grouping of consumers may be different) or estimating the network costs with purely geometric calculations that do not include constraints concerning electrical feasibility.

Although the data-gathering process is usually complicated and time-consuming, digital information is now more abundant than ever. GIS-based technology provides instant access to databases that contain geospatial information such as the location of roads, lakes, and the power grid. The High Resolution Settlement Layer (HRSL) estimates the human population with a resolution of approximately 30x30 meters (Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University, 2016), and digital metering could provide detailed consumption information. If the trend continues, the availability and accuracy of information will increase in the upcoming years.

Currently, it is possible (albeit usually difficult and expensive) to obtain information with a high level of resolution, and not considering it could have a substantial impact on the planning solution (International Energy Agency, 2018). A high level of modeling detail also leads to a more realistic outcome. The MIT-Comillas Universal Energy Access Laboratory (UEA Lab) has developed REM, which is state of the art in electrification planning, to take advantage of abundant digital information becoming quickly available.

In this thesis, we describe REM at its current state and present several developments that improved its performance. REM performs automated least-cost electrification design; it determines cost-optimal combinations of electrification modes for a given study region, including single building standalone systems, isolated grids with local electricity generation or mini-grids, and extensions of the existing distribution network. REM performs this task with a very high level of spatial granularity, producing detailed designs down to the individual consumer level. It prescribes network infrastructure layouts, local generation configurations, and storage options. These capabilities are intended to allow planners to make better-informed decisions about electrification modes, budget allocations, and bills of materials; ministries and regulators can get quantitative support for policy design; and developers can gain detailed insights into the potential for off-grid systems in a region. REM can also facilitate participatory planning approaches by providing references for least-cost electrification designs that can be evaluated by different stakeholders.

REM considers the specific demand profile of each consumer (incorporating residential, commercial, and industrial loads) and determines the least-cost grid/off-grid electrification plan by comparing a large number of clustering alternatives through a combination of heuristic optimization, mathematical algorithmic optimization, and simulation algorithms. These algorithms account for estimated yearly weather conditions and demand profiles, targets of quality of electricity supply, the reliability performance of local distribution lines, voltage and capacity constraints of lines and transformers, catalogs of power system components for grid-extension and off-grid systems, any existing limits or targets in the use of fossil fuels or renewables or carbon emissions, and implications of the topology of the terrain: forbidden areas, use of prescribed paths such as roads or streets, and extra costs due to factors like altitude or the slope of terrain being considered.

REM has been applied to multiple real electrification planning problems, ranging from cases representing small areas with hundreds of consumers to comprehensive analyses of entire countries with millions of them. Specifically, REM has been applied to develop master electrification plans in Rwanda, Uganda, Mozambique, and Indonesia.

REM results from the combined efforts of many former and current members of the UEA Lab. The UEA Lab is a joint research group of MIT, Cambridge, Massachusets, USA, and the Institute for Research in Technology (IIT), Madrid, Spain (IIT-Comillas, 2020).

The UEA Lab saw the potential of geospatial planning models and their application as decision support tools in the provision of universal access to energy. As a result, the UEA Lab developed several works related to this topic (González–García et al., 2014).

The goal of one such work was to develop a computer tool that performed large-scale automated electrification in developing countries, considering as viable solution the traditional extensions of the power grid and off-grid alternatives such as mini-grids and isolated systems.

The MIT team spearheaded the creation of such a model (with the collaboration of IIT), offering a master thesis to that end. The thesis was authored by Douglas Ellman and co-supervised by Claudio Vergara and Ignacio Pérez Arriaga. Initially, the group did not have a clear picture of how the finished tool would operate.

At that time, Claudio Vergara did a research stay of one year at IIT (from October 2013 to October 2014), and he learned about the Reference Network Model (RNM). RNM is a tool developed at IIT about 15 years ago that performs automated distribution network designs. The research team realized that using RNM as a routine that designs networks inside the abovementioned computer tool would enormously enhance the capabilities of the new model.

RNM designs the minimum-cost network that meets demand under quality-of-service specifications, using a user-provided catalog of equipment to specify distribution infrastructure down to the individual consumer level. RNM has been highly scrutinized: the Spanish distribution utilities have validated its results; RNM was then accepted by Spanish regulators as a decision-support model to determine appropriate remuneration figures for electric power distribution. RNM has been used for this same purpose in several other countries and many technical studies.

After a lot of hard work, Douglas Ellman presented his master thesis in June of 2015 (Ellman, 2015). The master thesis introduced the first prototype of the Reference Electrification Model (REM) and structured REM as a sequential process divided into several submodules. This high-level structure has stood the test of time, and it is still present in the current version of REM.

Douglas Ellman left the team a few months after completing his master thesis. After his departure, the IIT team performed the bulk of REM's development. The first prototype of REM provided inconsistent results, and the team realized that it needed a complete overhaul while maintaining its high-level logic and structure. Fernando de Cuadra and Pedro Ciller scrutinized the model's algorithms, conducting an in-depth analysis that shed light on the issues of this first version model. In the first phase, Pedro and Fernando complemented the in-depth review and precise detection of hidden software and logic errors with several enhancements and upgrades that turned the first prototype of REM into a robust and reliable tool. Pedro Ciller presented a master thesis, supervised by Fernando de Cuadra, in June of 2016 that describes several of these enhancements (Ciller Cutillas, 2016).

Once REM was in good shape to be applied in actual electrification planning projects, the development continued in a second phase with the addition of new capabilities that the practical application of the tool revealed to be essential; further detection and correction of performance issues was still necessary.

The Enel Foundation funded the first application of REM in a project, which involved two test cases with approximately 3,000 and 15,200 consumers in regions of Colombia and Kenia, respectively (IIT-Comillas, 2016). This project was developed during the second half of 2016.

In the years that followed, REM played a crucial role in many large-scale electrification projects. The model was applied in the South Service Territory in Uganda in 2017 to identify areas that are best suited for electrification with mini-grids (IIT-Comillas, 2017a). The UEA Lab developed the electrification master plans of Rwanda (IIT-Comillas, 2017b) and Mozambique (IIT-Comillas, 2019) between 2017 and 2019. Andrés González García was vital in the success of these projects, contributing to REM's development and its customization to the needs of each project.

In parallel with the projects, several MIT students contributed to different aspects in the development of REM, while the work on the core algorithms of the model continued being developed by Pedro Ciller and Fernando de Cuadra at IIT, with Sara Lumbreras joining later and Ignacio Pérez Arriaga as director of the overall UEA Lab team. At MIT, Vivian Li developed a particular REM configuration, named LREM (Local REM), which provides detailed electrification designs where all consumers are connected to the same mini-grid (Li, 2016). Turner Cotterman and Matthew Brusnahan applied LREM in India, Nigeria, and Rwanda (Cotterman, 2017; Brusnahan, 2018). Turner Cotterman also explored the impact of upstream reinforcements in REM's final electrification solution (Cotterman, 2017). Cailinn Drouin studied the incorporation of topography into REM (Drouin, 2018).

All of these MIT students have spent several weeks in Madrid in the framework of the collaboration established between IIT and MIT. As one more example of this joint effort, Olamide Oladeji (an MIT student) worked with Pedro Ciller and Fernando de Cuadra on a clustering algorithm that determines the consumers that REM should electrify with extensions of the power grid (Oladeji, 2018). This clustering algorithm is described in chapter 6 of this thesis.

Table 1-1 shows the primary affiliation and position that the abovementioned members of the UEA Lab held during the development of this thesis. This list is by no means exhaustive (i.e., the UEA Lab has more former and current members that have been involved in the development of REM), and it does not necessarily describe the current affiliation or position of the members of the team. Reference (MIT & IIT-Comillas Universal Energy Access Lab, 2019) presents a complete list of former and current members of the UEA Lab as well as an updated list of the projects, publications, and theses related to REM.

Ignacio Pérez Arriaga, Professor at IIT-Comillas and MIT	
IIT-Comillas	MIT
Fernando de Cuadra, Professor	Claudio Vergara, Postdoctoral Associate
Sara Lumbreras, Associate Professor	Douglas Ellman, Master Student
Pedro Ciller, Ph.D. Candidate	Olamide Oladeji, Master Student
Andrés González García, Ph.D. Candidate	Matthew Brusnahan, Master Student
	Vivial Li, Master Student
	Cailinn Drouin, Master Student
	Turner Cotterman, Master Student

Table 1-1: Affiliation and position held by several UEA Lab members during the development of this thesis.

Members of the UEA Lab recently created a company named WAYA, which provides consulting services regarding the application and commercialization of REM in projects that involve electrification master plans in developing countries, among other things. Several projects and research works that involve REM are currently ongoing (and hopefully will continue for many years).

The work presented in this thesis takes the reference (Ellman, 2015) as a starting point. The first part of this thesis focuses on improving the logic of the main algorithms of REM, upgrading the first prototype of REM until it exhibited a robust performance so that it could be applied in actual electrification projects. Some of the improvements are presented in reference (Ciller Cutillas, 2016). Chapter 3 presents an overview of REM and the list of enhancements implemented in the first prototype of REM. Chapter 4 delves deep into the optimization of off-grid generation designs for large-scale areas that REM performs.

The improvements implemented into the first prototype of REM were aligned with the algorithms presented in reference (Ellman, 2015) (i.e., the goal was to ensure that the existing algorithms of REM worked adequately). However, there was plenty of room to develop new algorithms that followed different approaches.

The second part of this thesis focuses on the development of novel algorithms that apply other strategies. One of them is the design of an algorithm that provides a quick estimation of the network cost of any potential mini-grid of a case study, which is presented in chapter 5. Two new clustering algorithms have also been developed, which are described in chapter 6.

1.5. The challenges of high-resolution modeling

Working with a high level of resolution has several drawbacks. Firstly, it is necessary to devote a substantial amount of time to gather and process the data. Secondly, the computation time needed to obtain the planning solution of a large-scale area is significant. There are several "dimensions" concerning the level of resolution that are important in electrification planning. This thesis focuses on the spatial and temporal aspects of high-resolution modeling.

The spatial resolution refers to the location of the loads and the existing power grid. We can distinguish among several levels of aggregation of consumers: the consumer itself, villages, settlements, or cells. Similarly, the power grid can be modeled considering the routes of transmission and distribution lines and the location of the transformers, although less detailed representations could be considered. Figure 1-3 shows different spatial resolutions regarding the location of consumers.

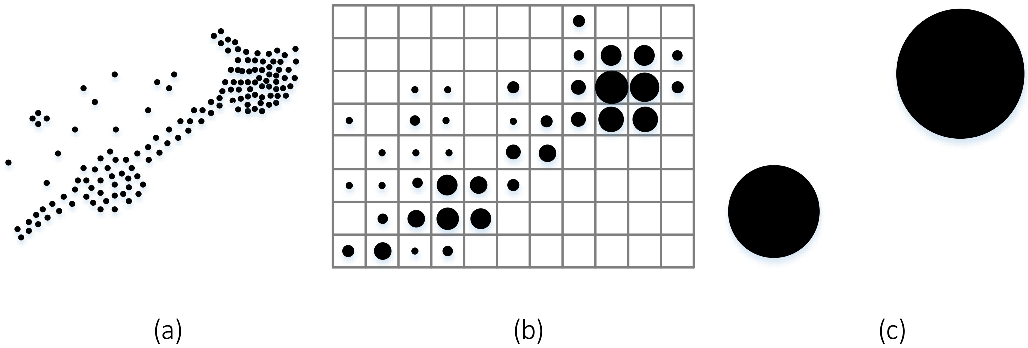


Figure 1-3: Example of the consumers with three different levels of spatial resolution: (a) the consumer itself, (b) cells, and (c) villages.

Some problems are inherent to a high level of spatial resolution. If a model operates with villages, then it will not be able to calculate the network layout that connects each consumer to the power grid or the generation site of the mini-grid. Moreover, it may not be necessary to group the consumers into viable mini-grids or grid extensions if a model operates at the village level because the villages by themselves could be considered as viable mini-grids or grid extensions.

The temporal resolution refers to the demand profiles and potential of renewable technologies (such as solar irradiance or wind speed). For example, we could consider hourly or daily demand profiles for consumers. The level of temporal resolution considered has an impact on the designs of off-grid systems (Stenzel et al., 2016). Figure 1-4 shows different temporal resolutions regarding demand profiles.

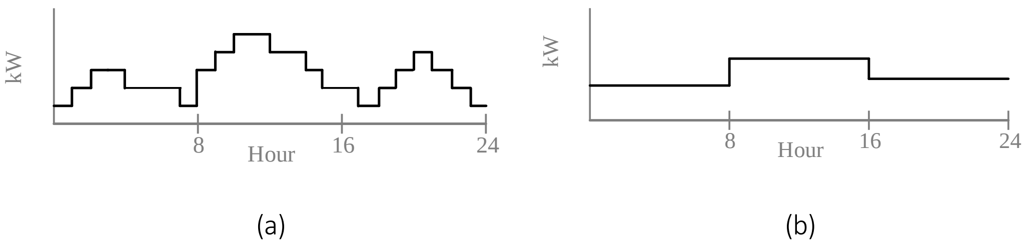


Figure 1-4: Example of a daily demand profile with two different levels of temporal resolution: (a) one hour, and (b) eight hours.

Some problems arise when the temporal level of resolution is high. For example, it is not possible to optimize the hourly dispatch of a mini-grid if its hourly demand is completely unknown (unless we assume a hypothesis such as that the demand is constant). In that case, only quick estimations based on rules of thumb or similar methods could be obtained.

However, if a model assigns an hourly demand profile to each consumer, then it is possible to implement an optimization method to size the generation design of the (potentially thousands of) mini-grids that appear in a large-scale case study, considering the hourly dispatch of the systems.

There are other levels of resolution that are relevant in electrification planning, although their impact is not analyzed in this thesis. One example is related to the number of load types (or consumer types) considered. We could consider that all the consumers of a case are residential, but that would neglect the impact of productive loads (such as hospitals or schools) in the electrification solution. Another example is related to the network and generation catalogs and the maximum number of elements that they can contain: a network catalog could be limited to one line for each voltage level, or it can contain several lines for each voltage level.

It is reasonable to wonder if modeling the problem with a high-resolution level is worth the effort considering the uncertainty related to the input parameters. The electrification solution can change drastically if we assume that all the consumers in a region have the same demand profile instead of considering that there are productive loads with different demand profiles (Ciller et al., 2019a). Similarly, the optimal solution for a village may include a smart combination of different electrification alternatives (AC individual systems, mini-grids, and grid extensions), but it is not possible to obtain such a combination unless we consider a spatial resolution more detailed than the village level shown in Figure 1-3. Therefore, it is worth to model the problem with a high-resolution level.

It is necessary to develop or apply techniques that balance accuracy and computation time to deal with problems inherent with a high level of spatiotemporal resolution. These "new" problems have motivated the research question that has fueled the development of this thesis:

How can large-scale electrification planning balance detailed modeling with feasible computational resources?

The notion of feasible computational resources implies that the algorithms should work effectively in cases of realistic size in an ordinary computer. Electrification master plans are often developed at the national level (Korkovelos et al., 2017; Rwanda Energy Group (REG), 2019), so we consider that a case of realistic size refers to a large-scale region such as an entire country, but not a continent. We also assume that an ordinary computer refers to a personal computer (i.e., an unspecialized computer intended for the individual user), but not clusters of computers or supercomputers.

A proper balance between detailed modeling and computational resources narrows the methods that we can consider. There are computationally intensive approaches that succeed in a village-scale problem but would run into an excessive computation time if extrapolated to regional planning. The techniques based on classical optimization such as mixed-integer linear programming cannot handle the electrification planning problem as a whole, although they can play an important role in solving specific subproblems.

This thesis contributes to the electrification planning problem and the development of REM with the improvement or the creation of several heuristic algorithms, which keep a proper balance between optimality and computation time.

We provide computation times several times among the thesis. All the computation times were obtained with an HP computer with 16 GB of RAM and the Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz 1.99GHz processor. The operative system of the computer is Windows 10 (64 bits).

1.6. Thesis structure

The first part of this thesis focuses on ensuring that the algorithms that were present in the first prototype of REM worked adequately, and it presents substantial contributions to the main algorithms of the model. The second part of this thesis focuses on the development of new algorithms. This thesis also reviews the methods and tools that address the large-scale electrification planning problem from the techno-economic angle.

Chapter 2 reviews the methods and tools that address the techno-economic planning problem at a large-scale. We present a conceptual formulation of the problem that lays the ground for classifying the tools according to their modeling complexity.

Chapter 3 presents an overview of the high-level structure of REM and its algorithms, including a description of the financial model of REM. We also describe several enhancements that were implemented in the first prototype of REM, increasing its robustness and optimality substantially. We also present some improvements that increase the capabilities of the model, such as the addition of solar kits as viable electrification solutions or handling multiple types of consumers.

Chapter 4 describes the method that REM uses to optimize the generation designs of mini-grids and standalone systems. We also analyze why a single-system tool or method (i.e., a tool or method that optimizes the generation design of an individual mini-grid) is not directly applicable to large-scale planning.

Chapter 5 introduces a method that quickly estimates the network cost of all potential mini-grids of a case study. This method outperforms the rules of thumb that most electrification tools apply.

Chapter 6 presents two novel clustering algorithms. The first algorithm (exhaustive clustering) calculates the optimal grouping of consumers into mini-grids, and it applies the method introduced in chapter 5 to estimate the network cost of the mini-grids.

The second algorithm (top-down clustering) determines which consumers are better electrified with extensions of the power grid and which should be electrified with off-grid systems. The top-down clustering was jointly developed in cooperation with an MIT student, Olamide Oladeji.

Chapter 7 presents the conclusions and future research lines that would expand the capabilities of REM or improve its performance.

1.7. The evolution of REM

We can distinguish among three different versions of REM. The initial version corresponds to the first prototype of REM, which was mainly developed at MIT and is described in (Ellman, 2015). This first prototype of REM is the starting point of this thesis, and its high-level structure and algorithms are described in section 3.1.

The intermediate version of REM includes all the upgrades and improvements presented in section 3.2 and chapter 4. These enhancements redesign critical parts of the algorithms of the first prototype of REM and add new functionalities such as the capability of handling several types of consumers and considering solar kits as a viable electrification solution.

The current REM version incorporates the new algorithms presented in chapters 5 and 6, and the clustering submodule is different as two additional clustering algorithms are included in REM. We explain the changes related to the clustering submodule in chapter 6.

Figure 1-5 shows the evolution of REM along with this thesis (the dates are approximate). The top-down clustering presented in chapter 6 was developed cooperatively among MIT and IIT. The remaining developments and contributions of this thesis were primarily developed at IIT.

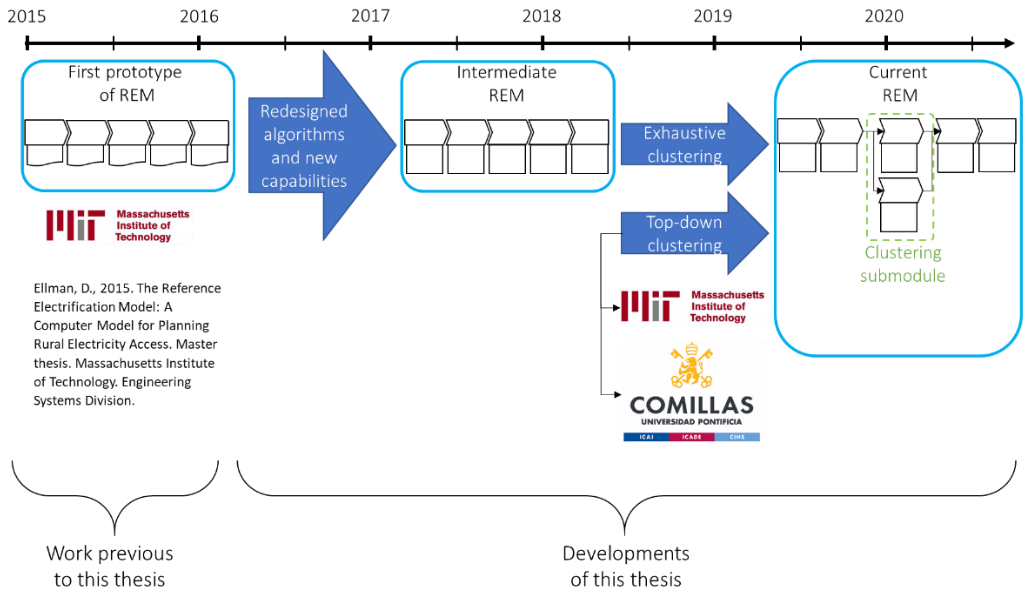


Figure 1-5: Evolution of REM.

“All models are wrong; some models are useful.”
George Box

2

A REVIEW OF ELECTRIFICATION PLANNING METHODS AND TOOLS

This chapter discusses the methods and tools that address the techno-economic planning problem in an underserved region. We present a conceptual formulation that sets the ground for the analysis of methods and tools, which are classified according to their modeling complexity.

The content of this chapter has been published in the following paper:

Ciller, P., Lumbreras, S., 2020. Electricity for all: The contribution of large-scale planning tools to the energy-access problem. Renewable and Sustainable Energy Reviews 120, 109624. https://doi.org/10.1016/j.rser.2019.109624

The rest of this chapter is structured as follows: section 2.1 presents several subproblems that are inherent to large-scale planning in an underserved region. Section 2.2 introduces a conceptual formulation of the techno-economic planning problem, which is used in section 2.3 to classify the methods and tools according to their modeling complexity. Section 2.4 presents future lines of development for electrification planning tools.

2.1. The techno-economic planning problem

This section briefly describes three subproblems that are part of the techno-economic planning problem, which was introduced in chapter 1, and whose goal is to determine the optimal combination of standalone systems, mini-grids, and grid extensions that electrify an area.

The three subproblems that comprise the techno-economic planning problem are (a) the generation sizing problem, (b) the network design problem, and (c) the clustering problem. Figure 2-1 shows the most frequent decisions involved in the three subproblems, as well as their objective functions.

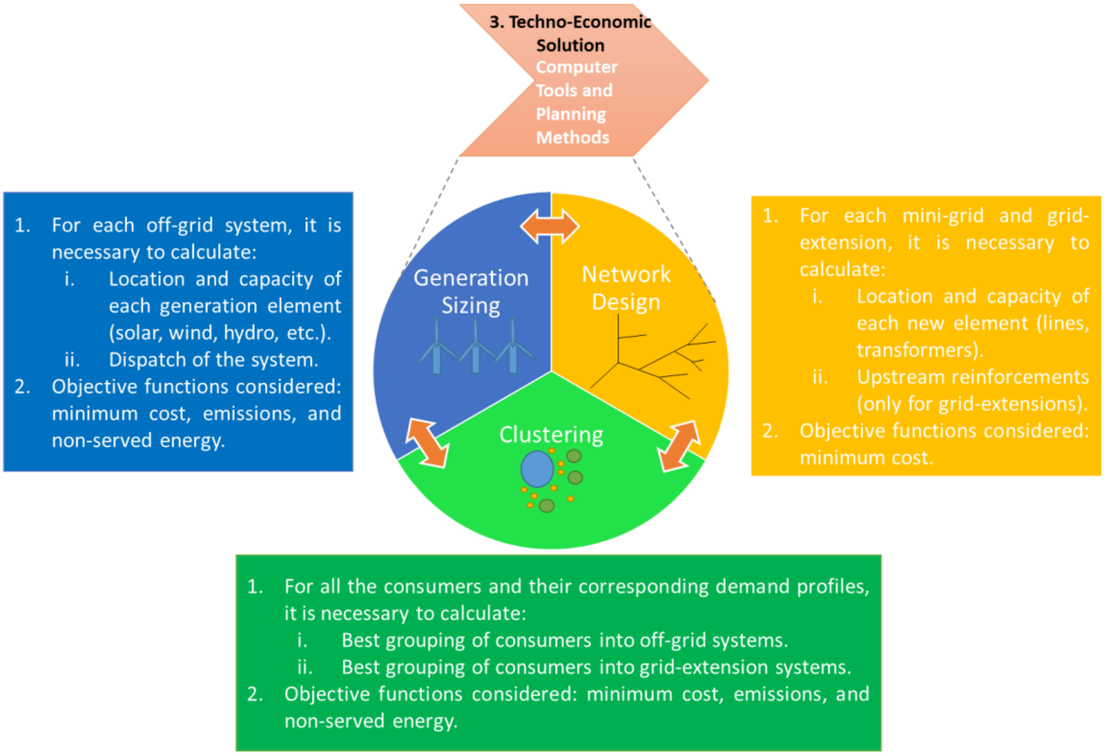


Figure 2-1: The techno-economic electrification planning subproblems.

The three subproblems are interconnected as their solutions are not independent of each other.

- Regarding mini-grids, the generation sizing and network design problems determine the generation and network designs of a mini-grid, but the clustering obtains the consumers that belong to the mini-grid first. Similarly, the clustering groups the consumers into mini-grids, but the optimal grouping of consumers into mini-grid depends on the generation and network costs, which are calculated solving the generation sizing and network design problems, respectively.
- Regarding grid extensions, the network design problem determines the layout and cost of an extension of the power grid. Still, the clustering identifies the consumers that belong to that extension of the power grid first. In a similar manner, the clustering groups the consumers into grid extensions, but it needs accurate estimations of the network costs of grid extensions, which are provided by the network design problem.

The generation sizing and the network design problems need to be solved a significant number of times in the large-scale electrification planning problem, so a direct application of methods that were designed to address a single instance of these problems may fail for computational reasons.

2.1.1. Clustering

The clustering problem consists in grouping the consumers into the best candidate systems (isolated, mini-grids, and grid extensions). There are few references related to clustering applications to the techno-economic electrification planning problem in the literature. Reference (Parreno Jr and Del Mundo, 2015) is such an example, although this process has manual steps and cannot be adapted easily to large-scale cases. Reference (Govender et al., 2001) is another example, but this reference does not consider the off-grid technologies and their costs in the process.

Most tools and methodologies require the planner to define the clusters beforehand and introduce them as inputs (villages, settlements, or cells), although a few tools such as GEOSIM, the Open Source Spatial Electrification Tool (OnSSET) and REM include a clustering algorithm. Figure 2-2 shows a clustering example obtained with REM, where consumers that belong to the same clusters are represented with the same color. The location of consumers often follows roads or rivers, which is reflected in their spatial representation.

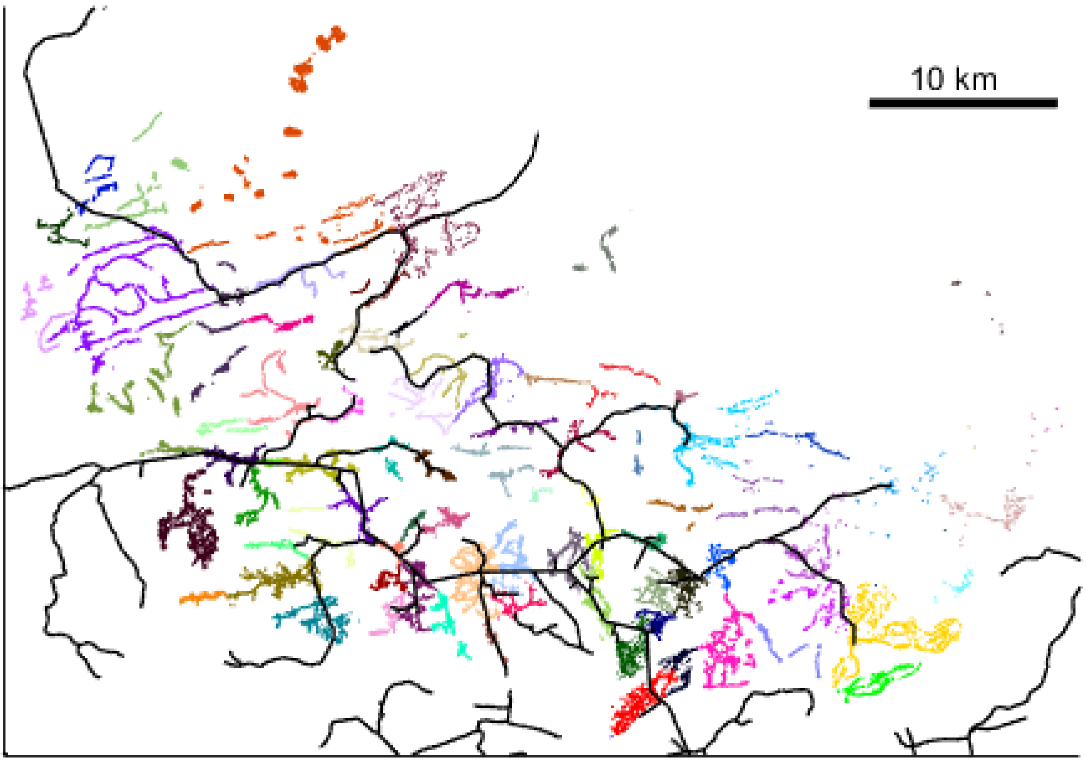


Figure 2-2: Clustering of consumers obtained with REM. The already-existing distribution network is represented with black lines.

2.1.2. Generation sizing

The off-grid generation design or generation sizing problem consists in calculating the best generation design for a given off-grid system and its corresponding demand. The inputs of the tools that aim at this problem usually include demand profiles and a catalog of components that includes the techno-economic parameters of the generation technologies. There are reviews related to the methods used to sizing hybrid energy systems (Luna-Rubio et al., 2012) as well as their configurations and control methods (Upadhyay and Sharma, 2014). The main software tools that aim at sizing hybrid energy systems are discussed in (Sinha and Chandel, 2014). Some of the most widely-known mini-grid generation design tools are the Hybrid Optimization Model for Multiple Energy Resources (HOMER) (Lambert et al., 2006) and the Distributed Energy Resources - Customer Adoption Model (DER-CAM) (Lawrence Berkeley National Laboratory, 2017).

The generation sizing problem also aims at determining the optimal dispatch of the off-grid system (i.e., the real-time operation so that the operating costs are minimized). The dispatch can be obtained with classical optimization methods, although most tools apply heuristic strategies that require a reduced computational time at the expense of losing optimality (Neves and Silva, 2015; Alramlawi et al., 2019).

Figure 2-3 shows the daily dispatch that REM provides for a mini-grid applying a load following strategy. The load following strategy is a heuristic dispatch where the resources used to meet the demand always follow the same order: solar, battery, and diesel. If there is solar energy available after meeting the demand, then it will be used to charge the battery. The diesel generator is not used to charge the battery in the load following strategy. REM may not use the diesel generator to meet the demand if the marginal cost of diesel is higher than the Cost of Non-Served Energy (CNSE).

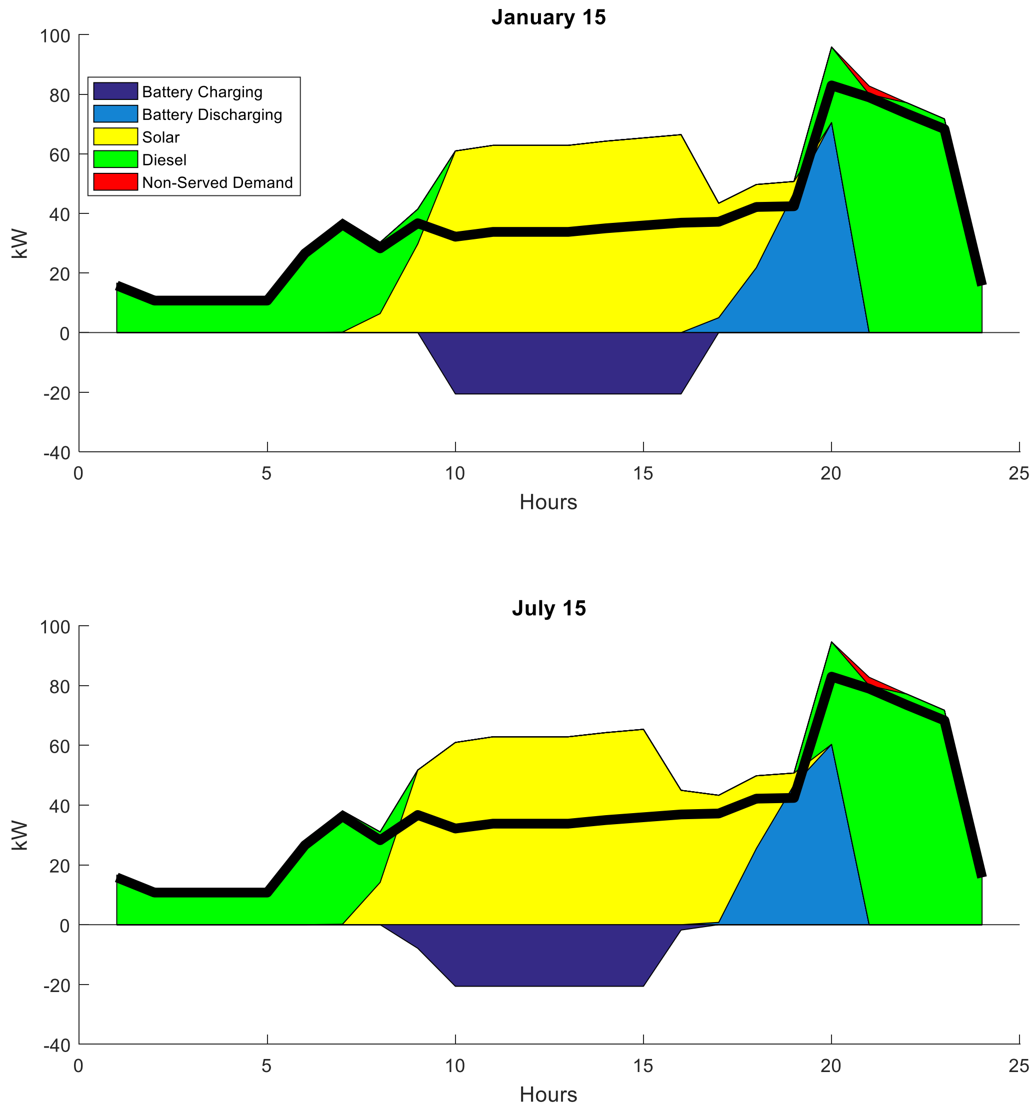


Figure 2-3: REM's daily sample dispatch of a mini-grid. The black line represents its total demand.

The dispatch shown in Figure 2-3[2] uses solar energy to meet the demand and charge the batteries during the day. The energy stored in the batteries is used to meet the demand in the evening, and the diesel generator meets the demand at night. However, the diesel generators make noise when they operate, which is particularly unpleasant during the night. A cycle-charging dispatch strategy could be implemented to avoid using the diesel generator during the night.

The circle-charging dispatch strategy operates the diesel generator at its maximum capacity when

[2] *REM operates with a temporal resolution of one hour, and it interpolates among the hourly values of the demand and mini-grid components when it plots a dispatch. The dispatch can give the impression that the battery is being charged and discharged simultaneously, although this does not happen.*

The solar dispatch only includes the amount of solar energy used to charge the battery, which is limited by the state of charge and the speed of charge of the battery. The dispatch does not show the total available solar energy.

it is used, and it charges the battery with the energy left after meeting the demand. With the cycle-charging strategy, the diesel generator would operate at its maximum capacity to meet the demand and charge the batteries during the evening, and the batteries would meet the demand during the night.

2.1.3. **Network design**

The network design problem aims at obtaining the best power distribution network for a mini-grid or a grid extension. The tools that address this problem require the location and the demand of the consumers, the location of the existing power grid (only for grid-extension designs), and a catalog of network components (lines and transformers).

Village Power Optimization model for Renewables (ViPOR) is a network design tool that obtains the distribution network of a single mini-grid applying a simulated annealing algorithm (Lambert and Hittle, 2000). However, ViPOR is currently unsupported and it does not incorporate electric constraints in the calculation of network designs.

A review of the methodologies and tools used to calculate the power distribution network has been recently published (Georgilakis and Hatziargyriou, 2015). One of the most advanced existing network design tools is the RNM (Mateo Domingo et al., 2011), which is used in REM to calculate network designs for grid extensions and mini-grids. Figure 2-4 shows the projection onto Google Earth Pro of a grid-extension design calculated with RNM, which calculates the layout of the Medium Voltage (MV) and Low Voltage (LV) lines, and the location of the transformers.

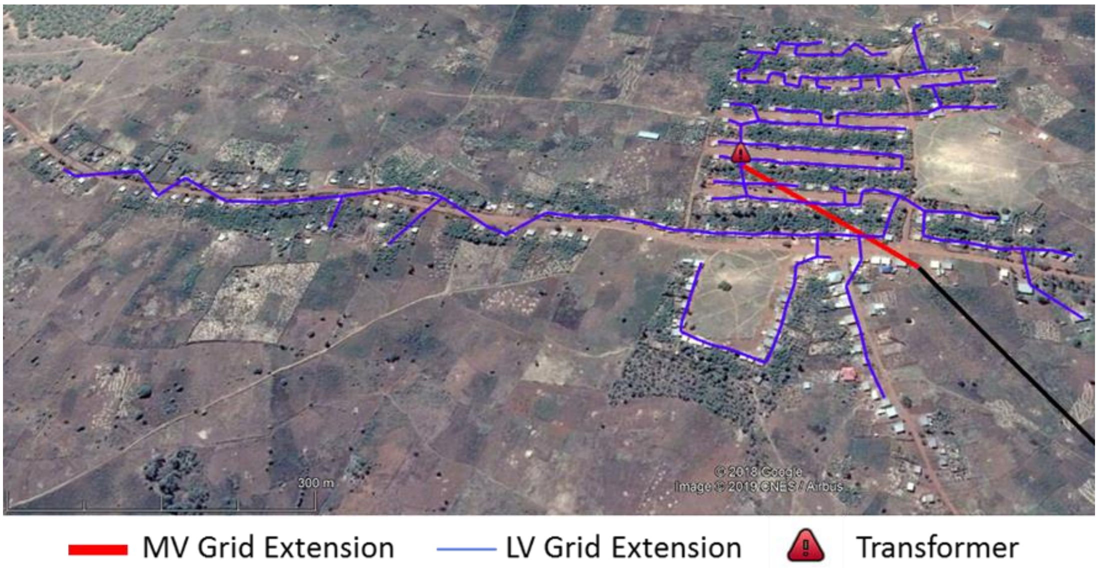


Figure 2-4: REM's network design for a grid extension. The already-existing distribution network is represented with black lines.

This review focuses on tools and methodologies that address the large-scale techno-economic electrification planning problem as a whole. This implies that tools such as HOMER and ViPOR are not included in the review since they only deal with specific parts of the problem, and they operate at the small-scale level (i.e., an individual village or settlement).

There are several reasons to include only large-scale tools and methods in the review. In the first place, some reviews already deal with the methods (Luna-Rubio et al., 2012) and tools (Sinha and Chandel, 2014) at a small-scale level. In the second place, the nature of the problem is different since, in most cases, there is no need to cluster the consumers when a village or a settlement is electrified because all the village or settlement is electrified altogether as a single system. Finally, the computational resources needed to solve a small-scale problem are lower, allowing the use of classical optimization techniques or computationally intensive procedures that would fail in a large-scale problem.

The next section introduces a conceptual formulation of the techno-economic electrification planning problem, which is used in section 2.3 to classify the tools and methodologies that address this problem at a large-scale scope.

2.2. A conceptual formulation

The existing mathematical formulations available describe the electrification planning problem partially as they focus on a single village or community (Ferrer-Martí et al., 2013; Triadó-Aymerich et al., 2016), or include a reduced level of detail (Zeyringer et al., 2015). A general formulation –which was missing in the literature– is very valuable to build comparisons and focus discussions.

Large-scale tools deal with an extensive area such as a region or a country and consider only a purely economic criterion for computational reasons. On a small scale, it is possible to consider factors that go beyond cost, such as environmental impact, social impact (measured by the amount of Non-Served Energy (NSE) or job creation), and regulatory conditions (such as subsidies or tax reduction for certain types of technologies).

In addition, it is vital for the success of a project to consider stakeholder preferences. Often, this is undertaken by generating a small list of non-dominated, feasible solutions. Then, the preferred solution is identified as the candidate solution with the highest acceptance among stakeholders (Santos Pérez, 2015). However, section 2.3 shows that no electrification planning tool or methodology applies multicriteria optimization techniques.

The formulation proposed has been developed to serve as an illustration for the general statement of the problem and its most important features. The following simplifications, accurate but useful, have been adopted:

1. The formulation considers only generation costs and disregards the cost of other elements such as inverters and charge controllers, which are comparatively much cheaper (Bhattacharyya, 2015; Azimoh et al., 2017).
2. Only one line can be installed between each pair of nodes. Parallel lines are not considered in most projects, as demand profiles are usually low compared to line capacities, so that parallel lines are not usually useful.
3. The cost of losses is calculated in a simplified way.
4. The formulation assumes that upstream impacts are negligible. This includes changes in the system-wide dispatch or needs for reinforcement in the transmission network.

5. Standard network constraints such as voltage drops and reliability constraints are not considered, given that they increase the complexity considerably. They can be included in subsequent analyses.
6. The capacity of the generation elements of the mini-grid is considered continuous, and linear piecewise functions approximate the investment cost of these elements.
7. Generators installed in a mini-grid can be located only in consumers that belong to the mini-grid.
8. The efficiency of the diesel generators is considered to be constant.
9. The existing distribution network is discretized into a finite number of candidate connection points, which are considered for grid-extension designs.
10. All networks are three-phase and balanced.
11. The traditional DC power flow assumptions hold: the voltage of each node is one (in p.u.), the difference of voltage angles between nodes is small, and the reactance of each line is much larger than its resistance (Qi et al., 2012). The last assumption may not hold in MV and LV distribution networks, but the error introduced may be acceptable provided that the reactance of the lines is high enough (Purchala et al., 2005). If this is not the case, then non-linear approximations can provide high accuracy (Baradar and Hesamzadeh, 2015).

Regional planning methods apply similar simplifications to deal with the techno-economic planning problem at a large-scale level (Zeyringer et al., 2015; Abdul-Salam and Phimister, 2016a).

Although the electrification planning problem has a dynamic and stochastic nature, this section presents a static formulation, which is consistent with the approach found in the full list of the literature surveyed. For the time being, no large-scale tool or methodology can handle multiple time horizons or include robust optimization methods to effectively cope with uncertainties, which are frequent in developing countries due to the lack of reliable data.

2.2.1. **Sets**

N	supply and demand nodes;
$I \in N$	demand nodes (households, schools, health centers, etc.);
$S \in N$	supply nodes (already existing network candidate connection points);
Y	off-grid elements (generation technologies and batteries) and grid extension components (MV/LV substations);
$G \in Y$	off-grid elements (generation technologies and batteries) and on-grid generation;
$Z \in Y$	other elements related to the grid-extension designs (namely MV/LV substations);
$O \in G$	components of the off-grid system that are not generation technologies (namely batteries);
$T \in G$	generation elements, both off-grid (namely diesel, solar and wind) and on-grid (where the category used will be grid extension);
C	available elements of the catalog for each generation technology (including lines for grid extension);

H	hours for the period considered (usually one year: 8760 hours);
V	voltage levels (namely LV and MV).

2.2.2. **Parameters**

$annualfactor_y$	fraction of the cost of the y-th element that is amortized in one fiscal year. It depends on its lifetime, the discount rate considered and its total cost;
cap_{yc}	capacity of the c-th element of the catalog associated with the element y [kW];
$cinv_{yc}$	investment cost required to acquire an element y of capacity c [$];
$cfuel$	market price for diesel [$/liter];
cl_{ijvc}	investment cost of the c-th catalog line of voltage v between nodes i and j [$];
$cO\&M_y$	annual operation and maintenance cost of element y expressed as a fraction of investment cost [p.u.];
$demand_{ih}$	demand of node i at hour h [kWh];
$emissions$	metric tonnes of equivalent CO_2 diesel emissions per liter [tCO_2eq /liter];
$\varepsilon diesel$	average efficiency of diesel generators [liters/kW];
$\varepsilon network$	network efficiency [p.u.]. This parameter is used to account for the network losses;
$\varepsilon charge$	efficiency of battery charge [p.u.];
$\varepsilon discharge$	efficiency of the battery discharge [p.u.];
M	big-M parameter (sufficiently large number used in disjunctive equations);
$\overline{capacity}_y$	maximum capacity of generation, batteries, and substations installed [kW];
$\underline{capacity}_y$	minimum capacity of generation, batteries, and substations installed [kW]. These parameters are usually set to 0, which corresponds to no generation;
$\overline{charge}_i$	maximum hourly charge for the battery located in node i [kWh];
$\underline{charge}_i$	minimum hourly charge for the battery located in node i [kWh];
$\overline{soc}$	maximum allowed state of charge of the battery [p.u.]. This parameter is usually set to 1, which corresponds to a fully charged battery;
$\underline{soc}$	minimum allowed state of charge of the battery [p.u.];
$\overline{profile}_{iht}$	maximum generation profile in node i at hour h for generation technology t [kWh];
$\underline{profile}_{iht}$	minimum generation profile in node i at hour h for generation technology t [kWh];
$plim_{ijvc}$	capacity limit of the c-th catalog line of voltage v between nodes i, j[kW];
$sbase$	base power [kW];
x_{ijvc}	reactance of the c-th catalog line of voltage v between nodes i, j [p.u.].

2.2.3. **Variables**

CAP_{iy} positive variable that accounts for the capacity of the y-th element installed at node i [kW];

$CDEC_{iyc}$ positive variable used to express the capacity cost $CAP_{i,y}$ with a piecewise linear function;

$BINC_{iyc}$ binary variable that takes the value 1 if $CDEC_{iyc}$ (and therefore CAP_{iy}) is between $cap_{y(c-1)}$ and cap_{yc};

$\mathrm{B_{iz}}$ binary variable that takes the value 1 if an only if there is a z element (namely an MV/LV substation) located in node i;

$CHARGE_{ih}$ positive variable that accounts for how much the battery is charged in node i at hour h;

$DISCHARGE_{ih}$ positive variable that accounts for how much the battery is discharged in node i at hour h;

$CINV_{iy}$ positive variable that accounts for the investment cost associated with element y located at node i;

$TOTCINV_y$ positive variable related to the total investment cost associated with element y;

$CFUEL_i$ positive variable related to the fuel cost associated with the diesel generator located at node i;

$\Xi_{ih(t=genset)}$ binary variable that describes the commitment of a diesel generator (it takes the value 1 if the generator at node i is on at hour h, and 0 otherwise);

$TECHGEN_{ihg}$ positive variable related to the real power generated at node i by technology g at hour h [kWh];

GEN_{ih} positive variable related to the real power generated at node i in hour h, either with a mini-grid generator or with a grid extension of the distribution network [kWh];

P_{ijhvc} free variable related to the directed power flow that goes through node i to node j at hour h through the c-th catalog line of voltage v [kWh];

P^{+}_{ij} positive variable that takes the maximum value of P_{ijhvc} for each hour h, voltage v and catalog element c if P_{ijhvc} is positive. Otherwise, this variable takes the value 0;

SOC_{ih} positive variable related to the state of charge of the battery located at node i in hour h [p.u.];

Θ_{ih} phase angle at node i at hour h [radians];

XL_{ijvc} binary variable that takes the value 1 if nodes i, j are connected with the c-th catalog line of voltage v.

2.2.4. **Equations**

The equations are classified into several groups, which are described in Table 2-1.

Group	Description
Distribution-network equations	They determine the layout and capacities of lines and transformers included in distribution networks of mini-grid and grid extensions. They include geometric considerations (equations 2-1-2-5) and electric criteria (equations 2-6-2-13) that are frequently considered in the network design problem shown in Figure 2-1.
Generation equations	They determine the capacities of the off-grid equipment (equations 2-14- 2-17) and the hourly dispatch (equations 2-18-2-24) of each mini-grid and isolated system. They are related to the generation sizing problem shown in Figure 2-1.
Cost equations	They model the way that costs are computed (equations 2-25-2-31) for each isolated system, mini-grid, and grid extension design. The costs included account for the investment and operation cost of the elements of the network and generation equipment.

Table 2-1: Formulation equations.

2.2.4.1. Distribution-network equations

Demand nodes can host generation technologies different from grid extension:

$$CAP_{ig} = 0 \quad \forall i \in I, g = \{grid\ extension\} \in G \tag{2-1}$$

Two nodes can only be connected through one line:

$$\sum_{v,c \in VC} XL_{mnvc} \leq 1 \quad \forall (m,n) \in N^2 \tag{2-2}$$

If a demand node does not have an MV/LV substation installed, then it cannot belong to the MV distribution network:

$$XL_{mnvc} \leq B_{mz} \quad \forall (m,n,c,z) \in N^2CZ, v = \{MV\} \in V \tag{2-3}$$

$$XL_{mnvc} \leq B_{nz} \quad \forall (m,n,c,z) \in N^2CZ, v = \{MV\} \in V \tag{2-4}$$

If a supply node does not have an MV/LV substation, then it cannot be connected to a low-voltage line:

$$\sum_{c \in C} XL_{snvc} \leq B_{nz} \quad \forall (s,n) \in SN, v = \{MV\} \in V \tag{2-5}$$

The balance of generation minus demand at each node and the power used to charge batteries is equal to the power transmitted to the nodes that are connected to it (Kirchhoff's First Law):

$$\frac{GEN_{ih} + DISCHARGE_{ih} - CHARGE_{ih} - demand_{ih} + NSE_{ih}}{\varepsilon network} = \sum_{j,h,v,c \in NHVC} P_{ijhvc} \ \forall\ i,h \in NH \tag{2-6}$$

Power flow equations must be satisfied (Kirchhoff's Second Law):

$$-plim_{ijvc} \cdot \left(1 - XL_{ijv}\right) \leq P_{ijhvc} - \frac{\Theta_{ih} - \Theta_{jh}}{x_{ijvc}} \cdot sbase \leq plim_{ijvc} \cdot \left(1 - XL_{ijv}\right) \forall (i,j,h,v,c) \in N^2HVC \tag{2-7}$$

$$\Theta_{ih} = 0 \quad \forall h \in H, i = \{1\} \in N \tag{2-8}$$

Equation 2-7 provides an estimation of the power flow based on the traditional DC power flow assumptions, and therefore the resistance of each line is not considered.

Power flows are bounded by capacity:

$$-plim_{ijvc} \cdot XL_{ijvc} \leq P_{ijhvc} \leq plim_{ijvc} \cdot XL_{ijvc} \;\; \forall (i,j,v,c) \in N^2VC \qquad \text{2-9}$$

The variable $\mathrm{B_{iz}}$ takes the value 1 if and only if CAP_{iz} is strictly greater than zero:

$$\frac{CAP_{nz}}{M} \leq B_{nz} \leq M \cdot CAP_{nz} \;\; \forall (n,z) \in NZ \qquad \text{2-10}$$

The capacity of each MV/LV substation (if installed) is greater than or equal to the maximum power flow that flows from it:

$$CAP_{iz} \geq \sum_{j \in N} P_{ij}^{+} - M \cdot (1 - B_{iz}) \;\; \forall (i,z) \in NZ \qquad \text{2-11}$$

$$P_{ij}^{+} \geq P_{ijhvc} \;\; \forall (i,j,h,v,c) \in N^2HVZ \qquad \text{2-12}$$

Demand at each node bounds NSE:

$$NSE_{ih} \leq demand_{ih} \;\; \forall (i,h) \in IH \qquad \text{2-13}$$

2.2.4.2. Generation equations

The total generated power is the sum of the power generated by technology:

$$GEN_{ih} = \sum_{t \in T} TECHGEN_{iht} \;\; \forall (i,h) \in IH \qquad \text{2-14}$$

Limits for the capacity of the generation elements and the battery:

$$\underline{capacity}_g \leq CAP_{ig} \leq \overline{capacity}_g \forall (i,g) \in IG \qquad \text{2-15}$$

The maximum generation allowed for each technology depends on the capacity installed and a generation profile that depends on technical parameters of the generation technology and resource availability (such as solar irradiance or wind speed, among others). In the cases of MV/LV substations and diesel generators, the generation profile may be constant:

$$TECHGEN_{iht} \leq CAP_{it} \cdot \overline{profile}_{iht} \;\; \forall (i,h,t) \in IHT \qquad \text{2-16}$$

The minimum generation allowed for each generation technology depends on the capacity installed and a minimum generation profile. This bound might not be useful if there is no actual minimum generation for a capacity:

$$CAP_{it} \cdot \underline{profile}_{iht} \leq TECHGEN_{iht} \;\; \forall (i,h,t) \in IH(T \backslash \{genset\}) \qquad \text{2-17}$$

The state of charge of the battery at any hour depends on the state of charge of the previous hour and on how much the battery has been charged or discharged in the previous hour:

$$SOC_{ih} = SOC_{i(h-1)} + \frac{\varepsilon charge \cdot CHARGE_{i(h-1)} + \frac{DISCHARGE_{i(h-1)}}{\varepsilon discharge}}{maxcapacity_o} \;\; \forall (i,h) \in I(H \backslash \{1\}), o = \{battery\} \in O \qquad \text{2-18}$$

Constraints on the state of charge of the batteries:

$$\underline{soc} \leq SOC_{ih} \leq \overline{soc} \; \forall (i,h) \in IH \qquad \text{2-19}$$

Limits for the speed of charge/discharge:

$$\underline{charge}_i \cdot CAP_{ig} \; \leq CHARGE_{ih} \leq \overline{charge}_i \cdot CAP_{ig} \; \forall (i,h) \in IH, g = \{battery\} \in G \qquad \text{2-20}$$

$$\underline{discharge}_i \cdot CAP_{ig} \; \leq DISCHARGE_{ih} \leq \overline{discharge}_i \cdot CAP_{ig} \;\; \forall (i,h) \in IH, g = \{battery\} \in G \qquad \text{2-21}$$

If a diesel generator is installed, there is a minimum load restriction that must be met if the generator is on:

$$TECHGEN_{iht} \geq CAP_{it} \cdot \underline{profile}_{iht} - M(1 - \Xi_{iht}) \; \forall (i,h) \in IH, t = \{genset\} \in T \qquad \text{2-22}$$

$$TECHGEN_{iht} \leq M \cdot \Xi_{iht} \; \forall (i,h) \in IH, t = \{genset\} \in T \qquad \text{2-23}$$

The diesel generator produces emissions when it operates:

$$DE_{ih} = GEN_{tih} \cdot \varepsilon diesel \; \cdot emissions \;\; \forall (i,h) \in IH, t = \{genset\} \in T \qquad \text{2-24}$$

The operation of a diesel generator produces several polluting elements such as carbon oxides (COx), nitrogen oxides (NOx), and sulfur oxides (SOx) (Sothea and Kim Oanh, 2019). The term that accounts for the diesel emissions in the formulation is the total annual equivalent CO_2 emissions. The equivalent CO_2 emissions of a gas is calculated by determining the amount of CO_2 that produces the same global warming potential as the gas (myclimate, 2020). This metric allows us to measure and compare the climate effects of several gases, and it has been used in the literature when optimizing the generation of a mini-grid (Dufo-López et al., 2011).

2.2.4.3. Cost equations

As shown in Figure 2-5, the investment cost related to generation technologies and batteries is interpolated using linear piecewise functions. These piecewise linear functions should reflect the economies of scale in generation technologies.

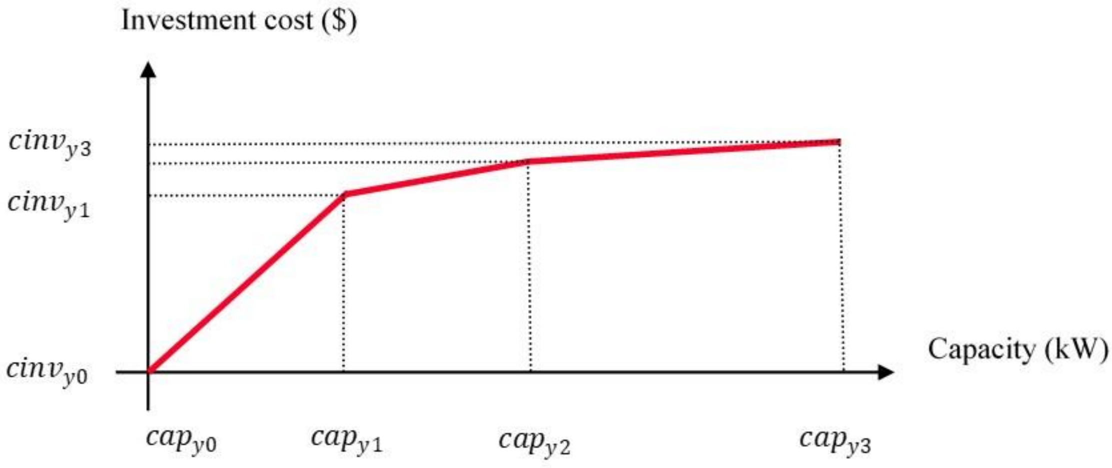


Figure 2-5: Linear interpolation of the investment costs.

Where $cinv_{yc}$ is the investment cost of the y-th generation technology that has capacity c (which is the element cap_{yc}). The formulation uses a binary variable $BINC_{iyc}$ that takes the value 1 if y-th generation technology installed at node i has a capacity that lies between $cap_{y(c-1)}$ and cap_{yc}. The formulation also uses a positive variable $CDEC_{iyc}$ that is equal to the capacity of the y-th generation technology installed at node i if that capacity lies between $cap_{y(c-1)}$ and cap_{yc}, and takes the value zero in the remaining cases. The remaining of this section describes the cost-related equations.

Exactly one of the cost segments is active:

$$\sum_{c \in C} BINC_{iyc} = 1 \ \forall (i, g) \in I(Y \backslash \{grid\ extension\}) \qquad \text{2-25}$$

Only the capacity that is related to the active cost segment is effective:

$$cinv_{y(c-1)} \cdot BINC_{iyc} \leq CDEC_{iyc} \leq cinv_{yc} \cdot BINC_{iyc} \ \forall (i, y, c) \in I(Y \backslash \{grid\ extension\})C \qquad \text{2-26}$$

The capacity of an element is equal to its active capacity:

$$CAP_{iy} = \sum_{c \in C} CDEC_{iyc} \ \forall (i, m) \in I(Y \backslash \{grid\ extension\}) \qquad \text{2-27}$$

The investment cost of an element is calculated using linear interpolation between the nearest two elements of the catalog:

$$CINV_{iy} = \sum_{c \in C} (cinv_{y(c-1)} \cdot BINC_{iyc} + (CDEC_{iyc} - cinv_{y(c-1)} \cdot BINC_{iyc}) \cdot \frac{cinv_{yc} - cinv_{y(c-1)}}{cap_{yc} - cap_{y(c-1)}}) \ \forall (i, y) \in I(Y \backslash \{grid\ extension\}) \qquad \text{2-28}$$

The total investment cost associated with an element is the sum of the corresponding investment costs through all nodes:

$$TOTCINV_y = \sum_{i \in I} CINV_{iy} \ \forall y \in Y \backslash \{grid\ extension\} \qquad \text{2-29}$$

The investment cost of the lines is calculated using the binary variables related to their connections:

$$CINV_y = \sum_{m,n,v,c\in N^2VC} \left(XL_{ijvc} \cdot Cl_{ijvc}\right),\ y = \{grid\ extension\} \in G \tag{2-30}$$

The fuel cost depends on the diesel cost, the efficiency of the generator, and how much power is generated:

$$CFUEL_i = \left(\sum_{h\in H} TECHGEN_{iht}\right) \cdot cfuel \cdot \varepsilon diesel\ \forall i \in I, t = \{genset\} \in T \tag{2-31}$$

2.2.4.4. Objective functions

This formulation considers three terms in the objective function, which represent the economic, social, and environmental factors associated with the electrification solutions. The economic term in the objective function minimizes the investment and running costs of the electrification plan.

The useful life of elements can vary significantly (i.e., the useful life of a line can be over 20 years, whereas the useful life of a battery is usually around 5). To account for the different useful lives, the economic parameter that is minimized in this formulation is the total annualized cost, which is composed of investment, operation and maintenance, and fuel costs (which result from the use of diesel generators).

$$f_1 = \sum_{y\in Y} \frac{CINV_y}{annualfactor_y} + \sum_{y\in Y} CINV_y \cdot cO\&M_y + \sum_{i\in I} CFUEL_i \tag{2-32}$$

The social term weights in NSE, which accounts for the loss of welfare caused by leaving some demand unserved:

$$f_2 = \sum_{n,h\in NH} NSE_{nh} \tag{2-33}$$

The environmental term incorporates the effects of emissions:

$$f_3 = \sum_{n,h\in NH} DE_{nh} \tag{2-34}$$

It is clear that there is a degree of conflict among the three objective functions, and analyzing the trade-offs among them would involve the application of multicriteria optimization techniques that require solving several instances of a mono-objective problem (Ballestero and Romero, 1998). This is not affordable in the case of large-scale planning, so tools and methods generally consider a cost-related objective function that may include economic penalties for non-served energy or emissions.

This conceptual formulation of the problem can serve as a common ground to classify tools and methodologies for electrification planning. This classification is presented in the next section.

2.3. Classification of tools and methodologies

In this section, the tools and methodologies are classified according to their modeling complexity. They range from *first-pass* tools, which offer information for a pre-feasibility analysis with a short computation time, to the *detailed analysis* tools, which provide a solution that includes detailed generation and network designs for all the grid extensions and mini-grids.

Figure 2-6 provides an overview of the complexity levels. The boundaries among levels are not clear-cut, so there could be cases where a tool or methodology could arguably belong to one or another level. A high modeling complexity comes at the expense of high computation time.

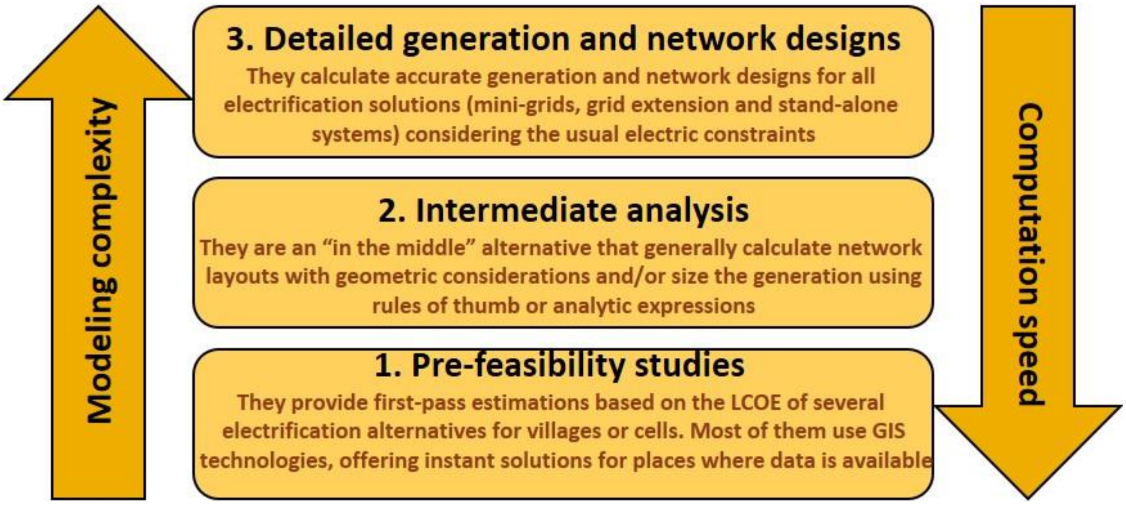


Figure 2-6: Classification of tools by modeling complexity.

We propose a classification framework based on the formulation introduced in section 2.2. This framework applies a score to each tool and methodology in three different dimensions associated with its generation sizing procedure, its network design algorithm, and its objective function. Each tool is then ranked on each dimension with an average weighted sum of the corresponding criteria, which is scaled to the interval [0, 1] (zero corresponds to the worst possible score, and one corresponds to a perfect score). Table 2-2 shows the criteria considered for each dimension and their corresponding weights, which are based on the authors' expertise.

Dimension	Aspects	Weight
Generation sizing	Generation technologies considered	5
	Generation sizing method	5
	Location of the generation site	3
	Mini-grid architecture	3
	Criteria considered	3
Network design	Layout calculation method	5
	Network levels considered	3
	Topography	2
	Criteria considered	3
Objective function	Multicriteria approach	2
	User-defined terms	2
	Criteria considered	3

Table 2-2: The classification framework.

Regarding generation sizing, the most frequently considered technologies include solar panels, wind, diesel generator, hydro, and biomass. The methodology also values how easy it is to incorporate a new generation technology into the tool or methodology. The generation sizing methods range from estimating the Levelized Cost of Electricity (LCOE) to classical optimization methods that guarantee optimality at the expense of a high computation cost. The possibility of considering several architectures (that could be in Alternating Current (AC) and Direct Current (DC)) or customizing it is very valuable. The generation equipment may be located in several spots or a single place, and the most common criteria considered when solving this problem are cost, NSE, and emissions.

Regarding network design, network layout calculation methods range from estimating the LCOE without designing the layout to using classical optimization techniques that ensure optimality at the expense of a very high computation cost. Some tools consider usual electric constraints such as maximum voltage drop allowed when designing the network of a mini-grid or a grid extension, and they ensure the electric feasibility of the network performing power flows. Others rely entirely on geometric calculations, but their network designs would need to be checked before implementation since they could be electrically unfeasible. Most tools calculate the layout of the distribution networks when extending the power grid, but connecting many consumers to the current distribution network could imply reinforcing the distribution and transmission network upstream. Although upstream reinforcements may be neglected in small-scale projects, they may significantly impact large-scale electrification plans. Similarly, the viability and cost of a network design may be heavily influenced by the terrain slopes and forbidden zones, such as the lakes and mountains of the analysis region. In addition, future expenses may be avoided if the networks of mini-grids are grid-compatible. The most common criterion considered when solving this problem is cost.

Regarding the objective function, there are two ways of dealing with a multicriteria problem. The first one consists in defining a single objective function that includes all the criteria involved (i.e., the total cost plus penalties for non-served energy and emissions). The second one involves working with several objective functions and applying multicriteria optimization techniques to obtain the Pareto frontier or set of non-dominated solutions. Some tools such as Logiciel d'Aide à la Planification d'Électrification Rurale (LAPER) and GEOSIM include elements in their objective function that are associated with non-techno-economic factors such as political and development criteria for villages. This allows the user to perform sensitivity analyses and measure the impact of these factors in the electrification solution. Most tools and methodologies perform a cost minimization, but other criteria such as minimizing emissions and NSE are relevant too.

The current electrification planning tools fail to deal with uncertainty, although it plays a critical role in energy planning (Mirakyan and De Guio, 2015). This limitation can be mitigated using scenario analysis, which involves defining several scenarios where the uncertain parameters take several values and analyzing the corresponding impact on the planning solution. However, this is only a partial solution (DeCarolis et al., 2017).

The critical uncertainty in the problem is related to demand. It is hard to obtain accurate demand profiles in developing countries, and they are critical in electrification planning. The demand growth is also difficult to estimate as it depends on several socio-economic factors such as population growth. An inaccurate forecast of demand growth could translate into an oversizing of the systems. It could also lead to an undersizing of the systems, which will have lower reliability than expected (Louie and Dauenhauer, 2016).

Uncertainties are also relevant in off-grid generation technologies. Diesel generation has uncertainties related to the fuel cost and fuel transportation from the nearest supply point to the mini-grid (Fioriti et al., 2018). Wind profiles are very locational and can be modeled following a Weibull distribution, whereas solar irradiance can be modeled following a Beta distribution (Khatod et al., 2010). There is plenty of literature that deals with these uncertainties when optimizing the generation size of a single mini-grid (Sharma et al., 2012), but those techniques are yet to be implemented into large-scale electrification planning tools. Uncertainty has been thoroughly studied in the related problems of distribution planning (Gholizadeh-Roshanagh et al., 2016) or transmission planning (Lawson et al., 2016; Tsamasphyrou et al., 2000). These methods could also be adapted to electrification planning tools.

The remaining of this section is devoted to analyzing the features of tools and methodologies that belong to each level with the proposed framework. As the modeling complexity increases, so does the number of equations considered from the conceptual formulation.

2.3.1. **Pre-feasibility studies**

Pre-feasibility tools operate with villages or cells, and they use LCOEs or similar economic indicators to estimate the best electrification alternative for each village or cell. The main advantages of pre-feasibility tools are low computation time and usability. Some of them exploit the benefits of GIS technologies to access data such as the location of power plants, the layout of the power grid, solar irradiance, wind speed, hydro potential, the layout of the roads, and population density (among others). Instant access to GIS databases eliminates the need to gather and process specific data from different sources, which is a time-consuming process.

Pre-feasibility tools can virtually consider any off-grid technology or combination of technologies if there are analytic expressions that allow calculation of the LCOE or a similar economic indicator. These tools provide an estimation of the cells that should be electrified with grid extensions, but they generally do not calculate layouts for the grid extension or generation designs for the off-grid systems.

These tools do not include most equations of the conceptual formulation introduced in section 2.2. Specifically, they only consider equation 2-1 from the distribution-network equations, and they do not include any generation equation (IntiGIS is an exception since it sizes off-grid generation with analytic expressions).

Perhaps the most advanced tool of this category is OnSSET (Mentis, 2017), which was developed at the Royal Institute of Technology (KTH) as an open-source tool. OnSSET can estimate the electrified consumers using GIS data, group the unelectrified consumers in square cells of 1 km^2 and calculate the LCOE of off-grid technologies that include solar, wind, hydro, and diesel. This model also incorporates topography when estimates the LCOE of electrifying a cell with grid extension, adding cost penalties on areas of high elevation and slope gradients. However, it does not consider alternative routes between the network and a cell that could lower cost. Reference (Nerini et al., 2016) shows detailed information related to the expressions involved in the LCOEs calculations. Recently, OnSSET has been updated to include a clustering algorithm that merges nearby cells and adapted its grid-extension cost estimations to deal with clusters (Korkovelos et al., 2019).

The OnSSET methodology has been applied to sub-Saharan Africa (Mentis et al., 2017a), Nigeria (Mentis et al., 2015), Ethiopia (Mentis et al., 2016), and Afghanistan (Korkovelos et al., 2020). This model has also been applied in combination with a more general energy modeling tool, Open Source Energy Modelling System (OSeMOSYS), to produce more detailed results in sub-Saharan Africa (Mentis et al., 2017b) and Kenya (Moksnes et al., 2017), and the Division of Energy System Analysis of KTH has carried out projects applying it to Nigeria, Tanzania, Zambia, Afghanistan (Korkovelos et al., 2017) and Benin (KTH Royal Institute of Technology and SNV Netherlands Development Organisation, 2018) in cooperation with the World Bank (WB), among others (KTH Royal Institute of Technology, 2018).

Another relevant tool that belongs to this level is IntiGIS (Pinedo Pascua, 2012), which was developed at Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT). SOLARGIS (Monteiro et al., 1998) and SOLARGIS2 (Domínguez Bravo et al., 2008) are preliminary versions of this tool. Reference (Amador and Domínguez, 2005) shows an application of the SOLARGIS methodology in Lorca (Spain). IntiGIS is a more advanced version with enhanced capabilities, which also groups consumers into cells and estimates their LCOE with analytic expressions. This model considers wind, solar, and diesel as off-grid technologies and the grid extension possibility as an on-grid solution. An application of IntiGIS to Latin America is presented in (Domínguez and Pinedo-Pascua, 2009). An upgraded version, IntiGIS 2.0, has been developed recently (Page Arias, 2015; Romero Otero, 2016), but we have not found publications in the literature related to it. The IntiGIS tool includes several hybrid-configurations as off-grid alternatives and applies analytic expressions to size their components.

We consider that LAPER (Fronius and Gratton, 2001) is a pre-feasibility tool because it requires the user to provide an initial grid-extension network. Afterward, the model starts disconnecting villages based on cost comparisons. LAPER also establishes an electrification schedule, ranking villages according to several criteria that are weighted with user preferences. Reference (Soler et al., 2003) shows an application of the LAPER model in Morroco.

The Rapid Electrification Screening Tool (REST) was created to provide a quick first-pass estimation that includes the detailed cost of several electrification solutions. Although it was developed to work at a local level (focusing on a single village or community), this tool has been included in this review as it can be applied iteratively to produce results similar to the ones provided by other tools included in this section. The REST tool was developed to fight against energy poverty in Uganda, Tanzania, and Kenya (Syngellakis et al., 2008).

There are also several methodologies in the literature that apply similar procedures. Electrification planning methods based on spatial analysis have been applied to Burkina Faso (Moner-Girona et al., 2016) and Africa (Szabó et al., 2011, 2013). Reference (Mahapatra and Dasappa, 2012) studies the influence of several factors in the best electrification mode, including the distance to the grid and the life cycle cost of the system. A summary of these tools is provided in Table 2-3. All of them minimize a cost-related function to obtain the solution, which is usually the LCOE although it may be a net present value or an annuity.

			Network Design		Generation Design					Objective Function	
Year	Tool	Granularity Level	NL	T	Solar	Wind	Hydro	Diesel	Biomass		Additional references
2017	OnSSET (Mentis et al., 2017a)	Cell	No	Yes	Yes	Yes	Yes	Yes	No	Cost	(Korkovelos et al., 2019; Mentis et al., 2015, 2016; Moksnes et al., 2017; Nerini et al., 2016)
2016	(Moner-Girona et al., 2016)	Cell	No	No	Yes	No	Yes	Yes	No	Cost	-
2013	(Szabó et al., 2013)	Cell	No	No	Yes	No	Yes	Yes	No	Cost	(Szabó et al., 2011)
2012	(Mahapatra and Dasappa, 2012)	Village	No	No	Yes	No	No	No	Yes	Cost (+)	-
2012	IntiGIS (Pinedo Pascua, 2012)	Cell	No	No	Yes	Yes	No	Yes	No	Cost	(Amador and Domínguez, 2005; Domínguez and Pinedo-Pascua, 2009; Page Arias, 2015; Romero Otero, 2016)
2008	REST (Syngellakis et al., 2008)	Village	No	No	Yes	Yes	No	Yes	No	Cost	-
2001	LAPER (Fronius and Gratton, 2001)	Village	Yes	No	Yes	Yes	Yes	Yes	No	Cost (*)	(Soler et al., 2003)

Table 2-3: Pre-feasibility tools and methodologies. NL=Network Layout, T=Topograhy.

(*) LAPER includes user-defined terms that measure political, financial resources and development criteria to calculate an electrification schedule, but the final electrification mode of each village is calculated solely with a cost minimization criteria.

(+) The impact of carbon emissions is considered, including cost terms in the calculations of the LCOEs.

Figure 2-7 shows the pre-feasibility classification according to the framework proposed in Table 2-2. LAPER is the most advanced tool in terms of network design and objective functions, although it does not provide geospatial capabilities that facilitate data acquisition. IntiGIS is the most remarkable one regarding generation designs since it includes several hybrid configurations among its off-grid candidate solutions, and it uses analytic expressions to size off-grid components, going one step beyond other pre-feasibility tools.

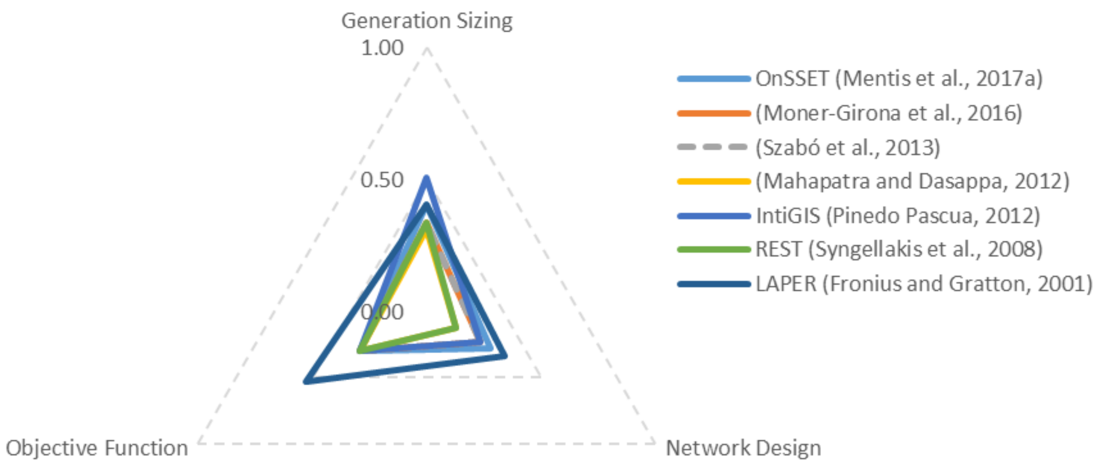


Figure 2-7: Pre-feasibility tools classification according to the classification framework.

Figure 2-7 shows that there is room for improvement in the objective-function dimension. There is a trade-off between accuracy and computation time in the generation sizing and network design dimensions, but the addition of user-defined criteria in a similar manner as LAPER and reference (Mahapatra and Dasappa, 2012) should improve the tools without a significant increase in computation time. However, the planer can also consider those criteria later, post-processing the electrification solution that the tools provide.

2.3.2. **Intermediate analysis tools**

Intermediate analysis tools generally calculate the layout of the grid extensions with heuristic methods based on minimum-spanning-tree algorithms, although they do not guarantee the electric feasibility of the designs. Any tool that calculated off-grid generation designs with an algorithm that went beyond rules of thumb that pre-feasibility tools apply would belong to this category too. Unfortunately, we have only identified one methodology with this level of complexity in generation designs (Blechinger et al., 2019), even though incorporating an iterative application of generation sizing algorithms seems reasonably straightforward (Luna-Rubio et al., 2012; Upadhyay and Sharma, 2014).

As these tools are at an intermediate level, there is a partial correspondence between the aspects they consider and the formulation introduced in section 2.2. They usually consider several distribution-network equations (2-1-2-5) and some generation equations (2-14, 2-15, a stylized version of equations 2-16 and 2-17, and 2-18). Most intermediate analysis tools focus on cost minimization and consider neither NSE nor emissions.

Reference (Blechinger et al., 2019) presents the most dominant method in this category, developed by researchers from the Renier Lemoine Institut (RLI). It generates a raster map to determine the layout of the network. The raster map is created considering topography information such as terrain elevation, slopes, and location of protected areas and forests. This method applies an iterative process to determine the generation design of each mini-grid, simulating one year on an hourly basis for each candidate generation design. The method also includes an initial clustering algorithm, and it can divide the plan into several implementation phases to facilitate project execution. Researchers from the RLI previously developed similar (albeit less sophisticated) methodologies and applied them in Nigeria (Bertheau et al., 2016) and sub-Saharan Africa (Bertheau et al., 2017).

Network Planner (Columbia University, 2017), which was developed by the University of Columbia, is an open-source tool that considers diesel, solar, and batteries as off-grid technologies. The model sizes the off-grid generation with rules of thumb, such as setting the battery capacity to a multiple of the solar capacity.

Network Planner uses an iterative algorithm to calculate grid extension layouts (Parshall et al., 2009). This algorithm compares the best off-grid solution for a village with its internal grid-extension cost, which consists of a transformer and the low-voltage network cost, but it does not include the cost of the MV line that connects the village to the power grid or another grid-electrified village. For those villages with an internal network cost lower than its best off-grid cost, the model calculates the maximum length that this MV line could have, so that grid extension is the best option. Then, the model iterates connecting villages where the maximum length is larger than its distance to the network or a village that was connected in a previous iteration. Network Planner has been applied in Nigeria (Ohiare, 2015; World Bank, 2016), Liberia (Modi et al., 2013), and Ghana (Kemausuor et al., 2014a).

The iterative procedure that Network Planner uses to connect villages to the grid is also applied in (Banks et al., 2000), which was one of the first electrification planning methodologies proposed in the literature. A similar method is presented in (Deichmann et al., 2011), which introduces an algorithm that tries to emulate the development of the power systems to determine the expansion of the power grid.

Another important tool that belongs to this level is GEOSIM, which was developed by Innovation Energie Développement (IED). This tool labels some villages as Development Poles (DPs) according to their inner potential, which is calculated based on health, local economy and education indicators, as well as the distances to the remaining villages. Then, it creates clusters of villages that are electrified together around DPs using an algorithm based on the Huff model (Huff, 1963). The idea of clustering villages is entirely coherent when working with low-populated settlements that are close to each other, and it is not present in most tools of this category or the previous one.

GEOSIM has been applied in several countries (Innovation Energie Developpement, 2018), developing projects and selling licenses. Particularly, it was applied in Benin, Burkina Faso, Cameroon, Ethiopia (Innovation Energie Développement, 2007), Madagascar, Mali, Niger, Tanzania, Cambodia, and Lao People's Democratic Republic (Innovation Energie Développement, 2010).

A summary of the tools and methodologies that correspond to this section is provided in Table 2-4. A few intermediate analysis tools use classical optimization to obtain the layout of the network, but this comes at the expense of high computation times and a reduced number of villages that the methodology can process.

One of these methodologies (Abdul-Salam and Phimister, 2016b) considers three objective functions using hierarchical lexicographic programming (the first level maximizes the aggregated demand electrified with grid extension, the second level minimizes the sum of the distances between grid extension villages and the power grid, and the third level minimizes the inter-distance among grid extension villages). This methodology requires knowing beforehand the number of villages that are electrified with grid extensions, which limits its applicability.

Another methodology that applies classical optimization to calculate the network layout is (Abdul-Salam and Phimister, 2016a), minimizing a cost function and comparing the results with the heuristic presented in (Parshall et al., 2009). The main drawback of this methodology is that it represents the existing grid. Another methodology that uses classical optimization is (Zeyringer et al., 2015), which applies a Mixed Integer Linear Programming (MILP) model to determine which cells should be electrified extending the transmission network and which ones should have solar panels. However, the size of the cells is considerably large (2000 km^2), and the off-grid generation is limited to solar panels.

There are several general-purpose models such as Energy Flow Optimization Model (EFOM), MARKet ALlocation (MARKAL), or The Integrated MARKAL-EFOM System (TIMES) that also use classical optimization techniques. These models were developed to provide assistance in a wide range of energy planning problems and allow the user to customize specific parts to develop a tailormade application for their specific problem. Some of these models could deal with the techno-economic planning problem up to a certain extent (Howells et al., 2005), and they can be combined with the tools described in this review as in the case of OSeMOSYS and OnSSET (Mentis et al., 2017b; Moksnes et al., 2017) mentioned in section 2.3.1. Although these models have been applied to small-scale problems (i.e., a problem with a single village or settlement) (Fuso Nerini et al., 2015), they have not been applied to large-scale planning cases in the literature. Therefore, these models are beyond the scope of this review.

There are algorithms that design networks for the grid-connected villages (Kocaman et al., 2012), but they are not included in Table 2-4 since they do not deal with the planning problem as a whole (i.e., they do not consider off-grid alternatives as viable electrification solutions).

			Network Design							Generation Design							Objective Function	
Year	Tool	Granularity Level	NL	TE	DL	UR	EF	T	Layout calculation	Solar	Wind	Hydro	Diesel	Biomass	Biodiesel	Sizing Method		Additional references
2019	(Blechinger et al., 2019)	Cell	Yes	No	MV	Yes (<)	No	Yes	MST-based calculations	Yes	No	No	Yes	No	No	Iterative process	Cost	(Bertheau et al., 2016, 2017)
2016	(Abdul-Salam and Phimister, 2016b)	Village	Yes	No	MV	No	No	No	Classical Optimization (hierarchical lexicographic programming) + MST-based calculations	Yes	Yes	No	No	No	No	Exogenous calculations	Multiple criteria (*)	-
2016	(Abdul-Salam and Phimister, 2016a)	Village	Yes	No	MV	No	No	No	Classical Optimization (MINLP)	Yes	No	No	Yes	No	No	Exogenous calculations	Cost	-
2015	(Zeyringer et al., 2015)	Cell	Yes	Yes	No	No	No	No	Classical Optimization (MILP)	Yes	No	No	No	No	No	Classical Optimization	Cost	-
2013	Network Planner	Village	Yes	No	MV	No	No	No	MST-based calculations	Yes	No	No	Yes	No	No	Analytic Expressions	Cost	(Parshall et al., 2009; Ohiare, 2015; World Bank, 2016; Modi et al., 2013; Kemausuor et al., 2014a)
2011	(Deichmann et al., 2011)	Village	Yes	Yes	HV/ MV	No	No	No	MST-based calculations	Yes	Yes	No	Yes	No	Yes	Exogenous calculations	Cost	-
2010	GEOSIM	Village	Yes	No	MV	No	No	No	MST-based calculations	Yes	No	Yes	Yes	Yes	No	No (+)	Cost	(Innovation Energie Développement, 2007, 2010)
2000	(Banks et al., 2000)	Village	Yes	No	MV	No	No	No	MST-based calculations	Yes	No	No	No	No	No	Exogenous calculations	Cost	-

Table 2-4: Intermediate analysis tools and methodologies. NL=Network Layout, TE=Transmission Expansion, DL=Distribution Layout, UR=Upstream Reinforcements, EF=Electric Feasibility, T=Topography, HV=High Voltage, MINLP=Mixed Integer Non-Linear Programming. MST=Minimum Spanning Tree

(<) This method includes a cost input parameter, in $/kWh, to account for power generation and transmission costs. This parameter may be used to include an estimation of the reinforcement costs in an approximate way.

(*) Maximize demand covered by the grid, minimize distance among villages and the grid, and minimize dispersion among villages connected to the grid

(+) It estimates the LCOE to provide the best off-grid supply alternative, but it does not provide optimal generation sizes

Figure 2-8 shows the intermediate analysis classification according to the framework proposed in Table 2-2. Some intermediate analysis tools rank higher in network design than pre-feasibility tools, but this does not happen in generation sizing.

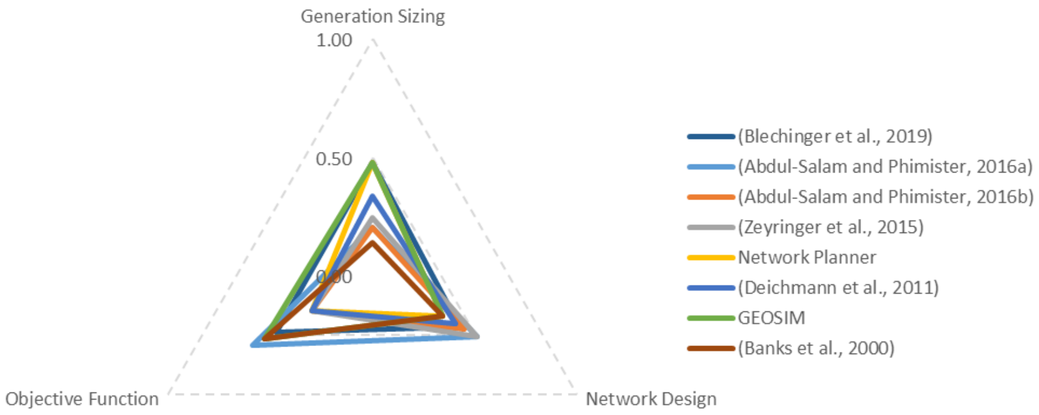


Figure 2-8: Intermediate analysis tools classification according to the classification framework.

2.3.3. **Detailed generation and network designs**

Detailed generation and network design tools operate with a very high level of modeling detail. The network designs consider the usual electric constraints (such as maximum voltage drop allowed) and topographical features of the region (such as terrain slopes and forbidden zones). The optimization of generation designs simulates the dispatch of the system, accounting for seasonalities and non-served energy. The main drawbacks of these tools are the long computation times and the need for a significant amount of input data, which are considerably difficult to obtain.

REM, which is further explained in chapter 3 and is the cornerstone of this thesis, is the only tool that arguably can be classified at this level (Ciller et al., 2019a). The modeling detail achieved in REM closely matches the conceptual formulation provided in section 2.2. All distribution-network and generation equations are considered except equation 2-24, which incorporates diesel emissions. Regarding the objective functions, REM minimizes the annual investment and operation cost plus a penalty for NSE. REM has been applied to case studies in the Vaishali district in India (Ellman, 2015), Kenya and Colombia (IIT-Comillas, 2016; Mwalenga et al., 2016), Rwanda (IIT-Comillas, 2017b), Uganda (IIT-Comillas, 2017a) and Peru (Gonzalez-Garcia et al., 2016).

The core of this tool has three submodules that perform sequential tasks. The first submodule (*mini-grid generation*) calculates generation designs for a set of representative off-grid systems using an ad-hoc adaptation of the *Hooke and Jeeves* algorithm (Hooke and Jeeves, 1961) and simulating the dispatch of the system for each candidate design that is evaluated. The generation technologies considered are limited to solar and diesel.

The results of the mini-grid generation submodule are used in the second submodule (*clustering*) to estimate the generation cost of off-grid systems. REM's clustering groups the consumers into clusters that represent optimal combinations of standalone systems, mini-grids, and grid extensions. The clustering uses quick estimations of costs because the calculation of detailed designs for each potential cluster is computationally unaffordable.

The clustering of REM generates a hierarchical structure of clusters where each consumer belongs to three clusters: an isolated cluster (the consumer on its own), an off-grid cluster, and a grid-extension cluster. The hierarchical structure of clusters is evaluated in the third submodule (*final designs*), which performs cost-comparisons among clusters to determine the final electrification solution. REM obtains detailed designs for each cluster to perform the cost-comparisons. REM goes one step further than other tools when incorporating topography, introducing topographical considerations such as terrain slopes and forbidden zones when optimizing the network layouts with RNM (Drouin, 2018).

RNM was not designed for planning in underserved regions, but to help regulators of developed countries estimate the distribution network cost. This implies that RNM could design networks that are not optimal in this context. However, network design methods described in section 2.3.2 are not as sophisticated as RNM.

If many consumers are electrified with extensions of the power grid, it may be necessary to reinforce the upstream distribution and transmission networks. There is plenty of literature related to the calculation of the reinforcements needed to cope with additional demand, but this problem has not been studied in the context of electrification in an underserved region. Several strategies have already been considered to include in REM the upstream reinforcement costs in the electrification solution (Cotterman, 2017), although they are in a preliminary stage. At this point, no electrification tool incorporates an accurate calculation of costs related to these reinforcements to the best of our knowledge.

Figure 2-9 shows the detailed generation and network designs tool classification according to the framework proposed in Table 2-2.

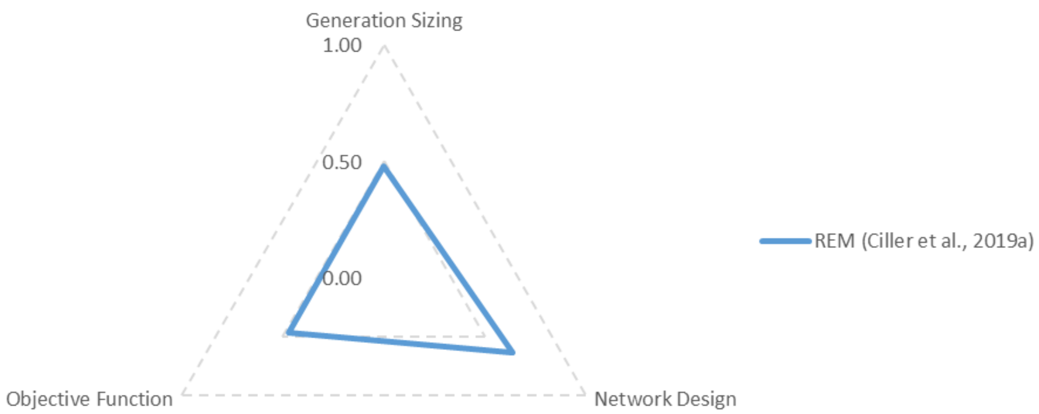


Figure 2-9: Detailed generation and network design tools classification according to the classification framework.

REM calculates generation designs more accurately than most tools analyzed here (the procedure presented in (Blechinger et al., 2019) also simulates the behavior of the system for a whole year), but it includes a limited number of generation technologies (solar and diesel) so REM does not outrank other tools regarding generation designs. However, REM obtains a high score in network designs since it considers electric constraints and considers the topographical features of the terrain when it optimizes the network layouts.

REM is the only tool that lies in the category of detailed generation and network designs, and it is reasonable to question whether the level of modeling detail in REM is justified given the uncertainty of the input data.

REM can be used at different scales, and for each one the level of uncertainty is different. When applied in a small region, it is possible to survey each building and obtain its precise location, whether it is already electrified or not, the proximity of existing network lines, and its electricity demand. The input has little uncertainty as well as the output, and the level of detail in REM is justified. However, REM was designed to tackle large-scale electrification planning where the input data cannot be trusted.

We have checked (International Energy Agency, 2018), and we are still working on confirming our findings, that the REM results are very sensitive to the level of modeling detail. The output of REM differs considerably when we consider that all buildings in some territory have the same average demand profile from the case where each building (residential, school, church, shop, mine, etc.) has its own profile (Ciller et al., 2019a). If we clump all district buildings in a single point (i.e., a village or a cell), it is impossible (unless strong assumptions are made) to differentiate the three electrification modes. Similar conclusions would be obtained if we reduced the hourly demand over a year to just the annual energy and the peak demand. Therefore, the high-level of modeling detail in REM is necessary.

REM takes a long time (several days) to run one case for a sub-Sahara African country of average size. Turning REM into a fully probabilistic model is out of the question today. However, REM is a powerful tool in the hand of an experienced planner who understands where the input uncertainty lies and who can run several scenarios, find out how robust the solutions obtained are regarding this uncertainty and make a planning decision, which ultimately is – at least for the time being – in the hands of the human planner.

2.4. Conclusions

This chapter briefly described the three main subproblems (generation sizing, network designs, and clustering) that comprise the large-scale techno-economic planning problem in developing countries. The generation sizing problem aims at finding the optimal generation design and dispatch of each off-grid system. The goal of the network design problem is to optimize the network layout of mini-grids and grid extensions. Finally, the objective of the clustering problem is to group the consumers into mini-grids and grid extensions.

We also introduced a conceptual formulation of the large-scale techno-economic planning problem. This formulation, which was missing in the literature, allowed us to classify the main methods and tools that address this problem according to their modeling complexity. The classification includes three categories.

The first category (pre-feasibility tools) includes methods and tools that provide quick solutions at the expense of low computation time and a reduced level of modeling detail. Pre-feasibility tools generally exploit the advantages of GIS technologies to provide instant access to databases that contain essential inputs for the problem. Most of these tools operate with villages or cells, and they minimize the LCOE of several electrification alternatives to determine the best option for each village or cell. Perhaps OnSSET is the most dominant tool in this category.

The second category (intermediate analysis tools) encompasses methods and tools that operate at an in-the-middle level of modeling complexity. They tend to use rules of thumb or simple methods to size the generation of mini-grids, and they calculate the layout of the grid extensions with geometric techniques based on the calculation of the MST. Reference (Blechinger et al., 2019) presents the most advanced methodology in this category.

The third category (detailed generation and network designs) refers to methods and tools that address the planning problem with very detailed modeling. These tools optimize the generation designs of off-grid systems with optimization methods that simulate the hourly dispatch, providing an accurate estimation of the amount of non-served energy. They optimize the network layouts using complex algorithms that include power flow calculations and the standard electric constraints such as voltage drops. For the time being, REM is the only tool that arguably belongs to this category.

Regarding additional developments, it would be interesting that a tool could perform multiobjective optimization (for instance, considering cost, emissions, or NSE), allowing the planner to analyze the trade-offs among several objective functions. Instead, most tools minimize the total investment and operation cost (or a similar economic indicator). Some tools include terms into the total cost that account for other factors (i.e., a penalty for NSE).

We consider that no tool can calculate the impact of grid-connections in the upstream distribution and transmission network accurately. The impact may be negligible in small-scale plans, but upstream reinforcements could be a significant part of the total electrification cost in large-scale projects. New developments are necessary to measure the cost of reinforcing the distribution and transmission networks accurately.

Additionally, the existing generation sizing algorithms that simulate the dispatch of the mini-grids consider a limited number of generation technologies (solar panels, batteries, and a diesel generator). These algorithms should include more generation technologies, although their computation time would increase.

Regarding network designs, the network design algorithms either neglect electric constraints or were not developed to electrify underserved regions (this is the case of RNM, which is used in REM). It would be interesting to develop network design algorithms aimed explicitly at the electrification of underserved regions that can cope with voltage regulations and power-flow constraints.

Eventually, future large-scale electrification planning tools should be better equipped to deal with uncertainty as there is a scarcity of data in developing countries, and the existing information is not always reliable. Although there are methods in the literature that successfully address the uncertainty of specific parameters, they were designed to deal with a single mini-grid or grid-extension. However, tools deal with a vast number of mini-grids and grid extensions in large-scale planning so a straightforward application of these methods may fail for computational reasons.

"We are gambling on our vision, and we would rather do that than make 'me-too' products."
Steve Jobs

3

THE REFERENCE ELECTRIFICATION MODEL

Section 3.1 presents an overview of REM's first prototype (Ellman, 2015), describing its sequential structure and its algorithms. Section 3.2 describes the most significant upgrades implemented in the first prototype of REM, which was conceptually sound at a high-level but provided inconsistent results. These upgrades turned REM into a robust computer tool that produces reliable solutions and has been applied to develop several master electrification plans and analyses. Section 3.3 presents a case study where the results that REM provides before and after implementing the improvements presented in this chapter are compared. Section 3.4 describes the main conclusions of this chapter.

The content of section 3.1, section 3.2.1.4, and section 3.2.3.4 have been adapted from the following paper:

Ciller, P., Ellman, D., Vergara, C., Gonzalez-Garcia, A., Lee, S.J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., Palacios, R., Stoner, R., de Cuadra, F., Perez-Arriaga, I., 2019. Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM). Proceedings of the IEEE 107, 1872–1905. https://doi.org/10.1109/JPROC.2019.2922543

A substantial part of section 3.2.2 comes from the following master thesis:

Ciller Cutillas, P., 2016. Clustering-related improvements in the Reference Electrification Model. Master thesis. Universidad Pontificia Comillas, Madrid. School of Engineering.

The content of section 3.2.1.3, section 3.2.1.1, and the input data of the case study (and part of section 3.3) have been adapted from the following paper:

Ciller, P., de Cuadra, F., Lumbreras, S., 2019. Optimizing Off-Grid Generation in Large-Scale Electrification-Planning Problems: A Direct-Search Approach. Energies 12, 4634. https://doi.org/10.3390/en12244634

3.1. The first prototype of REM

Section 3.1.1 introduces the high-level structure of REM's first prototype (Ellman, 2015), which is essentially maintained in the current version of REM. This structure consists of several submodules that operate following a sequential process. Sections 3.1.2, 3.1.3, and 3.1.4 describe the three blocks that constitute the main algorithms of the first prototype of REM. Section 3.1.5 presents the financial model of REM, describing how the first prototype of REM performs economic calculations.

3.1.1. REM in a nutshell

REM represents demand at a much higher level of detail than the remaining existing models. Instead of relying on aggregations, REM can deal at the individual building or consumer level. This allows REM to calculate detailed generation designs for all off-grid systems and network layouts for each mini-grid and grid extension. REM operates at a very high spatial (individual consumer) and temporal resolution (it simulates the hourly dispatch of off-grid systems to optimize generation designs).

REM groups the individual consumers into electrification clusters so that total system costs are minimized. These clusters may denote groups of consumers to be connected to the same mini-grid systems, groups to be connected to the existing grid, or clusters of single consumers to be supplied with standalone systems.

The potential number of reasonable electrification clusters is unmanageable (for our purposes, “reasonable” means that, for a given consumer, only the nearby consumers are considered candidates to be electrified in the same system) in large-scale planning. The problem is intricate as we do not know beforehand the least-cost electrification mode of each consumer (isolated system, mini-grid, or grid extension).

REM groups the consumers into clusters following a two-step method. The first step (off-grid clustering) momentarily assumes that consumers will be electrified exclusively with off-grid systems. The clusters at the end of this step are the off-grid clusters. The second step (grid-extension clustering) begins from the existing network and the off-grid clusters and introduces the possibility of grid extensions. The clusters at the end of this step are the grid-extension clusters.

REM does not decide the final electrification modes in the clustering process. This is determined by performing cost-comparisons among the electrification clusters, which form a hierarchical structure with three levels (see Figure 3-1). The grid-extensions clusters form the on-grid level, the off-grid clusters form the off-grid level, and the individual consumers form the isolated level.

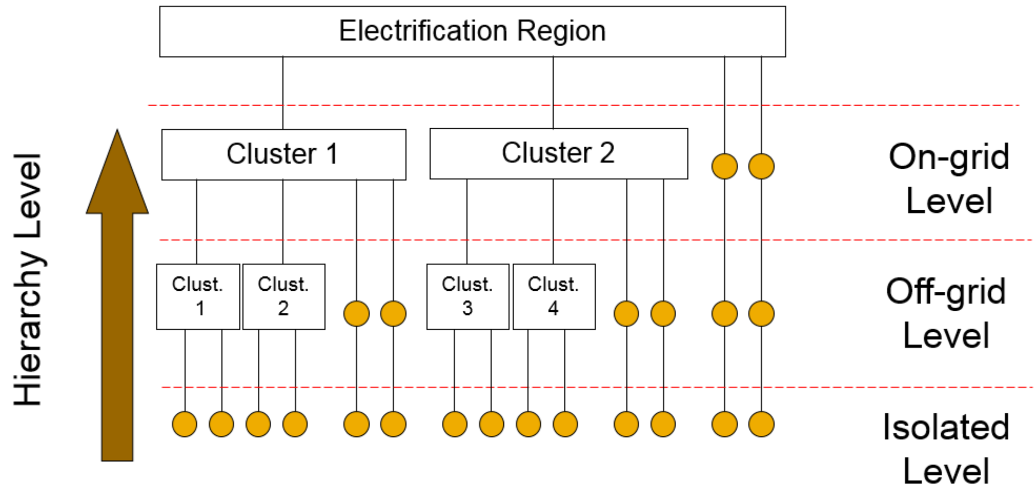


Figure 3-1: Hierarchical structure of clusters. Source: adapted from (Ciller Cutillas, 2016).

REM calculates the electrification solution comparing the cost of each grid-extension cluster to the sum of costs of all the off-grid clusters that belong to it, and the least-cost alternative is the one included in the final proposed electrification solution. A grid-extension cluster could end up being electrified with off-grid systems if the added detailed cost of the off-grid clusters underneath it is lower than the cost of a grid-extension design for the whole grid-extension cluster. Similarly, the consumers of an off-grid cluster could be electrified with a grid-extension design.

REM can obtain detailed generation and network designs to perform the cost-comparisons among the electrification clusters. Still, it is not computationally affordable to calculate such designs for each potential combination of consumers that is evaluated in the clustering process so REM uses approximations of the generation and network costs when calculating the electrification clusters.

Specifically, REM only optimizes from scratch the generation cost of a reduced subset of off-grid systems, which should be representative of the situations to be encountered in the case being studied. This process is performed before the clustering, and the results are stored in a look-up table so that the generation cost of the remaining off-grid systems can be interpolated quickly.

The network costs are estimated based on geometric and electric considerations such as distances among clusters (or a cluster with the power grid) and the peak demand of clusters. The estimations of the network costs are calculated during the clustering process.

Figure 3-2 shows the high-level structure of REM, which mostly corresponds to the procedure we have described in this section so far. This structure consists of five submodules that operate sequentially. The first and fifth submodules are related to the processing of inputs and outputs, respectively. The second, third, and fourth submodules represent the computational core of REM and the contributions of this thesis focus on them.

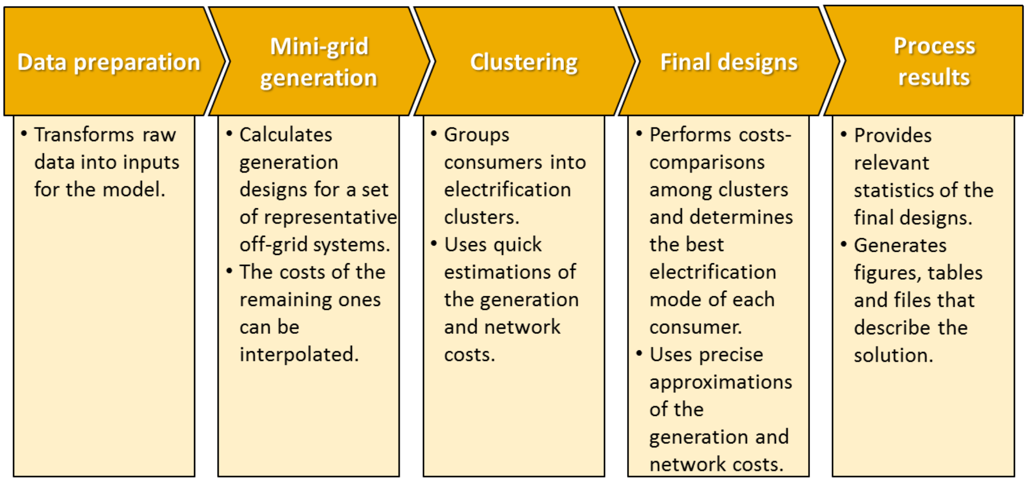


Figure 3-2: High-level structure of REM.

The goal of the first submodule (data preparation) is to obtain all the data that REM needs and convert it into the formats that REM uses. Such data includes the location of the consumers and the power grid, the demand profiles of the consumers, network and generation catalogs, techno-economic parameters such as discount rates, and the pieces of information needed to estimate the potential of generation technologies (i.e., the solar irradiance in the case of solar generation).

In the second submodule (mini-grid generation), REM optimizes from scratch the generation designs of a reduced number of off-grid systems, storing the results in a look-up table. The generation cost of the remaining off-grid systems, if needed, can be quickly interpolated.

In the third submodule (clustering), REM groups the consumers into clusters that represent the optimal grouping of consumers into standalone systems, mini-grids, and grid extensions. REM uses estimations of the generation and network costs to calculate the electrification clusters.

In the fourth block (final designs), REM explores the hierarchical structure of clusters, performing cost-comparisons to determine the electrification mode of each cluster (a combination of standalone systems, a mini-grid, or an extension of the power grid). REM calculates detailed network designs to perform the cost-comparisons among the clusters. Although REM could optimize from scratch generation designs to perform the cost-comparisons, the generation costs are generally obtained interpolating in the look-up table for computational reasons. The accuracy of interpolated generation designs is precise enough for large-scale planning.

The fifth submodule (process results) generates files that contain detailed information about the electrification solution. Such information includes cost breakdowns for each standalone system, mini-grid, and grid-extension design. REM also provides a graphical output, which includes the projection of the final electrification solution onto Google Earth.

The next section describes the second submodule of REM, mentioning how the model optimizes the generation design of a representative off-grid system and how REM handles the calculation of the look-up table.

3.1.2. **Mini-grid generation**

The design of mini-grid generation depends on the demand, available technology, and local conditions such as the cost of fuel and the hourly solar irradiance. REM assumes that mini-grids have a centralized generation and operate in islanded mode. The adopted general architecture for any off-grid system in REM is shown in Figure 3-3. REM does not include all the components in every generation design. For instance, if a mini-grid has only a diesel generator, then charge controllers and inverters will not be necessary.

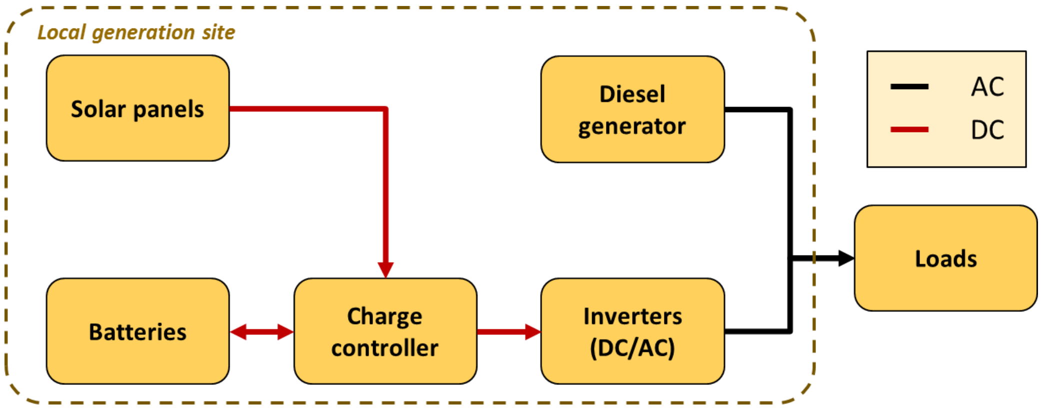


Figure 3-3: Off-grid generation architecture in REM. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

3.1.2.1. The single-system generation sizing algorithm

The problem of optimizing the generation design of a single mini-grid has been thoroughly studied in the literature (Luna-Rubio et al., 2012). One of the procedures used to deal with this problem involves applying classical optimization modeling techniques such as mixed-integer linear programming (Moretti et al., 2019). However, these methods are computationally intense and may require substantial resources to optimize a significant number of generation designs, which is often the case in the applications of REM. Heuristic techniques provide the right balance between optimality and computation time, so they are a good fit for REM.

The algorithm that the first prototype of REM uses to optimize the generation design of an individual off-grid system is a variation of the Hooke and Jeeves algorithm (Hooke and Jeeves, 1961), which is a heuristic method that starts from an initial central point and iterates performing exploratory moves and pattern moves. The algorithm moves in a search space with as many dimensions as variables are involved in the optimization process. It does not need to calculate derivatives of the objective function.

Exploratory moves evaluate the objective function in several neighboring points around the central point, obtaining information about directions where the objective function could improve. Pattern moves explore the search space in the most promising direction obtained in the exploratory moves, attempting to fasten the optimization process.

The Hooke and Jeeves algorithm is well-known in the literature, and it has been applied to many different problems. Some fields of application include electric motors (Li and Rahman, 1990; Tutelea and Boldea, 2010) and mechanical engineering (Kirgat and Surde, 2014), among many others.

The optimization that REM performs happens in a tridimensional search space, where the dimensions are: (1) diesel generator capacity, (2) total capacity of the solar panels, and (3) battery capacity. REM sizes the remaining components of the design afterward. The boundaries of the search space are calculated considering the aggregated demand of the off-grid system.

The objective function that REM considers is the total cost, including investment and operation cost plus a penalty for non-served energy (i.e., the estimated cost of the lack of electricity supply to the end consumer). Some components of the cost of the candidate designs depend on the operation of the mini-grid (i.e., fuel cost), so the model evaluates the objective function for each candidate generation design by performing an annual simulation of the operation of the mini-grid, adopting some generation dispatch strategy.

The first prototype of REM defined an initial central point and step lengths for the variables involved, which were used to calculate neighbor points around the central point. Then, the algorithm evaluated the cost of several neighbor points, moving along the least-cost direction.

If the algorithm could not reduce the cost moving along any of the direction evaluated, then it reduced the lengths of the steps and evaluated neighbor points that were closer to the central point. This procedure continued until the steps could not be reduced any longer. Figure 3-4 shows the flow diagram of the generation sizing algorithm of the first prototype of REM.

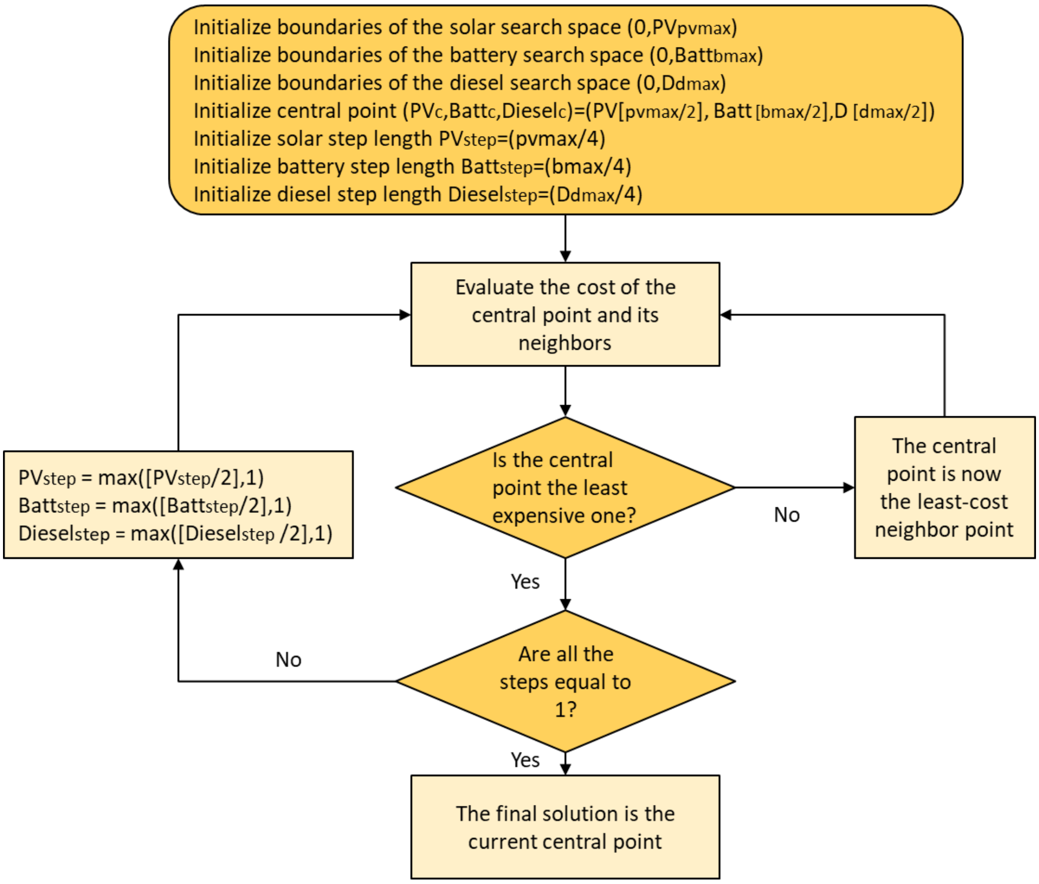


Figure 3-4: Flow diagram of the generation sizing algorithm present in the first prototype of REM.

The next section describes the dispatch strategy that the first prototype of REM used to simulate the hourly dispatch of a generation design.

3.1.2.2. The dispatch algorithm

The first prototype of REM simulates the hourly dispatch of each candidate generation design evaluated by the algorithm described in section 3.1.2.1. To do that, it applied an hourly dispatch strategy named *battery valuation*, where the marginal costs of the generation components determine the order of resources used to meet the demand.

The marginal solar cost is set to 0 $/kWh (solar panels generate electricity at an approximately zero cost once they have been installed). The marginal cost of the diesel is calculated assuming that it operates at half-load, and it depends on its efficiency and the fuel cost. The costs of non-critical and critical non-served energy are input parameters of the model.

The battery valuation dispatch assigns a marginal cost to the use of the battery that depends on its state of charge, the marginal costs of the remaining elements, and a correction that accounts for losses and battery use. The idea is that the marginal cost of using the battery increases as its state of charge decreases and vice versa.

In most cases, the marginal cost of diesel is lower than the CNSE so that REM provides solutions where off-grid systems serve a high percentage of their demands. By definition, the cost of critical non-served energy is equal to or higher than the cost of non-critical non-served energy. Hence, the marginal costs of the elements considered in the dispatch (except the battery) usually satisfy equation 3-1:

$$Solar_{mc} < Diesel_{mc} < NonCriticalNSE_{mc} < CriticalNSE_{mc} \qquad 3\text{-}1$$

Where $Solar_{mc}, Diesel_{mc}, NonCriticalNSE_{mc}, CriticalNSE_{mc}$ are the marginal costs of solar generation, diesel generation, non-critical non-served energy, and critical non-served energy, respectively. Table 3-1 shows the most frequent relationship between the state of charge of the battery (SOC) and the resource that determines its marginal cost in the battery valuation strategy[3].

State of charge interval	Resource that sets the marginal cost of the battery
$0.5 \leq SOC \leq 0.625$	Critical non-served energy
$0.625 < SOC \leq 0.75$	Non-critical non-served energy
$0.75 < SOC \leq 0.875$	Diesel generator
$0.875 < SOC \leq 1$	Solar (0 $/kWh)

Table 3-1: Most frequent assignation of the marginal cost of the battery.

The marginal cost of the battery in each hour is usually calculated as the marginal cost of the corresponding resource provided in Table 3-1 plus a correction that accounts for losses in the mini-grid and battery use. Table 3-2 shows a frequent order of resources used to meet the demand with the battery valuation strategy.

State of charge interval	Order of resources used to meet the demand
$0.5 \leq SOC \leq 0.625$	Solar, diesel generator, non-critical non-served energy, critical non-served energy, battery
$0.625 < SOC \leq 0.75$	Solar, diesel generator, non-critical non-served energy, battery, critical non-served energy
$0.75 < SOC \leq 0.875$	Solar, diesel generator, battery, non-critical non-served energy, critical non-served energy
$0.875 < SOC \leq 1$	Solar, battery, diesel generator, non-critical non-served energy, critical non-served energy

Table 3-2: Most frequent order of resources used to meet the demand.

Once the dispatch has determined the order of resources that meet the demand, the battery valuation strategy determines which resources are used to charge the battery. This decision depends on the marginal cost of the battery, and the marginal cost and availability of the remaining elements of the mini-grid.

Specifically, if the marginal cost of solar or diesel (plus a correction for losses and using battery lifetime) is lower than the marginal cost of the battery, then that resource is used to charge the battery (if it is still available after meeting the demand).

REM considers the kinetic battery model described in reference (Manwell and McGowan,

[3] *We assume that the minimum state of charge allowed for the battery is 0.5 for the sake of simplicity, although it is an input parameter of the model. If the minimum state of charge allowed is $\underline{soc}$, then the interval [$\underline{soc}$, 1] is divided into four subintervals of equal length to determine the ranges of the SOC where each resource sets the marginal cost of the battery.*

1993) to determine the maximum allowed charge and discharge rates of the battery in an hour. The kinetic battery model, which is also used in HOMER (HOMER Energy LLC, 2019a), represents the battery with two tanks. The first tank contains energy available for direct use, and the second tank contains energy that cannot be used immediately. There are time constraints that determine the amount of energy that can flow from one tank to the other in a time period.

3.1.2.3. The look-up table

The task of calculating accurate generation designs for all the candidate mini-grids is computationally unfeasible in a large-scale problem. This implies that the iterative application of a generation sizing tool such as HOMER would not work for this application.

What REM does to solve this problem is to calculate only a few generation designs for a representative number of candidate mini-grids and, if it needs the generation cost of another design, the model will obtain it using linear interpolation. Figure 3-5 shows a small electrification planning problem that we will use to illustrate this concept. There are 32 residential consumers in this problem with the same demand profile.

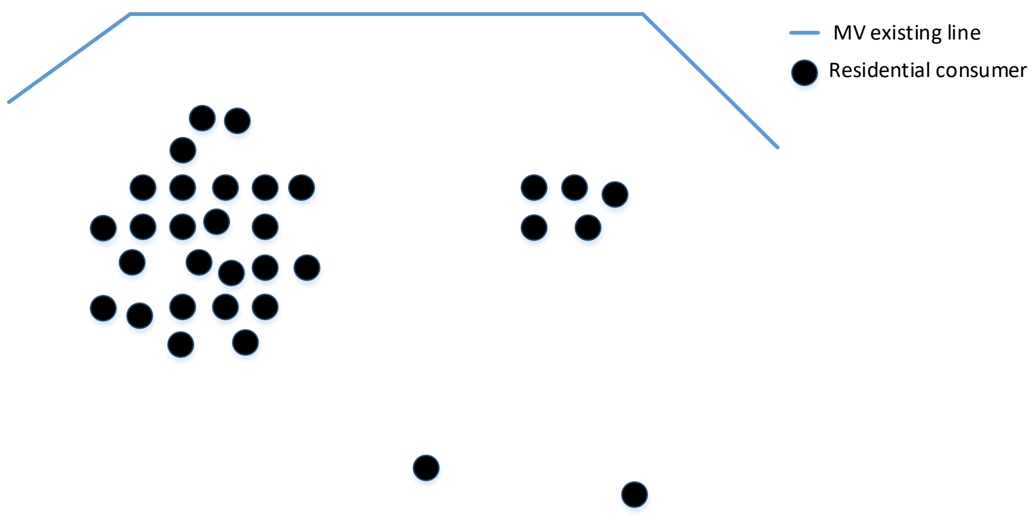


Figure 3-5: Electrification problem example. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

Since all the consumers have the same demand, there are also 32 candidate off-grid systems with different aggregated demands (those with 1, 2, ..., 31, 32 consumers). Instead of calculating these 32 generation designs, which would be a feasible strategy for this toy example but not for a large-scale problem, REM could calculate the generation designs related to 1, 2, 5, 15 and 32 consumers. If it needs the generation cost of a generation design with, for instance, 25 consumers, the model will interpolate the values, using the data from the designs with 15 and 32 consumers.

In this example, the look-up table would have a single axis with five representative points related to residential consumers. The first prototype of REM was limited to one single consumer type, so it was not possible to include productive loads in the cases. The first prototype of REM automatically selected the look-up table's representative points, including logarithmically spaced points up to the number of consumers of the case.

3.1.3. **Clustering**

The goal of the clustering process is to determine which consumers should be electrified together (i.e., with the same system). Evaluating all the possible combinations of consumers is computationally infeasible in large-scale problems, so the first prototype of REM uses a Minimum Spanning Tree (MST) to obtain the potential connections among consumers.

Figure 3-6 adds the candidate connections to the example shown in Figure 3-5. Using the consumer identifiers of Figure 3-6, it is clear that consumers 1 and 2 could be electrified together, but a direct connection between them and consumers 31 and 32 is not worth considering. However, consumers 1 and 32 could be electrified in a single mini-grid or grid extension if economies of scale justify the gradual aggregation of more consumers until all the consumers shown in Figure 3-6 belong to the same cluster.

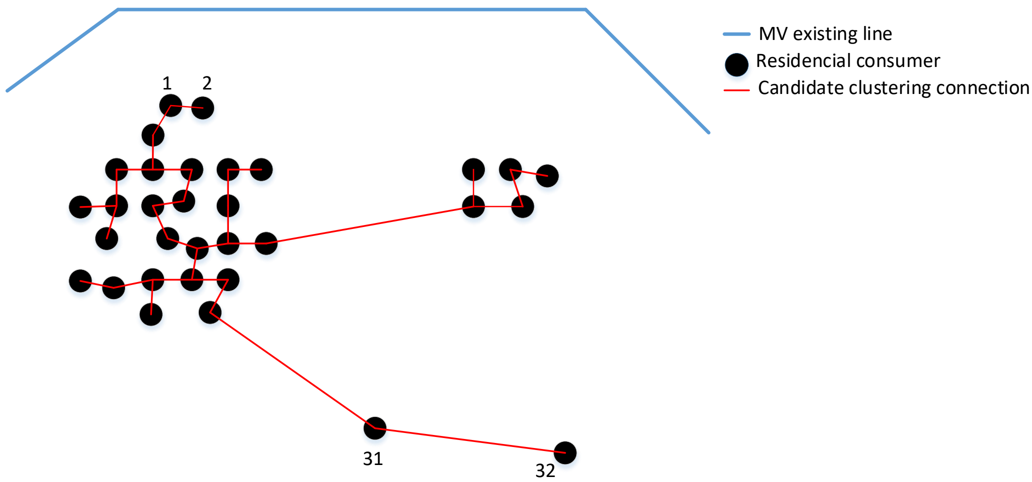


Figure 3-6: Clustering candidate connections. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

3.1.3.1. Off-grid clustering process

The first step of the clustering process (off-grid clustering) temporarily assumes that all the consumers will be electrified with off-grid systems. The algorithm makes consumer-grouping decisions based on two conflicting drivers: (1) the savings in generation costs due to economies of scale inherent to large mini-grids, versus (2) the increment of network costs associated with grouping consumers together.

REM evaluates first the arcs of the MST that are more likely to be activated by joining the corresponding clusters, i.e., from the shortest to the longest arc. In each evaluation, the model compares the costs of the configurations shown in Figure 3-7 to determine if the connection should be activated. In the figure, triangles represent generation sites.

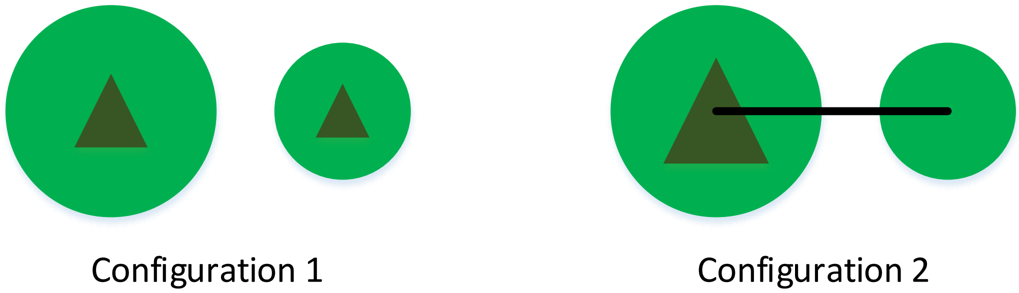


Figure 3-7: Off-grid clustering configurations. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

Configuration 1, where the clusters are electrified separately, has a higher generation cost than configuration 2. However, configuration 2 has a higher network cost than configuration 1. REM estimates the cost difference between both configurations (generation costs are obtained from the look-up table described in section 3.1.2, and the line shown in configuration 2 approximates the incremental network costs between both configurations) and joins the clusters if configuration 2 is less expensive. Figure 3-8 shows a possible off-grid clustering solution with seven off-grid clusters for the example under consideration and the arcs of the MST that have not been activated.

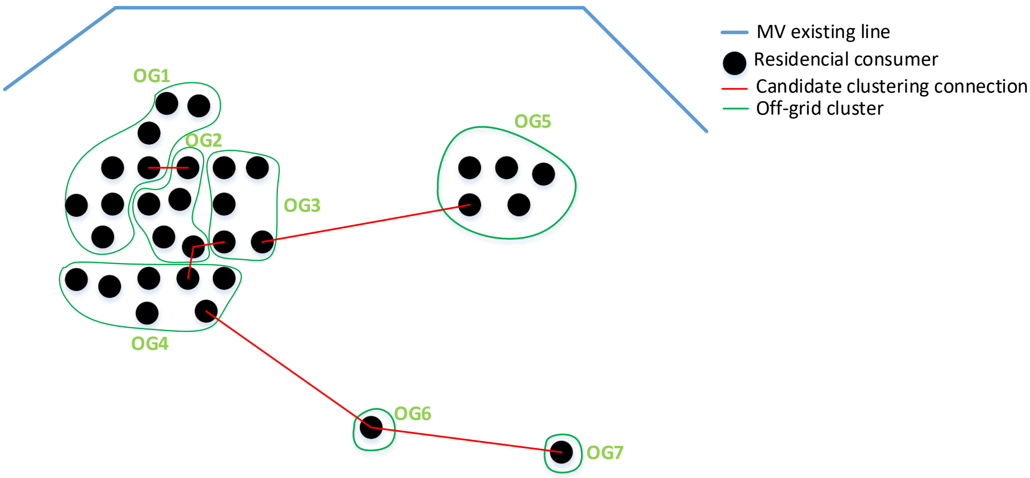


Figure 3-8: Off-grid clustering example. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

The arcs of the MST are used as potential clustering connections; therefore, some of them may be redundant (if they link the same pair of clusters) and just ignored. The clusters at the end of this step are the off-grid clusters, and they are the starting point of the grid-extension clustering process.

3.1.3.2. Grid-extension clustering process

The second step of the clustering process (grid-extension clustering) starts from the existing network and the off-grid clusters obtained in the first clustering step. Note that REM does not decide the final electrification modes in the clustering process. This is determined in the final design phase, described in section 3.1.4, when the different alternatives derived from the clustering process are examined and final detailed comparisons are made.

The grid-extension clustering evaluates the arcs of the MST that join pairs of two different off-grid clusters. This algorithm calculates the cost of several configurations to determine if it is worth joining both clusters, under the assumption that at least one of them will be connected to the grid. In the first set of configurations, which is shown in Figure 3-9, both clusters are connected with a line (triangles represent here MV/LV transformers; thick lines are MV lines and thin ones are LV lines). This implies that REM will join both clusters if a configuration from this set ends up being the least-cost one.

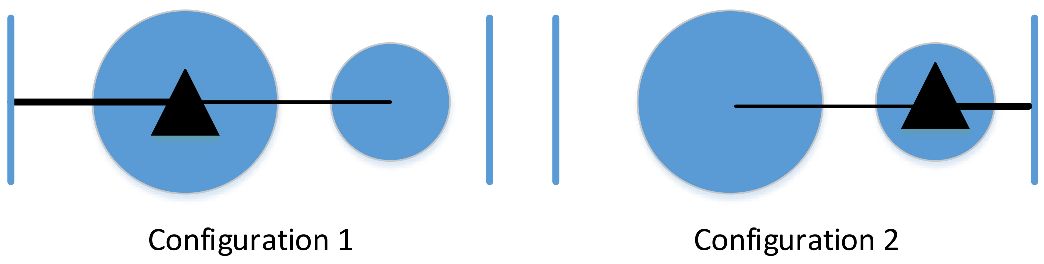


Figure 3-9: Set of configurations that support merging grid-extension clusters. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

Figure 3-10 shows several configurations with the clusters electrified separately. In configurations 3 and 4, one of the clusters is electrified with an off-grid system (here triangles are transformers or generation sites). If a configuration from Figure 3-10 is the least-cost one, then REM will not connect both clusters.

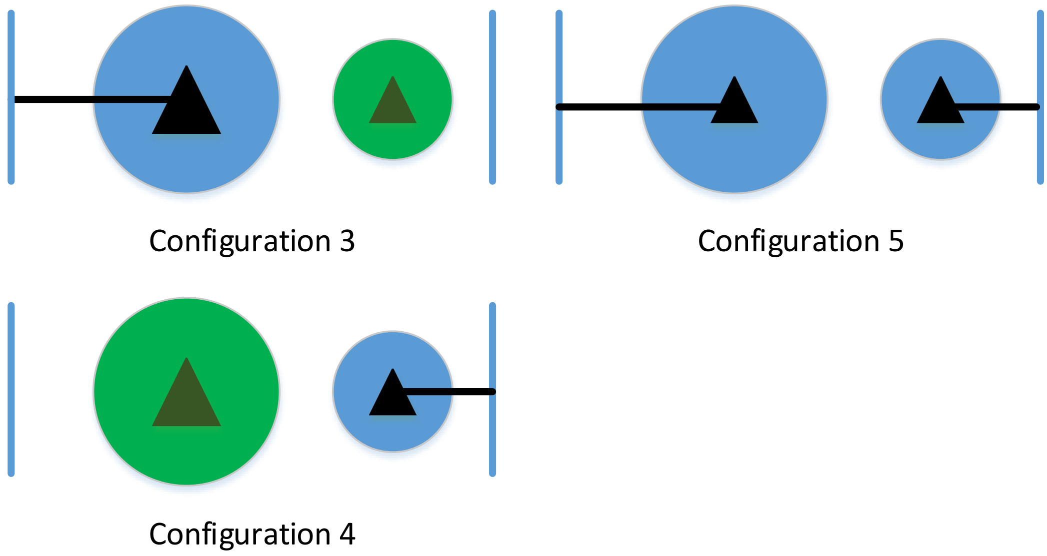


Figure 3-10: Set of configurations that support keeping grid-extension clusters separate. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

The clusters at the end of this step are named grid-extension clusters (admittedly a confusing term, since many of these clusters may end up not being grid-connected). Following with the example, Figure 3-11 shows the corresponding grid-extension clusters at the end of the grid-extension clustering process and the arcs of the MST that have not been activated. In Figure 3-11, none of the clusters appears as connected to the MV grid, as the purpose of this grid-extension phase is not to decide on grid connection, but to get a better set of clusters that will be thoroughly analyzed in the final designs phase.

Note that, in contrast to other electrification planning methods that are rule-based, REM may find off-grid electrification solutions for consumers that are close to the network if the off-grid solution is less expensive. This typically happens when the aggregated demand of these consumers is so low that – following a cost minimization logic – it does not justify the investment in the minimum size transformer in the catalog and the corresponding wiring cost needed for an extension of the power grid.

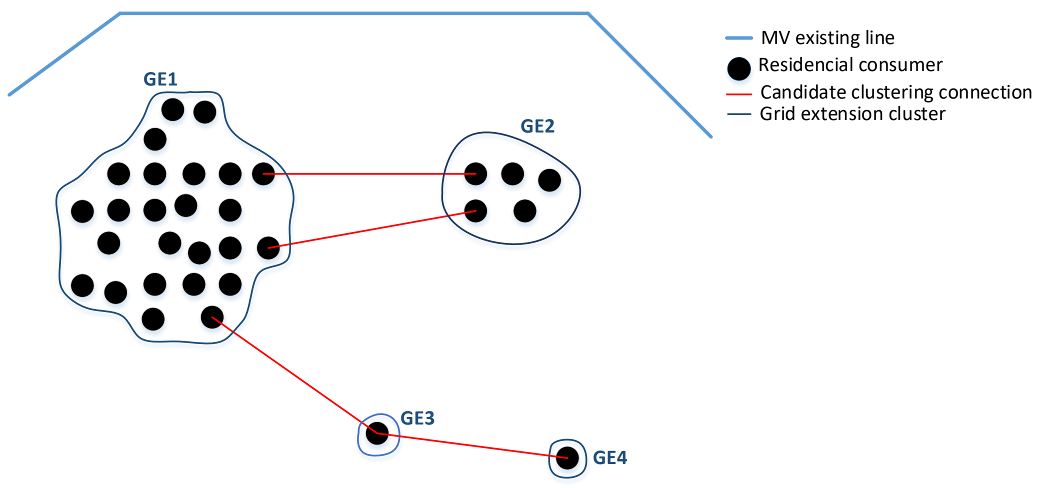


Figure 3-11: Grid-extension clustering example. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

The clustering process creates a hierarchical structure of clusters where the first level contains the grid-extension clusters, the second level contains the off-grid clusters, and the third level contains the individual consumers. Figure 3-12 shows the structure that corresponds to the example.

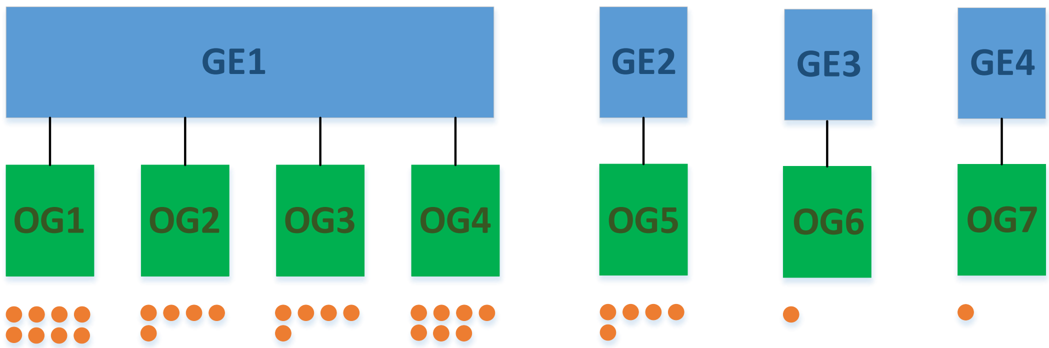


Figure 3-12: Hierarchical clustering of consumers. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

This cluster structure is used to determine the electrification mode of each consumer, and the goal of the clustering processes is to deliver a quasi-optimal structure of clusters to be thoroughly explored in the final design phase.

3.1.4. **Final designs**

In the final design process, REM exploits the hierarchical structure of clusters in Figure 3-12 to determine the best electrification mode for each consumer, which belongs to three nested clusters: the individual consumer, its off-grid cluster, and its grid-extension cluster. REM calculates mini-grids for the off-grid clusters, and extensions of the power grid for the grid-extension clusters.

REM obtains the least-cost electrification solution for a cluster by comparing its least-cost electrification mode with the sum of the best electrification solutions of the clusters that are in the immediately lower level in Figure 3-12. Therefore, cost evaluations are propagated bottom-up in the structure of clusters.

In the example of Figure 3-12, let us consider the structure below GE1. First, REM would find the cost of electrifying individually the consumers below OG1, providing each consumer with an individual generation set. This solution is the temporary optimum solution for the OG1 set of consumers, and it is compared with the electrification cost for OG1 with a single mini-grid. The least-cost solution becomes the temporary optimum solution for the OG1 set of consumers.

The same process is applied to the consumers in OG2, OG3, and OG4, respectively. The resulting group of least-cost solutions becomes the temporary optimum solution for the GE1 set of consumers (note that this solution may include a combination of isolated consumers and mini-grids).

The final step is to compare this temporary optimum solution with electrifying GE1 as a single extension of the power grid. The least-cost solution becomes the final optimum solution for the GE1 set of consumers.

Figure 3-13 shows a possible final electrification solution for the example under consideration. In this case, the cluster GE1 is electrified with a grid extension design, whereas the remaining grid-extension clusters have lower costs when electrified with off-grid systems that are coherent with the hierarchical structure.

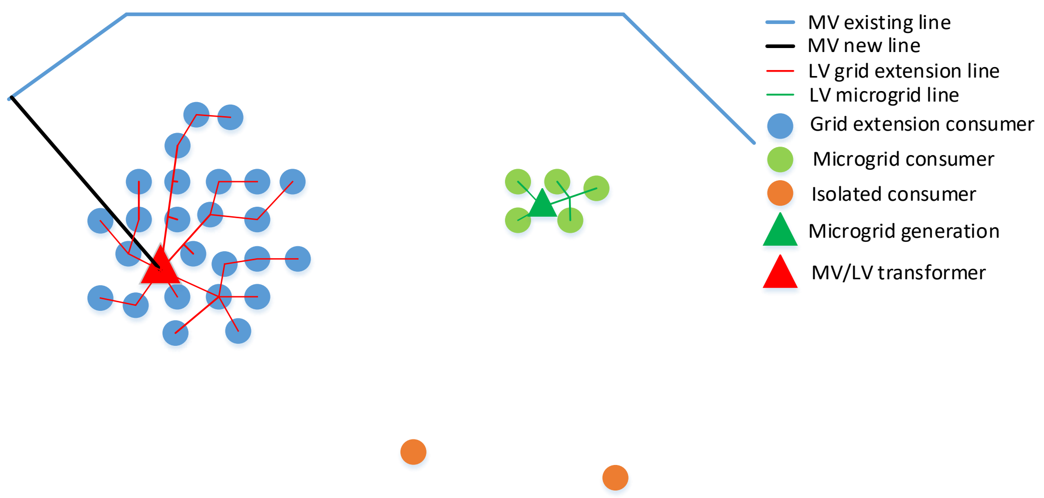


Figure 3-13: Final electrification solution. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

In this stage, accurate network designs are calculated for the cost comparisons performed to determine the final electrification mode of each consumer. REM uses RNM to obtain the optimal network layouts and the corresponding costs.

3.1.4.1. RNM as a network designer

RNM is a flexible model that can design a balanced three-phase quasi-optimal distribution network from scratch, calculating the corresponding costs (Mateo Domingo et al., 2011). RNM can create the entire distribution network starting from the transmission/HV distribution substations, or only the medium and LV components, or only the LV network. RNM needs the location of the transmission substations and the consumers, as well as techno-economic information related to the catalog of components (mainly lines and substations/transformers). If RNM is used to design only a part of the distribution network, the locations of the corresponding upstream substations are required too.

RNM minimizes cost, subject to the usual electrical constraints, such as maximum allowed voltage drop and maximum capacity. The model selects the best elements among a defined catalog of components, and it considers the influence of topography when calculating a network layout. RNM also allows forbidden and penalized zones.

3.1.4.2. Network design for mini-grids and grid extensions

The initial prototype of REM used RNM to design a single network for the mini-grid, and it determined if it was LV or MV according to the number of transformers that RNM located: if RNM placed one transformer, then REM assumed that it was the generation site of the mini-grid, and only included the cost of LV lines. If RNM placed several transformers, then REM assumed that the mini-grid had an MV design, and its network cost included the cost of transformers, MV, and LV lines.

REM should provide LV designs for small mini-grids and MV designs for large ones, where the “size” of a mini-grid here must be understood as a combination of distance, number of consumers, and total load. Figure 3-14 shows examples of both types of networks, one with only LV and the other one with MV and LV.

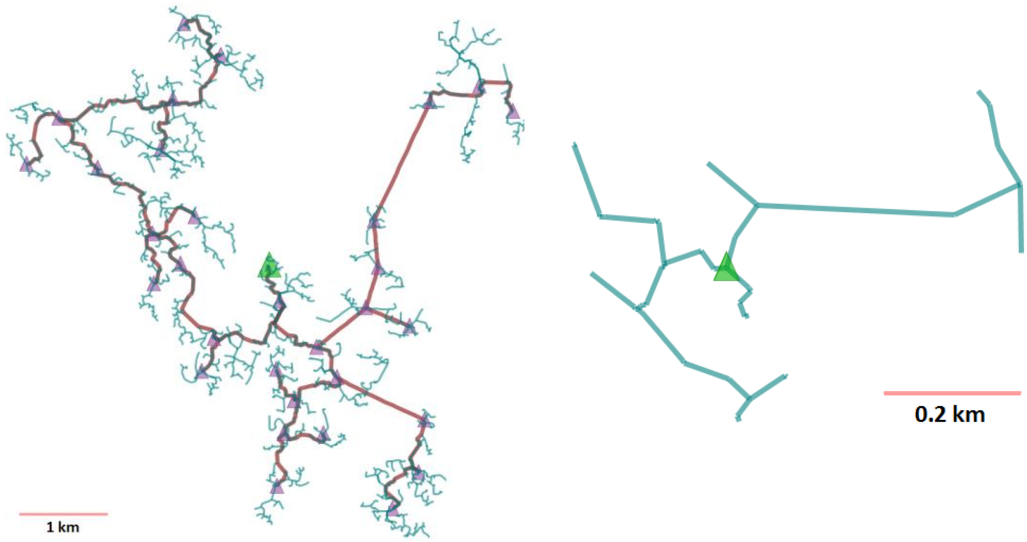


Figure 3-14: Two examples of mini-grid network layouts. MV line (red line), LV line (blue line), generation (green triangle), MV/LV transformer (purple triangle). © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

REM uses RNM only once when calculating a grid extension design. RNM computes the MV and LV distribution networks of the corresponding grid extension. Since REM uses RNM to calculate distribution networks and RNM provides three-phase networks, the electrification solutions that REM provides always include three-phase networks.

3.1.5. **Financial model**

This section describes how REM performs economic calculations. REM considers direct monetary costs and indirect societal costs in the economic evaluation of electrification plans.

3.1.5.1. Direct monetary costs

The direct monetary costs include investments in physical assets and diverse kinds of operating expenditures. The costs are categorized as investment costs, operations and maintenance costs, and energy costs. REM discounts future expenditures based on the appropriate discount rate for each technology to account for the time value of money. By allowing different technologies to be discounted with independent discount rates, REM accounts for the diverse ownership structures and risk profiles that are possible in the studies to be performed with the model. For instance, utility-owned grid-extension projects would typically have a lower discount rate than privately-owned solar home systems.

REM is a static optimization planning model, which determines the minimum-cost solution for just a future snapshot situation, i.e., one year in the future. Due to the wide range of equipment lifetimes used in the electrification space, an annuity for that future year is computed for each technology. This allows accounting for shorter-lived products, such as batteries or solar home systems, and assets with long economic lives, like lines and transformers in the distribution networks.

For expenditures occurring on a non-annual basis, the expenditure is converted to a yearly annuity A:

$$A = C \cdot r/(1 - (1 + r)^{-L}) \tag{3.1}$$

Where C is the periodic expenditure, r is the discount rate, and L is the period. Direct monetary costs include investment, operation and maintenance, and energy costs:

- Investment costs, or Capital Expenditures (CAPEX), are determined directly from the system design and associated cost catalog. For grid extension, this is often dominated by the cost/km of the distribution network, whereas for mini-grids, the $/kW for solar PV and $/kWh for battery storage are often the most significant components of the total cost of supply. As mentioned above, these capital expenditures are converted to annuities to compare projects in the "static optimization" REM.

- Each equipment type is assigned an annual operational and maintenance cost based on the local equipment characteristics and necessary expenditures to maintain equipment in working condition. For distribution lines, this is defined by ($/km)/year, but for transformers, batteries or diesel generation sets, this is defined simply as $/year for a given piece of equipment.

- Direct energy costs happen in grid-extension projects, and in mini-grids with diesel generation. For grid-extension projects, the cost per kWh of energy is the wholesale electricity price when delivered at the MV level (which includes the price of energy at the wholesale level, plus transmission and HV distribution costs). REM currently assumes a flat cost of electricity regardless of hour-of-day or time-of-year, and the amount of energy demanded. REM accepts as different input values of this wholesale electricity cost at different connection points, as lines may need diverse reinforcements. For mini-grid systems with diesel generation, the price of diesel fuel is the only energy cost.

3.1.5.2. Non-monetary costs

The least-cost optimization performed by REM also includes non-monetary costs. The main societal cost is the CNSE, which is associated with the reliability of supply. REM imposes this penalty on a per-kWh basis for every unit of energy demand that is not supplied. This penalizing factor ensures that system reliability is adequately accounted for while making sure that supply does not become prohibitively expensive since the direct monetary costs quickly grow with higher reliability levels.

A single value of CNSE cannot capture the diversity of situations of supply failure, as perceived by consumers with different needs and at different times. As a reasonable approximation to this complex reality, REM distinguishes between critical and non-critical loads for all consumers and applies a different value of CNSE to the curtailment of critical and non-critical demand. Critical and non-critical demand profiles are specific for each type of consumer and demand pattern, although REM cannot apply different values of CNSE for different consumer types. Determination of the appropriate value for CNSE is a difficult task, which would require extensive and well-designed surveys of the involved consumers. An alternative approach is modifying the values of CNSE as inputs to REM until the model delivers an acceptable combination of cost and reliability of supply.

3.2. A kaleidoscope of improvements

This section presents the most significant improvements made in this thesis to turn the basic first prototype of REM (Ellman, 2015), which conceptually contained all the basic ingredients of the approach but was incapable of producing a workable design, into a fine-tuned computer tool that – after overcoming a number of major challenges – produces reliable results for the electrification of territories of basically any size. The improvements are classified according to the three algorithmic blocks of Figure 3-2, which represent the main algorithms in REM (mini-grid generation, clustering, and final designs). We place the symbols , , and near the titles of the subsections to refer to improvements related to the block of mini-grid generation, clustering and final designs, respectively.

This section also describes improvements that expanded the functionalities of the first REM prototype, such as adding solar kits as a viable electrification solution or enabling REM to deal with several types of consumers. We place the symbol near the titles of the subsections that describe improvements related to the algorithms of the first prototype of REM, and we set the symbol close to the titles of subsections that present new functionalities such as the addition of solar kits or enabling REM to work with several consumer types.

3.2.1. **Mini-grid generation**

Two improvements were implemented to enhance the optimality of the results. The first one (master-slave decomposition) modified the hierarchy of variables of the algorithm that optimizes the generation design of a point of the look-up table from scratch. The second one (dispatch strategy) changed the dispatch strategy in REM. These two improvements were essential for the proper performance of REM.

This section presents two additional improvements related to expanding the capabilities of REM. The first one (addition of multiple types of consumers) increases the number of consumers types REM can handle. The second one (synthetic demand patterns), currently under development, focuses on reducing the number of representative off-grid systems that REM needs to calculate to compute the look-up table. Synthetic demand patterns will be useful in cases with many consumer types once the work is complete.

3.2.1.1. Master-Slave decomposition

The generation sizing algorithm of the first prototype of REM operates in a three-dimensional search space where all technologies considered (solar panels, batteries, diesel generator) are considered equally important in the algorithm.

In the search space, there is a significant difference among the axis related to the diesel capacity and the remaining axes. REM selects multiples of an individual solar panel or battery to optimize the generation design of an off-grid system, allowing the operation of panels and batteries in parallel racks. The possible values that the solar panels and the batteries could take in the search space are given by combining several units or a single solar panel or battery in a row, so the capacities of solar panels and batteries are multiples of a single element.

However, the use of diesel generators in parallel is more complex, and REM does not allow this possibility. Therefore, the diesel capacities available are limited to the units available in the generation catalog, and they are generally not multiples of a single diesel generator. Moving from a diesel generator to the immediately bigger or smaller one could produce significant variations of capacity (i.e., from 4 kW to 5 kW, or from 5 kW to 10 kW).

REM handles this difference among the off-grid technologies with a master-slave decomposition where the master level controls the diesel capacity, and the slave level explores a solar-battery plane with a fixed diesel capacity, from an adequate starting search point provided by the master level. The strategy of exploiting the structure of an optimization problem with a nested decomposition has been successfully applied to other problems in the literature (Liu and Zhang, 2014; Prada y Nogueira, 2017).

The master algorithm starts with a 100% renewable solution (no diesel), and the slave problem finds the least-cost design in the no-diesel plane. Then, the master problem increases the diesel capacity to the next diesel generator available on the catalog, and the slave problem finds the combination of solar and battery that better fits the demand for that diesel generator. This procedure continues until the master problem has considered all diesel generators available in the diesel search space, which includes all diesel generators between the no-diesel solution and the smallest diesel generator that can meet all the demand. The starting search point in terms of solar and battery capacity is the best solution obtained at the previous iteration (for the previous diesel generator). Figure 3-15 provides an example with diesel generators of 0, 5, 8, 10, and 15 kW.

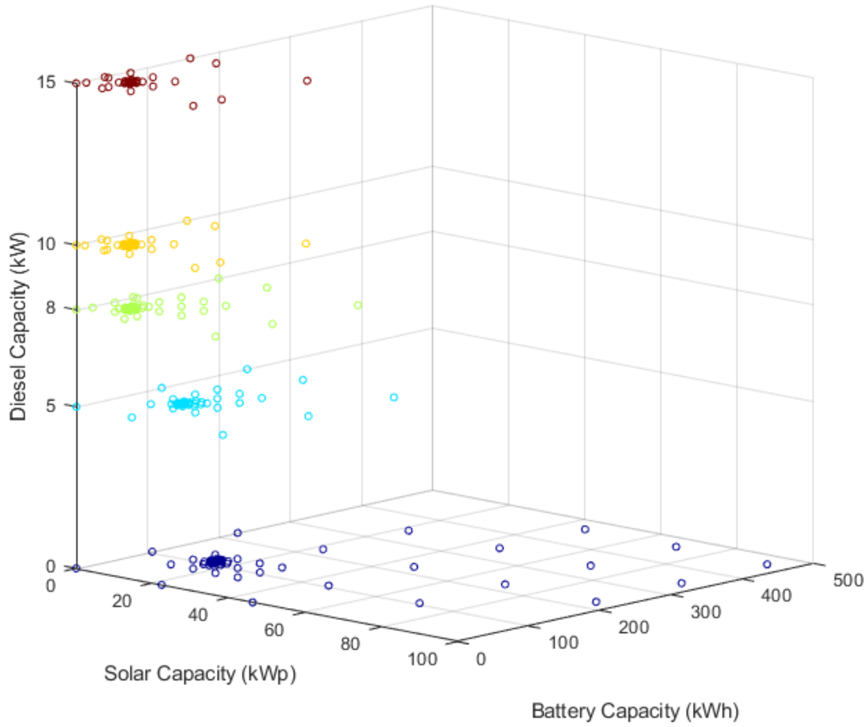


Figure 3-15: Candidate generation designs that the algorithm considers. As the diesel capacity increases, candidate designs have fewer solar panels and batteries. Source: (Ciller et al., 2019b).

Figure 3-16 shows the flow diagram of the master problem. The master problem performs an exhaustive search as it goes through all the available diesel generators in the diesel search space (whose upper bound is the smallest diesel generator that can meet all the demand on its own). This procedure requires a substantial amount of computation time in cases where the number of diesel generators to be evaluated is high.

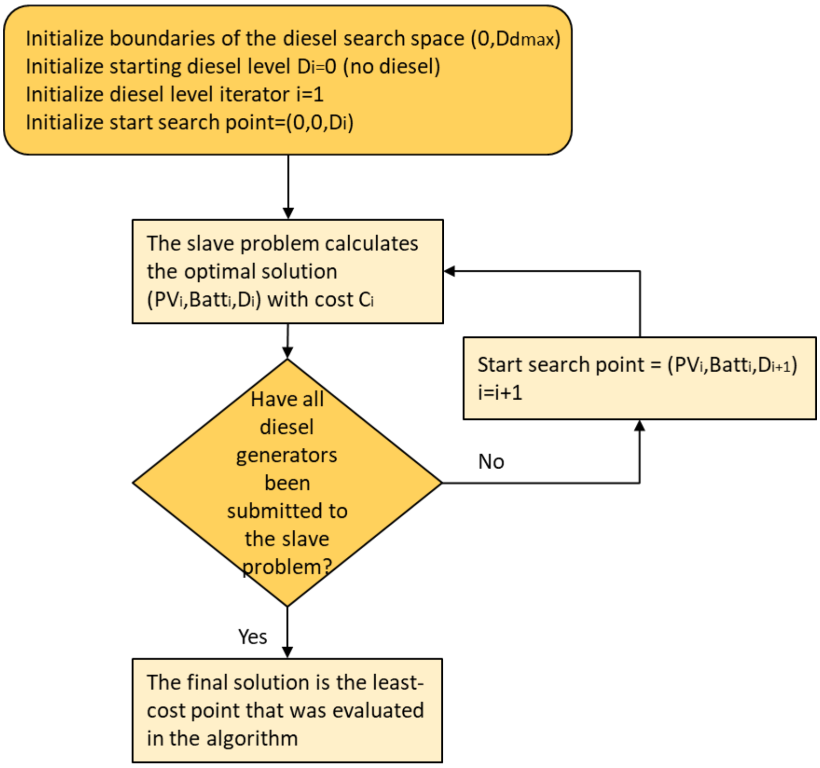


Figure 3-16: Flow diagram of the master problem. D_{dmax} is defined as the smallest diesel generator that can meet the demand of the system. Source: (Ciller et al., 2019b).

The slave problem searches in the neighborhood of an initial central point provided by the master problem, moving towards the neighbor point with minimum cost. If no neighboring point improves the current solution, the algorithm reduces the step size until its value is below a pre-specified threshold. Figure 3-17 shows an example of the slave problem, which corresponds to the first iterations for the no-diesel plane shown in Figure 3-15.

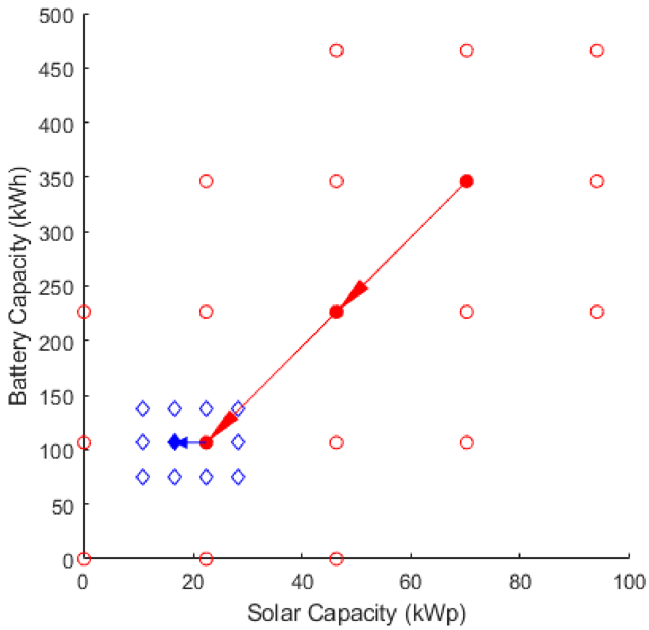


Figure 3-17: Slave problem example. The algorithm moves two times to a neighbor point with lower cost, reduces the step length, and moves another time. Source: (Ciller et al., 2019b).

Figure 3-18 shows the flow diagram of the slave problem. The slave problem, in practice, behaves as a gradient-descent method with discrete derivatives since it always moves in the lowest-cost direction.

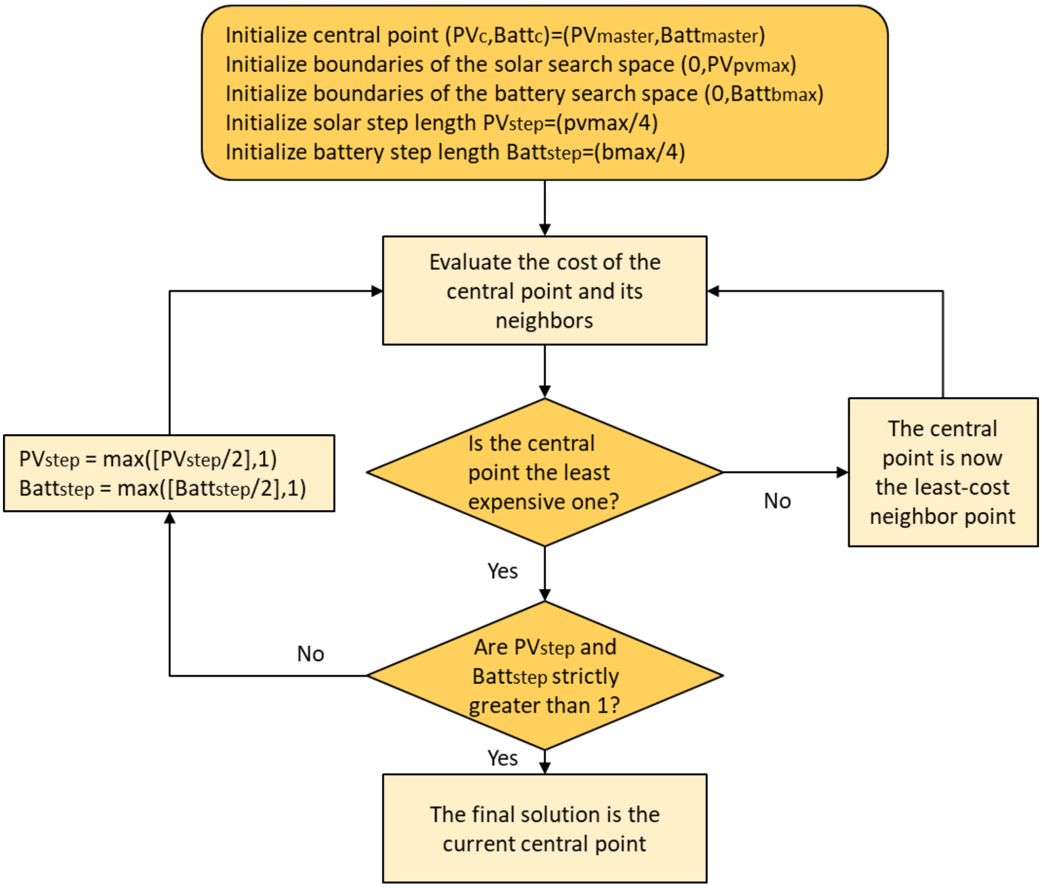


Figure 3-18: Flow diagram of the slave problem. The upper boundaries of the solar and battery search spaces are defined as ten times the average demand. Source: (Ciller et al., 2019b).

Table 3-3 shows several generation designs obtained with the generation sizing algorithm that REM's first prototype applies. The generation cost per consumer for 100 residential consumers is significantly higher than the total cost per consumer for 50 and 150 residential consumers. Similarly, the generation cost per consumer for 200 residential consumers is substantially higher than the total cost per consumer for 150 and 250 residential consumers. The master-slave decomposition will show that the generation sizing algorithm of the first prototype of REM reaches a local minimum when it optimizes the generation design of 100 and 200 residential consumers.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	1	0	273.48	10.81
5	1.46	13.32	0	1	0	147.05	17.32
10	2.93	24.42	0	1	0	129.87	27.19
50	5.36	4.44	10	1	0.71	84.62	30.67
100	14.63	73.26	15	1	0.46	94.59	34.56
150	16.09	17.76	30	1	0.7	77.67	46.19
200	29.25	144.30	30	1	0.46	92.79	37.44
250	26.81	26.64	50	1	0.71	77.56	41.94
300	31.10	31.08	60	1	0.71	77.26	45.29
500	52.55	48.84	100	1	0.71	73.86	69.58
1,000	104.03	139.86	200	1	0.69	70.32	64.92
3,000	254.48	419.58	600	1	0.7	69.46	70.07
7,500	777.56	1,052.28	1,500	1	0.69	69.23	61.95

Table 3-3: Designs with the initial prototype of REM.

Table 3-4 shows several generation designs, which correspond to the same residential consumers as the designs of Table 3-3, obtained with the master-slave decomposition presented in this section. The generation cost per consumer decreases as the number of residential consumers grows, which is a coherent result that shows that the master-slave decomposition offers a more robust performance than the generation sizing algorithm of the first prototype of REM.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	1	0	273.48	**20.54**
5	1.37	11.10	0	0.99	0	140.07	**32.65**
10	2.73	24.42	0	1	0	124.68	**31.33**
50	3.90	8.88	9	1	0.68	78.34	**114.34**
100	9.75	13.32	20	1	0.68	75.74	**188.57**
150	14.63	17.76	30	1	0.69	74.59	**189.75**
200	19.50	22.20	40	1	0.69	74.6	**201.62**
250	24.38	26.64	50	1	0.7	74.68	**212.58**
300	31.10	31.08	60	1	0.7	74.45	**233.41**
500	47.78	73.26	90	1	0.68	69.27	**260.89**
1,000	85.12	139.86	175	1	0.69	67.64	**310.87**
3,000	282.75	419.58	600	1	0.68	66.69	**392.18**
7,500	706.88	1,052.28	1,500	1	0.68	66.55	**432.32**

Table 3-4: Designs with the master-slave decomposition.

The master-slave decomposition requires higher computation times than the generation sizing algorithm of the first REM prototype because it performs an exhaustive search of the diesel generators available, evaluating more diesel generators than the generation sizing algorithm of the first prototype of REM. However, the computation times related to the master-slave decomposition are still very affordable, and results show that the increase in computation time is justified.

Figure 3-19 compares the generation costs provided in Table 3-3 and Table 3-4 up to 500 residential consumers, showing that results improved by changing the generation sizing algorithm of the first prototype of REM, which assigned the same importance to all the components involved in the optimization process, to a master-slave nested decomposition where the diesel capacity is controlled in the master level and the capacities of solar panels and batteries are controlled in the slave level.

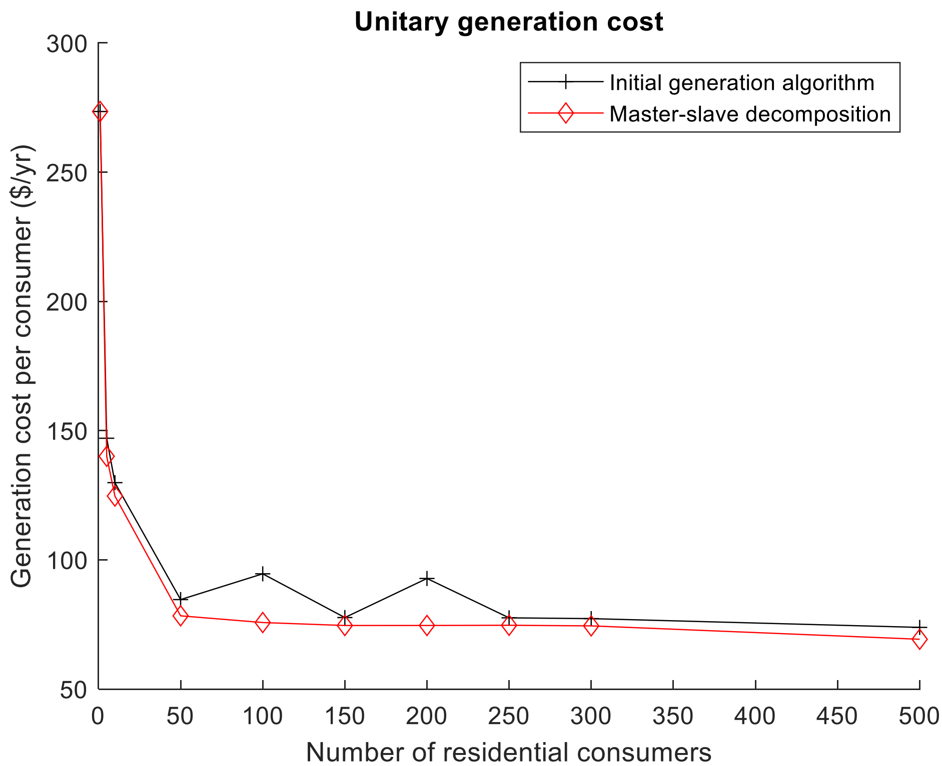


Figure 3-19: Generation sizing comparison of results.

The load following strategy is applied to determine the hourly dispatch of the candidate generation designs that REM evaluates to obtain the generation designs shown in Table 3-3 and Table 3-4.

The load following procedure uses first the solar energy to meet the demand and, if the battery is not fully charged, the remaining solar energy is used to charge it. If there is not enough solar energy, and the battery is not fully discharged, then the battery is used to meet the demand. If there is still unserved demand after using the battery, then REM uses the diesel generator or allows some non-served energy (least-cost decision, depending on the penalties for non-served energy and the marginal diesel cost).

In the next section, we perform a thorough cost-comparison among the battery valuation strategy and the load following dispatch.

3.2.1.2. Dispatch strategy

We thoroughly evaluate the cost differences among the battery valuation strategy and the load following dispatch. For each cost-comparison, the generation design and the aggregated demand are the same to ensure that the different dispatch strategies are the only reason behind the different costs.

We perform the cost-comparisons by applying the master-slave decomposition introduced in section 3.1.2.1 to optimize the generation designs for the combinations of residential consumers presented in Table 3-4. REM computes the generation cost of each candidate generation design with the battery valuation strategy and the load following dispatch. As expected, generation designs that do not include batteries lead to the same operation of the system with both dispatch strategies so they are not included in the cost-comparisons.

Equation 3-2 defines the relative cost difference between the two dispatch strategies. Equation 3-2 implies that the relative cost difference is positive when the load following dispatch outperforms the battery valuation strategy, and the relative cost difference is negative when the battery valuation strategy performs better than the load following dispatch.

$$Relative\ cost\ difference = \frac{(Cost\ battery\ valuation - Cost\ load\ following)}{Cost\ load\ following} \tag{3-2}$$

Figure 3-20 shows the relative cost difference among the two dispatch strategies for all the generation designs that REM evaluated. Most generation designs perform better with the load following dispatch strategy than with the battery valuation strategy. However, the battery valuation strategy leads to a lower total cost than the load following dispatch in some cases.

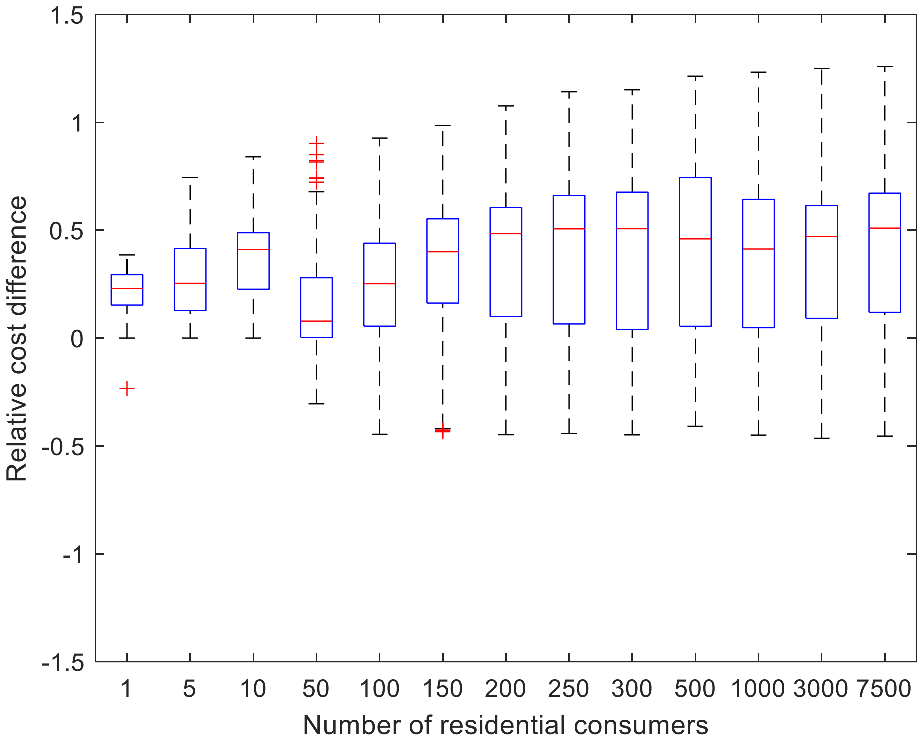


Figure 3-20: Cost-comparisons among the dispatch strategies.

We have analyzed the generation designs where the battery valuation strategy performs better than the load following dispatch, and they generally correspond to designs that cannot meet all the demand unless the diesel generator is used to charge the battery. In these cases, using the diesel generator to charge the battery (as the battery valuation strategy may do) reduces the amount of non-served energy, leading to a better dispatch than the load following strategy (which never uses the diesel generator to charge the battery). For example, a generation design that includes a diesel generator and batteries but no solar panels will never use the batteries with the load following dispatch. Still, the battery valuation strategy will benefit from having batteries available.

Table 3-5 shows the number of evaluated generation designs for each combination of residential consumers and the fraction of designs where the load following dispatch provides a better cost than the battery valuation strategy. The load following dispatch performs better than the battery valuation strategy in at least 78% generation designs.

Residential Consumers	1	5	10	50	100	150	200	250	300	500	1,000	3,000	7,500
Number of designs	33	79	71	278	482	515	567	591	662	710	927	1,154	1,259
Fraction of designs with a positive relative cost difference	0.97	1	1	0.78	0.86	0.86	0.82	0.82	0.81	0.83	0.83	0.86	0.86

Table 3-5: Number of designs and fraction of designs where the load following dispatch outperforms the battery valuation strategy.

We also apply the master-slave decomposition introduced in section 3.1.2.1 to optimize the generation designs for the combinations of residential consumers of Table 3-4. This time, REM simulates the hourly dispatch of each candidate generation design with the battery valuation strategy[4]. Table 3-6 shows the corresponding results.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	0.86	0	370.51	**20.68**
5	1.27	8.88	0	0.77	0	264.43	**29.50**
10	1.37	0.00	5	0.98	0.73	146.83	**35.06**
50	4.88	4.44	9	1	0.74	80.64	**127.80**
100	10.73	6.66	20	1	0.74	79.33	**220.48**
150	14.63	17.76	30	1	0.73	76.66	**231.53**
200	21.45	33.30	40	1	0.72	75.95	**237.49**
250	26.81	39.96	50	1	0.72	75.39	**257.27**
300	31.10	46.62	60	1	0.73	75.06	**267.80**
500	52.55	73.26	90	1	0.73	70.63	**342.46**
1,000	104.03	139.86	175	1	0.72	67.65	**408.62**
3,000	311.03	139.86	600	1	0.75	70.21	**482.31**
7,500	777.56	1,052.28	1,500	1	0.73	69.32	**540.85**

Table 3-6: Designs with the master-slave decomposition and the battery valuation strategy.

Figure 3-21 compares the generation costs shown in Table 3-4 and Table 3-6. The master-slave decomposition provides better results with the load following dispatch than with the battery valuation strategy for one, five, and ten residential consumers. In these cases, the battery valuation strategy leaves a significant amount of non-served demand if it does not include a diesel generator.

[4] *There is an essential difference between the procedures followed to obtain Figure 3-20 and Table 3-6. The procedure used to obtain Figure 3-20 uses the operational cost that the load following dispatch provides to evaluate the generation, and the cost related to the battery valuation strategy is not used in the algorithm (i.e., the load following dispatch "guides" the master-slave decomposition). However, the procedure used to obtain Table 3-6 uses the battery valuation strategy to obtain the operational cost (i.e., the battery valuation strategy "guides" the master-slave decomposition).*

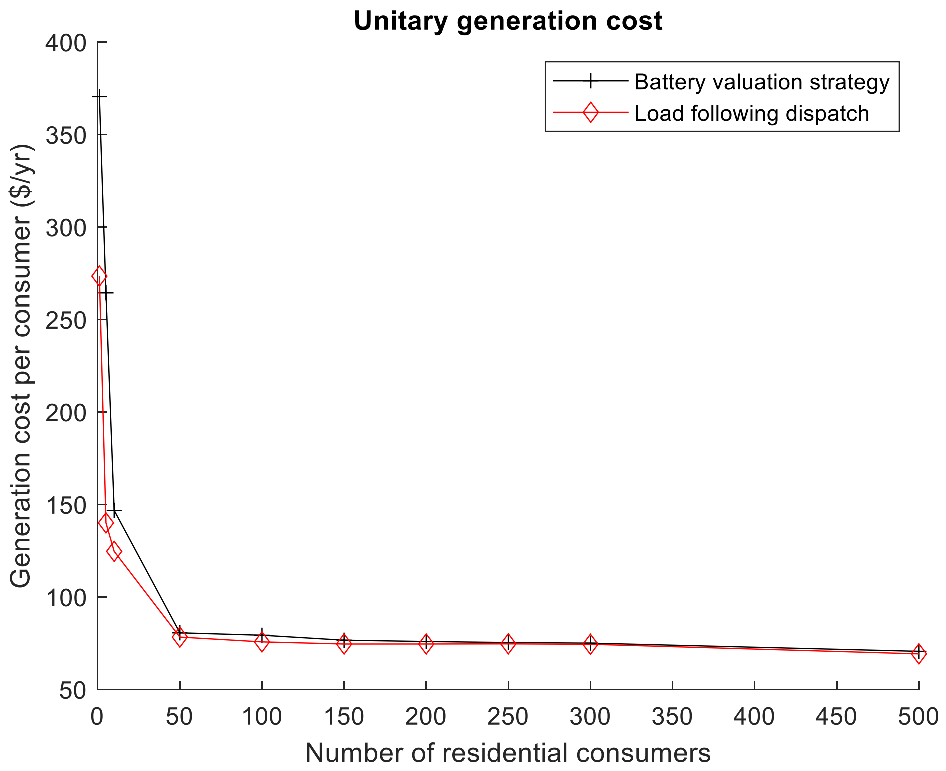


Figure 3-21: Master-slave decomposition with both dispatch strategies.

The battery valuation strategy also has disadvantages that are not present in the load following dispatch. One such disadvantage is the oscillating behavior (the diesel generator and the battery alternate to meet the demand in consecutive hours) that may appear in some generation designs, and the diesel generator charges the battery in the hours it is used to meet the demand (see Figure 3-22). It should be noted that REM interpolates among the hourly values of the demand and mini-grid components when it plots a dispatch, giving a false impression about the battery being charged and discharged simultaneously.

This behavior is not only suboptimal (it would be more economical to use only the battery or the diesel generator to meet the demand during that timeframe) but also difficult to defend from a practical point of view.

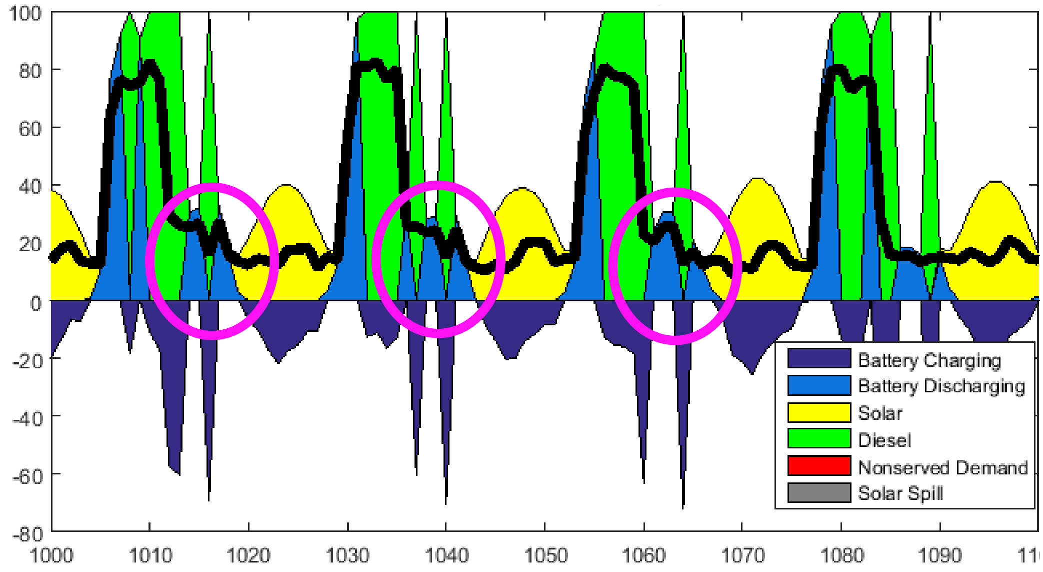


Figure 3-22: Hourly dispatch with the battery valuation strategy. Timeframes where the diesel generator and the battery are used alternatively to meet the demand are highlighted with pink circles. The x-axis is measured in hours, and the y-axis is measured in kW.

The analysis provided in this section justifies changing the battery valuation strategy with the load following dispatch, which is a robust method that has been thoroughly used in generation sizing tools such as HOMER (HOMER Energy LLC, 2019b). The dispatch change was a practical decision that improved REM's performance, but the two dispatch strategies studied in this section are valuable heuristic methods.

3.2.1.3. Addition of multiple types of consumers

The first prototype of REM operates with one type of consumer, which limits significantly the scope of the analysis performed as productive loads cannot be included appropriately. The look-up table was extended to include several types of consumers at the expense of increasing the number of representative off-grid systems that REM needs to calculate.

Specifically, adding a new consumer type to the look-up table implies adding one more dimension in the space of consumers, and the total number of generation designs that REM would need to calculate to be able to perform a satisfactory interpolation could be significantly larger.

Figure 3-23 shows an example with three consumer types, where the representative mini-grids are obtained as combinations of 0, 300, 700, and 1000 residential consumers, 0, 1, 5, and 10 hospitals, and 0, 1, 5, and 10 schools. REM would calculate generation designs for 4 × 4 x 4 = 64 representative off-grid systems, which correspond to the points in Figure 3-23.

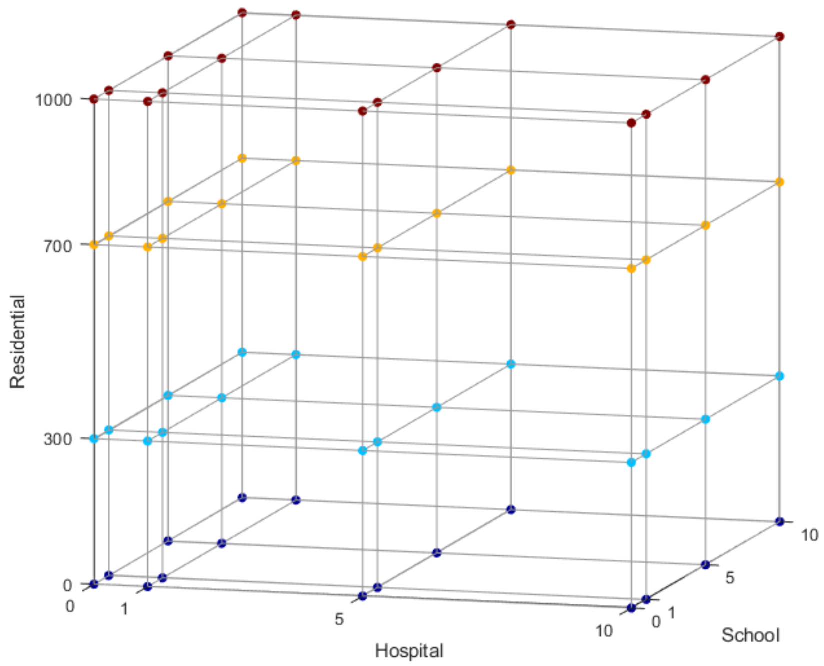


Figure 3-23: Look-up table example with three consumer types. Source: (Ciller et al., 2019b).

The number of consumer types that REM can process limits the look-up table approach. A case example with five different consumer types would imply a penta-dimensional look-up table that would require computing too many points. This limitation can be overcome by realizing that generation designs are related to the aggregated demand of candidate mini-grids and not to specific combinations of consumer types. REM can associate the axes of the look-up table to "demand patterns" instead of consumer types. By doing so, REM can operate with a number of consumer types that is higher than the number of axes of the look-up table, as far as the demand of any consumer type can be expressed as a linear combination of a set of "basic" demand patterns.

For example, we could have two different types of residential households (big and small) and assume that the demand of a big household is five times the demand of a small household. Then REM could have one demand pattern related to the small household profile, and the point "5" of the look-up table could be either five small households or one big household since the demand is the same in both cases.

It seems logical to explore this idea, using dimension-reduction techniques that synthesize a large number of demand profiles related to consumer types into a few basic demand patterns associated with axes of the look-up table. Both the basic demand patterns and the linear combinations are specific inputs to REM. Although the user can manually calculate basic demand patterns, the task is far from easy in cases with several productive loads. Therefore, it is interesting to devote some efforts to develop a computational procedure that distills a reduced number of basic demand patterns from the demand profiles of the consumers.

3.2.1.4. Synthetic demand patterns (currently under development)

We use the term “synthetic demand patterns” to refer to the basic demand patterns when they are obtained with an automated dimensionality reduction technique. In contrast to the load profiles of consumers, the synthetic demand patterns may not have a physical meaning because they are only used as a tool to limit the number of dimensions of the look-up table. Once a few – three, for instance – synthetic demand patterns have been computed, they would allow REM to create a look-up table with a reduced number of dimensions – three, in our example – even in cases with a significant amount of load types.

An attempt to calculate synthetic demand patterns, which is currently under development, relies on applying Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that extracts a set of linearly independent features (which are the principal components) that summarize the data. PCA is a technique that has been applied to problems from a wide range of fields, such as load forecasting (Manera and Marzullo, 2005; Yingying and Dongxiao, 2010; Xiao-fei and Li-qun, 2016).

We can apply PCA to decompose the demand profiles of the consumers into a linear combination of principal components (which, in this particular case, are the synthetic demand patterns) plus an average term. Figure 3-24 shows the PCA approximation of a residential profile with three synthetic demand patterns (which are the principal components), where PCA was applied to a set of 120 different demand profiles.

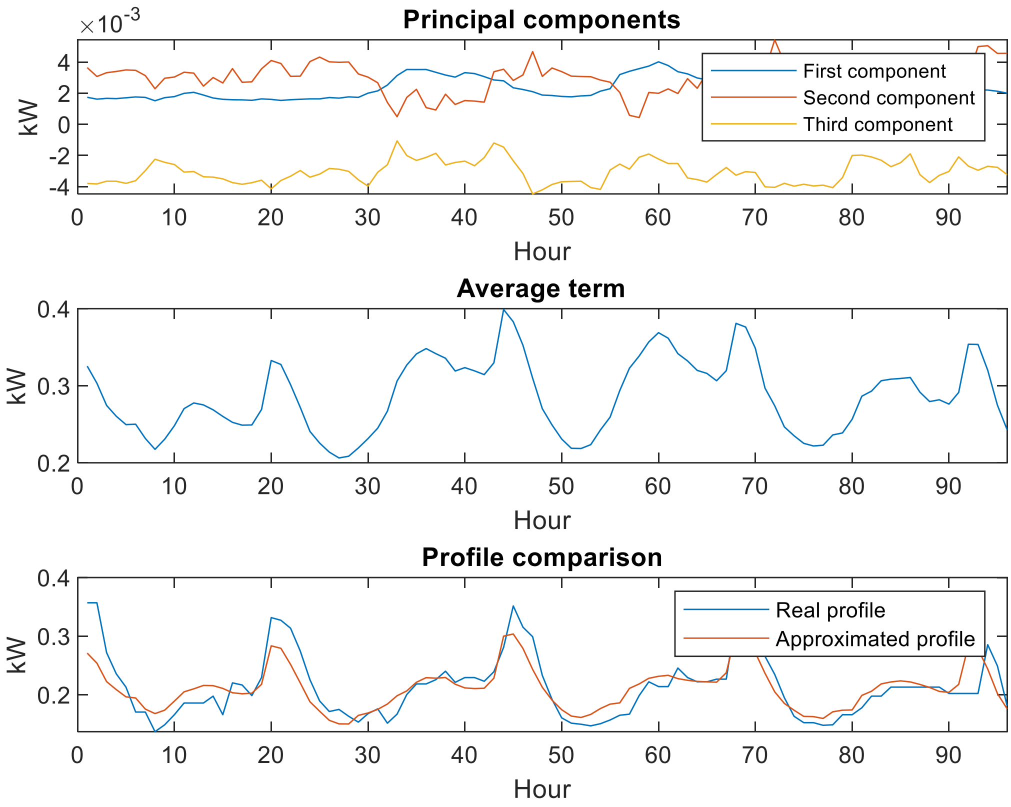


Figure 3-24: Calculation of synthetic demand patterns with PCA. The figure shows the principal components (top), the average term (middle), and the real and approximated profiles (bottom). The approximated profile is a linear combination of the principal components plus the average term.

The synthetic demand patterns obtained with PCA present one major flaw: the demand consumption may be negative at some hours (see the third component in Figure 3-24). There is not a simple way of dealing with a negative demand when simulating the hourly dispatch of a mini-grid, so the points of the look-up table need to be carefully selected to ensure that none of them has a negative demand. The automated determination of the points of the look-up table poses an additional challenge that is yet to be overcome, as the user determines the representative off-grid systems whose generation costs are stored in the look-up table in the current version of REM.

3.2.2. **Clustering**

The clustering step plays a crucial role in the final electrification solution, and many efforts were devoted to analyze the first clustering algorithm and improve its performance. The enhancements presented in this section are classified into three groups: the modeling improvements, the improvements to the logic of the algorithm, and the interpolation improvements.

The interpolation improvements are the most relevant ones, as they avoid clustering solutions with many small clusters that are rather suboptimal. The case study presented in section 3.3 shows that the final solution depends heavily on the clustering results and the impact of a suboptimal clustering in the final electrification solution.

3.2.2.1. Modeling improvements

The modeling improvements include:

- The addition of technical network losses in the cost calculation of lines that connect two clusters or a cluster with the power grid. This enhancement improves the level of realism considered when calculating the cost of lines.

- The addition of management costs as part of the total cost of the clusters. The management costs are not only considered in the clustering but also considered in the final electrification solution.

- The use of peak demands and center-to-center distances to size the network elements. The first prototype of REM used average demands and minimum consumer-to-consumer distances.

- The suppression of an alternative clustering logic that was present in the first prototype of REM. This logic omitted configurations in the grid-extension clustering where one cluster is electrified with an off-grid system.

3.2.2.1.1. Addition of technical network losses

The first prototype of REM did not include technical network losses when estimating the cost of a line that joins two clusters or a cluster to the power grid. The addition of the cost of losses is translated into a more realistic cost estimation, although technical losses generally account for a small percentage of the total energy (less than 10% in developed countries and around 20% in developing countries (Mahmood et al., 2014)). Equation 3-3 provides the cost related to power losses in a three-phase line for a year (Pande and Ghodekar, 2012):

$$C_{loss} = 3 \cdot I^2 \cdot r \cdot L \cdot 8760 \cdot load\ loss\ factor \cdot cost\ of\ losses \qquad \text{3-3}$$

Where C_{loss} is the cost associated with power losses (\$/year), I is the maximum load of the line (A), r is the resistance of the line per unit of length (Ω/km), L is the length of the line (km), $cost\ of\ losses$ is the costs of losses parameter (\$/kWh), and the $load\ loss\ factor$ term is the load loss factor of the line. The cost of losses is set to the energy cost of the power grid in the case of a line that joins two grid-extension clusters or a grid-extension cluster to the grid, and it is interpolated in the look-up table in the case of a line that joins two off-grid clusters.

The load loss factor is defined as the ratio between the average power losses ($P_{average}$) and the losses at maximum load (P_{max}) in a period T (Pande and Ghodekar, 2012). The load loss factor is calculated with equation 3-4:

$$load\ loss\ factor = \frac{P_{average}}{P_{max}} = \frac{1}{P_{max}} \cdot \frac{\int_0^T P(t)dt}{T} \qquad (3\text{-}4)$$

Where $P(t)$ the instantaneous demand power losses and $\int_0^T P(t)dt$ the energy losses of the system during T.

Figure 3-25 shows the technical energy losses as a fraction of the total energy for the activated connections (lines) in the off-grid clustering of the case study presented in section 3.3. Losses account for less than 20% of the total energy for approximately 95% of the lines.

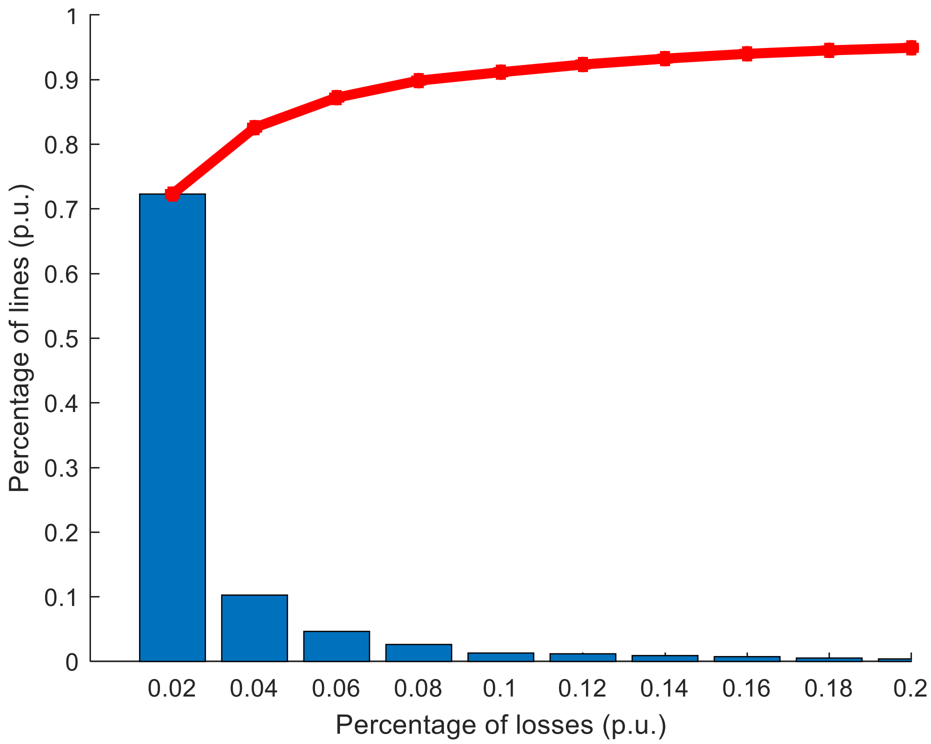


Figure 3-25: Losses of a line as a fraction of the energy of the line.

3.2.2.1.2. Addition of management costs

Management costs, which were not considered in the first prototype of REM, were included in the cost-comparisons performed in the clustering algorithm. Annual management costs differ by system type and size due to the nature of different pieces of equipment and different business structures. When modeling the management cost for grid-extension projects and solar home systems, REM assumes that economies of scale have been reached, and the marginal management cost of each additional consumer is uniform. When considering the management cost associated with mini-grids, REM assumes that each mini-grid will have some fixed management cost, plus a monotonically decreasing marginal cost per additional consumer. In this way, the model acknowledges the economies of scale associated with mini-grids of increasing size.

In the case of off-grid systems, the user introduces the per-consumer management cost of three off-grid systems (they are expected to have a low, medium, and high number of consumers) and how many consumers they have (it is not necessary to introduce the number of consumers of the large off-grid system as REM can use the total number of consumers in the case). Table 3-7 shows an example of the input data related to management costs.

Number of consumers small-size off-grid system	1
Number of consumers medium-size off-grid system	150
Annual management per-consumer cost of the small-size off-grid system ($/yr)	60
Annual management per-consumer cost of the medium-size off-grid system ($/yr)	16
Annual management per-consumer cost of the large-size off-grid system ($/yr)	9

Table 3-7: Example of input parameters related to management costs.

REM uses the parameters shown in Table 3-7 to adjust the parameters A, B, k of the function $f(m)$ introduced in equation 3-5, whose independent variable is the number of consumers m of an off-grid system.

$$f(m) = \frac{A\left(1-e^{-\frac{m}{k}}\right)}{m} + B \qquad \text{3-5}$$

Equation 3-5 is similar to the one presented in reference (Carrasco et al., 2013) to estimate the management costs, and it approximates other methodologies regarding the calculation of management costs as a function of the number of consumers (Comisión de Regulación de Energía y Gas, 2020).

Figure 3-26 shows an example of the management cost curve calculated with the numeric values from Table 3-7. The management cost curve approaches 9 $/yr asymptotically since REM associates that cost to an infinitely large mini-grid.

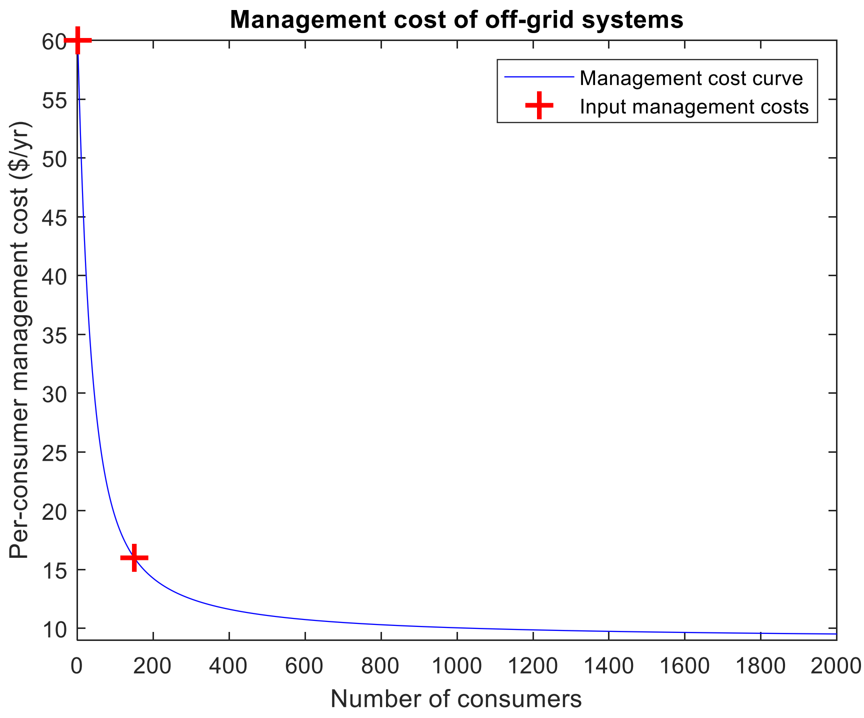


Figure 3-26: Example of the management cost curve.

For grid-extension clusters, REM assumes that the management cost per consumer is constant and equal to the management cost per consumer of an infinitely large mini-grid. Following the example, REM would consider that the management cost of each consumer in a grid-extension cluster is equal to 9 $/yr.

The management costs were also introduced in the cost-comparisons that REM performs to explore the hierarchical structure of clusters in the final designs stage.

3.2.2.1.3. Use of peak demands and center-to-center distances

The initial prototype of REM used the average demand and the shortest consumer-to-consumer distance to size the lines that estimate the incremental network cost. The peak demand of the clusters replaced the average demand, and the center-to-center distance replaced the shortest consumer-to-consumer distance.

3.2.2.1.4. Suppression of alternative logic for the grid-extension clustering

The initial prototype of REM included two logics for the grid-extension clustering algorithm. The first logic evaluated the cost related to configurations 1-5 from Figure 3-9 and Figure 3-10. The second logic only considered configurations 1, 2, and 5 (so it did not include the configurations where a cluster can be electrified with a mini-grid in the grid-extension clustering).

The addition of configurations that reflect viable electrification should lead to better clustering results. REM can electrify nearby clusters with different electrification modes in the final solution, so configurations where a mini-grid electrifies one cluster and a grid extension electrifies the other cluster are always included in the grid-extension clustering now. Therefore, the second logic is not present in the current REM version, although it provided better solutions than the first logic in some cases (Ellman, 2015).

The second logic occasionally provided better results because the clustering process of the first prototype of REM presented several flaws, which are identified and corrected in section 3.2.2. The suppression of several configurations masked the issues in some cases, but it did not help overcome the real problems.

3.2.2.2. Improvements to the logic of the algorithm

The improvements to the logic of the algorithm aim at enhancing the robustness of the algorithm, and they include:

- the addition of more candidate connections

- the addition of more loops through the candidate connections

- the inclusion of more configurations in the grid-extension clustering

3.2.2.2.1. Addition of candidate connections

In the first prototype of REM, the clustering candidate connections are the arcs of the MST that connects all the consumer. However, the MST often misses candidate connections among nearby consumers that are worth considering. REM now obtains the clustering candidate connections with the Delaunay triangulation.

The Delaunay triangulation is formed by a set of non-overlapping triangles whose vertices are the consumers, and that covers the convex hull (i.e., the convex set of minimum area) of the consumers. Besides, the Delaunay triangulation has several properties that capture the vicinity relations among consumers, being the most relevant properties the following ones (Peco, 2001):

1. The minimum spanning tree of a set of consumers is contained in the Delaunay triangulation of that set of consumers. This property guarantees that the potential connections that the first prototype of REM considers are still considered when the Delaunay triangulation is applied.
2. The arcs of the Delaunay triangulation do not intersect.
3. Each consumer is connected with its nearest consumers.

The Delaunay triangulation has already been used to identify potential connections in clustering algorithms related to distribution networks (Mateo Domingo et al., 2011; Navarro and Rudnick, 2009).

If we have a case with n consumers, then the minimum spanning tree provides $n-1$ candidate connections and the Delaunay triangulation provides $3n-3-h=3(n-1)-h$ connections, being h the number of arcs that form the convex hull of the set of consumers (Peco, 2001).

In the case described in section 3.3, which has 6,688 consumers, the minimum spanning tree provides 6,687 candidate connections and the Delaunay triangulation provides 20,037 candidate connections, so this improvement almost triples the number of candidate connections. Figure 3-27 shows a comparison among the clustering candidate connections obtained with the Delaunay triangulation and the MST for the case study described in section 3.3.

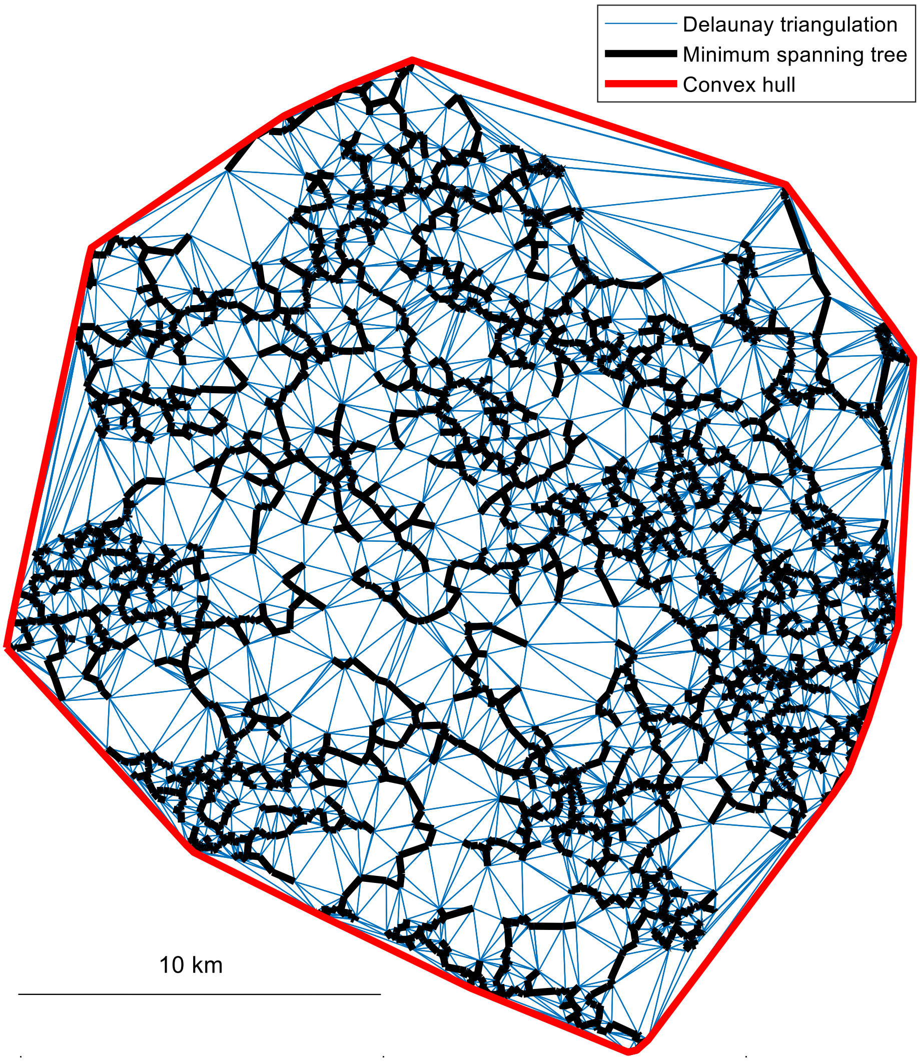


Figure 3-27: Example that compares the clustering candidate connections provided by a Delaunay triangulation and the MST of the consumers.

3.2.2.2.2. Additional loops through the candidate connections

The off-grid and on-grid clustering algorithms now loop through all the unconnected candidate connections until no new connection is activated in a loop. Each time a connection is activated, the loop starts again from the beginning. In the initial clustering, the candidate connections were evaluated only once, not considering connections worth activating later.

The off-grid clustering evaluates 9,319 potential connections among clusters in the case study presented in section 3.3 if it performs a single loop among the Delaunay triangulation arcs. In contrast, the off-grid clustering evaluates 18,209 potential connections if it performs multiple loops, restarting the loop each time a connection is activated. Therefore, this improvement approximately doubles the number of candidate connections evaluated for the case study of section 3.3.

We should note that the clustering algorithm may not loop through all the candidate connections available even if it performs a single loop among them. As clusters grow, multiple candidate connections may connect two clusters, and if the evaluation of one such connection implies the merge of the clusters, then it is not necessary to evaluate the remaining connections among these two clusters. Hence, it makes sense that the off-grid cluster only considers 9,319 potential connections when it performs a single loop among the Delaunay triangulation arcs, which includes 20,037 potential connections.

3.2.2.2.3. Additional configurations in the grid-extension clustering

Finally, two configurations were included in the grid-extension clustering: configurations 1’ and 2’ (see Figure 3-28) were not considered in the initial clustering algorithm. Those configurations account for cases where the grid-extension clusters are connected with an MV line, and each cluster has its own transformer.

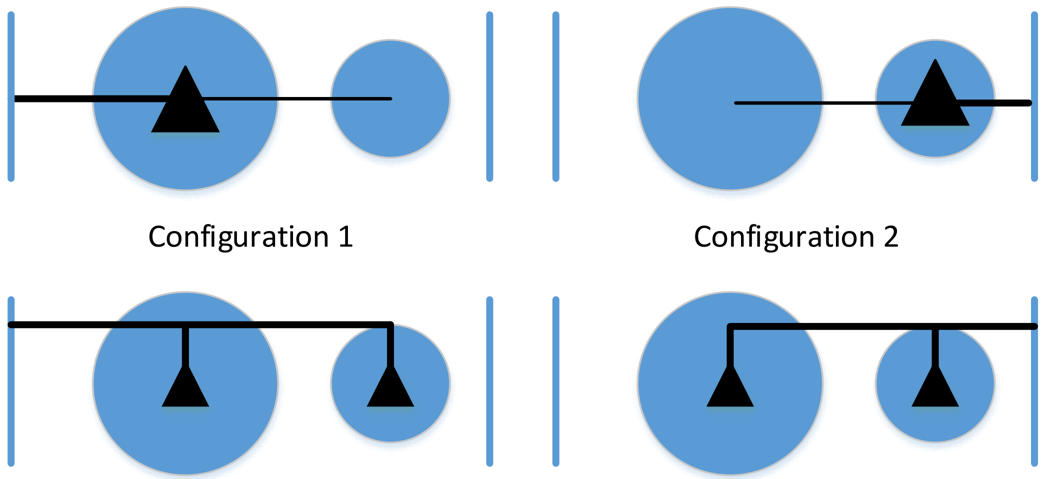


Figure 3-28: Set of configurations that support merging grid-extension clusters in REM after the improvements. © 2019 IEEE. Reprinted, with permission, from (Ciller et al., 2019a).

3.2.2.3. Interpolation improvements

REM’s clustering follows a bottom-up approach where each consumer starts as a single cluster, and nearby clusters may be connected based on a local cost-comparison. In the off-grid clustering, the trade-offs between generation, management, and network cost are evaluated to determine if two nearby clusters are better electrified together. Large off-grid clusters benefit from savings related to economies of scale in generation and management, but they have significant network costs. If the initial cost-comparisons among small clusters determine that it is better not to join the clusters, then large off-grid clusters will not be created even if they are part of the optimal clustering solution.

The interpolation improvements aim at measuring the trade-offs among costs accurately so that the off-grid clustering does not reach a local minimum with a solution that includes many small clusters. The interpolation improvements include the initial smoothing of the generation costs and the use of a continuous network catalog for the calculation of the incremental network costs.

- The first improvement (smoothing) focuses on economies of scale in generation. The impact of the generation costs in the off-grid clustering is thoroughly explained in chapter 4, but we describe in this section the initial fixes that ensured that a non-monotonous unitary generation cost did not lead to issues in the clustering.

- The second improvement (continuous network catalog) focuses on incremental network costs, which are estimated with a line that connects the clusters, or a cluster with the power grid. The use of a continuous network catalog aims at improving the estimation of the incremental network cost that was used in the first prototype of REM.

3.2.2.3.1. Initial smoothing

The curve obtained by interpolating the unitary generation costs stored in the look-up table may not be monotonic, even after the implementation of the improvements described in section 3.2.1. This section introduces two methods to ensure that the unitary generation costs curve used in the clustering follow a monotonic behavior.

The first method is simple but effective. REM checks that the second point of the look-up table has a unitary generation cost strictly lower than the first point of the look-up table. If this condition does not hold, REM sets the unitary generation cost of the second point to 99% of the unitary generation cost of the first point (so the unitary generation cost of the second point is strictly lower than the unitary generation cost of the first point). Then, REM evaluates whether the third point of the look-up table has a unitary generation cost strictly lower than the second point of the look-up table. If this is not the case, REM sets the unitary generation cost of the third point to 99% of the unitary generation cost of the second point. This procedure continues looping through the remaining points of the look-up table until REM ensures that the $i-th$ point of the look-up table has a unitary generation cost strictly lower than the $(i-1)-th$ point.

Figure 3-29 shows the results of applying the first smoothing method to the generation results shown in Table 3-3 (up to 500 residential consumers), which were obtained before implementing the master-slave decomposition presented in section 3.2.1.1. The smoothing procedure guarantees that the unitary generation cost strictly decreases when the number of residential consumers increases.

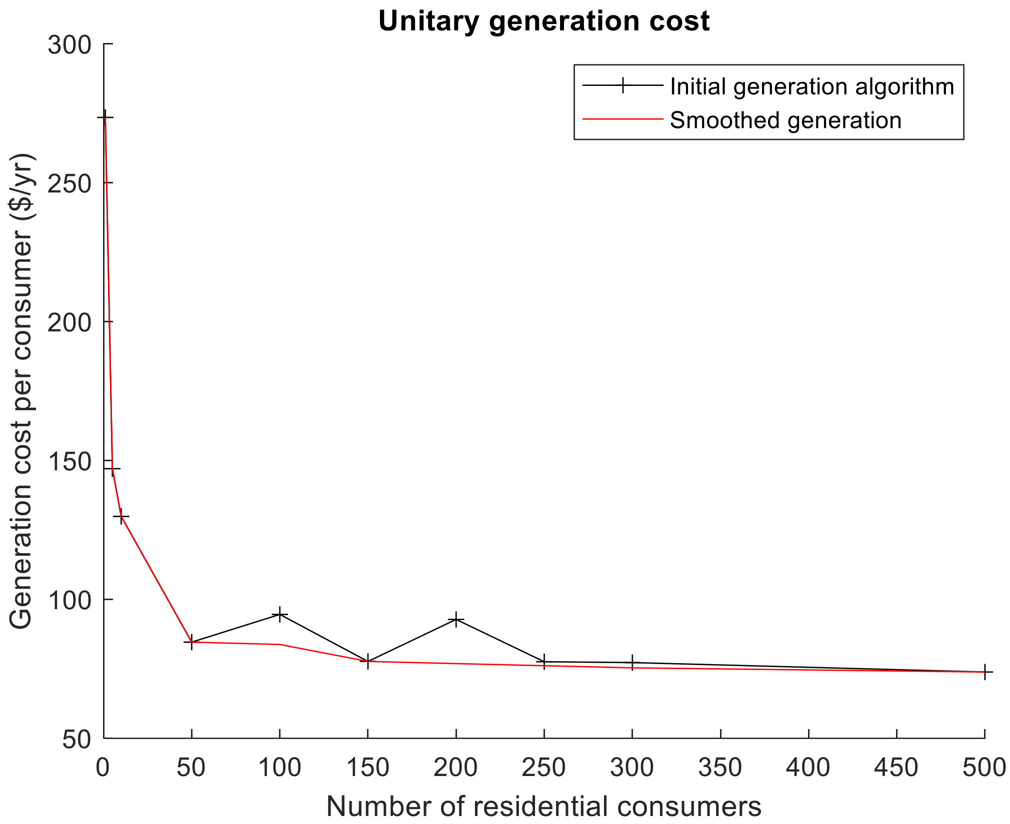


Figure 3-29: Application of the first smoothing method.

The second method adjusts the unitary generation costs with a piecewise exponential function whose coefficients were calculated with the points of the look-up table. Although the idea of adjusting the unitary generation cost with a strictly decreasing function is in the right direction, the initial attempts did not provide consistent results in all cases. The current version of REM can use a different type of curve to smooth the generation costs; further details are provided in chapter 4 of this thesis.

Figure 3-30 shows an example where the first method ("Partially smoothed") and the second method ("Final smoothed") are applied. The first prototype of REM could only operate with one type of load, so the extrapolation of these methods for several types of loads was out of the scope at that moment.

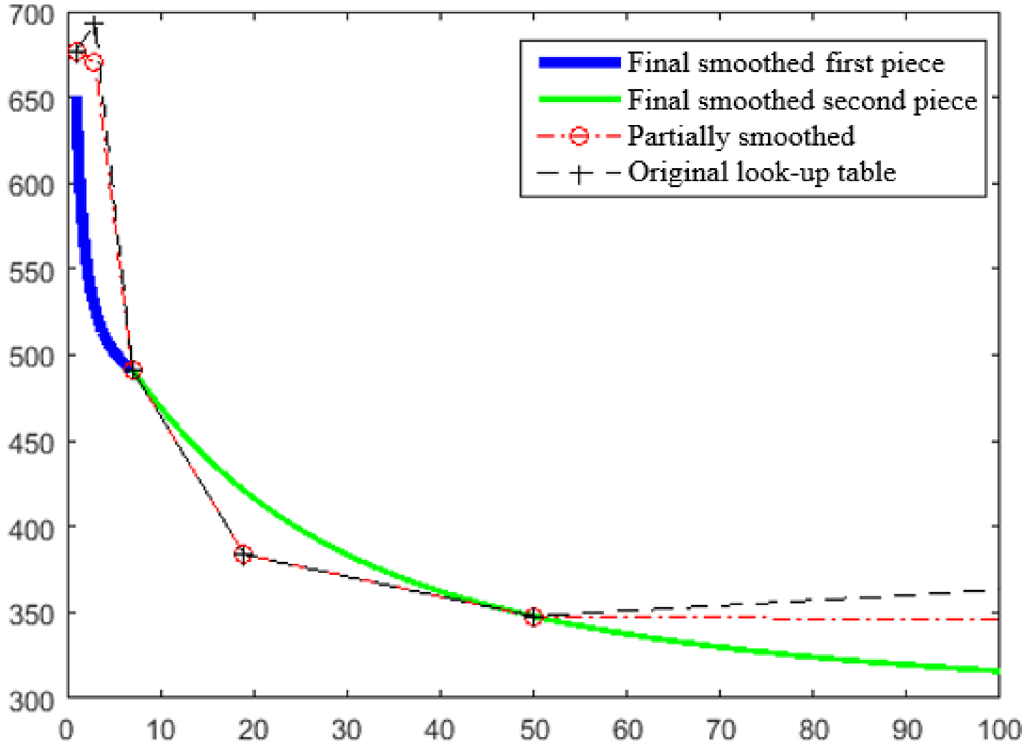


Figure 3-30: Initial attempts to smooth the unitary generation costs. Source: adapted from (Ciller Cutillas, 2016). The x-axis measures the number of residential consumers, and the y-axis measures the unitary generation cost ($/yr).

3.2.2.3.2. Continuous network catalog

We refer to the network catalogs used to calculate the network layouts of mini-grids and grid-extension designs as "discrete" because it includes a finite number of elements. The lowest-capacity line of the discrete network catalog is generally too expensive to compensate for the generation savings of two small clusters being electrified together. If that happened, the initial connections would never be activated in the off-grid clustering, and the clustering would never explore solutions that benefit from economies of scale in generation.

The use of a continuous network catalog prevents this issue (we refer to the network catalog obtained interpolating and extrapolating the costs of the discrete network catalog as "continuous" because it has elements of any desired positive capacity). Figure 3-31 shows the annual cost of an LV line with a discrete and continuous catalog. The least-cost line in the discrete catalog has a capacity of 135 A, which is oversized for residential demands in developing countries.

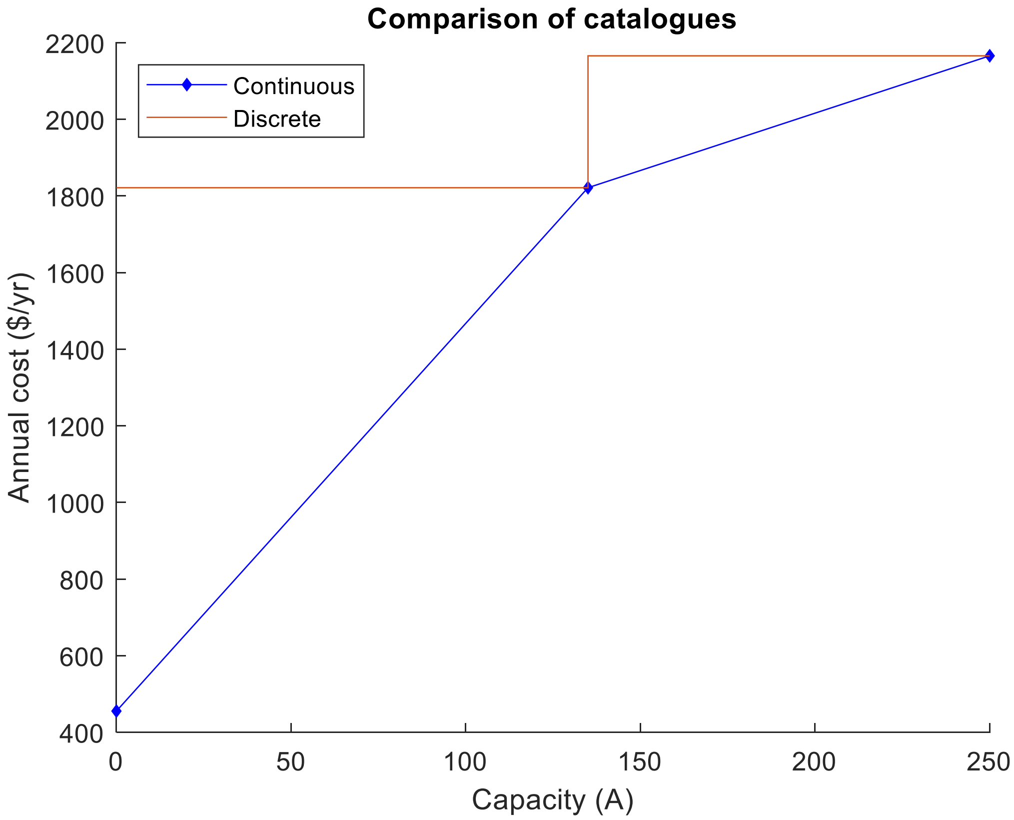


Figure 3-31: Costs of the discrete and continuous network catalogs of LV lines in the clustering.

REM does not set the cost of the zero-capacity LV line to zero to reach a balance between the clustering logic (which is not compatible with a discrete network catalog) and the real cost of the networks (which are calculated with a discrete catalog). Specifically, the zero-capacity LV line cost is set to one-fourth of the cost of the lowest-capacity LV line from the discrete catalog. This number was adjusted evaluating the performance of the clustering in many cases, and it has provided reasonable results so far. The optimization of the value that this number takes is beyond the scope of this thesis.

The clustering algorithm uses quick estimations of the network cost to determine whether to join the clusters, and we acknowledge that the cost of a line that connects two clusters or a cluster with the power grid is not the best method to estimate the incremental network cost. It was necessary to develop a new method (which is presented in chapter 5) that estimates the network costs accurately to continue enhancing the clustering of REM.

The continuous catalogs of transformers and MV lines were calculated from the corresponding discrete catalogs following the same procedure.

The continuous catalogs are a reasonable approximation to the real (discrete) catalogs, and they are used only for clustering purposes. The network designs that appear in the final electrification solutions are calculated with RNM and discrete components, so using a continuous catalog in the clustering does not decrease the level of modeling realism of the final solution that REM provides.

3.2.3. **Final designs**

Four enhancements were implemented in the calculation of final designs. The first three aim at improving the optimality of REM's algorithms. The first one (evaluation of off-grid clusters with grid extensions) calculates grid-extension designs for off-grid clusters. The second one (calculation of LV and MV mini-grids) changes the use of RNM to calculate the networks for the mini-grids. The third one (more candidate connection points) includes additional candidate connection points for the network calculation of grid extensions.

The fourth improvement presented in this section (addition of solar kits) enhances REM capabilities, including DC solar kits as a viable electrification solution.

3.2.3.1. Evaluation of off-grid clusters with grid extensions

The first prototype of REM evaluated the hierarchical structure of electrification clusters considering only grid-extension designs for grid-extension clusters, mini-grid designs for off-grid clusters, and isolated systems for isolated clusters (which are the individual consumers). However, REM now allows a broader exploration of the clustering solution, where off-grid clusters may also be evaluated as grid-extension designs.

Designing extensions of the power grid for the off-grid clusters may seem counterintuitive at first (off-grid clusters were calculated without considering the grid), but it is better to proceed this way. Off-grid clusters may include productive loads such as large factories, and their best electrification solution could be a grid extension.

Although the exploration of more alternatives implies increasing the computation time, it offers a more robust behavior as some small clusters may include productive loads with high levels of consumption, and they could be better electrified with grid-extension designs. Table 3-8 presents the computation times related to evaluating the hierarchical structure of clusters of the case study presented in section 3.3. The solution of the case study that REM provides after the improvements only includes off-grid systems, so the evaluation of off-grid clusters as grid extensions does not reduce the final electrification cost in this case.

Off-grid clusters are not evaluated with grid extensions	Off-grid clusters are evaluated with grid extensions
14 min, 05.60 sec	17 min, 36.30 sec

Table 3-8: computation times needed to evaluate the hierarchical structure of clusters.

REM minimizes the additional computation time by quickly obtaining a lower bound of the cost of a grid extension for an off-grid cluster. If the lower bound is higher than the cost of electrifying the cluster with off-grid alternatives, it is clear that a grid extension will not be the least-cost solution for the off-grid cluster. In that case, REM does not apply RNM to calculate the grid extension's detailed network layout (this calculation accounts for the majority of the computation time needed to obtain the grid extension).

3.2.3.2. Calculation of LV and MV mini-grids

The initial prototype of REM used RNM once to calculate the network design of an individual mini-grid, and its voltage level (i.e., LV or MV) depended on how many transformers RNM placed in the design.

The initial strategy caused inconsistencies, such as having a mini-grid where cost decreases when its demand increases because REM places a larger-capacity transformer instead of several lower-capacity transformers. Figure 3-32 shows an example where REM interpreted the network design of a mini-grid as MV because RNM includes two transformers in the network design (left). However, the network design includes only one transformer when the demand was multiplied by five, and REM interpreted it as an LV network design (right).

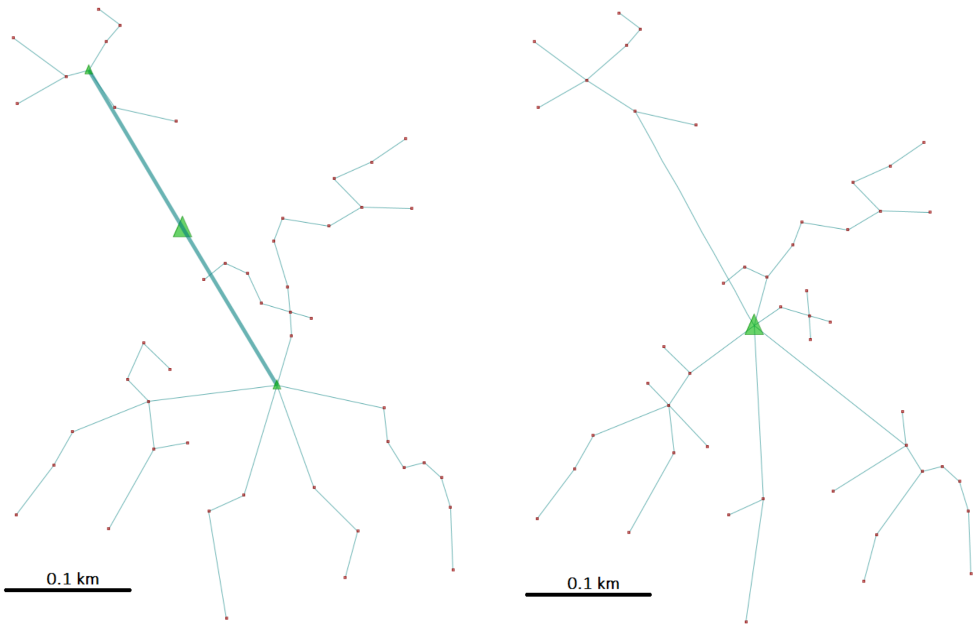


Figure 3-32: Mini-grid designs for two demand levels. Big green triangles represent the generation sites, small green triangles represent transformers, thick blue lines represent MV lines, thin blue lines represent LV lines, and red dots represent consumers.

Table 3-9 includes the costs that REM's initial prototype considered for the network designs shown in Figure 3-32. The total network cost decreases 38.30% when the demand is multiplied by five because the first prototype of REM did not apply RNM properly to calculate the network designs of mini-grids.

	Demand x1	Demand x5	Δ (%)
LV network cost ($/yr)	514	1,485	188.91
Distribution transformers cost ($/yr)	759	0	-100
MV network cost ($/yr)	1,134	0	-100
Total network cost ($/yr)	2,407	1,485	-38.30

Table 3-9: Evolution of the network cost that the first prototype of REM considered when a mini-grid demand is multiplied by five. The last column contains the percentual increment between the first and second columns of the table.

To avoid potential issues such as the one shown in Figure 3-32, REM now uses RNM twice with different configuration parameters to obtain the least-cost network design for a mini-grid. REM assumes that all the mini-grids have an LV generation system and evaluates two possible networks:

- LV network. REM assumes that the mini-grid has an LV distribution network. Generation is also connected in LV, so no transformers are needed.
- MV and LV network. REM assumes that the mini-grid has an MV network, and MV/LV transformers with LV sub-networks to reach the consumers. Generation is assumed to be connected at LV, so an extra MV/LV transformer is needed to feed the MV network.

There are two additional improvements regarding the network designs of mini-grids:

- When RNM calculates an LV network design, there should be a transformer in the catalog that can satisfy the demand of the entire mini-grid. If RNM cannot find a transformer that meets the demand of the mini-grid, then it will gradually decrease the peak demand of the consumers until it can obtain a solution (which may not be electrically feasible since RNM reduced the peak demands). The current version of REM includes this transformer in the catalog when needed.

- REM always assumes that the generation site operates in LV, so it is necessary to add the cost of an additional transformer located in the generation site to raise the voltage level in MV mini-grids. The current version of REM includes the cost of this transformer, which was not considered in the first prototype of REM.

3.2.3.3. More candidate connection points for grid extensions

The number of candidate network points considered for grid-extension designs was limited to 10 in the first prototype of REM. It was increased to 150, and the distance between any pair of candidate connection points has to be higher than a pre-specified threshold (100 m). The addition of more candidate network points improved the grid-extension designs for big clusters, which sometimes need to be connected to different lines.

Figure 3-33 shows an extension of the power grid where REM considers ten candidate connection points, although additional connection points would be worth considering. The already-existing MV lines are represented with black lines, and the location of the candidate connection points considered is highlighted with a thick black segment.

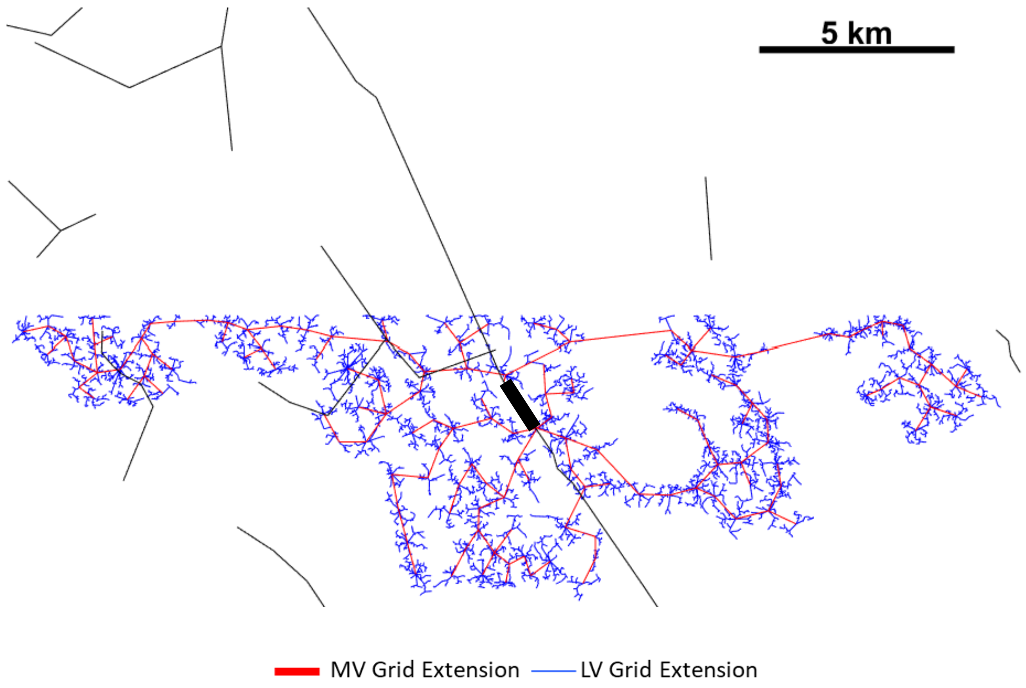


Figure 3-33: Case example with ten candidate connection points.

Figure 3-34 shows the same case example with 150 candidate connection points to the existing power grid. The additional candidate connection points allow REM to consider several lines to calculate the power grid extension, providing a realistic design.

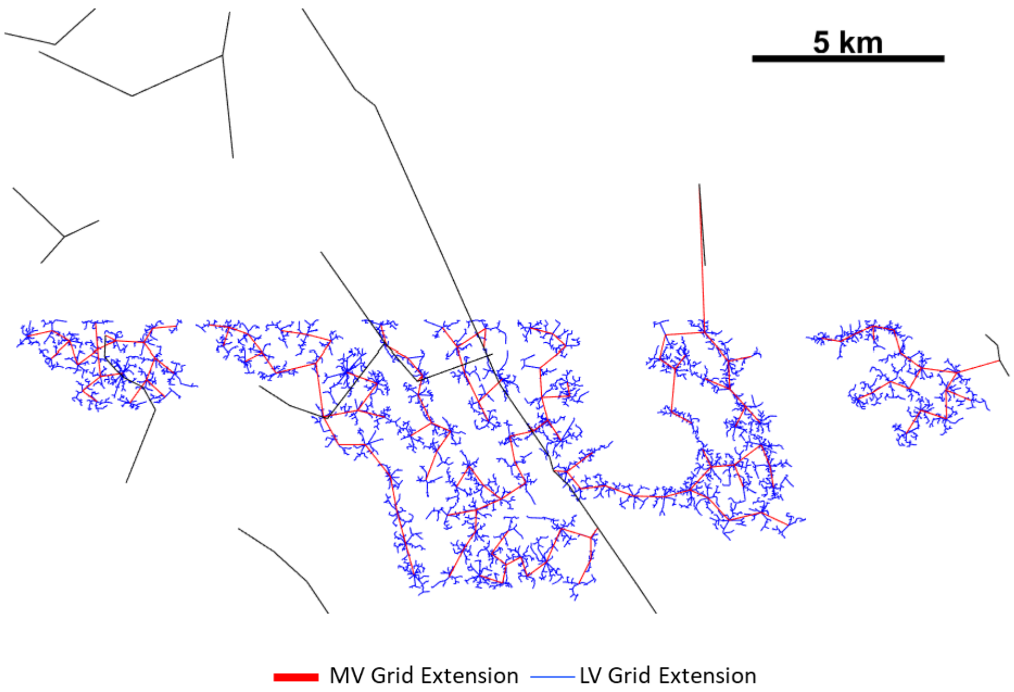


Figure 3-34: Case example with 150 candidate connection points.

Table 3-10 compares the network designs shown in Figure 3-33 and Figure 3-34. This example illustrates the importance of adequately selecting the candidate connection points for grid extensions because the MV network cost decreases by almost 15% when the number of candidate connection points rises from 10 to 150.

	10 candidate connection points	150 candidate connection points	Δ (%)
LV network cost ($/yr)	183,404	183,404	0
Distribution transformers cost ($/yr)	171,561	171,561	0
MV network cost ($/yr)	205,177	175,219	-14.60

Table 3-10: Network designs with 10 and 150 candidate connection points. The last column contains the percentual increment between the first and second columns of the table.

3.2.3.4. Addition of solar kits

The first prototype of REM evaluates the cost of electrifying individual consumers with AC generation systems when the model performs cost-comparisons among the hierarchical structure of clusters, but the current version of REM can also consider DC solar kits as an electrification alternative for individual consumers.

For low levels of demand, DC solar kits could be preferred to AC generation systems as the electrification solution for small isolated consumers. Although AC standalone systems can provide more energy, solar kits are more portable and less expensive, especially when it comes to operation and maintenance costs and they could suffice for the small demands of many poor households. This implies that solar kits are an option that is worth considering in an electrification plan (Sun, 2017).

Solar kits generally meet the demand of essential household appliances, so they tend to provide a lower level of utility to the consumer than the remaining electrification alternatives (i.e., AC individual systems, mini-grids, and extensions of the power grid). The CNSE is critical in solar kits, which have a low investment and operation cost (although the cost per kWh of demand served is usually high in solar kits).

Solar kits have limitations that are not present in remaining electrification solutions. One such limitation is their inability to handle demand growth. Extensions of the power grid and mini-grids can cope with additional demand, reinforcing the upstream network (in the case of grid extensions) or including additional generation modules (in the case of AC individual systems and mini-grids) if needed. However, solar kits have a fixed generation that cannot cope with the additional demand.

An additional limitation of solar kits concerns the availability of supply. Solar kits, whose components are usually a solar panel and a battery, generally provide less energy than AC systems, and only for a few hours a day. The energy consumption of solar kits users typically differs from their expected demand, and solar kit users adapt their electricity consumption to the availability of energy in terms of peak energy and total available hours of use. The remaining electrification alternatives generally do not impose such availability of supply constraints on the users.

As a consequence of these limitations, we consider that the supply service that a solar kit provides is different from the remaining electrification alternatives, and REM allows the user to introduce a different non-served energy penalty only for solar kits to represent this difference.

There is an additional practical reason that justifies that solar kits use a different CNSE, which is to allows the planner to control the penetration of solar kits in an electrification plan. The optimal penetration of solar kits may depend on political constraints, budget restrictions, and subsidizes.

The solar kits option is not fully compatible with a bottom-up clustering strategy because they may break the monotonicity of the economy of scale function that makes clusters grow. If REM applies the clustering algorithm described in section 3.1.3 (with the improvements included in section 3.2.2), then it only considers solar kits as an electrification option in the final designs phase (when deciding the best electrification mode of isolated consumers). Chapter 6 presents a new off-grid clustering algorithm that can handle solar kits and other constraints that distort the monotonicity of economies of scale in generation.

Note that the comparison between solar kits and the other two delivery modes is not straightforward. The choice of solar kits cannot be determined solely based on the single parameter of the CNSE. It can be expensive to achieve with solar kits' reliability levels that are comparable to those that can be more economically obtained with mini-grids or a reliable grid connection. However, solar kits with what many households might consider an acceptable reliability level exist at modest prices and with attractive financing schemes. Solar kits are individually managed without any external interference, and the availability of power can be focused on the individual household priorities, but they generally supply low electricity intensity appliances and their application to community or productive loads is less extended.

3.3. Case study

We apply REM (before and after the improvements) to a case study located in the region of Cajamarca, in Northern Peru. This region has an area of approximately 400 km^2, around 6,700 buildings, and several potential connection points to the projected network of 11 kV. Figure 3-35 shows a map of Peru and the Cajamarca region.

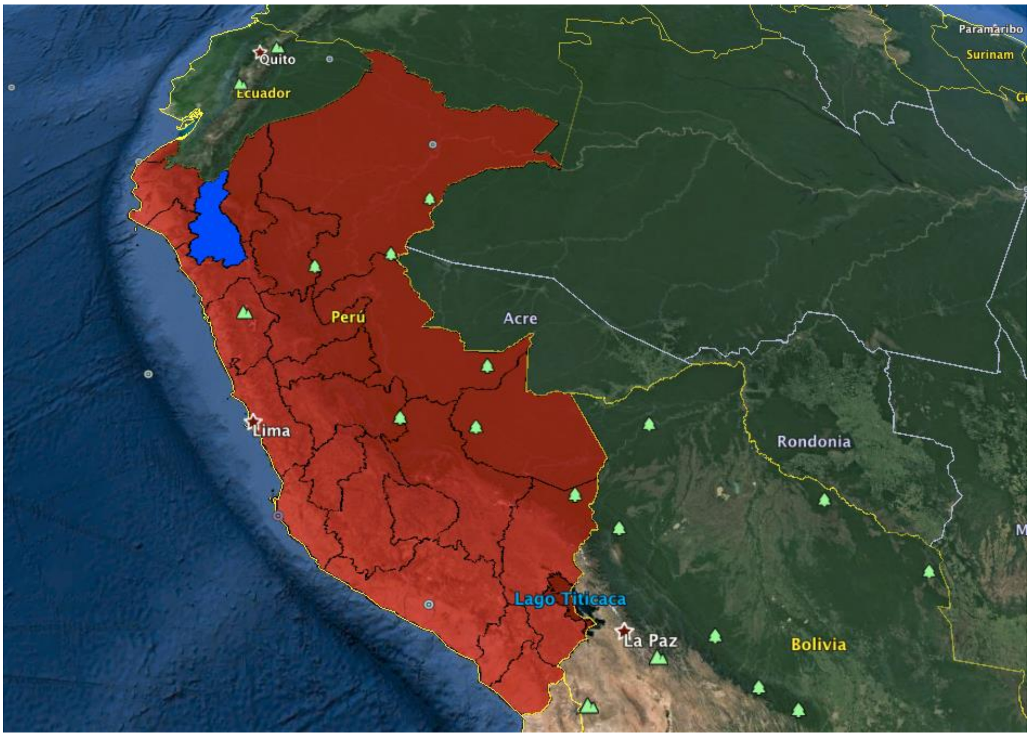


Figure 3-35: Peru map and location of the department of Cajamarca (blue region). Map data: Google, Digital Globe, GADM 2015. Source: (Gonzalez-Garcia et al., 2016).

The Ministry of Energy and Mines approved the National Plan for Rural Electrification in Peru, which considers the period 2014-2022 and has the goal of achieving universal access by the end of this timeframe. The location of the buildings was obtained by manual identification using imagery from Google Earth (see Figure 3-36), and the location of the projected network of 11 kV comes from the National Rural Electrification Plan of Cajamarca.

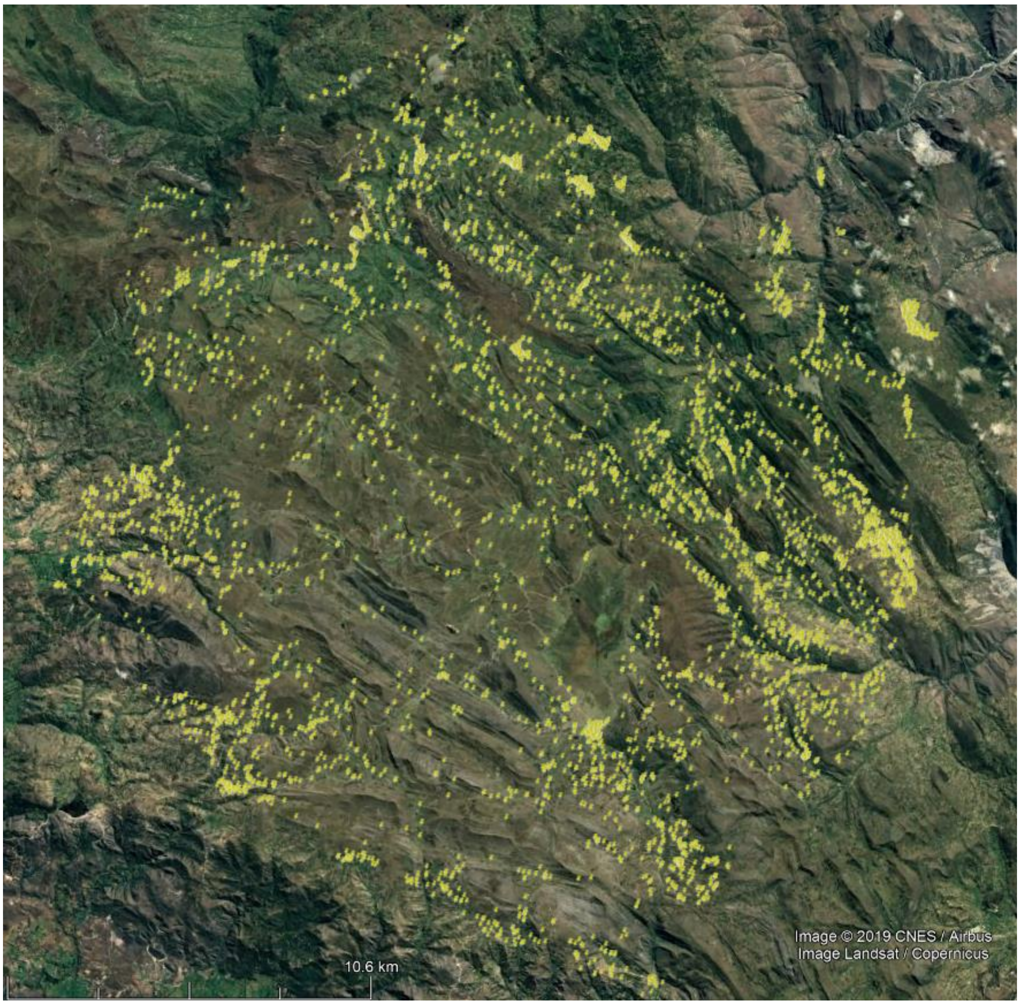


Figure 3-36: Location of the consumers for the case study. Map data: Google, CNES/Airbus, Landsat/Copernicus. Source: (Ciller et al., 2019b).

The grid energy cost is 0.045 $/kWh, and the reliability of the power grid is 100% (Gonzalez-Garcia et al., 2016). The network catalog is based on reference (Gonzalez-Garcia et al., 2016), where a similar case was analyzed using REM. All the consumers have the demand profile shown in Figure 3-37, which was estimated by dividing the aggregated demand profile presented in reference (Villanueva Saberbein and Aye, 2012) by the corresponding total number of consumers.

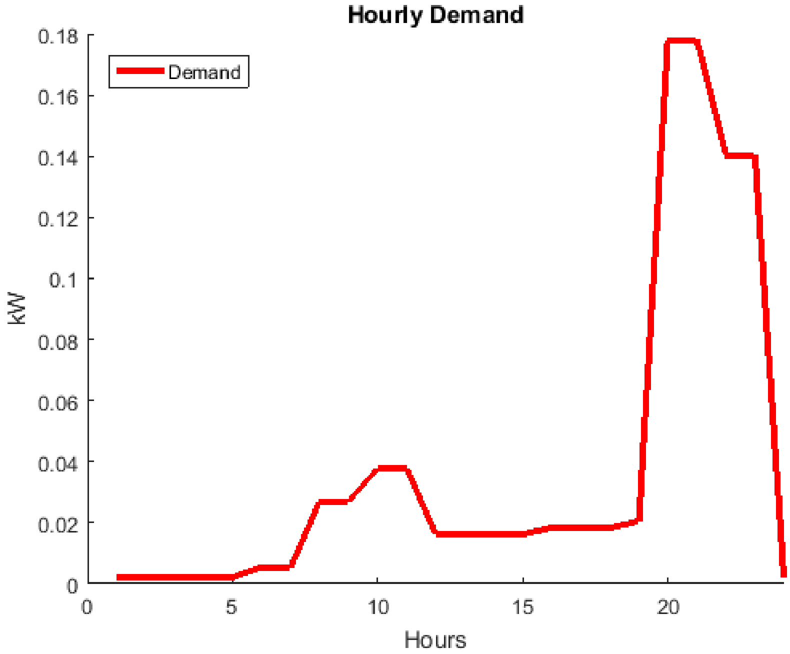


Figure 3-37: Residential demand profile. Source: (Ciller et al., 2019b).

Table 3-11 shows the off-grid generation components used in the case study. The solar irradiance was obtained from reference (NREL, 2017), and the average diesel price is 0.5 $/l[5].

Component	Model	Rated capacity(ies)	Rated voltage	Reference
Diesel generator	-	10 kW, 100 kW, 200 kW, 600 kW, and 1,500 kW	-	The data are based on the expertise of the UEA Lab, field trips, and interviews
Solar panel	First Solar FS-497	97.5 W	-	(First Solar, 2017)
Battery	Lead-acid Trojan L16RE	1,021 AH	2 V	(Trojan Battery Company, 2017)
Inverter	Sunny boy 5000TL	4.6 kW	-	(SMA Solar Technology AG, 2017a)
Charge controller	Sunny island 6.0H	4.6 kW	-	(SMA Solar Technology AG, 2017b)

Table 3-11: Available components for the off-grid systems.

The discount rate is 10%, and the CNSE is 1.5 $/kWh (Gonzalez-Garcia et al., 2016). Although management costs are a substantial improvement, we set them to zero to establish a fair comparison between REM before and after the improvements. Similarly, we do not include solar kits in the case study.

The improvements implemented in REM enhanced the results related to the look-up table

[5] *The average diesel price is similar to the ones we have considered in studies developed in African countries. Although this value may seem low, it is realistic. We should note that the diesel fuel could be subsidized, which could lead to a very inexpensive average diesel price.*

and the optimization of generation designs. Table 3-12 and Table 3-13 show the generation designs for the representative combination of residential consumers calculated with REM before and after the improvements. The first prototype of REM automatically determined the representative combinations of residential consumers, including logarithmically spaced points up to the number of consumers of the case. However, the user determines the combinations of residential consumers in the current version of REM so the representative combinations of residential consumers shown in Table 3-12 and Table 3-13 are different.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	0.86	0	370.51	12.03
2	0.59	4.44	0	0.84	0	298.95	14.06
3	0.78	6.66	0	0.95	0	222.69	15,57
4	1.07	8.88	0	0.93	0	221.21	20.90
9	2.34	15.54	0	0.74	0	264.71	20.12
19	5.07	33.30	0	0.74	0	256.69	18.33
39	4.68	0.00	10	1	0.79	128.80	21.70
82	8.58	79.92	10	0.98	0.61	131.46	26.00
170	19.89	0.00	100	1	0.79	115.64	25.12
355	40.95	0.00	100	1	0.79	80.10	27.44
740	84.24	0.00	200	1	0.79	76.23	25.00
1,541	130.75	1,172.16	200	1	0.7	89.31	25.08
3,210	363.87	0.00	600	1	0.79	66.47	26.10
6,688	756.99	0.00	1,500	1	0.79	69.07	45.28

Table 3-12: Designs obtained with the first prototype of REM.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	1	0	273.48	20.25
5	1.37	11.10	0	0.99	0	140.07	31.00
10	2.73	24.42	0	1	0	124.68	36.69
50	5.36	4.44	10	1	0.70	105.14	47.30
100	27.30	233.10	0	1	0	109.67	50.23
150	40.95	310.80	0	0.99	0	108.39	51.62
200	19.50	66.60	100	1	0.60	99.44	47.47
250	21.94	53.28	100	1	0.67	88.90	47.92
300	28.28	46.62	100	1	0.67	81.53	48.77
500	47.78	73.26	100	1	0.68	67.82	61.18
1,000	94.58	139.86	200	1	0.68	65.99	69.90
3,000	282.75	419.58	600	1	0.68	64.98	83.22
7,500	706.88	1,052.28	1,500	1	0.68	64.83	81.42

Table 3-13: Designs with the master-slave decomposition.

The unitary (per-consumer) generation costs provided in Table 3-13 do not monotonically decrease when the number of residential consumers increases. We analyze the reasons behind this behavior in chapter 4.

Figure 3-42 shows the generation cost per consumer that REM before and after the improvements considers for the clustering algorithm. REM before the improvements directly considers the generation costs shown in Table 3-12 and performs linear interpolation among them, whereas REM after the improvements applies the first smoothing procedure described in section 3.2.2.3.1 to the generation costs shown in Table 3-13.

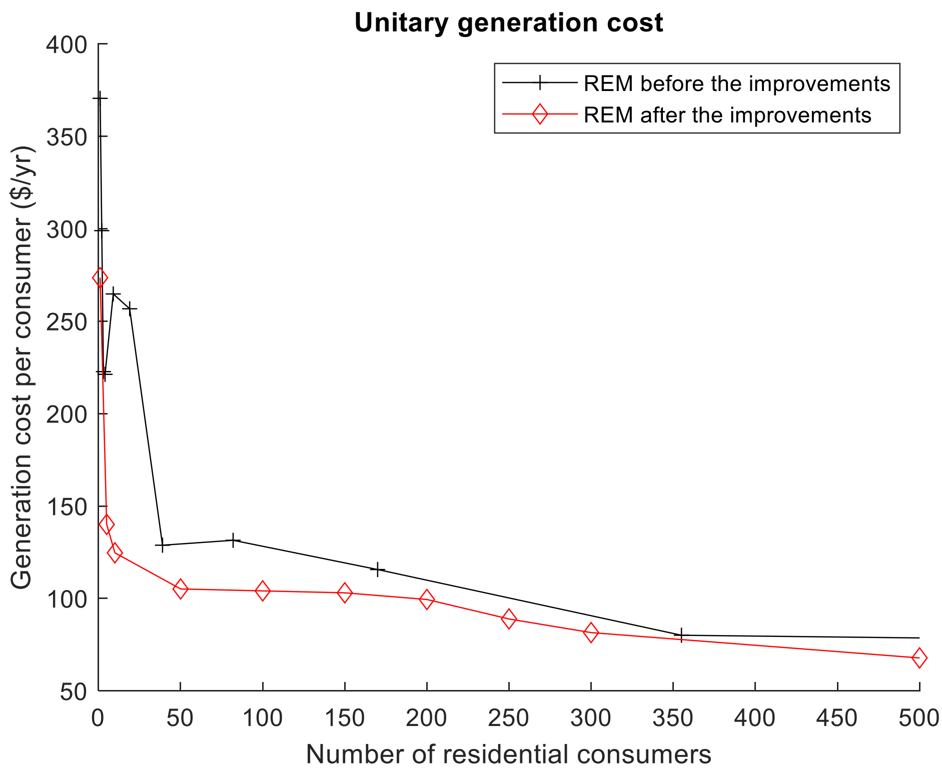


Figure 3-38: Generation costs that REM considers before and after the improvements.

The only representative combination of residential consumers that appears in both Table 3-12 and Table 3-13 corresponds to one (isolated) residential consumer, so this is the only generation design that is directly comparable. Table 3-14 compares the characteristics of a generation design for an isolated consumer that REM provides before and after the improvements. The generation design is the same in terms of capacities, although there are significant differences regarding the total cost (which drops from 370.51 $/yr to 273.48 $/yr) and the amount of served energy (which rises from 86% to approximately 100%). These differences are caused by the improvements related to the dispatch strategy presented in section 3.2.1.

	First prototype of REM	REM after the improvements	Δ (%)
Number of consumers	1	1	0
Peak demand (kW)	0.178	0.178	0
Average demand (kW)	0.039	0.039	0
Solar capacity (kW)	0.293	0.293	0
Battery capacity (kWh)	2.22	2.22	0
Generator capacity (kW)	0	0	-
Fraction of demand served	0.86	1	16.28
Investment and operation cost per demand served ($/kWh)	0.88	0.80	-9.09
Total cost per demand served ($/kWh)	1.09	0.81	-25.69
Total cost ($/yr)	370.51	273.48	-26.19

Table 3-14: Generation designs for an isolated consumer. The last column contains the percentual increment between the first and second columns of the table.

Figure 3-39 shows the daily dispatch of the generation design that the first prototype of REM provides for an isolated consumer. The generation design meets the daily demand with solar energy and charges the battery during the day to meet the nightly demand. However, there are some nights where all the demand is not met.

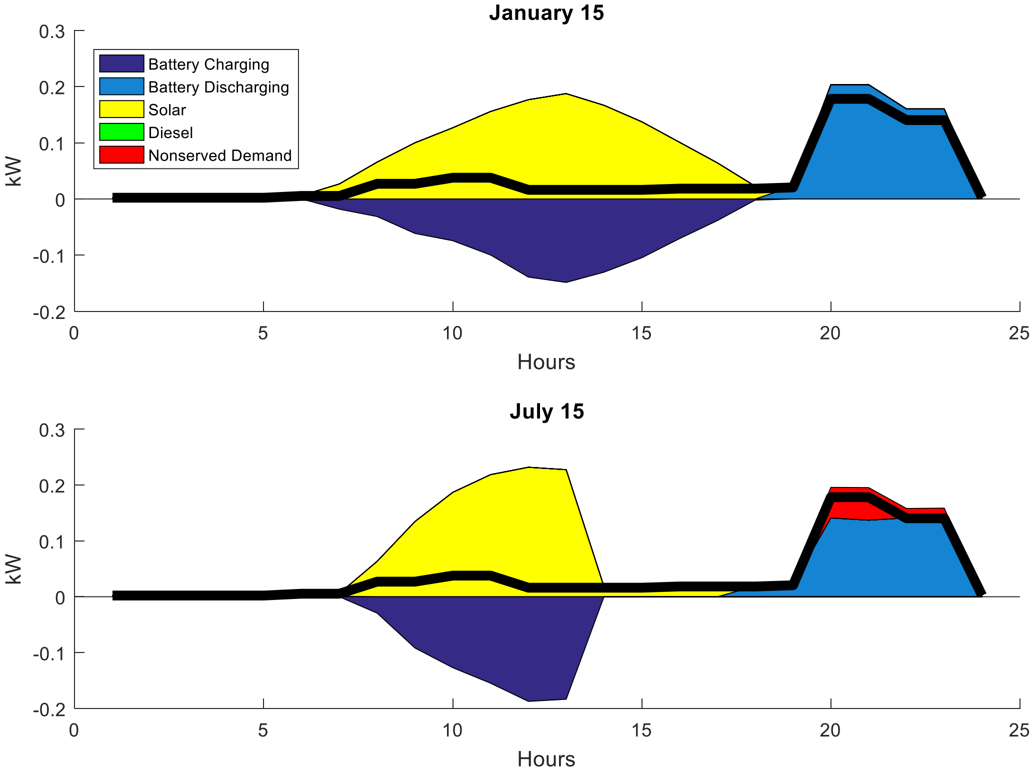


Figure 3-39: Daily dispatch of an isolated consumer (first prototype of REM). The black line represents the total demand.

Figure 3-40 shows the daily dispatch of the generation design for an isolated consumer that REM provides after the improvements. The behavior of the dispatch is very similar, although the percentage of non-served energy has been reduced.

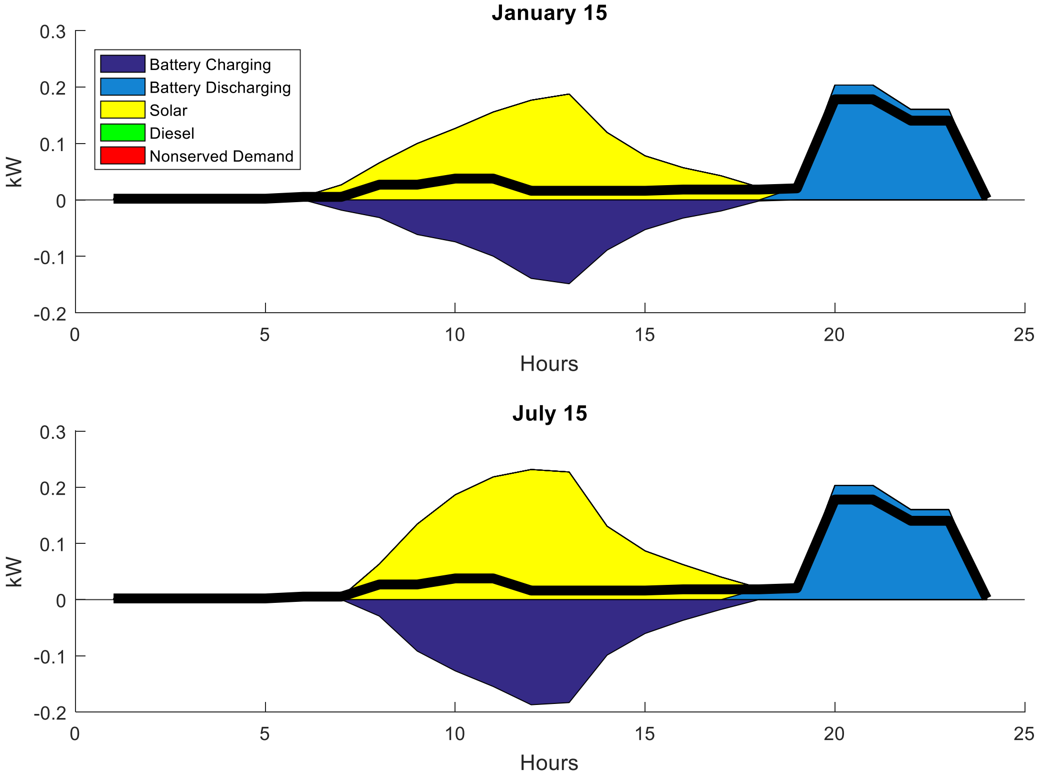


Figure 3-40: Daily dispatch of an isolated consumer (REM after the improvement). The black line represents the total demand.

The battery discharge showed in Figure *3-39* and Figure *3-40* is larger than the total demand during nighttime to account for the mini-grid losses. It should be reminded that the solar dispatch only shows the solar energy used to charge the battery, which depends on the charge equations of the battery.

The generation costs have a substantial impact on the clustering solution. Figure 3-41 shows the cumulative number of consumers per cluster for the off-grid clusters, which is significantly different before and after implementing the improvements. 99.52% of off-grid clusters have less than ten consumers with REM's first prototype, whereas that number drops to 15.97% after the improvements. Similarly, 100% of off-grid clusters have less than thirty-five consumers with REM's first prototype, and this figure drops to 55.56% after the improvements.

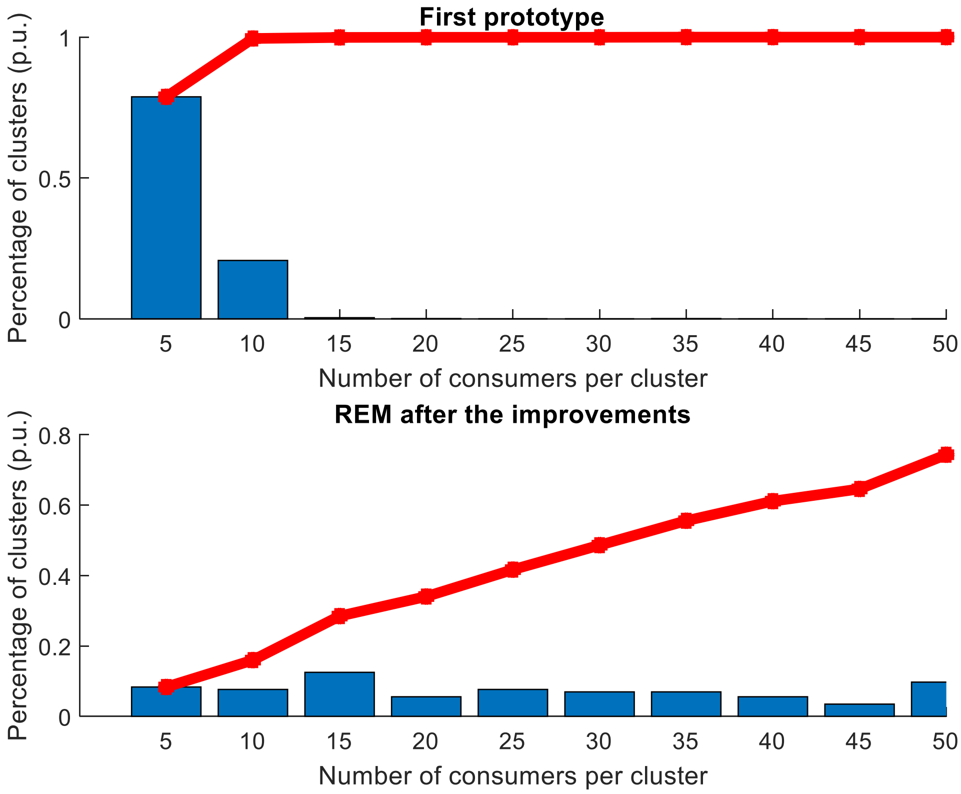


Figure 3-41: Number of consumers per cluster for the off-grid clustering.

Figure 3-42 shows the cumulative number of consumers per cluster for the grid-extension clusters, which is also significantly different. For example, 93.33% of the grid-extension clusters have less than fifty consumers with REM's first prototype, and this figure drops to 74.31% after the improvements. However, 94.44% of the grid-extension clusters have less than one hundred consumers, and this figure rises to 95.83% after the improvements. This should not be interpreted as a correct behavior of the grid-extension clustering when obtaining large clusters in the first REM prototype: the size of some grid-extension clusters is not coherent with the corresponding off-grid clustering, and some cascading effects kept connecting the clusters beyond a reasonable size. The cascading effect sometimes produces significantly large clusters (which correspond to the grid extension designs in Figure 3-43), masking the inefficient behavior of the off-grid clustering in the first prototype of REM.

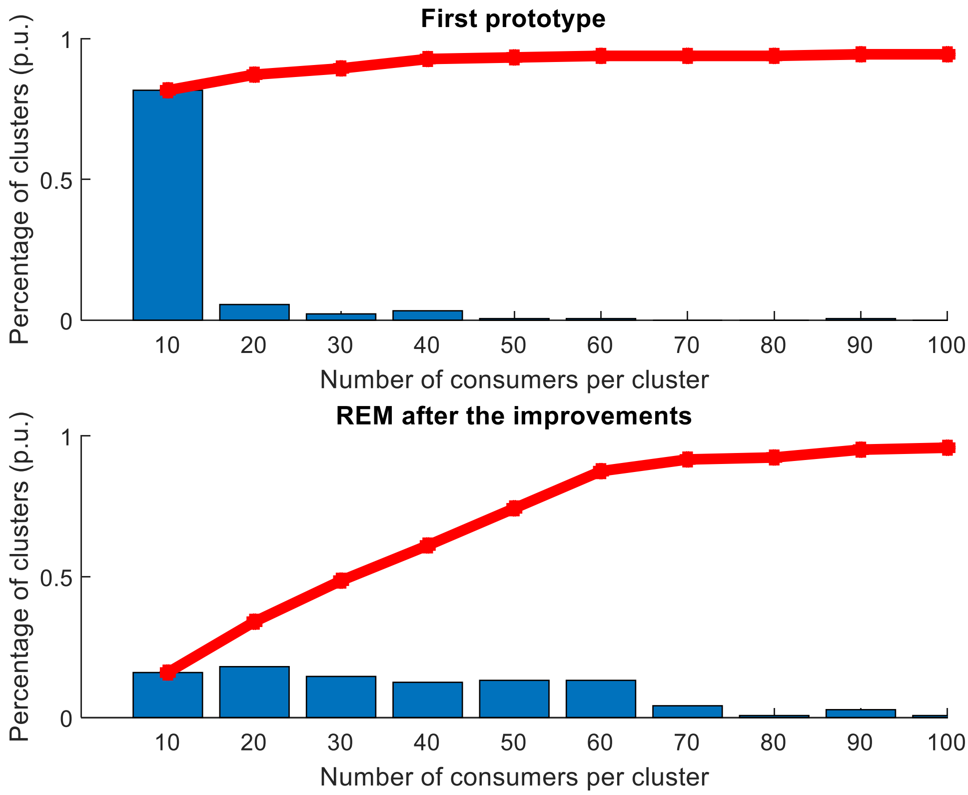


Figure 3-42: Number of consumers per cluster for the grid-extension clustering.

The off-grid clusters that REM provides after the improvements have more adequate sizes and fewer generation costs than the off-grid clusters that the first prototype of REM calculates, so they are more competitive and swift the final electrification solution towards off-grid alternatives. Figure 3-43 shows the electrification solutions that REM provides before and after implementing the improvements described in section 3.2. The first prototype of REM tends to electrify consumers near the power grid with grid extensions and the consumers that are further from the network with off-grid alternatives. However, REM provides an electrification solution without any grid extension after implementing the improvements.

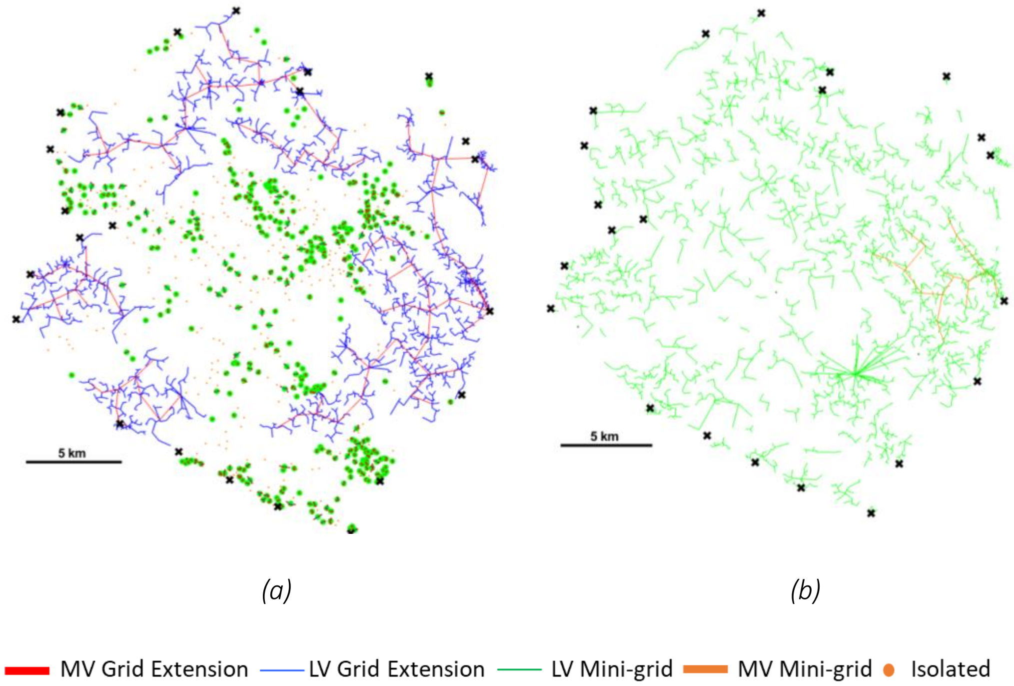


Figure 3-43: Electrification solution provided by (a) the first prototype of REM and (b) REM after the improvements described in this chapter.

Table 3-15 shows the cost summary of REM's electrification solutions before and after implementing the improvements. The total annual electrification cost drops from 1,111,674 $/yr to 909,338 $/yr, which is a significant improvement.

	First prototype of REM				REM after the improvements				Δ All (%)
	Mini-grids	Isolated	Grid extension	All	Mini-grids	Isolated	Grid extension	All	
Number of consumers	1,425	245	5,018	6,688	6,685	3	0	6,688	0
Fraction of consumers	0.21	0.04	0.75	1	1	0	0	1	0
Investment and operation cost per consumer ($/yr)	196.24	303.30	132.97	152.69	134.68	276	0	134.74	-11.76
Management cost per consumer ($/yr)	0	0	0	0	0	0	0	0	-
CNSE per consumer ($/yr)	51.40	70.28	0	13.52	1.23	1	0	1.23	-90.90
Final cost per consumer ($/yr)	247.64	373.57	132.97	166.22	135.9	277	0	135.97	-18.20
Total investment and operation cost ($/yr)	279,641	74,308	667,262	1,021,211	900,309	828	0	901,137	-11.76
Total management cost ($/yr)	0	0	0	0	0	0	0	0	-
Total CNSE ($/yr)	73,245	17,218	0	90,463	8,198	3	0	8,201	-90.93
Final cost ($/yr)	352,886	91,525	667,262	1,111,674	908,507	831	0	909,338	-18.20

Table 3-15: Electrification solution summary for the case study. The last column contains the percentual increment between the “All” columns of the table.

In this case study, REM calculates the generation costs of the final solution shown in Table 3-15 interpolating among the generation designs of the look-up table (which corresponds to Table 3-12 and Table 3-13 in the cases of REM before and after the improvements, respectively). This implies that most generation designs of the final electrification solution do not include real, discrete generation components because the interpolation leads to approximated designs. For example, REM (after the improvements) would provide a generation design for a mini-grid with fifteen consumers that includes a diesel generator of 1.25 kW due to linear interpolation among the designs for ten and fifty consumers shown in Table 3-13.

A final electrification solution with interpolated generation designs is accurate enough for large-scale planning, where the purpose of a plan is two-fold. On the one hand, a plan should establish the least-cost electrification mode that meets some minimum service standards in the considered region (i.e., where the least-cost electrification solution consists of extending the grid and where off-grid systems should be built). On the other hand, the plan should provide a reasonable approximation to the overall cost and the bill of materials needed to accomplish the electrification project. It is not the purpose of a large-scale plan to provide detailed engineering designs ready for construction.

The next chapter will delve into REM’s mini-grid generation block and how discrete generation components affect the clustering algorithm. We will present further developments that go beyond the improvements introduced in section 3.2.1.

3.4. Conclusions

This chapter provides an overview of REM's first prototype, which is the starting point of this thesis. The first prototype of REM presented a high-level structure, which is still present in the current version of REM, which consists of five blocks that operate sequentially.

REM's first prototype provided inconsistent results that could not be applied in real cases, and it was necessary to perform an in-depth analysis of the algorithms to determine the issues behind the incoherent results. This analysis was complemented with several algorithmic improvements.

This chapter describes the main algorithmic improvements implemented in the first prototype of REM, turning it into a robust tool that produces reliable results. We also describe several upgrades that expanded the capabilities of the model, such as the incorporation of solar kits as a feasible electrification solution or the addition of multiple consumer types. The enhancements presented in this chapter are classified according to the main algorithmic blocks of REM.

We show the impact of the improvements in a realistic case study located in Cajamarca (Peru). We compare the electrification results provided by REM before and after the upgrades, and there is a severe improvement in terms of the final electrification cost (which is reduced by 18.2%). We also show how some intermediate results related to off-grid generation designs and clustering change drastically after upgrading the model.

We can conclude that the improvements introduced in this chapter substantially enhanced the performance of REM. Specifically, the interpolation improvements introduced in section 3.2.2.3 heavily impact the electrification solution as they ensure that the clustering avoids local minimum solutions with many small off-grid clusters. The master-slave decomposition and the dispatch strategy described in sections 3.2.1.1 and 3.2.1.2 also played a critical role in REM's evolution as they lead to accurate cost calculations of the off-grid generation designs. Finally, the proper use of RNM to design networks for mini-grids (section 3.2.3.2) is essential for the robustness of the final electrification solutions.

"I learned that in the face of a void or in the face of any challenge, you can choose joy and meaning." *Sheryl Sandberg*

4

OFF-GRID GENERATION IN LARGE-SCALE PLANNING

This chapter is a continuation of the developments presented in section 3.2.1, where we analyzed and solved several issues present in the first prototype of REM concerning the optimization of generation designs of off-grid systems. Further analysis allowed us to conclude that modeling the capacities of some elements with discrete variables may lead to trouble when REM groups the consumers into clusters. We developed two procedures to overcome the difficulties. The content of this chapter has been published in the following paper:

Ciller, P., de Cuadra, F., Lumbreras, S., 2019. Optimizing Off-Grid Generation in Large-Scale Electrification-Planning Problems: A Direct-Search Approach. Energies 12, 4634. https://doi.org/10.3390/en12244634

Section 4.1 briefly describes how regional planning tools and single-system tools optimize the generation designs of off-grid systems. Section 4.2 analyzes the drawbacks of a direct application of the master-slave decomposition introduced in section 3.2.1.1 in regional planning, and it presents a new smoothing method. Section 4.3 introduces a new algorithm that REM applies to optimize the generation design of an off-grid system from scratch, which is based on continuous variables. Results and conclusions are provided in Sections 4.4 and 4.5, respectively.

Figure 4-1 shows how the smoothing method introduced in section 4.2 and the algorithm based on continuous variables presented in section 4.3 fit the overall REM procedure. After selecting the representative off-grid systems (which is now done by the user), REM can either apply a combination of the master-slave decomposition presented in section 3.2.1.1 and the smoothing method introduced in section 4.2 or directly use the continuous-component implementation described in section 4.3.

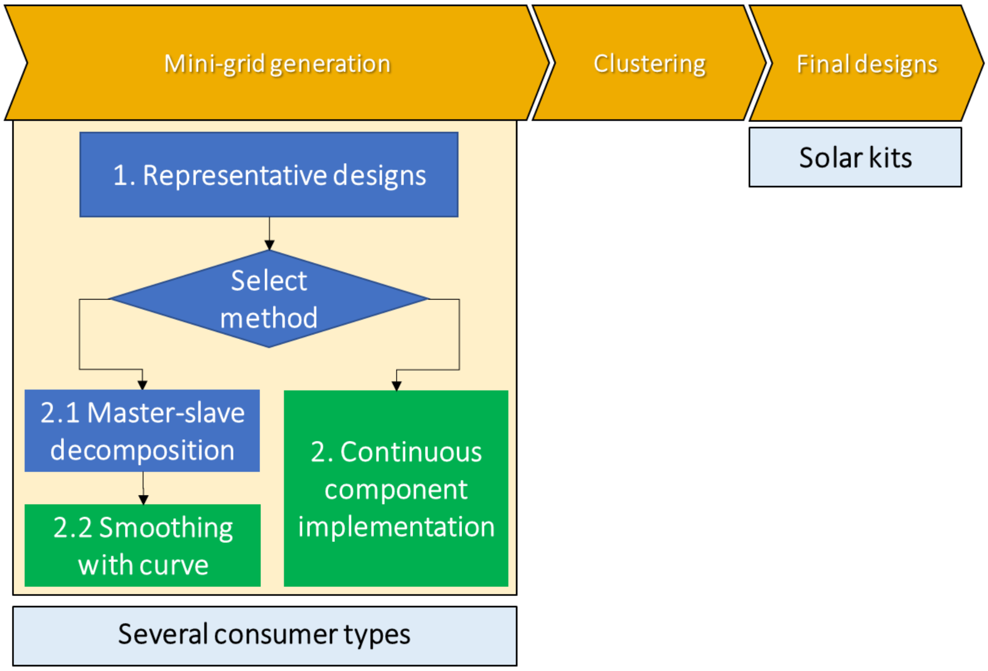


Figure 4-1: Incorporation of the smoothing method and the continuous implementation into REM's algorithmic structure.

4.1. Generation designs

The optimization of off-grid generation designs is critical for the soundness of an electrification plan, but most regional planning models calculate the generation capacity for off-grid systems with methods that lack modeling accuracy. These methods provide valuable first-pass information, but they lack the level of detail required for a sound electrification plan.

For example, Network Planner (Kemausuor et al., 2014b) sizes the generation designs with analytical expressions based on rules of thumb, and OnSSET (Korkovelos et al., 2019; Mentis et al., 2017a) uses analytical expressions to estimate the LCOE of off-grid systems. These models do not consider the temporal operation of the system, neglecting the impact of seasonality in renewable energy generation, which could be translated into periods where a significant amount of the demand is not supplied.

On the other hand, the optimization of generation designs is a widely-studied problem from the perspective of an individual off-grid system (Luna-Rubio et al., 2012; Upadhyay and Sharma, 2014). Some methods are based on classical optimization techniques such as Mixed Integer Programming (MIP) (Domenech et al., 2018), whereas others apply heuristic algorithms (Brivio et al., 2017), metaheuristic techniques (Borhanazad et al., 2014; Maleki and Askarzadeh, 2014; Merei et al., 2013; Mohammed et al., 2018), or artificial intelligence methods (Lujano-Rojas et al., 2013). Most methods minimize the cost of the system, although some methods include other criteria such as minimizing carbon emissions (Tawil et al., 2018).

Several software tools optimize the generation design of a single off-grid system (Sinha and

Chandel, 2014), being HOMER the most widely used. HOMER has been thoroughly applied both in developed (Rahman et al., 2016) and developing countries (Al-Rubaye et al., 2018; Gebrehiwot et al., 2019; Micangeli et al., 2017; Sen and Bhattacharyya, 2014). Other relevant tools are DER-CAM (Hartvigsson et al., 2018), which considers a MILP formulation, and the improved Hybrid Optimization by Genetic Algorithms (iHOGA) (Ganguly et al., 2017), which uses a genetic algorithm.

Although some of these methods and tools are based on sophisticated optimization techniques and detailed models, they are not directly applicable to regional planning. It is necessary to optimize the generation designs of many mini-grids in large-scale planning, and applying a computationally intensive approach for each design is not feasible. Single-system methods also assume that the optimal electrification solution is an individual mini-grid that electrifies all the consumers of the village or settlement. In contrast, the optimal number of off-grid systems and which consumers belong to each system need to be determined in regional planning.

REM optimizes the generation designs of off-grid systems applying a method that combines the high level of modeling detail of single-system tools with the massive electrification scope of regional planning tools. To the best of our knowledge, there is only one regional planning methodology that applies a similar level of modeling detail regarding generation designs (Blechinger et al., 2019). This tool is described and analyzed in chapter 2.

4.2. A first approach to detailed regional planning

REM groups the consumers into electrification clusters minimizing the total costs of the systems. The clustering algorithm that REM applies has two steps. In the first step (off-grid clustering), REM assumes temporarily that only off-grid alternatives (standalone systems and mini-grids) are the only viable electrification solutions. The best grouping of consumers into off-grid systems depends on the trade-offs between the costs involved, being the generation cost one of them. If the estimations of the generation costs fail to capture the economies of scale in generation equipment, then the off-grid clustering results may be far from optimal.

We now provide an illustrative example with one load type (residential) that shows the limitations of the master-slave decomposition presented in section 3.2.1.1 for regional planning, which is what REM essentially applied to optimize the generation designs after the improvements presented in chapter 3. Any tool or method that aims at optimizing the generation design with discrete generation components, such as those that deal with a single mini-grid or village, would present the same limitations. The problems found, the solutions proposed and the overall conclusions are fully applicable to cases with several types of loads, but they are not used here for the sake of simplicity.

There are two diesel generators available with capacities of 10 kW and 100 kW, and the representative mini-grids correspond to 1, 5, 10, 50, 100, 150, 200, 250, 300, and 500 residential consumers. Although the number of available diesel generators may seem low, it is realistic. The logistics of dealing with an extensive catalog of diesel generators in a regional planning project complicates the implementation phase, and some planners prefer to limit the available diesel options, purchase specific generators in bulk and benefit from volume discount pricing.

Figure 4-2 shows the minimum cost per consumer for these representative combinations of residential consumers, showing the partial optima obtained for the three diesel options (0, 10, and 100 kW) and the minimum cost curve (i.e., the minimum-cost design for each point).

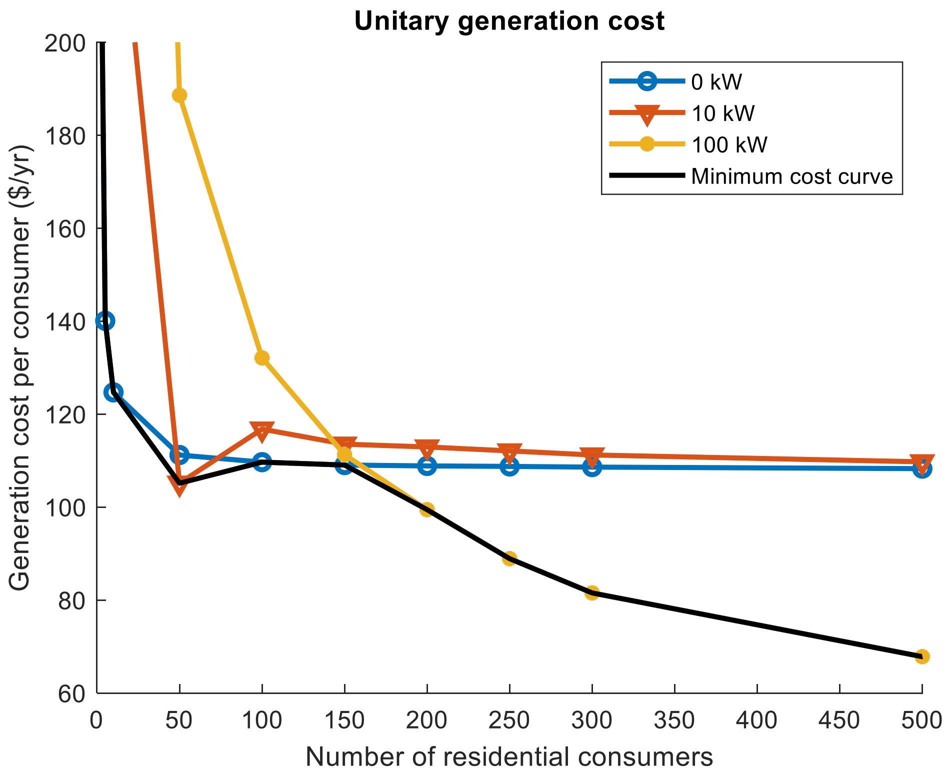


Figure 4-2: Generation cost per consumer obtained with the discrete algorithm. Source: adapted from (Ciller et al., 2019b).

This case illustrates two effects of having discrete diesel options. The first one is the instability of the generation mix concerning the demand. In this case, mini-grids with less than 50 consumers do not include a diesel generator and mini-grids with a range of consumers that lay between 50 and 100 consumers include a 10 kW diesel generator. However, mini-grids between 100 and 150 consumers do not include the diesel generator because the 10 kW diesel generator is too small. In comparison, the 100 kW diesel generator is too big and expensive for this particular range of consumers. For mini-grids larger than 150 consumers, the generation solution includes the 100 kW diesel generator.

The second effect is that the unitary generation cost is not a monotonically decreasing function of the number of residential consumers. This second effect causes issues in the off-grid clustering of REM because the algorithm considers the trade-off between the additional network cost and the generation and management savings (due to economies of scale) to determine whether to join two clusters. Hence, if the management savings are small and the generation cost per consumer starts increasing at some point, the algorithm would not find the best solution. In the case shown in Figure 4-2, the algorithm could stop joining clusters at sizes of about 50 households, missing the economies of scale beyond 150 households.

To overcome this difficulty, we can smooth the generation-cost curve by adjusting the coefficients of a family of curves that guarantee a monotonic behavior so the generation cost per consumer always decreases when the number of consumers increases. Equation 4-1 defines this family of curves.

$$\mathrm{P}_{\alpha,\beta,\gamma}(\mathrm{x}) = \frac{\alpha}{\mathrm{x}^{\gamma}} + \beta \tag{4-1}$$

Where x is the number of residential consumers; α, β, γ are non-negative parameters that REM adjusts, and $\mathrm{P}_{\alpha,\beta,\gamma}(\mathrm{x})$ is the approximated unitary generation cost for x residential consumers. Each curve of the form $\mathrm{P}_{\alpha,\beta,\gamma}(\mathrm{x})$ is a decreasing convex function. Note that the smooth curves must be replaced by smooth hypersurfaces in cases with more types of loads.

The curve introduced in equation 4-1 is an improvement over the initial smoothing methods presented in section 3.2.2.3.1. The first method iteratively loops through the points of the look-up table. If the unitary generation cost of the $i-th$ point is higher than the unitary generation cost of the $(i-1)-th$ point, then the unitary generation cost of the $i-th$ point is set to 99% of the unitary generation cost of the $(i-1)-th$ point. This method does not guarantee that the incremental reduction of the unitary generation cost decreases monotonously when the number of residential consumers increases.

The second method was based on a piecewise exponential function with two pieces. The use of an exponential family of curves sometimes led to trouble. If only one exponential function was used, then the approximation sometimes failed to capture the economies of scale in generation for the initial points of the look-up table, which are critical. If two or more exponential functions were used (such as in Figure 3-30), then the behavior of the piecewise function in the neighborhoods of the hinge points (i.e., the points where two exponential functions meet) sometimes was not smooth.

In the case study presented in this chapter, REM uses the smooth curve only to determine the off-grid clusters, but it is not used to compute the final generation cost of the off-grid systems in the final electrification solution. Therefore, the use of the smooth curve does not involve any loss of realism in the final electrification solution of the case study presented in this chapter.

4.3. The continuous-component implementation

Smoothing the cost values has some limitations. It may be difficult to smooth generation costs if we are working on a case with several types of loads. For example, it is debatable whether all types of loads should have equal importance when smoothing their generation costs. Residential loads are more frequent, but productive loads have a substantial impact on the final electrification solution.

It can be concluded from section 4.2 that modeling the capacities of diesel generators with discrete variables is problematic. In this section, we propose a new logic for the master problem that treats the diesel capacity as a continuous variable. When the diesel capacity is treated as a continuous variable, the results of the master problem are not so heavily influenced by the diesel generators available as the algorithm can interpolate among them to obtain a diesel generator of any desired capacity.

The proposed logic for the master level performs a search by trisecting an interval (i.e., dividing an interval into three segments of the same length with four points), which is shortened by discarding the diesel capacity that is further from the current best design and trisected again. The process continues until the length of the interval is lower than a pre-specified tolerance. A similar trisection procedure is applied in (Kong and Muzathik, 2012) to determine the optimal point of the I-V and P-V characteristics of a solar panel. The slave problem presented in section 4.2 has provided satisfactory results so far, so it has not been necessary to modify it.

We present an illustrative example in Figure 4-3. In this case, we assume that the pre-specified tolerance is 2 kW and the minimum capacity that meets the aggregated demand is 12 kW, so the boundaries of the diesel capacity are [0 kW, 12 kW]. The first set of points is evaluated by trisecting this interval, which yields the 4 kW and 8 kW generators, and we assume that the lowest-cost point corresponds to the 4 kW diesel generator. Hence, the highest-capacity generator (12 kW) is discarded and the second set of points is obtained trisecting the interval [0 kW, 8 kW], which yields the 2.67 kW and 5.34 kW generators, and we assume that the lowest-cost solution for the second set of points corresponds to the 5.34 kW diesel generator. The lowest-capacity generator (0 kW) is therefore discarded, and the interval [2.67 kW, 8 kW] is trisected to obtain the third set of points, yielding the 4.45 kW and 6.22 kW generators. The lowest-cost point of the third set of points is the 4.45 kW generator, which is the final solution provided by the algorithm since $|4.45\text{-}2.67| < 2$.

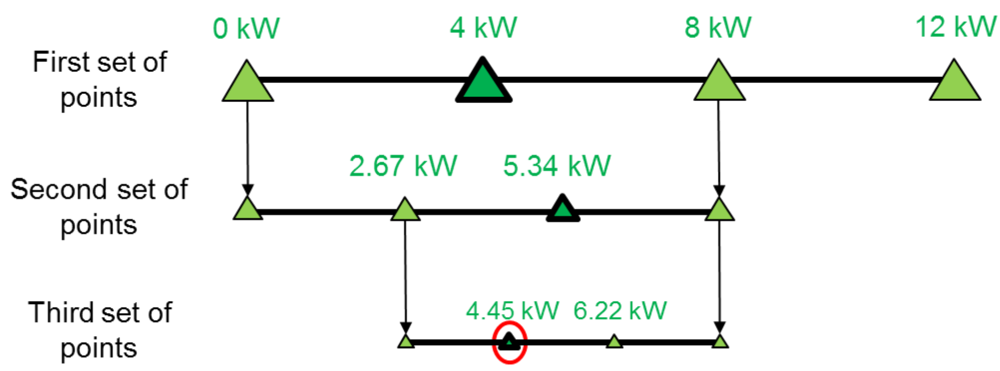


Figure 4-3: Example of the continuous algorithm. The best solution for each set of points is marked with a triangle of a wider boundary, and a circle surrounds the final solution that the algorithm provides. Source: (Ciller et al., 2019b).

The slave problemneeds to calculate generation designs for two different diesel capacities for each iteration, since the first and the last points of the i-th set of points also belong to the (i-1)-th set of points. Figure 4-4 shows the flow diagram of the master problem presented in this section.

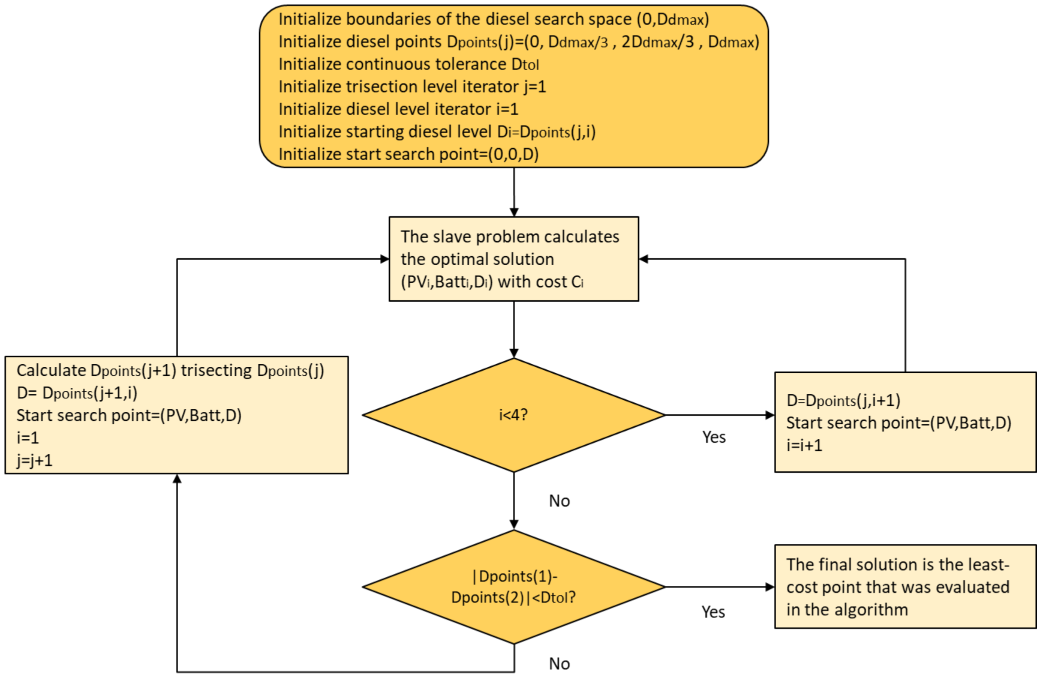


Figure 4-4: Flow diagram of the master (continuous) problem. Source: (Ciller et al., 2019b).

The next section presents a case study where we show that a direct application of the master-slave decomposition presented in section 3.2.1.1 leads to an inefficient grouping of consumers in the off-grid clustering, which reaches a local minimum. We also show that the smoothed curve introduced in section 4.2 and the continuous-component implementation presented in this section enhance the clustering results.

In the case study presented in this chapter, the continuous diesel generators are only considered to estimate the generation costs used to group the consumers into mini-grids (such as the smoothed curve presented in section 4.2), but they are not used to calculate the generation costs included in the final electrification solution. Therefore, continuous diesel generators do not involve any loss of realism in the final electrification solution of the case study presented in this chapter.

In the case studies presented in this thesis, REM calculates the generation costs of the final electrification solution by performing linear interpolation among the points of the look-up table. This implies that the final electrification solution includes generation designs with interpolated diesel generators even if the look-up table is calculated with discrete components.

4.4. Case study

The case study and the input parameters are very similar to the ones presented in section

3.3, so we will not provide a thorough description here. The capacities of diesel generators available are 10 kW, 100 kW (as in the previous example in section 4.2), 200 kW, 600 kW, and 1,500 kW. The generation designs calculated correspond to 1, 5, 10, 50, 100, 150, 200, 250, 300, 500 (as in the previous example in section 4.2), 1,000, 3,000, and 7,500 residential consumers.

Table 4-1 and Table 4-2 show the look-up table obtained with discrete and continuous diesel capacities, respectively (Table 4-1 provides the same generation results as Table 3-13; we also provide these results here for the sake of clarity). As shown in Table 4-1 and Figure 4-2, using discrete capacities is translated into a non-monotonic behavior in the size of diesel generators in the range from 50 to 200 households. In contrast, the diesel capacity increases consistently with the number of residential consumers when handled as a continuous variable.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	1	0	273.48	20.25
5	1.37	11.10	0	0.99	0	140.07	31.00
10	2.73	24.42	0	1	0	124.68	36.69
50	5.36	4.44	10	1	0.70	105.14	47.30
100	27.30	233.10	0	1	0	109.67	50.23
150	40.95	310.80	0	0.99	0	108.39	51.62
200	19.50	66.60	100	1	0.60	99.44	47.47
250	21.94	53.28	100	1	0.67	88.90	47.92
300	28.28	46.62	100	1	0.67	81.53	48.77
500	47.78	73.26	100	1	0.68	67.82	61.18
1,000	94.58	139.86	200	1	0.68	65.99	69.90
3,000	282.75	419.58	600	1	0.68	64.98	83.22
7,500	706.88	1,052.28	1,500	1	0.68	64.83	81.42

Table 4-1: Designs with discrete diesel capacities.

Residential consumers	Solar capacity (kWp)	Battery capacity (kWh)	Generator capacity (kW)	Fraction of demand served (p.u.)	Fraction of demand served with diesel (p.u.)	Total cost per consumer ($/yr)	Computation time (sec)
1	0.29	2.22	0	1	0	273.48	25.53
5	1.37	11.10	0	0.99	0	140.07	146.55
10	2.73	24.42	0	1	0	124.68	203.22
50	5.36	4.44	8.9	1	0.70	100.97	261.90
100	9.75	13.32	18.8	1	0.68	87.76	343.35
150	14.63	17.76	26.8	1	0.69	81.87	323.43
200	19.50	22.20	35.7	1	0.69	79.08	356.33
250	24.38	26.64	44.7	1	0.70	76.97	365.97
300	28.28	31.08	53.6	1	0.70	74.96	389.68
500	47.78	73.26	99.1	1	0.68	67.84	397.13
1,000	85.12	139.86	178.2	1	0.69	64.70	458.40
3,000	254.48	419.58	534.6	1	0.69	63.83	508.14
7,500	636.19	1,052.28	1336.2	1	0.69	63.63	631.48

Table 4-2: Designs with continuous diesel capacities. The pre-specified tolerance of the continuous algorithm was set to 0.1 kW.

The continuous algorithm requires a higher computation time than the discrete algorithm because it iterates several times until it reaches the pre-specified threshold (0.1 kW in this case study). The initial iteration evaluates four diesel generators, whereas each additional iteration requires the evaluation of two additional diesel generators. The discrete algorithm only needs to evaluate a reduced number of diesel generators in this case study (six diesel generators at most, including the no-diesel solution).

The computation time is still manageable with the continuous algorithm, although it could be reduced with parallel computing (i.e., several generation designs could be optimized simultaneously as the optimization of one generation design is independent of the optimization of the remaining ones).

Figure 4-5 shows the minimum cost design curve (obtained with the master-slave decomposition described in section 3.2.1.1 and performing linear interpolation), the smoothed curve (obtained applying the procedure described in section 4.2 to the minimum cost design curve) and the continuous results (obtained with the method presented in section 4.3 and performing linear interpolation). The continuous implementation captures better the trend of economies of scale in generation.

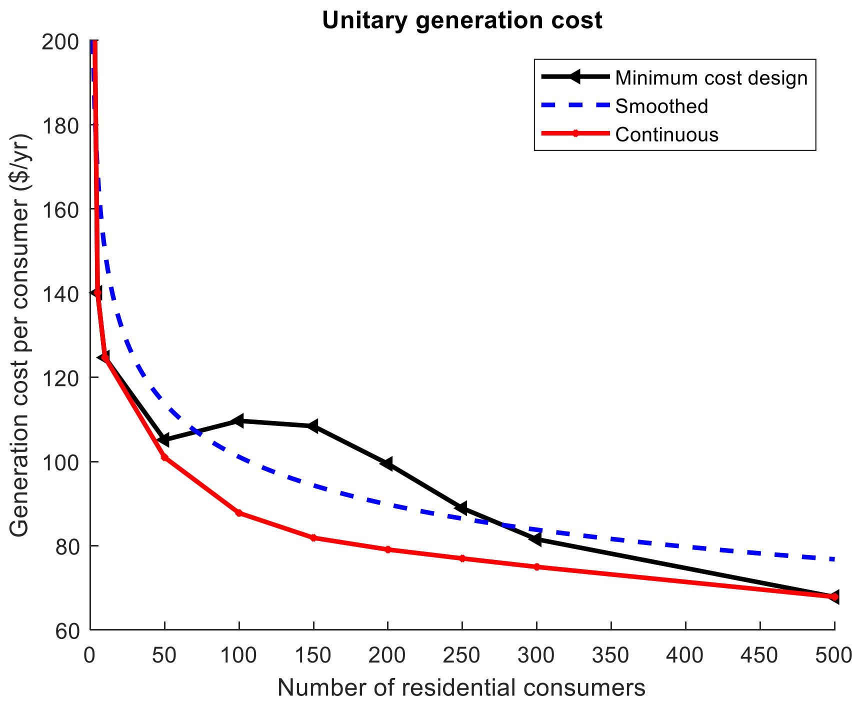


Figure 4-5: Generation cost obtained with the master-slave decomposition presented in section 3.2.1.1 and the methods described in sections 4.2 and 4.3. Source: (Ciller et al., 2019b).

Figure 4-6 and Figure 4-7 show the cost breakdown of the generation designs obtained with discrete and continuous diesel capacities, respectively. As expected, designs that include a diesel generator have an Operational Expenditure (OPEX) that accounts for a much more significant amount of the generation cost (the OPEX of diesel generators accounts for a much larger percentage of the total cost than the OPEX for solar panels and batteries). This effect especially stands out in Figure 4-6 as the design for 50 consumers includes a 10 kW diesel generator but designs for 100 and 150 residential consumers do not include a diesel generator.

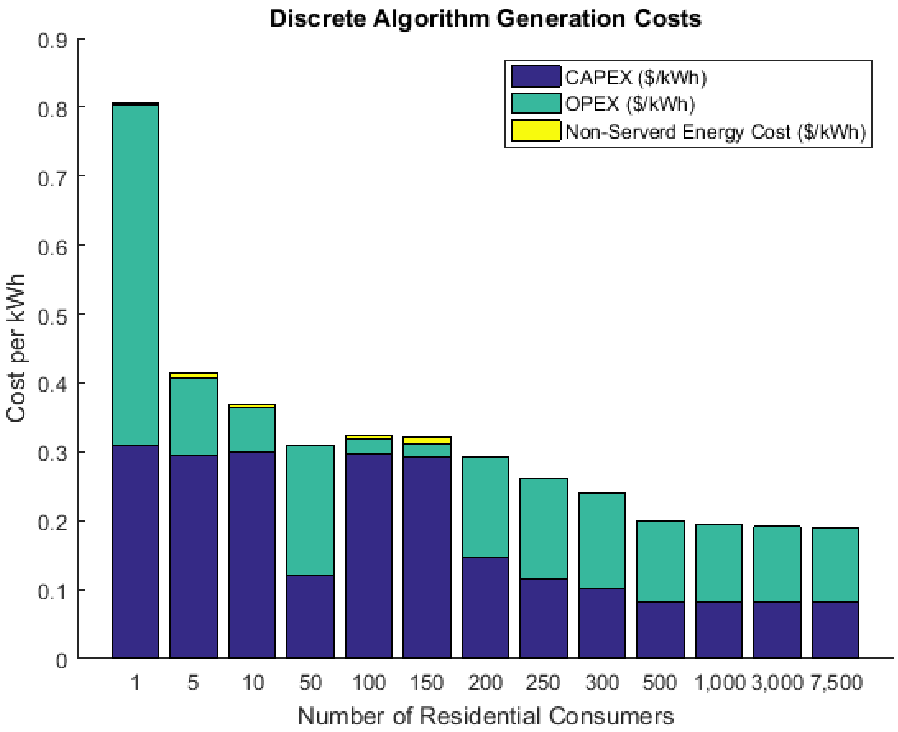


Figure 4-6: Generation costs obtained with the discrete algorithm. Source: (Ciller et al., 2019b).

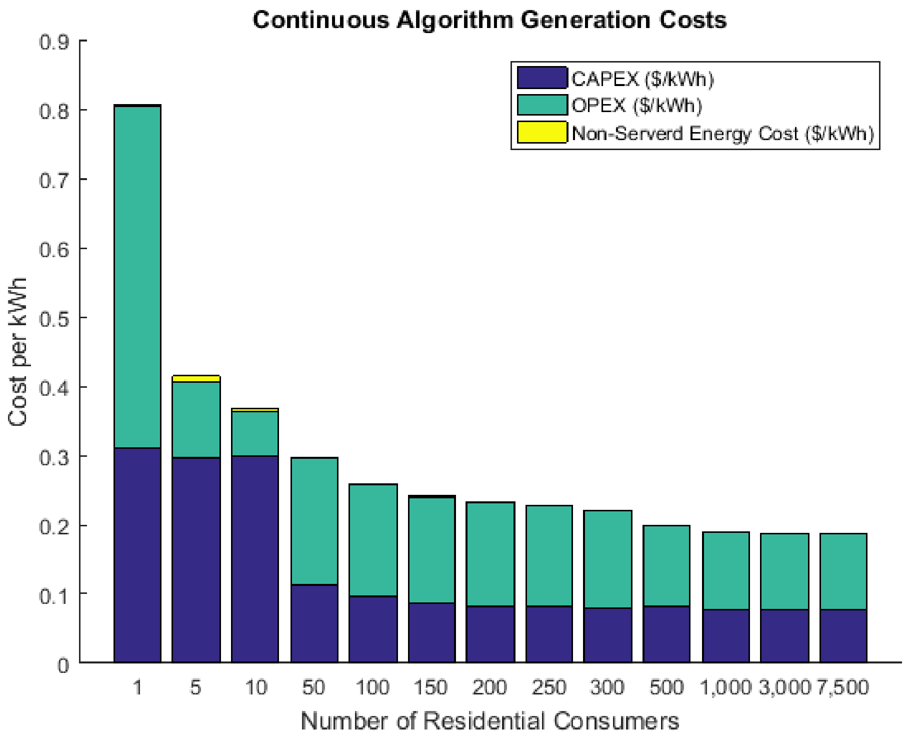


Figure 4-7: Generation costs obtained with the continuous algorithm. Source: (Ciller et al., 2019b).

Figure 4-8 shows the sizes (number of consumers) of off-grid clusters obtained with discrete diesel capacities, the corresponding smooth curve, and continuous diesel capacities. The off-grid clusters obtained with the smooth curve and continuous generators are similar, but they are significantly different when generation costs are estimated directly from a look-up table calculated with discrete diesel capacities.

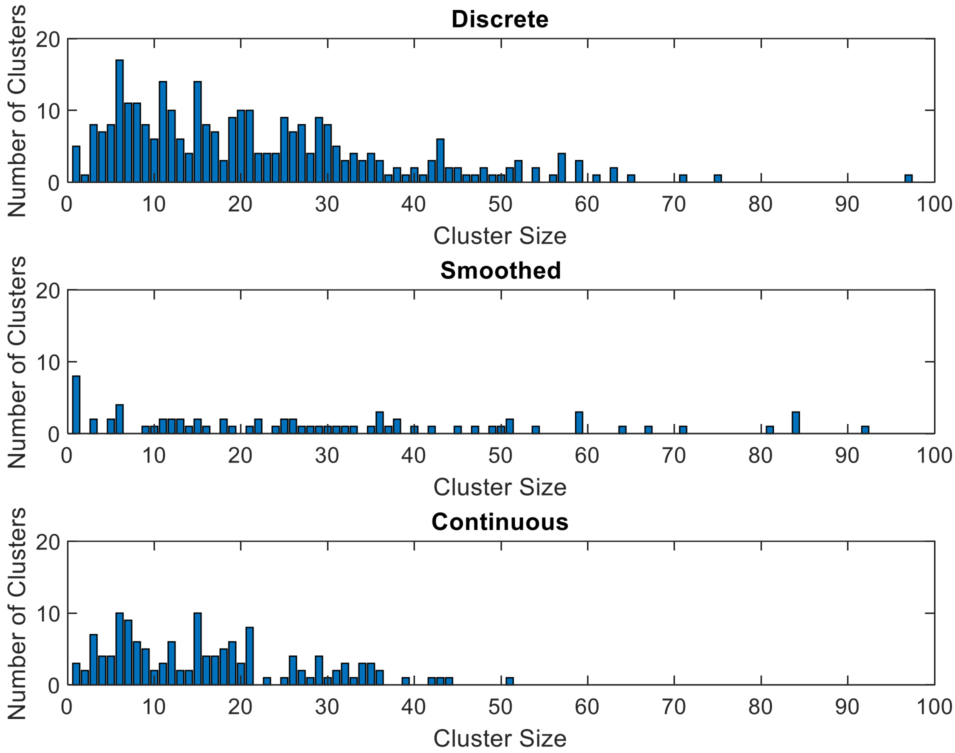


Figure 4-8: Histogram that shows the number of clusters for each cluster size obtained with the different generation algorithms. There are also fourteen clusters with more than 100 consumers in the smoothed case (the biggest one has 965 consumers) and twelve clusters with more than 100 consumers in the continuous case (the biggest one has 813 consumers), which are not shown for the sake of clarity. Source: (Ciller et al., 2019b).

All electrification clusters have less than 100 consumers when generation designs are obtained with discrete diesel capacities. In this case, the generation design for 100 residential consumers has a higher cost per consumer than the generation design for 50 consumers. This causes the clustering algorithm to reach a local optimum, and large mini-grids with low unitary generation costs are never created. This issue, however, does not happen when the smoothed curve or continuous diesel capacities are used. Indeed, there are a few off-grid clusters with almost 1,000 residential consumers in those cases (beyond this point, the economies of scale in generation are negligible).

The off-grid clusters have a strong impact on the electrification solution. Figure 4-9 shows the electrification solutions obtained when the generation costs used to calculate the off-grid clusters are obtained with discrete capacities, the smoothed curve, and continuous capacities. All the solutions use mini-grids to electrify most consumers. Still, their number of consumers is significantly lower when discrete diesel capacities are considered to calculate the off-grid clusters.

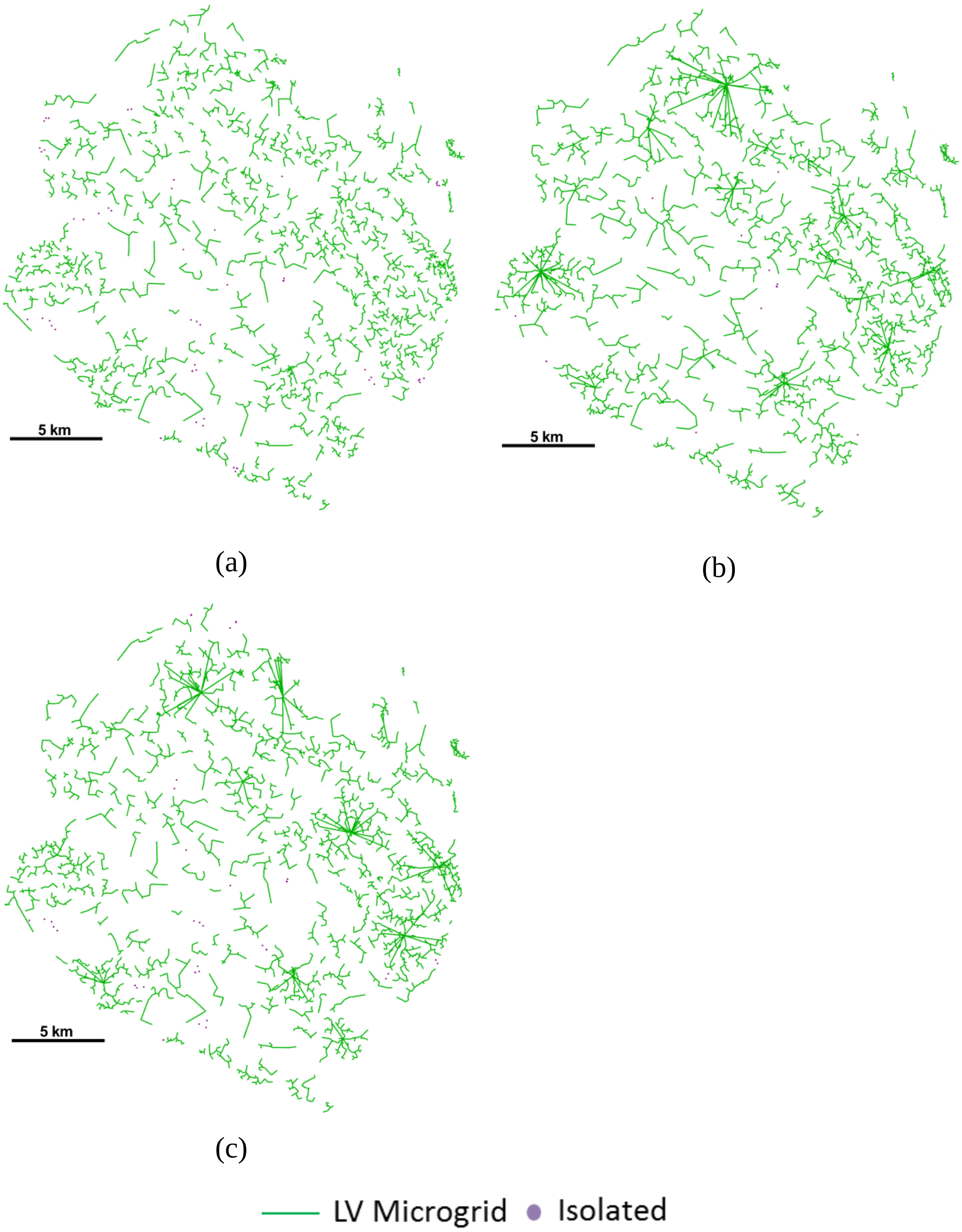


Figure 4-9: Electrification solutions where off-grid clusters are calculated with (a) discrete diesel capacities, (b) the smoothed curve, and (c) continuous diesel capacities. Source: (Ciller et al., 2019b).

Table 4-3 shows the electrification costs obtained with the methods described in this chapter. The generation costs of the off-grid systems that appear in Figure 4-9 are calculated using the generation cost curve based on discrete components to establish a fair comparison among the three methods (i.e., the smoothed curve and the continuous-component implementation are only applied to calculate the off-grid clusters). REM calculates the generation costs of the final solution with the generation designs of Table 4-1, interpolating among the generation designs optimized from scratch when needed (as in the case study presented in section 3.3).

	Consumers grouped considering discrete diesel capacities			Consumers grouped considering the smoothed curve			Consumers grouped considering continuous diesel capacities			Δ All (smoothed)	Δ All (continuous)
	Mini-grids	Isolated	All	Mini-grids	Isolated	All	Mini-grids	Isolated	All		
Number of consumers	6,629	59	6,688	6,674	14	6,688	6,644	44	6,688	0	0
Fraction of consumers	0.99	0.01	1	1	0	1	0.99	0.01	1	0	0
CAPEX per consumer ($/yr)	76.28	108.24	76.56	77.35	108.24	77.41	74.26	108.24	74.48	1.11	-2.72
OPEX per consumer ($/yr)	54.89	167.46	55.89	47.72	167.46	47.97	48.78	167.46	49.56	-14.17	-11.33
CNSE per consumer ($/yr)	0.97	0.85	0.97	0.63	0.85	0.63	0.57	0.85	0.57	-35.05	-41.24
Final cost per consumer ($/yr)	132.15	273.48	133.39	125.70	273.48	126.01	123.61	273.48	124.60	-5.53	-6.59
Total CAPEX ($/yr)	505,677	6,386	512,063	516,232	1,515	517,747	493,385	4,762	498,147	1.11	-2.72
Total OPEX ($/yr)	363,886	9,880	373,766	318,479	2,344	320,823	324,096	7,368	331,465	-14.16	-11.32
Total CNSE ($/yr)	6,444	50	6,494	4,210	12	4,222	3,799	37	3,837	-34.99	-40.91
Final cost ($/yr)	876,006	16,135	892,141	838,920	3,829	842,749	821,281	12,033	833,313	-5.54	-6.59
Fraction of demand served (p.u.)	0.998	0.998	0.998	0.999	0.998	0.999	0.999	0.998	0.999	0.10	0.10
Cost per kWh of demand served ($/kWh)	0.387	0.804	0.391	0.369	0.804	0.37	0.363	0.804	0.366	-5.37	-6.39

Table 4-3: Electrification solution summary for the three different generation algorithms. The last column contains the percentual increment between the "All" columns of the discrete and continuous diesel components cases. The penultimate column contains the percentual

increment between the “All” columns of the discrete diesel components and the smoothed curve cases.

As expected, the final electrification cost is higher when REM calculates the off-grid clusters considering discrete diesel capacities, and it is lower when REM calculates the off-grid clusters considering continuous diesel capacities. Although the exact numbers depend on the network or generation catalog, the crucial point is to notice that grouping the consumers into mini-grids using generation costs based on a model with only discrete generation components may be problematic.

4.5. Conclusions

In this chapter, we introduced further developments of the content presented in section 3.2.1, which deals with the optimization of generation designs in off-grid systems and presents solutions to some issues of the first prototype of REM. An analysis introduced in this chapter studies the consequences of modeling with discrete variables the capacities of generation and storage components. It can be concluded that components with discrete behavior could distort economies of scale in generation, which leads to a problematic behavior of the off-grid clustering.

The chapter introduced two procedures that mitigated the impact of components with discrete behavior on the off-grid clustering of REM. The first one approximated the generation costs with a smooth curve, and the second one modeled the capacity of elements that could alter the economies of scale with continuous variables. Both methods ensured that larger mini-grids benefited from economies of scale in generation when grouping the consumers into mini-grids, but the method based on continuous variables was directly applicable to cases with several types of loads.

The case study shows that a straightforward application of any model based only on discrete components (such as single-system methods or tools) could lead to suboptimal solutions when clustering the consumers into off-grid systems. We can conclude that the two procedures introduced in this chapter led to the better grouping of consumers into mini-grids.

Regarding additional developments, the method presented in this chapter has two significant limitations. Firstly, the number of generation technologies was limited to solar panels and diesel generators, and renewable energies such as wind or hydro should be included in future developments. However, the addition of generation technologies involves dealing with more dimensions when optimizing the generation design of an off-grid system from scratch, which would increase the computation time. Secondly, demand profiles are considered deterministic input parameters, whereas there is much uncertainty about demand in developing countries. Hence, future work should aim at developing a more robust method that can deal with uncertainties.

"Once again I had to put myself in a vulnerable position in order to become stronger."
Cédric Villani

5

ESTIMATION OF THE NETWORK COST IN MINI-GRIDS

The first part of this thesis (chapter 3 and chapter 4) focused on turning the initial prototype of REM into a robust tool that provided consistent results. It took substantial efforts to scrutinize the algorithms of the model, finding the most critical issues, and overcoming them.

This chapter initiates the second part of the thesis, where we focus on developing new algorithms from scratch. REM uses a look-up table to store the generation costs of a reduced number of off-grid systems, and the generation costs of the remaining off-grid systems can be interpolated in the clustering algorithm. It is very logical to wonder if REM could apply a similar procedure with the network costs, optimizing from scratch the network layout of a few systems and somehow using the corresponding results to estimate the network costs of the remaining systems in the clustering.

We explored several ideas until we developed a satisfactory method that calculates a "look-up table" for the network costs of LV mini-grids, which is presented in this chapter. The method provides very accurate estimations of the network costs, and it is applied in chapter 6 to develop a new off-grid clustering algorithm.

This chapter is structured as follows. Section 5.1 briefly describes some network optimization methods and tools. Section 5.2 describes the mini-grid metrics that could be useful in estimating network cost, and section 5.3 presents the network cost estimation method. Section 5.4 introduces a case study, and the results of our method are compared with an estimation aligned with the methods of regional planning tools. Section 5.5 includes the conclusions as well as suggestions for additional developments.

5.1. Network optimization tools and methods

There is plenty of literature related to the optimization of network designs in distribution network planning (Georgilakis and Hatziargyriou, 2015), although these methods were not explicitly designed for planning in underserved regions. One such example is RNM, which REM applies to design distribution networks for mini-grids and extensions of the power grids. RNM was designed to determine appropriate remuneration figures for electric power distribution, and it took some efforts (see section 3.1.4) to use it as a subroutine in REM.

Other distribution network models (Gómez et al., 2013) are similar to RNM, although they generally do not include topographical considerations such as forbidden zones in the optimization process. Most distribution network models apply heuristic methods or metaheuristic techniques, and the use of classical optimization techniques is limited to small networks as the optimization of a distribution network is a computationally-intensive process. ANETO (Garcia Conejo et al., 2007) and Network Performance Assessment Model (NPAM) (Larsson, 2005) are distribution network models that are similar to RNM, and they include the usual electric constraints in their calculations.

Some tools calculate the network design of a single off-grid system, which is generally an underserved village or settlement. ViPOR is a tool that calculates the network of an individual mini-grid by applying a simulated annealing algorithm (Lambert and Hittle, 2000). Reference (Steve Nolan et al., 2017) presents a method that calculates the network layout of a mini-grid in the context of rural electrification. However, the scope of these methods is limited to small mini-grids, and it is unclear if they could be directly extrapolated to regional planning.

Few methods in the literature aim at large-scale network cost optimization in rural electrification planning. Reference (Kocaman et al., 2012) introduces a technique that optimizes from scratch an MV and LV distribution network, calculating the location of the transformers. This technique applies heuristic algorithms based on geometrical considerations, but it does not include electric notions such as power flows or voltage drops.

Regional planning models presented in chapter 2 generally estimate the network costs of mini-grids and grid extensions with analytic expressions or geometric calculations based on the computation of MSTs. They generally provide fast results at the expense of not considering electric constraints (such as power flows or maximum voltage drop allowed) and topographical restrictions (such as forbidden zones and terrain altitudes). These tools usually consider only one line and transformer for each voltage level, limiting the scope of the results severely.

Oppositely, REM accurately calculates the distribution network layout and its corresponding costs for each grid extension design and mini-grid, providing very detailed information at the expense of substantial computation time. REM cannot afford to optimize the network layout of each potential mini-grid and grid extension that the clustering algorithm evaluates, and the model uses quick estimations based on distances and peak demands of clusters. Experience has shown that these estimations are not always accurate, and it is useful to devote effort to improve them.

The rest of this chapter presents a “look-up-table” methodology that captures how the network cost of LV mini-grids behaves without optimizing all the network layouts from scratch. The method selects a set of representative candidate mini-grids, optimizes their networks using RNM, and uses this information to estimate the network cost of the remaining mini-grids.

5.2. Mini-grid metrics

This section describes the metrics that our method considers. The metrics capture the electric and geometric properties that we expect to be representative of the network cost of a mini-grid. Our method does not necessarily use all the metrics described in this section, but it selects the ones that are better estimators of the network cost for each case study.

If several mini-grids are very similar in terms of consumers and demand, they could have similar network costs. Figure 5-1 provides an illustrative example, showing the distribution network of several mini-grids, which have been labeled 1, 25, and 82 in the example. Mini-grids 1 and 25 are very similar (same number of consumers, same demand, and very similar spatial distribution of consumers), and mini-grid 82, although different, could have a similar network cost.

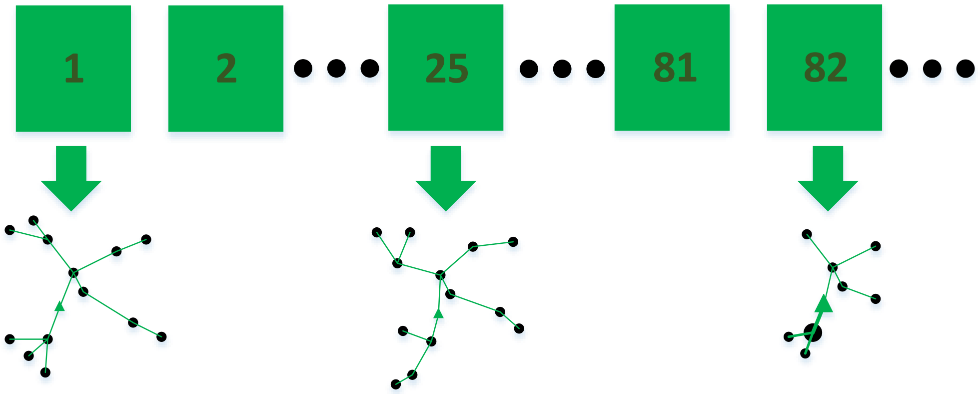


Figure 5-1: Example with similar mini-grids. The black dots in the network layouts represent the consumers. The green triangles represent the generation sites, and the green lines represent the LV distribution network.

There is a clear correlation between several metrics of a mini-grid and its network cost. If we compare two mini-grids that are identical in every aspect but peak demand, then the one with the greater demand is expected to have a higher network cost. Similarly, if we account for the network cost of two mini-grids that only differ in the location of their consumers, then the one with more disperse consumers is expected to be more expensive. Size and demand are, therefore, two important cost drivers that our method considers.

However, the network cost of a mini-grid does not change if we move all its consumers a certain distance or if we rotate them around a certain point. The following properties formalize these intuitive ideas about the metrics:

Cost-monotonicity: when the metric increases its value, the network cost also increases or stays at the same level. All the metrics considered satisfy this property.

Translation-invariance: the network cost of a mini-grid does not change if all the consumers of that mini-grid are translated a specific distance in the same direction. All the metrics satisfy this property.

Rotation-invariance: the network cost of a mini-grid does not change if all its consumers are rotated a specific angle around the same point.

Scale-monotonicity: If we scale a mini-grid so that its consumers are more disperse, network cost will increase so the values of the metrics should also increase. All the metrics satisfy this property.

Table 5-1 presents the metrics, which are described in the remaining of this section.

Type	Variable
Electric	Electric moments (central)
	Electric moments (rotation)
	Aggregated and peak demand
Spatial	Minimum-spanning-tree length
	Minimal area rectangle (min{width, height}, area, perimeter)
Other	Number of consumers

Table 5-1: Mini-grid metrics.

5.2.1. **Electric Metrics**

In this section, we briefly introduce the electric metrics that our method considers. These metrics include the electric central moments, the electric rotation moments, and the demand-related metrics.

5.2.1.1. Electric Moments (central)

The (p, q) central electric moment is defined by the expression:

$$\mu_{p,q} = \iint (x - \bar{x})^p (y - \bar{y})^q f(x, y)\, dxdy \qquad 5\text{-}1$$

Where the integral limits are given by the boundaries of the mini-grid and $f(x, y)$ is the peak demand of the consumer (x, y). In practice, we have a discrete number of consumers c so the integral becomes a summation when computing the moments.

$$\mu_{p,q} = \sum_{i=1}^{i=c} (x_i - \bar{x})^p (y_i - \bar{y})^q P_i \qquad 5\text{-}2$$

The $(n + 1)$ moments of order n are given by the solutions of the equation $n = p + q$, being p, q nonnegative integers. The moments of odd order decrease when a consumer with negative coordinates $(-x, -y)$ is included in a mini-grid. This causes undesired effects since there is no difference in the final network cost between adding a consumer with coordinates (x, y) or $(-x, -y)$ to a symmetric (with respect to the origin of coordinates) mini-grid. Therefore, all moments of odd order and moments of even order where variables have an odd exponent do not satisfy the cost-monotonicity property, and we do not consider them useful predictor variables.

Hence, we include only the central moments of even orders where all the variables have even exponents. In practical terms, it is enough to include central moments of orders 2, 4, and 6. The inclusion of additional central moments does not tend to improve the accuracy of the models and generally produces collinearity among the variables.

The central moments included meet all the properties described in this section except rotation invariance. The electric rotation moments are included to compensate for that.

5.2.1.2. Electric Moments (rotation)

Reference (Ming-Kuei Hu, 1962) introduces the rotation moments for image recognition, and they are translation and rotation invariant.

$$I_1 = \mu_{2,0} + \mu_{0,2} \tag{5-3}$$

$$I_2 = (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \tag{5-4}$$

$$I_3 = (\mu_{3,0} - 3\mu_{1,2})^2 + (3\mu_{2,1} - \mu_{0,3})^2 \tag{5-5}$$

$$I_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{2,1} + \mu_{0,3})^2 \tag{5-6}$$

$$\begin{aligned} I_5 = (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] + (3\mu_{2,1} \\ - \mu_{0,3})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] \end{aligned} \tag{5-7}$$

$$\begin{aligned} I_6 = (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{2,1} \\ + \mu_{0,3}) \end{aligned} \tag{5-8}$$

$$\begin{aligned} I_7 = (3\mu_{1,2} - \mu_{3,0})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] \\ + (\mu_{0,3} - 3\mu_{2,1})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 \\ - (\mu_{2,1} + \mu_{0,3})^2] \end{aligned} \tag{5-9}$$

Reference (Flusser, 2000) highlights that these moments are do not form a complete or independent basis and adds another third-order rotation moment:

$$\begin{aligned} I_8 = \mu_{1,1}\left[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2\right] - (\mu_{2,0} - \mu_{0,2})(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} \\ + \mu_{2,1}) \end{aligned} \tag{5-10}$$

Since I_1 is a linear combination of electric central moments of order two we will not consider it as a candidate variable, and we will include the remaining rotation moments $I_2 - I_8$. The rotation moments included meet all the properties described in this section.

5.2.1.3. Demand

Although the peak demands of individual consumers are already considered in the calculation of electric moments, the aggregated and peak demands of a mini-grid may provide valuable information, and they are included in the metrics considered. The demand-related metrics included meet all the properties described in this section.

5.2.2. **Spatial Metrics**

In this section, we briefly describe the spatial metrics that our method considers. These metrics include the MST length and several metrics related to the minimal area rectangle.

5.2.2.1. Length of the MST

The length of the MST linking all consumers is calculated with the consumers of the mini-grid and the generation site located at the demand-weighted center of the mini-grid (REM always locates the generation site at this spot). This metric meets all the properties described in this section.

5.2.2.2. Minimal Area Rectangle

The term Minimal Area Rectangle (MAR) refers to the rectangle of minimum area that contains all the consumers of the mini-grid. The main geometric attributes of the minimum rectangle of a set of points (consumers) are rotation-invariant, translation-invariant, and scale-variant, meeting all the desirable properties that we identified beforehand. Besides, any increase in these metrics should lead to a more substantial network cost. Figure 5-2 shows an example with several points and their corresponding MAR.

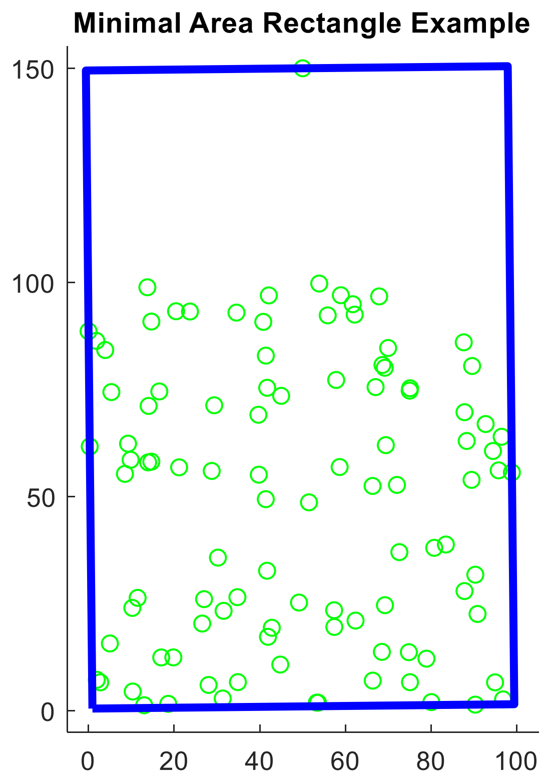


Figure 5-2: Minimal area rectangle example.

MAR is sensitive to extreme values. The example provided in Figure 5-2 has one extreme value with coordinates (50, 150) that increases the values of its area and perimeter significantly. Since the perimeter depends linearly on the width and the height of the rectangle, the metrics considered include area, perimeter, and the minimum between its height and width.

5.2.3. **Other metrics**

This section describes additional metrics that are useful to consider. For the time being, the only metric that belongs to this section is the number of consumers of the mini-grids.

5.2.3.1. Number of consumers

We included several metrics that measure the "size" of a mini-grid (such as the length of the MST or the aggregated demand). The number of consumers of a mini-grid is an additional metric correlated to its size, and it is also considered. This metric meets all the properties described in this section.

5.3. Method

Our method assumes that all mini-grids deploy an LV distribution network, which is consistent with the usual regional planning results. Our method also assumes that the topographical features of the terrain are not critical in the cost of the distribution networks (i.e., for the time being, our method does not explicitly consider the impact of topography). Figure 5-3 shows a flow chart of the method, which follows three sequential steps. The first step is *network assignment,* which calculates the number of linear regression models needed and assigns the mini-grids to the models (a data point represents each one). The second step is *clustering*, which applies a k-medoids algorithm to obtain a set of representative mini-grids for each linear regression model. The final step is the *calibration of linear models,* which calculates network designs for the representative mini-grids, and determines the metrics and coefficients of the linear regression models.

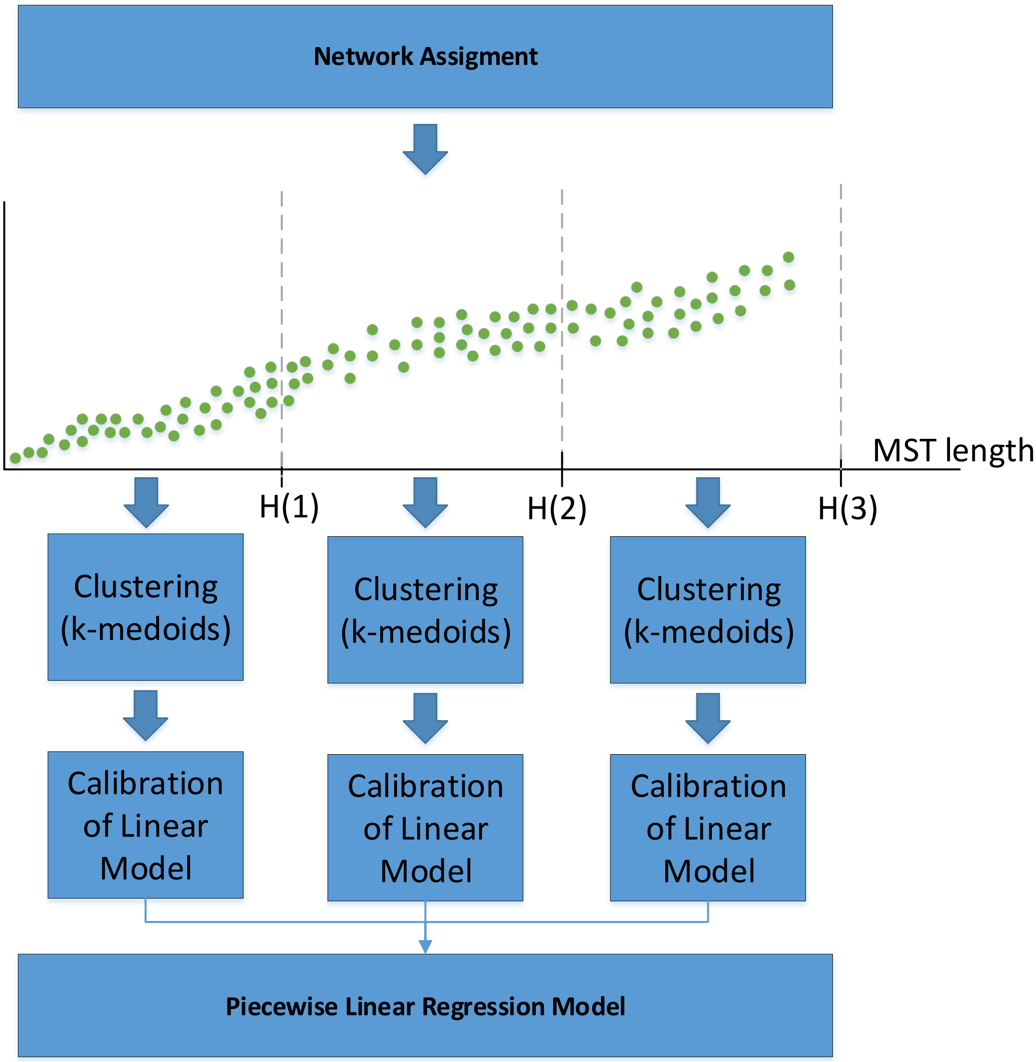


Figure 5-3: Flow diagram of the network cost estimation algorithm.

One advantage of using several linear models to estimate the network cost of the mini-grids is that our method can select different metrics and coefficients for each linear model.

For example, a linear model with small mini-grids that supply only residential consumers could consider only the length of the MST of the mini-grids to estimate their network cost accurately. However, another linear model that estimates the network cost of large mini-grids that include productive loads could also include several electric metrics to capture the impact of productive loads and provide accurate estimations of the network costs.

Even if the metrics of several linear models are the same, their coefficients will be different, and the cost estimation will be better than the one obtained with a single linear model. The rest of this section describes our method in detail.

5.3.1. **Network assignment: determination of the number of models to fit and their correspondence with particular mini-grids**

The *network assignment* step determines the number of linear regression models to use, and it distributes all mini-grids of the case study among the models. Two mini-grids are assigned to the same linear model if and only if their network costs are of the same order of magnitude. One way of achieving this is to ensure that the quotient between the lengths of the MSTs of any pair of mini-grids that belong to the same model is lower or equal than a threshold Q. Equation 5-11 forces this constraint for each model $n \in \{1, 2, \dots, N\}$:

$$\frac{max(length(MST_n))}{min(length(MST_n))} \leq Q \tag{5-11}$$

Where MST_n refers to the MSTs of the networks assigned to the $n-th$ model. The total number of models N depends on each case and must be calculated. Let u and U be the minimum and maximum lengths of all the MSTs (in km), respectively. We consider the sequence:

$$\{u, Q \cdot u, Q^2 \cdot u, \dots, Q^{N-1} \cdot u, Q^N \cdot u\} \tag{5-12}$$

It is clear that the quotient between any pair of consecutive terms in the sequence is equal to Q, so N linear models are necessary and sufficient[6] as long as N is the minimum natural number that satisfies equation 5-13:

$$U \leq Q^N \cdot u \tag{5-13}$$

Dividing both sides by u and taking logarithms yields:

$$log_Q(U/u) = log_Q(U) - log_Q(u) \leq N \tag{5-14}$$

So equation 5-15 provides the minimum number of linear models needed in terms of U, u and Q.

$$N = \lceil log_Q(U) - log_Q(u) \rceil \tag{5-15}$$

Where $\lceil x \rceil$ is the lowest integer that is greater than or equal to x. However, equation 5-15 implies that the number of models tends to infinite if u tends to zero, so equation 5-15 is only applied if $u \geq 1/Q$ to avoid potential trouble. Otherwise, the number of models is determined

---

[6] *If equation 5-13 holds and N is the minimum natural number that satisfies equation 5-13, then we have $u < Q \cdot u < Q^2 \cdot u < \dots < Q^{N-1} \cdot u < U \leq Q^N \cdot u$ and we could group the mini-grids in N linear models whose MST lengths lie in the ranges $[u, Q \cdot u)$, $[Q \cdot u, Q^2 \cdot u)$,…, $[Q^{N-1} \cdot u, Q^N \cdot u]$. Similarly, if equation 5-13 does not hold we would need at least $N + 1$ linear models to group all the mini-grids ensuring that equation 5-11 is satisfied.*

considering that a linear model could cover the range of MST lengths $[u, 1]$, and $N - 1$ models could cover the range $[1, U]$. Equation 5-15 is applied to the range $[1, U]$ to determine $N - 1$, which yields:

$$N = \lceil log_Q(U) - log_Q(1) \rceil + 1 = \lceil log_Q(U) \rceil + 1 \tag{5-16}$$

Equation 5-16 is also valid if $1/Q < U \leq 1$ because in that case only one model is needed to cover the range $[u, U] \subset [u, 1]$, and $\lceil log_Q(U) \rceil = 0$ so N is set to the correct value. If $U \leq 1/Q$, then N is directly set to 1. Equation 5-17 comprises all the expressions used to calculate N in terms of U, u and Q.

$$N = \begin{cases} \lceil log_Q(U) - log_Q(u) \rceil & if \;\; U > u \geq 1/Q \\ \lceil log_Q(U) \rceil + 1 & if \;\; U > 1/Q > u \\ 1 & if \;\; 1/Q \geq U > u \end{cases} \tag{5-17}$$

Once the number of models has been determined, the mini-grids are distributed among the linear models. This process initializes each model with the same number of mini-grids. It calculates the quotients between the maximum and minimum lengths of MSTs for each model and, for those models that do not satisfy equation 5-11, it reassigns mini-grids from the closest models (if they meet equation 5-11) to reduce their quotient. This process goes on iteratively until equation 5-11 holds for each model (the first model may not be forced to satisfy this equation if $u < 1/Q < U$). Figure 5-4 shows a stylized flow diagram of the network assignation.

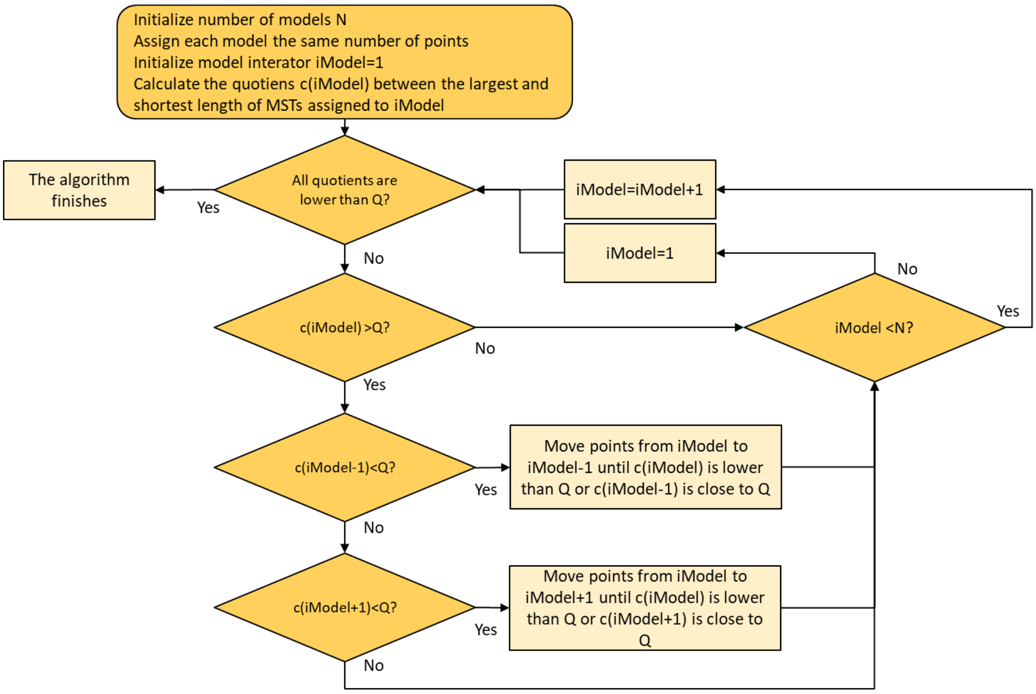


Figure 5-4: Stylized flow diagram of the network assignment.

The threshold Q is set to ten in the case study (Weisstein, 2019), although other values are possible because equation 5-11 implicitly assumes that there is a perfect linear correlation between the MST and the network cost, which is not true (although the correlation is usually very high). Setting Q to a value slightly lower than ten could mitigate the impact of this assumption because the maximum cost difference among the mini-grids that belong to the same linear model would be reduced.

5.3.2. **Clustering (k-medoids)**

Once the algorithm has assigned the mini-grids to linear models, the clustering step obtains a representative set of mini-grids for each linear model. Detailed network designs are calculated later for representative mini-grids, so the outcome of the clustering should be real mini-grids that exists in the case study. To that end, the clustering step applies a k-medoids algorithm, which also turns out to be more robust to outliers than other methods such as k-means (Razavi Zadegan et al., 2013).

There are several implementations of the k-medoids algorithm in the literature. The Partitioning Around Medoids (PAM) calculates an initial solution, and then performs all possible swaps among medoids and non-medoids to improve the solution (Kaufman and Rousseeuw, 1990). The main drawback of PAM is that it is a computationally-intensive process that does not perform well when dealing with large datasets. Clustering for LARge Applications (CLARA) tries to overcome this drawback by applying PAM only on a reduced sample of the original dataset, trading optimality for computation speed. The Clustering LARge Applications based on RANdomized Search (CLARANS) also samples only a part of the original dataset, but the sample is not selected beforehand (Ng and Jiawei Han, 2002).

Our method applies the Matlab build-in function to perform the k-medoids algorithm (MathWorks, 2014), which uses the PAM algorithm if the number of input points is lower than 3,000. If the number of input points is between 3,000 and 10,000, then it applies a method based on reference (Park and Jun, 2009). If the number of input points is larger than 10,000, then it evaluates a subset of the data following a procedure similar to CLARANS.

K-medoids is applied several times with an increasing number of clusters, and the sum of point-to-medoid distances are computed. The process stops when the marginal gain of increasing the number of clusters drops below a pre-established threshold (Bholowalia, 2014).

The clustering step always considers the length of the MST (a representative spatial metric) and the aggregated demand (a representative electric metric) to determine the representative mini-grids. These two metrics are since they are measured with different units. The length of the MST is scaled with the average cost of the LV lines of the catalog, which is a reasonable estimation of how much the network cost would increase for a given increment of the MST length. The aggregated demand is scaled by an estimation of the network cost per kWh in the analysis region, which could be obtained by expert advice or looking at previous reports or publications that deal with electrification planning projects in the corresponding region.

5.3.3. **Calibration of linear models**

This step calculates accurate network designs for the representative mini-grids using RNM, and it determines the most representative metrics for each model and their coefficients applying hierarchical regression (Pedhazur, 1997), which has been successfully applied in many fields (Megherbi et al., 2006; Moller et al., 2003).

This technique starts with an initial linear model and adds blocks of variables sequentially. The process continues until the addition of variables does not improve the model, or there are no more variables to add. The hierarchical levels determine the order followed to include variables in the model: the first hierarchical level corresponds to the initial linear model, the second level contains the first block of variables added to the model, and so on.

The expertise of the analyst plays a critical role in hierarchical regression, as he or she needs to determine which variables are assigned to each hierarchical level. The most relevant variables should be included in the initial levels, and the less important ones should be incorporated in later stages.

We also considered applying stepwise regression, which is a similar procedure that starts with an initial linear model, and variables are introduced or removed sequentially. In stepwise regression, the sequential order of variables is determined by their statistical significance (i.e., the computer determines the order by calculating p-values).

Results are hard to replicate with stepwise regression, as small variations in the dataset could lead to different regression models (Lewis, 2007). This would be a significant issue since we are obtaining our representative networks with a k-medoids algorithm (whose outcome may depend on the initial solution).

Table 5-2 shows the variables included in each hierarchical level (a hierarchical level also includes all the variables of the previous levels).

Hierarchical level	Variables added
1	Length of the MST
2	Number of consumers
3	Central moments of order 2
4	Central moments of order 4
5	Central moments of order 6
6	Minimum area rectangle metrics
7	Aggregated demand
8	Rotation moments

Table 5-2: Models used in the hierarchical regression.

We consider the length of the MST as the most relevant variable, followed by the number of consumers of the mini-grid and its central moments, which are grouped by their order. In this way, levels 1-5 include the spatial, electric, and “other” metrics that we consider paramount. Levels 6-8 include the remaining metrics.

To avoid collinearity, our method applies the Belsley collinearity test (Belsley et al., 1980) and removes the collinear variables from their corresponding hierarchical level. We also compute the p-values to check at which point it is not worth adding more variables to the model.

5.4. Case study

This section presents an application to a case study located in Rwanda, comparing the exact network cost of mini-grids with the approximation that our method provides. The location of the consumers and their demand profiles are based on the case presented in (Rwanda Energy Group (REG), 2019). The location of residential consumers was obtained combining information from the HRSL (Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University, 2016), a report from SOFRECO (SOFRECO, 2013), and the expected population growth for 2024 (National Institute of Statistics of Rwanda (NISR) and Ministry of Finance and Economic Planning (MINECOFIN) [Rwanda], 2012). Energy Development Corporation Limited (EDCL) provided the location of the productive loads and their peak and average demands. The hourly demand profile of the loads was estimated according to in-the-field surveys conducted in the village of Gicumbi (Li, 2016; Santos Pérez, 2015).

The network and generation catalogs are based on the experience of the Universal Access Laboratory, which is based on their participation in several projects combined with field trips and conducted interviews.

The case study has 1,598,842 unelectrified consumers, which REM has grouped into 24,381 candidate mini-grids (off-grid clusters). Many candidate mini-grids (97.73%) have at most 200 consumers, and the biggest candidate mini-grid has 647 consumers. The location of the consumers and the candidate mini-grid sizes are shown in Figure 5-5.

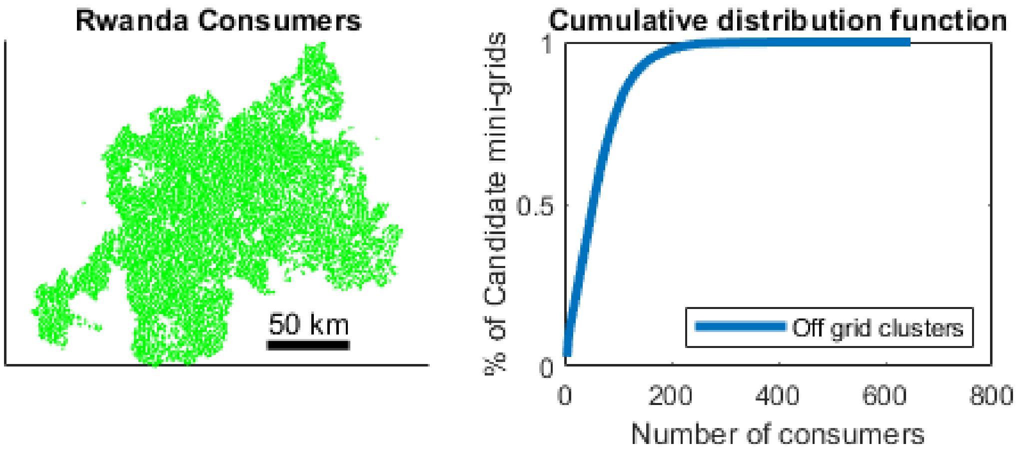


Figure 5-5: Rwanda unelectrified consumers (left) and candidate mini-grid sizes (right).

Our method groups the mini-grids into three linear models, and the lowest and the largest

MST lengths are 0.035 km and 13.64 km, respectively. Figure 5-6 shows the mini-grids that our method assigns to the linear models.

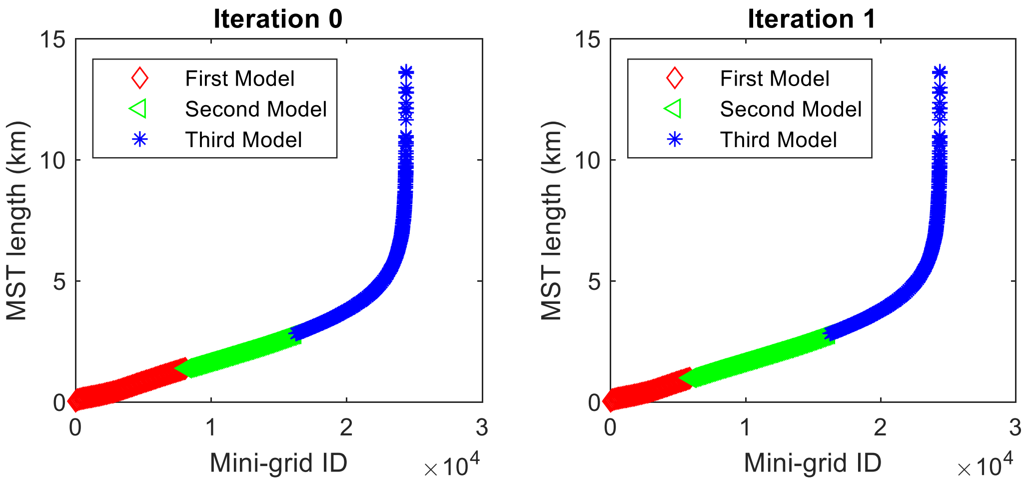


Figure 5-6: Network assignation. This procedure firstly assigns the same number of networks to each model and then reassigns networks from the first to the second model. The networks are sorted according to the length of the MST.

Table 5-3 shows the number of mini-grids that are assigned to each model.

	Condition for assignation (km)	Number of mini-grids
1st Model	$length(MST) < 1$	5,890
2nd Model	$1 \leq length(MST) \leq 2.832$	10,353
3rd Model	$2.832 < length(MST)$	8,138

Table 5-3: Mini-grids assigned to each model.

For each model, the procedure selects a few representative mini-grids. The number of mini-grids is determined imposing that the marginal gain of having more mini-grids drops below 10% and the candidate number of mini-grids considered are 25, 31, 40, 50, 63, 79, 100, 126, 159 and 200 (ten points logarithmically spaced between 25 and 200). We scale the MST length with a factor of 10,132.3 $/km and the aggregated demand with a factor of 0.4 $/kWh.

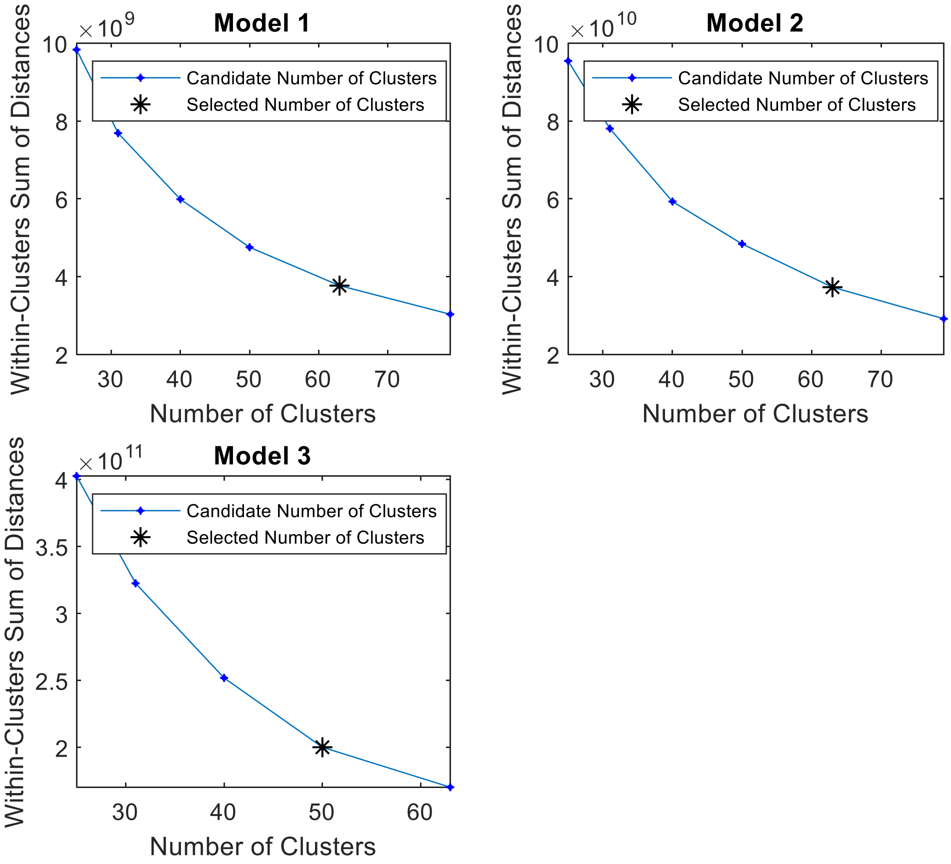


Figure 5-7: Number of clusters selection. Our method selects sixty-three clusters for the first and the second model, and fifty clusters for the third one.

The method adjusts three linear models calculating the network costs of the representative mini-grids and using hierarchical regression. We consider that it is not worth to add more variables into the model when the marginal gain of the adjusted R^2 is less than 10%, which corresponds to the third hierarchical model in all cases (see Figure 5-8).

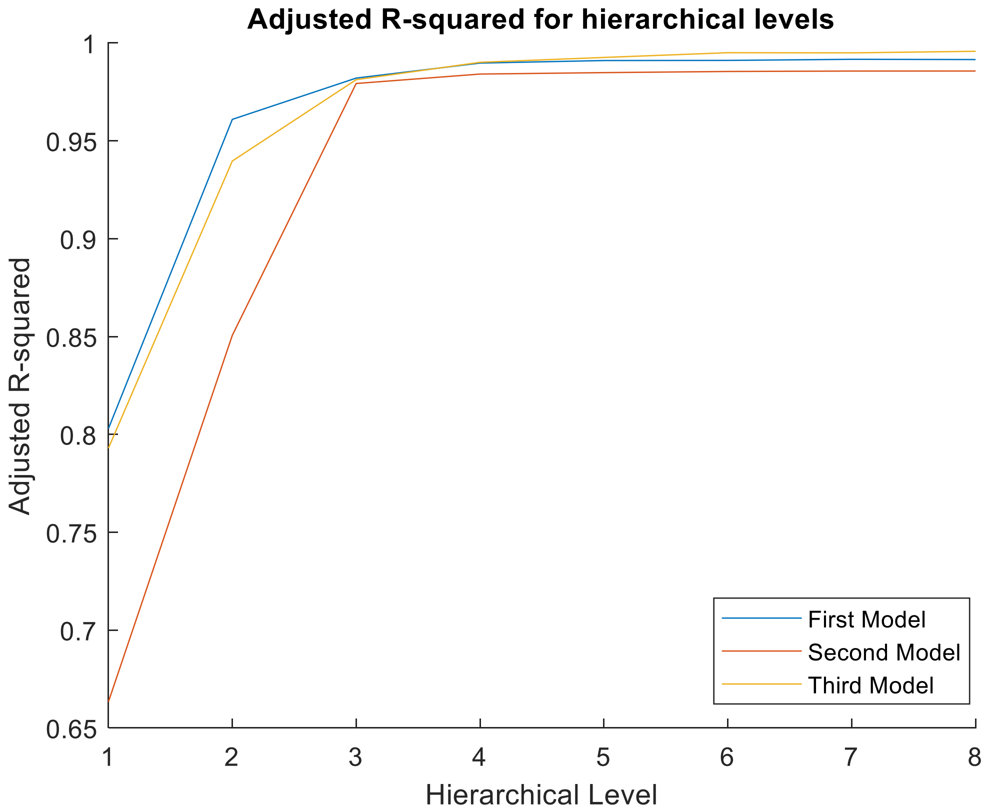


Figure 5-8: Adjusted R^2 for each hierarchical level.

Although the three linear models have the same explanatory variables (length of the MST, number of consumers, and central moments of order two), it is still advantageous to have different linear models as the coefficients are different.

We compare the exact and approximated network cost of the 24,381 candidate mini-grids to obtain the relative linear error (which is defined as the absolute value of the quotient between the difference of costs and the real cost) that we incur with the model. Figure 5-9 provides the corresponding results.

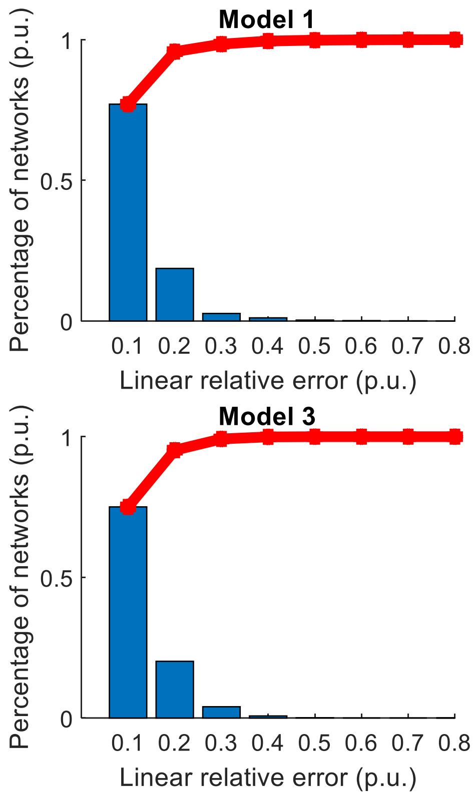


Figure 5-9: Linear relative error (p.u.) for each model.

We also compare our method with a linear model that only considers the MST length to estimate the network cost (in this case, we directly force the model to use 200 representative networks instead of using the marginal gain procedure described in section 5.3.2). Most regional planning tools apply techniques based on the calculation of an MST to calculate the network costs.

Figure 5-10 shows the linear error (p.u.) obtained for the 24,381 networks with both procedures. The naïve estimation provides a linear relative error lower than 20% for 56.46% of the networks, whereas our method obtains a linear relative error lower than 20% for 96.74% of the networks.

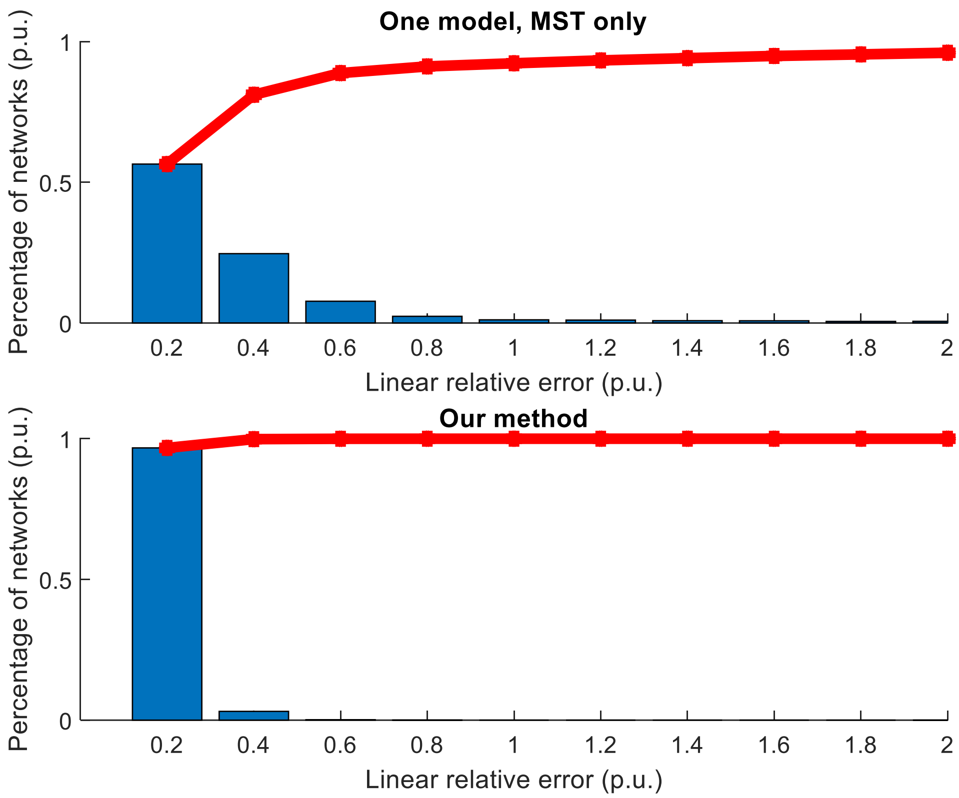


Figure 5-10: Linear relative error comparison.

Figure 5-11 shows the linear absolute error obtained with both procedures (which is defined as the absolute value of the difference of costs). Our method obtains a linear absolute error lower than 100 $/yr for 84.36% of the networks, and the more straightforward procedure obtains the same error for only 57.93% of the networks.

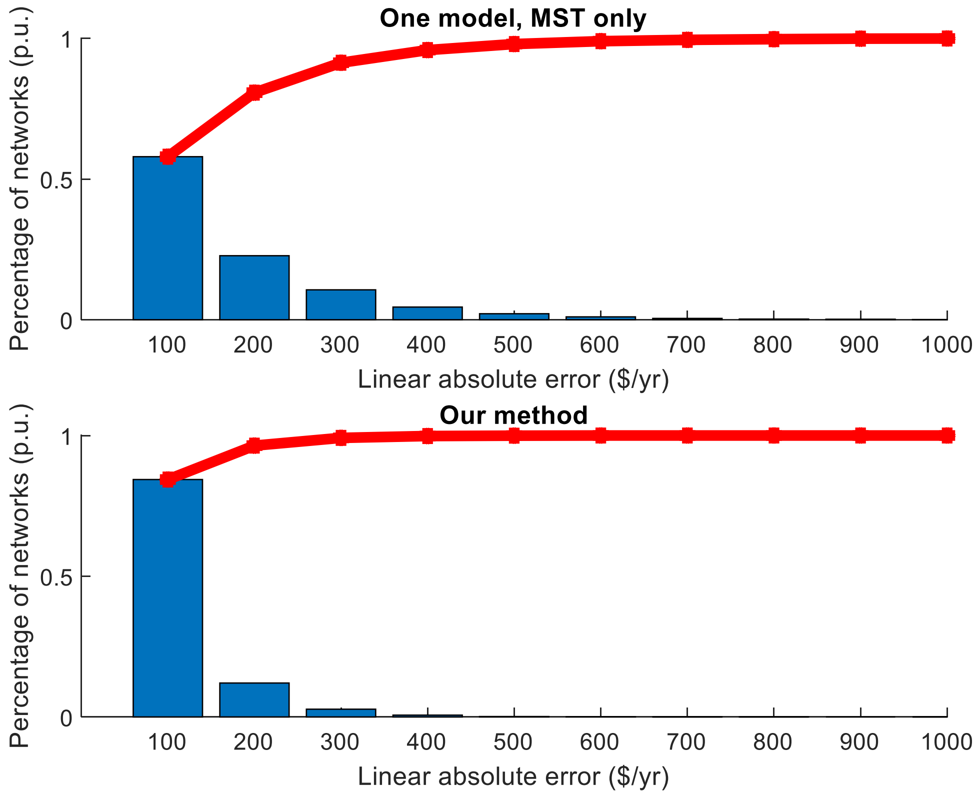


Figure 5-11: Linear absolute error comparison.

Results show that our method significantly improves naïve approximations without increasing computation time significantly, as it requires only the calculation of accurate designs for less than 1% of the total mini-grids in this case study. Table 5-4 shows the computation times required to calculate accurate network designs for the representative mini-grids and all the mini-grids.

Representative mini-grids	All mini-grids
15 min, 28.62 sec	23 hours, 52 min, 38.48 sec

Table 5-4: Computation times needed to calculate the network designs.

The computation time needed to calculate the network designs for the representative mini-grids is 1.08% times the computation time required to calculate the network designs for all the mini-grids.

5.5. Conclusions

REM calculates detailed network designs for each mini-grid and grid extension that appears in the final electrification solution. However, the model cannot afford to optimize the network layout of each potential clustering solution that REM evaluates. In the clustering algorithm, REM uses quick estimations of the incremental network cost that are based on peak demands and distances among clusters (or a cluster with the power grid). These estimations do not capture the inherent complexity of distribution networks, and it is useful to improve them as they play a critical role in the clustering of REM.

This chapter presents a method that estimates the network cost of all the potential LV mini-grids of a large-scale electrification case. The algorithm distributes the mini-grids among several linear regression models, performs a clustering algorithm, and only optimizes the network designs of the representative mini-grids. A hierarchical regression method selects the mini-grids attributes that are adequate to estimate the network cost for each linear regression model. The explanatory variables considered include both spatial and electric metrics, and the number of consumers of the mini-grid.

We compare the exact network costs of mini-grids with the estimations that our method provides in a realistic large-scale case study, where our method optimizes the network designs of less than 1% of the mini-grids to obtain the estimations. The computation time required to optimize the networks of all the mini-grids is approximately one day. In contrast, the computation time needed to optimize the networks of the representative mini-grids is around fifteen minutes.

We also compare our method with a more straightforward estimation that is aligned with the rules of thumb that some regional planning tools apply. Our method provides an estimation where the linear relative error is lower than 20% for 96.74% of the mini-grids, whereas the straightforward method provides an estimation where the linear relative error is lower than 20% for only 56.46% of the mini-grids. We also compare the linear absolute errors, and our method obtains an error lower than 100 $/yr for 84.36% of the mini-grids, and the more straightforward procedure obtains the same error for 57.93% of the mini-grids. We can conclude that a straightforward method (such as the ones that regional planning tools apply) may lead to significant errors when estimating the network costs.

Regarding additional developments, the method should be expanded to estimate the cost of MV mini-grids too. Most of the steps of the current approach could hold for MV mini-grids, but it may be necessary to change the hierarchical order of explanatory variables and add new variables because transformers could account for a significant amount of the total network cost.

It would also be interesting to explore the use of different hierarchical orders of variables for the different linear models. Some variables could be critical for small mini-grids with a reduced number of consumers and a low aggregated demand profile. Still, they may be less important for large mini-grids with a significant amount of consumers and a high aggregated demand profile.

"If I'd observed all the rules, I'd never have got anywhere."
Marilyn Monroe

6

CLUSTERING IN LARGE-SCALE ELECTRIFICATION PLANNING

REM's clustering improved significantly after implementing the improvements described in section 3.2.2, but it still presented some weaknesses. We use the term *standard clustering* to refer to the clustering algorithm presented in section 3.1.3 with the improvements described in section 3.2.2.

The standard clustering applies quick estimations of the network cost when determining whether nearby clusters should be joined. These estimations are crucial to the optimality of the clustering, and they do not capture the complex behavior of distribution networks.

The economies of scale in generation also play a crucial role in the standard off-grid clustering of REM. In chapter 4, we analyzed the impact of modeling the capacities of diesel generators with discrete variables in the final electrification solution. REM's standard clustering cannot deal with elements or constraints that distort the economies of scale in generation.

In this chapter, we apply the network cost estimator introduced in chapter 5 to develop a new off-grid clustering (exhaustive clustering) that overcomes the two main limitations of the off-grid clustering: the use of inadequate network cost estimations and the inability to handle elements or constraints that are incompatible with a monotonous behavior of the economies of scale in generation.

This chapter also describes a new grid-extension clustering (top-down clustering) that was jointly developed with Olamide Oladeji, which was an MIT student at that time. The top-down clustering starts designing an extension of the power grid that electrifies all the consumers and then performs cost-comparisons to "disconnect" parts of the network, electrifying the downstream consumers with off-grid systems.

Part of the content related to the top-down clustering comes from the following paper:

Oladeji, O., Ciller, P., de Cuadra, F., Perez-Arriaga, I. Partitioning Distribution Networks: An Approach to Integrated Electrification Planning. IEEE Transactions on Power Systems. Submitted.

The rest of this chapter is structured as follows: section 6.1 briefly describes some clustering methods in the context of large-scale electrification. Section 6.2 explains the limitations of the standard off-grid clustering of REM, introduces the exhaustive clustering and its application to a case study. Section 6.3 introduces some limitations of the grid-extension clustering, presents the top-down clustering and its application to a case study. Finally, section 6.4 presents the conclusions and suggestions for additional developments.

6.1. Clustering methods

The clustering of consumers into combinations of standalone systems, mini-grids and grid extensions plays a key role in determining the best techno-economic plan in an underserved region, and many models perform crude simplifications to address it (Morrissey, 2019). Most tools do not optimize the clustering of consumers into systems, and they consider the villages, settlements, or cells as the natural clustering of consumers. However, the use of administrative or artificial divisions as clusters may lead to inefficient solutions from the techno-economic point of view. For example, the electrification of a village with off-grid systems could be less expensive if the solution includes a smart combination of mini-grids and standalone systems instead of a single mini-grid that electrifies all its consumers.

There are few clustering applications in the context of electrification planning in an underserved region. Reference (Parreno Jr and Del Mundo, 2015) introduces a clustering algorithm based on a weighted MST and applies it to an individual village. The GEOSIM tool clusters villages around *development poles* (villages that are assigned a high score based on several indicators) using an algorithm based on the Huff model (Huff, 1963). OnSSET incorporates a clustering algorithm that merges adjacent cells into clusters (Korkovelos et al., 2019). Similarly, reference (Blechinger et al., 2019) presents an electrification planning methodology that generates buffers around populated areas, polling units, and schools. Then, it clusters overlapping zones together. These methods are based on geospatial considerations, and their main drawback is that they do not consider the costs involved in the final electrification solution and the trade-offs among them when clustering the consumers.

There is some literature concerning the clustering of consumers in distribution network planning. Reference (González-Sotres et al., 2013) introduces an algorithm that applies a k-means clustering to optimize the size and location of transformers in distribution planning. The RNM (Mateo Domingo et al., 2011), which REM uses to design distribution networks for mini-grids and grid extensions, groups consumers into settlements following proximity criteria. NPAM (Larsson, 2005) includes a bottom-up clustering where a cluster starts with a single element, and nearby elements are included if they meet specific criteria. The ANETO model (Garcia Conejo et al., 2007) divides the analysis region into cells and groups adjacent cells together. However, models and methods that aim at distribution network planning electrify all the consumers with extensions of the power grid, and they do not include off-grid alternatives as viable electrification solutions. This implies that it may not be straightforward to extrapolate these methods to the electrification of underserved regions, where mini-grids and standalone systems play a key role in planning.

6.2. Exhaustive clustering

This section first recapitulates the off-grid clustering, presenting its current limitations. Then, it introduces a new algorithm, *exhaustive clustering*, that is based on the off-grid clustering, but it overcomes its main weaknesses. Figure 6-1 shows how the exhaustive clustering fits in the current version of REM.

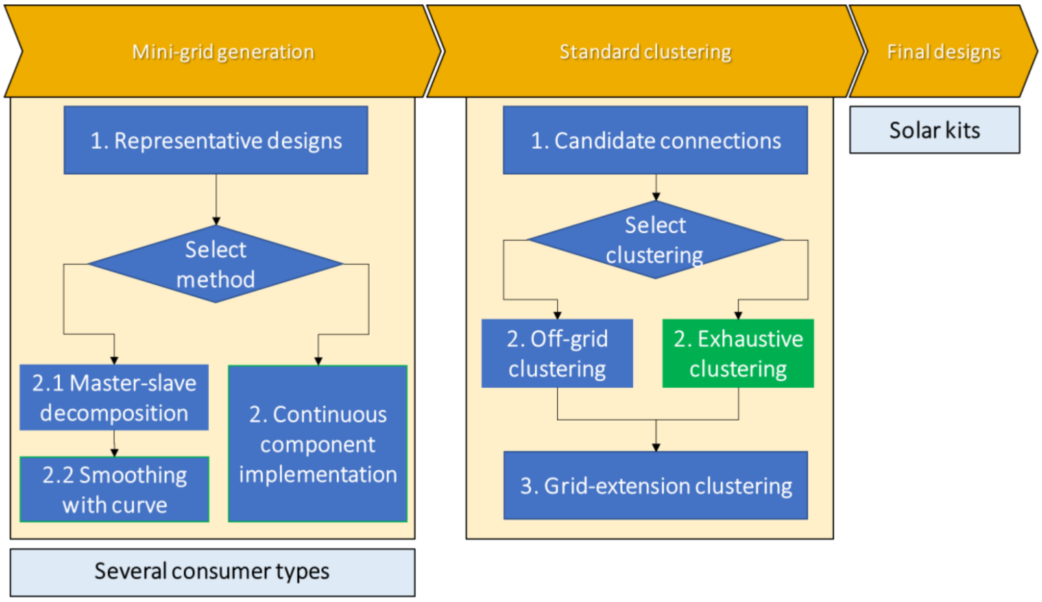


Figure 6-1: Incorporation of the exhaustive clustering into REM's algorithmic structure.

The exhaustive clustering is an alternative to the off-grid clustering: both algorithms group the consumers into off-grid clusters, which represent the candidate off-grid solutions for a case.

6.2.1. The limitations of REM's off-grid clustering

The standard clustering starts with each consumer being its own individual off-grid system. Then, it performs cost-comparisons among nearby clusters to determine if it is worth electrifying them with the same off-grid system (and hence it joins the corresponding clusters). The cost-comparisons are based on trade-offs: large mini-grids (i.e., with a high aggregated demand and many consumers) benefit from economies of scale in generation and management costs, but they have substantial network costs.

The generation and management costs considered in the off-grid clustering accurately represent the costs of the systems. However, the incremental network cost incurred if the clusters are merged is estimated with a line that connects the centers of both clusters.

There are two reasons for using the cost of a line to estimate the incremental network cost. On the one hand, it is not computationally feasible to obtain accurate network designs for the three possible off-grid systems (the first cluster, the second cluster, and the merged cluster) each time a cost comparison is performed. On the other hand, the off-grid clustering starts with all consumers isolated and then joins nearby consumers generating bigger clusters. The bottom-up, irreversible logic could easily reach a local minimum if the incremental network costs are higher than the management and generation savings in the initial cost-comparisons. This could happen if accurate network designs were calculated for all off-grid systems since the lines of a network catalog are usually oversized for a few consumers with residential demand profiles in unelectrified regions. This effect is mitigated by allowing interpolation among the lines of the network catalog to estimate the incremental network cost (the use of a “continuous” catalog in the clustering is explained in section 3.2.2.3.2).

The cost of an interpolated line, however, is not always a good estimator of the incremental network cost among off-grid systems. Network designs follow a non-linear behavior as they have to comply with electric constraints such as Kirchoff's laws or maximum voltage drop allowed. It is necessary to calculate power flows to optimize a network design, and the task is more laborious if a wide range of network components is available. An erroneous estimation of the incremental network cost may be translated into merging clusters that are better electrified separately, or not merging clusters that would be better electrified with the same system.

The bottom-up, irreversible logic of the current off-grid clustering in REM also poses challenges when incorporating some constraints that impact the final electrification solution (and therefore should be modeled in the clustering). This logic requires that generation and management costs behave monotonically: the generation and management cost of two separated clusters cannot be lower than the generation and management cost of the two clusters joined. However, constraints that establish (for example) that solar kits must be used to electrify isolated residential consumers are not easy to incorporate in such logic because they distort the monotonicity of generation costs.

Figure 6-2 shows two cost curves that we will use to illustrate this concept. The first curve corresponds to a case where AC standalone systems are used to electrify isolated residential consumers, and the second curve corresponds to a case where isolated residential consumers are electrified with DC solar kits. Chapter 4 provides a thorough explanation of how these cost curves are calculated, but the key idea is that REM optimizes the generation designs of a representative set of off-grid systems, and then interpolates or extrapolates the generation costs of the remaining ones if it needs them in the off-grid clustering.

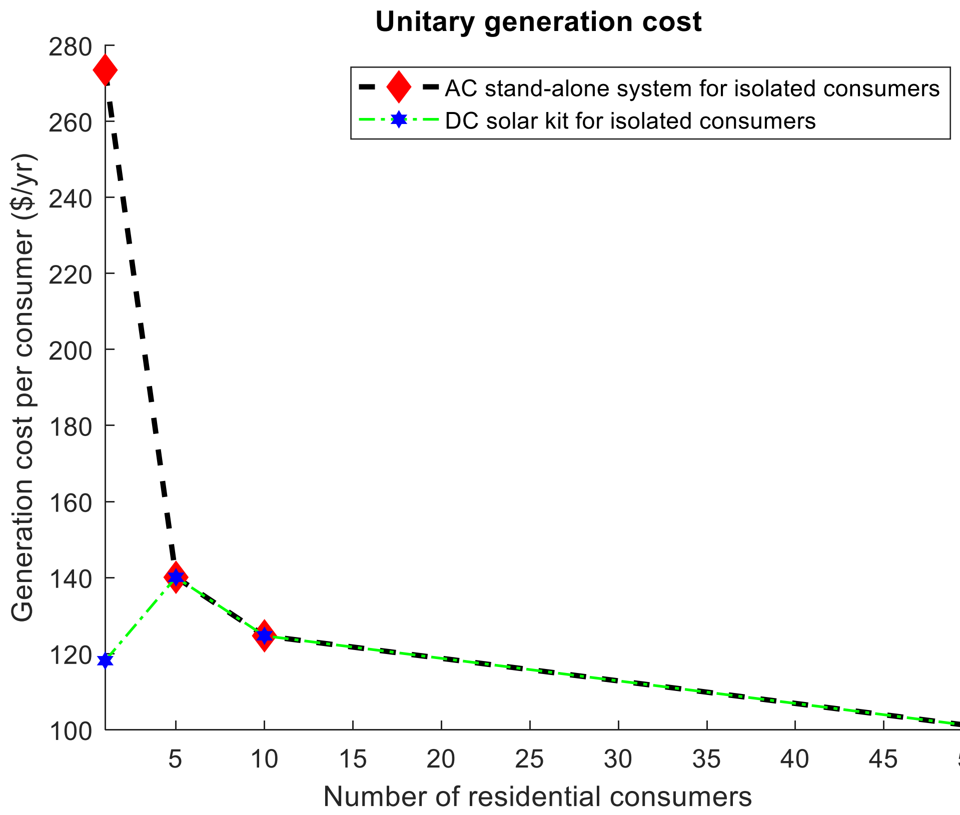


Figure 6-2: Generation cost curves where AC standalone systems and DC solar kits are used to electrify isolated consumers.

Solar kits are low-cost electrification solutions that are not compatible with the monotonicity of the generation cost curve (i.e., the unitary or per-consumer generation cost is higher for an AC mini-grid with five consumers than for an isolated consumer with a single solar kit). If the generation cost curve that corresponds to solar kits in Figure 6-2 is introduced into the standard clustering in REM, then the final solution would possibly be a solar kit for all the consumers. However, it could be better to have large mini-grids that benefit from economies of scale in generation (that usually go beyond 50 residential consumers). Large clusters would not be created because it is not worth to join two consumers into a single cluster since the generation and network costs of the joined cluster would be higher than the sum of the generation and network cost for the separated consumers.

To sum up, the standard off-grid clustering of REM has two significant drawbacks. The first one is that it uses an estimator of the network cost that fails to capture the complexity of network designs, causing clusters to merge when they should not and keeping clusters separated that would be better electrified with the same off-grid system. The second one is related to its logic: the clustering follows a bottom-up logic where connections among clusters depend on greedy, irreversible decisions. This logic is sensitive to economies of scale in generation and management, hindering the incorporation of constraints that somehow alter their monotonicity.

6.2.2. **The Exhaustive clustering algorithm**

This section presents a novel clustering algorithm (*exhaustive clustering*) that extends the off-grid current clustering of REM to overcome its two main flaws. Exhaustive clustering uses a sophisticated process to estimate the network cost of a mini-grid, which is based on a piecewise linear model that considers geometric and electric metrics as explanatory variables of the network cost. The network designs of only a few representative mini-grids are calculated accurately, and the network cost of the remaining mini-grids can be quickly estimated. Further details about this method are provided in chapter 5.

Exhaustive clustering performs a wide exploration of the space of candidate solutions for the clustering problem, creating a hierarchical structure of clusters that is later evaluated to determine the off-grid clusters. The creation and evaluation of a hierarchical structure of clusters enables the incorporation of constraints that distort the monotonicity of economies of scale in generation and management, which are adequately handled.

Figure 6-3 shows the flow diagram of the algorithm. The goal of the exploratory clustering is to generate a subset of mini-grids that is representative of the case study, which is later used to adjust the network cost estimator and to create a hierarchical structure of clusters. Finally, the exhaustive clustering performs a cost evaluation of the hierarchical structure of clusters to obtain the final clustering of consumers. The rest of this section provides a full description of the algorithm.

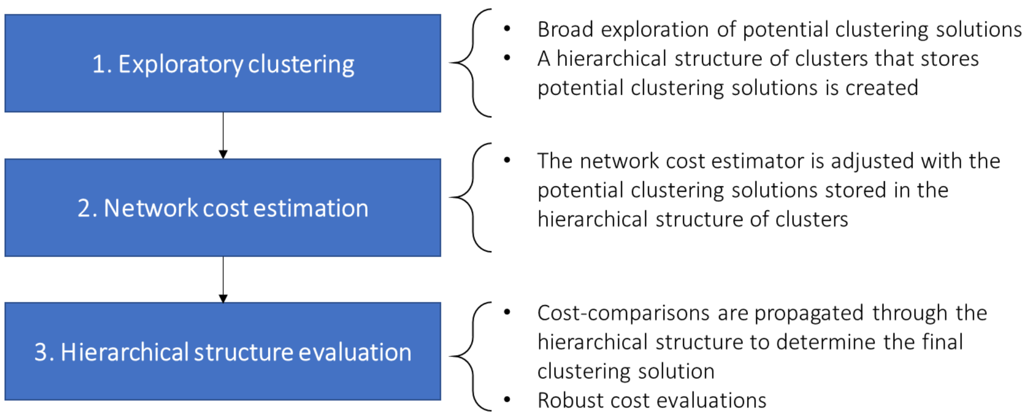


Figure 6-3: Exhaustive clustering flow diagram.

The standard off-grid clustering in REM follows a bottom-up approach where every single consumer is initially an independent cluster, and nearby clusters merge into larger clusters in an iterative procedure. The intermediate clustering solutions generated through the process turn out to be very useful, as they constitute the result of an extensive exploration of the space of all reasonable groupings of consumers (i.e., only consumers that are close to each other can be clustered together) into off-grid systems.

However, most of the space of clustering solutions remain unexplored because the standard off-grid clustering in REM terminates when the savings in generation and management do not seem to compensate for the incremental network cost related to merging two clusters. It could be interesting to explore this part of the space of solutions because the incremental network cost estimator (i.e., the cost of a line that joins the centers of the clusters) may be erroneous, and the off-grid clustering could have reached a local minimum. Therefore, the standard off-grid clustering in REM has been modified to ensure a complete exploration of the space of solutions, forcing clusters to merge as long as they do not grow large enough so that an LV mini-grid would not be a viable electrification solution. Figure 6-4 shows an example with several intermediate clustering solutions that the exhaustive clustering could consider.

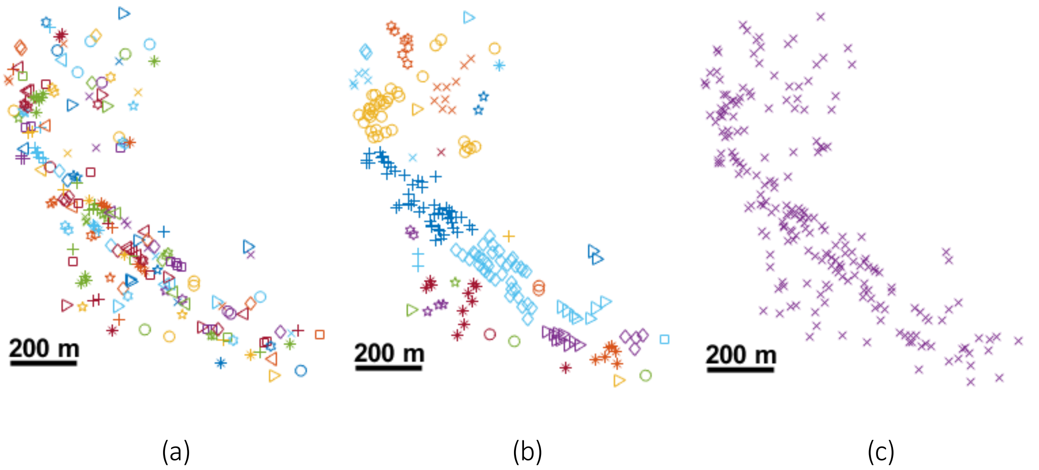


Figure 6-4: Intermediate solutions where (a) most consumers are isolated and clusters are small, (b) there are larger clusters and only a few isolated consumers, and (c) all the consumers belong to the same cluster. Nearby consumers that belong to the same cluster are represented with the same symbol and color.

Each intermediate solution that is stored constitutes one layer of the hierarchical structure of clusters, which is used first to adjust the network cost estimator and later evaluated to obtain the clustering solution. For the time being, the exhaustive clustering stores the intermediate solutions that are multiples of a certain number.

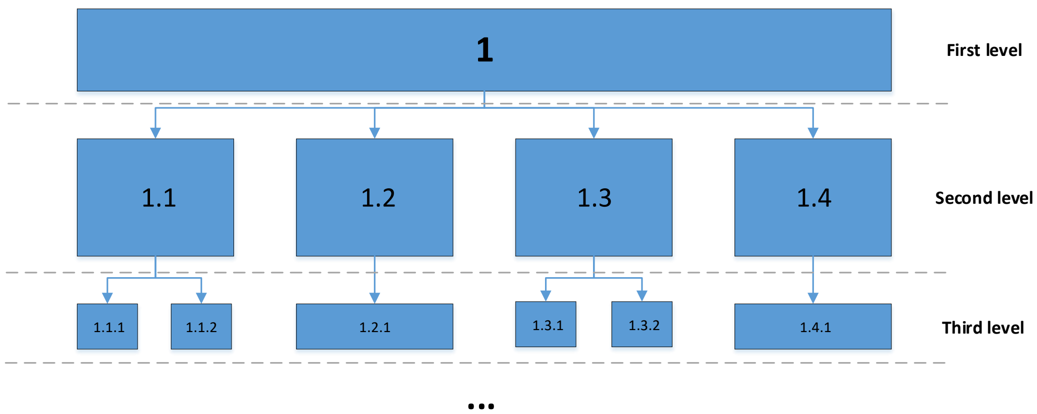


Figure 6-5: Hierarchical structure of clusters example.

Figure 6-5 shows an example of the hierarchical structure of clusters with three levels. The electrification cost of cluster 1.1 (which is the least-cost solution between a mini-grid or a combination of isolated systems for all its consumers) would be compared with the sum of electrification costs of clusters 1.1.1 and 1.1.2. Similarly, the electrification cost of the remaining clusters that belong to the second level of the hierarchical structure would be compared with the sum of the electrification cost of the corresponding clusters from the third level. These comparisons yield an off-grid electrification solution that could combine clusters from the second and the third level. Finally, the cost of this electrification solution would be compared with the electrification cost of cluster 1, and the least-cost alternative would be the electrification solution. This evaluation of the hierarchical structure of clusters is very similar to the one that REM performs to determine the best electrification mode of each consumer (see section 3.1.4).

Exhaustive clustering performs a robust cost evaluation of the hierarchical structure of clusters, overcoming the two main weaknesses of the standard clustering of REM. The algorithm applies the network cost estimator presented in chapter 5 to evaluate the network costs of mini-grids. Elements or constraints that distort the monotonicity in economies of scale in generation, such as solar kits, are also considered in the cost evaluation. By doing so, the algorithm considers accurate representations of all the costs involved.

6.2.3. **Case study**

We consider the case of Cajamarca to compare the off-grid clustering and the proposed exhaustive clustering. The input parameters of the case study are similar to the ones used in section 3.3, and the catalog of lines and transformers is the same used for the case study presented in section 5.4. The only electrification solutions considered are mini-grids and isolated systems. The exhaustive clustering stores the intermediate solutions that are multiples of one hundred for the case study.

The generation designs are calculated with the continuous implementation introduced in section 4.3, and the generation results of the case are the ones presented in section 4.4 that correspond to the continuous implementation. Therefore, we will only deal with the clustering and the final results in this section.

In this case study, the generation costs considered in the clustering and the final solution are calculated with the continuous algorithm. This implies that the clustering algorithm does not commit errors regarding the estimation of generation costs because the generation cost of each cluster exactly matches its final generation cost.

Therefore, we can evaluate the accuracy of the incremental network cost estimator that the standard clustering applies. We can also assess the capabilities of the standard clustering handling elements (such as solar kits) that are not compatible with a monotonic generation cost curve.

Figure 6-6 shows the electrification solution with the standard clustering and the exhaustive clustering. Both solutions seem very similar, but mini-grids tend to be bigger with the standard clustering. As an example, we have surrounded a part of the solution with a red shape. REM electrifies the surrounded consumers with two mini-grids when it applies the standard clustering. Still, the model uses twelve mini-grids and several standalone systems to electrify the same consumers when it applies the exhaustive clustering.

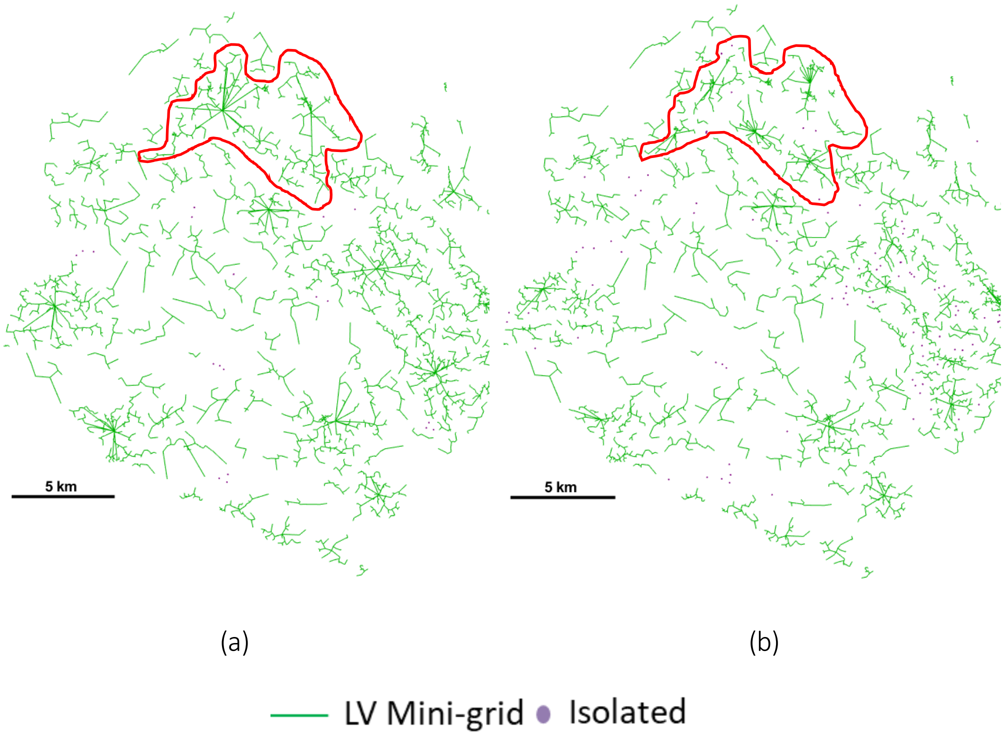


Figure 6-6: Electrification solutions where candidate mini-grids are calculated with (a) the off-grid clustering and (b) the exhaustive clustering.

Figure 6-7 shows the cumulative number of consumers per cluster for clusters with less than 250 consumers (we limit the number of consumers for the sake of clarity). The distributions are very similar. Approximately 8.25% of the clusters have 250 consumers or more with the standard clustering, whereas that figure drops to 1.49% with the exhaustive clustering (those clusters are not shown in Figure 6-7). These numbers suggest that the standard clustering is generating a few oversized clusters, which exhaustive clustering replaces by smaller clusters.

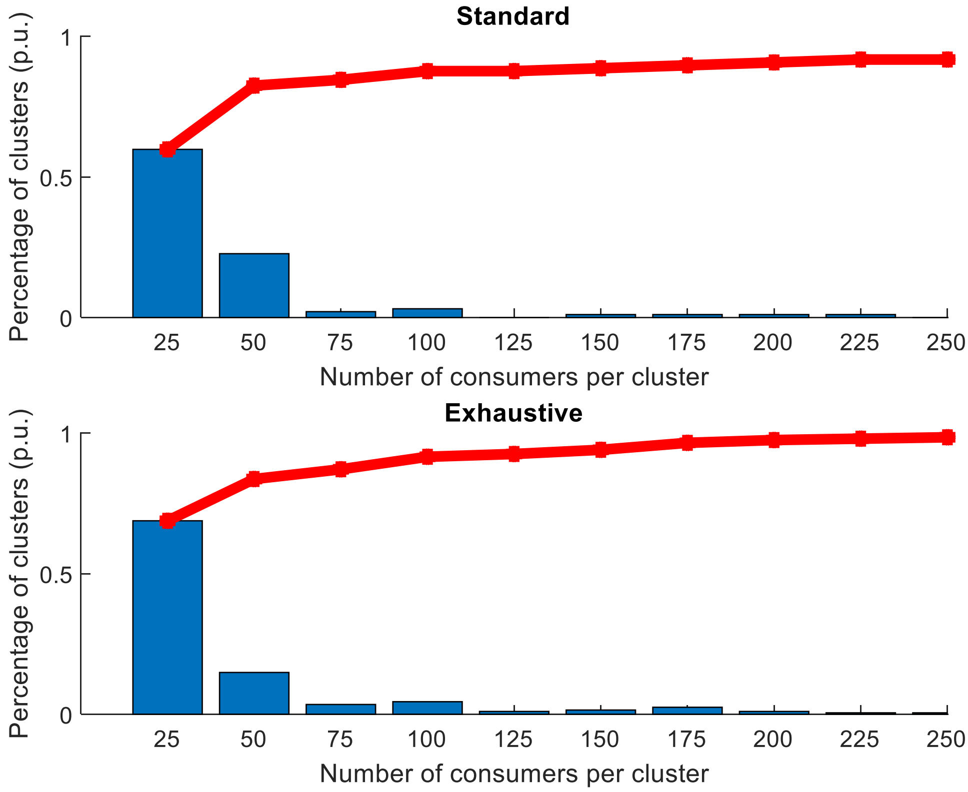


Figure 6-7: Number of consumers per cluster for clusters of less than 250 consumers.

Table 6-1 shows the electrification costs obtained using the standard clustering and exhaustive clustering. The final cost is 6% more expensive with the standard clustering than with the exhaustive clustering, which is a significant improvement.

	Consumers grouped with the standard clustering			Consumers grouped with the exhaustive clustering			Δ All (%)
	Mini-grids	Isolated	All	Mini-grids	Isolated	All	
Number of consumers	6,666	22	6,688	6,584	104	6,688	0.00
Fraction of consumers	1	0	1	0.98	0.02	1	0.00
CAPEX per consumer ($/yr)	77.72	108.24	77.82	62.09	108.24	62.81	-19.29
OPEX per consumer ($/yr)	50.61	167.46	51	56.39	167.46	58.12	13.96
CNSE per consumer ($/yr)	0.31	0.85	0.32	0.46	0.85	0.47	46.88
Final cost per consumer ($/yr)	128.65	276.54	129.13	118.95	276.54	121.4	-5.99
Total CAPEX ($/yr)	518,097	2381	520,478	408,818	11,257	420,075	-19.29
Total OPEX ($/yr)	337,380	3684	341,064	371,288	17,416	388,704	13.97
Total CNSE ($/yr)	2,093	19	2111	3047	88	3,135	48.51
Final cost ($/yr)	857,570	6,084	863,654	783,154	28,761	811,914	-5.99
Fraction of demand served (p.u.)	0.999	0.998	0.999	0.999	0.998	0.999	0.00
Cost per kWh of demand served ($/kWh)	0.378	0.804	0.379	0.349	0.804	0.356	-6.07

Table 6-1: Electrification solution summary for the two clustering algorithms. The last column contains the percentual increment between the "All" columns of the table.

The network cost estimator introduced in chapter 5 allows us to rerun the exhaustive clustering while keeping track of the estimated electrification cost of all the intermediate solutions that the clustering evaluates (regardless of whether they are stored in the hierarchical structure of clusters that the exhaustive clustering generates). This is quite interesting, as now we can estimate how the electrification cost varies when clusters merge.

Figure 6-8 shows how the electrification estimated cost changes in the latter connections of the exhaustive clustering. The blue line represents the connections that the standard clustering performs, and the red line is related to additional connections that the exhaustive clustering activates to ensure that all the space of potential clustering solutions is searched. When two large off-grid clusters are joined, the network cost of the resulting cluster may be substantially high due to the maximum voltage drop constraint. This effect causes a steep increment of the electrification estimated cost in the red curve shown in Figure 6-8.

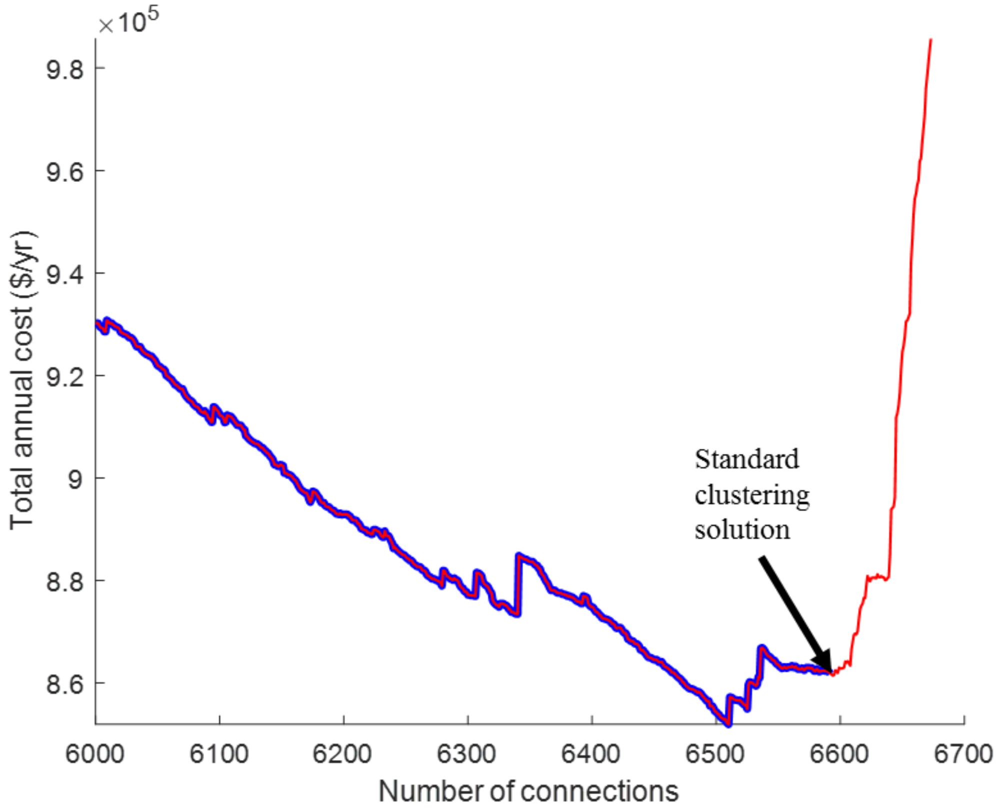


Figure 6-8: Variation of the estimated electrification cost in the exhaustive clustering (base case).

The electrification estimated cost at the end of the standard clustering is 861,798 $/yr. It is very close to the final cost provided in Table 6-1 for the standard clustering (863,654 $/yr), which is calculated by designing detailed network designs for all mini-grids with RNM. This implies that the network cost estimator is providing very accurate estimations of the designs that RNM calculates later.

The standard clustering evaluates several configurations with a lower estimated cost than 861,798 $/yr (for example, the estimated cost is 851,929 $/yr after 6,510 iterations). The electrification cost obtained with the exhaustive clustering (811,914 $/yr, see Table 6-1) is lower than the cost of any of the intermediate solutions evaluated in the exhaustive clustering, which implies that it is a combination of several levels of the hierarchical structure.

We perform a sensitivity analysis modifying two key parameters: the use of solar kits to electrify isolated consumers and the minimum number of consumers that an off-grid cluster needs to have so that a mini-grid is considered viable. This last parameter is set to 5, 10, 15, 25 and 50 consumers for the sensitivity, and the input parameters of the solar kit are provided in Table 6-2.

Investment and operation cost ($/yr)	172.8
Management cost ($/yr)	0
Per-consumer cost ($/yr)	0
Solar capacity (kWp)	0.18
Battery capacity (kWh)	0.978
Lifetime (yr)	5
CNSE ($/kWh)	0.98

Table 6-2: Parameters of the solar kit.

Table 6-3 shows the final electrification cost (which includes the investment and operation cost plus the penalty for non-served energy) obtained with the standard clustering. As expected, the final cost rises when the minimum number of consumers that a cluster needs to reach to be a mini-grid is increased. The reduction of the cost that happens when solar kits are introduced is higher when the minimum number of consumers to consider as mini-grid is higher.

		Minimum number of consumers to consider mini-grid				
		5	10	15	25	50
Solar kits	No	863,654	871,105	898,450	936,605	1,039,905
	Yes	860,067	860,146	863,800	869,586	893,834

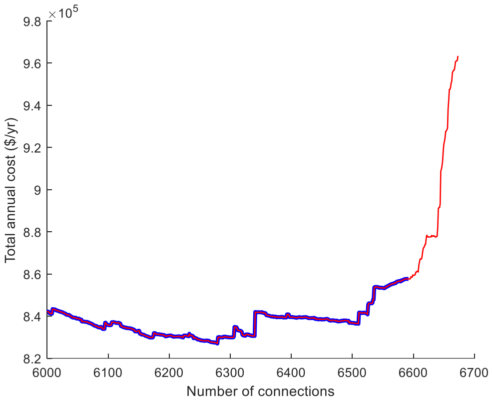
Table 6-3: Final costs ($/yr) obtained with the off-grid clustering.

Table 6-4 shows the final electrification cost obtained with the exhaustive clustering, which outperforms the standard clustering in all the cases. Specifically, the electrification cost is approximately between 5% and 8% better with the exhaustive clustering.

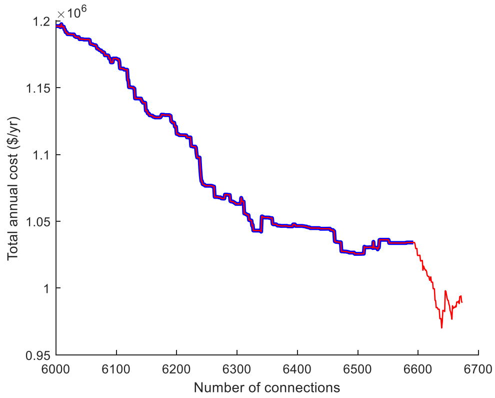
		Minimum number of consumers to consider mini-grid				
		5	10	15	25	50
Solar kits	No	811,914	822,643	855,597	892,245	961,782
	Yes	799,830	801,641	806,993	816,453	853,864

Table 6-4: Final costs ($/yr) obtained with the exhaustive clustering.

Figure 6-9 shows how the estimated cost changes when the exhaustive clustering connects consumers in the case (a) solar kits, the minimum number of consumers to consider mini-grid is 10, and (b) no solar kits, the minimum number of consumers to consider mini-grid is set to 50. As in Figure 6-8, the blue line represents the connections that the standard clustering performs, and the red line is related to additional connections that ensure that all the space of potential clustering solutions is searched.



(a)



(b)

Figure 6-9: Variation of the estimated electrification cost in the exhaustive clustering. (sensitivity analysis).

In case (a), the standard clustering activates too many connections, and the clustering solution provided is worse than some intermediate solutions explored (as happens in the base case, although this time the effect is more exaggerated). In case (b), the standard clustering fails to activate connections that would produce a better electrification solution.

6.3. Top-down clustering

This section first recapitulates the grid-extension clustering, presenting its current limitations. Then, it introduces a new algorithm, *top-down clustering*, that that aims at overcoming the main weaknesses of the grid-extension clustering.

The top-down clustering starts calculating an extension of the power grid that connects all the consumers. Then, this method sequentially evaluates if it is worth to disconnect or prune an element of the network (i.e., a line or a transformer) and electrify the downstream consumers with off-grid systems. There is a planning tool in the literature that applies a similar strategy: LAPER examines the MV lines of an initial network provided by the user, and it performs cost-comparisons to determine whether to disconnect villages from the grid and electrify them off-grid (Fronius and Gratton, 2001). However, LAPER operates with villages instead of consumers and it deals with the planning problem with a lower level of modeling detail, which somehow limits the scope of its approach.

Figure 6-10 shows how the top-down clustering fits in the current version of REM, which is the final product of this thesis. The current REM version starts from the initial prototype of REM described in section 3.1. It includes the enhancements described in section 3.2, the generation sizing methods presented in chapter 4, and the two clustering algorithms described in this chapter.

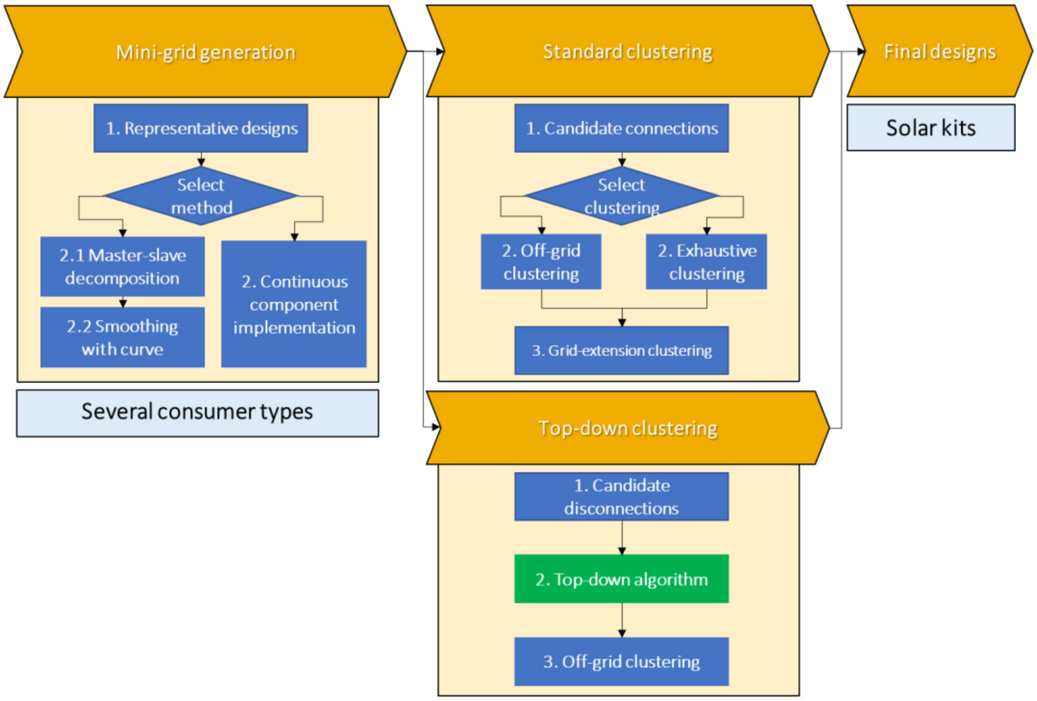


Figure 6-10: Incorporation of the top-down clustering into REM's algorithmic structure.

The top-down clustering, which is an alternative to the standard clustering, operates by applying three sequential steps. In the first step, it calculates an initial extension of the power grid that connects all the consumers of the case using RNM. The elements of that initial grid extension represent the potential disconnections, which are evaluated in the second step of the top-down clustering (*top-down algorithm*).

The top-down algorithm step performs cost-comparisons to determine if it is worth to disconnect an element from the initial network and electrify its downstream consumers with off-grid alternatives. At the end of this step, the top-down clustering provides a list of the consumers that should be electrified extending the power grid, which are the ones that are still connected to the initial power grid. The remaining consumers should be electrified with off-grid solutions, and they are grouped into off-grid systems with a straightforward application of the off-grid clustering in the third step.

6.3.1. **The limitations of REM's grid-extension clustering**

The grid-extension clustering in REM, which is described in section 3.1.3.2, also has several drawbacks. Firstly, it starts operating from the off-grid clusters, which were calculated without considering the characteristics power grid. The off-grid clusters are not necessarily the best starting point for the grid-extension clustering process.

The procedure that some planners follow to determine where the power grid should be expanded and which regions should be electrified with off-grid alternatives follows a straightforward logic that overcomes this first drawback. First, the procedure determines where the power grid should be expanded by applying criteria based on distances among consumers and the power grid (i.e., each consumer closer than 2 km to the power grid is electrified with an extension of the power grid). The consumers that do not satisfy the criteria are electrified with off-grid solutions.

There is an idea behind this straightforward procedure that the top-down clustering follows: it determines the grid-extension clusters first, and the off-grid clusters are calculated later. By doing so, the characteristics of the power grid are considered in the calculation of the off-grid clusters.

The grid-extension clustering also uses approximations to estimate the network costs similar to the ones that the off-grid clustering applies. Specifically, the grid-extension clustering estimates the incremental network costs with lines that connect the centers of two clusters or the center of a cluster with the power grid.

The top-down clustering starts calculating a detailed extension of the power grid that connects all the consumers, so the exact cost of each line and transformer of the initial network is known. This implies that the top-down clustering has a good estimation of the cost of a line or a transformer when the algorithm is evaluating if it is worth to disconnect an element of the network and electrify the downstream consumers with off-grid systems.

Finally, the grid-extension clustering cannot cope successfully with topographical features of the terrain, such as altitudes and forbidden zones. This is the most critical limitation of the grid-extension clustering that the top-down clustering overcomes.

It would be necessary to evaluate several paths to determine the least-cost route each time the grid-extension clustering calculates the cost of a line connecting two clusters or a cluster with the power grid. However, these evaluations are computationally intensive and cannot be performed each time the grid-extension clustering evaluates if it is worth connecting two nearby clusters.

Figure 6-11 shows an example of a route that connects the center of a cluster with the power grid when topography is considered. One method to determine the least-cost route involves calculating a mesh and evaluating the cost of several lines that go inside the mesh. The number of minimum-length paths that connect the center of the cluster and a point of the network has a combinatorial nature, and it increases significantly with the number of cells of the mesh (Better Explained, 2020).

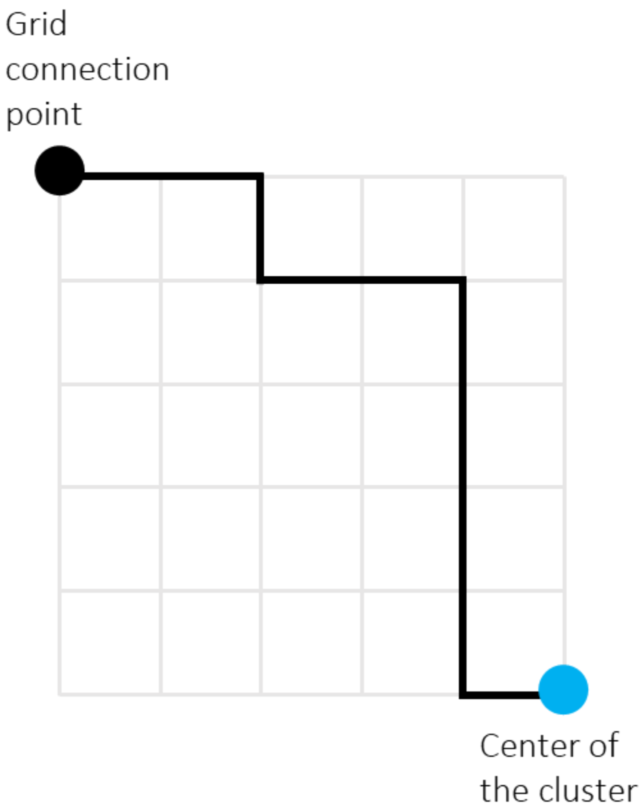


Figure 6-11: Example of a minimum-distance route that connects the center of a cluster with the grid following the lines of a mesh.

REM uses RNM to calculate the initial extension of the power grid that connects all the consumers, and RNM can handle topographical elements when it optimizes a distribution network. The algorithms of RNM adapt the layout of the networks so that lines do not cross forbidden zones, and the network routing is calculated considering the additional costs due to terrain elevation and slopes.

The initial extension of the power grid that the top-down clustering obtains can be calculated incorporating topographical considerations, ensuring that no element of the network crosses a forbidden zone and that the extra costs due to altitudes are included in the clustering process.

6.3.2. **The top-down clustering**

The top-down method starts by designing a distribution network that connects all the consumers to the existing grid. RNM calculates the initial network, which has a radial topology so it can be represented with a tree structure. Figure 6-12 shows a distribution network and the corresponding tree representation.

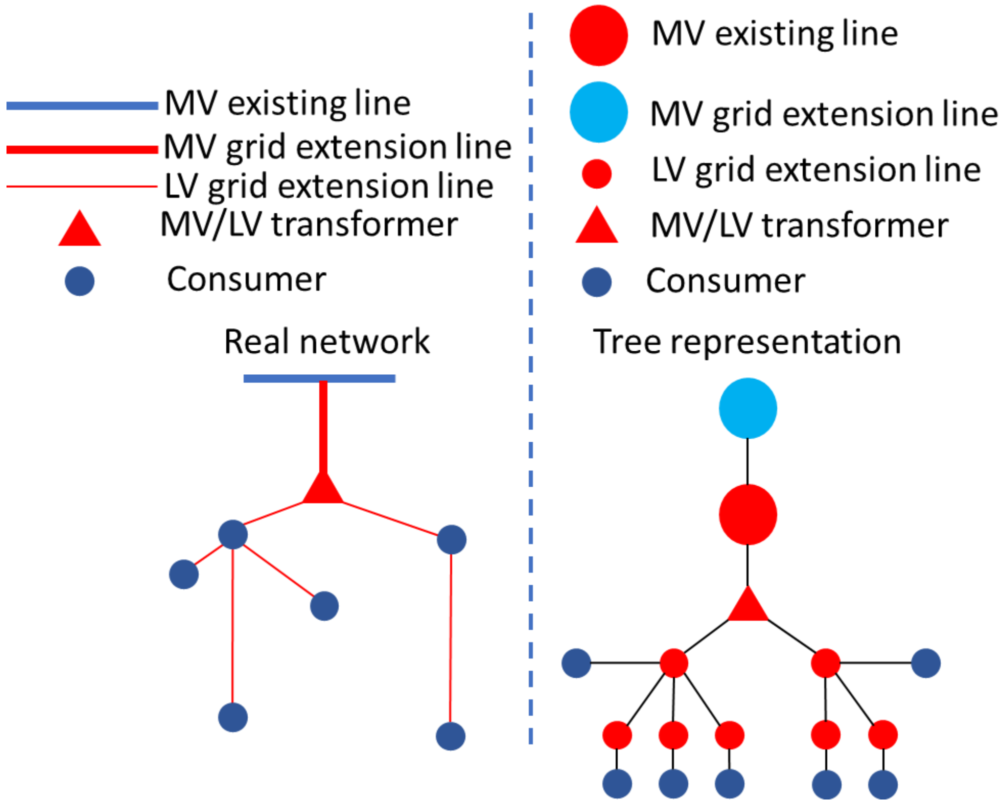


Figure 6-12: Distribution network (left), and tree representation (right).

We classify the nodes of the initial distribution network into three types: (a) Substations or transformers (HV/MV substations and MV/LV transformers, the MV existing power grid is modeled as a set of HV/MV substations), (b) Lines (the MV and LV distribution lines), and (c) consumers.

Once the algorithm has generated the tree representation of the network, the top-down method loops through each element of the tree to determine if it is best to "prune" the element and electrify all the consumers downstream with off-grid solutions. We refer to the node that is under consideration for pruning as the *evaluated element*, even if the best local decision is not to prune the tree.

The algorithm performs a cost comparison between the two configurations shown in Figure 6-13 to determine if it is best to prune the evaluated element of the network and electrify the downstream consumers with off-grid solutions. All the consumers that are below the evaluated element considered are electrified as part of a grid extension in configuration A, whereas they are electrified with a mini-grid in configuration B. The costs involved downstream from the evaluated element in configuration A are grid energy cost, network cost, management cost, and CNSE The costs considered downstream the evaluated element in configuration B are generation cost, network cost, management cost, and CNSE.

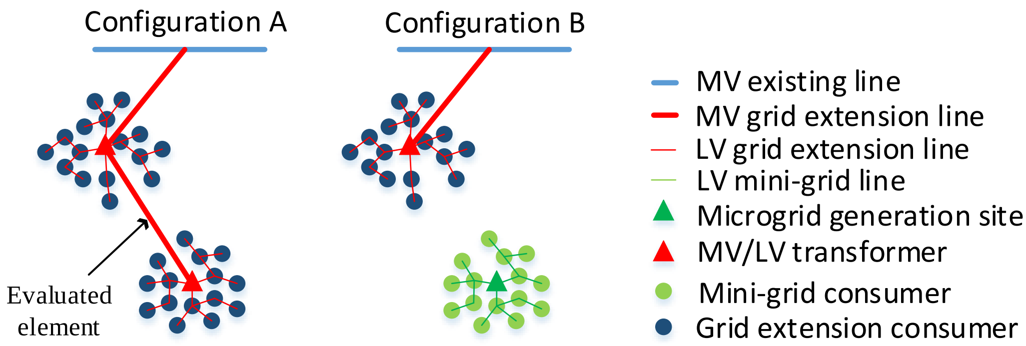


Figure 6-13: Top-down clustering configurations.

It is straightforward to calculate all the costs involved but the network costs, and the top-down algorithm assumes that the network cost of the mini-grid in configuration B is approximately equal to the cost of the network downstream the evaluated element in configuration A. The assumption of an equal network cost might not hold for every case, but it avoids the need for estimating the network cost of a mini-grid each time that an element is evaluated. Further development of the top-down algorithm could be to use the network cost estimator presented in chapter 5 to estimate the network cost of the mini-grid in configuration A.

If there are several transformers downstream from the evaluated element in configuration A, then the mini-grid of configuration B is MV, and the cost of an additional transformer located at the generation site is included.

The capacities of some elements of the network could decrease if the evaluated element is pruned because the consumers downstream would be electrified with off-grid systems. The incremental cost related to the capacity reduction is measured using continuous catalogs of lines and transformers (continuous catalogs were introduced in section 3.2.2.3.2 to estimate the incremental network cost in the clustering).

A question that arises naturally which order should be considered to evaluate the nodes of the tree. In the bottom-up clustering, the arcs of the Delaunay triangulation are evaluated from the shortest to the longest, so applying a distance-based criterion seems reasonable. The peak demand of consumers also seems relevant, and the initial network provides additional information, such as the tree structure.

The top-down algorithm determines the order of evaluation of the nodes according to a criterion based on voltage drops, which depend on geometric and electric parameters. The voltage drop at node i is calculated with equation 6-1 in a three-phase system:

$$\Delta V_i = \sqrt{3} \cdot z_i \cdot L_i \cdot I_i \qquad 6\text{-}1$$

Where z_i, L_i and I_i are the impedance per unit of length (Ω/km), the length (km) and the current (A) of node i, respectively.

We define the accumulated voltage drop of node i as the maximum voltage drop between that node and any terminal node that is downstream node i (we say that a node is terminal if it does not have any node downstream). Figure 6-14 shows an example where the voltage drops and accumulated voltage drops of lines are shown. Consumers and transformers have zero length, so their voltage drop is zero and they do not increase the accumulated voltage drops.

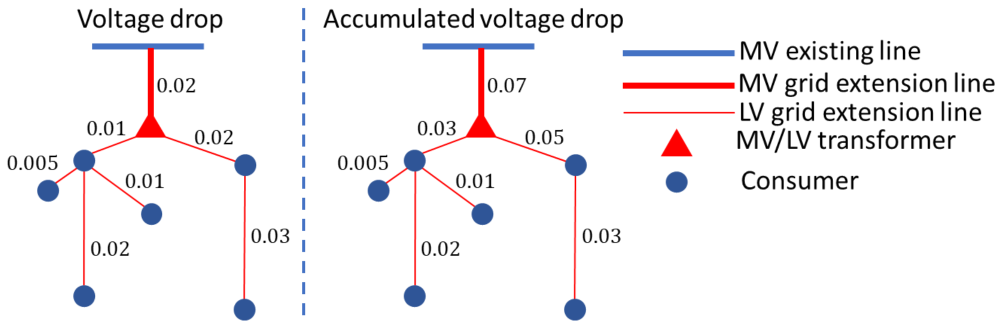


Figure 6-14: Example of the accumulated voltage drop.

The accumulated voltage drop at node i can be calculated recursively using equation 6-2:

$$\Delta AV_i = \Delta V_i + max\ \{\Delta AV_{ch1}, \Delta AV_{ch2}, \dots, \Delta AV_{chn}\} \qquad 6\text{-}2$$

Where $ch1, ch2, \dots, chn$ correspond to the n children nodes of node i (i.e., the nodes that are downstream i, and directly connected to i).

The electric moment of a node i is calculated with equation 6-3:

$$M_i = L_i \cdot S_i \qquad 6\text{-}3$$

Where L_i and S_i are the length (km) and the power (VA) of node i, respectively. The power of a node is $\sqrt{3}$ times the product of its voltage (V) by its current (A):

$$S_i = \sqrt{3} \cdot U_i \cdot I_i \qquad 6\text{-}4$$

If we replace I_i into equation 6-1 we obtain:

$$\Delta V_i = \sqrt{3} \cdot z_i \cdot L_i \cdot \frac{S_i}{\sqrt{3} \cdot U_i} = \frac{z_i}{U_i} \cdot L_i \cdot S_i = \frac{z_i}{U_i} \cdot M_i \qquad 6\text{-}5$$

Equation 6-5 implies that the voltage drop in a node is directly proportional to its moment. We can define the accumulated moment of a node i as we did with the accumulated voltage drop in equation 6-2.

$$AM_i = M_i + max\ \{AM_{ch1}, AM_{ch2}, \dots, AM_{chn}\} \qquad 6\text{-}6$$

As voltage drops are directly proportional to moments, equation 6-6 can be used to sort the nodes according to their accumulated voltage drop.

The top-down algorithm sorts the nodes according to their accumulated moments (from the highest to the lowest) using equation 6-6. Then, it evaluates the first node that does not have unevaluated nodes downstream. This process continues until all the nodes have been evaluated. When an element is pruned, the accumulated moments of its upstream nodes are recalculated to account for the changes in the network.

6.3.3. **Case study**

We also consider the Cajamarca case to compare the standard clustering of REM and the top-down clustering. The case study is the same as in section 6.2.3 with the exception of grid extension being a viable electrification solution. The energy cost of the power grid is set to 0.045 $/kWh, and the reliability of the grid is 100% (Gonzalez-Garcia et al., 2016).

In order to establish a fair comparison between both clustering algorithms, REM applies the same off-grid clustering (which determines the grouping of consumers into candidate off-grid systems) in the standard clustering and the top-down clustering, and it corresponds to the method presented in section 3.1.3. The integration of the top-down clustering and exhaustive clustering is a future line of research that is yet to be developed as they perform complementary tasks, although both clustering methods are valuable on its own.

Figure 6-15 shows the electrification solution with the standard clustering and the top-down clustering (the connections to the power grid are shown with black crosses). Both solutions electrify most consumers with grid extensions, but they seem to be more abundant with the top-down clustering.

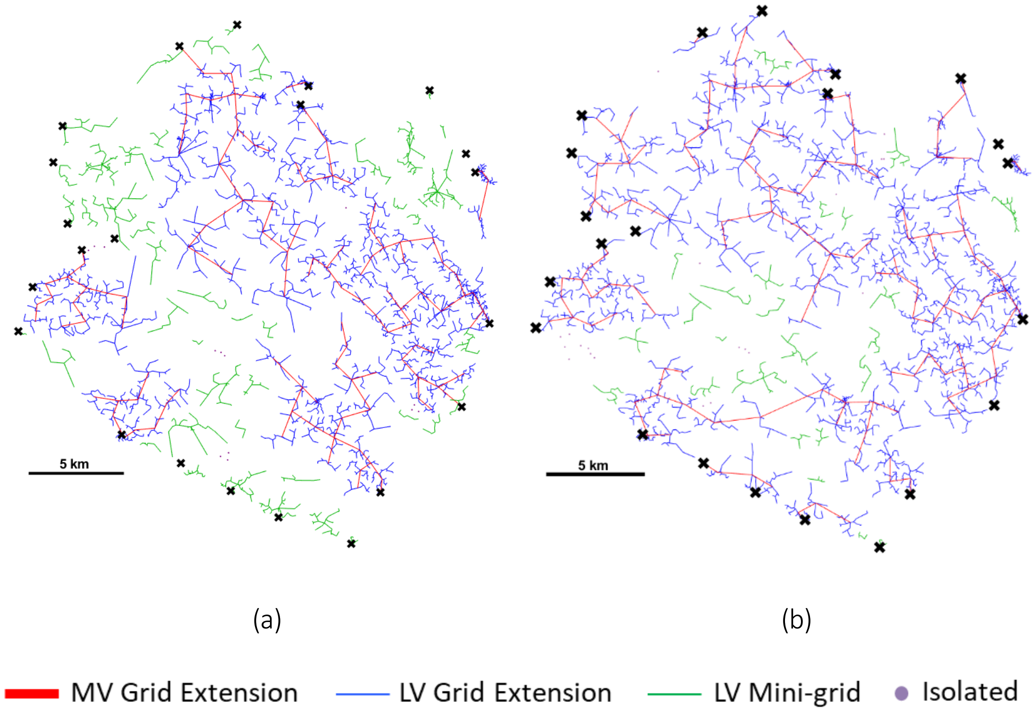


Figure 6-15: Electrification solutions where clusters are calculated with (a) the standard clustering and (b) the top-down clustering.

Table 6-5 shows the electrification costs obtained using the standard clustering and the top-down clustering. The electrification cost obtained with the top-down clustering is approximately equal to the electrification cost obtained with REM's standard clustering. It can be concluded that several quasi-optimal solutions are very similar in terms of cost. Still, they differ significantly in the combination of electrification alternatives (for example, the percentage of consumers electrified with grid extension is 84% with the standard clustering, but it rises to 93% with the top-down clustering).

	Consumers grouped with the standard clustering				Consumers grouped with the top-down clustering				Δ All (%)
	Mini-grids	Isolated	Grid extension	All	Mini-grids	Isolated	Grid extension	All	
Number of consumers	1,052	14	5,622	6,688	418	20	6,250	6,688	0
Fraction of consumers	0.16	0	0.84	1	0.06	0	0.93	1	0
CAPEX per consumer ($/yr)	83.53	108.24	49.16	54.69	92.45	108.24	51.93	54.63	-0.11
OPEX per consumer ($/yr)	59.6	167.46	43.9	46.63	55.57	167.46	45.39	46.39	-0.51
CNSE per consumer ($/yr)	0.75	0.85	0	0.12	1.02	0.85	0	0.07	-41.67
Final cost per consumer ($/yr)	143.89	276.54	93.07	101.44	149.04	276.54	97.32	101.09	-0.35
Total CAPEX ($/yr)	87,872	1,515	276,403	365,791	38,646	2,165	324,556	365,367	-0.12
Total OPEX ($/yr)	62,704	2,344	246,809	311,858	23,227	3,349	283,680	310,256	-0.51
Total CNSE ($/yr)	792	12	0	804	428	17	0	445	-44.65
Final cost ($/yr)	151,369	3,872	523,213	678,453	62,300	5,531	608,236	676,068	-0.35
Fraction of demand served (p.u.)	0.998	0.998	1	1	0.997	0.998	1	1	0
Cost per kWh of demand served ($/kWh)	0.422	0.804	0.274	0.298	0.437	0.804	0.286	0.297	-0.34

Table 6-5: Electrification solution summary for the two clustering algorithms. The last column contains the percentual increment between the "All" columns of the table.

We also perform a sensitivity analysis to compare the top-down clustering and the standard clustering modifying the reliability of the network and the fuel cost. Specifically, the reliability of the network is set to 100%, 95%, and 90%. The cost of fuel is set to 0.5 $/l, 0.6 $/l, and 0.7 $/l. Table 6-6 shows the final electrification cost obtained with the standard clustering. The final cost tends to rise when the grid reliability drops and the fuel cost is more expensive, although there are small fluctuations when the fuel cost rises from 0.6 $/l to 0.7 $/l.

		Grid reliability (%)		
		90	95	100
Fuel cost ($/l)	0.5	865,320	812,292	678,453
	0.6	883,042	813,935	696,293
	0.7	881,069	811,576	698,429

Table 6-6: Final costs ($/yr) obtained with the standard clustering.

Table 6-7 shows the final electrification cost obtained with the top-down clustering, which outperforms the standard clustering when the grid reliability is 100%. However, the standard clustering provides better results with lower reliability levels.

		Grid reliability (%)		
		90	95	100
Fuel cost ($/l)	0.5	944,163	825,813	676,068
	0.6	978,459	837,029	675,922
	0.7	983,610	837,029	675,922

Table 6-7: Final costs ($/yr) obtained with the top-down clustering.

The cases with grid reliability of 100% and diesel costs of 0.6 $/l and 0.7 $/l have the same electrification solution with the top-down clustering, which happens because the off-grid systems of the solution do not include a diesel generator. The cases with 95% grid reliability and diesel costs of 0.6 $/l and 0.7 $/l also have the same electrification solution for the same reason.

It is interesting to look at the percentages of consumers that are electrified with grid extensions and mini-grids (isolated systems are marginal in the cases considered for the sensitivity analysis). Table 6-8 shows the percentage of consumers electrified with grid extensions obtained with the standard clustering. The percentages stay still with grid reliabilities of 95% and 100%, but they fluctuate when the grid reliability is 90%.

		Grid reliability (%)		
		90	95	100
Fuel cost ($/l)	0.5	0.12	0.63	0.84
	0.6	0.25	0.63	0.84
	0.7	0.18	0.64	0.84

Table 6-8: Fraction of consumers (p.u.) electrified with grid extensions obtained with the standard clustering.

Table 6-9 shows the percentage of consumers electrified with grid extensions obtained with the top-down clustering. The top-down clustering electrifies a higher percentage of consumers with grid extensions in all the cases, and the percentages are quite different from the ones that the standard clustering provides.

		Grid reliability (%)		
		90	95	100
Fuel cost ($/l)	0.5	0.40	0.88	0.93
	0.6	0.80	0.90	0.95
	0.7	0.81	0.90	0.95

Table 6-9: Fraction of consumers (p.u.) electrified with grid extensions obtained with the top-down clustering.

Although top-down clustering did not outperform the standard clustering, it constitutes a valuable tool. It requires more research and development to improve its performance, but it presents structural advantages that are missing in the grid-extension clustering of REM such as considering the characteristics of the power grid when determining which consumers are electrified with off-grid systems and incorporating the topographical features of the terrain (further details were provided in section 6.3.1).

6.4. Conclusions

The performance of REM's off-grid clustering was substantially enhanced with the implementation of the improvements described in section 3.2.2, but it still presented two main flaws: the use of inadequate estimations of the incremental network costs and the inability to cope with elements or constraints that somehow alter the monotonicity of economies of scale in generation.

In this chapter, we introduce a novel off-grid clustering algorithm (exhaustive clustering) that overcomes the two hurdles abovementioned. The exhaustive clustering applies the network cost estimator introduced in chapter 5, and it performs a broad exploration of the space of potential clustering solutions to overcome the hurdles related to non-monotonous economies of scale in generation. We perform a sensitivity analysis that shows that the exhaustive clustering systematically outperforms the standard clustering of REM.

We also present a new grid-extension clustering (top-down clustering) that starts from a grid extension design that connects all the consumers. Then, the top-down algorithm systematically goes through all the elements of the network performing cost-comparisons to determine if an element should be "disconnected" and the downstream consumers should be electrified with off-grid systems.

The case study shows that the top-down clustering provides better results than the standard clustering of REM in some cases, but it performs worse in others. The top-down clustering tends to electrify more consumers with grid-extension designs than the standard clustering of REM. However, the top-down clustering presents some structural advantages over the standard clustering of REM such as incorporating topographical features.

Regarding additional developments, the exhaustive clustering presented only considers LV mini-grids (as the network cost estimator cannot deal with MV mini-grids for the time being). Therefore, it would be interesting to expand both the network cost estimator and the exhaustive clustering so that they consider MV mini-grids in their calculations.

The method that the exhaustive clustering applies to determine which intermediate solutions should be stored should be improved. The current procedure is straightforward, but it could also consider how the estimated cost of the clustering configurations varies to determine if is worth storing an intermediate solution. For example, a clustering solution that has a cost that is lower than any of the previous solutions could be automatically included in the hierarchical structure of clusters.

Regarding the top-down clustering, a question that requires more research is whether RNM should design additional distribution networks from scratch for the remaining grid-connected consumers after the pruning of several elements. The top-down clustering only uses RNM at the beginning to design the initial network, and continuous catalogs are used to estimate the incremental network saving incurred when an element is pruned. The accuracy of these estimations diminishes as the number of disconnected elements increase, leading to erroneous pruning decisions. However, the optimization of more distribution networks and the calculation of the corresponding trees could increase the computation time substantially.

The top-down clustering has been applied to cases with a relatively low number of consumers (less than 10,000) so far. The optimization of the initial network and the calculation of the corresponding tree are computationally intensive processes that limit the applicability of the top-down clustering. It is necessary to develop a robust implementation that relies on parallel computing techniques and smart simplifications so that the top-down clustering can be applied in large-scale cases.

The top-down clustering also needs more research to understand why it tends to electrify many more consumers with grid extension designs than the standard clustering of REM in cases where the standard clustering of REM provides a better solution.

It would also be interesting to merge the top-down clustering and exhaustive clustering as they complement each other. The top-down could determine first which consumers are electrified with an extension of the power grid and the exhaustive clustering could later group the remaining ones into off-grid systems.

“Success is not final, failure is not fatal: it is the courage to continue that counts.”
Winston Churchill

7

CONCLUSIONS AND FURTHER RESEARCH

This chapter presents the main conclusions and introduces the lines of future research that this Ph.D. thesis has laid the ground for. It also summarizes the contributions of this thesis and the publications derived from this work.

Large-scale electrification planning should balance the level of modeling detail necessary to attain an acceptable accuracy level with a reasonable computational burden. A model that requires to run an unmanageable amount of computational resources cannot be used in practice, whereas a tool that operates with a low level of modeling detail will provide results that have scarce value.

This thesis has focused on three inherent challenges in large-scale electrification planning, where an appropriate balance among modeling accuracy and computation requirements is critical. The first one is generation sizing, and it aims at calculating accurate generation costs for each potential off-grid system in a region. The second one is clustering, and its goal is to group the consumers into standalone systems, mini-grids, and grid extensions. The third one is the network design problem, and it deals with the optimization of the network layout of mini-grids and grid extensions.

7.1. Summary and conclusions

This section presents a summary of this thesis and the main conclusions. Both the summary and the conclusions are classified according to the chapter they belong to. Chapter 2 has presented a review of large-scale electrification planning tools and methodologies. Chapter 3 has presented an overview of REM and several improvements that were implemented into the first prototype of REM, enhancing its robustness and consistency. Chapter 4 has described the methodology that REM applies to design the generation of off-grid systems. Chapter 5 has introduced a method that can estimate the network cost of any potential low-voltage mini-grid in a case without the need to calculate its layout. Chapter 6 has presented two clustering algorithms and it has compared them with the standard clustering of REM.

7.1.1. Large-scale electrification planning tools

Several tools and methods address the problem of finding the best techno-economic plan for a large-scale area in developing countries. Generally, those plans combine traditional grid extensions with off-grid alternatives such as mini-grids and solar kits.

Chapter 2 has introduced a conceptual formulation of the electrification planning problem from the techno-economic side, which was missing in the literature. The formulation is used to classify the main methods and tools according to the degree of modeling complexity that they consider. Specifically, tools and methods are grouped into three categories: pre-feasibility tools, intermediate analysis tools, and detailed generation and network design tools.

Pre-feasibility tools operate aggregating the consumers into villages or cells. They generally calculate the LCOEs for several electrification alternatives to estimate the electrification solution of each village or cell. Most pre-feasibility tools take advantage of GIS-based technologies instantly access input databases, and provide quick solutions at the expense of a low level of modeling detail.

Intermediate analysis tools also group the consumers into villages or cells, and they estimate the network layout of the grid extensions using algorithms that are based on the calculation of an MST. However, they do not include electric considerations such as power flows when designing the networks. Most of these tools size the generation design with rules of thumb or relatively simple methods that do not measure seasonalities or the amount of non-served energy in the dispatch.

Detailed generation and network design tools consider a very high level of modeling detail, although this comes at the expense of high computational times and the need for many input data. REM is the only tool that belongs to this category.

7.1.2. Domain-based improvements

Chapter 3 has provided an overall description of the first prototype of REM (Ellman, 2015), including a description of its algorithms and its financial model. REM calculates the least-cost electrification plan of a region, which generally includes a combination of standalone systems, mini-grids with its own generation system, and extensions of the distribution power grid. REM operates at a very high level of spatial granularity, providing detailed network designs at the consumer level. REM also operates at a high level of temporal resolution, considering hourly demand profiles for the consumers (which include residential and productive loads) and simulating the dispatch of mini-grids and standalone systems. REM incorporates the topographical features of the terrain when optimizing the network designs of mini-grids and extensions of the power grid (Drouin, 2018).

REM applies several heuristic algorithms to determine the final electrification solution, which is calculated by minimizing a combination of the actual cost plus a penalty for the non-served energy. The actual cost includes investment, operation and maintenance, and management costs. The penalty for the non-served energy accounts for the non-served demand that could result from reliability failures of the main grid or insufficient generation or storage capacity of off-grid solutions. Reference (MIT & IIT-Comillas Universal Energy Access Lab, 2019) presents a complete list of projects, theses, and papers related to REM.

Chapter 3 has also described the main improvements that have been implemented in the first prototype of REM, enhancing the robustness and consistency of its solutions. Examples of such improvements include enhancements in the optimization of generation designs of mini-grids and their dispatch, upgrades that drastically improved the clustering of consumers into systems, and a robust calculation of the distribution networks of mini-grids and extensions of the power grid.

Chapter 3 has also introduced several upgrades that expanded the capabilities of the first prototype of REM. Examples of such additional capabilities are the ability of REM to handle multiple types of consumers and the inclusion of solar kits as a possible electrification solution.

All the improvements and upgrades presented in chapter 3, which have been classified according to the part of the model that they apply to, have significantly enhanced REM's performance.

7.1.3. **Mini-grid generation design**

Most regional planning models estimate the generation cost of off-grid systems using rules of thumb or analytical expressions. Although these quick estimates provide valuable information, they are far from a sound recommendation since they ignore essential factors such as the hourly solar irradiance and explicit recognition of the CNSE.

Chapter 4 has described the strategy that REM uses to obtain the generation cost of any off-grid system. The straightforward strategy of optimizing from scratch the generation design of any potential mini-grid fails for computational reasons in regional planning. In order to alleviate the computational burden, REM optimizes from scratch the generation designs of several representative combinations of consumers and stores the results in a look-up table so that generation designs for other combinations of consumers can be interpolated quickly.

This chapter has investigated the impact of modeling the capacities of generation components with discrete variables in regional planning, performing an analysis that was missing in the literature. We have concluded that a direct application of a single-village-based method or tool that is based on discrete components may lead to issues in a large-scale planning case because it may not capture the trend of economies of scale in generation.

This chapter has proposed two novel methods to overcome the issues that discrete components may cause. The first one approximates the generation costs of off-grid systems with a curve that guarantees that economies of scale in generation are adequately captured. The second method, which is the one that REM currently uses, initially handles the capacities as continuous variables to estimate the economies of scale in generation. REM can later adjust the generation designs present in the final solution so that they have real, discrete components

7.1.4. **Estimation of the network cost in mini-grids**

Large-scale planning models tend to estimate the network costs considering only geometric criteria (generally based on the calculation of MSTs). However, a sound network design should be calculated considering power flows and the usual electric constraints (such as the maximum voltage drop allowed).

REM uses a different model (RNM) to optimize the network layout of mini-grids and grid extensions. RNM is a mature model that designs distribution networks considering the topographical features of the terrain (forbidden zones and raster map of altitudes), network catalogs that include several lines with different capacities for each voltage level, and the usual electrical constraints.

However, using RNM to optimize the network layout of each candidate mini-grid and grid extension is time-consuming. It is useful to develop a methodology that keeps a proper balance between accuracy and computation speed, going beyond the rules of thumb that most regional planning models apply, but estimating the network cost accurately without the computational burden related to optimizing the layout from scratch.

Chapter 5 has described a procedure that estimates the network cost of any potential LV mini-grid in a large-scale case without the need for optimizing its network layout from scratch. The method optimizes the network designs of a representative set of mini-grids using RNM. Then, it adjusts the coefficients of a piecewise linear model to estimate the cost of the network of the remaining mini-grids.

The method considers a wide range of spatial and electrical metrics (such as the length of the MST that connects all the consumers and the generation site, and central electric moments) as candidate explanatory variables, and it applies hierarchical regression to determine which explanatory variables are relevant for each part of the piecewise linear model.

This chapter has presented a case study where our method estimates the network cost accurately, and it only needs to optimize from scratch the networks of less than 1% of the mini-grids. The method has been compared with a more straightforward approach based on the calculation of an MST, and it can be concluded that it provides significantly better results.

7.1.5. **Clustering**

Regional planning tools generally do not operate directly with consumers but with aggregations of consumers such as villages, settlements, or cells. These tools do not try to group the consumers into systems, and they consider the villages, settlements, or cells as the natural clustering of consumers. However, the use of administrative or artificial divisions as clusters of consumers may lead to inefficient solutions from the techno-economic point of view. For example, the electrification of a village with off-grid systems could be less expensive if the solution includes a smart combination of mini-grids and standalone systems instead of a single mini-grid that electrifies all its consumers.

Chapter 6 has presented a clustering algorithm (exhaustive clustering) that groups the consumers into candidate off-grid systems, aiming at the best possible solution from the techno-economic point of view. The algorithm considers all the costs involved and the trade-offs among them to reach the final solution. Besides, the cost estimations are based on detailed modeling of the systems, which is crucial to obtain realistic results. The clustering method we have proposed is an extension of the standard clustering of REM. The exhaustive clustering performs an in-depth evaluation of the space of clustering solutions, storing several potential solutions that form a hierarchical structure of clusters that is later evaluated to determine the least-cost clustering configuration. We have presented a sensitivity analysis that shows that the exhaustive clustering systematically outperforms the standard clustering of REM. The exhaustive clustering also behaves more robustly than the standard clustering because it can cope with constraints that distort the economies of scale in generation such as the use of solar kits.

Chapter 6 has also introduced a clustering algorithm (top-down clustering) that determines which consumers should be electrified with grid extension designs, which has been jointly developed with an MIT student (Olamide Oladeji) (Oladeji, 2018). The top-down clustering starts calculating a grid extension that electrifies all the consumers, and then it prunes elements of the network (lines and transformers) if the consumers downstream are better electrified with off-grid systems. We have presented a case study that shows that the top-down clustering provides better results than the standard clustering of REM in some cases, but the standard clustering of REM performs better in other cases. However, the top-down clustering presents some structural advantages such as considering the location of the power grid when determining the off-grid clustering and involving RNM in the estimation of costs related to network elements.

7.2. Future research

This section describes the future research lines that derive from the work presented in this thesis, which are classified according to the chapter they belong to. There is an additional section that includes future research lines related to other aspects of REM.

The future research lines regarding the exhaustive and the top-down clustering (chapter 6) should be prioritized. The clustering algorithms play an essential role in REM, and the application of the exhaustive clustering substantially enhanced the model. Additional developments concerning the clustering could lead to a similar improvement in the final electrification solution.

7.2.1. Mini-grid generation design

This section includes two future research lines regarding the calculation of off-grid generation designs in REM: (a) the addition of more generation technologies, and (b) the implementation of a robust procedure that allows REM to cope with uncertainties. Although both lines of research are important, we believe that the addition of new generation technologies is critical and should be prioritized.

7.2.1.1. Addition of generation technologies

The number of generation technologies that REM currently considers is limited to solar and diesel, and more technologies such as mini-hydro or wind should be incorporated in future versions of REM. There has been a first attempt to incorporate wind into REM (Schröder, 2019). The adaptation of the algorithm that optimizes the generation design of a single off-grid system has been successful, but there is still work to be done: wind generation is very locational (Liu et al., 2019), and this poses additional challenges. Topographical features such as the roughness, the altitude, and the contour of the terrain have a strong influence on the wind profile (Bansal et al., 2002), and they need to be incorporated in the calculation of the look-up table.

One straightforward approach would be to add an axis to the look-up table that is related to the hourly wind profile. For example, we could consider three possibilities: low wind profile, medium wind profile, and high wind profile. Then, the locational features of each candidate mini-grid would be translated into one value for the wind profile, which would be used to interpolate in the look-up table. However, the addition of several axes to the look-up table leads to a substantial increment in computation time, so this approach has limited applicability.

A different approach is based on dimensionality reduction, and it would involve working in the space of attributes where the addition of new generation technologies or features does not necessarily imply adding more axes to the look-up table. This approach is not straightforward, so it would require a significant amount of work.

Another possibility would be to run a case without considering wind generation, and selecting the off-grid systems where wind generation is promising afterwards. Then, the generation designs could be recalculated for these systems, including the possibility of wind generation.

The addition of mini-hydro generation in REM would pose similar challenges because it also depends on locational features such as the distance of the off-grid system to the hydro site of interest. It should also be highlighted that the addition of wind and hydro implies the need for accurate data regarding both technologies. It would be necessary to obtain detailed wind maps (Elliott, 2002), the location of the potential mini-hydro plants, and the flow rates of the rivers (among other things) (Mandelli et al., 2013), and it is usually complicated to collect accurate and detailed input data.

7.2.1.2. Dealing with uncertainties

The current version of REM is not properly equipped to deal with uncertainties because it is based on deterministic heuristic methods. Regarding the optimization of generation designs, there is uncertainty in the load because it is complicated to forecast future demand in an underserved region where there is currently no access to electricity (Mandelli et al., 2016). There is also uncertainty in the renewable generation, which is related to the solar irradiance in the case of solar panels (Arun et al., 2009).

There are already methods in the literature that cope with uncertainties when optimizing the generation design of a single mini-grid, some of them are based on stochastic optimization techniques (Hajipour et al., 2015) or metaheuristic algorithms (Fioriti et al., 2018) combined with Monte-Carlo simulations. A direct application of these methods into REM may fail because of computational reasons, but it would be interesting to explore an adaptation that only considers uncertainty in the load. The uncertainty could be modeled with a reduced number of scenarios concerning the load (i.e., we could consider several scenarios with high levels of demand, several scenarios with a medium-level of demand, and some scenarios with a low-level of demand).

7.2.2. Estimation of the network cost in mini-grids

This section includes future research lines regarding the network cost estimator introduced in chapter 5, which should be extended so that it can deal with (a) MV mini-grids, (b) topographical features, and (c) extensions of the power grid. We consider that the inclusion of MV mini-grids should be reasonably straightforward and should be prioritized. Extending the network cost estimator so that it can include topographical features and grid extensions seems more complicated and could require significant efforts.

7.2.2.1. Inclusion of MV mini-grids

The method that estimates the network cost of mini-grids should be extended to consider MV mini-grids too. MV mini-grids include MV/LV transformers in their distribution networks, and their cost should be accounted for (Domenech et al., 2018). It could be necessary to readjust some parts of the method to measure the impact of MV/LV transformers in the total network cost.

7.2.2.2. Addition of topography

It would also be necessary to improve the network cost estimator so that it can account for topographical features such as terrain slopes and forbidden areas, which may play an important role in the network design of a mini-grid (Müller et al., 2016; Shrestha et al., 2016). An initial idea could be to add a third coordinate to the spatial metrics to account for the terrain altitudes of the consumers and the generation site of the mini-grid.

We could include some spatial variables that measure the impact of forbidden areas in the final cost. For each forbidden zone, we could compute the distance between the center of the forbidden zone and the mini-grid and the total area of the forbidden zone (among other things).

7.2.2.3. Extrapolation to extensions of the power grid

It would be interesting to explore if the network cost estimator can be extrapolated to estimate the network cost of a grid extension. It would be necessary to include metrics that account for the location of the candidate connection points and their distance to the cluster, and the grid energy cost (RNM uses this parameter when designing a distribution network so it should be included in the method).

The extrapolation of the network cost estimator so that it can deal with grid extensions does not seem straightforward, and it could require substantial efforts to make it work. However, it could be directly applied in the grid-extension clustering, replacing the current network cost estimator. It could also be the starting point of a new grid-extension clustering that followed a logic similar to the exhaustive clustering.

7.2.3. **Clustering**

This section includes future research lines that are related to the two clustering algorithms introduced in chapter 6: (a) the inclusion of MV mini-grids in the exhaustive clustering, (b) the development of a new procedure to store clustering solutions in the exhaustive clustering, (c) a general, robust implementation of the top-down clustering, and (d) the combination of the exhaustive clustering and the top-down clustering into a single algorithm. The inclusion of MV mini-grids and the development of a new method to store solutions in the exhaustive clustering should be prioritized as they are reasonably straightforward, and their development would lead to a substantial enhancement of the exhaustive clustering.

7.2.3.1. MV mini-grids in the exhaustive clustering

The exhaustive clustering should be expanded to consider MV mini-grids (this depends on the network cost estimator). Although most mini-grids have an LV distribution network, it could be reasonable to have MV mini-grids if the economies of scale in generation justify the extra network cost, or the peak demand of consumers or their dispersion is high enough (Domenech et al., 2018).

First, it would be necessary to extend the capabilities of the network cost estimator so that it can handle MV mini-grids. Then, the exhaustive clustering could also explore the clustering solutions that include MV mini-grids and evaluate their cost. Once the network cost estimator is upgraded to include MV mini-grids, including them in the exhaustive clustering should be straightforward.

7.2.3.2. Solutions stored in the exhaustive clustering

The method that the exhaustive clustering uses to determine which intermediate solutions are stored in the hierarchical structure of clusters should be improved to include rules based on cost variations. For the time being, the exhaustive clustering stores an intermediate clustering solution if and only if a certain number of clusters have joined (for example, the intermediate solutions after 100, 200, 300, ..., clusters have joined).

A first approach would involve storing not only the intermediate solutions that appear when a certain number of clusters have joined, but also storing each clustering solution that has a lower cost than any previous solution that has already been stored. By doing so, we would at least guarantee that the lowest-cost solution is stored. If this procedure consumes too much computer memory in large-scale cases, then we could only store solutions that have a cost that is 2% (for example) lowest than any of the previous solutions.

7.2.3.3. Robust implementation of the top-down clustering

The current implementation of the top-down clustering has only been tested in cases with a relatively low number of consumers (less than 10,000 consumers), and the computation time could be a challenge in cases where the number of consumers is more substantial. A first approach to alleviate the computational burden would be to parallelize the algorithm: after

RNM has designed the initial network that electrifies all the consumers, REM could identify the specific lines of the power grid that RNM has used to design the network. Then, REM could execute one independent top-down clustering for each line involved.

Another possibility that would reduce the computation time implies joining nearby consumers together if they are closer than a specified threshold distance (i.e., we could join the consumers that are closer than five meters) before starting the top-down clustering. RNM already applies this strategy when designing a distribution network, but REM could consider a different threshold distance.

Finally, it is necessary to run the top-down clustering in more cases and scrutinize the results to gain an in-depth understanding of the algorithm and improve its performance. The top-down clustering seems to electrify more consumers with extensions of the power grid than the standard clustering of REM, and this calls for further research.

7.2.3.4. Combining the top-down and the exhaustive clustering

The top-down and the exhaustive clustering should be merged into a single clustering algorithm as they are complementary: the top-down clustering determines which consumers should be electrified with grid extensions and the exhaustive clustering groups into mini-grids the remaining ones.

This implementation should be straightforward, but we believe that it would be better to continue developing both algorithms separately for a while before considering the possibility of merging them.

7.2.4. The future of REM

There are several future research lines related to enhancing the functionality of the model or improving the performance of the current algorithms. This section describes what we consider to be the remaining short-term priorities regarding REM development: (a) an accurate calculation of the upstream reinforcements, (b) a robust implementation of the solar kits, and (c) the automated calculation of synthetic demand patterns. The robust implementation of solar kits should be prioritized since it would require a reduced effort to complete it.

7.2.4.1. Upstream reinforcements

The impact of grid-connections on the upstream distribution and transmission network should be better modeled. For the time being, REM assumes that the cost of upstream reinforcements is included in the constant input parameter that accounts for the energy cost of the power grid.

A first approach consists in applying a brownfield[7] model to the extensions of the power grid to calculate the reinforcements required upstream. This approach would lead to

[7] *We should distinguish between greenfield and brownfield models. A greenfield model designs a distribution network from scratch, and a brownfield model reinforces the already-existing network to account for new consumers or an increase of the demand. The RNM that REM currently applies to calculate the networks is a greenfield model, but there is also a brownfield RNM model.*

suboptimal solutions because an extension of the power grid could stop being the least-cost electrification alternative once the cost of upstream reinforcements is calculated. However, the approach can be extended to an iterative process in which the cost of upstream reinforcements is used to update the input parameter that accounts for the energy cost of the power grid. Then, REM obtains a new electrification solution. These ideas have already been tested (Cotterman, 2017), but they are still in a preliminary stage and performing an iterative process could be excessively time-consuming in a large-scale case.

A different alternative involves the top-down clustering algorithm. When the top-down clustering designs the initial network, REM could apply a brownfield RNM model to calculate the cost of the reinforcements needed. Then, those costs would be included in the cost-comparisons that determine which consumers would be better electrified with off-grid alternatives. We consider that this line of research is critical, and substantial efforts should be devoted to developing a robust and efficient method that could be applied in REM.

7.2.4.2. Solar kits

Solar kits should be better modeled in REM. The current version of REM determines whether to use a solar kit or an AC system to electrify an isolated consumer considering only the peak demand of the consumer (i.e., the solar kit is used if and only if the peak demand of the consumer is below a specific value). REM also calculates the amount of unserved demand related to solar kits with a quick estimation, instead of simulating the hourly dispatch of the system.

The future version of REM should model solar kits accurately, which could be achieved by simulating the hourly dispatch of solar kits to calculate the amount of non-served demand. It would be necessary to include a generation catalog for the solar kits in REM to simulate the dispatch. REM could then compare the cost of the solar kit with the cost of an AC system to determine the electrification solution for an isolated consumer, instead of doing a quick calculation with the peak demand.

REM could also include several solar kits (the current version of REM only includes one solar kit) and perform a cost-comparison to determine the least-cost individual system for each consumer type.

7.2.4.3. Synthetic demand patterns

Large-scale planning cases have multiple types of loads, such as residential households, schools, hospitals or telecom towers (among many others). Each type of load has a demand profile that is considered in the calculation of the look-up table and could be independent of the remaining profiles.

REM allows the user to define basic demand patterns to narrow the number of dimensions of the look-up table. The demand profile of each type of load is expressed as a linear combination of the basic demand patterns, limiting the number of dimensions of the look-up table to the number of basic demand patterns.

However, the manual calculation of the basic demand patterns is complicated in cases with many types of loads. It is useful to develop a procedure that automatically calculates the basic demand patterns out of the load profiles, and we use the term "synthetic demand pattern" to refer to a basic demand pattern calculated with a dimensionality reduction method.

We are currently working on the calculation of synthetic demand patterns applying PCA, but more work is required to obtain a solution that can be implemented in REM. The following points need further research:

- Synthetic demand patterns do not have a physical meaning, and they could take negative values. It would be useful to find a method that produces synthetic patterns without negative demands.
- The user currently determines the representative off-grid systems whose generation designs are calculated from scratch, but it is unclear if this is compatible with the automated calculation of synthetic demand patterns. It could be necessary to develop a robust method that automatically determines the representative off-grid systems whose generation designs are calculated from scratch.
- PCA allows the use of weights so that the profiles of some load types are approximated accurately, but it comes at the expense of losing precision in the approximation of the remaining load types. Further research is needed to determine how to assignate weights to demand profiles of consumers when calculating the synthetic demand patterns.

7.3. Contributions

This section summarizes the main contributions of this thesis. The contributions are classified according to the chapters of this work.

7.3.1. Large-scale electrification planning tools

The contributions presented in chapter 2 are:

- The definition of the electrification problem from a techno-economic perspective, including a formulation that is used to compare tools and methodologies.
- The critical review of large-scale electrification planning methods and tools, incorporating considerations about both modeling and solution methods.
- The list of the main challenges concerning electrification planning tools that should be addressed by the electrification community.

7.3.2. Domain-based improvements

The contribution presented in chapter 3 is:

- An in-depth analysis that has led to the transformation of the first prototype of REM – conceptually sound at a high level, but dysfunctional when applied to actual electrification cases – into a sound, reliable, and fully functional computer tool that has been already used in developing the master electrification plans of various countries, from Indonesia and Cambodia to Rwanda and Mozambique, plus specific analysis in territories of Nigeria and India, among other cases. This has been achieved by identifying the existing shortcomings in the initial algorithms and overcoming them. This has resulted in significant improvements in the performance of (a) the optimization of generation designs for off-grid systems, (b) the clustering of consumers into candidate mini-grids and extensions of the power grid, and (c) the calculation of the final electrification mode of each cluster. The analysis has also led to enhancements that expanded the capabilities of the model, allowing REM to consider different consumers types and including solar kits as a viable electrification alternative.

7.3.3. Mini-grid generation design

The contributions presented in chapter 4 are:

- The study of the impact of discrete off-grid generation components (such as diesel generators) in the standard off-grid clustering of REM, which groups the consumers into mini-grids according to the generation costs (among others). This study has shown that a method that optimizes the generation designs modeling all the capacities with discrete variables (such as methods that operate with an individual village or settlement) may lead to suboptimal solutions when grouping the consumers into mini-grids.
- Two alternative methods that mitigate the impact of discrete off-grid generation components in the standard clustering of REM. The first one approximates the generation costs with a smooth curve, and the second one models the capacity of elements that could alter the economies of scale with continuous variables. Both methods ensure that larger mini-grids benefit from economies of scale in generation when grouping the consumers into mini-grids, but the method based on continuous variables is directly applicable to cases with several types of loads.

7.3.4. Estimation of the network cost of mini-grids

The contribution presented in chapter 5 is:

- A method that can estimate the network cost of any potential LV mini-grid in a region. The method optimizes the network design of a representative subset of mini-grids, and it uses the results to determine the explanatory variables and the coefficients of a piecewise linear model, which can estimate the network cost of any LV mini-grid. We have presented a real case study where the network cost of each mini-grid is calculated accurately, and less than 1% of the network designs are optimized from scratch.

7.3.5. **Clustering**

The contributions presented in chapter 6 are:

- A clustering algorithm, named exhaustive clustering, that aims at determining the best grouping of consumers into off-grid systems. This method applies the network cost estimator presented in chapter 5, and it presents substantial advantages over the standard off-grid clustering of REM. On the one hand, the exhaustive clustering is more robust than the standard off-grid clustering because it performs a broader exploration of the space of potential clustering solutions. On the other hand, the standard clustering follows a logic that requires monotonous economies of scale in generation (i.e., the quotient between the generation cost and the total demand of a mini-grid cannot increase when the total demand of that mini-grid increases), but the exhaustive clustering can handle elements or constraints that are not compatible with monotonous economies of scale in generation (such as solar kits).
- A clustering algorithm, named top-down clustering, that determines which consumers should be electrified with extensions of the distribution power grid. This method presents several advantages over the standard clustering of REM. For example, the top-down clustering can handle topographical features such as terrain slopes or forbidden zones that distribution lines cannot cross, and the standard clustering of REM is not fully compatible with topography. Moreover, the cost estimations of network elements (lines and transformers) are better in the top-down clustering than in the standard clustering.

7.4. Publications

Part of the work presented in this thesis has been published in the following papers:

Ciller, P., Lumbreras, S., 2020. Electricity for all: The contribution of large-scale planning tools to the energy-access problem. Renewable and Sustainable Energy Reviews 120, 109624. https://doi.org/10.1016/j.rser.2019.109624

Ciller, P., Ellman, D., Vergara, C., Gonzalez-Garcia, A., Lee, S.J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., Palacios, R., Stoner, R., de Cuadra, F., Perez-Arriaga, I., 2019. Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM). Proceedings of the IEEE 107, 1872–1905. https://doi.org/10.1109/JPROC.2019.2922543

Ciller, P., de Cuadra, F., Lumbreras, S., 2019. Optimizing Off-Grid Generation in Large-Scale Electrification-Planning Problems: A Direct-Search Approach. Energies 12, 4634. https://doi.org/10.3390/en12244634

We expect that part of the work presented in this thesis will be published in the following paper:

Oladeji, O., Ciller, P., de Cuadra, F., Perez-Arriaga, I. Partitioning Distribution Networks: An Approach to Integrated Electrification Planning. IEEE Transactions on Power Systems.

Submitted.

Some of the developments presented in this thesis were applied in cases that were analyzed for a collaboration in the following two books:

International Energy Agency, 2018. World Energy Outlook 2018. International Energy Agency, Paris. URL https://www.oecd-ilibrary.org/content/publication/weo-2018-en

International Energy Agency, 2019. Africa Energy Outlook 2019. International Energy Agency, Paris. URL https://www.iea.org/reports/africa-energy-outlook-2019

In the first book mentioned (International Energy Agency, 2018), REM was applied to analyze the potential for clean cooking in a representative area of Africa. REM was also applied to study the importance of accurate demand projections and detailed spatial designs that go down to the building level in a region located in the South Service Territory in Uganda.

In the second book mentioned (International Energy Agency, 2019), REM was used to determine the impact of the grid reliability in the final electrification solution in a region located in Rwanda.

References

Abdul-Salam, Y., Phimister, E., 2016a. How effective are heuristic solutions for electricity planning in developing countries. Socio-Economic Planning Sciences 55, 14–24. https://doi.org/10.1016/j.seps.2016.04.004

Abdul-Salam, Y., Phimister, E., 2016b. The politico-economics of electricity planning in developing countries: A case study of Ghana. Energy Policy 88, 299–309. https://doi.org/10.1016/j.enpol.2015.10.036

Alramlawi, M., Mohagheghi, E., Li, P., 2019. Predictive active-reactive optimal power dispatch in PV-battery-diesel microgrid considering reactive power and battery lifetime costs. Solar Energy 193, 529–544. https://doi.org/10.1016/j.solener.2019.09.034

Al-Rubaye, R.T.A., Al-Rubaye, A.T.A., Al-Khuzaie, M.M., 2018. Optimal Design of Hybrid Renewable Energy System off grid in Al-Diwaniyah, Iraq. IOP Conference Series: Materials Science and Engineering 454, 012103. https://doi.org/10.1088/1757-899X/454/1/012103

Amador, J., Domínguez, J., 2005. Application of geographical information systems to rural electrification with renewable energy sources. Renewable Energy 30, 1897–1912. https://doi.org/10.1016/j.renene.2004.12.007

Arun, P., Banerjee, R., Bandyopadhyay, S., 2009. Optimum sizing of photovoltaic battery systems incorporating uncertainty through design space approach. Solar Energy 13.

Azimoh, C.L., Klintenberg, P., Mbohwa, C., Wallin, F., 2017. Replicability and scalability of mini-grid solution to rural electrification programs in sub-Saharan Africa. Renewable Energy 106, 222–231. https://doi.org/10.1016/j.renene.2017.01.017

Ballestero, E., Romero, C., 1998. Multiple Criteria Decision Making and its Applications to Economic Problems. Springer US.

Banks, D.I., Mocke, F., Jonck, E.C., Labuschagne, E., Eberhard, R., 2000. Electrification planning decision support tool, in: Domestic Use of Energy Conference. Citeseer.

Bansal, R.C., Bhatti, T.S., Kothari, D.P., 2002. On some of the design aspects of wind energy conversion systems. Energy Conversion and Management 43, 2175–2187. https://doi.org/10.1016/S0196-8904(01)00166-2

Baradar, M., Hesamzadeh, M.R., 2015. AC Power Flow Representation in Conic Format. IEEE Trans. Power Syst. 30, 546–547. https://doi.org/10.1109/TPWRS.2014.2326980

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons.

Bertheau, P., Cader, C., Blechinger, P., 2016. Electrification Modelling for Nigeria. Energy Procedia 93, 108–112. https://doi.org/10.1016/j.egypro.2016.07.157

Bertheau, P., Oyewo, A., Cader, C., Breyer, C., Blechinger, P., 2017. Visualizing National Electrification Scenarios for Sub-Saharan African Countries. Energies 10, 1899. https://doi.org/10.3390/en10111899

Better Explained, 2020. Navigate a Grid Using Combinations And Permutations [WWW Document]. URL https://betterexplained.com/articles/navigate-a-grid-using-combinations-and-permutations/ (accessed 8.7.20).

Bhattacharyya, S.C., 2015. Mini-grid based electrification in Bangladesh: Technical configuration and business analysis. Renewable Energy 75, 745–761. https://doi.org/10.1016/j.renene.2014.10.034

Bholowalia, P., 2014. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. International Journal of Computer Applications 105, 8.

Blechinger, P., Cader, C., Bertheau, P., 2019. Least-Cost Electrification Modeling and Planning–A Case Study for Five Nigerian Federal States. Proceedings of the IEEE 107, 1923–1940. https://doi.org/10.1109/JPROC.2019.2924644

Blodgett, C., Dauenhauer, P., Louie, H., Kickham, L., 2017. Accuracy of energy-use surveys in predicting rural mini-grid user consumption. Energy for Sustainable Development 41, 88–105. https://doi.org/10.1016/j.esd.2017.08.002

Borhanazad, H., Mekhilef, S., Gounder Ganapathy, V., Modiri-Delshad, M., Mirtaheri, A., 2014. Optimization of micro-grid system using MOPSO. Renewable Energy 71, 295–306. https://doi.org/10.1016/j.renene.2014.05.006

Brivio, C., Moncecchi, M., Mandelli, S., Merlo, M., 2017. A novel software package for the robust design of off-grid power systems. Journal of Cleaner Production 166, 668–679. https://doi.org/10.1016/j.jclepro.2017.08.069

Brusnahan, M.D., 2018. Minigrids for Electrification: Policies to Promote Industry Growth. Master thesis. Massachusetts Institute of Technology. Institute for Data, Systems, and Society.

Cader, C., Pelz, S., Radu, A., Blechinger, P., 2018. Overcoming data scarcity for energy access planning with open data – the example of Tanzania. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W8, 23–26. https://doi.org/10.5194/isprs-archives-XLII-4-W8-23-2018

Carrasco, L.M., Narvarte, L., Lorenzo, E., 2013. Operational costs of A 13,000 solar home systems rural electrification programme. Renewable and Sustainable Energy Reviews 20, 1–7. https://doi.org/10.1016/j.rser.2012.11.073

Ciller Cutillas, P., 2016. Clustering-related improvements in the Reference Electrification Model. Master thesis. Universidad Pontificia Comillas, Madrid. School of Engineering.

Ciller, P., Ellman, D., Vergara, C., Gonzalez-Garcia, A., Lee, S.J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., Palacios, R., Stoner, R., de Cuadra, F., Perez-Arriaga, I., 2019a. Optimal Electrification Planning Incorporating On- and Off-Grid Technologies: The Reference Electrification Model (REM). Proceedings of the IEEE 107, 1872–1905. https://doi.org/10.1109/JPROC.2019.2922543

Ciller, P., de Cuadra, F., Lumbreras, S., 2019b. Optimizing Off-Grid Generation in Large-Scale Electrification-Planning Problems: A Direct-Search Approach. Energies 12, 4634. https://doi.org/10.3390/en12244634

Ciller, P., Lumbreras, S., 2020. Electricity for all: The contribution of large-scale planning tools to the energy-access problem. Renewable and Sustainable Energy Reviews 120, 109624. https://doi.org/10.1016/j.rser.2019.109624

Columbia University, 2017. Network Planner Website [WWW Document]. URL http://qsel.columbia.edu/network-planner/ (accessed 4.29.18).

Comisión de Regulación de Energía y Gas, 2020. Remuneración de la prestación del servicio de energía eléctrica mediante soluciones individuales fotovoltaicas. (In Spanish). URL http://apolo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/030202e046240b58052585b6005e48ee/$FILE/DOCUMENTO%20REMUNERACI%C3%93N%20DE%20LA%20PRESTACI%C3%93N%20DEL%20SERVICIO%20DE%20ENERG%C3%8DA%20EL%C3%89CTRICA%20MEDIANTE%20SOLUCIONES%20INDIVIDUALES%20SOLARES%20FOTOVOLTAICAS.pdf

Cotterman, T., 2017. Enhanced techniques to plan rural electrical networks using the Reference Electrification Model. Master thesis. Massachusetts Institute of Technology. Institute for Data, Systems, and Society.

DeCarolis, J., Daly, H., Dodds, P., Keppo, I., Li, F., McDowall, W., Pye, S., Strachan, N., Trutnevyte, E., Usher, W., Winning, M., Yeh, S., Zeyringer, M., 2017. Formalizing best practice for energy system optimization modelling. Applied Energy 194, 184–198. https://doi.org/10.1016/j.apenergy.2017.03.001

Deichmann, U., Meisner, C., Murray, S., Wheeler, D., 2011. The economics of renewable energy expansion in rural Sub-Saharan Africa. Energy Policy 39, 215–227. https://doi.org/10.1016/j.enpol.2010.09.034

Domenech, B., Ranaboldo, M., Ferrer-Martí, L., Pastor, R., Flynn, D., 2018. Local and regional microgrid models to optimise the design of isolated electrification projects. Renewable Energy 119, 795–808. https://doi.org/10.1016/j.renene.2017.10.060

Domínguez Bravo, J., Pinedo Pascua, I., Augusto González, J., 2008. Herramientas SIG para la integración de energías renovables en electrificación rural. IntiGIS., in: Hernández Calvento, L., Parreño Castellano, J.M. (Eds.), . Presented at the XIII Congreso Nacional de Tecnologías de la Información Geográfica (in Spanish), Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria.

Domínguez, J., Pinedo-Pascua, I., 2009. GIS Tool for Rural Electrification with Renewable Energies in Latin America, in: 2009 International Conference on Advanced Geographic Information Systems & Web Services. Presented at the Web Services, IEEE, Cancun, Mexico, pp. 171–176. https://doi.org/10.1109/GEOWS.2009.25

Drouin, C., 2018. Geospatial Cost Drivers in Computer-aided Electrification Planning: The Case of Rwanda. Master thesis. Massachusetts Institute of Technology. Department of Mechanical engineering.

Dufo-López, R., Bernal-Agustín, J.L., Yusta-Loyo, J.M., Domínguez-Navarro, J.A., Ramírez-Rosado, I.J., Lujano, J., Aso, I., 2011. Multi-objective optimization minimizing cost and life cycle emissions of stand-alone PV–wind–diesel systems with batteries storage. Applied Energy 88, 4033–4041. https://doi.org/10.1016/j.apenergy.2011.04.019

Elliott, D., 2002. Assessing the world's wind resources, in: 2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.02CH37309). Presented at the Winter Meeting of the Power Engineering Society, IEEE, New York, NY, USA, pp. 346–347. https://doi.org/10.1109/PESW.2002.985013

Ellman, D., 2015. The Reference Electrification Model: A Computer Model for Planning Rural Electricity Access. Master thesis. Massachusetts Institute of Technology. Engineering Systems Division.

Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University, 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe. [WWW Document]. URL https://www.ciesin.columbia.edu/data/hrsl/ (accessed 9.11.19).

Ferrer-Martí, L., Domenech, B., García-Villoria, A., Pastor, R., 2013. A MILP model to design hybrid wind–photovoltaic isolated rural electrification projects in developing countries. European Journal of Operational Research 226, 293–300. https://doi.org/10.1016/j.ejor.2012.11.018

Fioriti, D., Giglioli, R., Poli, D., Lutzemberger, G., Micangeli, A., Del Citto, R., Perez-Arriaga, I., Duenas-Martinez, P., 2018. Stochastic sizing of isolated rural mini-grids, including effects of fuel procurement and operational strategies. Electric Power Systems Research 160, 419–428. https://doi.org/10.1016/j.epsr.2018.03.020

First Solar, 2017. First Solar FS-497 datasheet [WWW Document]. URL http://www.firstsolar.com/-/media/First-Solar/Documents/PD-5-401-04_Series_4_Datasheet.ashx (accessed 3.2.17).

Flusser, J., 2000. On the independence of rotation moment invariants. Pattern Recognition 33, 1405–1410. https://doi.org/10.1016/S0031-3203(99)00127-2

Fronius, R., Gratton, M., 2001. Rural electrification planning software (LAPER). Presented at the 16th International Conference and Exhibition on Electricity Distribution (CIRED 2001), IEE, Amsterdam, Netherlands, pp. v5-20-v5-20. https://doi.org/10.1049/cp:20010884

Fuso Nerini, F., Dargaville, R., Howells, M., Bazilian, M., 2015. Estimating the cost of energy access: The case of the village of Suro Craic in Timor Leste. Energy 79, 385–397. https://doi.org/10.1016/j.energy.2014.11.025

Ganguly, P., Kalam, A., Zayegh, A., 2017. Design an optimum standalone hybrid renewable energy system for a small town at Portland, Victoria using iHOGA, in: 2017 Australasian Universities Power Engineering Conference (AUPEC). Presented at the 2017 Australasian Universities Power Engineering Conference (AUPEC), IEEE, Melbourne, VIC, pp. 1–6. https://doi.org/10.1109/AUPEC.2017.8282487

Garcia Conejo, J.A., Tevar Bartolome, G., Gomez Exposito, A., Rodriguez Montanes, M., 2007. ANETO: A System for the Automatic Generation of Theoretical Network Models. 9th International Conference on Electrical Power Quality and Utilisation. https://doi.org/10.1109/EPQU.2007.4424189

Gebrehiwot, K., Mondal, Md.A.H., Ringler, C., Gebremeskel, A.G., 2019. Optimization and cost-benefit assessment of hybrid power systems for off-grid rural electrification in Ethiopia. Energy 177, 234–246. https://doi.org/10.1016/j.energy.2019.04.095

Georgilakis, P.S., Hatziargyriou, N.D., 2015. A review of power distribution planning in the modern power systems era: Models, methods and future research. Electric Power Systems Research 121, 89–100. https://doi.org/10.1016/j.epsr.2014.12.010

Gholizadeh-Roshanagh, R., Najafi-Ravadanegh, S., Hosseinian, S.H., 2016. Optimal robust integrated power distribution network planning under load demand uncertainty. Journal of Power Technologies 96, 115–123.

Gómez, T., Mateo, C., Sánchez, Á., Frías, P., Cossent, R., 2013. Reference Network Models: a computational tool for planning and designing large-scale smart electricity distribution grids, in: Siddhartha Kumar, K., Gupta, A. (Eds.), High Performance Computing in Power and Energy Systems. Springer Berlin Heidelberg, Berlin, pp. 247–279.

González–García, A., Pérez–Arriaga, I.J., Moreno-Romero, A., 2014. MASTER4all: A decision support model for roadmaps towards sustainable energy universal access for policymakers and utilities, in: Innovating Energy Access for Remote Areas: Discovering Untapped Resources - Symposium. Berkeley, USA.

Gonzalez-Garcia, A., Amatya, R., Stoner, R., Perez-Arriaga, J.I., 2016. Evaluation of universal access to modern energy services in Peru. Case study of scenarios for Electricity Access in Cajamarca. Enel Foundation Working Papers. URL https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/15559/IIT-16-155A.pdf?sequence=1&isAllowed=y

González-Sotres, L., Mateo, C., Sánchez-Miralles, Á., Miró, M.A., 2013. Large-Scale MV / LV Transformer Substation Planning Considering Network Costs and Flexible Area Decomposition. Power Delivery, IEEE Transactions on 28, 2245–2253.

Govender, K.M., Meyer, A.S., Dwolatzky, B., 2001. The automatic clustering of uniformly distributed loads for the use in rural electrification planning, in: Transmission and Distribution Conference and Exposition, 2001 IEEE/PES. IEEE, pp. 450–455.

Hajipour, E., Bozorg, M., Fotuhi-Firuzabad, M., 2015. Stochastic Capacity Expansion Planning of Remote Microgrids With Wind Farms and Energy Storage. IEEE Trans. Sustain. Energy 6, 491–498. https://doi.org/10.1109/TSTE.2014.2376356

Hartvigsson, E., Stadler, M., Cardoso, G., 2018. Rural electrification and capacity expansion with an integrated modeling approach. Renewable Energy 115, 509–520. https://doi.org/10.1016/j.renene.2017.08.049

HOMER Energy LLC, 2019a. Kinetic Battery Model [WWW Document]. URL https://www.homerenergy.com/products/pro/docs/latest/kinetic_battery_model.html (accessed 10.27.20).

HOMER Energy LLC, 2019b. Load Following Strategy [WWW Document]. URL https://www.homerenergy.com/products/pro/docs/latest/load_following_strategy.html (accessed 9.27.19).

Hooke, R., Jeeves, T.A., 1961. "Direct Search" Solution of Numerical and Statistical Problems. J. ACM 8, 212–229. https://doi.org/10.1145/321062.321069

Howells, M.I., Alfstad, T., Victor, D.G., Goldstein, G., Remme, U., 2005. A model of household energy services in a low-income rural African village. Energy Policy 33, 1833–1851. https://doi.org/10.1016/j.enpol.2004.02.019

Huff, D.L., 1963. A Probabilistic Analysis of Shopping Center Trade Areas. Land Economics 39, 81–90. https://doi.org/10.2307/3144521

IIT-Comillas, 2016. REM: Test Cases Colombia and Kenya [WWW Document]. URL https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.es?nombre_abreviado=ENEL%20Kenya-Colombia%20Test%20Cases (accessed 4.4.18).

IIT-Comillas, 2017a. Promotion of mini-grids for rural electrification [WWW Document]. URL https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.es?nombre_abreviado=2017%20GIZ%20Uganda (accessed 4.4.18).

IIT-Comillas, 2017b. Assistance to the Government of Rwanda to develop a National Electrification Strategy (NES) and related investment plans [WWW Document]. URL https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.es?nombre_abreviado=Rwanda%20NES%2017 (accessed 4.4.18).

IIT-Comillas, 2019. Mozambique geospatial options analysis towards universal electrification

[WWW Document]. URL https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.en?nombre_abreviado=MIT-NEP-MOZAMBIQUE-2018 (accessed 9.27.20).

IIT-Comillas, 2020. Institute for Research in Technology webpage [WWW Document]. URL https://www.iit.comillas.edu/index.php.en (accessed 10.16.20).

Innovation Energie Développement, 2007. Capacity building for Off-grid rural electrification planning at the regional level in Ethiopia. URL http://www.euei-pdf.org/en/seads/capacity-building/capacity-building-for-off-grid-rural-electrification-planning (accessed 4.28.18).

Innovation Energie Développement, 2010. CAP-REDEO results report. URL https://ec.europa.eu/energy/intelligent/projects/en/projects/cap-redeo (accessed 4.28.18).

Innovation Energie Developpement, 2018. GEOSIM projects [WWW Document]. URL http://www.geosim.fr/index.php?page=references-en (accessed 4.29.18).

International Energy Agency, 2018. World Energy Outlook 2018. International Energy Agency, Paris. URL https://www.oecd-ilibrary.org/content/publication/weo-2018-en

International Energy Agency, International Renewable Energy Agency, United Nations Statistics Division, World Bank, World Health Organization, 2019. Tracking SDG7. The Energy Progress Report 2019. URL https://trackingsdg7.esmap.org/data/files/download-documents/2019-Tracking%20SDG7-Full%20Report.pdf (accessed 8.10.19).

International Energy Agency, 2019. Africa Energy Outlook 2019. International Energy Agency, Paris. URL https://www.iea.org/reports/africa-energy-outlook-2019

Kaufman, L., Rousseeuw, P.J., 1990. Finding groups in data: an introduction to cluster analysis, Wiley series in probability and mathematical statistics. Wiley, Hoboken, N.J.

Kemausuor, F., Adkins, E., Adu-Poku, I., Brew-Hammond, A., Modi, V., 2014a. Electrification planning using Network Planner tool: The case of Ghana. Energy for Sustainable Development 19, 92–101. https://doi.org/10.1016/j.esd.2013.12.009

Kemausuor, F., Adkins, E., Adu-Poku, I., Brew-Hammond, A., Modi, V., 2014b. Electrification planning using Network Planner tool: The case of Ghana. Energy for Sustainable Development 19, 92–101. https://doi.org/10.1016/j.esd.2013.12.009

Khatod, D.K., Pant, V., Sharma, J., 2010. Analytical Approach for Well-Being Assessment of Small Autonomous Power Systems With Solar and Wind Energy Sources. IEEE Transactions on Energy Conversion 25, 535–545. https://doi.org/10.1109/TEC.2009.2033881

Kirgat, G.S., Surde, A.N., 2014. Review of Hooke and Jeeves Direct Search Solution Method Analysis Applicable To Mechanical Design Engineering. Internationanl Journal of Innovations in Engineering Research and Technology 1, 1–14.

Kocaman, A.S., Huh, W.T., Modi, V., 2012. Initial layout of power distribution systems for rural electrification: A heuristic algorithm for multilevel network design. Applied Energy 96, 302–315. https://doi.org/10.1016/j.apenergy.2012.02.029

Kong, K.C., Muzathik, A.M., 2012. New Approach on Mathematical Modeling of Photovoltaic Solar Panel. Applied Mathematical Sciences 6, 21.

Korkovelos, A., Bazilian, M., Mentis, D., Howells, M., 2017. A GIS approach to planning electrification in Afghanistan. World Bank, Washington D.C. URL https://energypedia.info/images/d/d6/A_GIS_approach_to_electrification_planning_in_Afghanistan.pdf (accessed 8.16.19).

Korkovelos, A., Khavari, B., Sahlberg, A., Howells, M., Arderne, C., 2019. The Role of Open Access Data in Geospatial Electrification Planning and the Achievement of SDG7. An OnSSET-Based Case Study for Malawi. Energies 12, 1395. https://doi.org/10.3390/en12071395

Korkovelos, A., Mentis, D., Bazilian, M., Howells, M., Saraj, A., Fayez Hotaki, S., Missfeldt-Ringius, F., 2020. Supporting Electrification Policy in Fragile States: A Conflict-Adjusted Geospatial Least Cost Approach for Afghanistan. Sustainability 12, 777. https://doi.org/10.3390/su12030777

KTH Royal Institute of Technology, SNV Netherlands Development Organisation, 2018. Electrification Pathways to Benin. URL http://www.onsset.org/electrification-pathways-for-benin.html (accessed 8.16.19).

KTH Royal Institute of Technology, 2018. OnSSET Projects [WWW Document]. OnSSET Projects. URL http://www.onsset.org/projects.html (accessed 3.27.18).

Lambert, T., Gilman, P., Lilienthal, P., 2006. Micropower system modeling with HOMER. Integration of alternative sources of energy 1, 379–418.

Lambert, T.W., Hittle, D.C., 2000. Optimization of autonomous village electrification systems by simulated annealing. Solar Energy 68, 121–132.

Larsson, M.B.-O., 2005. The Network Performance Assessment Model: A new framework for regulating the Electricity network Companies. Ph.D. thesis. KTH Royal Institute of Technology. Department of Electrical Engineering., Stockholm, Sweden.

Lawrence Berkeley National Laboratory, 2017. Distributed Energy Resources Customer Adoption Model (DER-CAM) | Building Microgrid [WWW Document]. URL https://building-microgrid.lbl.gov/projects/der-cam (accessed 2.19.17).

Lawson, A., Goldstein, M., Dent, C.J., 2016. Bayesian framework for power network planning under uncertainty. Sustainable Energy, Grids and Networks 7, 47–57. https://doi.org/10.1016/j.segan.2016.05.003

Lewis, M., 2007. Running head: Stepwise versus Hierarchal Regression. Presented at the Annual meeting of the Southwest Educational Research Association, San Antonio, TX, USA, p. 30.

Li, C., Rahman, A., 1990. Three-Phase Induction Motor Design Optimization Using the Modified Hooke-Jeeves Method. Electric Machines & Power Systems 18, 1–12. https://doi.org/10.1080/07313569008909446

Li, V., 2016. The Local Reference Electrification Model: A Comprehensive Decision-Making Tool for the Design of Rural Microgrids. Master thesis. Massachusetts Institute of Technology. Institute for Data, Systems, and Society.

Liu, F., Sun, F., Liu, W., Wang, T., Wang, H., Wang, X., Lim, W.H., 2019. On wind speed pattern and energy potential in China. Applied Energy 236, 867–876. https://doi.org/10.1016/j.apenergy.2018.12.056

Liu, X., Zhang, Z., 2014. A hybrid reliability approach for structure optimisation based on probability and ellipsoidal convex models. Journal of Engineering Design 25, 238–258. https://doi.org/10.1080/09544828.2014.961060

Louie, H., Dauenhauer, P., 2016. Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability. Energy for Sustainable Development 34, 30–43. https://doi.org/10.1016/j.esd.2016.08.002

Lujano-Rojas, J.M., Dufo-López, R., Bernal-Agustín, J.L., 2013. Probabilistic modelling and analysis of stand-alone hybrid power systems. Energy 63, 19–27. https://doi.org/10.1016/j.energy.2013.10.003

Luna-Rubio, R., Trejo-Perea, M., Vargas-Vázquez, D., Ríos-Moreno, G.J., 2012. Optimal sizing of renewable hybrids energy systems: A review of methodologies. Solar Energy 86, 1077–1088. https://doi.org/10.1016/j.solener.2011.10.016

Mahapatra, S., Dasappa, S., 2012. Rural electrification: Optimising the choice between decentralised renewable energy sources and grid extension. Energy for Sustainable Development 16, 146–154. https://doi.org/10.1016/j.esd.2012.01.006

Mahmood, M., Shivam, O., Kumar, P., Krishnan, G., Scholar, U., Scholar, P., 2014. Real Time Study on Technical Losses in Distribution System. International Journal of Advanced Research in Electrical, Electronics and Intrumentation Engineering 3, 7.

Maleki, A., Askarzadeh, A., 2014. Artificial bee swarm optimization for optimum sizing of a stand-alone PV/WT/FC hybrid system considering LPSP concept. Solar Energy 107, 227–235. https://doi.org/10.1016/j.solener.2014.05.016

Mandelli, S., Colombo, E., Redondi, A., Bernardi, F., Saanane, B.B., Mgaya, P., Malisa, J., 2013. A small-hydro plant model for feasibility analysis of electrification projects in Rural Tanzania, in: 2013 IEEE Global Humanitarian Technology Conference (GHTC). Presented at the 2013 IEEE Global Humanitarian Technology Conference (GHTC), IEEE, San Jose, CA, pp. 11–16. https://doi.org/10.1109/GHTC.2013.6713646

Mandelli, S., Brivio, C., Colombo, E., Merlo, M., 2016. Effect of load profile uncertainty on the optimum sizing of off-grid PV systems for rural electrification. Sustainable Energy Technologies and Assessments 18, 34–47. https://doi.org/10.1016/j.seta.2016.09.010

Manera, M., Marzullo, A., 2005. Modelling the load curve of aggregate electricity consumption using principal components. Environmental Modelling & Software 20, 1389–1400. https://doi.org/10.1016/j.envsoft.2004.09.019

Manwell, J.F., McGowan, J.G., 1993. Lead acid battery storage model for hybrid energy systems. Solar Energy 50, 399–405. https://doi.org/10.1016/0038-092X(93)90060-2

Mateo Domingo, C., Gómez San Román, T., Sanchez-Miralles, Á., Peco Gonzalez, J.P., Candela Martinez, A., 2011. A Reference Network Model for Large-Scale Distribution Planning With Automatic Street Map Generation. IEEE Transactions on Power Systems 26, 190–197. https://doi.org/10.1109/TPWRS.2010.2052077

MathWorks, 2014. k-medoids clustering - Matlab kmedoids - Mathworks [WWW Document]. URL https://es.mathworks.com/help/stats/kmedoids.html?lang=en (accessed 12.11.19).

Megherbi, H., Seigneuric, A., Ehrlich, M.-F., 2006. Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. European Journal of Psychology of Education 21, 135–147. https://doi.org/10.1007/BF03173573

Mentis, D., Welsch, M., Fuso Nerini, F., Broad, O., Howells, M., Bazilian, M., Rogner, H., 2015. A GIS-based approach for electrification planning—A case study on Nigeria. Energy for Sustainable Development 29, 142–150. https://doi.org/10.1016/j.esd.2015.09.007

Mentis, D., Andersson, M., Howells, M., Rogner, H., Siyal, S., Broad, O., Korkovelos, A., Bazilian, M., 2016. The benefits of geospatial planning in energy access – A case study on Ethiopia. Applied Geography 72, 1–13. https://doi.org/10.1016/j.apgeog.2016.04.009

Mentis, D., 2017. Spatially explicit electrification modelling insights (PhD Thesis). KTH Royal Institute of Technology. Division of Energy Systems Analysis., Sweden.

Mentis, D., Howells, M., Rogner, H., Korkovelos, A., Arderne, C., Zepeda, E., Siyal, S., Taliotis, C., Bazilian, M., de Roo, A., Tanvez, Y., Oudalov, A., Scholtz, E., 2017a. Lighting the World: the first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa. Environmental Research Letters 12, 085003. https://doi.org/10.1088/1748-9326/aa7b29

Mentis, D., Howells, M., Rogner, H., Korkovelos, A., Arderne, C., Siyal, S., Zepeda, E., Taliotis, C., Bazilian, M., de Roo, A., Tanvez, Y., Oudalov, A., Scholtz, E., 2017b. Linking the open source, spatial electrification tool (ONSSET) and the open source energy modelling system (OSeMOSYS), with a focus on Sub-Saharan Africa. Presented at the EGU General Assembly Conference Abstracts, p. 15758.

Merei, G., Berger, C., Sauer, D.U., 2013. Optimization of an off-grid hybrid PV–Wind–Diesel system with different battery technologies using genetic algorithm. Solar Energy 97, 460–473. https://doi.org/10.1016/j.solener.2013.08.016

Micangeli, A., Del Citto, R., Kiva, I., Santori, S., Gambino, V., Kiplagat, J., Viganò, D., Fioriti, D., Poli, D., 2017. Energy Production Analysis and Optimization of Mini-Grid in Remote Areas: The Case Study of Habaswein, Kenya. Energies 10, 2041. https://doi.org/10.3390/en10122041

Ming-Kuei Hu, 1962. Visual pattern recognition by moment invariants. IEEE Transactions on Information Theory 8, 179–187. https://doi.org/10.1109/TIT.1962.1057692

Mirakyan, A., De Guio, R., 2015. Modelling and uncertainties in integrated energy planning. Renewable and Sustainable Energy Reviews 46, 62–69. https://doi.org/10.1016/j.rser.2015.02.028

MIT & IIT-Comillas Universal Energy Access Lab, 2019. UEA Lab webpage [WWW Document]. URL http://universalaccess.mit.edu/ (accessed 1.5.19).

Modi, V., Adkins, E., Carbajal, J., Shepa, S., 2013. Liberia power sector capacity building and energy master planning, Final Report. Phase 4: National Electrification Master Plan. URL http://sel.columbia.edu/assets/uploads/blog/2013/09/LiberiaEnergySectorReform_Phase4Report-Final_2013-08.pdf (accessed 12.18.16).

Mohammed, O., Amirat, Y., Benbouzid, M., 2018. Economical Evaluation and Optimal Energy Management of a Stand-Alone Hybrid Energy System Handling in Genetic Algorithm Strategies. Electronics 7, 233. https://doi.org/10.3390/electronics7100233

Moksnes, N., Korkovelos, A., Mentis, D., Howells, M., 2017. Electrification pathways for Kenya–linking spatial electrification analysis and medium to long term energy planning. Environmental Research Letters 12, 095008. https://doi.org/10.1088/1748-9326/aa7e18

Moller, N., Fouladi, R., J. McCarthy, C., D. Hatch, K., 2003. Relationship of Attachment and Social Support to College Students' Adjustment Following a Relationship Breakup. https://doi.org/10.1002/j.1556-6678.2003.tb00262.x

Moner-Girona, M., Bódis, K., Huld, T., Kougias, I., Szabó, S., 2016. Universal access to electricity in Burkina Faso: scaling-up renewable energy technologies. Environmental Research Letters 11, 084010. https://doi.org/10.1088/1748-9326/11/8/084010

Moner-Girona, M., Puig, D., Mulugetta, Y., Kougias, I., AbdulRahman, J., Szabó, S., 2018. Next generation interactive tool as a backbone for universal access to electricity. Wiley Interdisciplinary Reviews: Energy and Environment 7, e305. https://doi.org/10.1002/wene.305

Monteiro, C., Saraiva, J.T., Miranda, V., 1998. Evaluation of electrification alternatives in developing countries-the SOLARGIS tool, in: Electrotechnical Conference, 1998. MELECON 98., 9th Mediterranean. Presented at the Electrotechnical Conference, 1998. MELECON 98., 9th Mediterranean, pp. 1037–1041. https://doi.org/10.1109/MELCON.1998.699387

Moretti, L., Astolfi, M., Vergara, C., Macchi, E., Pérez-Arriaga, J.I., Manzolini, G., 2019. A design and dispatch optimization algorithm based on mixed integer linear programming for rural electrification. Applied Energy 233–234, 1104–1121. https://doi.org/10.1016/j.apenergy.2018.09.194

Morrissey, J., 2019. Achieving universal electricity access at the lowest cost: A comparison of published model results. Energy for Sustainable Development 53, 81–96. https://doi.org/10.1016/j.esd.2019.09.005

Müller, M.F., Thompson, S.E., Kelly, M.N., 2016. Bridging the information gap: A webGIS tool for rural electrification in data-scarce regions. Applied Energy 171, 277–286. https://doi.org/10.1016/j.apenergy.2016.03.052

Mwalenga, L., Amatya, R., González-García, A., Stoner, R., Pérez-Arriaga, J.I., 2016. A Comprehensive Computer-Aided Planning Approach for Universal Energy Access. Case study of Kaloleni, Kilifi County, Kenya. Enel Foundation Working Papers, Submitted.

myclimate, 2020. What are CO_2 equivalents? [WWW Document]. URL https://www.myclimate.org/information/faq/faq-detail/what-are-co2-equivalents (accessed 9.14.20).

National Institute of Statistics of Rwanda (NISR), Ministry of Finance and Economic Planning (MINECOFIN) [Rwanda], 2012. Fourth Rwanda Population and Housing Census [WWW Document]. URL http://www.statistics.gov.rw/publication/rphc4-population-projections (accessed 2.17.20).

Navarro, A., Rudnick, H., 2009. Large-Scale Distribution Planning — Part II : Macro-Optimization With Voronoi's Diagram And Tabu Search. IEEE Transactions on Power Systems 24, 752–758.

Nerini, F.F., Broad, O., Mentis, D., Welsch, M., Bazilian, M., Howells, M., 2016. A cost comparison of technology approaches for improving access to electricity services. Energy 95, 255–265. https://doi.org/10.1016/j.energy.2015.11.068

Neves, D., Silva, C.A., 2015. Optimal electricity dispatch on isolated mini-grids using a demand response strategy for thermal storage backup with genetic algorithms. Energy 82, 436–445. https://doi.org/10.1016/j.energy.2015.01.054

Ng, R.T., Jiawei Han, 2002. CLARANS: a method for clustering objects for spatial data mining. IEEE Transactions on Knowledge and Data Engineering 14, 1003–1016. https://doi.org/10.1109/TKDE.2002.1033770

NREL, 2017. PVWatts Calculator [WWW Document]. URL http://pvwatts.nrel.gov/ (accessed 3.2.17).

Ohiare, S., 2015. Expanding electricity access to all in Nigeria: a spatial planning and cost analysis. Energy, Sustainability and Society 5, 1–18. https://doi.org/10.1186/s13705-015-0037-9

Oladeji, O., 2018. Network Partitioning Algorithms for Electricity Consumer Clustering. Master thesis. Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science.

Page Arias, A., 2015. IntiGIS 2.0: Objetivos y propuesta metodológica. Análisis de competitividad tecnológica en Ghana. Master thesis (in Spanish). Universidad Complutense. Faculty of Geography and History, Madrid.

Pande, S., Ghodekar, D.J.G., 2012. Computation of Technical Power Loss of Feeders and Transformers in Distribution System using Load Factor and Load Loss Factor 3, 4.

Park, H.-S., Jun, C.-H., 2009. A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications 36, 3336–3341. https://doi.org/10.1016/j.eswa.2008.01.039

Parreno Jr, R.P., Del Mundo, R., 2015. Clustering Algorithm As A Planning Support Tool For Rural Electrification Optimization. International Journal Of Scientific & Technology Research 4, 112–120.

Parshall, L., Pillai, D., Mohan, S., Sanoh, A., Modi, V., 2009. National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. Energy Policy 37, 2395–2410. https://doi.org/10.1016/j.enpol.2009.01.021

Peco, J., 2001. Modelo de cobertura geográfica de una red de distribución de energía eléctrica. PhD thesis (in Spanish). Universidad Pontificia Comillas, Madrid. School of Engineering.

Pedhazur, E.J., 1997. Multiple regression in behavioral research: explanation and prediction, 3rd ed. Harcourt Brace College Publishers, Forth Worth.

Pinedo Pascua, I., 2012. Intigis: propuesta metodológica para la evaluación de alternativas de electrificación rural basada en SIG. PhD thesis (in Spanish). Universidad Politécnica de Madrid. School of Agricultural Engineering, Madrid.

Prada y Nogueira, I., 2017. Aerodynamic design optimization based on Multi-Attribute Structured Hybrid DirectSearch. Application to industrial problems. Ph.D. thesis (in Spanish). Universidad Pontificia Comillas, Madrid. School of Engineering.

Purchala, K., Meeus, L., Van Dommelen, D., Belmans, R., 2005. Usefulness of DC power flow for active power flow analysis, in: IEEE Power Engineering Society General Meeting, 2005. Presented at the IEEE Power Engineering Society General Meeting, 2005, IEEE, San Francisco, CA, USA, pp. 2457–2462. https://doi.org/10.1109/PES.2005.1489581

Qi, Y., Shi, D., Tylavsky, D., 2012. Impact of Assumptions on DC Power Flow Model Accuracy. Presented at the 2012 North American Power Symposium (NAPS), Champaign, IL, pp. 1–6. https://doi.org/10.1109/NAPS.2012.6336395

Rahman, M.M., Khan, M.M., Ullah, M.A., Zhang, X., Kumar, A., 2016. A hybrid renewable energy system for a North American off-grid community. Energy 97, 151–160. https://doi.org/10.1016/j.energy.2015.12.105

Razavi Zadegan, S.M., Mirzaie, M., Sadoughi, F., 2013. Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets. Knowledge-Based Systems 39, 133–143. https://doi.org/10.1016/j.knosys.2012.10.012

Romero Otero, L., 2016. Sistemas de información geográfica y electrificación rural. Análisis, desarrollo y estudio de caso con IntiGIS. Master thesis (in Spanish). Universidad Complutense. Faculty of Geography and History, Madrid.

Rwanda Energy Group (REG), 2019. The National Electrification Plan: Report on definition of technologies (On-grid and off-grid) at village level. URL http://www.reg.rw/fileadmin/user_upload/Report_of_the_Design_of_the_National_Electrification_Plan_in_Rwanda.pdf (accessed 1.24.20).

Santos Pérez, F.J., 2015. Metodología de ayuda a la decisión para la electrificación rural apropiada en países en vías de desarrollo. PhD thesis (in Spanish). Universidad Pontificia Comillas, Madrid. School of Engineering.

Schröder, B., 2019. Improvement of the Reference Electrification Model (REM) through wind power generation. Master thesis. Universidad Pontificia Comillas, Madrid. School of Engineering.

Sen, R., Bhattacharyya, S.C., 2014. Off-grid electricity generation with renewable energy technologies in India: An application of HOMER. Renewable Energy 62, 388–398. https://doi.org/10.1016/j.renene.2013.07.028

Sharma, N., Varun, Siddhartha, 2012. Stochastic techniques used for optimization in solar systems: A review. Renewable and Sustainable Energy Reviews 16, 1399–1411. https://doi.org/10.1016/j.rser.2011.11.019

Shrestha, A., Jha, S.K., Shah, B., Gautam, B.R., 2016. Optimal grid network for rural electrification of Upper Karnali Hydro Project affected area, in: 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). Presented at the 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, Agra, India, pp. 1–5. https://doi.org/10.1109/R10-HTC.2016.7906799

Sinha, S., Chandel, S.S., 2014. Review of software tools for hybrid renewable energy systems. Renewable and Sustainable Energy Reviews 32, 192–205. https://doi.org/10.1016/j.rser.2014.01.035

SMA Solar Technology AG, 2017a. Sunny Boy 5000TL datasheet [WWW Document]. URL http://files.sma.de/dl/15330/SB5000TL-21-DEN1551-V20web.pdf (accessed 3.2.17).

SMA Solar Technology AG, 2017b. Sunny Island 6.0H datasheet [WWW Document]. URL http://files.sma.de/dl/2485/SI_6H8H-AEN131411W.pdf (accessed 3.2.17).

SOFRECO, 2013. Electricity Access Scale-Up Roll-Out Program (No. 11.07.029/3324/11/EARP-CTX-DIR/YM/j.p.s.). Rwanda.

Soler, R., Thomas, F., Dhaiby, N.-E., Bakri, M., 2003. Optimizing the place of PV systems in rural electrification planning in Morocco, in: Photovoltaic Energy Conversion, 2003. Proceedings of 3rd World Conference On. IEEE, pp. 2574–2577.

Sothea, K., Kim Oanh, N.T., 2019. Characterization of emissions from diesel backup generators in Cambodia. Atmospheric Pollution Research 10, 345–354. https://doi.org/10.1016/j.apr.2018.09.001

Stenzel, P., Linssen, J., Fleer, J., Busch, F., 2016. Impact of temporal resolution of supply and demand profiles on the design of photovoltaic battery systems for increased self-consumption, in: 2016 IEEE International Energy Conference (ENERGYCON). Presented at the 2016 IEEE International Energy Conference (ENERGYCON), IEEE, Leuven, Belgium, pp. 1–6. https://doi.org/10.1109/ENERGYCON.2016.7514010

Steve Nolan, Scott Strachan, Puran Rakhra, Damien Frame, 2017. Optimized Network Planning of Mini-Grids for the Rural Electrification of Developing Countries. Presented at the IEEE PES-IAS PowerAfrica, Africa.

Sun, P.T., 2017. Solar kits to change the situation in Cameroon rural areas [WWW Document]. Plug The Sun. URL https://www.plugthesun.com/blog/rural-electrification-cameroon/ (accessed 3.30.18).

Syngellakis, K., Ribeiro, A., Arvidson, A., Nordström, M., Marcel, J.-C., 2008. ENABLE Project Summary Report. URL https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/enable_final_report.pdf (accessed 4.28.18).

Szabó, S., Bódis, K., Huld, T., Moner-Girona, M., 2011. Energy solutions in rural Africa: mapping electrification costs of distributed solar and diesel generation versus grid extension. Environmental Research Letters 6, 034002. https://doi.org/10.1088/1748-9326/6/3/034002

Szabó, S., Bódis, K., Huld, T., Moner-Girona, M., 2013. Sustainable energy planning: Leapfrogging the energy poverty gap in Africa. Renewable and Sustainable Energy Reviews 28, 500–509. https://doi.org/10.1016/j.rser.2013.08.044

Tawil, T.E., Charpentier, J.F., Benbouzid, M., 2018. Sizing and rough optimization of a hybrid renewable-based farm in a stand-alone marine context. Renewable Energy 115, 1134–1143. https://doi.org/10.1016/j.renene.2017.08.093

Triadó-Aymerich, J., Ferrer-Martí, L., García-Villoria, A., Pastor, R., 2016. MILP-based heuristics for the design of rural community electrification projects. Computers & Operations Research 71, 90–99. https://doi.org/10.1016/j.cor.2016.01.010

Trojan Battery Company, 2017. Trojan L16RE2V datasheet [WWW Document]. URL http://www.trojanbattery.com/pdf/datasheets/L16RE2V_TrojanRE_Data_Sheets.pdf (accessed 3.2.17).

Tsamasphyrou, P., Renaud, A., Carpentier, P., 2000. Transmission Network Planning Under Uncertainty with Benders Decomposition, in: Nguyen, V.H., Strodiot, J.-J., Tossings, P. (Eds.), Optimization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 457–472. https://doi.org/10.1007/978-3-642-57014-8_30

Tutelea, L., Boldea, I., 2010. Induction Motor Electromagnetic Design Optimization: Hooke Jeeves Method Versus Genetic Algorithms. Presented at the 12th International Conference on Optimization of Electrical and Electronic Equipment, Basov, pp. 485–492.

Upadhyay, S., Sharma, M.P., 2014. A review on configurations, control and sizing methodologies of hybrid energy systems. Renewable and Sustainable Energy Reviews 38, 47–63. https://doi.org/10.1016/j.rser.2014.05.057

Villanueva Saberbein, K., Aye, L., 2012. Technical and Financial Feasibility of a Stand-alone Photovoltaic System for Rural Electrification in the Andean South Region of Peru. Journal of Sustainable Development 5. https://doi.org/10.5539/jsd.v5n11p32

Weisstein, E.W., 2019. Order of Magnitude [WWW Document]. URL http://mathworld.wolfram.com/OrderofMagnitude.html (accessed 12.28.19).

World Bank, 2016. Achieving universal access in the Kaduna electric service area. URL http://documents.worldbank.org/curated/en/782491487138851242/Achieving-universal-access-in-the-Kaduna-electric-service-area (accessed 4.28.18).

World Bank, 2020. Access to electricity data [WWW Document]. URL https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS (accessed 3.3.20).

World Population Review, 2020. Africa Population [WWW Document]. URL http://worldpopulationreview.com/continents/africa-population/ (accessed 3.3.20).

Xiao-fei, L., Li-qun, S., 2016. Power system load forecasting by improved principal component analysis and neural network, in: High Voltage Engineering and Application (ICHVE), 2016 IEEE International Conference On. IEEE, pp. 1–4.

Yingying, L., Dongxiao, N., 2010. Application of Principal Component Regression Analysis in power load forecasting for medium and long term, in: Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference On. IEEE, pp. V3–201.

Zeyringer, M., Pachauri, S., Schmid, E., Schmidt, J., Worrell, E., Morawetz, U.B., 2015. Analyzing grid extension and stand-alone photovoltaic systems for the cost-effective electrification of Kenya. Energy for Sustainable Development 25, 75–86. https://doi.org/10.1016/j.esd.2015.01.003