University of Liège

Doctoral Thesis

Optimal design and deployment of isolated energy systems: The Bolivian pathway to 100% rural electrification

Author:
Sergio Luis Balderrama Subieta

Supervisor:
Sylvain Quoilin

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Thermodynamics Laboratory
Aerospace and Mechanical Engineering Department

April 24, 2022

Abstract

During the past decades, the planet has undergone increased environmental pressure. This has led to a clear momentum towards the creation of a more sustainable world. In this context, the challenge to provide energy access for all in a fair and sustainable way is an enormous task. The international agency of energy has estimated that yearly investments of 55 billion USD are needed to reach set targets. To optimize the limited resources, researchers have focused on the use of geographical information systems (GIS) to better capture the spatial dimension and define least-cost pathways to universal energy access. The size of the deployment problem imposes to model dispersed energy demands and isolated energy systems in a simplified manner, which can lead to suboptimal solutions. In consequence, there is a need to capture the diversity of conditions in which these systems are deployed.

The goal of this thesis is to contribute to the modeling of rural electrification processes through tailored models and methods. These tools are integrated into a coherent modeling framework, covering the whole value chain between accurate characterization of household demand to the macroscopic (national) planning of rural electrification. The models related to each relevant scale are soft-linked by defining common variables of interest. Then, methods to integrate the results of the more detailed models into the higher-level model are introduced. This approach provides additional technical insights and a better spatio-temporal optimum. The structure of this thesis reflects this bottom-up approach. It is organized in three parts.

The first part deals with energy demand modeling in a rural context and presents the Bolivian case study. It introduces two methods to create stochastic load profiles depending on the available data (measurements or surveys). In addition, it explores the components of an ideal rural community and frames appliance ownership according to surveys in rural communities. Finally, demand curves at the household and community level are generated using an ad-hoc stochastic bottom-up profile generation model.

The second part presents and applies an optimal sizing and operation framework for isolated energy systems in different contexts. The operational data from an existing microgrid in Bolivia is used as a benchmark and as a test case to test the model. Different sizing methods and formulations are compared, leading to the conclusion that a compromise must be found between system reliability and computational tractability. Finally, the trade-off between cost and lost load probability in single households equipped with a solar home system is analyzed.

The third part deals with the creation of surrogate models for microgrid design and its use in GIS-based electrification models. The limitations of existing GIS tools are discussed and surrogate models are proposed as a solution to increase accuracy without compromising the solving time of the model. A methodology to create and validate surrogate models for rural electrification is introduced. Then, the OnSSET model is adapted and improved to integrate this new formulation. Finally, different electrification scenarios are computed for the case of Bolivia, where both hybrid microgrids and solar home systems proved to be essential technologies for the cost-optimal electrification of remote communities.

Acknowledgements

These past years have been characterized by learning, discovery and challenges. This would not have been possible without the contributions of so many people that I am afraid that I would forget to mention someone. This said, Hereunder I would do a non-exhaustive list.

Certainly the biggest contributor to all the good things that have happened has been my thesis supervisor Sylvain Quoilin. There are no words that I can say to thank you for the support that you have given me. To Professor Vincent Lemort, it has been a pleasure to be part of the Lab and I thank you for your good will and the great atmosphere that you have created. To the rest of the PhD thesis jury, Prof. Pierre Dewallef, Prof Bertrand Cornelusses, Prof Emanuela Colombo and Prof Pedro Crespo del Grando, I thank you for accepting to revise and attend my PhD thesis defense.

To all the colleagues that I have met, I thank you for your support and discussions. Profs Torrico and Walter, thank you for your help and wise advice. Evelyn, Clau and Rober you made my stages in Bolivia full of interesting conversations and fun. Francesco, Nicolo and Gabriela thank you for your hard work, interesting ideas and help during the course of this work. Certainly, my PhD thesis will look completely different without your contributions.

To the members of the LaboThAp, These past years have been some of the best years of my life. I want to thank especially Sam, Nico, Thibaut and Remi for their friendship and support. Cami, thank you for allowing me to stay in your office, you are certainly the best. Javier, thank you for all the good times and the support to play games or do BBQ in random moments of the year.

In the extrange period COVID, certainly the people which you live makes the difference. To the people on Rue bassenge 13, thank you for everything. Oli, you are a great friend and always with amazing ideas and the will to do them. Nats, thank you for your friendship and kind words. Geoff, thank you for the advice and help in random tasks during the last year. Cendric, all my conversations with you have been more than interesting.

Certainly, the biggest asset for a successful PhD, it is friends. I am certain that I have one of the best collections in the world. Gabriel and Gabriela, your conversations and support either on the distance or face to face have been of big help during this time. Ioannis, thank you for all the help, you are one of the best persons that I have ever met. You are tied with Miguel, whom has help more than I can thank. Queralt, Meli, Andres, Lizzet, Elena, Fabrizia, Pep your friendship has light my day even in the darkness Liege times.

To my parents which have supported me during my whole life, in the good and especially in the bad moments. Andrea and Karen, your kind words has been the energy that has moved me in the most difficult times.

List of Publications

Articles in peer-reviewed academic journals

1. Navia, M.; Orellana, R.; Zarate, S.; Villazon, M.; **Balderrama, S.**; Quoilin, S. (2022) Energy Transition Planning with High Penetration of Variable Renewable Energy in Developing Countries: The Case of the Bolivian Interconnected Power System. Energies, 15, 968.
2. **Balderrama Subieta, S**. L., Lombardi, F., Stevanato, N., Gabriela, P., Colombo, E., & Quoilin, S. (2021). Surrogate models for rural energy planning: Application to Bolivian lowlands isolated communities. Energy, 232.
3. Fernandez, M., Cardozo, E., Zambrana, J., Pena, G., **Balderrama Subieta, S.** L., Sanchez, C., Soto, S., & Quoilin, S. (2021, June 01). Evaluacion del costo de electrificacion rural en Bolivia para alcanzar el ODS 7. Journal Boliviano de Ciencias.
4. Balderrama, G., **Balderrama Subieta, S. L.**, Lombardi, F., Stevanato, N., Sahlberg, A., Howells, M., Colombo, E., & Quoilin, S. (2020). Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET). Energy for Sustainable Development, 56, 98 - 118.
5. Stevanato, N., Rinaldi, L., Pistolese, S., **Balderrama Subieta, S. L.**, Quoilin, S., & Colombo, E. (2020, October). Modeling of a Village-Scale Multi-Energy System for the Integrated Supply of Electric and Thermal Energy. Applied Sciences.
6. Stevanato, N., Lombardi, F., Guidicini, G., Rinaldi, L., **Balderrama Subieta, S. L.**, Pavicevic, M., Quoilin, S., & Colombo, E. (2020, July). Long-term sizing of rural microgrids: Accounting for load evolution through multi-step investment plan and stochastic optimization. Energy for Sustainable Development.
7. **Balderrama Subieta, S. L.**, Lombardi, F., Riva, F., Canedo, W., Colombo, E., & Quoilin, S. (2019). A two-stage linear programming optimization framework for isolated hybrid microgrids in a rural context: The case study of the "El Espino" community. Energy, 188.
8. Rojas Candia, R., **Balderrama Subieta, S. L.**, Adhemar Araoz Ramos, J., Vicente Senosiain, M., Pena Balderrama, G., Jaldin Florero, H., & Quoilin, S. (2019). Techno-economic assessment of high variable renewable energy penetration in the Bolivian interconnected electric system. International Journal of Sustainable Energy Planning and Management, 22.
9. Francesco, L., **Balderrama Subieta, S. L.**, Quoilin, S., & Emanuela, C. (2019). Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. Energy, 177, 433-444.

Papers at international scientific conferences, published in full proceedings

1. Soto, A., **Balderrama Subieta, S. L.**, Cardozo, E., Fernandez, M., Zambrana, J., & Quoilin, S. (2021). Exploring the Tradeoff between Installed Capacity andUnserved Energy in Rural Electrification. PROCEEDINGS OF ECOS 2021.

2. Zarate, S., Villazon, M., Navia, M., **Balderrama Subieta, S. L.**, & Quoilin, S. (2021). Modeling hydropower to assess its contribution to flexibility services in the Bolivian power system. Proceedings of the 16th SDEWES Conference.

3. **Balderrama Subieta, S. L.**, Pena Blderrama, G. J., Lombardi, F., Stevanato, N., Sahlberg, A., Colombo, E., & Quoilin, S. (2019). Model-Based cost evaluation of Microgrids systems for rural Electrification and energy planning purposes. PROCEEDINGS OF Solar World Congress 2019.

4. **Balderrama Subieta, S. L.**, Lombardi, F., Stevanato, N., Pena, G., Colombo, E., & Quoilin, S. (2019). Automated evaluation of levelized cost of energy of isolated micro-grids for energy planning purposes in developing countries. PROCEEDINGS OF ECOS 2019.

5. Lombardi, F., **Balderrama Subieta, S. L.**, Nicolo, S., Pistolese, S., Colombo, E., & Quoilin, S. (2018). Modelling of a village-scale Multi-Energy System (MES) for the integrated supply of electric and thermal energy. Proceedings of SSB 2018, 10th International Conference on System Simulation in Buildings.

6. **Balderrama Subieta, S. L.**, Haderspock, F., Canedo, W., Renan, O., & Quoilin, S. (2018). Techno-economic evaluation of rural electrification in Bolivia: lessons learned from the "El Espino" micro-grid. Proceedings of ECOS 2018 - the 31th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems.

7. Antonio Rojas Candia, R., Adhemar Araoz Ramos, J., **Luis Balderrama, S.**, Pena Balderrama, G., Alejo Espinoza, L., Senosiain, V., & Quoilin, S. (2018). Techno-economic assessment of high renewable energy source penetration in the Bolivian interconnected electric system. Proceedings of the 31st ECOS Conference.

8. **Balderrama Subieta, S. L.**, Canedo, W., Lemort, V., & Quoilin, S. (2017). Impact of Diesel generator limitations in the robust sizing of isolated hybrid Microgrids including PV and batteries. Proceedings of ECOS 2016 - the 30th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems.

9. **Balderrama, S.**, Canedo, W., Lemort, V., & Quoilin, S. (2016). Techno-economic optimization of isolate micro-grids including PV and Li-Ion Batteries in the Bolivian context. Proceedings of ECOS 2016 - the 29th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems.

Contents

List of Figures

List of Tables

Introduction and contributions

Chapter 1

Introduction

1.1 Context

During the past decades, the planet has undergone increased environmental pressure, which has led to more awareness and a clear momentum towards a more sustainable world for all. This momentum has been translated, for example, into a worldwide agreement to mitigate and adapt to the impacts of climate change [1]. At the same time, it is acknowledged that reduction of poverty, inequality and wildlife conservation are also key issues to be tackled. To that aim, 17 different sustainable development goals (SDGs) have been defined. These objectives range from gender equality to the appropriate use of terrestrial ecosystems. In particular, SDG 7 is defined as "Ensure access to affordable, reliable, sustainable and modern energy for all". This objective involves that, in the next 10 years, the majority of the 800 million (10 % of the total world population) without access to electricity and the 2 billion people without access to clean fuels (e.g. for cooking) will have to gain access to them [2].

SDG 7 must be achieved in such a way that the share of renewable energy is increased, energy efficiency is improved and the final prices for the consumer are maintained accessible. All these points have several implications at the economical, political, environmental and social levels, and the fulfillment of these ambitious targets will involve significant planning efforts and project implementation capacities in the years to come.

The challenge to provide energy access to all in a fair and sustainable way is an enormous task. The international agency of energy (IEA) has estimated that yearly investments of 55 billion USD each year are needed to reach the set targets [3]. This endeavor becomes more challenging when the locations of these people are taken into account. As shown in Figure 1.1, most people lacking access to modern forms of energy are located in rural locations of developing countries. This means that the information, the available economical resources, the infrastructure, etc. are substantially reduced compared with urban zones in developed economies.

Proper energy planning tools are required to achieve the above challenges. In particular, the planning activities for rural electrification in a specific territory involve locating underserved regions or villages, analyzing their expected demand, calculating the cost of different electrification technologies, or simulating the most suitable deployment pathways (Figure 1.2).

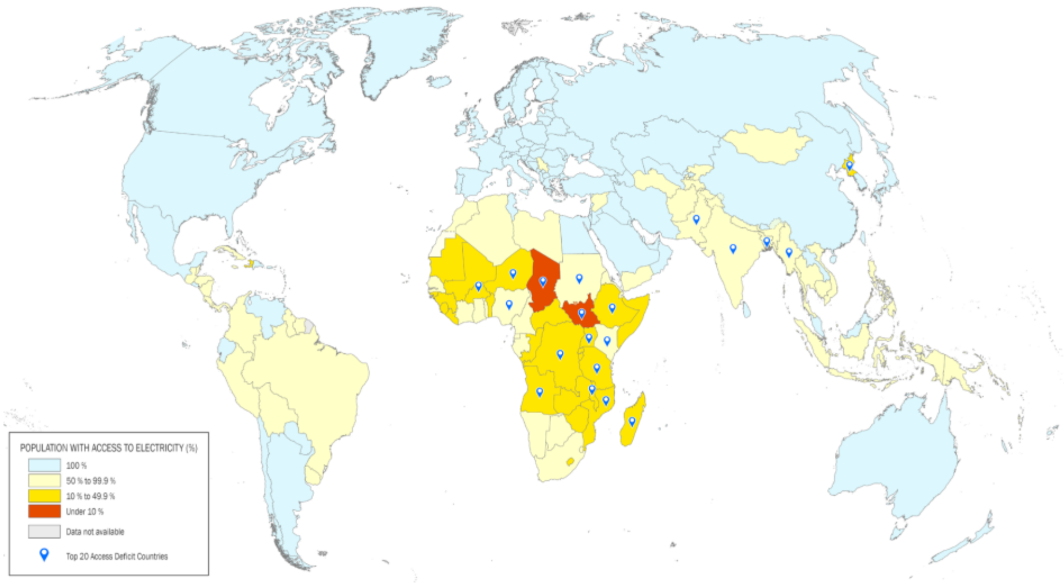

FIGURE 1.1: Electrification rate per country, taken from [4].

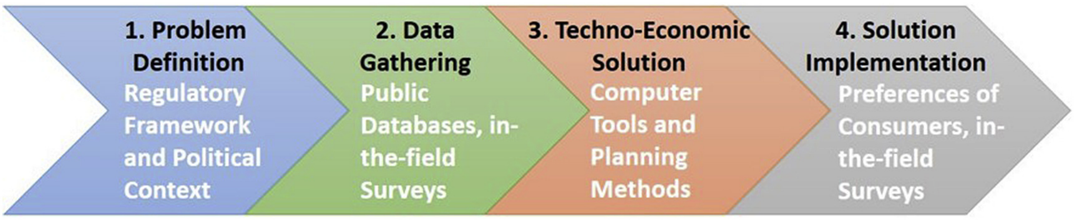

FIGURE 1.2: Electrification planning process, taken from [5].

1.2 Model-based energy planning in developing countries

Energy planning is a complex decision problem, highly relevant in both developed and developing countries. In industrialized countries, investment decisions have typically been informed using quantitative planning models, and researchers have used modeling to explore policy questions for decades. The scientific literature includes a rich collection of models that address a variety of energy policy concerns for developed countries. Modeling tools for Energy systems are generally distinguished in terms of their sector of interest, spatio-temporal resolution, level of technical detail or simulation horizon. Despite the wide variety of modeling approaches, two categories have emerged for energy systems modeling at the country level (a more comprehensive review is proposed in Collins et al [6]:

- Operational power system models, which optimize the operation of the power system but do not consider investment. These are typically unit commitment and optimal dispatch (UCED) models, which describe the constraints of the power system with high level of accuracy and have the capacity to model rapid variations in renewable generation, forecast errors or reserve markets.
- Long-term planning energy system optimization models (ESOM), which are used to generate scenarios for the long-term evolution of the energy system. These models include investments and optimize the system over multiple years.

They are generally not restricted to the power sector (i.e. they have an endogenous representation of all the sectors in the energy systems) and can cover large geographical areas.

Both types of models are optimization models, often based on linear programming (LP) or mixed-integer linear programming (MILP). These models are well adapted to industrialized countries, usually characterized by high electrification rates, modern forms of energy and high consumption densities. Traditional modeling tools however present significant limitations for energy planning in developing countries. In Figure 1.3, Lenai (2013) summarizes the specific characteristics of these regions, insufficiently captured by state-of-the-art optimization models [7].

rural - urban divide	decentralized supply options
reliance on traditional energy (biomass, firewood)	prevalence of inequity and poverty
informal sector activities (barter, in-kind payments)	technological change
technology diversity (ability to leapfrog)	technology diffusion
transition to modern energy (increased consumption pattern and rising energy intensity due to modernization, urbanization, employment demand)	sector reform/structural change and competition in emerging liberalized markets
spatial difference and divergence in consumption/ disaggregated demand by income and location	environmental implications of energy use (sustainability)
low data availability for modeling	long-term uncertainties
economic growth and corresponding energy implications	demand-side options
energy shortage/poor performance of utilities	financial status of utilities
low energy access and rates of electrification	resource depletion
institutional issues like corruption	

FIGURE 1.3: Features of developing countries not commonly included in energy models, taken from [7]

In this thesis, two important aspects are identified as key limitations of state-of-the-art energy planning tools in developing countries:

1. **The spatial dimension**. Addressing decentralized capacity expansion such as microgrids or solar home systems is a challenge for start-of-the art power system models. The models are typically designed for a well interconnected system and assume 100% electrification rate, which is not representative of the situation in many developing countries. Previous works demonstrate that grid expansion is not necessarily the least-cost solution, especially in countries characterized by a low consumption density and unfavorable geographical conditions [8]. In order to address these issues, models with a high level of spatial disaggregation (e.g. GIS-based models) are required. OnSSET is an example of such a model, but cannot address the overall problem since no detailed consideration of the power system or of the long-term trends are considered [9].

2. **Multidisciplinarity**. The representation of cross-sectoral and cross-disciplinary interactions is insufficiently addressed in current energy models. In this thesis, "sectors" are understood in a broad sense, i.e. not limited to energy-only sectors. Sector coupling thereby encompasses interactions between, for example, the power sector and the water sector (water-energy nexus) or interaction between the bioenergy sector and the food sector (food-energy nexus). These interactions are particularly relevant in the case of developing countries, potentially characterized by water scarcity, resource depletion, deforestation or local pollution [10]. The main difficulty when linking models originating from

different research fields is the harmonization between model formulations and input datasets. New methods therefore need to be devised to integrate e.g. social sciences aspects into energy models. In a previous work [8], the partners of this project proposed a tool to translate qualitative survey data into quantitative demand levels and curves. The data was however only applied to specific cases and was too partial to be considered into large-scale energy planning models. Translating modeling outputs into socially-tailored messages is another challenge and requires close links with education, governance, social participation, indigenous autonomous management, and others.

As an answer to these limitations, researchers have focused on the use of geographical information systems (GIS) to better capture the spatial dimension and define least-cost pathways to universal energy access [5]. The analysis coming from these tools have shown the potential of isolated energy systems for rural electrification [8, 9]. In these approaches, the size of the optimization problem imposes to consider dispersed energy demands and isolated energy systems in a simplified manner (e.g. a single system configuration for the whole country) [11]. This can lead to sub-optimal solutions and more work is therefore needed to capture the diversity of the technical, geographical, socio-cultural, demographic conditions in which these systems are deployed [8].

1.3 Aim and objectives

Under the circumstances described in previous sections, the following research question arise:

How to capture the diversity of the technical, geographical, socio-cultural, demographic conditions of rural communities and include them into GIS-based electrification tools without impacting the computational tractability of the problem?

This thesis suggests to tackles this challenge by including an intermediate step between the sizing process of microgrids and the use of rural electrification GIS tools (Figure 1.4). This step resorts to the creation and validation of surrogate models, able to estimate the optimal design parameters of off-grid electrification systems.

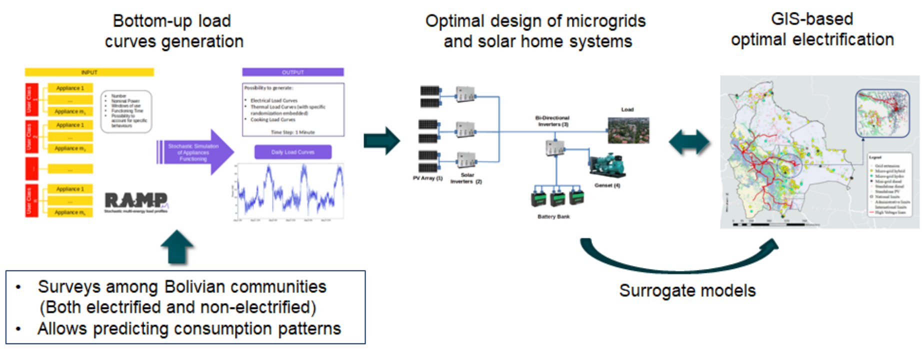

FIGURE 1.4: Main contribution of this work

To that aim, different contributions to the modeling of rural electrification processes are proposed, relying on different specialized models and tools. Each of them captures a relevant scale of the problem. These tools are integrated into a coherent modeling framework, covering the whole value chain between accurate characterization of household energy demand to the macroscopic (national) planning of rural electrification.

To that aim, it is first necessary to develop and calibrate relevant models for each considered level of aggregation. This includes:

- The bottom-up modeling of demand curves at the household and community levels.
- The optimization of isolated power generation systems adapted to the local context, including individual solar home system and microgrids.
- The generation of optimal rural electrification pathways at the country level through the deployment of the central grid and of decentralized technologies.

The models related to each relevant scale should then be soft-linked by carefully defining common variables of interest, and by devising methods to integrate the results of more detailed models into higher-level models, thus providing additional insights and a better spatio-temporal optimum.

1.4 Novelties & contribution

The contributions of this thesis relate to the development and parametrization of the individual models, but also to their integration into a coherent modeling framework. At the household level, a methodology to translate survey and interview data into highly detailed stochastic demand time series is proposed, together with proper aggregation methods to model energy demands at the community level. This demand data is then used as input for a stochastic microgrid sizing and operation tool. This model is parametrized and tested using the monitoring data of a real microgrid in Bolivia. It is then run for a wide range of possible settings and machine learning methods are used to derive surrogate models predicting the cost and the optimal configuration of the system. These surrogate models are finally integrated into a GIS model to define optimal electrification pathways at the country level. These specific contributions are summarized in Table 1.1.

A special emphasis has been put on the flexibility and adaptability of the presented contribution. This is a key feature since the high uncertainty associated with the energy resources, demands, and local constraints might require slight adjustments in the methodology in such a way that it can be reused in other contexts or countries.

1.5 The importance of open data and software

At the present time, most state-of-the art energy system models are subject to licensing or are not publicly released. This can constitute a barrier for practitioners to get access to suitable design tools for microgrids and rural electrification in general, especially in the Global South. At the same time, in the past years, the open-source community has grown significantly in importance in the energy field, and the free and open models are now competing with the traditional closed-source commercial software. A recent review of existing tools concludes that: "The result shows that available open source tools (...) are mature enough based on a function comparison with commercial or proprietary tools for serious use." [12].

TABLE 1.1: Summary of novelties and where they can be found in this dissertation.

Novelty	Discussed in
A method to translate survey and interview data into highly detailed stochastic demand time series	Chapters 2, and 3.
An optimization framework to size and control microgrids under uncertainty	Chapter 4, 5, and 6.
An analytic method to account for the trade-off between lost load probability and cost	Chapter 6
A method to create surrogate models for different microgrid design variables	Chapter 7 and 8
The implementation of surrogate models into GIS-electrification tools	Chapter 8 and 9
An optimal rural electrification pathway for the specific case of Bolivia	Chapter 9

For the above reasons, all the developments, methods and models produced in this PhD thesis are released as open-source together with ad-hoc online documentation and the needed data to reproduce the results. They can therefore be freely downloaded, re-used, adapted or modified. This open-science approach is also selected to increase the transparency and reproducibility of the proposed methods [13]. GitHub is used by default to host source code, more information can be found in annex A.

1.6 Organization of the thesis

The thesis is organized into three main parts:

- The first part deals with energy demand modeling in a rural context and introduces the Bolivian case study. Chapter 2 discusses different demand modeling methodologies. It introduces two methods to create stochastic load profiles depending on the available data: measurements or surveys. In addition to these, it explores the components of an ideal rural community and frames appliance ownership according to the energy sufficiency concept. Chapter 3 describes the energy reality of Bolivia, divides the country into representative regions and creates demand profiles for each of them.
- The second part presents and applies a new optimal sizing and operation framework. In chapter 4, the mathematical model is presented and its peculiarities are explored. Chapter 5 analyses data from an existing microgrid in Bolivia and uses the sizing model in different contexts. Conclusions on the advantages and disadvantages of applying different models are drawn in this chapter. Finally, chapter 6 explores the trade-off between cost and energy not served for a particular household using a Solar home system (SHS).
- The third part deals with the creation of surrogate models for microgrid design and its use in GIS-electrification base models. In chapter 7, GIS-electrification tools are presented, their limitations discussed and surrogate models are proposed as solutions to increase accuracy without compromising the solving

time. A methodology to create and validate such surrogate models is presented in chapter 8. The OnSSET model is adapted and improved in chapter 9 to account for the individual design parameters of isolated systems.

Part I

Rural energy demand

Chapter 2

Rural energy demand characteristics and modeling methods

This chapter is largely based on contributions from the two following publications:

Lombardi, F., ***Balderrama, S.****, Quoilin, S., & Colombo, E. (2019). Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. Energy, 177, 433-444.*

Sanchez, C., ***Balderrama, S.****, Stevanato, N., ,Quoilin, S., (2021). Energy Sufficiency for Rural Communities: The Case of The Bolivian Lowlands. Proceedings of the 16th SDEWES Conference 2022, Dubrovnik.*

2.1 Introduction

As acknowledged in the Agenda 2030 of the United Nations, sustainable development is linked to the simultaneous accomplishment of interlinked objectives with synergies between them. From all the Sustainable Development Goals (SDGs), access to energy (SDG goal 7) [14] is particularly intertwined with other dimensions of socio-economic development. Studies show a link between poverty reduction and energy access [15]; likewise, electricity access is interconnected with income generating activities, household economy, health, education, habits, and social networks [16]. The combination of electricity and heating systems can lead to higher energy independence of rural communities in diverse contexts [17, 18]. Furthermore, the link between energy, water and food is particularly relevant and is increasingly explored in the scientific literature [19, 20].

To foster these synergies, high levels of energy access and thus of energy generation capability are needed. In rural settings, these conditions are rarely met, partly due to the associated uncertainty in electricity demand evolution. In an Indian village case study, Riva et al. [21] show that different scenarios of long-term demand can considerably impact the final design of the off-grid power system. Inaccurate demand projections can lead to unexpected situations, as showcased by Ulsrud et al.

[22]. They report how the consumers of a solar mini-grid started to draw more electricity than initially forecast. A similar case is reported by Zhao et al. in a case-study on the design and development of a real microgrid system in Dongfushan Island, where a significant error in the prediction of energy consumption arose due to intensive use of air conditioners [23]. Diaz et al. [24] report examples of microgrids that were initially designed with PV only, but had to be changed due to higher energy consumption than forecast during nights. In the same study, a generator had to be added to a hydro-diesel system due to the new connection of a village. Also, Kobayakawa and Kandpal [25] analyzed microgrids struggling to meet the demand due to unplanned connections of additional users to the system. Riva et al. [26] stress the need to introduce an appropriate modeling framework for assessing long-term projections of electricity demand within rural energy planning. In [16], the nexus between evolution of electricity demand and local rural development is conceptualized, suggesting system dynamics as an adequate approach to investigate this issue from a quantitative point of view.

This mismatch between generation and demand has different origins, as highlighted in [27], and there are several techniques to tackle this issue. In general, decision-making with respect to renewable energy systems has been driven by two main factors: technology development and investment [28]. Model-based scenarios from a purely techno-economic point of view have thus received much attention recently, to identify pathways towards achieving decarbonisation targets. However, there is a "lack of attention to the actors, their decisions, interactions and learning processes, and how these shape twisted transition pathways" [29].

As a first step to ensure electricity coverage in the rural communities of the global south, a minimum energy access must be settled. To that aim, the concept of energy sufficiency is introduced. In [30], the authors propose that energy sufficiency is a state in which people's basic needs for energy services are met equitably and ecological limits are respected. The focus is on services that meet needs for shelter, health, work, mobility and communication. As noted above, these needs vary according to local conditions. For example, concepts in health, shelter, mobility and work are being rethought along "sufficiency" lines and tested in different contexts. Ideally, unelectrified communities should move from the current low energy consumption state to a position where energy consumption is enough to ensure a continuous development without jeopardizing the environment they live in.

In order to evaluate these aspects from a quantitative standpoint and to include them into the proposed modeling framework of this thesis, the following specific objectives are proposed:

1. To describe available demand modeling methods.
2. To define an ideal rural community composition.
3. To discuss possible appliance ownership of each member of a rural community.

2.2 Modeling the electric demands

As previously detailed, energy modeling for unelectrified rural communities is a hard task due to the uncertain demand evolution pathways in these settings. Mandelli et al [32] define energy consumption modeling as: "the domain of models able

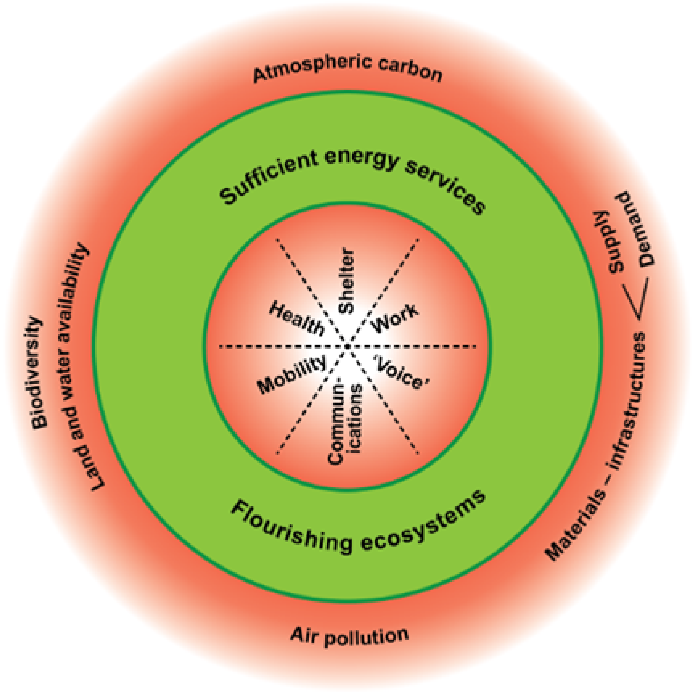

FIGURE 2.1: Graphical description of energy sufficiency [31].

to support energy-related policy decisions.". According to Swan et Ugursal, it can be divided in two categories (Figure 2.2):

1. Top down approaches: These models focus time series analyses of aggregated demand, they explore different variables such as gross domestic product per capita, poverty level, ethnicity and their impact on the demand across the years. Another application of these models is the creation of stochastic scenarios based on statistical methods using measurements of the analyzed systems.

2. Bottom up approaches: These models focus on the analysis of the individual components of the system. The data used can have different levels of granularity, ranging from the analysis of the different sectors to the usage patterns of all appliances in a household.

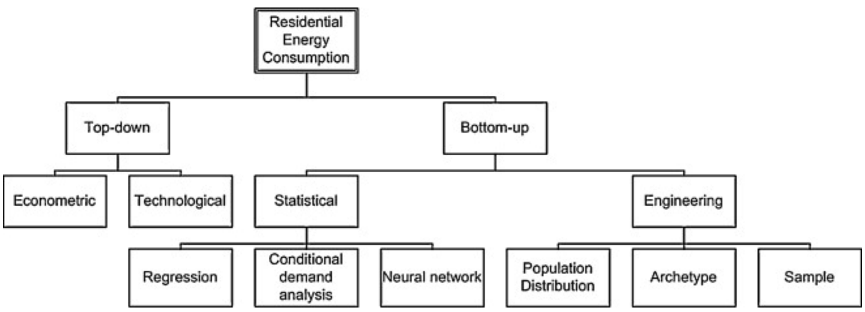

FIGURE 2.2: Energy modeling techniques, taken from [33].

Both approaches are valid depending on the needs and limitations encountered by the practitioners in their specific projects. It is noteworthy that hybrid methods combining both approaches are also possible, as explored by Lombardi et al [34]. In that work, stochastic bottom-up generated loads were combined to the top-down Italian electricity demand to explore the impact of a transition from gas stoves to electrical stoves on the energy mix.

In the present work, both approaches are implemented and tested depending on the considered case study and on the availability of input data. The details of both methods are detailed in the next sections.

2.2.1 Top-down stochastic demand energy modeling

Future scenarios of energy demand time series can usefully be deducted from historical monitoring data. However, in many situations, historical data cannot account for the uncertainty and the stochasticity of the load, and is therefore not suitable for probabilistic models. In this thesis, a new methodology is proposed, in which historical monitoring data is used to calibrate a stochastic model of the load variations around its averaged daily profile.

The goal is to generate stochastic demand curves which are different from historical data, but present similar characteristics in terms of peak load, load duration curve and variability. Such a demand curve generator can the be used to define a high number of realistic time series to be used in energy system optimization tools. The proposed methodology consists of the following successive steps:

1. Average the historical data into average daily profiles for each month and for each household.
2. Compute the logarithmic error between the data and the averaged values.
3. Generate stochastic time series calibrated with the characteristic of the log-normal noise.
4. Apply this stochastic noise to the averages historical profiles to generate realistic yearly time series.

A log-normal distribution of the noise is selected because its skewness matches well with that of the error between the load and the average curve in the data. The logarithmic error is computed by:

$$LE = \log\left(\frac{Load_{hist}}{Load_{mean}}\right) \tag{2.1}$$

The purpose is to generate stochastic times series whose characteristics are close to that of this monitored noise. One of the main characteristics to be conserved among all the generated time series is the maximum load throughout the year because it conditions the peak installed capacity of the power generation system.

To that aim, an algorithm adapted from the "Iterated Amplitude Adjusted Fourier Transform (IAAFT)" [35, 36] has been implemented. This algorithm generates samples of a random process conforming to given auto-covariance and probability density function. The advantage of this approach is that it gives more insight into the underlying process, it can take as input only the marginals and autocorrelations, and it can generate time series of any length [37].

As a first step, a random realization of a given probability distribution function (PDF) is created at the desired temporal sampling interval x_0. Then an iterative process starts, in which the realization is shuffled in order to match the given (two-sided) power spectral density (PSD) (S_{xx}). The PDF is not affected by the temporal reordering. In every iteration the Fourier amplitudes of S_{xx} are compared with the ones of x_0. This procedure is summarized below [35]:

1. The phases of x_0 are calculated.
2. A new signal x is created with the same phases but with the amplitudes of S_{xx} via the IAAFT transformation.
3. x lost its marginal information (became gaussian). For that reason, a zero mean nonlinear transformation is applied. In this procedure the values of the signal with the correct distribution, xo, are shuffled to match the rank of x.
4. The act of shuffling the x alters the Fourier amplitudes and hence the PSD. The above steps are iterated until the final step matches the rank from the previous iteration.

The final result of this methodology is shown in Figure 2.3, where a base demand scenario of 1 year was used to create 20 different scenarios. As it can been seen, they maintain the principal characteristics of the main load (shape), at the same time as exhibiting a stochastic behavior around it.

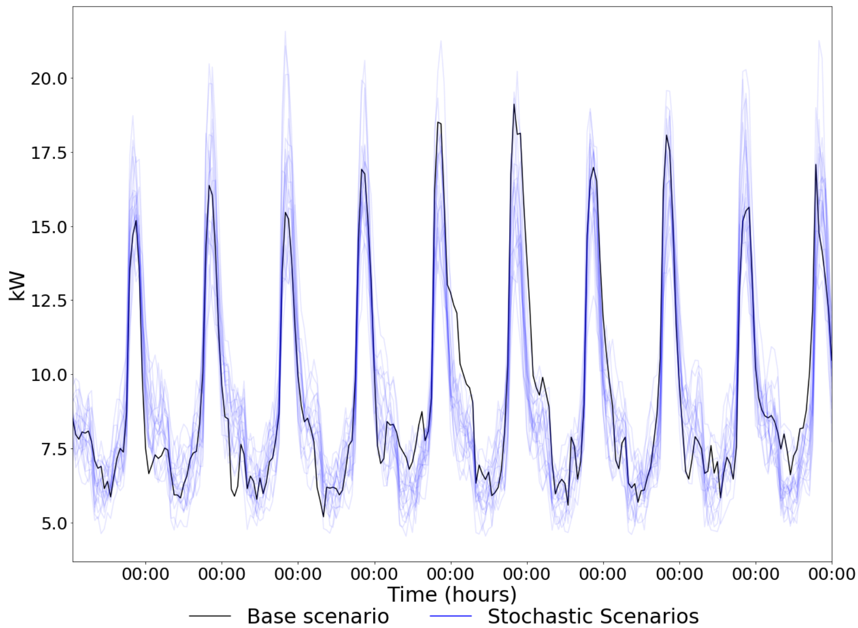

FIGURE 2.3: Demand from a base scenario with 20 synthetic profiles.

2.2.2 Bottom-up demand modeling

The second approach for demand prediction implemented in this thesis can be referred to as bottom-up. It is well adapted to situations in which no historical load data is available (for example in the case of a non-yet electrified community), but in which a detailed description of the users and of their appliances can be built.

In this work, the RAMP model is selected to generate stochastic bottom-up load profiles [38] and the procedure proposed in Stevanato et al. [39] is followed for its application in the context of a remote community. The RAMP model is based on the definition of several User Classes, each of which is associated with a set of appliances. Each appliance (e.g. TVs, lights bulbs, phone chargers) is defined by a nominal absorbed power, a total functioning time along the day, and possible time frames of use, in addition to some further optional features. Based on this information, which is subject to stochastic variation between pre-defined ranges to account for uncertainty and random users' behaviour, the model allows computing the total load curve of a village (Figure 2.4).

The advantage of this approach is the possibility to create synthetic village demand curves in a bottom-up manner from limited information. The required data is obtained through surveys and interviews within the community members and the identification of possible services providers and of income generating activities. At this point, a set of plausible scenarios can be generated stochastically. Non-existing behaviours or appliances can also be introduced to the model for future scenarios to explore the impact that future changes of the load curves would entail on the sizing of energy systems [8].

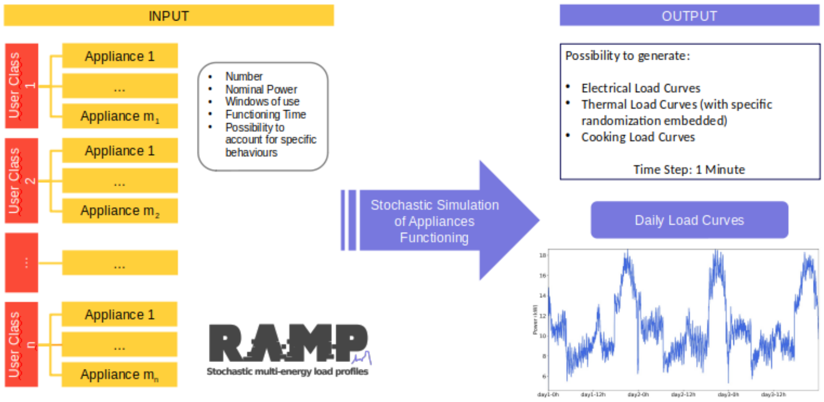

FIGURE 2.4: RAMP model logic.

From a conceptual point of view, RAMP is based on three main layers of modelling, namely: i) the User type; ii) the User; and iii) the Appliance layers (Figure 2.5). The higher layer consists in the definition of a set of arbitrary User types (e.g. Household, Commercial activities, Public offices, Hospitals, etc.), whose level of discretisation depends on the modeler's needs; for instance, when more precise information is available, a "Household" User type may be further subdivided by income classes or building type. Each user type is subsequently characterized in terms of the number of individual Users associated to that category (second layer) and in terms of Appliances owned by each of those users (third layer). As shown in Figure 2.5, the three-layer structure allows to independently model the behavior of each j_{ik}-th Appliance, so that each individual j_i-th User within a given i-th User type will have a unique an independent load profile compared to the other Users of the same type. The aggregation of all independent User profiles ultimately results in a total load profile, which is uniquely generated at each model run. Multiple model runs generate different total load profiles, reproducing the inherent randomness and unpredictability of users' behavior and allowing to obtain a series of different daily profiles.

All the inputs required to run the model are summarised in Table 2.1, and consist of information that can be obtained from common field surveys, in analogy with and expanding those defined by Mandelli et al. [32].

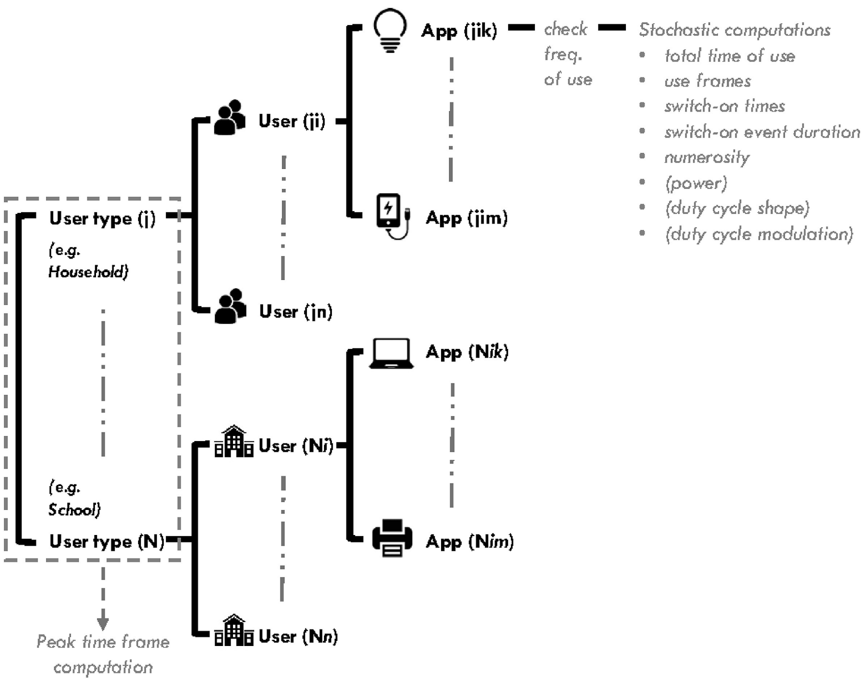

FIGURE 2.5: Energy modeling techniques, taken from [38].

Core stochastic algorithm

From a mathematical point of view, the stochastic algorithm that constitutes the core of the model (without including the Appliances' optional attributes mentioned in Table 2.1) is articulated in the following steps:

1. Identify the expected peak time frame;
2. For each $Usertype_j$, for each $Users_{ij}$ and for each $Appliance_{jik}$, check if the appliance is used based on the weekly frequency of use ($frequency_{jik}$). If not, ignore the appliance; otherwise, compute:
   (a) The randomized total time of use $TotUse_{jik}$.
   (b) The randomized vector of time frames in which the appliance can be on $UseFrames_{jik}$.
   (c) Compute a random switch-on time (with random switch-on even duration $tmin_{jik}$) within the available use frames.
   (d) Compute the randomized power required by the appliance for the switch-on event under consideration P_{ijk}.
   (e) Compute the actual power absorbed by $Appliance_{jik}$ during the switch-on event considering a random numerosity in the range: 0, m_{jik}.

Repeat steps 2.c to 2.e until the sum of the durations of all the switch-on events equals the randomised $TotUse_{jik}$;

3. Aggregate all profiles in a total load profile.

The identification of a peak time frame allows differentiating between off- and on-peak switch-on events, which are associated with different probability distributions for the computation of the random numerosity. To this end, a theoretical peak time

TABLE 2.1: Summary of the input data required by the RAMP model.

User type and Users	
$Usertype_j$	Name of the User type (e.g. "households", "commercial activities", etc.)
n	Number of $Users_{ij}$ (for i = 1 : n) within $Usertype_j$
Appliances	
$Appliance_{jik}$	Name of the k-th Appliance associated with the j-th User type and the i-th User
m_{jik}	Numerosity of $Appliance_{jik}$ (e.g. numerosity of "indoor light bulbs")
P_{jik} (W)	Power absorbed by a single item of $Appliance_{jik}$ (i.e. assuming numerosity = 1)
$TotUse_{jik}$ (min)	Total time of use of the $Appliance_{jik}$ in a day
$tmin_{jik}$ (min)	Minimum time that the $Appliance_{jik}$ is kept on after a switch-on event
$\delta_{tmin,jik}$ (%)	Percentage random variability applied to $tmin_{jik}$
$UseFrames_{jik}$	Time frames in which a random switch-on of $Appliance_{jik}$ can occur
$\delta_{frames,jik}$ (%)	Percentage random variability applied to use $frames_{jik}$
Appliances' optional attributes	
$cycle_{jik}$	Duty cycle of $appliance_{jik}$ (up to 3 per appliance)
$\delta_{frames,jik}$ (%)	Percentage random variability applied to the duration of the segments composing $cycle_{jik}$
$CycleMod_{jik}$	Association between time frames and different duty cycles
$frequency_{jik}$ (%)	Weekly frequency of use of $Appliance_{jik}$ (<100% for "occasional-use" appliances)
$FixedNum_{jik}$	Constraint for all the m_{jik} appliances to always switch-on simultaneously
$\delta_{p,thermal}$ (%)	Porcentage random variability applied to P_{jik}, conceived for thermal appliances

frame is identified, as proposed in [32], as the time frame associated with the maximum load in a virtual total load profile resulting from the fictitious assumption that each Appliance is always switched-on with maximum power and numerosity during all of its potential time frames of use. Within such theoretical peak time frame, a unique peak time (t_{peak}) is hence randomly sampled with uniform distribution. Finally, an actual expected peak time frame is defined in equation 2.2.

$$peaktimeframe = [t_{peak} - k, t_{peak} + k] \tag{2.2}$$

where k is the product of a random sampling with normal distribution around t_{peak} and standard deviation equal t_{peak}, δ_{peak}. By default, δ_{peak} is set to 15 % of t_{peak}, but it represents a potential calibration parameter that allows to modulate the extension of the peak time frame and may serve to simulate, for instance, a different social behavior during holidays or weekends.

Peak-load periods correspond to periods in which a large share of *Users* is interested by intensive activity patterns and when, consequently, they are more likely to switch-on multiple Appliances of the same kind (e.g. "Households" might be likely to switch-on multiple indoor lights simultaneously in the evening, when they also cook, watch TV, etc.). To this regard, the model acts on the modulation of the "coincident numerosity" factor, defined by Equation 2.3.

$$f = \frac{M_{ON,ijk}}{M_{ijk}} \tag{2.3}$$

where $m_{ON;ijk}$ represents the numerosity of appliances that are simultaneously switched on during a switch-on event related to $Appliance_{ijk}$. Such factor can assume values ranging from $1/m_{ijk}$ to 1 During off-peak periods, $m_{ON,ijk}$ is randomly chosen based on Equation 2.4, i.e. by relying on a uniform distribution. During peak-load periods, conversely, $m_{ON,ijk}$ is randomly chosen based on a Gaussian distribution (Equation 2.5).

$$offpeak : m_{ON,ijk} = max[1, unif(0, m_{ijk})] \tag{2.4}$$

$$onpeak : m_{ON,ijk} = max[1, norm(\mu_{\%} \cdot m_{ijk}, \sigma_{\%} \cdot m_{ijk})] \tag{2.5}$$

where the parameters $\mu_{\%}$ and $\sigma_{\%}$ are set by default in such a way to obtain, respectively, a mean value of the on-peak distribution that is the average of 0 to m_{ijk} and a standard deviation that reaches the extremes of the range. Indeed, $\mu_{\%}$ and $\sigma_{\%}$ represent two further calibration parameters that can be manipulated by the modeler to reproduce behavioral patterns that are typical of its context of application, as well as to force the model towards the generation of "extreme" profiles, which may be required by robust optimization tools [27].

Optional stochastic attributes

As shown in Table 2.1, RAMP offers the possibility to define several optional Appliances' attributes, which allow to further enhance the customisability and the stochasticity of the model.

Modular duty-cycles and cooking cycles

Key optional attributes are those allowing to model predefined duty cycles and to modulate (if needed) the behavior of such cycles throughout the day. For instance, a predefined duty cycle maybe set to reproduce the behavior of a fridge; however, considering that actual fridge cycles are not fixed but rather dependent on the temperature and on user's activity patterns [40, 41], different duty cycles (e.g. standard, intensive, etc.) can be modeled and associated with different time frames to follow the variation of such parameters during the day (Figure 2.6). Alternatively, duty cycles segments can be allowed to randomly vary within a user-defined range, to reproduce the behavior of highly random and subjective tasks, such as cooking (Figure 2.7).

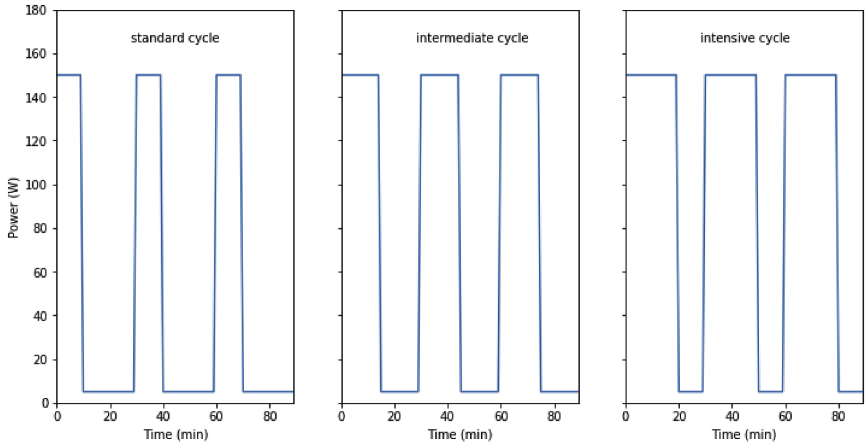

FIGURE 2.6: Example of duty cycle modulation throughout the day for a fridge.

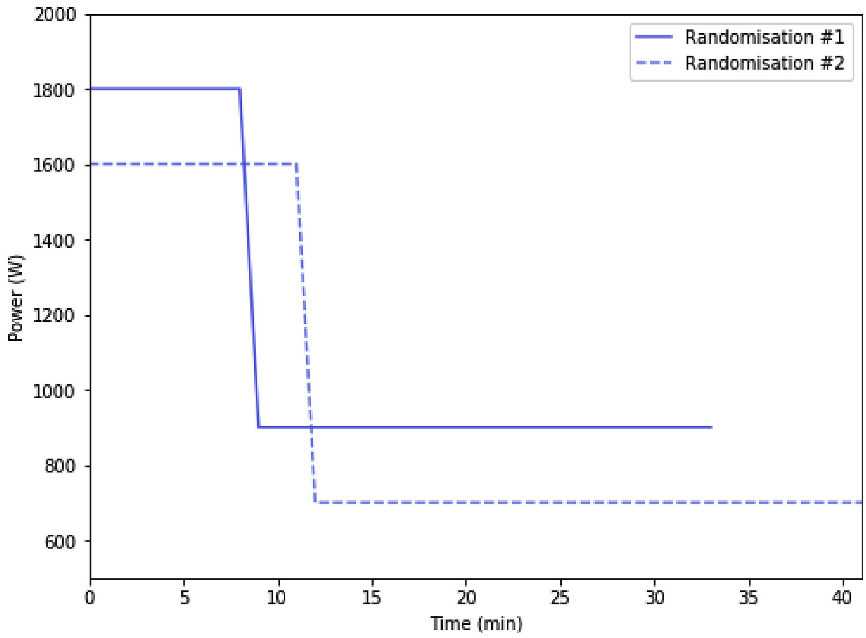

FIGURE 2.7: Example of two different randomization of a cooking cycle (in this case rep-resenting a boiling task followed by a simmering period)

Frequency of use

It is also possible to mark Appliances as "occasionally-used": in this case, the latter will be included in the set of Appliances that the i-th User will switch on during the day only conditionally to a random probability check (Equation 2.6), independently evaluated for each *User*.

$$if\ frequency_{ijk} > unif(0,1) \rightarrow \exists appliance_{ijk} \tag{2.6}$$

As a result, on a given day (i.e. a single model run) some of the *Users* of a given type may use them, while others may not; this functionality is conceived to reproduce the real patterns of use of appliances such as irons or mixers, and strengthens the unique random characterisation of each individual *User*. Figure 2.8 shows an example of different daily load profiles for a single household owning iron and using it with a frequency of 3 days a week over 7 days period; the appliance is used only occasionally, and its relative weight on the weekly average profile is thus opportunely represented.

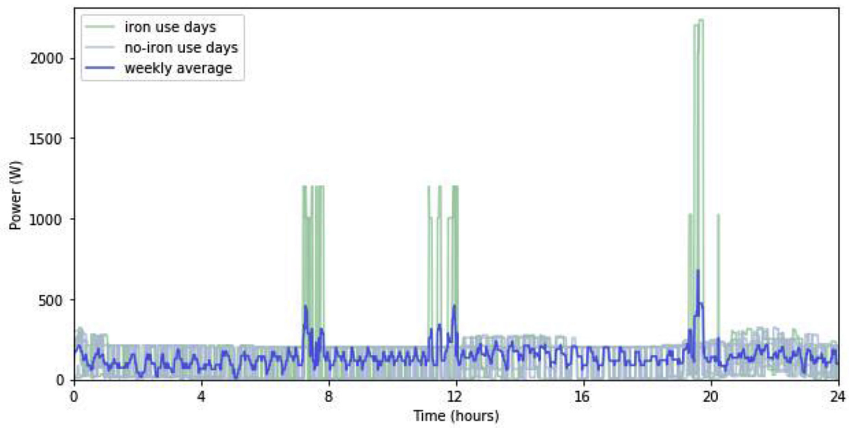

FIGURE 2.8: Example of 7 different stochastic daily profiles for a single household using iron with an average frequency of 3 days a week, modeled by the "occasional-use" attribute. Some of the stochastic daily profiles (in light green) include iron use, while others (in grey) do not.

Thermal appliances and random power regulation

A special functionality is included in the model to better simulate the behavior of thermal appliances. Those, in fact, are typically characterized by a high degree of variability in terms of absorbed power, which is a function of subjective and random preferences, for example in terms of hot tap water temperature. Such variability's embedded in the model by allowing to set a percentage random variability for thermal appliances' power (dP; thermal), which RAMP exploits to uniquely characterize each switch-on event, as shown in the example in Figure 2.9. The possibility to randomly variate appliances' power is nonetheless useful for modeling any other kind of appliance that allows for power regulation (e.g. electric heating stoves, ovens, etc.) as already shown in Figure 2.7 for the cooking cycle example.

2.3 The composition of rural communities

In [42], the authors abstain from defining the minimum amount of energy needed to meet basic needs, quantitatively. The reason is that basic needs are normative and vary significantly geographically, depending on the climate, social customs, norms and other factors inherent to the region and society. In fact, governments and policy makers in some countries made efforts to define basic or lifeline energy entitlements

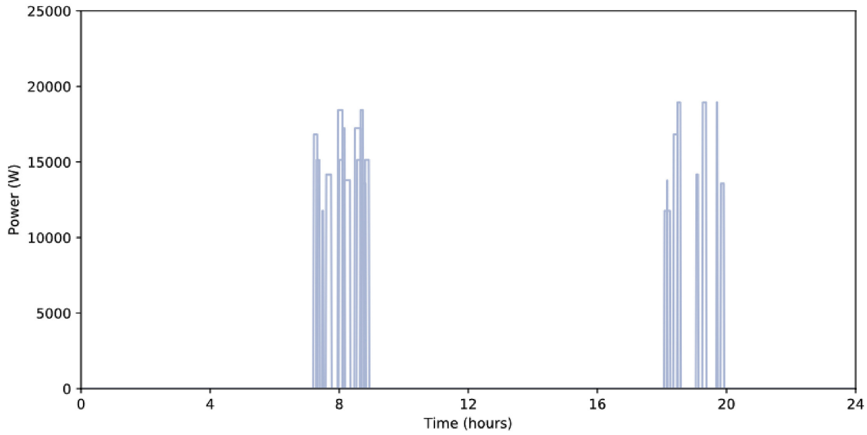

FIGURE 2.9: Example of multiple stochastic runs (10) for a thermal appliance, in this case reproducing a "shower" task: the model variates not only switch-on times and shower duration, but also the absorbed power (i.e. hot water temperature).)

for their poorest citizens to cover basic lighting, communications and entertainment needs. However, the defined entitlements fall far below what is required for income generating activities to empower local growth and development.

From the human rights perspective, the following has to be considered: 1) An adequate standard of living, including access to food, clothing and housing and to the continuous improvement in living condition, 2) the highest attainable standard of physical and mental health, 3) work, 4) education, 5) a healthy environment to live and access to basic public services [43].

In this context, [44] considers six categories to clearly identify the energy needs of rural communities in developing countries. These categories are: lighting in conjunction with information and communication technologies (ICT), refrigeration, cooking, process power and water pumping. Clearly, these categories are in line with the human rights perspective outlined above. Several studies defined sectors associated in one way or another with the above-mentioned categories [26, 43, 44, 45]. Therefore, three sectors can be identified: 1) residential, 2) community and 3) income generating activities.

Within the residential sector, several studies have been carried out to understand the complexity of household electricity use. Nevertheless, the understanding of behavior and energy consumption patterns remains limited, specially in rural areas and developing countries [46]. The determinants of household energy use can be summarized into endogenous and exogenous factors, according [46]. The former refers to economic and non-economic characteristics, as well as cultural and behavioral characteristics. Exogenous factors comprise the physical environment, policies, energy supply factors and the characteristics of energy appliances. Basically, they are the characteristics of the household and conditions outside the household. The public sector (community) is associated with services necessary to satisfy the population's right to education and good health, among others.

Energy holds the opportunity to contribute to the population economy. Not taking into account the energy needs for income generating activities, can improve the risk of energy peripheralization in rural communities [47]. There is evidence that access to electricity for small enterprises has an impact on the economy, although it is small compared to the impact on village life due to the provision of new services and products. However, it has been shown that, even after access to electricity, the majority of economic activity remains in agriculture in these areas, so the transformation processes of harvested products could represent an opportunity for growth and diversification of the community economy [48]. Table 2.2 summarizes the sectors that are taken into account in this thesis with thier associated users and activities.

TABLE 2.2: Sectors that need to meet their energy needs.

	SECTOR	USER	ACTIVITIES
1	Residential	Household	Lighting Cooking Space heating Space air-conditioning Food preservation Other Studying/working Water pumping Communication
2	Community (public institutions)	Medical centre	Lighting Space heating Space air-conditioning Powering appliances Water pumping
		School	Lighting Water pumping
		Church	Lighting Powering appliances
		Sports field	Lighting
		Public infrastructure	Public lighting Water pumping
3	Income Generating Activities (IGA)	Agriculture	Watering
		Livestock	Livestock watering
		Industry/transformation processes	Wool harvesting Crop processing
		Comerce (Small shops)	Food refrigeration
		Comerce (restaurants)	Cooking Food refrigeration
		Workshops	Repair/construction

While it is possible to define energy-intensive activities, for each sector, going into more detail depends on specific contextual factors. For example, the adoption of electrical appliances and patterns of appliance use are subject to several population characteristics. Such factors can define the 'energy culture' of a community [49].

2.4 Conclusions

The aim of the present chapter was to introduce the demand energy estimation methods used as input to the further developments of this thesis. Two main paradigms are identified. The first one makes use of aggregated historical data and reconstitutes the stochastic variations around the mean (top down). It is well adapted to the analysis of existing energy systems where it is possible to obtain the required data. The second one focuses on the individual behavior of the entities composing the system (bottom-up). It allows to analyze the impact of changes in its composition or changes in the usage of appliances. It allows the design of plausible demand scenarios beyond the current trends in the communities.

The composition of a community is also explored as a significant feature of consumption patterns. In general, three main user groups are distinguished. The residential one comprises all the households present in the community and can be characterized by different factors (cultural, size, socio-economic, etc). The community services group includes all institutions and services needed for the achievement of a certain quality of life for the inhabitants of the village. They can be public or private and can have different composition depending on the particular needs of the community. Finally, income generating activities encompass the constituents that generate local economic activity. Depending on their characteristics, they can represent an important part of the total demand of the system. It is important to highlight the heterogeneity of these components: depending on the reality of each country, they can have completely different appliances or use them in diverse ways.

Under the light of this classification, it is clear that for communities currently without electricity, the most suitable approach is the bottom-up strategy. It allows to create different plausible community compositions and explores the impacts of contrasting demand scenarios.

Chapter 3

The Bolivian case study

This chapter is largely based on:
*Peña, J., **Balderrama, S.**, Lombardi, F., Stevanato, N., Sahlberg, A., Howells, M., Colombo, E., & Quoilin, S. (2020). Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET). Energy for Sustainable Development, 56, 98-118.*

*Sanchez, C., **Balderrama, S.**, Stevanato, N., ,Quoilin, S., (2021). Energy Sufficiency for Rural Communities: The Case of The Bolivian Lowlands. In Proceedings of the SDEWES Conference, Dubrovnik.*

Bolivia is a landlocked country located in the center of the south American continent. It is one of the poorest countries in the western hemisphere. Its territory covers an area of 1,098,581 km^2 of unique geography with contrasting climatic zones. Its main altitudinal classification divides the territory in the highlands (up to 6500 m.a.s.l) and the lowlands (< 800 m.a.s.l). Climatically, the lowlands of Bolivia are characterized by a monsoon and tropical savanna climate; while the highlands experience large variations, from warm humid subtropical to cold desert climate [50]. Bolivia has currently a population of 11 million inhabitants, from which 67.3 % live in urban areas and 32.7 % live in rural areas [51]. In less than two decades, the electrification rate in Bolivia increased from 64 % in 2000 to 93 % in 2018 [52]. In the same period, the electrification rate in urban areas increased from 85 % to 98 % and from 25 % to 78 % in rural areas. The government of Bolivia has set a goal to reach universal access to electricity by 2025, requiring a national strategy to guide investment needs for grid-extension and off-grid solutions.

The political constitution of Bolivia establishes that every person has the right to universal and equitable access to electricity, and it is the duty of the government to provide all basic services through public, cooperatives or mixed entities [53]. Consequently, due to the high levels of demand growth and low coverage in rural areas [54], the Bolivian government planned to reach full coverage by the year 2025 [55].

However, rural electrification planning is a complex task, in particular because of the difficulty to evaluate the demands, which are necessary inputs to all energy planning tools. The load time series are hard to obtain for multiple reasons: (1) some communities are still unelectrified and have therefore no current load curves; (2) the load can significantly evolve from one year to the next, as demonstrated later in this thesis; (3) Bolivian communities are very diverse in terms of size, location, socio-economic characteristics and cultural context. In this thesis, to assess these demand curves, the methodology described in Chatper 2 is applied to the specific case of Bolivia. The objectives are:

1. To describe the components of the Bolivian energy system, with a particular focus on the rural areas.
2. To propose typical village configuration depending on geographical and socio-economic characteristics.
3. To propose appliance ownership and usage patterns base on on-site surveys.
4. To generate demand time series for Bolivian communities, directly usable by other energy system models.

3.1 The Bolivian Energy System

Despite its great renewable energy potential, Bolivia mainly relies on natural gas as its primary energy source. In 2000, natural gas represented 57 % of primary energy produced, and in 2010 this percentage rose up to 80 % as a consequence of significant growth in natural gas exploitation. During the period 2000-2010, non-renewable energy production increased by 208 % while renewable energy generation only increased by 21 % [56]. By 2016, the Bolivian primary energy production structure was constituted mainly by natural gas (81.02 %), followed by condensed oil and gasoline (13.15 %), traditional biomass (5.14 %), hydro-energy (0.68 %), and alternative renewable sources (wind and solar) with 0.02 % [57, 58]. The Bolivian electric system comprises the National Interconnected System (SIN, Sistema Interconectado Nacional) which supplies the main cities and the isolated systems (SA, Sistemas Aislados) that provide electricity to remote places.

Figure 3.1a and 3.1b illustrate the population size and electrification rate in the near 19 300 communities of Bolivia. The highest concentration of fully electrified communities is closer to the capital cities and close to the high-voltage network, being mostly dense populated areas. Small populations near and far away from the high-voltage grid have the lowest electrification rates.

3.1.1 Interconnected electric system of Bolivia

The SIN consists of generation, transmission and distribution facilities operating coordinately to supply the electricity consumption of eight departments representing 96 % of the national demand [59]. The Bolivian system is divided into four well-defined areas as shown in Figure 3.2: North (La Paz and Beni), Oriental (Santa Cruz), Central (Oruro and Cochabamba) and Sur (Potosi, Chuquisaca and Tarija). The high voltage transmission system is part of the SIN and includes 230, 115 and 69 kV transmission lines. The SIN generation fleet is composed of:

1. Hydroelectric power plants consisting of run-of-river units, reservoir plants and a complex whose operation is link to water supply for the Kanata metropolitan area.
2. Thermal units consist of open-cycle natural gas turbines, steam turbines that operate with sugarcane bagasse, natural gas engines and Dual Fuel units that use natural gas or diesel.

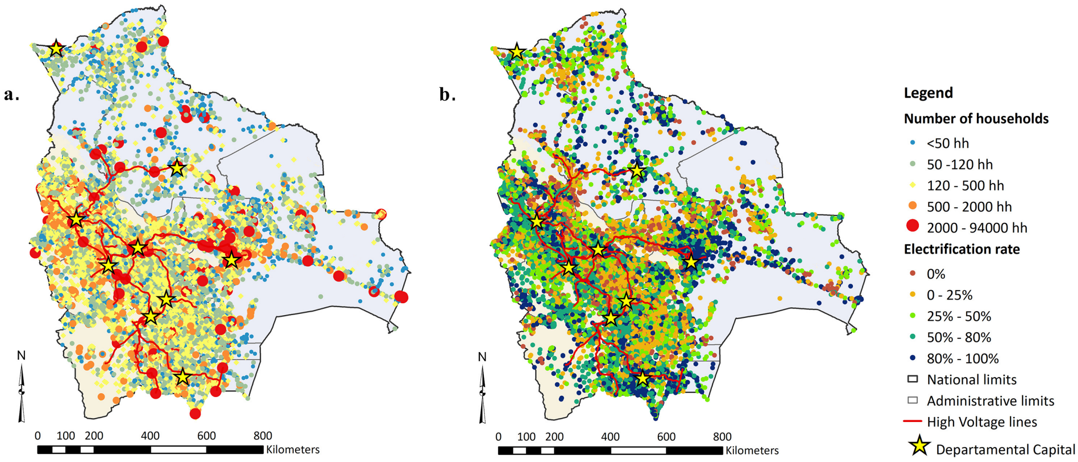

FIGURE 3.1: Overview of the electrification status in communities of Bolivia. Note that the size of the symbols is not representative of the area. a. Classification of population size in each community and high-voltage transmission lines in 2018. Population extrapolated from National Census 2012. b. Electrification rate in communities of Bolivia in 2012.

3. Combined cycle steam turbines that use the exhaust gases of natural gas turbines.
4. Wind-onshore turbines are located in Qollpana.
5. From 2020, 2 PV power plants started operation and in 2021 it is expected that 3 new eolic parks will be connected to the SIN.

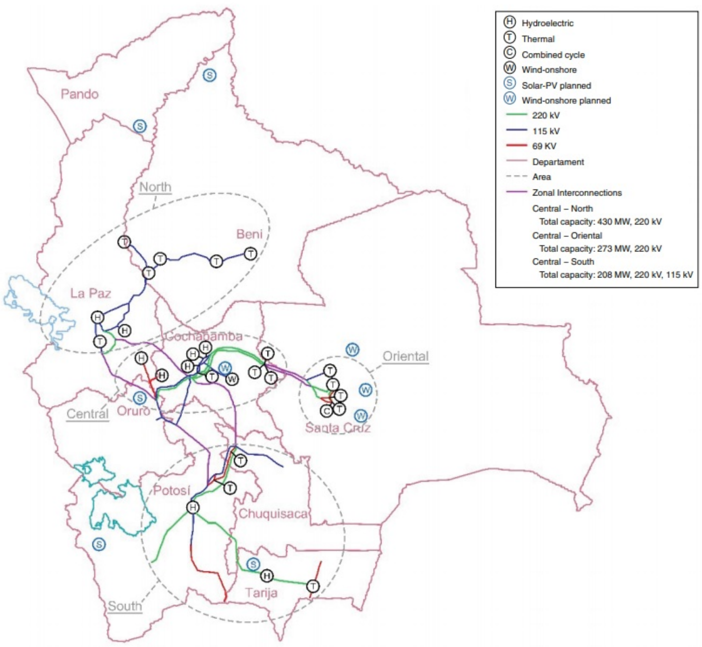

FIGURE 3.2: The SIN layout at 2016 and VRES projects planned up to 2021-2022 [59, 60].

The demand is divided into: Regulated consumers, mostly residential, who are served by distribution companies, and non-Regulated large consumers which are large industrial enterprises that directly participate in the electrical market [59]. The consumption is highest in the Oriental area with 37.8 %, followed by North with 24.3 %, Central with 21.4 % and South with 17.2 % [55]. The electric consumption of the country is mainly residential. In 2014 this segment demanded 38 % of the required energy, followed by industrial with 27 %, public services (street lighting, hospitals, public institutions, etc.) with 24 % and mining sector with 11 % [61]. In recent years, the demand has experienced a strong growth: In the period 2000-2006, an average increase rate of 4 % was registered, reaching 4.4 TWh in 2006. In 2007-2012 the increase rate was 9 % with 6.6 TWh for 2012 [55]. In 2016 the total consumption reached 8.4 TWh and for a demand of 12.4 TWh is foreseen in 2021 [59].

3.1.2 Isolated systems

In 2018, the isolated Systems supplied electricity to nearly 10 % of the total electrified households (211 thousand households) and made the 6.8 % of the total installed capacity. As isolated systems can vary largely in size (from kilowatts to megawatts), a distinction between isolated systems and microgrids is made in this thesis. Microgrids refer in this study to smaller systems with small non-regulated distribution networks. On the contrary, existing isolated systems in Bolivia have a size in the order of megawatts with regulated distribution networks.

The current isolated system installed capacity is 180 MW, with an energy mix of 66 % gas, 25 % diesel, 6 % hydropower and 4 % solar [62]. In recent years, several isolated systems have been incorporated to the national grid, reducing their carbon footprint by being dispatched as peak technologies [55]. Only a handful of microgrids have been implemented in Bolivia, serving small communities. Table 3.1 summarizes the installed microgrids until 2020.

TABLE 3.1: Installed microgrids in Bolivia until 2020, taken from [63].

Community	Installed capacity (kW)	Households
El espino	60	125
El sena	2018	426
El remanso	166.5	175
Puerto villazon	156.4	95

3.1.3 Solar potential in Bolivia

The potential of renewable energy is distributed throughout the territory. Solar energy is feasible in all regions, but mainly in the Andean highlands sector. Wind energy predominates in the departments of Santa Cruz, Cochabamba and in some parts of the highlands. The geothermal sources are located southwest of the department of Potosi. Finally, important biomass resources are available in the eastern and northern part of the country [60]. PV and battery microgrids have been the most deployed technologies to solve the electrification problem throughout the years [64]. This is explained in part by the high irradiation levels in the country.

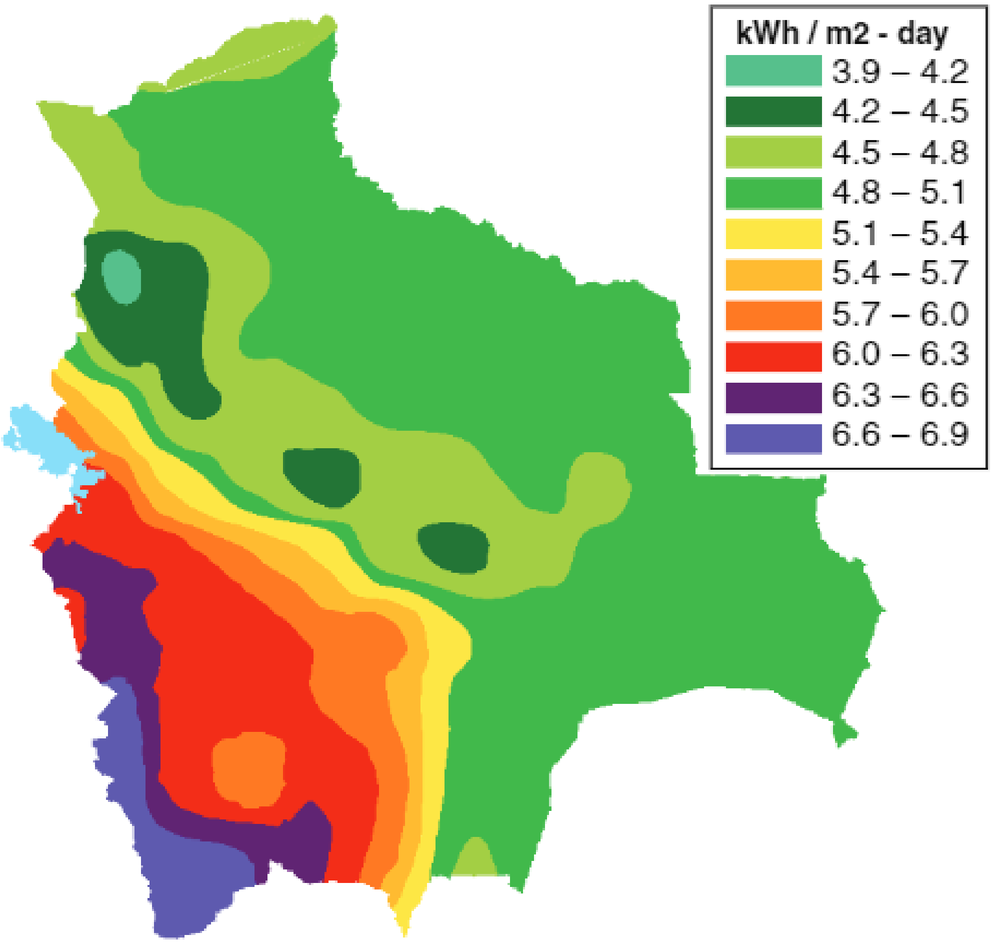

FIGURE 3.3: Horizontal global solar radiation in Bolivia (annual average), taken form [65, 66].

3.2 Bolivian rural electricity demand

Forecasting demand in a rural community is a complex task due to the high uncertainty associated with the different features of energy consumption. In the present work, this uncertainty is tackled by simulating a variety of expected demand scenarios. To this end, a series of plausible villages configurations are proposed in the next sections and the related demand curves are generated according to the methodologies proposed in Chapter 2.

One important source of information for this purpose is the geo-referenced data from the latest National Census on Population and Households. The database contains geo-referenced information of the number of households, electrification status, electricity source (grid, mini-grid, PV panel, diesel generator) and exact geographic location of Bolivian communities [67].

3.2.1 Rural Bolivian communities without access to electricity

In Bolivia, low electricity consumption has been witnessed as a characteristics of low income households. To increase energy consumption, measures have been taken by the government to increase the affordability of electricity access. One of these measures is the "dignity tariff", which was implemented in 2006 and provides an average discount of 25 % to the electricity bill for households with consumption levels lower than 70 kWh/month. For rural communities this value was decrease to 50 kWh/month due to the low consumption of these locations. It is worthwhile to note that, when the tariff was implemented, 70 kWh only represented a little more than a few light bulbs and a television working for a few hours. Nowadays, thanks to the progresses in terms of energy efficiency, the same amount can power more appliances, as shown later in this chapter.

A preliminary assessment of the Census database reveals that larger communities are usually connected to the grid. Only 14 communities with > 550 un-electrified households do not have any initial connection to the grid. On the contrary, small communities usually present high shares of low-income households and exhibit demand levels that are too small (as small as 1 MWh per settlement per year) to justify grid-extension.

These characteristics have direct implications on the least-cost electrification solutions. An initial connection to the grid and an accumulated high demand is anticipated to make grid extension a relevant alternative. In contrast, low demand and large distance (larger than 50 km) from the medium-voltage grid favors standalone systems as most cost-effective solution. For communities characterized by high demands and large distances from the grid, either isolated systems or grid extension can be considered as suitable (an indicative calibration of these relations is given by Fuso-Nerini et al. [68]).

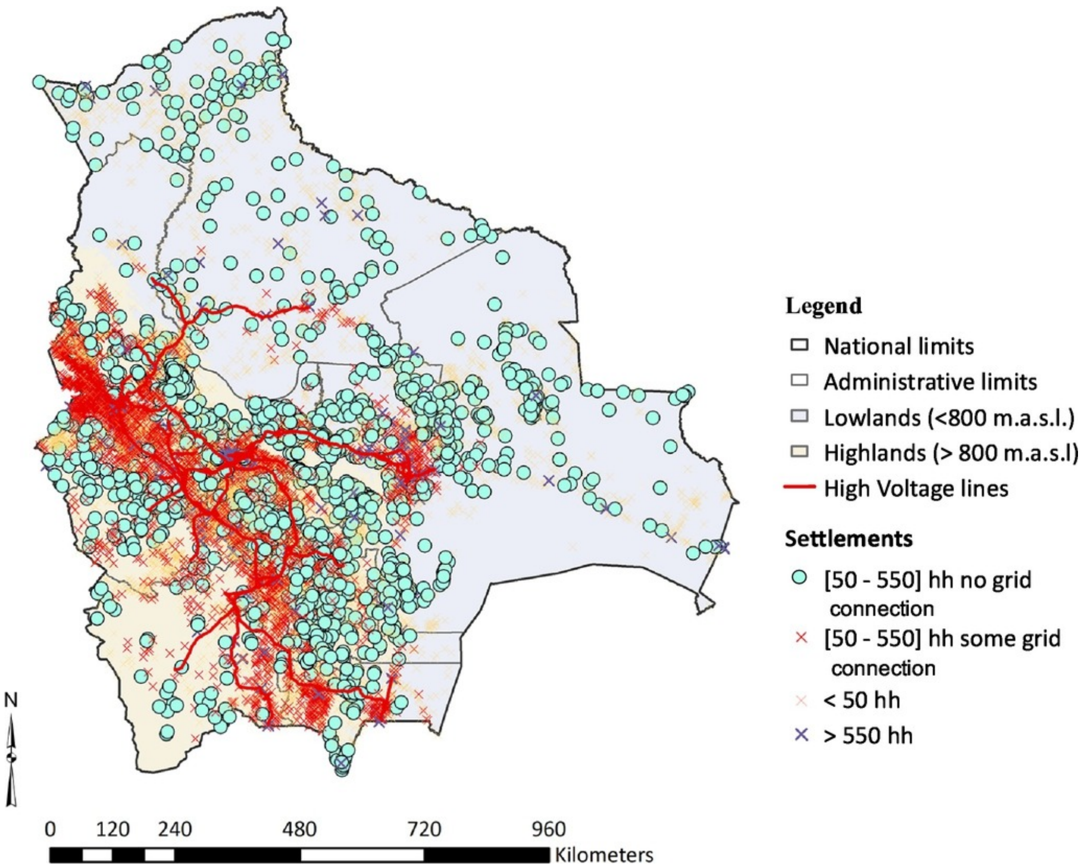

FIGURE 3.4: Population and grid connection status of Bolivian communities.

The Census dataset reports communities with a wide range of sizes, from a handful of households to a few thousands. In this work, the load curve generation methodology presented in Chapter 2 is restricted to communities with a minimum of 50 households and a maximum of 550 households. Below 50 households, the settlements are usually very scattered and not suitable for community-based power generation facilities such as microgrids. Communities above 550 usually have an existing connection to the grid and should therefore be included into the analysis of the SIN, which is outside the scope of this thesis. Figure 3.4 illustrates the location of the communities with the selected population threshold.

3.2.2 Energy demand in Bolivian rural communities

Rural municipalities in Bolivia have been grouped into different groups, of which two are distinguished for rural areas: extremely rural poor (ERP) and rural poor (RP). The communities within these municipalities reach high levels of poverty, which have been measured according to the degree of unsatisfied basic needs. The percentages of unsatisfied basic needs (UBN) inthese groups reach 96%, and 90%, respectively [69]. As detailed in the previous chapter, rural electricity demand evolution is multifactorial. In this thesis, the following factors are taken in account:

1. Community composition (Residential, community services and IGA).
2. Appliance ownership for each member of the community.
3. Location of the community (High or low lands)

3.2.3 Residential sector

Bolivian rural communities are mainly composed of lower income inhabitants that have agricultural activities as their main sustenance activity. They possess a minimal quantity of appliances which in part explain their low consumption. In addition to these ones, it is possible that a small fraction of high income households are present. They will normally have small business or are public servants in the school or health center. They will have more or higher power rate appliances, this will allow them to have a more comfortable life.

In this context, the demand will normally be modeled with a large share of Low income inhabitants. To analyse the impact of this strategy, two types of users are defined for this sector: low-consumption households (LC) and high-consumption households (HC). The latter reaches the "decent" consumption set by the government (more than 70 KWh/month), which in this study, is considered as "enough".

As explained before, Generating bottom-up demand curves requires associating a number of appliances with each user type. To that end, two field surveys from two communities have been used: the villages of "El Espino" (19.22^o S, 63.26^o W) and "Toconao" (23.24^o S, 68.2^o W). The two communities are representative of the lowlands and the highlands in Bolivia respectively. The first survey (El Espino) was carried out in the scope of this thesis in the region of Santa Cruz, Bolivia. The second survey (Toconao) was available from another project carried out in Chile, in proximity to the Bolivian border. Although from a different country, the socio-economic characteristics of the Toconao community are deemed representative enough to characterize appliance ownership in the Bolivian highlands (Figure 3.5).

Besides collecting techno-economic information, the surveys focused on the social and behavioural aspects of the villagers, including the energy-related habits and time-of-use. Finally, the division between lowlands and highlands is considered representative enough of the Bolivian rural population in terms of the cultural diversity, climate conditions and geographical areas. The survey templates for the households are provided in annexes B (El Espino) and C (Toconao). Similar surveys where taken to public institutions and generating activities. A total of 86 responses were collected, 36 in El Espino and 50 in Toconao. The surveys were taken to a member of each house or from the entity (public or income generator) .

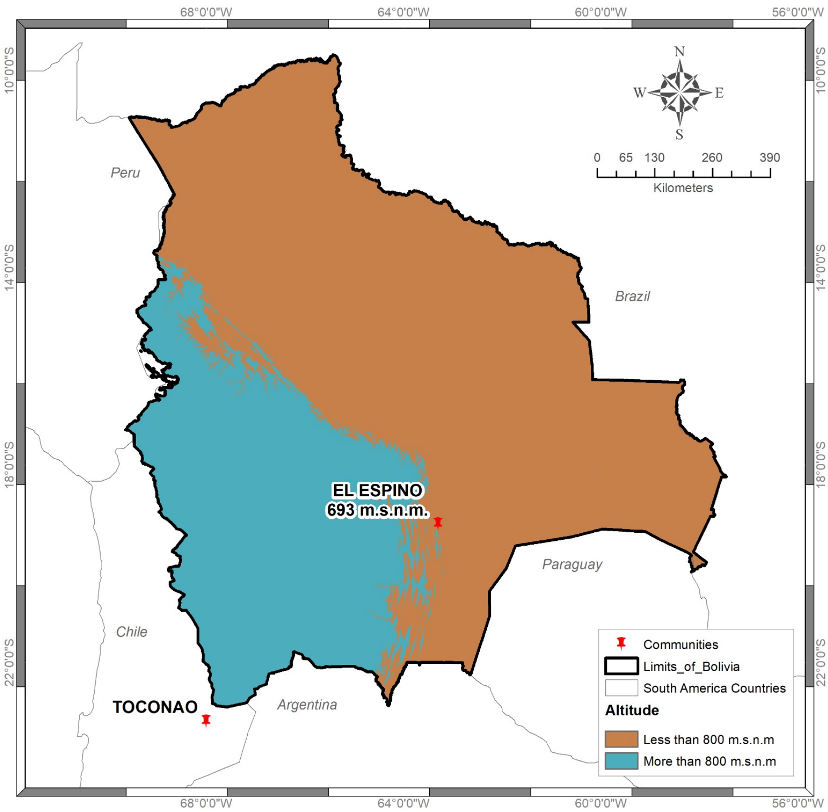

FIGURE 3.5: Bolivian high and low lands.

The ownership of appliances for the different groups is shown in Table 3.2. HC households have a significantly higher amount of appliances than LC. The main difference in terms of energy consumption is related to the presence of high power appliances (fridge and iron) in the HC group. Although there is an important difference in quantity, the light bulbs represent a limited energy consumption because of their low rated power. Interestingly, in all socio-economic levels and regions, entertainment plays an important role in the community. All households have a TV to use at different hours of the day, for entertainment, but also for information purposes. It is also worthwhile to note that, in some communities the education of children is complemented by radio or TV shows.

3.2.4 Community services sector

Public services (mainly education, health and religious service) are usually poorly equipped in small Bolivian communities, but they are however characterized by a minimum set of appliances to properly function. Specifically, an average of 3 fridges

TABLE 3.2: Appliance ownership in rural communities for the highlands and lowlands of Bolivia. The complete information of all appliance and sectors can be found in annex D.

	Highlands		Lowlands		Health	School	Church
	HC	LC	HC	LC	center		
Indoor bulb (units)	7	5	6	2	12	8	10
Outdoor bulb (units)	1	1	2	1	1	6	7
TV (units)	2	1	2	1	0	1	0
Radio (units)	1	1	0	0	0	0	0
Phone charger (units)	4	2	5	2	8	5	0
Fridge	1	0	1	0	3	1	0
Laptop	1	0	0	0	0	0	0
Iron	1	1	0	0	0	0	0
DVD	0	0	1	1	0	1	0
Mixer	0	0	1	0	1	0	0
Antena	0	0	1	1	0	0	0
PC	0	0	0	0	2	18	0
Printer	0	0	0	0	0	1	0
Stereo	0	0	0	0	0	1	0
Speaker	0	0	0	0	0	0	1

are present in the health centers were for the storage of vaccines, medications, and other heat-sensitive products. In addition, there is usually one or two desktop computers for data collection and processing task. In the schools, the computational equipment for similar purposes, but are also used to teach pupils the use of new technologies. The church has an important role for the community but only requires a few light bulbs and a speaker to conduct masses. Finally, public lighting is taken in account for each community.

3.2.5 IGA sector

Income-generating activities are all agricultural or non-agricultural activities that allow the inhabitants of a community to generate the necessary income for subsistence. In this thesis are not taken in account as they depend on different factors that are beyond the scope of this work.

3.2.6 Plausible Demand scenarios

Using these features, a series of plausible villages configurations are proposed and simulated. Survey data is used to generate aggregated demand time series using the open-source RAMP stochastic model, as proposed in Chapter 2. The synthetic demand time series are calculated for a period of 1 year and 15 villages archetypes are proposed. Each archetype describes a possible energy consumption pattern for

Bolivian villages. Figure 3.6 shows the selected (combination of) configurations for these settlements:

- The household socio-economic level is divided into two categories: LC and HC.
- Five different villages composition are simulated: A1) 90%, A2) 80%, A3) 70 %, A4) 60%, A5) 50% of LC households.
- Regarding public services, 3 situations are considered: B1) No public services, B2) School and B3) School plus medical center.
- All scenarios include public lighting and a church.
- The number of households in the community is varied between 50 and 550 with a step of 50.
- No IGA are considered in the communities.

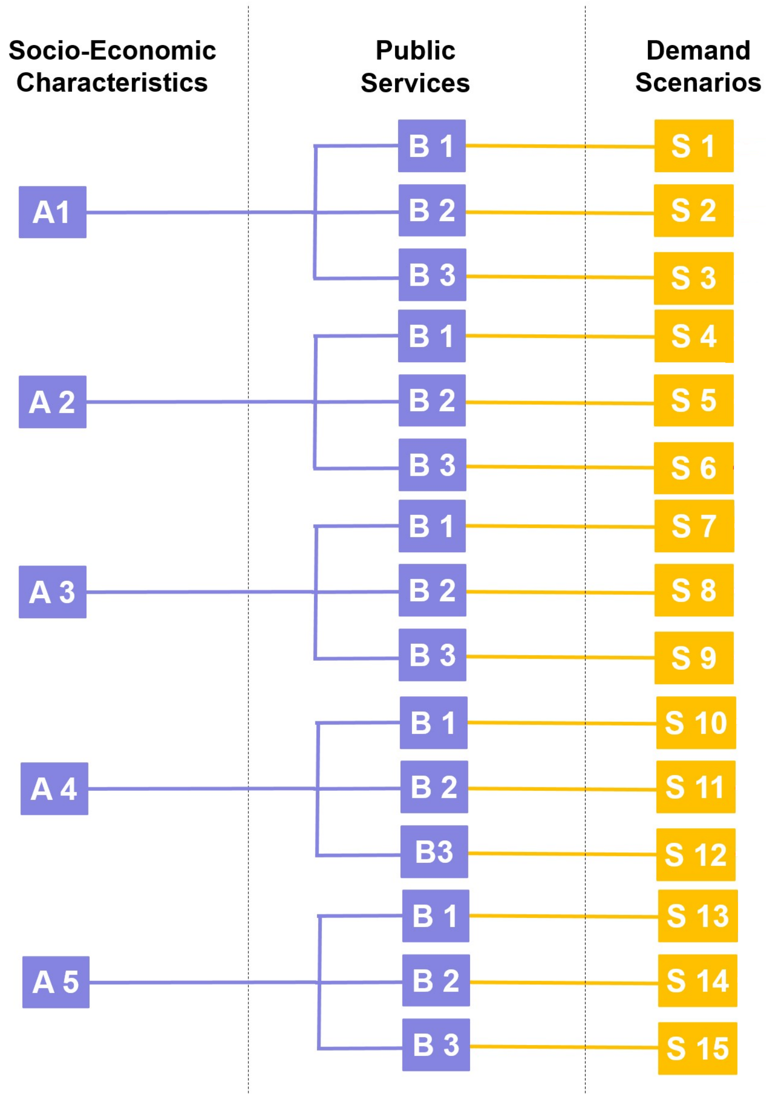

FIGURE 3.6: Construction of demand scenarios.

3.3 Results

3.3.1 Demand curves at the household level

Results at a household level are shown in Table 3.3 and Figure 3.7. A higher energy consumption is stated for HC households compared to LC households. This difference is mainly driven by the use of a fridge in higher socio-economic levels. Also, there is an important difference between the high and low lands, especially for LC households. This is explained by the presence of appliances such as an iron, which are not present in the lowlands.

The total demand for HC households is slightly higher than the maximum subsidized quantity of the dignity rate of 70 kWh. In contrast, LC family units exhibit a considerably lower consumption. This is partly explained by the usage of high efficiency electrical devices in the simulation, whereas the dignity tariff was designed in a time where light bulbs were inefficient and consumed between 50 to 100 W. In terms of life quality, the advantage of having a fringe cannot be minimized. As it allows better conservation of food and the purchase of big quantities of products that need cooling to maintain their properties. This led to an increase of several times the total consumption.

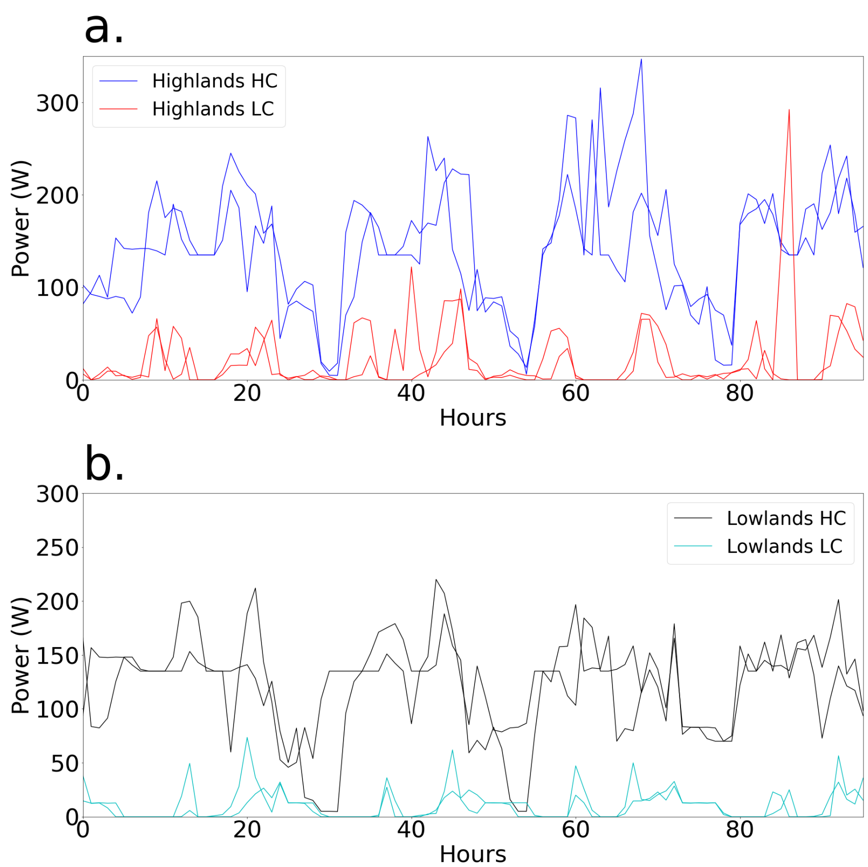

FIGURE 3.7: Simulated demand curves for a few households in the Highlands (a) and Lowlands (b)

TABLE 3.3: Summary of the household demands in the lowlands and highlands.

	Highlands HC	Highlands LC	Lowlands HC	Lowlands LC
mean (W)	131.4	20.5	112.1	10.0
std (W)	66.5	35.0	50.6	13.5
min (W)	4.6	0.0	4.6	0.0
max (W)	530.6	470.2	251.5	85.5
Total Energy (kWh)	95.9	15.0	81.9	7.3

3.3.2 Demand curves at the community level

A total of 330 stochastic load profiles were generated, building upon interview-based information from the two representative systems in the highlands and lowlands of Bolivia. The substantial differences between individual curves result from different appliance ownership and stochastic household activity patterns. In Figure 3.8, it can be appreciated the difference between including different levels of LC penetration for communities with minimum services (church and public lighting). A mean underestimation of 44 % of the total energy demand and 66 % of peak demand can happen if the comparison is made between S1 and S13 scenarios. In general, an increase in the high-income population percentage leads to an important growth in total demand and peak load. This is a consequence of the higher number of appliances owned by this segment of the population.

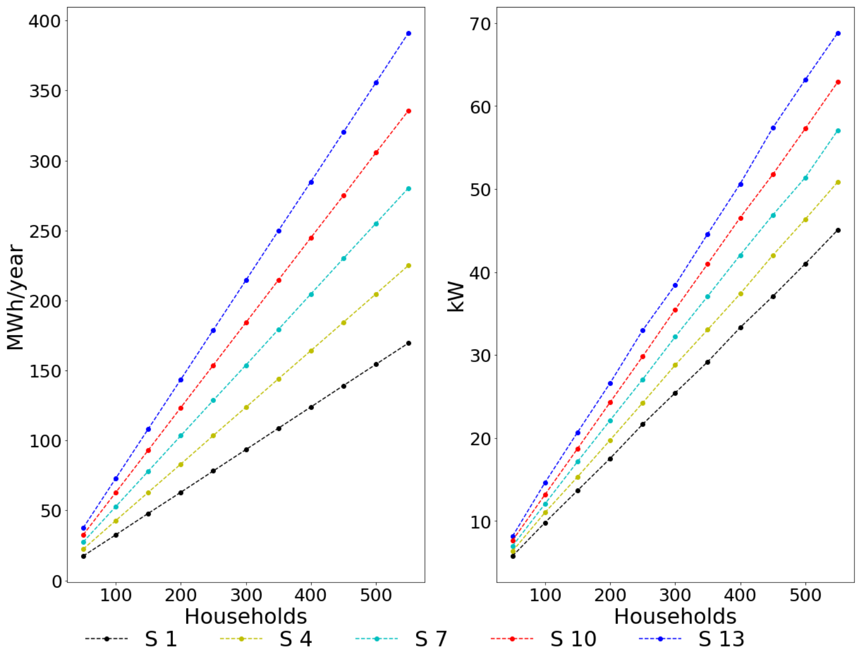

FIGURE 3.8: Total energy consumption for scenarios S1, S4, S7, S 10 and S 13 for different quantity of households for rural communities in the lowlands.

Figure 3.9 illustrates the contributions of the different energy consumption components in the case of a community with 50, 250 and 500 households with two different levels of LC penetration (50 and 90 %). The share of the school and the hospital in the total energy consumption decreases as the village size increases or with a higher number of HC. In addition to this, their contribution to the peak demand is low in all scenarios. Similar results are found for the highlands with the difference that the total and peak demand are higher due to the higher consumption of households with respect to the lowlands.

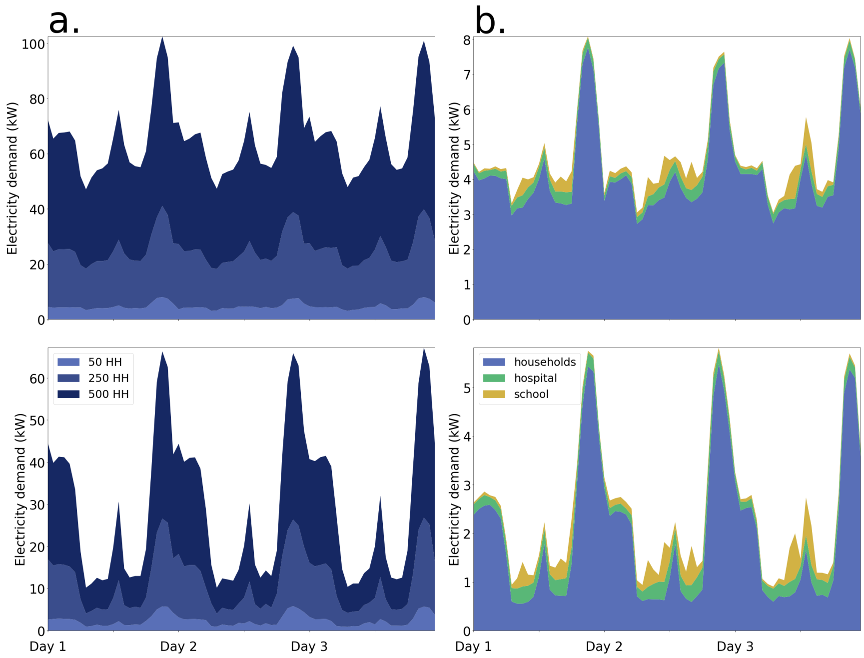

FIGURE 3.9: Demand profiles for the first days of March, Top Line: 50 % of low-income households and Bottom Line: 90 %. a) Demand profiles for communities of 50, 250 and 500 households. b) Disaggregated demand profiles for a community of 50 households.

3.4 Discussion

If demand is evaluated base on current poverty levels or ownership of appliance, there is a high probability of underestimating the demand or not be able to offer reliable energy for all normal activities in a household (SDG 7). On the other hand, it is possible that the highest energy scenarios are not met for different reasons. These scenarios can be more difficult to cover with microgrids and specially SHS, without increasing the LCOE. As explored in section 2.1, there are several situations where a microgrid was unable to meet demand. It is possible to solve this issue by installing a larger system but this can lead to higher LCOE. A trade-off therefore exists between the installed capacity and the maximum power or energy that the system is able to meet. This trade-off seems logical from an economical point of view, but there is a need to ensure adequate access to energy to individual households in such a way to ensure adequate conditions for individual development and minimum standards of living.

Although income generation activities and consumption are key assets for human development (cfr. section 2.3), they are not taken into account in the the present work due to the heterogeneity of Bolivian communities. Determining the likelihood of the presence of specific businesses in different locations is not possible without further research. Another important service not taken into account is water supply, either for human or agricultural purposes.

Finally, it is important to note that the recent evolution of demand patterns at the household level would deserve a specific analysis. In particular, the COVID-19 pandemic and the impossibility to gather people in closed spaces have entailed new organizational and occupational challenges, for example in terms of education. At the time of writing this thesis, online courses are still mandatory in Bolivia and it is expected that some form of online education will continue, even after the pandemia. This has and will have strong implications on the organization of rural life. Ensuring sufficient energy to develop these activities will be key in the future since access to education is classified as a human right [43].

Under this context, the concept of energy sufficiency (Figure 2.1) could be used as a framework to identify minimal states of energy that the system must provide. These amounts of energy will evolve during time (at household level), as the appliance adoption reaches urban levels. Moreover, it can also guide a higher limit where an increased supply of energy would not mean an equal boost of welfare in the community. This path is the contrary to one household in a developed country, where the objective is to decrease consumption. As a result of these trends, a position in the middle can be found where the needs of a community are achieved while maintaining an equilibrium with nature.

3.5 Conclusions

In this chapter, the energy context of Bolivia, and more specifically of its rural communities, is presented. The country main source of electricity is natural gas, even if recent efforts to promote renewable sources have been state in the past years a special effort to push RE has been in line with the policies of countries around the world. To accomplish this, the country has a plan to increase the installed capacity and connect as many households as possible to it. Despite the impressive increase of the electrification rate in Bolivia in the past years. It still has a relatively low 78 % electrification rate in rural areas, the majority of these people being located in remote areas hardly accessible via grid extension.

As the first step to calculate the cost to electrify Bolivia, plausible communities in the country are constructed. This is done based on information gathered in surveys from two villages located in the high and low lands of Bolivia. Each component of the village is characterized based on the parameters described in 2.2.2. A total of 330 demand instances are constructed from this information, and are further aggregated into community-centered load curves. The main driver of the obtained demand time series is residential consumption and is highly sensitive to the proportion of HC households in the total energy demand.

Finally, an analysis on the energy scenarios is proposed. The underlying idea is that each village is provided with basic infrastructure and the amount of appliances is based on surveyed communities. Although not all human needs are covered in the proposed scenarios, they mark a minimum energy level to be supplied. The obtained monthly quantity is in line with the current energy consumption in rural/poor electrified communities in Bolivia (around 70 kWh/month). In addition to this, the communities are assumed to be provided with infrastructure to ensure health, education and security for the inhabitants.

The stochastic demand curves generated in this chapter will be re-used in the next chapters as an input to size and operated isolated systems such as micro-grids or individual solar home systems.

Part II

Optimal isolated energy sizing and operation for rural electrification purposes

Chapter 4

Sizing and operations of Microgrids

This chapter is largely based on the following previous publication:

***Balderrama, S.**, Lombardi, F., Riva, F., Canedo, W., Colombo, E., & Quoilin, S. (2019). A two-stage linear programming optimization framework for isolated hybrid microgrids in a rural context: The case study of the "El Espino" community. Energy, 188, 116073.*

In rural electrification programs, when approaching 100 % of coverage, it gets increasingly hard to connect the last percentiles of the population. These are usually composed by remote, isolated communities, sometimes scattered and with low income. This challenge is illustrated, e.g., in the work of Gomez et al. [70], in their assessment of the Brazilian government efforts to provide electricity for all its citizen. The study shows that the remaining people without access to electricity are those with the highest distance from the grid and the lowest income in their respective regions.

In these critical areas, hybrid microgrid systems can offer reliable and potentially clean electricity. They involve different renewable and non-renewable energy sources and storage systems with complementary operational characteristics. The synergies between components, when optimally exploited, can lead to efficient and environmentally friendly systems, as shown by Diaz et al. [24] in the analysis of 28 isolated microgrids.

Under the right conditions, hybrid microgrids can provide energy with lower costs compared to traditional alternatives (grid extension, diesel-based microgrids and solar home systems). For instance, Mentis et al. [71] applied GIS techniques to calculate the cost of electrification in Nigeria, stressing how a higher cost of diesel can significantly strengthen the profitability of hybrid solutions. In [72], Nerini et al. expanded the cost model of the GIS approach and analyzed the impact of different factors on the levelized cost of energy (LCOE) for each electrification option. The results show that microgrids are best suited for communities that are located far away from the main grid.

4.1 Recent challenges in isolated microgrids modeling

Despite the advantages of hybrid microgrids, their potential is not yet fully exploited. One of the reasons is that their planning and operation face several challenges linked to the following aspects:

- the high degree of uncertainty associated with renewable energy potential forecasts
- the complex dynamics that governs current and future evolutions of electricity consumption in rural contexts [73]
- the imperfect mathematical representation of components operation [74]

4.1.1 Parametric uncertainty

The first challenges are related to long-term demand and renewables projections. The uncertainty associated with these parameters is typically referred to as "parametric uncertainty". Models are generally fed with these data in the form of exogenous parameters [75]. As explained in section 2.1, energy demand is highly stochastic. Depending on a variety of factors, the total energy demand, the peak load or the shape of the consumption curve can change drastically depending on the selected community or the geographical location, and can significantly evolve from one year to the next. These changes impact the sizing process, as explored by Riva et al. [76]. In that study, the authors soft-link a bottom-up method to develop long term projections of appliances in a community with an energy demand model generator and a sizing model. In addition, two other load curves are created (fixed demand and a constant demand growth rate for the duration of the project) and the resultant optimal designs are compared. The final conclusion is that if demand evolution of the system is not taken into account, there could be important impacts in the installed capacities and the viability of the project. In a posterior work, Riva et al. explore the different diffusion methods for the connection of households to microgrids as a function of time with similar conclusions regarding the impact of uncertainty on electrification projects with microgrids [77].

As regards long-term forecast of renewable energy availability, Diaz et al. [24] show that fuel consumption of a hybrid system can significantly vary from one year to another depending on the renewable energy output in that period. In [78], a whole micro wind turbine based electrification system had to be re-designed as a result of the inability of the first configuration to meet the demand for some households. The impact of mountains in the uniformity of wind resource was not considered, leading to a lower energy production than initially forecast.

4.1.2 Structural uncertainty

Another open issue relates to the mathematical formulation adopted for the sizing and architecture of microgrids, which necessarily requires a compromise between real-life relevance and computational efficiency and accuracy. In fact, microgrids components are often modeled in a simplified manner using constant efficiencies or neglecting technological constraints, either due to a lack of data or to ensure computational tractability. Altogether, this increases what is commonly referred to as "structural uncertainty" [75]. Diesel generator models, for example, often neglect decreased part-load efficiencies or minimum load constraints. As demonstrated in the next chapters, this can lead to significantly overestimated performance and therefore biased planning of the system.

In order to handle non-linear formulations within a reasonable computational time, Mandelli et al. [79] and Orosz et al. [80] adopted a heuristic enumerative optimization, that consists in testing all the possible combinations inside sets of nominal capacities from the analyzed technologies. The HOMER® software is another example of such approach and is widely used in practical project implementation, as highlighted by Neves et al. [81] within their review of hybrid microgrids in micro-communities. Nonetheless, enumerative approaches can be computationally intensive and only guarantee a local optimum. In order to reduce the time needed to solve the non-linear optimization problem, meta-heuristics techniques inspired by natural phenomena have been tested with success, e.g. by Sigarchian et al. [82] or Altes et al. [83].

Alternatively, the use of linear programming (LP) is very common in the scientific literature [26], since it allows to obtain the global optimum of very large problems in a computationally-efficient manner. However, their main disadvantage is the impossibility to model non-linear or discontinuous component characteristics. The use of mixed-integer linear programming (MILP) offers the chance to keep the main advantages of LP whilst offering a satisfying approximation of non-linear behaviour, at the expense of a lower computational efficiency [74]. Nonetheless, an accurate MILP characterization of components such as diesel generators requires high-quality data in terms of their real-life and site-specific operation, preferably based on real measurements.

4.1.3 Techniques for optimization under uncertainty

A first step for considering demand and resource uncertainties consists in performing sensitivity analyses on the uncertain parameters, as proposed by Brivio et al. [84]. The impact of the uncertainty of demand and energy generation can also be assessed using robust optimization, as explored by Khodaei et al. [85]. In this case, the microgrid components are still sized to meet the requirements of one scenario with a given demand and renewable energy time series.

Two-stage stochastic optimization is another way to solve problems under uncertainty. The generic formulation of the problem aims to optimize an objective function subjected to a set of constraints. This is done by determining the optimum value of the first-stage variables under the uncertainty generated by stochastic parameters on the second stage scenarios. The second-stage variables are the actions taken in each scenario to deal with the uncertainty. The two-stage stochastic framework is excellent to deal with the uncertainty in a microgrid, as shown by Zhou et al. [86] where a two-stage optimization problem is formulated to size a multi-energy distributed system. In their work, in addition to electricity, thermal and cooling demand are also met by different technologies. A meta-heuristic algorithm is used to iterate between various combinations of generator capacities, which are finally compared using an economic indicator. A MILP dispatch model is used for each scenario.

A multi-objective two-stage stochastic approach is presented by Gou et al. [87], with the objective to minimize the net present cost (NPC) and the pollutants emission using chance-constrained programming and a genetic algorithm as optimization techniques. A microgrid for an isolated island considering PV, wind turbines, batteries and diesel generators is sized. Under these conditions, the model also highlights the necessary tradeoff between the pollutant's emissions and economic objective functions. Lee et al. [88] explored how to implement a multi-stage stochastic optimization for energy systems and recommend approximate dynamic programming as one of the few methodologies that can deal with the high computational effort. More recently, stochastic optimization was adopted for the analysis of a multi-energy microgrid for the supply of both electricity and thermal energy [89]. The authors adopted a mixed-integer quadratic two-stage stochastic programming model to minimize the investment and operation costs of the system. A MILP two-stage stochastic programming model is used in [90] to design a distributed energy system and shows that uncertainty in demand impacts the sizing, while energy prices or renewable energy production have a lower effect. As an extension of the stochastic problem, Narayan et al. [91] modified the objective function to take in account the risk associated to the microgrid. They use the Markovitz method which consider variance of cost as an indicative of risk. The resulting model formulation is a non-linear optimization problem that takes in account the variance of cost in the objective function. More recently, Fioriti et al. [92] determined the nominal capacities for a microgrid using a similar approach than [86] but they did not use a unit commitment and dispatch model to calculate the optimal flows of energy. Instead, load-following (LF) and a rolling horizon (RH) dispatch strategies were used to calculate the operation costs of the system. Results shows that using RH or LF leads to similar nominal capacities of PV, batteries and inverters; however, RH strategy seems to halve the genset nominal capacity and decrease the total cost.

Research in two-stage optimization for sizing microgrids mainly considers the uncertainty in the demand and in the renewable generation. The main concern is to ensure that the proposed systems can operate under a variety of energy profiles. These profiles are typically generated in the form of time-series by means of statistical methods and historical data. The parametric uncertainty linked to this input data is particularly high in the case of remote regions, since no monitoring can be performed before the actual electrification. In addition, despite reports claiming that off-grid systems will be one of the major rural electrification options by 2030 [93], the use of two-stage stochastic approaches for sizing microgrids in an isolated rural context is not the focus on most of the literature. The lack of research in this field is reflected on systems that cannot handle the uncertainty associated with microgrids. These problems are clearly explained by Pansera [64] in his analysis of renewable energy in Bolivian rural areas. According to the author, Bolivian innovation in renewable energy lacks coordination between actors at different levels and an inadequate education system to foster research in this topic.

The main goal of this chapter is to describe different sizing techniques. In particular, the specific objectives are the following:

1. To provide the equations describing a generic microgrid for rural electrification.
2. To develop a framework for optimal sizing in a rural context. The proposed model should include state-of-the-art best practices in two-stage stochastic optimization and in microgrid optimal design and operation.
3. To ensure that models and constraints at the component level reproduce a close-to-reality microgrid operational behavior.
4. To test various components formulations with increasing computational complexity to capture the required tradeoffs between accuracy and tractability.

5. To include all models within an easy-to-reuse open-source library, thus ensuring reproducibility and transparency and facilitating its linking with other tools.

4.2 Two-stage MILP problem

The considered system comprises an electrical load supplied by renewable sources, an inverter, a battery bank and backup generators (Figure 4.1). The main optimization variables are divided into first-stage variables (rated capacities of each energy source) and second-stage variables (energy flows from the different components, on/off status of the generators). The optimization is implemented in Python using the Pyomo Library [94] and the CPLEX solver.

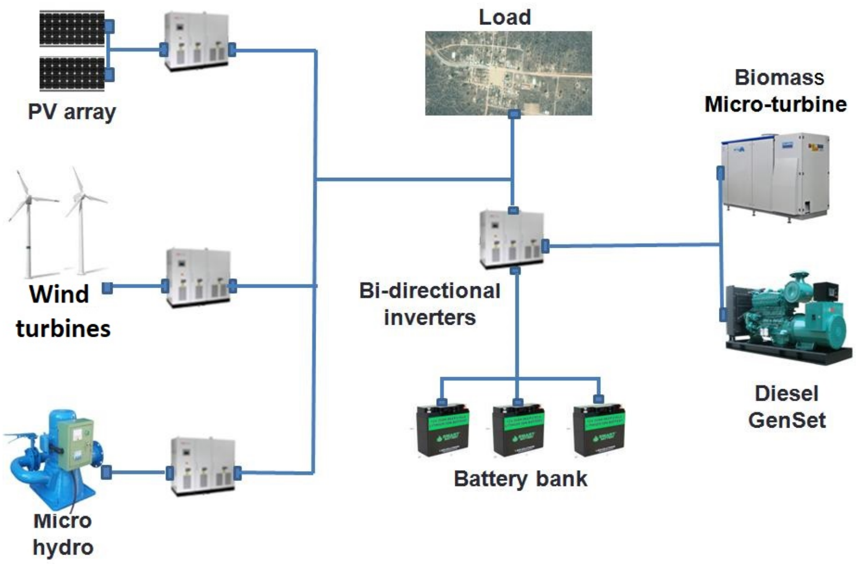

FIGURE 4.1: Proposed microgrid typology.

4.2.1 Renewable non-dispatchable energy modeling

The energy supplied by the renewable source (E^{re}) is calculated through equation 4.1, where η^{re} is the efficiency associated to the renewable source inverter, $E^{re,u}$ is the energy yield by one unit of the renewable source, N^{re} is the number of renewable units installed in the microgrid, s is the scenario being analyzed, r is the renewable source, and t is the time step.

$$E^{re}_{s,r,t} = \eta^{re}_{r} \cdot E^{re,u}_{s,r,t} \cdot N^{re}_{r} \tag{4.1}$$

4.2.2 Battery bank modeling

The battery bank is the storage system of the microgrid. Its state of charge (SOC) is computed in equation 4.2, where the energy coming into the battery is $E^{bat,ch}$ and the energy coming out of the battery is $E^{bat,dis}$. The charge and discharge efficiencies are denoted by η^{ch} and η^{dis}, respectively. In order to maintain an acceptable battery lifetime, an upper and lower limit in the quantity of energy that can be stored is enforced with equation 4.3, where C^{bat} is the nominal capacity of the battery and DOD is the maximum depth of discharge. Equations 4.4 to 4.7 are used to limit the energy flows, where $P^{bat,ch}$ and $P^{bat,dis}$ are the maximum charge and discharge power. $\Delta t^{bat,ch}$ and $\Delta t^{bat,dis}$ are the times required to fully charge or discharge the battery bank and Δt is the time step.

$$SOC_{s,t} = SOC_{s,t-1} + E^{bat,ch}_{s,t} \cdot \eta^{ch} - E^{bat,dis}_{s,t} / \eta^{dis} \tag{4.2}$$

$$C^{bat} \cdot DOD \leq SOC_{s,t} \leq C^{bat} \tag{4.3}$$

$$P^{bat,ch} = C^{bat} / \Delta t^{bat,ch} \tag{4.4}$$

$$P^{bat,dis} = C^{bat} / \Delta t^{bat,dis} \tag{4.5}$$

$$E^{bat,ch}_{s,t} \leq P^{bat,ch} \cdot \Delta t \tag{4.6}$$

$$E^{bat,dis}_{s,t} \leq P^{bat,dis} \cdot \Delta t \tag{4.7}$$

4.2.3 Diesel generator modeling

Diesel generators are designed to operate at their highest efficiency at nominal capacity (Figure 4.2). Part-load operation usually leads to lower efficiencies, which cannot be adequately modeled in a purely linear framework. This can however be done by adding integer variables to the problem, at the expense of a higher computational effort. In this work, and for the sake of comparison, both the LP and MILP formulations have been implemented to model the diesel generator.

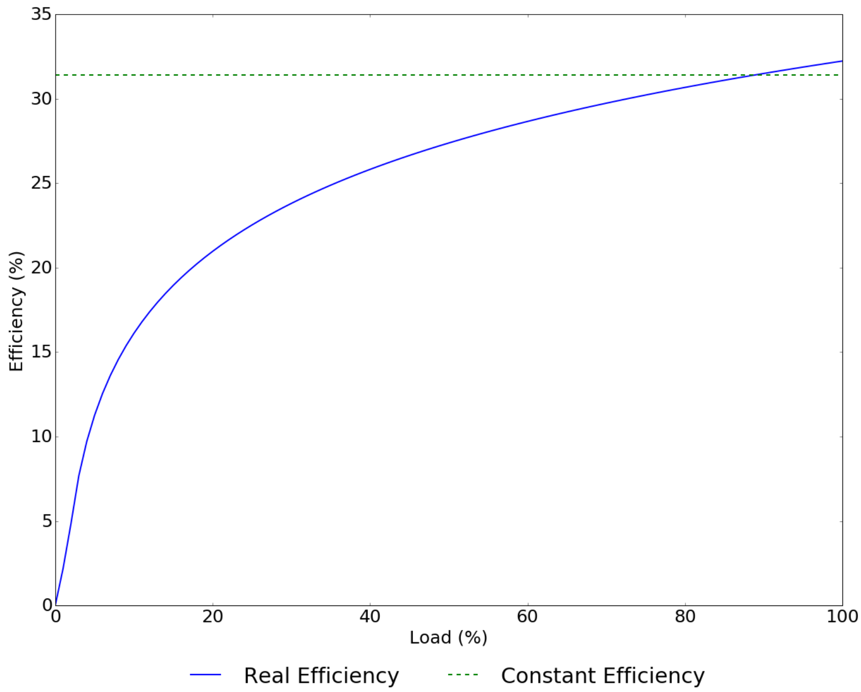

FIGURE 4.2: Efficiency vs load for a genset with constant and realistic efficiencies.

LP generator model

In an LP formulation, the generator can freely vary its output between 0 and 100 % without any penalty in partial load. The only limitation is therefore the maximum capacity of the unit. Equation 4.8 is used to limit the energy production of each generator (E^{ge}), where C^{ge} is the nominal capacity of the generator.

$$E^{ge}_{s,g,t} \leq C^{ge}_{g} \cdot \Delta t \tag{4.8}$$

In this LP model, the total cost of supplying energy with the gensets (Co^{Fuel}_{s}) is calculated from equation 4.9. b^{LP} is the unitary cost of the energy produced by the genset and is calculated with in equation 4.10. U^{fuel} is the unitary cost of the fuel; η^{ge} is the genset efficiency and LHV is the lower heating value of the fuel used by each generator.

$$Co^{Fuel}_{s} = \sum_{g=1}^{G} \sum_{t=1}^{T} b^{LP}_{g} \cdot E^{ge}_{s,g,t} \tag{4.9}$$

$$b^{LP}_{g} = U^{fuel}_{g} / (\eta^{ge}_{g} \cdot LHV_{g}) \tag{4.10}$$

MILP generator model

Some of the inherent characteristics of gensets cannot be accurately modeled in an LP framework but can be approximated using an MILP formulation. Microgrid operators normally limit the minimum energy output of the gensets (I^{ge}) to maintain acceptable average efficiencies during long periods of time. To approximate this behavior, the fuel costs of a generator are computed by summing a fixed cost when the generator is committed and a variable cost, proportional to the generated energy. This results in an efficiency curve similar to the one in Figure 4.2, with 0% efficiency when idling (zero load) and the a nominal efficiency which is only reached at full load.

Furthermore, for the sake of computational tractability, a clustered formulation is adopted, following [95]. It assumed that all N^{ge} generators have the same capacity and efficiency curve and are turned on one by one, up to their maximum capacity, to cover the demand at time t.

This, and a maximum energy output, are enforced in the model through equation 4.11, where $N^{ge}_{s,g,t}$ is the number of gensets of type g working at time t and in scenario s. Finally, the number of gensets ($N^{ge,int}$) in the system is limited by equation 4.12. With these set of constraints, the dispatch of gensets is coordinated in such a way that one generator is at full capacity before the next generator can be turned on.

$$C^{ge}_{g} \cdot I^{ge}_{g} + C^{ge}_{g} \cdot (N^{ge}_{s,g,t} - 1) \leq E^{ge}_{s,g,t} \leq C^{ge}_{g} \cdot N^{ge}_{s,g,t} \tag{4.11}$$

$$E^{ge}_{s,g,t} \leq N^{ge,int}_{g} \cdot C^{ge}_{g} \tag{4.12}$$

The fuel cost in each period (Co^{ge}) is calculated in equation 4.13, allowing to simulate efficiency drops, as described in [96, 97]. To calculate the interceptor (a^{ge}) and slope (b^{ge}), equations 4.14 and 4.15 are used, where I^{Cost} is the percentage of cost at nominal capacity from the LP model (used to transform the constant efficiency formulation into a variable one). With this approach, the MILP genset has higher unitary cost than its LP counterpart. The difference decreases with an increase of the power output and equalizes at nominal capacity (Figure 4.3). It is worthwhile to note that the total costs are also equal at nominal capacity. Finally, Co^{Fuel}_{s} is calculated in equation 4.16.

$$Co^{ge}_{s,g,t} = N^{ge}_{s,g,t} \cdot a^{ge}_{g} + b^{ge}_{g} \cdot E^{ge}_{s,g,t} \tag{4.13}$$

$$a_g^{ge} = b_g^{LP} \cdot C_g^{ge} \cdot I^{Cost} \tag{4.14}$$

$$b_g^{ge} = (b_g^{LP} \cdot C_g^{ge} - a_g^{ge}) / C_g^{ge} \tag{4.15}$$

$$Co_s^{Fuel} = \sum_{g=1}^{G} \sum_{t=1}^{T} Co_{s,g,t}^{ge} \tag{4.16}$$

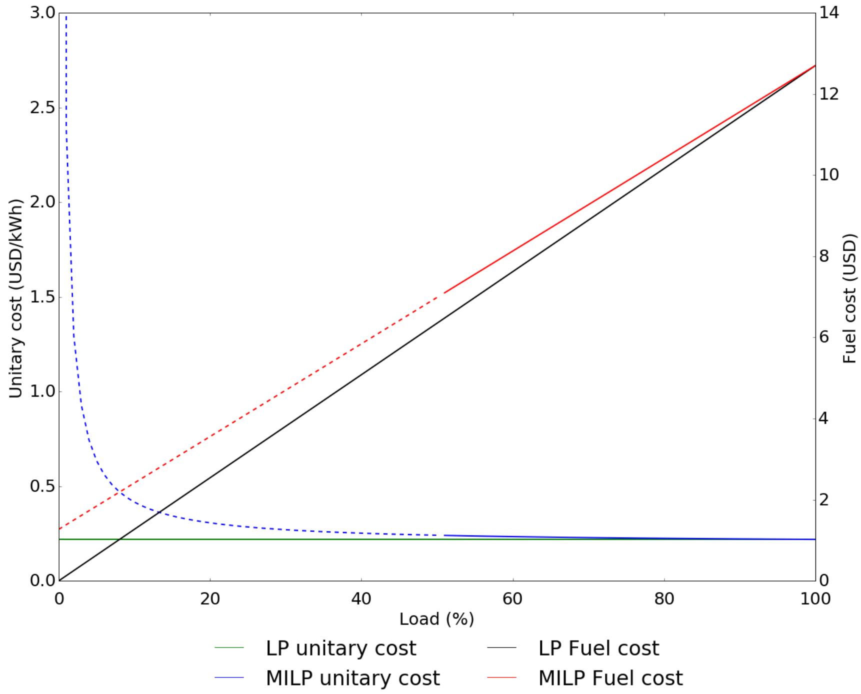

FIGURE 4.3: Unitary and fuel cost at a given power output for the LP and MILP models (The section where the genset cannot operate is displayed with a dashed line).

4.2.4 Energy constraints

The energy balance of the system is ensured with equation 4.17, where D is the energy demand of the system, E^{LL} is the energy that cannot be met by the system (LL refers to "Lost Load") and the energy that cannot be consumed or stored in the system is referred to as "Curtailment" ($E^{Curtailment}$).

$$D_{s,t} = \sum_{r=1}^{R} E_{s,r,t}^{re} + \sum_{g=1}^{G} E_{s,g,t}^{ge} - E_{s,t}^{bat,ch} + E_{s,t}^{bat,dis} + E_{s,t}^{LL} - E_{s,t}^{Curtailment} \tag{4.17}$$

The lost load probability (LLP) is imposed by equation 4.18, where $I^{occurrence}$ is the probability of occurrence attributed to each scenario.

$$\sum_{s=1}^{S} \left(\frac{\sum_{t=1}^{T} E_{s,t}^{LL}}{\sum_{t=1}^{T} D_{s,t}} \cdot I_s^{occurrence} \right) \leq LLP \tag{4.18}$$

Finally, two constraints are added to the model to allow imposing a minimum percentage of energy to be produced by non-dispatchable energy sources (equation

4.19) and to ensure a minimum amount of energy independence thanks to the battery in case of outage of other generators (equation 4.20). I^{re} is the minimum percentage of renewable energy in the system and N^{bat} is the number of consumption days that can be covered by the battery.

$$\sum_{s=1}^{S}\left(\sum_{r=1}^{R}\sum_{t=1}^{T} E^{re}_{s,r,t}\cdot I^{occurrence}_{s}\right)\cdot(1-I^{re}) \geq \sum_{s=1}^{S}\left(\sum_{g=1}^{G}\sum_{t=1}^{T} E^{ge}_{s,g,t}\cdot I^{occurrence}_{s}\right)\cdot I^{re} \quad (4.19)$$

$$C^{bat} \geq \sum_{s=1}^{S}\left(\frac{\sum_{t=1}^{T} D_{s,t}}{365}\cdot I^{occurrence}_{s}\right)\cdot\frac{N^{bat}}{(1-DOD)} \quad (4.20)$$

4.2.5 Objective function

The objective function of the sizing model is the expected net present cost of the project, as stated in equations 4.21 and 4.23. Inv is the total investment, as calculated in equation 4.24, YC is the yearly cost of supplying the demand, e is the discount rate and y is the total duration of the project.

$$Inv + \sum_{s=1}^{S}\left(\frac{YC_s}{CRF}\cdot I^{occurrence}_{s}\right) \quad (4.21)$$

$$CRF = \frac{e\cdot(1+e)^{y}}{(1+e)^{y}-1} \quad (4.22)$$

$$\sum_{s=1}^{s} I^{occurrence}_{s} = 1 \quad (4.23)$$

Inv is calculated in equation 4.24, where C^{re} is the nominal capacity of one unit of a renewable source. U^{re}, U^{bat} and U^{ge} are the unitary purchasing cost for the renewable units, battery and generator. In the LP version of the model, $N^{ge,int}$ is equal to 1.

$$Inv = \sum_{r=1}^{R} U^{re}_{r}\cdot N^{re}_{r}\cdot C^{re}_{r} + Fix^{re}_{r} + U^{bat}\cdot C^{bat} + Fix^{bat} + \sum_{g=1}^{G} U^{ge}_{g}\cdot N^{ge,int}_{g}\cdot C^{ge}_{g} \quad (4.24)$$

In order to capture the economies of scale for microgrids of different sizes, fixed costs Fix^{re} and Fix^{bat} are added to the cost function. These values are calculated in equations 4.25 and 4.26. The constant value encompass all the expenses that must be executed regardless of the project size, such as feasibility studies, pre-engineering, permitting, data recollection or environmental assessments. $N^{re,fix}$ and $N^{bat,fix}$ correspond to the binary decision variables of whether to invest in the technology or not..

$$Fix^{re} = \sum_{r=1}^{R} U^{re,fix}_{r}\cdot N^{re,fix}_{r} \quad (4.25)$$

$$Fix^{bat} = U^{bat,fix}\cdot N^{bat,fix} \quad (4.26)$$

Equations 4.27 and 4.28 are added to the model to allow the deployment of each technology only if the binary decision variables $N^{re,fix}$ and $N^{bat,fix}$ are *True*. M is defined as a large positive number. If the binary variables are *False* (i.e. equal to zero), equations 4.27 or 4.28 impose the installed capacity of the technology to 0 and the corresponding costs in equations 4.25 or 4.26 are also zero.

$$N_r^{re} \leq N_r^{re,fix} \cdot M \tag{4.27}$$

$$C^{bat} \leq N^{bat,fix} \cdot M \tag{4.28}$$

Since the fixed cost feature relies on binary variables, it is not compatible with the LP formulation. Equations 4.25, 4.26, 4.27 and 4.28 are therefore only activated in the MILP formulation of the model.

The yearly cost YC is calculated in equation 4.29, where Co^{OM}, Co^{Fuel}, Co^{rep} and Co^{LL} are the cost for operation and maintenance, the fuel costs, the cost of battery replacement and the cost of lost load, respectively.

$$YC_s = Co_s^{OM} + Co_s^{Fuel} + Co_s^{rep} + Co_s^{LL} \tag{4.29}$$

Co^{OM} is calculated in equation 4.30, where $I^{re,OM}$, $I^{Bat,OM}$ and $I^{ge,OM}$ are the percentages of the total investment cost spent each year in maintenance and operation.

$$Co_s^{OM} = \sum_{r=1}^{R} U_r^{re} \cdot N_r^{re} \cdot C_r^{re} \cdot I_r^{re,OM} + U^{bat} \cdot C^{Bat} \cdot I^{Bat,OM} + \sum_{g=1}^{G} U_g^{ge} \cdot N_g^{ge,int} \cdot C_g^{ge} \cdot I_g^{ge,OM} \tag{4.30}$$

Another important element to incorporate in the model is battery ageing. In this work, it is assumed that the battery lifetime has an inverse relation with the number of charging/discharging cycles performed by the battery. The maximum number of equivalent full discharge cycles (Cy^{bat}) is provided the battery manufacturer and is considered as an exogenous input to the model.

To calculate the replacement costs of the battery (Co^{rep}) linked to ageing, equation 4.31 is used. U^{rep} is the unitary battery replacement cost, expressed in USD/kWh, and is calculated in equation 4.32. It is assumed that power electronics is not subject to ageing. The replacement cost is therefore given by the difference between total battery cost (U^{bat}, in USD/kWh) and the cost for the power electronics (U^{elec}, in USD/kWh) divided by the total number of equivalent full cycles.

$$Co_s^{rep} = \frac{1}{2} \cdot \sum_{t=1}^{T} E_{s,t}^{bat,ch} \cdot U^{rep} + \frac{1}{2} \cdot \sum_{t=1}^{T} E_{s,t}^{bat,dis} \cdot U^{rep} \tag{4.31}$$

$$U^{rep} = \frac{U^{bat} - U^{elec}}{Cy^{bat} \cdot (1 - DOD)} \tag{4.32}$$

Finally, Equation 4.33 is used to calculate Co^{LL}, where U^{LL} is the value of lost load, defined as an exogenous input to the model.

$$Co_s^{LL} = \sum_{t=1}^{T} E_{s,t}^{LL} \cdot U^{LL} \tag{4.33}$$

4.3 Conclusions

In this chapter, a two-stage stochastic optimization framework for isolated microgrids in a rural context is presented. The problem is formulated as an LP/MILP optimization that allows sizing the components of the microgrid under uncertainty. It allows integrating the uncertain parameters and the particular context in which each project is carried out, which is key in rural electrification planning.

In the following chapters, this modeling framework will be declined and adapted for different situations, ranging from the optimization of individual solar home systems to the simulation of many microgrid topologies across the country.

Chapter 5

Case Study: The "El Espino" microgrid

A preliminary version of this chapter was published in:

Balderrama, S.*, Haderspock, F., Canedo, W., Orellana, R., & Quoilin, S. (2018). Techno-economic evaluation of rural electrification in Bolivia: lessons learned from the "El Espino" micro-grid. In ECOS 2018-Proceedings of the 31st International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (pp. 164-164). ECOS Association.*

Balderrama, S.*, Lombardi, F., Riva, F., Canedo, W., Colombo, E., & Quoilin, S. (2019). A two-stage linear programming optimization framework for isolated hybrid microgrids in a rural context: The case study of the "El Espino" community. Energy, 188, 116073.*

For the last ten years, Bolivia has gone through a period of economic growth and political stability without precedent in its 200 hundreds years of history. Under this context, the government has established an agenda to improve the infrastructure of the country. One of the most ambitious objectives is reaching 100 % of rural electrification coverage by 2025. This plan counts with the support of different international and national organizations. Furthermore, it has a comprehensive approach with solutions varying between grid extension, grid densification, the use of PV home systems and the use of hybrid systems for the electrification of rural isolated communities.

Hybrid microgrids are one of the most promising technologies for electrification of isolated rural communities in a reliable and efficient way. The use of renewable energy reduces the environmental impact and increases the village resilience to external conditions [17]. Despite its numerous advantages, this technology has never been used in Bolivia for rural electrification prior to 2015. For this reason, the Bolivian government has established a pilot plant in a rural community to assess its viability.

The community of "El Espino" was selected as beneficiary for the unit. The system design was led by a private entity with several years of experience in isolated systems. One of the main objectives of the sizing process was to ensure between 60 and

70 % renewable energy penetration and a battery capacity providing one day of autonomy. In this chapter, an *ex-post* analysis of the hybrid system is performed. The main objectives are the following:

1. To provide and describe historical monitoring data for an existing microgrid system in Bolivia.
2. To derive suitable component models from the data.
3. To propose an optimal dispatch strategy and compare it with the current one.
4. To apply a two-stage stochastic optimization which considers the main demand and resource uncertainties.
5. To test different approaches to deal with uncertainties and establish best practice guidelines.
6. To illustrate typical errors at the moment of sizing and operating an isolated system.

The monitoring data covers the time period from January, 1st 2016 to July 31st, 2017 with a time step of 5 minutes. The model presented in chapter 4 is used to analyze the system and compare it to an ideal case with optimal sizing and control.

5.1 The "El Espino" hybrid system

The first hybrid isolated microgrid for rural electrification in Bolivia is located in the rural community of “El Espino" (-19.13,-63.20) and it is part of the comprehensive plan to reach a 100% of rural electrification by the year 2025. It is installed in the southeast part of the country, 207 km away from the city of Santa Cruz de la Sierra by car. One of the purposes of the microgrid is to act as a proof-of-concept before engaging in more ambitious projects. The system was designed following the modus operandis of relying on foreign know-how [64] and built by a national Bolivian company.

"El Espino" is a rural Guarani community composed of 124 households, it comprises a small clinic, administrative offices and a school. There are no important productive activities [38], although there is a long-standing plan for the construction of a poultry farm.

Due to the remote location and its rural conditions, the average family income is low, with approximately 1 USD per day. Most of the community relies on agricultural activities to generate this income. There are however some public servants (teachers and medical staff) with higher income. Prior to the installation of the hybrid microgrid, "El Espino" was equipped with a diesel-only microgrid with a 14 kVA genset. The old generator was used for 3 hours every night to cover domestic demands but the service was discontinued in 2008 due to high operation and maintenance costs. The system was managed by a community committee, with a fixed fee of 7.2 USD/month per user. This amount was paid regardless the energy consumed or peak demand of the different consumers.

The new project involves the participation of several key players throughout its lifetime. The project planning and supervision was carried out by the Inter-American Bank of Development; the system design was established by an international expert; the Santa Cruz departmental government and the central government provided the financing; the construction of the hybrid system was handled by a national company and, finally, the operation and maintenance is performed by a local company ("Cooperativa Rural de Electrificacion").

During the project assessment phase, the expected energy demand of the system was 90 MWh/year for the first year and a peak demand close to 40 kW. By the end of the project, the expected demand was 248 MWh/year with the a peak load close to 66 kW and 235 connected households.

5.1.1 System Description

The microgrid consists of a PV array connected to an inverter, a battery bank connected to a bi-directional inverter and a diesel generator. The nominal capacities are detailed in Table 5.1 and the layout of the system is shown in Figure 5.1. The microgrid started operation in September 2015, but the available data used in this study covers the period from 01/01/2016 to 31/07/2017. The time resolution of the measurement is one hour, except for the diesel consumption, which is collected once every day. A summary of the measured variables is provided in Table 5.2

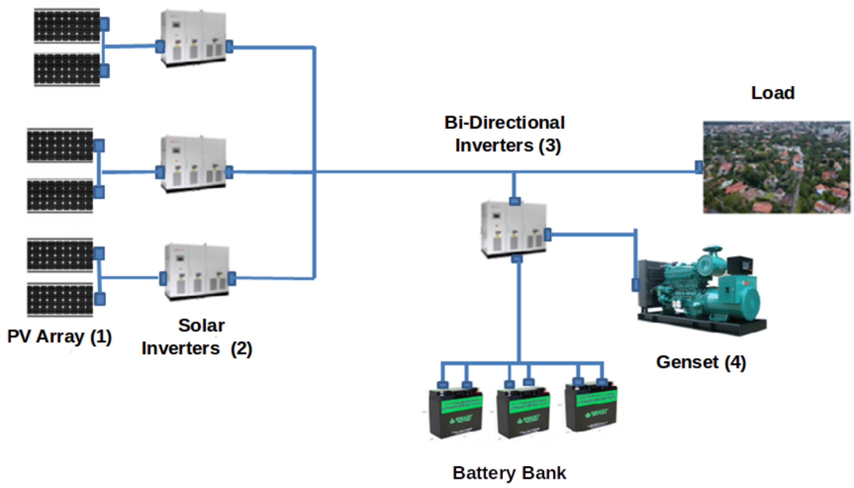

FIGURE 5.1: Layout of the microgrid "El Espino". The numbers refer to the collection data points in Table 5.2.

TABLE 5.1: “El Espino" microgrid: Technical information.

Equipment	Total capacity	manufacturer	model
PV	60 kW	Yingli Solar	YL250P-29b
Battery	464 kWh	BAE	Secura 24 OPzS 3000
Genset	58 kW	Cummins	C80 D5
Bi-directional inverter	99 kW	SMA	Sunny Island 8H
Solar inverter	51 kW	SMA	SUNNY TRIPOWER 17000TL

The next sections of this chapter are organized as follows: section 5.2 analyzes the energy flows and efficiency of the microgrid, section 5.3 proposes an optimal dispatch strategy and section 5.4 analyzes different sizing models based on this case study.

TABLE 5.2: Data collection points.

Recollection Point	Data collected
PV array (1)	Module temperature (^oC)
Solar inverter (2)	PV power (W)
Bi-directional inverter (3)	Battery power (W)
	Genset power (W)
Genset (4)	Diesel (l)
Meteorological station	Ambient temperature(^oC)
	Radiation (W/m^2)

5.2 Data Analysis

Although the system regulation strategy is not known, a careful analysis of the historical data allows to infer the following rule-based control strategy for the diesel generator and the charge/discharge of the batteries:

- Priority is given to solar energy.
- If there is PV energy surplus, it is used to charge the battery.
- If solar cannot meet the demand, the battery is discharged until it reaches 50 % of DOD.
- When both the battery and solar energy cannot meet the demand, the genset is used for sometime.
- In order to ensure an acceptable efficiency of the generator, the minimum energy output is set to 80 %.
- The genset can charge the batteries if the demand is lower than its minimum power output.

This control strategy is visible in Figure 5.2, showing the historical energy flows for a week of December 2016. It appears that the peak power of the diesel generator is well above the peak load, which limits its utilization to a few hours per day. This oversizing is explained by the fact that the Bolivian government subsidizes diesel fuel costs for isolated microgrids, at the condition that the genset has a high efficiency and that it can supply 100 % of the energy demand. The sizing has therefore been carried for an expected worst case scenario, which did not realize in practice.

With the above control strategy, the genset starts operating around 3 am and is mainly used to fill the battery due to the low demand at this time of the day. This leads to a half-full battery when the PV array starts to produce energy, obliging to curtail PV generation when the battery is full.

5.2.1 Energy Demand

As indicated in Figure 5.1, the measured load curve can be evaluated as a sum of the measurement points 2, 3 and 4. This is done in equation 5.1, where P^{De} is the demand, P^{PV} is the PV power and P^{ge} is the generator power. P^{bat} is the power of the battery and is positive if it is discharging and negative when charging.

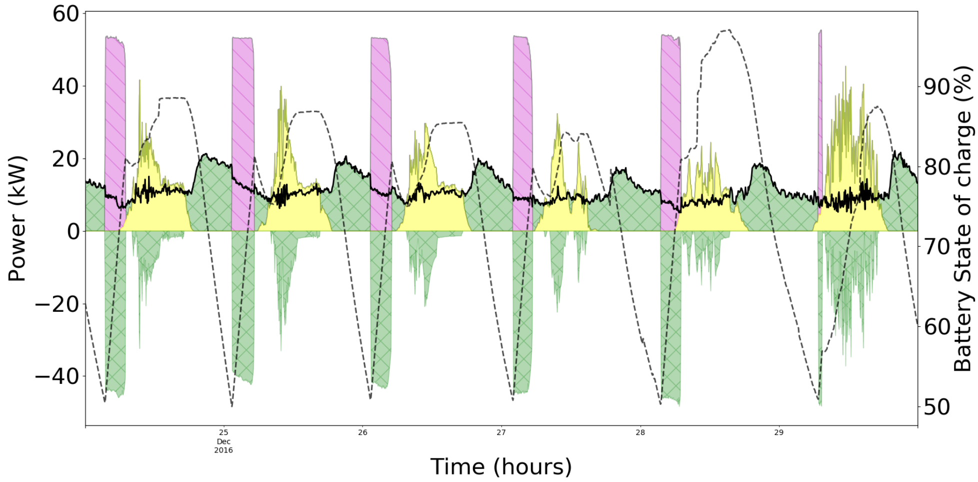

FIGURE 5.2: Real Energy Flow for “El Espino" microgrid

The load duration curve (LDC) and average daily demand are shown in Figure 5.3. During the design phase of the microgrid, a first-year peak demand of 41.8 kW and a last-year peak demand of 66.3 kW were forecast. After more than one year of operation, the maximum measured demand is 25.3 kW, indicating a clear an overestimation of the demand.

The energy consumption starts to rise at 8 a.m., and this trend continues until a period (2 p.m. to 5 p.m.) of relatively stable demand, after which the consumption experiences an abrupt increase until 20 p.m. From this point on, the energy supplied to the village decreases until 7 a.m. of the next day. The energy consumption for the 2016 year was 84.4 MWh, which is in line with the expected demand in the initial forecast.

$$P_t^{De} = P_t^{PV} + P_t^{ge} \pm P_t^{bat} \tag{5.1}$$

5.2.2 Diesel genset

The diesel generator is the backup energy source of the isolated system. It is used at times when the PV array and the batteries cannot meet the energy demand. This happens 10.4 % of the time, as shown in Figure 5.4. In order to maintain an acceptable efficiency, the diesel generator does not operate below 80 % of its nominal capacity. The genset normally operates between 7 pm to 9 am, with the highest operation at 5 am. To calculate the efficiency of the diesel generator, the daily diesel consumption data from 01/01/2017 to 30/06/2017 is used. With this information and the energy produced by the genset during this period, an average efficiency of 31.1 % is calculated (Table 5.3). As the southern hemisphere winter solstice approaches and the PV radiation drops, a higher usage of the genset is stated to meet the demand.

5.2.3 Battery Bank

A lead-acid battery bank is installed as storage technology for the "El Espino" microgrid. According to the measurements, it is charging 42.6 % of the time, discharging

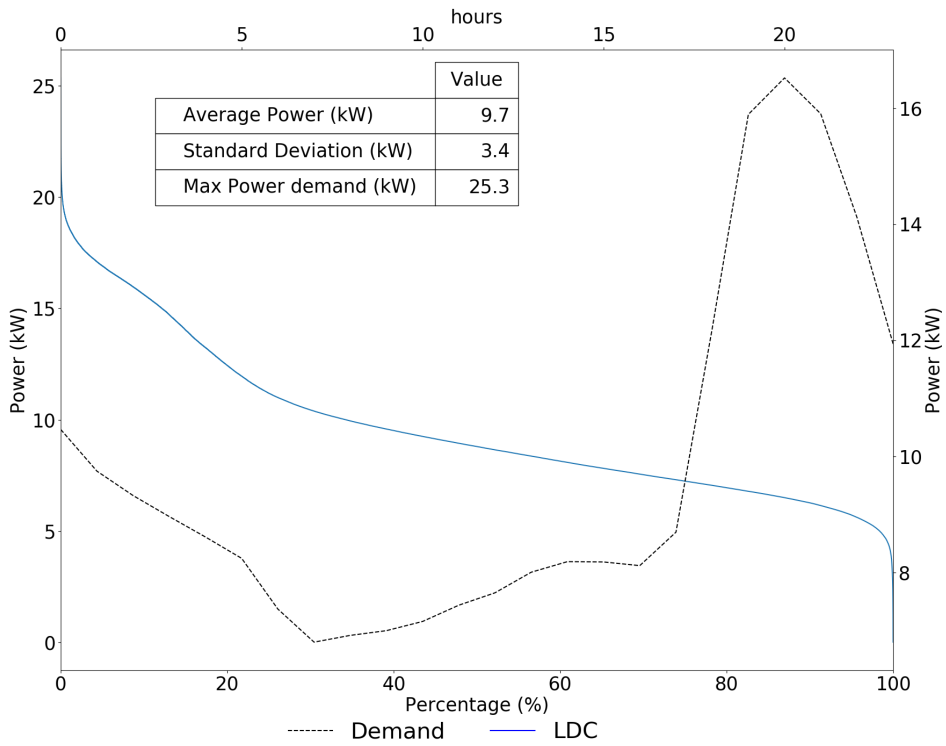

FIGURE 5.3: Average daily and load duration curves for the period 01/01/2016 to 31/07/2017.

TABLE 5.3: Performance data for the period 01/01/2017 to 30/06/2017

Month	Energy (kWh)	Diesel Comsuption (l)	Efficency (%)
January	1667	688	24.5
February	2223	740	30.3
March	3395	1129	30.4
April	3853	1150	33.8
May	5107	1518	34
June	4538	1359	33.7

57 % of the time and it is unused 0.4 % of the time. On average, the battery starts to charge at 3 am until 5 pm. At this moment the discharge phase starts (Figure 5.5). The state of charge (SOC) cannot go below 50 %; this value is set by the grid operator to avoid excessive ageing. The maximum SOC is around 3 pm and the minimum at 3am (Figure 5.5). Finally, the calculated average round trip efficiency of the battery is 74.2 %.

5.2.4 PV array

The PV array with batteries has an average power output of 13.8 kW, the maximum power reached is 51 kW and there is energy production 48.8 % of the time. If the average energy output of the PV array is compared with the average solar irradiation profile, there is a mismatch in the peaks of both profiles (Figure 5.6). This mismatch

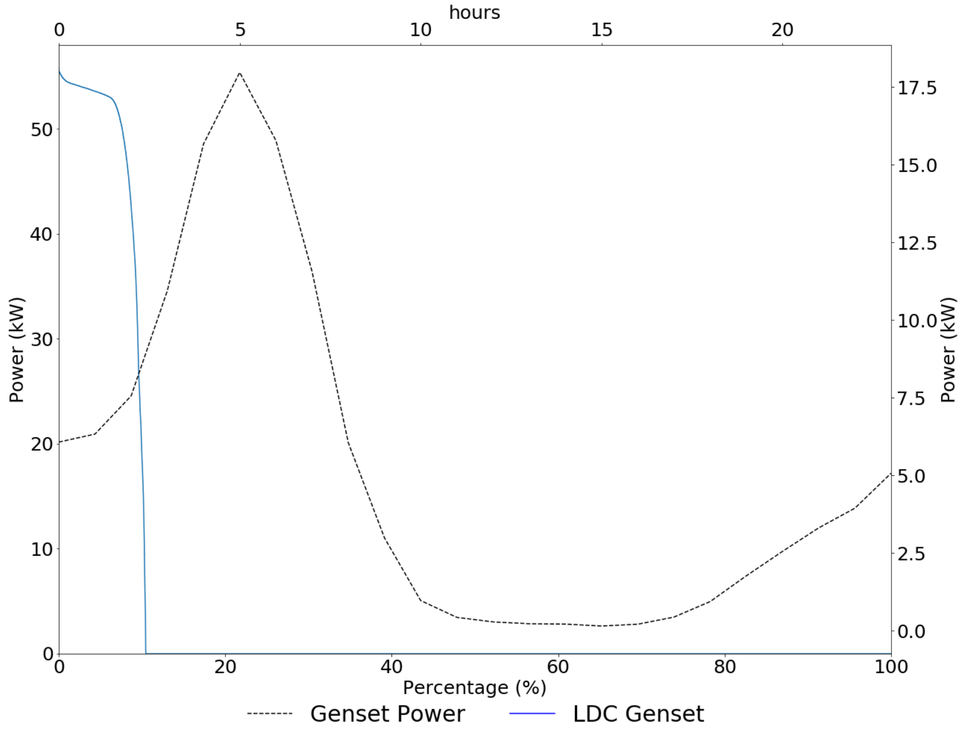

FIGURE 5.4: Load duration curve (LDC) and average power production of the genset for the period 01/01/2016 to 31/07/2017.

happens when the PV power output is larger than the demand and the SOC of the battery bank is close to 100 %. In that case, the MPPT trackers moves away from the point of maximum PV generation and therefore curtails the excess PV generation.

The curtailment of energy is also illustrated in Figure 5.2: until the state of charge of the battery reaches 92 % the PV production is higher than the demand, but from this point it decreases suddenly and equalizes with the demand.

The five-minute historical efficiency data (calculated according to equation 5.2) is displayed in Figure 5.7 as a function of the state of charge of the battery and of the ambient temperature. It appears clearly that there is a lot of scattering in the data, which is due to the variations of the other parameters. Curtailment also appears very clearly in this figure: for some SOC values, the efficiency drops significantly until reaching zero.

$$\eta_{PV} = \frac{P_{PV}}{I^{glo} \cdot A^{PV} \cdot N^{PV}} \tag{5.2}$$

Equation 5.2 is applied to the whole historical data set. P^{PV} is the power of PV array in a given moment, I^{glo} is the global incident irradiation on the tilted PV array, N^{PV} is the number of panels and A^{PV} is the surface of one PV.

PV modeling

In this section, the historical PV data is used to fit a realistic data-based efficiency curve as a function of the main boundary conditions of the PV array. The monitoring

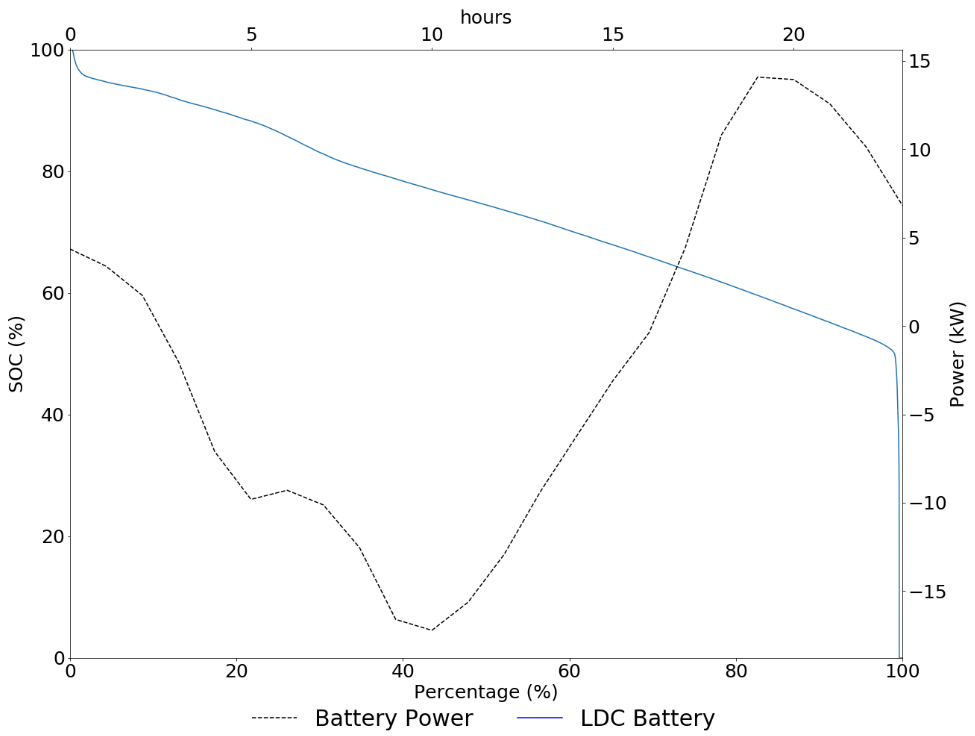

FIGURE 5.5: LDC and average power production of battery for the period 01/01/2016 to 31/07/2017.

data reveals that there are significant amounts of PV power curtailment in the system (Figure 5.7). This data should therefore be discarded to avoid biasing the fitting process. In order to select the most reliable data points, the following filtering rules are applied:

- The data points corresponding to high state of charge of the battery and therefore to a high probability of curtailment are discarded. A threshold of 85% is selected after a visual analysis of Figure 5.7
- The data points with unrealistically high efficiency (higher than 16%) are considered as outliers and discarded
- The data points corresponding to a low solar irradiation (< 250 W/m^2) are filtered out since the uncertainty on the calculated efficiency linked to measurement errors is high.

The selected observations are used to perform a non-linear regression with the efficiency as the dependent variable (*target*), and radiation, air mass and ambient temperature as the independent variables (*features*). In total, 18805 data points are used as training set.

A preliminary analysis of the data is performed using the covariance matrix and the scatter plot matrix (both provided in Annex E). The analysis reveals that the relevant variables to predict the PV efficiency are the solar irradiation, the pv module temperature, the ambient temperature and the hour of the day. The PV module temperature is not a variable which is directly available for the prediction of the system performance in random conditions since it itself depends on the irradiation and on

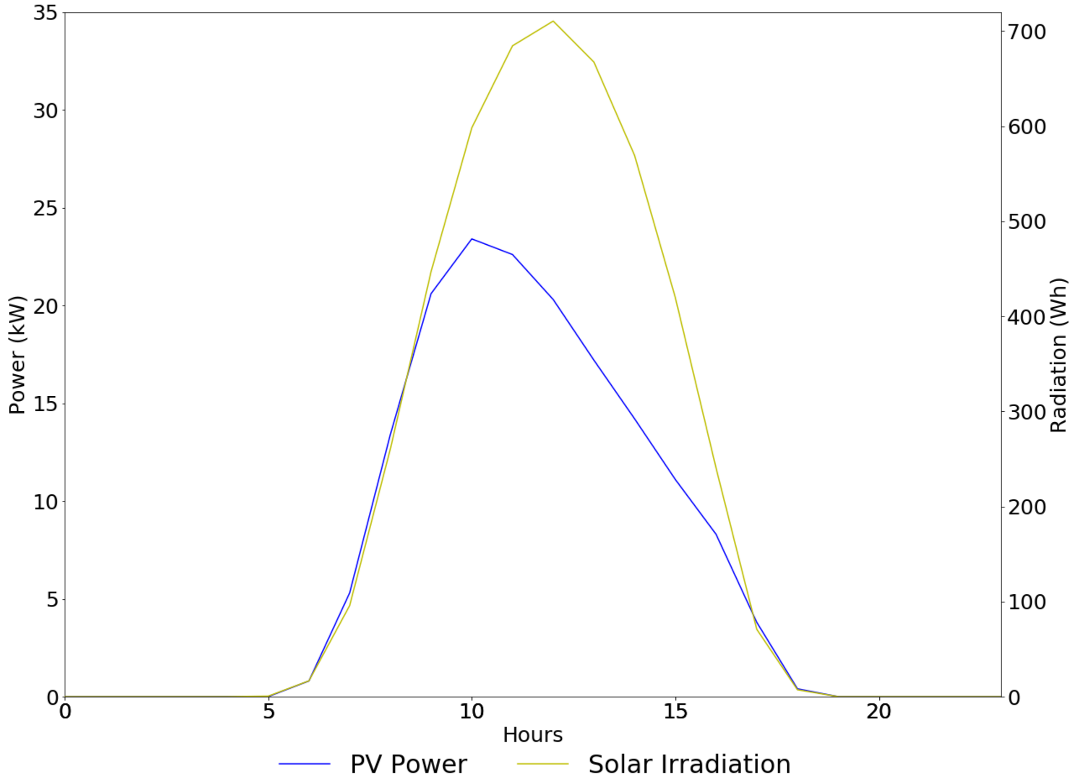

FIGURE 5.6: Average PV and solar irradiation for the period 01/01/2016 to 31/07/2017.

the ambient temperature. It is therefore removed from the input space. Furthermore the hour of the day only has an indirect influence on the PV performance: it influences the air mass and solar angles, which in turn have a direct influence on the PV efficiency. As a consequence, the hour of the day is not directly used as an input variable, but it is replaced by the air mass, which depends on the location (latitude, longitude), on the position of the sun and on the day of the year. It is calculated using the open-source *pvlib* Python library, which integrates the appropriate equations for calculation of the relevant solar angles, of the declination angle and of the equation of time [98].

The final formulation of the the efficiency curve should therefore have the generic form: $\eta_{PV} = f(G, T, AM)$, where G is the solar irradiation, T is the ambient temperature and AM is the air mass. In order to ensure a realistic shape of the efficiency curve, in line with previous experimental works on real systems, the following general efficiency equations adapted from [99] and used in this work:

$$\eta_{PV} = p \cdot \left[q \cdot \frac{G}{G_0} + \left(\frac{G}{G_0} \right)^m \right] \cdot \left[1 + r \cdot \frac{T - T_0}{T_0} + s \cdot \frac{AM}{AM_0} + \left(\frac{AM}{AM_0} \right)^u \right] \tag{5.3}$$

where the nominal irradiance $G_0 = 1000W/m^2$, the nominal temperature (in Kelvin) $T_0 = 298K$, and the reference air mass $AM_0 = 1.5$. p, q, m, r, s, u are the parameters of the equation to be fitted with the historical data.

Equation 5.3 allows ensuring a realistic behavior of the efficiency curve, even in the regions of the multidimensional input space where no reliable data was available

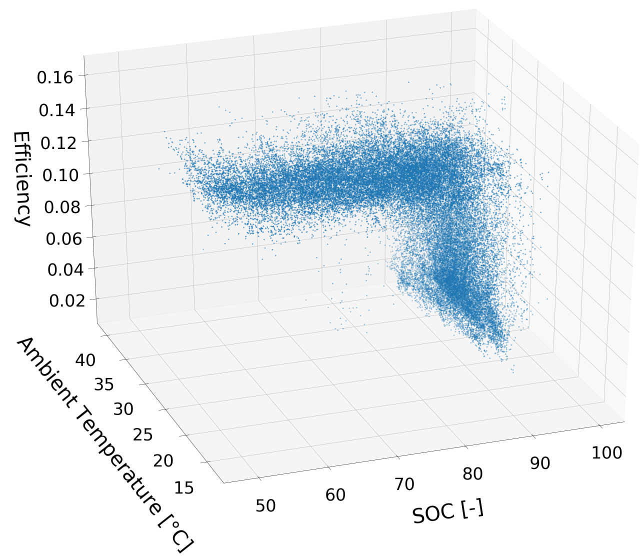

FIGURE 5.7: Measured PV efficiency as a function of the ambient temperature and of the battery state of charge (SOC)

for the regression process. This is especially true for the region of low solar irradiation, in which excessive scattering was stated in the data and which was therefore removed.

The non-linear regression is achieved solving the nonlinear least-squares problem on the training dataset. The *scipy* Python library is used to perform the regression and the obtained parameters are provided in Table 5.4.

TABLE 5.4: Obtained regression coefficients.

p	q	r	m	s	u
0.1496	-0.2688	0.0162	-0.4003	-1.0161	1.0043

The final efficiency curve is 3-dimensional. It is provided in Figure 5.8 by fixing the air mass to 1.5 and varying the two remaining features. The adverse effect of the temperature on the PV efficiency clearly appears, with a loss of about 1% efficiency point with a temperature increase of 30 K.

Applying the resulting fitted model to the whole dataset, it is possible to calculate the PV array power output in optimal operating conditions. The measured daily average PV power curve of "El espino" has its peak at 10 am and decreases faster after the peak in comparison to the fitted curve, which has a peak around midday and closely follows the solar irradiation. The mismatch between irradiation and measured data is caused by the energy curtailed after 10 am. The energy produced by the PV array is 93.3 MWh whereas the energy calculated with the regression is 126.1 MWh. In other words, the generated power from the PV array is 26 % lower than the power that would be generated if it was always operating at its maximum efficiency (Figure 5.9).

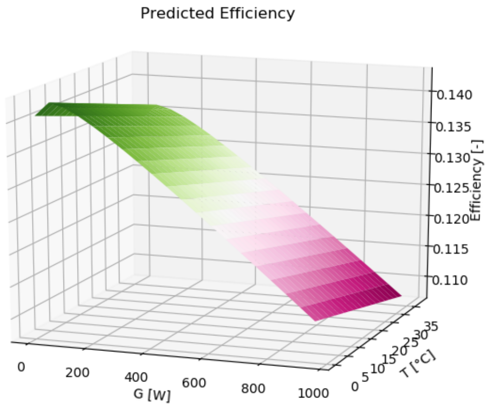

FIGURE 5.8: Predicted PV efficiency as function of the ambient temperature and solar irradiance for an air mass of 1.5

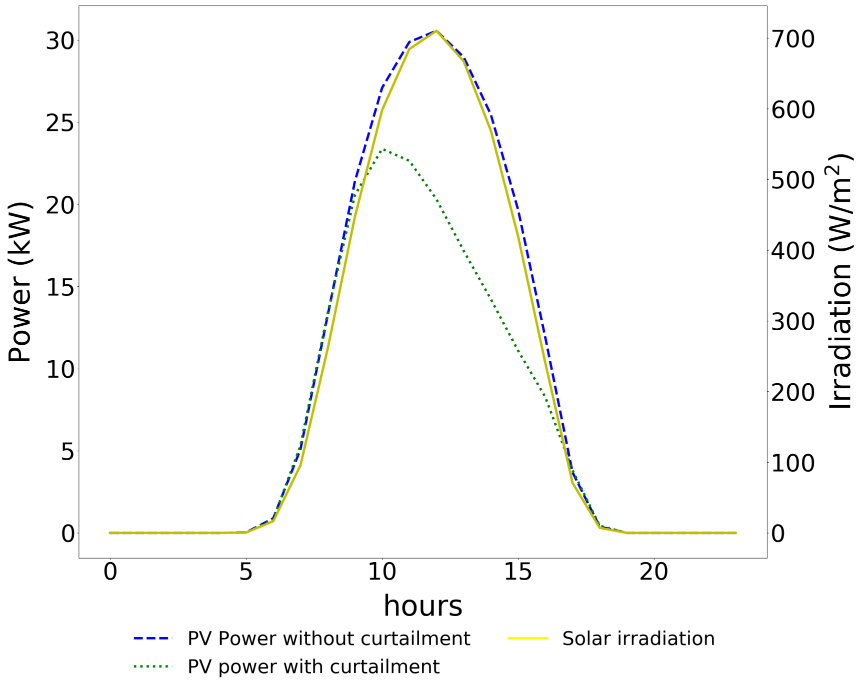

FIGURE 5.9: Mean daily PV power measure and regression and irradiation in the microgrid "El Espino".

5.2.5 Microgrid analysis

Once the demand and each of the components of the microgrid have been analyzed, their interactions within the whole system can be considered. Figure 5.10 shows the average daily energy flows during the analyzed period. Most of the day, the PV array produces enough energy to meet the demand and the surplus is used to charge the batteries. Batteries start their main discharge phase around 4 pm when the PV energy is insufficient to supply the demand. The energy stored in the battery bank is sufficient to meet the demand until 1 am. The genset works primarily during the night when the battery bank has reached its minimum state of charge. Interestingly, most of the genset power is used to charge the batteries, and only a small fraction is used to cover the demand (which is low at that time of the day) as shown in Figure 5.11. This has the consequence of partially filling the battery during the night which leads to PV energy curtailment during the day. This also has the effect of decreasing the total available energy, due to the low round trip efficiency of the battery.

From the total energy produced by the system, 57 % comes from the PV array and the rest from the genset. This value is below the objective of 70 % renewable penetration, in part due to the PV energy curtailment (26 % of the total energy produced). The PV energy that goes directly to the consumers is 32 % and 42 % to the battery. It is noteworthy that the analyzed values correspond to the first operations years of the system. With the expected increase in electricity demand in future years, it is likely that renewable penetration will decrease.

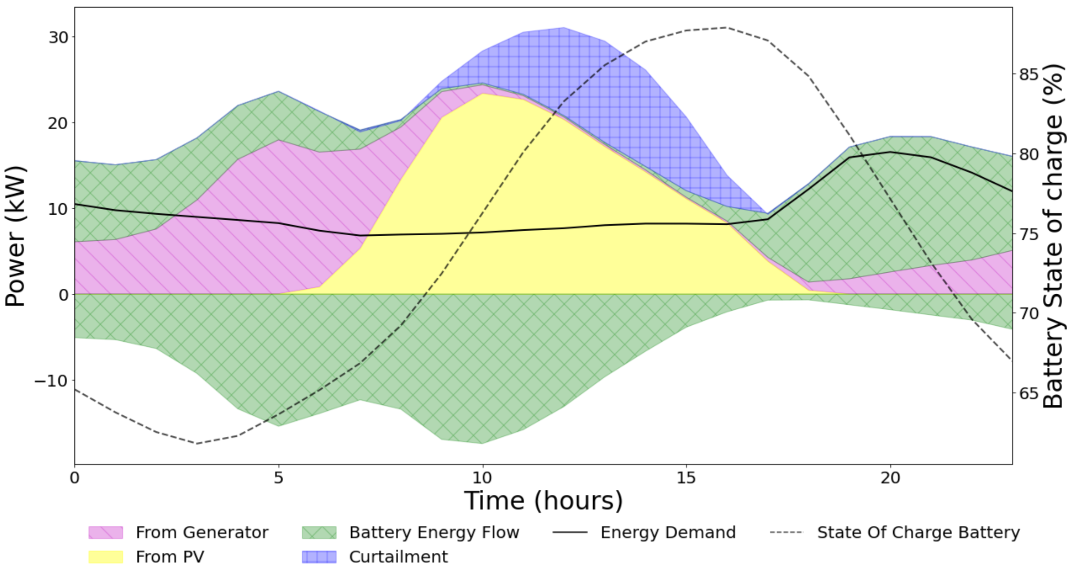

FIGURE 5.10: Average daily energy flows.

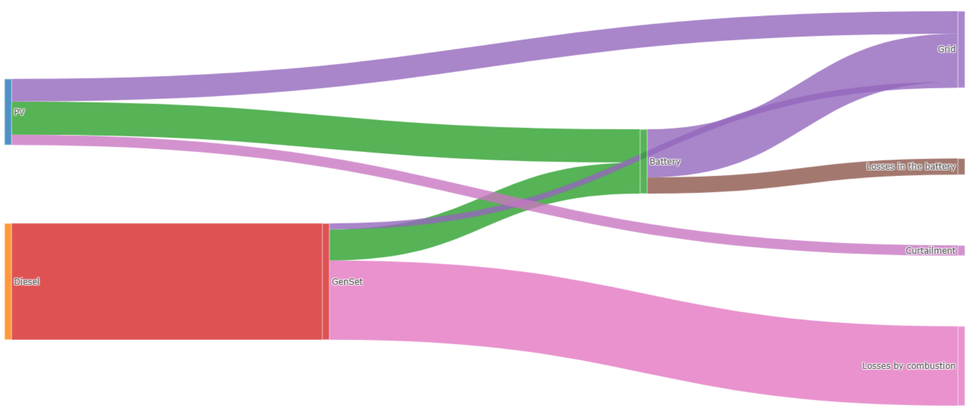

FIGURE 5.11: Sankey diagram for the energy flow in the "El Espino" microgrid from the 01/01/2017 to 30/06/2017.

5.3 Optimal dispatch for the "El Espino" hybrid MicroGrid

To assess the efficiency of the actual control strategy of the microgrid, the current dispatch of energy is compared to an ideal (optimal) case.

For that purpose, the objective function of the optimization model (4.21) is modified to take into account the variable operation cost only, as described in equation 5.4. This optimization uses the historical demand and PV (without curtailment) time series (from 01/01/2017 to 30/06/2017). The nominal capacities of the battery and genset are imposed to their nominal values in the actual microgrid.

Techno-economic data is summarized in Table 5.5. The operational cost of the battery is set to a small value to avoid unintended storage cycling, which can be seen as an additional curtailment of energy, as described in [100]. Although most isolated system in Bolivia benefit from a fixed, subsidized prices for diesel, in this work, a more representative price from the Bolivian market (0.53 USD/l) is adopted, plus 0.17 USD/l for the transportation of fuel to the village. The efficiency curve of the genset is taken from the manufacturer data sheet.

$$Co_s^{Fuel} + Co_s^{rep} + Co_s^{LL} \tag{5.4}$$

The optimization results indicate that the main change between the actual and optimal strategy is the operation time period of the genset, which is shifted to the hours of high demand, as shown in Figure 5.12. This avoids overcharging the battery in the morning and therefore results in lower curtailment.

The optimal dispatch saves 40 % of the diesel cost compared to the historical data, which is explained by a better exploitation of the solar resources. In some particular cases, the optimization even leads to curtailing energy from the genset rather than charging the battery. This is the result of a combination of the operation characteristics of the genset (in particular the 80% minimum load) and a limited battery round trip efficiency. Globally, because of a better timing of the generator operation, curtailment is significantly decreased, which indicates a clear margin for improvement in the current control strategy.

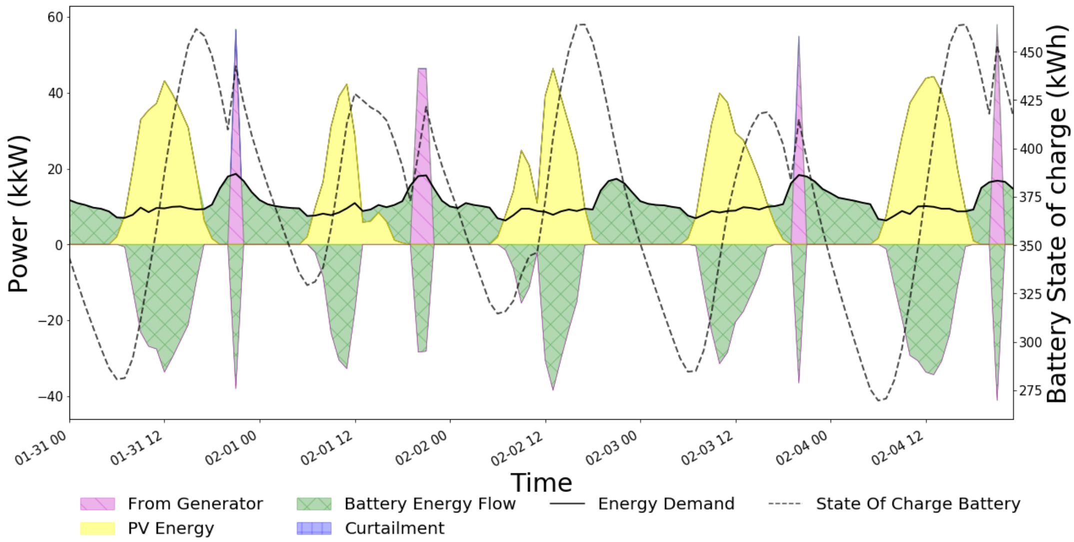

FIGURE 5.12: Optimal Energy Flow for “El Espino".

TABLE 5.5: Techno-economic parameters for the optimal dispatch

Parameter	Unit	Value
Periods (t)	hours	4344
δt	hours	1
^{LLP}	%	0
U^{LL}	USD	0
$\eta^{bat,dis}$	%	86
$\eta^{bat,ch}$	%	86
$\delta t^{bat,dis}$	%	4
$\delta t^{bat,ch}$	%	4
DOD	%	50
Cy^{bat}	cycles	1200
C^{bat}	kWh	464
η^{re}	%	100
U^{fuel}	USD/l	0.7
η^{ge}	%	39.1
LHV	kWh/l	9.9
$I^{min,ge}$	%	80
I^{cost}	%	10
$N_g^{ge,int}$	Units	1
C_g^{ge}	kW	58

TABLE 5.6: Results of the optimal dispatch model.

Variable	Optimal Dispatch 58 kW
Fuel Cost (USD)	2755
Curtailment (%)	0.7
Renewable Penetration (%)	85.3
Battery Usage (%)	58.5

5.4 Optimal sizing of the microgrid

As detailed in Chapter 4, there is an inherent uncertainty in the energy output of renewable sources and demand in rural villages. This thesis proposes a sizing framework optimization that takes this uncertainty into account and can be adapted to the needs of practitioners around the world. To illustrate its abilities, the "El Espino" microgrid is re-sized using the optimization model and then compared to the actual values, in an attempt to formulate best practice guidelines at the moment of sizing microgrids.

To remain as close to the real system as possible, the genset data from the manufacturer is used. The PV output is computed using the regression presented in section 5.2.4. The other techno-economic parameters required for the MILP sizing model are reported in Table 5.7. Table 5.8 summarizes the characteristics of each instanciation of the sizing model and assigns them an identifier.

In this optimization and in the further development of this thesis, unless it is explicitly written, equations 4.19 (Renewable energy penetration) and 4.20 (Battery autonomy) are not considered as constraints in the optimization.

For the sake of computational efficiency, the MILP formulation is tightened by bounding the variables in a way that only feasible results are considered. This is done by setting lower and higher values of the most relevant variables, as shown in Table 5.9. The maximum MILP gap is set to 0.8% and the maximum running time is set to 300 000 (s).

Finally, variations of the optimization problem are defined to assess their sensitivity on the sizing results. The parameters distinguishing the simulations are the following:

- LP/MILP formulation.
- Number of considered scenarios.
- Economic and technical conditions (2012 or 2018).
- Presence or not of exogenous constraints on the renewable penetration and on the battery size.

These differences can be recognized in the identifiers of the instances (Table 5.8), where the first word (MILP, LP) indicates the formulation, the word "Renewable" indicates the presence of renewable penetration target, the word "deterministic" indicates a formulation with one single scenario and the number (18 or 12) indicate the reference year for the cost parameters.

5.4.1 Energy demand scenarios

The selected period of the demand goes from the 2016-03-21 to 2017-03-20. From this information, 10 synthetic yearly profiles profiles are generated with a 1-hour time resolution using the methodology described in Chapter 2.2.1. In addition to the shape of the demand profile, an additional stochastic parameter is considered: the absolute value of the yearly electricity demand. The latter and its yearly growth indeed significantly impact the design process, as demonstrated by Rivas et al. [77]. To that aim, the times series are scaled by a factor whose probability distribution is averaged based on a survey sent to three local experts in rural electrification. The probabilities for the scaling factor values evaluated by these experts are reported in Figure 5.13. At the light of the latter, even among experts on the subject, there are wide uncertainties regarding the most likely demand evolution of a rural village.

5.4.2 Additional PV energy scenario

In order to consider more than one PV generation profile in the simulations, a second dataset from a meteorological station (located 29 km southeast of the microgrid) from the 2013 year is used. The timeseries cover the period from 2013-03-21 to 2013-31-12, follow by the period of 2013-01-01 to 2013-03-20. This is done to have a similar

TABLE 5.7: Techno-economic parameters for the MILP optimization.

Parameter	Unit	Value
Periods (t)	hours	8760
Project life time (y)	years	20
Δt	hours	1
e	%	12
^{LLP}	%	0
U^{LL}	USD/kWh	0
U^{bat}	USD/kWh	550
U^{elec}	USD/kWh	220
$I^{Bat,OM}$	%	2
η^{dis}	%	95
η^{ch}	%	95
$\Delta t^{bat,dis}$	hours	4
$\Delta t^{bat,ch}$	hours	4
DOD	%	20
Cy^{bat}	cycles	5500
C^{re}	kW	0.25
η^{re}	%	100
U^{re}	USD/kW	1500
$I^{re,OM}$	%	2
U^{ge}	USD/kW	1480
$I^{ge,OM}$	%	2
U^{fuel}	USD/l	0.7
η^{ge}	%	39.1
LHV	kWh/l	9.9
I^{ge}	%	50
I^{cost}	%	10
$U_r^{re,fix}$	USD	0
$U^{bat,fix}$	USD	0
C_1^{ge}	kW	15
C_2^{ge}	kW	30

temporal coverage with respect to the available measurement in "El Espino". In these time-series, the collected information is the direct and diffuse irradiation on a horizontal surface and the ambient temperature. An isotropic sky model is then used to derive the value of the total radiation on a tilted surface [98]. The radiation and

TABLE 5.8: Instance characteristics.

Tecno-economic charactecteristic	Type of optimization	Number of stochastic scenarios	Renewable penetration and battery autonomy equations	Name of the instance
18	MILP	10	No	MILP 18
18	MILP	10	Yes	MILP Renewable 18
18	MILP	1	No	MILP Deterministic 18
18	MILP	1	Yes	MILP Deterministic Renewable 18
18	LP	10	No	LP 18
18	LP	10	Yes	LP Renewable 18
12	MILP	10	Yes	MILP Renewable 12

TABLE 5.9: Lower and higher bounds for the variables in the MILP optimizations.

Variable	lower bound	upper bound
$N_g^{ge,int}$ (15 kW)	0	3
$N_{s,g,t}^{ge}$ (15 kW)	0	3
$N_g^{ge,int}$ (30 kW)	0	2
$N_{s,g,t}^{ge}$ (30 kW)	0	2
N_r^{re}	0	1000
C^{bat}	0	1000
$E_{s,t}^{bat,ch}$	0	1000
$E_{s,t}^{bat,dis}$	0	1000
$P^{bat,ch}$	0	1000
$P^{bat,dis}$	0	1000
$E_{s,t}^{Curtailment}$	0	500

PV temperature are used in conjunction with the fitted model developed in section 5.2.4 to calculate the PV energy output under these meteorological conditions.

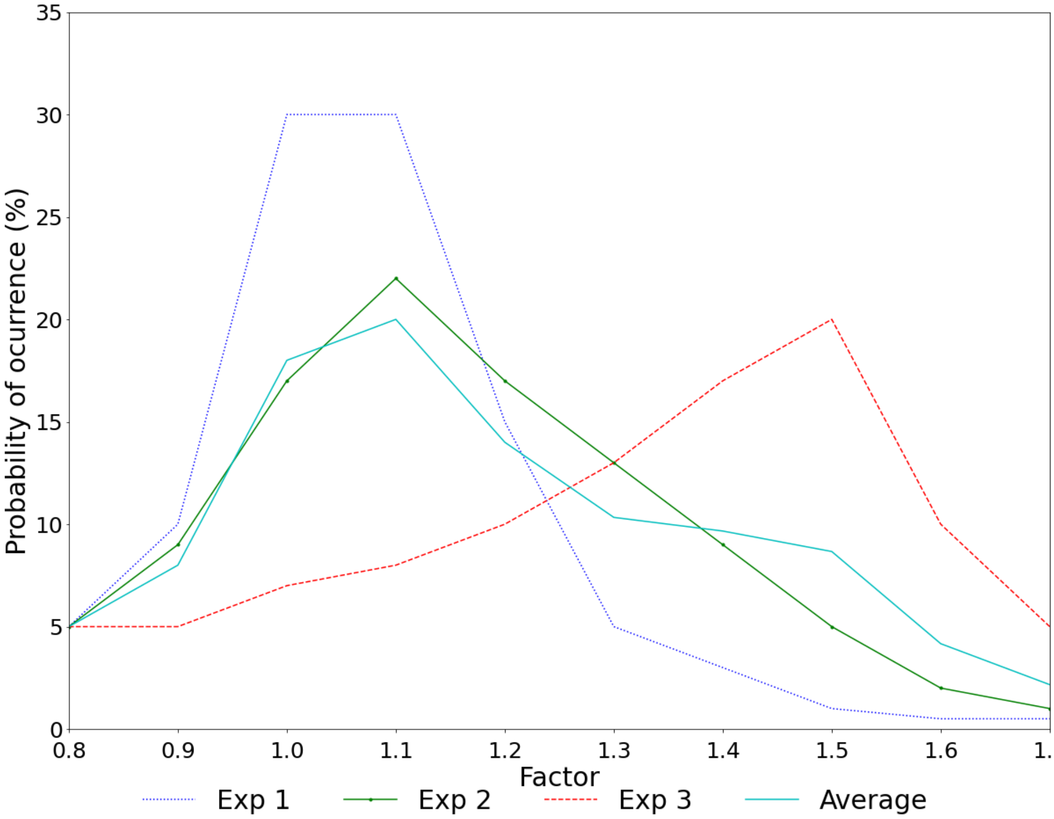

FIGURE 5.13: Probability of occurrence for the analyses scenarios from “El Espino" microgrid

5.5 Results and Discussion

The two PV energy outputs time series are combined with the 10 stochastic demands to create 10 plausible scenarios. In order to ensure computational tractability, and as suggested by Klotz and Newman [101], a partial solution (warm start) is provided to the solver. For scenarios without the renewable constraint the partial solution passed to the solver comprises 2 gensets of 15 kW and none of 30 kW. For the instances with the renewable constraint, the values passed to the solver are one genset of 15 kW and none of 30 kW.

5.5.1 Baseline simulation

In the baseline simulation, referred to as MILP 18 in Table 5.8, two possible generators nominal capacities are selected (15 and 30 kW) and their final number is left to the optimizer. The rest of the techno-economic parameters are those defined in Table 5.7. Li-Ion batteries are selected as the storage technology for these simulations since they have become one of the most relevant storage technologies in the recent years. Their advantages compared to the lead acid batteries installed in the current system include, among others high cycling capabilities, deep discharge and a recent sharp decrease in price.

Results for all MILP optimizations are shown in Table 5.10. For MILP 18, there is a renewable energy penetration of 22.2 % of the total demand in the village. Two gensets of 15 kW cover most of power demand in the system but cannot, however, cover the peak load in some scenarios. In these cases, the capacity deficit is covered by the batteries. These batteries supply 4.7 % of the demand.

TABLE 5.10: Results of the MILP optimizations.

Variable	MILP 18	MILP Renewable 18	MILP Deterministic 18	MILP Deterministic Renewable 18	MILP Renewable 12
PV rate (kW)	17	56.5	22.5	56.3	56.2
Battery rate (kWh)	44.4	357	42.8	357	571
15 kW genset number	2	1	1	1	1
30 kW genset number	0	0	0	0	0
NPC (Thousand USD)	225	417	200	414	585
LCOE (USD/kWh)	0.29	0.54	0.26	0.53	0.75
Curtailment (%)	0	0.9	0	0	0.8
Renewable penetration (%)	22.2	70.4	29.2	70.8	70
Battery Usage (%)	4.7	38.8	8	38	38.4
MILP gap(%)	1.2	0.3	0.8	0.2	0.7
Solving Time (s)	300 169	5 007	10 784	115	3 074

5.5.2 Effect of renewable and storage capacity targets

The microgrid of "El Espino" was designed to meet two objectives; to reach a battery storage capacity for one day of autonomy and a renewable energy penetration of 70 %. To enforce these objectives in the model, equations 4.19 and 4.20 are added to the optimization. In that case (MILP Renewable 18), the new optimal system heavily invests in PV and Batteries and only one genset of 15 kW is installed. The generator is undersized compared to the peak demand thanks to the possibility to cover most of it with the PV and storage.

The dispatch strategy in this scenario is shown in Figure 5.14. During the day, solar energy can meet most of the energy needs of the village and charges the battery when there is energy surplus. During the night, a combination of diesel generator and battery are used to meet the demand. The LCOE of the system increases significantly as a consequence of the two hard constraints added to the optimization. As shown in Figure 5.15, the system benefits from a better dispatch strategy. This can be seen by a lower flow from the genset to the Battery and smaller curtailment of solar energy, leading to a more efficient system overall.

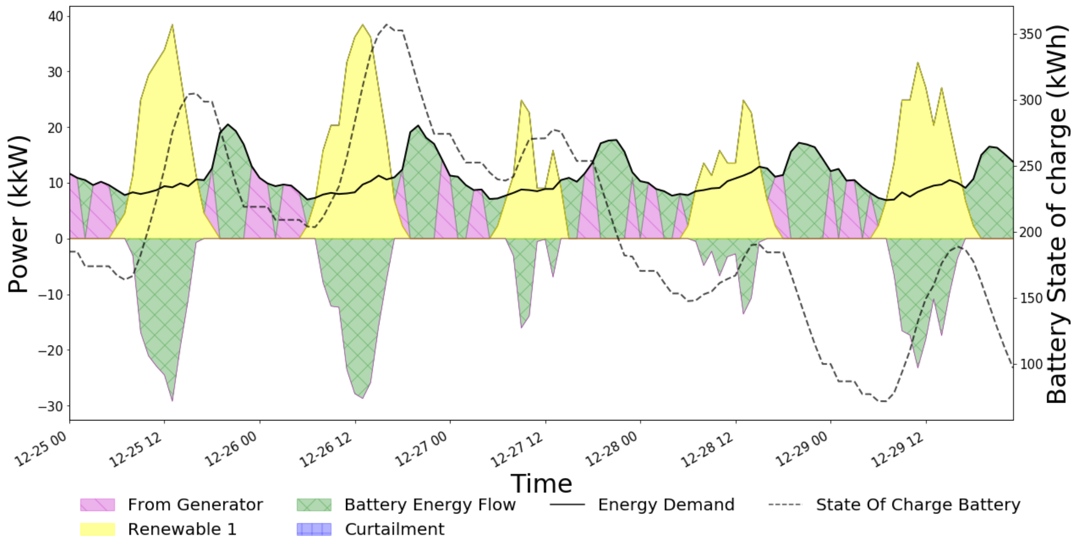

FIGURE 5.14: Energy Flow for MILP renewable 18 instance

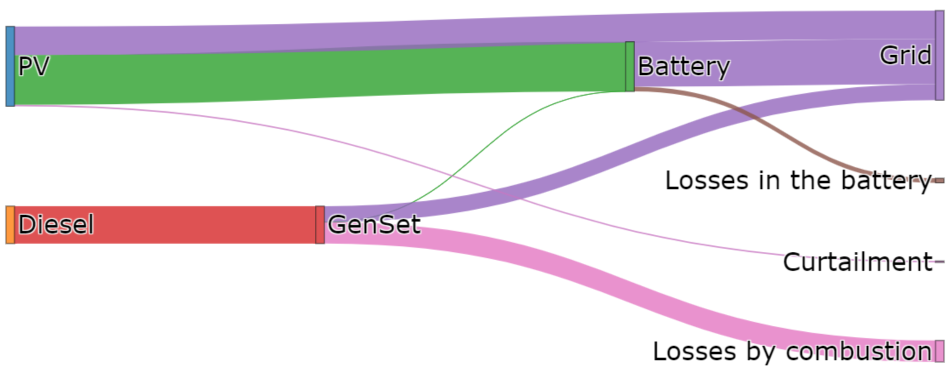

FIGURE 5.15: Sankey diagram for the energy flow in the Renewable 18 instance.

5.5.3 Renewable baseline scenario with older technology and costs

In 2012, when the real system was designed, the chosen technology was lead-acid batteries, as Li-Ion ones were not a mature technology. In comparison to Li-Ion, lead-acid batteries have lower cycling capacity and cannot perform deep discharge.

To ensure a fair comparison between the original design and the optimal sizing proposed in this thesis, an additional simulation is defined to analyze the impact of recent developments in battery technology. Technical and economic data are updated as shown in Table 5.11. Additionally, the PV investment is adapted to reflect the higher costs in 2012. Using the actualized data, the optimal system corresponds to a similar PV nominal capacity than the MILP Renewable 18 instance as the penetration constraint has to be respected. The lower technical capacities of the old storage technology leads to a higher capacity in 60 %. The higher operation cost of the Lead-acid technology and higher PV price generate an important increase in the LCOE and NPC.

When comparing the installed capacities of the PV and Batteries, it appears that the sizing model is quite close to the real values (Table 5.1). The main difference is the installed capacity of the diesel generator, where the two-stage stochastic model demonstrated that a lower rated power is able to better match the demand profiles while using the batteries to cover the peaks in high-demand scenarios. This, coupled with a more adaptive strategy, leads to a significantly smaller energy curtailment than the rule-based dispatch approach used by the grid operator.

TABLE 5.11: Updated Techno-economic parameters for the Renewable 12 instance.

Parameter	Unit	Value
DOD	%	50
Cy^{bat}	cycles	1200
U^{bat}	USD/kWh	440
U^{re}	USD/kW	1700

5.5.4 Impact of model formulation

The intensive computational effort needed to solve two-stage MILP problem can become a prohibitive constraint at the moment of sizing isolated systems. This is especially true in developing countries, where practitioners may not have the computational or technical resources to solve these complex problems. For this reason, in the following sections, we relax the two-stage MILP model to explore less computationally-intensive alternative formulations and evaluate their impact on the adequacy of the system.

Deterministic vs. Probabilistic formulation

Sizing the microgrid system in a deterministic framework involves calculating the expected demand and solar energy output of the system and solving the sizing problem with the number of scenarios set to one. The expected demand (D^{exp}) and renewable energy ($E^{re,exp}$) time series are calculated using equations 5.5 and 5.6. The sizing results (Table 5.10) show that only one genset is chosen instead of two. This is explained by the fact that extreme events are more easily overlooked with a single scenario. The optimal sizing in this case leads to a lower NPC, but also to a less robust system because of the impossibility to cover the peaks in high demand scenarios. In particular, the additional cost associated with a more robust system is 11 % of the base scenario cost. In the instance with high renewable penetration, the under-sizing is mitigated by the battery and PV size constraints. The decrease in NPC and LCOE remains limited in this latter case (0.6 %).

$$D_t^{exp} = \sum_{s=1}^{S} D_{s,t} \cdot I_s^{occurrence} \tag{5.5}$$

$$E_{r,t}^{re,exp} = \sum_{s=1}^{S} E_{s,r,t}^{re,u} \cdot I_s^{occurrence} \tag{5.6}$$

Linear programming vs mixed integer linear programming formulation

As stated before, generators' operational limitations have an impact on the sizing of the microgrid. To account for this, the model is formulated as MILP, which leads to a higher computational effort than solving the equivalent LP version of the problem. To assess the impact of these constraints, the same model is solved in its LP formulation, and the results are provided in Table 5.12. For the instance LP 18, there is a lower participation of renewable energy due to an decrease in installed PV capacity (- 37.9 %). Battery storage is not installed and its contribution to supply the demand drops from 4.7 % to 0 %. These effects are explained by the increased flexibility of the genset, which is now able to operate at lower power output and has a slightly larger installed capacity than the MILP formulation. For the LP renewable 18 instance, similar values are found if it is compared to the MILP renewable 18 instance. This is explained by the higher share of PV and battery capacity, and thus by the decreased importance of the generator in these simulations.

TABLE 5.12: Results of the LP optimizations.

Variable	LP 18	LP Renewable 18
PV rate (kW)	10.6	56.1
Battery rate (kWh)	0	357
Generator rate (kW)	41.1	13
NPC (Thousand USD)	210	412
LCOE (USD/kWh)	0.27	0.53
Curtailment (%)	0.1	0.8
Renewable penetration (%)	13.8	70
Battery Usage (%)	0	38.6
Solving Time (s)	142	144

5.6 Conclusions

In this chapter, historical operation data from the first isolated microgrid for rural electrification in Bolivia (“El Espino") is analyzed. A suboptimal microgrid design and un-flexible control strategy have led the system to curtail 26 % of the total solar energy available. Furthermore, the surveys to the experts on rural electrification show a high degree of uncertainty when forecasting the energy demand on rural microgrids.

An evaluation of and optimization of the control strategy and the energy dispatch is performed to assess the progress margin in terms of control strategy. Because of a non-optimal control in the actual system, the generator is forced to start at times of low demand, which leads to significant energy curtailment during daytime. This situation could be easily solved with the implementation of a more advanced control strategy.

To evaluate the capabilities of the model presented in Chapter 4, a new sizing of the existing system is realized with different objectives, technological parameters and methodologies. Overall, the optimal design results from a compromise between

NPC, peak installed capacity and flexibility (to balance variable generation). In low-demand scenarios, the microgrid mostly relies on renewable energy and batteries to cover the energy consumption of the village. In high-demand scenarios, gensets are used more intensively and the battery bank provides the needed flexibility to the microgrid. Using the original cost data, the optimal sizing leads to similar installed capacities for PV and batteries than in the real system. However, the installed generator appears to be significantly oversized for the demand. One or two lower-capacity units would lead to improved part-load efficiencies and increased microgrid flexibility.

Different approaches to size isolated microgrids are tested, with the conclusion that methods accounting for the uncertainty in the demand and renewable generation lead to a more robust configuration and also impact on the final levelized cost of electricity. The results also indicate that model formulation plays a key role, e.g. by overestimating the flexibility of diesel gensets in an LP framework, and that the limitations of all simplifying hypotheses should be well understood when sizing such a system.

Finally, the following lessons learned can be extracted:

1. MILP-based stochastic optimization models generate the most robust system configuration.
2. LP models provide good estimates of the cost of the system, but tend to overestimate the flexibility of generators, leading to lower installed capacities of PV and batteries. This effect can be minimized if the system is designed for high renewable penetration.
3. Deterministic system design consumes less computational resources than its probabilistic counterpart, but provides less reliable system designs. It also underestimates the costs since the expected demand does not have high power peaks as the higher energy consumption scenarios.
4. In all cases, the genset nominal capacity should be lower than the peak demand.

Rural electrification planning is a challenging task due to the uncertain parameters and to the particular context in which each project is carried out. For this reason, this work proposes a methodology to size the system using only the available, partial and uncertain data at hand. This data is deemed representative of the available information from practitioners around the world, as they usually only have access to limited demand time series and rely on their own knowledge of the context of the location.

Chapter 6

A bi-objective optimization approach for rural solar home system sizing

The content of this chapter is based on the following publication:

Soto, A., **Balderrama, S.**, Cardozo, E., Miguel, F., Jaime, Z., & Qouilin, S. (2021). Exploring the Trade-off between Installed Capacity and Unserved Energy in Rural Electrification. In ECOS 2021-Proceedings of the 34st International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (pp. 164-164). ECOS Association

In the scientific literature, multiple case studies can be found that consider solar home systems (SHS) covering almost 100 % of the total energy demand. This typically leads to a high LCOE due to the required oversizing of the batteries to cover worst-case periods of low solar irradiation throughout the year [102]. On the other hand, the energy non-served can have a significant impact on the household activities [103, 104] depending on its frequency and duration, thus jeopardizing the objective of universal access to energy [105].

Lost load probability (LLP) and LCOE are two antagonist objectives: reducing the former requires larger investments and therefore reduces the latter. The trade-off between LLP and LCOE of the system can be generalized as a multi-objective optimization problem. It can be visualized as a pareto front, where the trade-off between variables is clear. The selection of the final design point does not have a formal definition and often relies on subjective or non-analytical methods [106]. While this holds true, it seems reasonable to select a point close to the bulge of the curve (knee point) [107], as indicated in Figure 6.1. The target design point corresponds to a region in where the change of one target does not yield excessive changes on the other.

This chapter exposes the relationship between economic indicators (LCOE) and the Loss of Load Probability (LLP) at the moment of sizing a SHS for rural electrification. A method to choose a design point from the pareto front (LCOE vs LLP) is presented and will be re-used in the next chapters when comparing competing technologies in

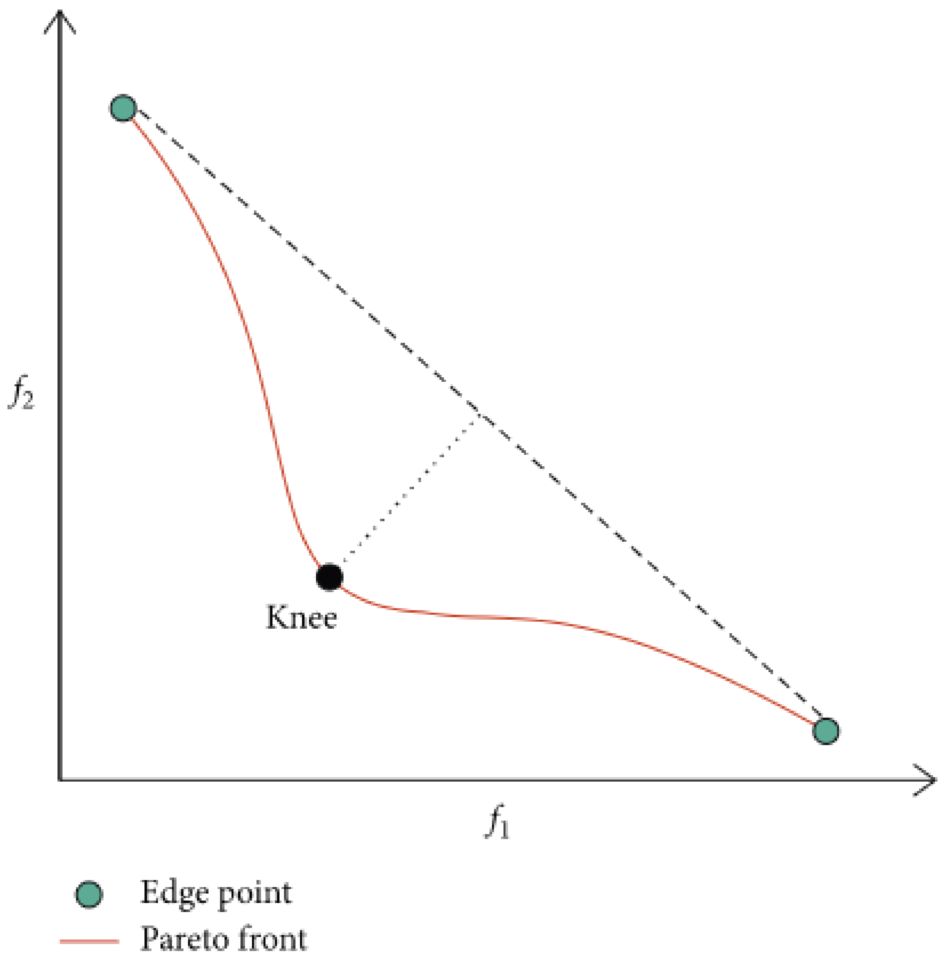

FIGURE 6.1: Pareto front with the knee point, taken from [108].

rural electrification applications. The implications of the chosen system for rural households are finally briefly discussed.

6.1 Methodology

The proposed methodology (Figure 6.2) considers the influence of the non-served energy in relation to the activities of each household. To that aim, socio-economic information is first gathered through different channels. This information is used to create energy demand time series for households, following the approach presented in chapter 3. Other important inputs include PV generation time series and the techno-economic data of each component present in the SHS.

The sizing optimization framework presented in Chapter 4 is then used to determine the PV/battery capacities that minimize the total cost of the system. Because individual systems are simpler than community-scale microgrids, the following simplifications with respect to the original framework are considered:

- The use of diesel is prohibited by enforcing a 100 % renewable system using equation 4.19.

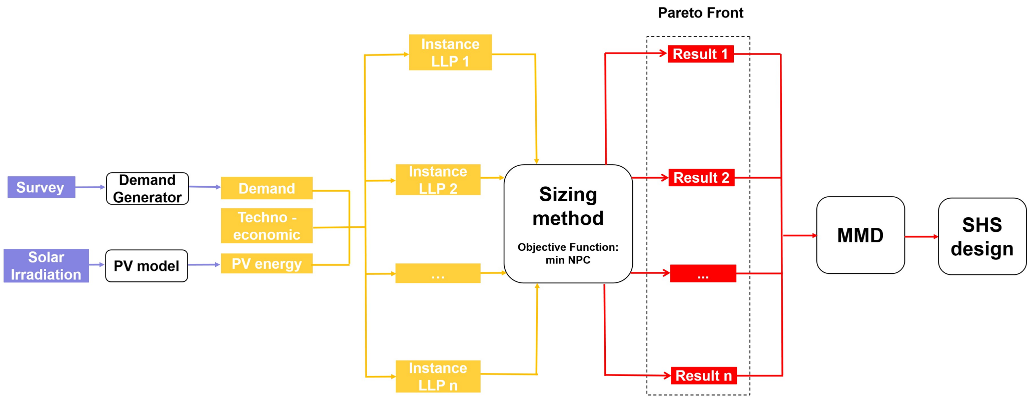

FIGURE 6.2: Proposed methodology for sizing SHS for rural applications.

- The two-stage stochastic LP version of the model is selected.

In that case, the main energy balance constraint (Eq. 4.17) for scenario s simplifies into:

$$D_{s,t} = E^{pv}_{s,t} - E^{bat,ch}_{s,t} + E^{bat,dis}_{s,t} + E^{LL}_{s,t} - E^{Curtailment}_{s,t} \tag{6.1}$$

Since there are no fuel costs and no operation costs are considered, the objective function is the total investment cost (Eq. 4.24), which simplifies into:

$$Inv = U^{PV} \cdot C^{PV} + Fix^{pv} + U^{bat} \cdot C^{bat} + Fix^{bat}s \tag{6.2}$$

As mentioned before, the problem to be solved can be seen as bi-objective: the first objective is the quality of the provided energy service is computed through LLP; the second objective is the cost for the electricity supply and is computed through the LCOE. To highlight the trade-off between these two competing objectives, a pareto front is built by running the optimization multiple times under different LLP targets (Figure 6.3). It appears that providing highly reliable energy services (LLP=0) results in prohibitively high LCOE, which would prevent the system from being deployed. As a consequence, a non-negligible amount of non-served energy should be allowed in such such systems, and will most likely be higher than those encountered in microgrid systems or in the central grid.

Determining an acceptable ratio of energy not served (or its counterpart, the lost load probability) is not trivial and almost inevitable involves arbitrary decisions. In this work, in order to minimize the level of arbitrariness, a methodology is proposed to define a trade-off point between cost indicators and LLP through the Maximum Distances Method (MDM) [109], described in the next section. The trade-off point is further referred to as ”Knee Point”.

6.1.1 Determining the knee point

The MDM method objective is to choose the longest segment of the possible perpendiculars to a line given by two points of the sizing process ($LLP_1, LCOE_1$ and $LLP_2, LCOE_2$). The length of the line is limited by the fitted pareto curve and the line

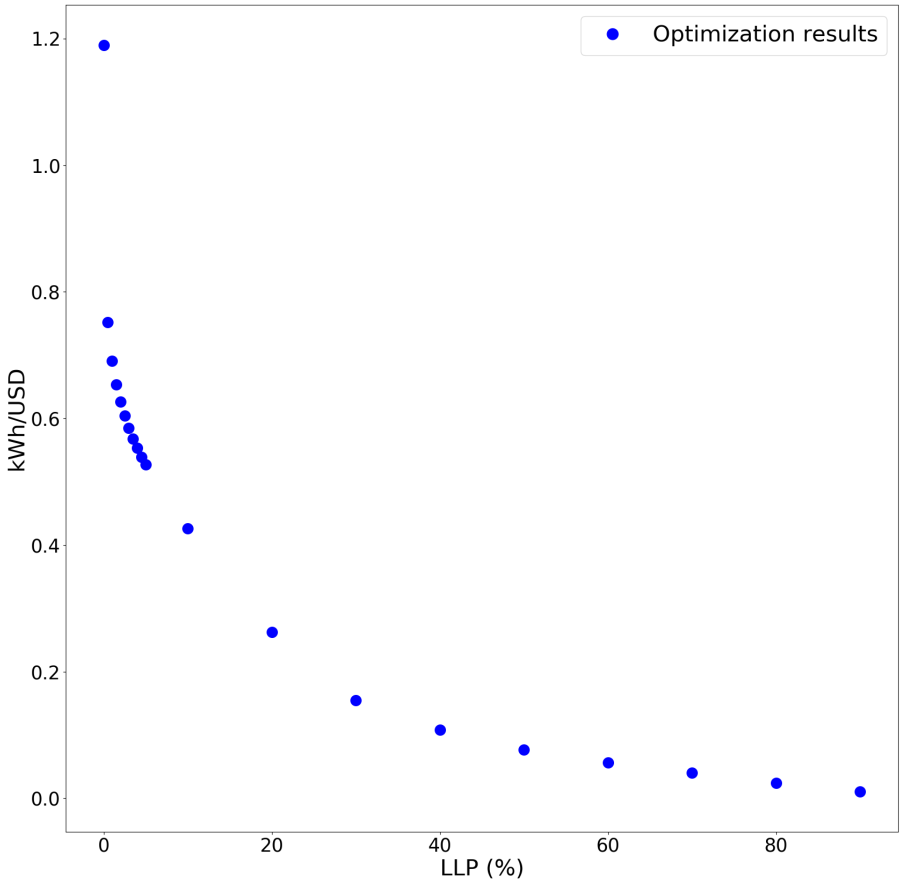

FIGURE 6.3: Pareto front with the results from the sizing method.

formed by the chosen points. Figure 6.4 describes the algorithm implemented for this approach.

First, the variability of the economic indicator (LCOE) from the sizing model as a function of the LLP is analyzed by fitting (Equation 6.3 to the pareto-optimal design points. The quality of the regression is quantified using the coefficient of determination (R^2), defined in equation 6.4. In that equation, $f(x)$ corresponds to the prediction and y is the real target value, N is the number of sizing data points and $\bar{y}$ is the average value of the independent variable.

$$LCOE = \frac{a}{(LLP + d)^b} + c \tag{6.3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - f(x_i))^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{6.4}$$

The extreme points of the pareto front (LLP_a, $LCOE_a$ and LLP_b, $LCOE_b$), are chosen

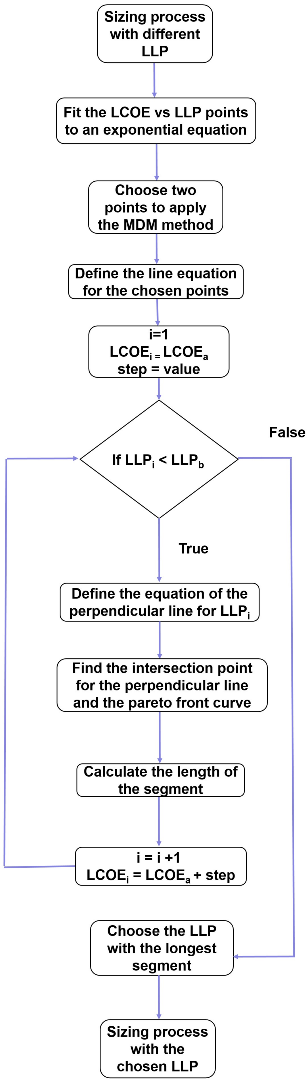

FIGURE 6.4: MDM algorithm.

to define a line (equation 6.5) whose slope (b_s) is calculated in equation 6.6. The offset a_s is then solved in equation 6.7 in terms of LLP_1 and $LCOE_1$.

$$LCOE = a_s + b_s \cdot LLP \tag{6.5}$$

$$b_s = \frac{LCOE_a - LCOE_b}{LLP_a - LLP_b} \tag{6.6}$$

$$a_s = LCOE_a - LLP_a \cdot b_s \tag{6.7}$$

A family of lines, represented by equation 6.8, can be drawn perpendicular to equation 6.5. The slopes (b_p) of these lines are constant and they are reciprocal negative values (equation 6.9) of b_s. With this information, an iterative process is defined to calculate the length of each segment circumscribed by equations 6.3 and 6.8. A step value ($step$) is defined, $LLP_i = LLP_a$ is set and the following steps are is carry out iteratively:

$$LCOE = a_i + b_p \cdot LLP \tag{6.8}$$

$$b_p = -\frac{1}{b_s} \tag{6.9}$$

- The interceptor of the perpendicular line is calculated by combining 6.5 and 6.8 as shown by equation 6.10.
- The value of $LCOE_i$ where the equation 6.5 and 6.8 intersect is found by using LLP_i in either equation.
- The point (LLP_{ii},$LCOE_{ii}$) where equation 6.3 and 6.8 intersect is found through a numerical approach as implemented in [110].
- The segment length is found using equation 6.11.
- $LCOE_i$ is updated by using equation 6.12.

$$a_p = b_s + (b_s - b_p) \cdot LLP_i \tag{6.10}$$

$$d_j = \sqrt{(LCOE_i - LCOE_{ii})^2 + (LLP_i - LLP_{ii})^2} \tag{6.11}$$

$$LCOE_i = LCOE_i + step \tag{6.12}$$

The process is carried out until $LCOE_i > LCOE_b$. At this point, the whole space has been explored and a system configuration can be selected. The LLP_i corresponding to the longest segment is chosen and it is used in the sizing model to calculate the final installed capacities and energy flows.

6.2 Application to the Bolivian case

A large-scale electrification program faces several logistic problems, including the purchase, assembly and transport of thousands of SHS. This plus the Bolivian heterogeneous cultural and geography characteristics, makes it impossible to size an optimal system for each household, community or region of Bolivia. Under these circumstances, the LP two-stage stochastic sizing model is used to design a generic system capable of meeting different demands under variable PV energy outputs and LLP scenarios. The final capacities are defined as the knee point obtained from the stochastically-generated LCOE/LLP curve (Section 6.1).

6.2.1 PV time series

The solar energy yield is highly dependent on the location since it is a result of the latitude, cloud cover and other climatic or geographic characteristics in the region.

For this reason, different time series are extracted from *Renewable.ninja* for the coordinates of the considered Bolivian village [111]. The selected year is 2012 and the tilt angle is set equal to the latitude. The conversion from solar irradiation to power is simulated by assuming a commercial PV model available over the whole territory (YL250P-29b) and applying a five parameters model as implemented in [98].

6.2.2 Demand time series

The demands were defined in chapter 3, with a total of 4 load curves of HI and LI population of each region in Bolivia. They are paired with 16 PV energy time series randomly chosen from Bolivian communities locations. The techno-economic parameters are shown in table 6.1 and the LLP is varied in steps 0.05 % from 0 to 5 % and steps of 10% from 10 % to 100 %.

TABLE 6.1: Techno-economic parameters for the MILP optimization.

Parameter	Unit	Value
Periods (t)	hours	8760
Project life time (y)	years	20
Δt	hours	1
e	%	12
U^{LL}	USD/kWh	0
U^{bat}	USD/kWh	550
U^{elec}	USD/kWh	220
$I^{Bat,OM}$	%	2
η^{dis}	%	95
η^{ch}	%	95
$\Delta t^{bat,dis}$	hours	4
$\Delta t^{bat,ch}$	hours	4
DOD	%	20
Cy^{bat}	cycles	5500
C^{re}	kW	0.25
η^{re}	%	95
U^{re}	USD/W	1500
$I^{re,OM}$	%	2
$U_r^{bat,fix}$	USD	0
$U_r^{bat,fix}$	USD	0

6.2.3 Results

Once the different instances of the SHS have been defined, the sizing model can be run. The results are presented in Figure 6.5 and table 6.2.

The regression of the pareto front describing the functional relationship between LCOE and the probability of the system not being able to supply a fraction of the required energy (LLP) is achieved with a determination coefficient higher than 0.99. Figure 6.5 then shows the selection of the knee points for the proposed scenario. The points $LPP = 0\%$ and $LLP = 100\%$ are chosen to parameterize equation 6.5. The iteration step for the maximum length search is set to 0.1 %.

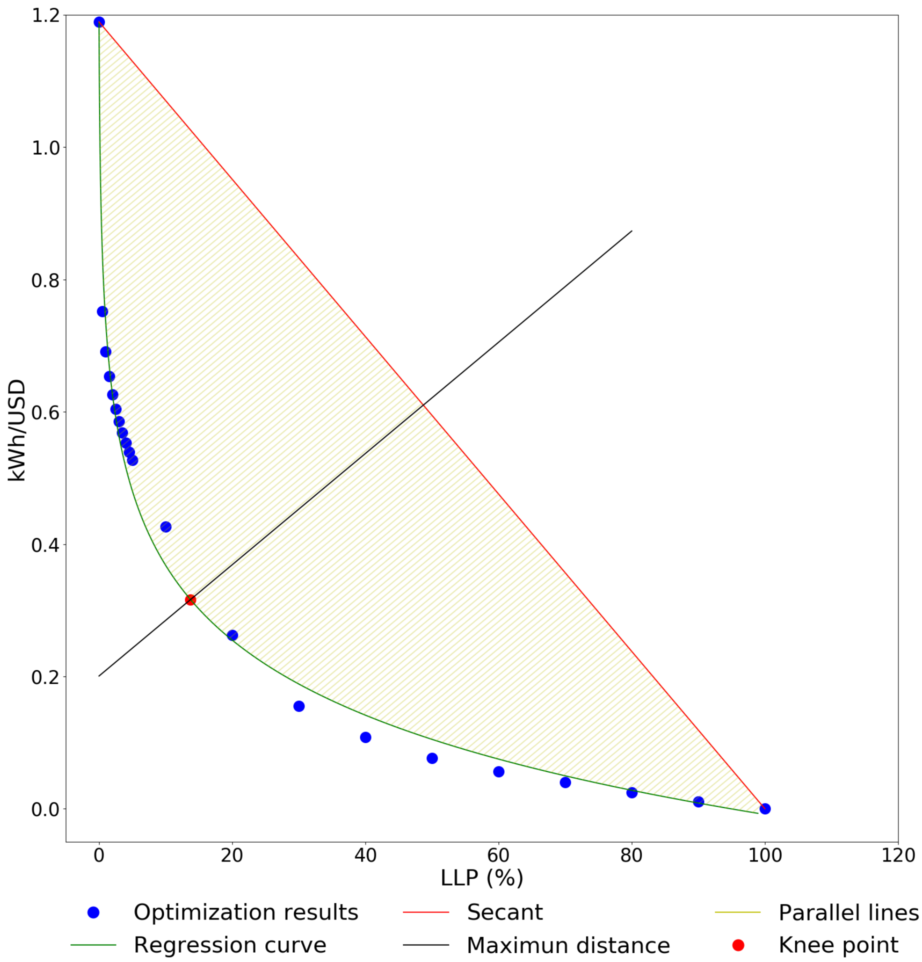

FIGURE 6.5: Left; Nonlinear adjustment of the variation of NPC as a function of LLP, graphical representation of MDM.

Table 6.2 provides the LLP at which the SHS covers the largest fraction of demand without compromising investment in system capacity. In this case a PV of 0.46 kW and a battery capacity of 1.16 kWh are part of the optimal system for households in Bolivia. It is important to note that, although the LCOE is high in comparison to the ones calculated in chapter 5, this system does not provide energy to all the activities in the household or for public services. The difference in price can be explained by the presence of the genset in microgrids, which allows smaller PV and Batteries installed capacities.

With the critical point obtained in the previous section, the SHS is considered to be sized at the point where the amount of energy supplied does not significantly affect

TABLE 6.2: Results of the sizing process.

	NPC (USD)	LCOE (USD/kWh)	Battery Capacity (kWh)	PV Capacity (kW)
Mean	1814	0.40	1.39	0.48
Standard deviation	1453	0.32	1.13	0.39
Max	5361	1.19	3.93	1.57
Min	0	0	0	0
Knee point	1633	0.36	1.16	0.46

the LCOE of the system. It is then relevant to analyze which appliances might be affected by the inability of SHS to supply the total of required demand.

Figure 6.6 shows the locations of the energy non served in a high consumption scenario and demonstrates that the latter is significant during night time. In contrast, most of the energy for LI households are met with this configuration as shown in Figure 6.7. From all the present appliances (Table 3.2), the refrigerator is the most critical with respect to the security of supply. Energy not served can lead to problems on the safety, quality and organoleptic properties of the store food. At the same time, the thermal inertia of the equipment can ensure that the temperature range stays in the limits for a certain period. A detailed analysis of these thermal effects is out of the scope of the present work, but would deserve specific attention in future works.

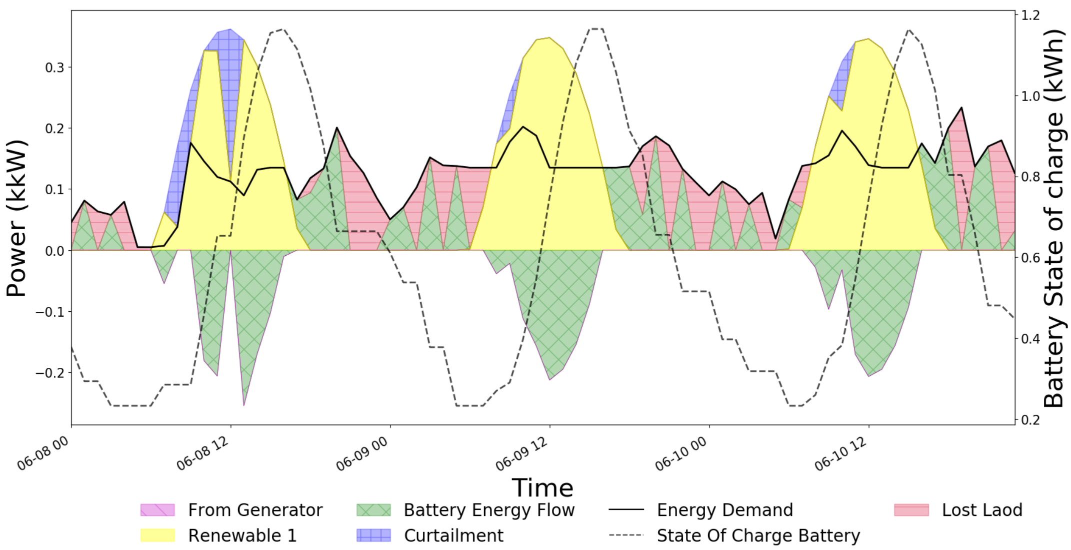

FIGURE 6.6: Energy dispatch for the knee point sizing in a HC households, in the days with peak demand.

6.3 Conclusion

In this chapter, a methodology to size a SHS for the remote population of a developing country taking into account the trade-off between non-served demand and the LCOE of the system is presented. For this purpose, electrical appliances available in rural households are considered through 16 stochastic scenarios. Through a non-linear regression, the relationship between the LCOE of the SHS and LLP is

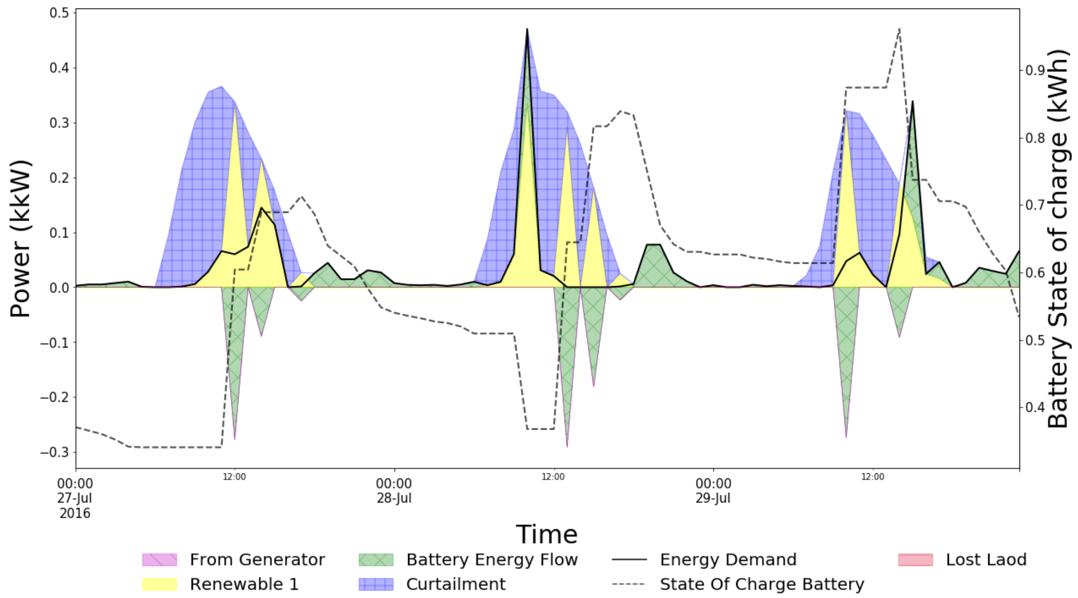

FIGURE 6.7: Energy dispatch for the knee point sizing in a LC household, in the days with peak demand.

determined after running multiple instances of the optimal sizing problem. Using geometric and numerical techniques, the knee point of the curve is determined and a SHS optimal design for rural electrification in Bolivia is proposed.

For the analyzed case study, it was found that the relationship between LCOE of the SHS and LLP has a general form that can easily be fitted by a first-order inverse function. This is verifiable by a coefficient of determination greater than 0.99. In order to make an acceptable trade-off between the energy cost and the energy supply, this final system is designed with the capability to provide an average of 86.3 % of the expected demand of rural households in Bolivia, under different conditions.

The practical implications of these findings demonstrate how difficult it is to reach energy sufficiency without backup sources such as diesel gensets. In contrast, it is clear that SHS allow to reach a minimal access for Bolivians living in the more isolated areas in the country. Although the real usage of these energy might diverge from the optimal dispatch calculated by the solver, these systems can arguably cover a significant share of the household demand and therefore cover most of the basic needs.

Part III

Optimal deployment of isolated energy systems

Chapter 7

Rural energy planning using Geographical information systems

This chapter content is largely based on the following publication:

*Peña, J., **Balderrama, S.**, Lombardi, F., Stevanato, N., Sahlberg, A., Howells, M., Colombo, E., & Quoilin, S. (2020). Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET). Energy for Sustainable Development, 56, 98-118.*

7.1 Introduction

Closing the electrification gap in developing countries while addressing the climate change goals is a complex task. The lack of information, high upfront investment costs, social challenges and planning capacities, between others, makes this a formidable challenge. Under this paradigm, Energy systems optimization models are typically adopted to support policy decisions, and their usage underwent a rapid increase in the past years [112].

However, as noted by Pfenninger et al. [113], several research gaps need to be addressed for energy modelling to provide effective support to meet global objectives. A key issue is the high complexity required by accurate and comprehensive representations of future energy systems, combined with the need to ensure computational tractability. Such trade-off between technical detail and computational tractability particularly emerges when evaluating multiple smaller-scale systems, such as micro-grids or stand-alone PV systems, within the broader picture of a country-wide power system. From the pool of available mitigation technologies, hybrid microgrids, either connected or disconnected to the main grid, offer an alternative to reduce GHG by harnessing locally-available renewable resources. This, in addition to their modularity and capacity to adapt to a specific context [27], makes them a key technology for the energy transition. Yet, despite their multiple advantages, their exact role is still to be clearly assessed and quantified. In the framework of rural electrification, their cost-competitiveness against grid-extension depends on a range of factors, such as the degree of energy access to be achieved, population density, local grid characteristics and local resources availability [68]. Different tools have been developed to determine, for a given country, the optimal mix of technologies to achieve full electrification, deciding between PV home systems, microgrids and grid extension. Such tools typically combine geospatial data and power system modelling to find the least-cost technology solutions to achieve universal access to electricity.

For instance, Ellman [114] developed an optimization tool (REM) with high spatial granularity that allows the evaluation of household consumption levels based on geospatial data. The resulting model is computationally-expensive since the grid extension is performed at household level, instead of a community level. This leads to an important increse in the quantity of individual calculations performed. OnSSET (the OpeN Source Spatial Electrification Toolkit) [9] is another electrification planning tool which finds the least-cost path to country-scale full electrification based on a limited, easy-to-gather set of input information. More precisely, the tool estimates plausible demand figures for each location relying on proxies such as night lights, road proximity and other GIS data. It compares and chooses the least-cost electrification alternative (between standalone systems, microgrids and grid extension) for each community, based on simplified cost functions for each category. Unlike REM, OnSSET focuses on solutions at a community level with a strong focus on limiting the CPU times. Cader et al. [115], in the context of the NESP project (Rural electrification modelling in the framework of the Nigerian Energy Support Program), developed a tool that includes the possibility to model hybrid microgrids at hourly resolution throughout an entire year. This approach optimizes each community individually and therefore leads to high computational resource usage when used for rural electrification planning.

From the available GIS-based tools for simulating the total cost of electrification in a developing country, OnSSET was selected in this work for its openness and adaptability [9]. The open-source license allows the implementation of new features in the source code, in an efficient and transparent manner. This chapter describes some aspects of the OnSSET electrification algorithm together with the implementation of new features aiming at improving the representation of isolated microgrids.

7.2 The OnSSET electrification algorithm

The OnSSET algorithm minimizes the cost of reaching 100 % of electrical coverage in a country. To that aim, it considers both the extension of the main grid and the off-grid solutions, i.e. microgrids and individual solar home systems (Figure 7.1). The fuel costs (most likely diesel) are first calculated in each community, taking in account the distance from the supply location. Then, the Levelized Costs of Electricity (LCOE) for all off-grid solutions are computed from the annual load, the peak load and the capacity factor of the analyzed technology. The LCOE of the grid densification and extension are further computed by summing the cost of extending the low, medium and high voltages lines. Finally, the lowest-LCOE technology is selected for each community and relevant outputs are computed, such as as the installed capacities or the total investment per community.

The above-mentioned technology options are economically efficient in different settings. Grid extension is advisable in areas close to existing transmission infrastructure, where electricity demand makes economic sense. Microgrids are often cost-effective in settlements outside the reach of the grid, with a sufficient density and diversity of users that is more cost-effective to connect together than supplying each

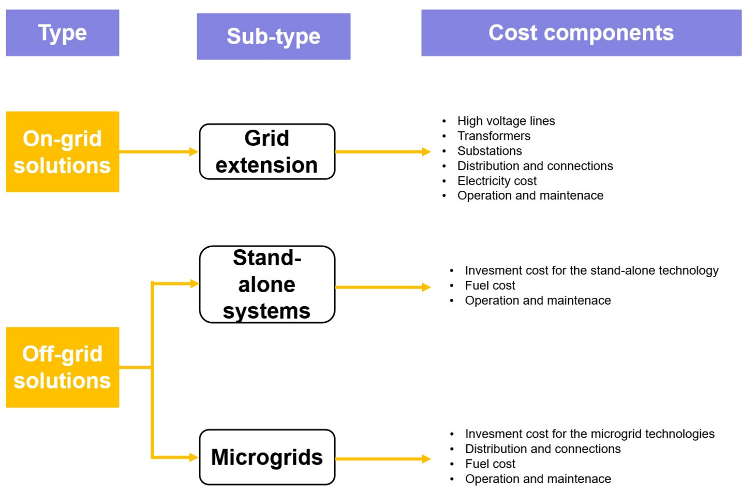

FIGURE 7.1: Taxonomy of OnSSET electrification alternatives, adapted from [116].

user with standalone systems. Lastly, standalone systems are the most cost-effective electrification solution for remote and low populated areas offering limited but life-changing electricity service. The economics of a technology option in a given settlement depend on site-specific characteristics such as demand, distance to the grid, renewable potential and added transportation costs to diesel price. This information together with other technology-specific data can be used to determine the LCOE of implementing various electrification options to supply identified electricity needs. The least-cost alternative that provides desired attributes on peak capacity and reliability is advised for investment.

7.2.1 Grid extension

The LCOE for grid-extension comprise the cost of electricity generation from the grid-connected power plants and the marginal cost of extending transmission and distribution lines [117]. The algorithm examines where it is less costly to extend the grid by medium voltage (MV) lines comparing to deploying off-grid technologies for each un-electrified settlement located within 50 km from the existing and planned high voltage (HV) network [117]. This iterative process determines if the connection of one settlement may lead to the cost-effective connection of neighboring settlements (all within a 50 km limit from HV lines) [117]. Extensions by MV lines for distances longer than 50 km may be limited by techno-economical aspects that are not considered in this model [118]. A comprehensive description of the sizing algorithms for HV and MV transmission lines, transformers and connections of the substations is provided in [118].

7.2.2 Microgrids

OnSSET evaluates demand on a yearly basis for specified household consumption levels, but it does not differentiate demands from small and large populations, which often present substantial differences in the load demand profiles. Contrary to stand-alone technologies, microgrids include a distribution network in the settlement. The length of the distribution network is determined with the settlement area, electricity demand and peak power demand [119]. Only single-source technologies are modeled, i.e. solar PV-battery and diesel-only microgrids. Microgrids are sized with a simple energy balance to meet an average peak demand using annual data on demand and renewable resources availability.

As previously mentioned, since no detailed reliability considerations are included in the sizing algorithms, intra-day and intra-seasonal effects are not captured. For PV-battery microgrids, OnSSET estimates the generation capacity required (PV panel) but does not explicitly calculate the size of the battery. Investment costs include the battery cost, assumed proportional to the generation capacity. Compared to diesel-only, solar-only microgrids require large batteries to supply electricity with comparable reliability.

7.2.3 Standalone systems

For small domestic consumers, standalone solar PV and diesel generators are often the most cost-effective solution in terms of total investment. These electrification technologies provide a few hours of essential electricity service to power small appliances. However, standalone systems cannot provide electricity with comparable reliability to micro-grid and grid-connected systems. In OnSSET, associated costs for non-served energy are not assessed in standalone systems. The LCOE of PV-standalone and diesel-standalone use location-specific data on annual solar irradiation and diesel costs, respectively.

7.2.4 LCOE Calculation in the original OnSSET algorithm

OnSSET has three distinct families of electrification technologies (standalone, microgrid, grid extention) and the difference in cost between them can be explained with the presented equations and the particularities of each technology (Figure 7.1). All the technologies operate almost in the same way: by ensuring that the installed capacity meets the peak demand and assuming the system can meet all the energy required in a year.

The classical OnSSET algorithm calculates the LCOE of each community by calculating a net present cost (NPC) and a levelized cost of electricity (LCOE) as stated in equation 7.3. Depending on the technology, this value includes different components (Figure 7.1). The most important characteristics are summarized in Table 7.1 for the four originally-considered technologies.

In Table 7.1, C_{HV}, C_{MV}, C_{LV} are the investment costs for the high, medium and low voltage grids, including the sub-stations and transformers. D_{max} is the peak load, D_{tot} is the total annual energy demand, d_{grid} is the shortest distance to the existing electric grid, N_{hh} is the number of households in the community, A is the area of the considered zone, CF is the capacity factor of the considered technology, loc is the location (latitude and longitude), Cap_{gen} is the installed capacity of the considered technology (in MW), C_{gen} is the investment cost, P_{fuel} is the fuel price in USD/MWh (the retail price of electricity in case of grid extension), C_{inv} is the total investment cost ($C_{inv} = C_{HV} + C_{MV} + C_{LV} + C_{gen} - C_{salvage}$), $C_{salvage}$ is the discounted value of the salvage cost after the project lifetime, C_{fuel} is the annual fuel cost and $C_{O\&M}$ is the annual operation & maintenance cost. $T_{project}$ and T_{tech} are the lifetimes of the project and of the installed technology, respectively.

TABLE 7.1: Calculation of the main costs components for each considered technology in the original OnSSET algorithm

	SHS	Diesel Microgrid	PV Microgrid	Grid extension
C_{HV}	0	0	0	$f(D_{max}, d_{grid})$
C_{MV}	0	$f(D_{tot}, D_{max}, A)$	$f(D_{tot}, D_{max}, A)$	$f(D_{tot}, D_{max}, A, d_{grid})$
C_{LV}	0	$f(N_{hh}, D_{tot}, D_{max}, A)$	$f(N_{hh}, D_{tot}, D_{max}, A)$	$f(N_{hh}, D_{tot}, D_{max}, A)$
Cap_{gen}	$\frac{D_{max}}{CF_{PV}}$	$\frac{D_{max}}{CF_{fuel}}$	$\frac{D_{max}}{CF_{PV}}$	0
CF	$f(loc)$	$Cste$	$f(loc)$	$Cste$
C_{gen}	$f(Cap_{gen})$	$f(Cap_{gen})$	$f(Cap_{gen})$	0
P_{fuel}	0	$f(P_{ref}, t_{travel}, C_{truck})$	0	$Cste$
C_{fuel}	0	$D_{tot} \cdot P_{fuel}$	0	$D_{tot} \cdot P_{fuel}$
$C_{O\&M}$	$\alpha \cdot C_{inv}$	$\alpha \cdot C_{inv}$	$\alpha \cdot C_{inv}$	$\alpha \cdot C_{inv}$
$C_{salvage}$	0	0	0	$f(T_{project}, T_{tech})$

The total investment C_{inv} has two components: the power generation technology and the grid costs. Depending on the technology life expectancy, (re-)investment can occur several times during the electrification process lifetime. The Capital investment is calculated in equation 7.1 in function of the yearly peak load and of the capacity factor of each technology. The installed generation capacity is calculated according to Table 7.1, where D_{max}, the peak load, is calculated according to Equation 7.2. The investment in the grid depends on different factors described in 7.2.1. The operation and maintenance is a yearly cost, calculated as a percentage of the total investment cost.

$$C_{gen} = Cap_{gen} \cdot CapCost_{gen} = \frac{D_{max}}{CF} \cdot CapCost_{gen} \tag{7.1}$$

$$D_{max} = \frac{D_{tot} \cdot (1 - \xi_{dist})}{r_{base,peak}} \tag{7.2}$$

where ξ_{dist} is the distribution losses factor and $r_{base,peak}$ is the base to peak load ratio.

The LCOE for each technology is finally computed by:

$$LCOE = \frac{C_{inv} + \sum_{y=0}^{n} (C_{O\&M} + C_{fuel}) / DF_y}{D_{tot} \cdot T_{project}} \tag{7.3}$$

where $DF_y = (1 + d)^y$.

To exemplify the above calculation, the computed values with the original OnSSET algorithm are provided for the particular case of "El Espino", described in Section 5, and assuming that no microgrid has been installed yet. The values correspond to the expected values for 2025, with the following main characteristics:

- Number of households: $N_{hh} = 142$
- Population: 455
- Peak Load: $D_{max} = 9.2kW$
- Yearly demand: $D_{tot} = 42809kWh$
- Project lifetime: $T_{project} = 15years$

- Traveling time to the community (for diesel costs): $Time_{travel} = 3h$

community comprises $N_{hh} = 142$ household, with a maximu

Table 7.2 displays the breakdown of the cost calculations. Due to the relative proximity of the community to the main grid, it appears that no new high voltage lines are required. This results in low investment costs for the grid extension and therefore gives a clear advantage to this technology compared to solar home systems or to microgrids. Grid extension will therefore be selected a the preferred technology for the "El Espino" community.

TABLE 7.2: Example calculation for the "El Espino" Community and the original OnSSET algorithm

	SHS	Diesel Microgrid	PV Microgrid	Grid Extension
$C_{HV}(USD)$	0	0	0	0
$C_{MV}(USD)$	0	0	0	10000
$C_{LV}(USD)$	0	92561	92561	92561
$Cap_{gen}(kW)$	45.1	13.2	45.1	0
$CF(-)$	20.4%	90%	20.4%	0.9
$C_{gen}(USD)$	228652	19494	157846	0
$P_{fuel}(USD/kWhe)$	0	0.61	0	0.1223
$C_{fuel}(USD/y)$	0	26031	0	5235
$C_{O\&M}(USD/y)$	4573	2241	5008	2051
$T_{tech}(years)$	15	20	20	30
$C_{salvage}(USD)$	0	8988	20085	13319
$LCOE(USD/kWh)$	0.92	1.03	0.91	0.49

7.3 Limitations of GIS electrification tools

Although GIS-based electrification planning tools have grown in popularity in the last years, several research gaps remain [5]. Many of them are related to the representation of microgrids, whose diversity is not properly accounted for, owing to the increased computational burden of simulating/optimizing each system individually. The main challenges for isolated energy systems can be summarized as follows:

1. Existing sizing algorithms that perform detailed optimization can only cope with a limited number of technologies. In the case of the OnSSET algorithm, microgrid technologies include PV-only or diesel-only systems, but do not consider hybrid microgrids in which the relative share of PV and diesel can vary according to the local conditions.

2. Because of the dimensions of the problem, the sizing of each individual system is very simple, as indicated in Table 7.1

3. Most GIS-based tools do not consider the evolution of the system or scenarios that include a hybrid solution in which a microgrid first operates as standalone, and is then connected to the main grid.

4. Many methodologies do not consider uncertainty in the evaluation of future demand levels, weather conditions or other technical parameters.

The Challenges mentioned above are tightly linked to the computational tractability of co-optimizing multiple technologies in a large number of microgrids and for multiple realizations of stochastic scenarios. As shown in chapter 5, even simple optimization models (LP 12) can consume significant CPU time. If there is a need of solving more complex models, more advanced strategies must be devised.

One way of improving the tractability of the problem without compromising the model complexity is to apply machine learning techniques (MLT) to approximate the optimization results. MLT have been successfully used to forecast or simulate different phenomena in energy systems (Mosavi et al. [120]). They can also be used to accurately predict energy consumption, as proven by Yildiz et al. [121].

The use of MLT in the long-term planning of microgrids has historically focused on the forecast of demand and renewable energy time series. However, in recent years, it has also been used to automate decision making and reduce computational effort by creating surrogate models from the results of a high number of optimizations. These surrogate models aim to estimate the value of a particular optimization outcome (e.g. total cost of the project, nominal capacities of the technologies) using the input conditions, as shown by Perera et al. [122]. In the latter study, an artificial neural network (ANN) surrogate model is trained to calculate the net present cost (NPC), grid interaction and unmet load fraction of an energy hub comprising various renewable energy sources, storage devices and internal combustion engines. The surrogate model is then used together with a heuristic optimization method to calculate the optimal nominal capacity of each technology. In another study [123], an ANN is trained on a database created from an operation and planning model at a national scale. The model takes multiple input parameters and returns the nominal capacities of the technologies and other crucial operation variables. The most promising aspect of this methodology is the possibility to change one of the assumptions of the optimization and obtain the new output variables with a low computational cost.

An alternative approach is proposed by Ciller et al. [124], in which a lookup table is constructed with the optimal costs for different communities sizes. Instead of resorting to a regression, each particular values is interpolated from the table. Lookup tables are however cumbersome and require large datasets in multi-dimensional environments. This limitation can possibly be tackled by using MLT methods, trained over a limited number of data samples in the multi-dimentional input space. As an example, Gaussian processes regression (GPR) has been used to estimate the performance of various thermal systems as a function of many input variables (features) with a higher accuracy than physical models. It further allows to perform feature selection and outlier detection, as shown by Quoilin & Schrouff [125]. In a previous work [126], it was shown that GPR are well suited to estimate the Levelized cost of electricity (LCOE) for isolated microgrids in a rural context. Up to 11 hypothetical villages sizes were created based on surveys and on a stochastic load profile generator. In total, 1100 optimizations were performed by varying the capital costs of the different technologies, the diesel cost, the village size and the PV energy output. Peña et al. [116] applied multi-variable linear regressions to calculate the NPC and LCOE for only diesel, PV/battery and hybrid microgrids in a large-scale geospatial electrification planning tool (OnSSET). The study revealed an important increase in the cost-competitiveness of micro-grids compared to previous analyses using simplified micro-grid sizing algorithms.

7.4 Conclusions

In this chapter, a brief description of existing tools and models for rural electrification planning is proposed. The problem is contextualized in regards of the current modeling capabilities, where GIS tools have proven to be a viable methodology to find cost efficient solutions to the problem. The limitations of these tools are also presented and mainly relate to the accurate representation and optimization of many dispersed systems without compromising the tractability of the problem.

Finally, the creation of surrogate models for microgrid sizing is suggested as a viable solution to decrease the computational time of the planning tools while maintaining a high accuracy in the estimations of the needed design parameters. This option will be explored in the next chapters.

Chapter 8

Surrogate model creation and validation

This chapter is largely based on the following publication:

***Balderrama, S.**, Lombardi, F., Stevanato, N., Peña, G., Colombo, E., & Quoilin, S. (2021). Surrogate models for rural energy planning: Application to Bolivian lowlands isolated communities. Energy, 121108.*

8.1 Introduction

As seen in previous chapters, sizing energy systems may require high computational resources depending on the objective of the process and on the available data. For this reason, there a trade-off must be found between technical detail and computational tractability when evaluating multiple smaller-scale systems within the broader picture of the planning process at the country level. This chapter builds upon the idea of training machine learning models to improve computational tractability by directly predicting the optimal design of isolated energy systems. The specific objectives can be summarized as:

1. To propose a standard training methodology for surrogate models capable of predicting the optimal microgrid design and cost as a function of multiple boundary conditions.
2. The inclusion of technical parameters (e.g. PV and battery capacities) as explanatory variables of the predicted LCOE.
3. A comparison of the performance of different MLTs over the same dataset.

8.2 Methodology

To develop and validate surrogate models for energy planning in a rural context, the studied system should first be defined. In this work, an isolated microgrid system is considered, composed by a PV array, a solar inverter, a battery bank, a bi-directional inverter and a diesel Genset. The system is designed to cover the whole electricity demand of a given community. In case of energy surplus, the batteries can be charged by the PV array or the Genset. Although the proposed system is relatively basic, it is important to mention that the proposed methodology can be applied to more complex systems with multiple renewable sources, combustion generators, connected or not to the main grid.

In Figure 8.1, the information flows and the most important tools implemented to create the surrogate models are shown. The demand of the village, the energy yield of the PV array and the techno-economic information constitute one optimization instance. The sizing method is used to determine the nominal capacities of the energy sources and different costs of the system for each instance. The results of each optimization are the dependent variables for the regression model. Using these input variables (features) and the selected dependent variables (targets), the regression process is carried out by tuning the hyper-parameters of the MLT model and computing some numerical performance indicators.

The surrogate models developed with this methodology will further be integrated into other energy models in the next chapter, with the objective to answer broader questions related to energy planning at a regional, national or trans-national level.

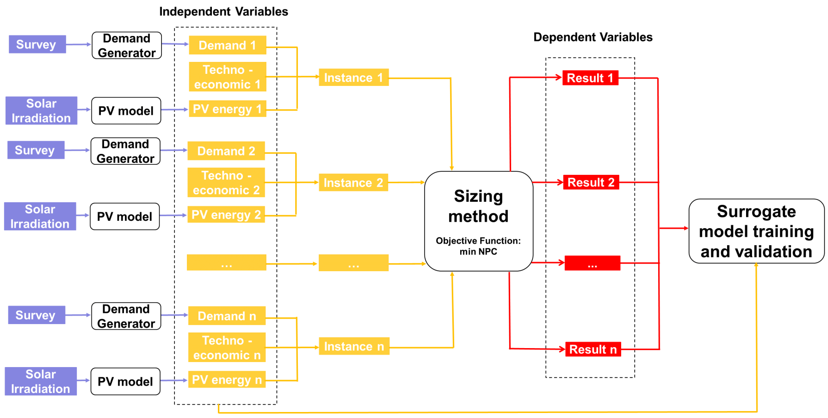

FIGURE 8.1: Methodology for the creation of the surrogate models.

Machine learning methods are divided into classification methods, which focus on dividing a data set into groups; and regression methods, which aim at creating the mapping function between one or more input variables (features) and a output variables (targets). In our specific case, the goal is to predict the optimal value of the different variables that minimize the NPC for a given set of features. In this work, the targets include both the objective function of the optimization process and some optimization variables such as nominal capacities, lost load in the system, etc.

The machine learning regression (MLR) is applied after running many optimizations over the full range of the input space. The overall process is shown in Figure 8.2 and can be subdivided in three main steps:

1. To ensure a random sampling of the test cases within the dataset, a shuffle technique is applied: the individual optimizations are first run in ascending order of the size of the community; the dataset is then shuffled and divided in folds for the cross validation.

2. For each dependent variable, a surrogate model is fitted using the MLR and the relevant independent variables.

3. The quality of the model is evaluated by computing numerical indicators over the testing set (which differs from the training set).

The first performance metrics is the mean absolute error (MAE), defined as the mean difference between the predicted target $f(x)$ and the real value (y), as presented in equation 8.1, where N is the number of inputs used in the MLR. The second is the coefficient of determination (R^2), computed in equation 6.4. The last indicator is the root mean square error (RMSE) and is defined in equation 8.2. In addition to the ability to predict the target values inside the training set, the model should be able to do it outside of the sampled data. In order to ensure this generalization ability and to avoid overfitting, a K-fold cross-validation method is selected. To that aim, the shuffled data set is divided into K sub-sets (folds) and the training is carried out K times. Each time, one fold is removed from the training set and is used as test set to compute the performance metrics. The MAE, R^2 and RMSE are finally averaged over all folds and reflect the capacity of the model to predict the independent variable for an unseen sample. In this study, different types of MLR are tested, and their hyperparameters are tuned to improve the quality of the regression.

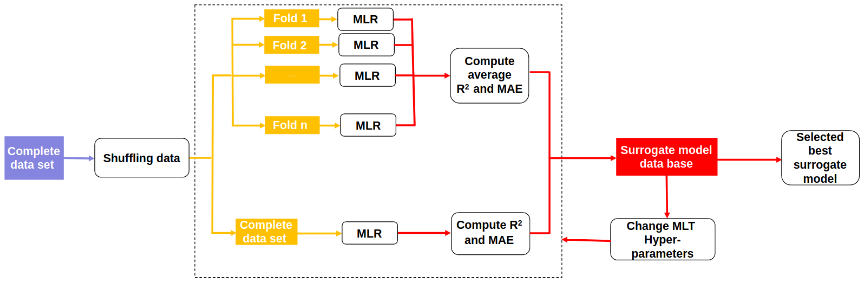

FIGURE 8.2: The methodology implemented for the training and validation of the surrogate models.

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^{N} |y_i - f(x_i)| \tag{8.1}$$

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - f(x_i))} \tag{8.2}$$

8.3 Case study

In this chapter, the model parameters are divided into two sets: unmutable and mutable. The first set contains the ones that do not vary between the different optimizations. These are the techno-economic parameters defined in Table 8.1. The other set can take different values in each instance and relates to some other techno-economic parameters or to the selection of the demand and PV time series. The generation of these stochastic time series is described in chapter 3.

8.3.1 Mutable techno-economic parameters

The challenge of providing clean, sustainable and affordable energy to isolated communities involves selecting the most suitable technology solutions for each situation. This means that, depending on the context, a lead-acid battery can be chosen over

TABLE 8.1: Unmutable model parameters

Parameter	Unit	Value
Periods in a year	hours	8760
Project life time	years	20
Time step	hours	1
Discount rate	%	12
Lost load probability	%	0
Unitary battery electronic cost	USD/kWh	222
Battery operation and maintenance cost	%	2
Battery charge efficiency	%	0.95
Battery discharge efficiency	%	0.95
Battery full discharge time	hours	4
Battery full charge time	hours	4
PV nominal capacity	W	250
PV inverter efficiency	%	97
PV operation and maintenance cost	%	2
Genset operation and maintenance cost	%	2
Minimum genset power output	%	50
Genset penalty cost for part load	%	9
Fixed cost PV/Battery	USD	15 000

lithium-ion or a bio-gas micro-turbine over a diesel unit. The ability to compare different solutions in a fast and reliable way is key for practitioners around the world. In this work, it is proposed to achieve this through surrogate models trained over a large range of usual boundary conditions. For that purpose, the parameters provided in Table 8.2 are varied, combined, and a optimization is run for each selected combination. To avoid intractable computational times, a Latin hypercube sampling (with 150 samples) is applied, covering the whole input space on which the optimization model is run. The variation ranges of each input are detailed in Table 8.2. As it is highly hazardous to perform an estimation of the peak demand due to the high uncertainty in the energy evolution of rural systems [127], the nominal capacity of the genset is set to a percentage of the higher demand in the dataset. Finally, the battery capacity/power output relationship is set to 4 hours.

8.3.2 Machine learning regression methods

The python library *scikit-learn* is selected to build and train the surrogate models [128]. It allows easily defining the optimization problem and includes different state-of-the-art built-in algorithms, which also allows to compare them. For this work, GPR and multi-variable linear regression (MVLR) are chosen to showcase the capabilities of the proposed methodology.

TABLE 8.2: Mutable parameters for the sizing process.

Parameter	Unit	Range
PV investment cost	USD/kW	1000 - 2000
Battery investment cost	USD/kWh	800 - 222
Depth of discharge	%	0 - 50
Battery Cycles	Cycles	1000 - 7000
Generator investment cost	USD/kW	1000 - 2000
Generator efficiency	%	10 - 40 %
Lower heating value	kWh/l	7 - 11
Fuel cost	USD/l	0.18 - 2
Generator Nominal capacity	kW	75 % of the peak demand
Energy Demand	kW	
PV unit energy production	kW	

Multi-variable linear regression

The MVLR is one of the simplest MLR methods to map the function (f) of a output variable (y) based on a set of input parameters (x). It can be described as follows:

$$f(x) = x^{\intercal} \cdot w \tag{8.3}$$

$$y = f(x) + \varepsilon \tag{8.4}$$

where $w \in R^m$ is a vector of weights or parameters of the model. To differentiate the observed values (y) from the target values ($f(x)$) an error term (ε) is introduced with an independent identically distributed Gaussian distribution with zero mean and variance σ_n^2 (equation 8.5).

$$\varepsilon \sim N(0, \sigma_n^2) \tag{8.5}$$

The interceptor of the linear equation can be included in w by adding a column of 1 in the input vector x. The values of w that minimizes the sum of the squared residuals are determined using the ordinary least squares method.

Gaussian process regression

Gaussian processes is a general-purpose machine learning algorithm that can be applied to regression or classification problems. It is constructed from a Bayesian analysis of the standard linear model (equations 8.3 and 8.4). The matrix that concatenates the n sample data points is defined as $X \in R^{nxm}$ and its respective target vector is $y \in R^n$. To calculate the probability density function, the Bayesian theorem is applied:

$$p(w|y,X) = \frac{p(y|X,w)p(w)}{p(y|X)} \tag{8.6}$$

In this framework, a prior probability distribution is defined according to the previous knowledge of the system. A prior with zero mean and a covariance matrix of Σ_p is used: $w \sim N(0, \Sigma_p)$. Finally, The prediction function f_* of a test case x_* can be found by averaging the outputs of all possible linear models with respect to the Gaussian posterior:

$$p(f_*|x_*, X, y) = \int p(f_*|x_*, w)p(w|X,y)dw \tag{8.7}$$

The Bayesian analysis of the linear model suffers from limited expressiveness. In order to overcome this, a projection to a higher dimensional space is achieved through a group of basis functions ($\phi(x)$) applied to the inputs. When applying the Bayesian analysis to this new formulation and using x and x′ as input vectors from two different target sets. It is possible to define the kernel (covariance) function:

$$k(x,x') = \phi(x)^{\intercal}\Sigma_p\phi(x') \tag{8.8}$$

The Gaussian process is defined by its mean function ($\mu(x)$) and kernel function. It is a collection of random variables, as shown in equation 8.9. In this work, a Radial-basis function (RBF) kernel is selected (Equation 8.10). The RBF has the peculiarity to assign one hyperparameter called lengthscale (l_i) to each independent variable. These hyperparameters are optimized to maximize the marginal likelihood, using the âĂŸL-BGFS-BâĂŹ algorithm, as implemented in [128]. One set of lengsthscales are found for each feature that needs to be estimated.

$$f(x) = GP(\mu(x), k(x,x')) \tag{8.9}$$

$$k(x_i, x_j) = exp\left(-\frac{1}{2}d(x_i/l, x_j/l)2\right) \tag{8.10}$$

For the sake of conciseness, the above equations only briefly describe Gaussian Processes regressions. The interested reader can refer to [129] for a more comprehensive explanation.

8.3.3 Optimization process implementation

To create a dataset of optimal microgrid configurations, many MILP sizing problems are solved. To this end, the algorithm shown in Figure 8.3 is proposed. It is divided into a MILP creation phase, a main loop and an inner loop. Each step is computed in the following manner:

- In the first phase, the abstract model of the optimization is created. Then, the unmutable parameters are incorporated into the MILP model. The mutable parameters are defined by their lower and upper bounds.

- The main loop is run for each village size (from N_{min} = 50 to N_{max} = 550 households, with a step of 50). In each case, a Latin hyper-cube is initialized, defining the sampling of the other mutable parameters.
- Inside the above loop, the demand and renewable generation profiles are generated for each of the 150 ($N_{optimizations}$) instances. All mutable parameters being set, the system is optimized and the process is repeated for each element of the Latin hypercube.

8.4 Results and Discussion

The eleven different village sizes (chapter 3) together with the 150 elements of the Latin hypercube result in 1650 different instances of the problem. The termination criteria for the optimization is a gap for the MILP problem of less than 1 % or a maximum solving time of 30 min. The optimizations were performed in 175 hours, with an average resolution time of 381 seconds per instance on a computer with 16 GB RAM and an Intel® Core™i7-8850H CPU @ 2.60GHz x 12. The time spent to optimize all instances shows the limitations of a per case approach, since, only in the lowlands of Bolivia, there are more than 3000 un-electrified villages and 903 of those are between 50 and 550 households without access to energy.

8.4.1 Optimization results

A summary of the optimization results is shown in Table 8.3. It is worthwhile to note that the considered search space of the techno-economic parameters is large, leading to exploring extreme situations where some of the technologies are heavily penalized or rewarded (Figure 8.4). Taking this into account, the average NPC for all optimization is 490 thousands of USD per village, which covers all electricity-related expenses for 20 years. The average LCOE is relatively high because of the penalization of the extreme cases (where grid extension or solar home systems will most likely be preferred to microgrids). Finally, the box plot of the LCOE (Figure 8.4) shows the importance of the economy of scale. Larger communities are characterized by a lower LCOE.

TABLE 8.3: Optimization results.

Variable	Average	Max	Min	Std deviation
NPC (thousands USD)	490	1 690	39	303
LCOE (USD/kWh)	0.44	1.18	0.1	0.16
PV nominal capacity (kW)	59	256	0	57
Battery nominal capacity (kWh)	186	1123	8	229
Renewable energy penetration (%)	54	99	0	35
Battery usage (%)	27	65	4	26
Energy curtailed (%)	9	36.7	0	8
Time (s)	381	2185	37	573

The nominal capacities of the different technologies are constrained during the optimization process. As mentioned before, the nominal capacity of the Genset is 75 % of

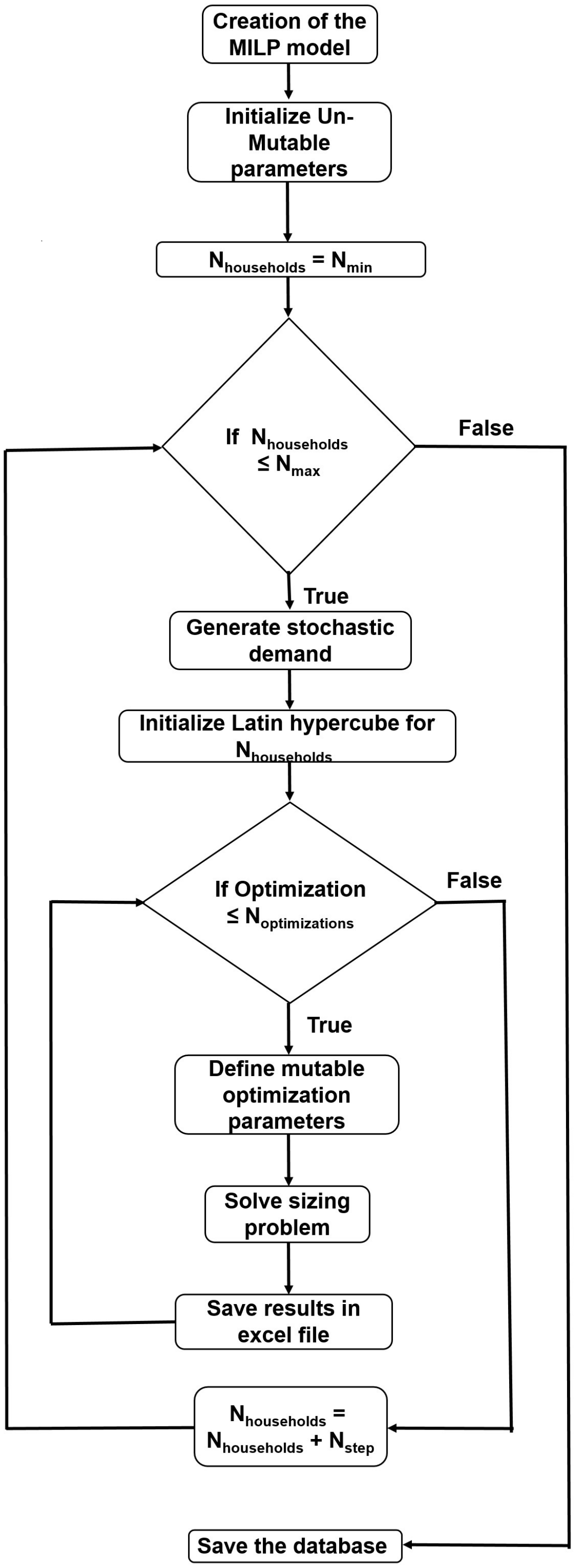

FIGURE 8.3: Algorithm for the dataset creation.

the maximum demand and it is always deployed to ensure a minimal quality of service. This forces the system to install a sufficient battery capacity to cover the peak demand. In general, it is possible to differentiate three main system configurations:

- The first one corresponds to a high battery and PV capacity, in which a large share of the consumption is covered by solar generation.
- The second one consists in using the battery to reach the peak demand and cover rapid changes in the load and in the PV generation. It corresponds to a low battery usage (equation 8.12), and low installed battery and PV capacities.
- The last configuration corresponds to the intensive use of the diesel generator and of batteries to cover the peaks. No PV is installed and the renewable penetration is thus null.

The transition between these three groups is clearly visible in Figure 8.5: the left of the plot corresponds to the systems with high PV and battery capacities, and therefore high renewable penetration (equation 8.11). The middle zone corresponds to limited PV capacity and the the right part corresponds to the case without PV generation and zero renewable penetration.

$$Renewable\ Penetration = \frac{\sum_{t=1}^{T} E_t^{re}}{\sum_{t=1}^{T} E_t^{re} + \sum_{t=1}^{T} E_t^{ge}} \tag{8.11}$$

$$Battery\ Usage = \frac{\sum_{t=1}^{T} E_t^{bat,dis}}{\sum_{t=1}^{T} D_t} \tag{8.12}$$

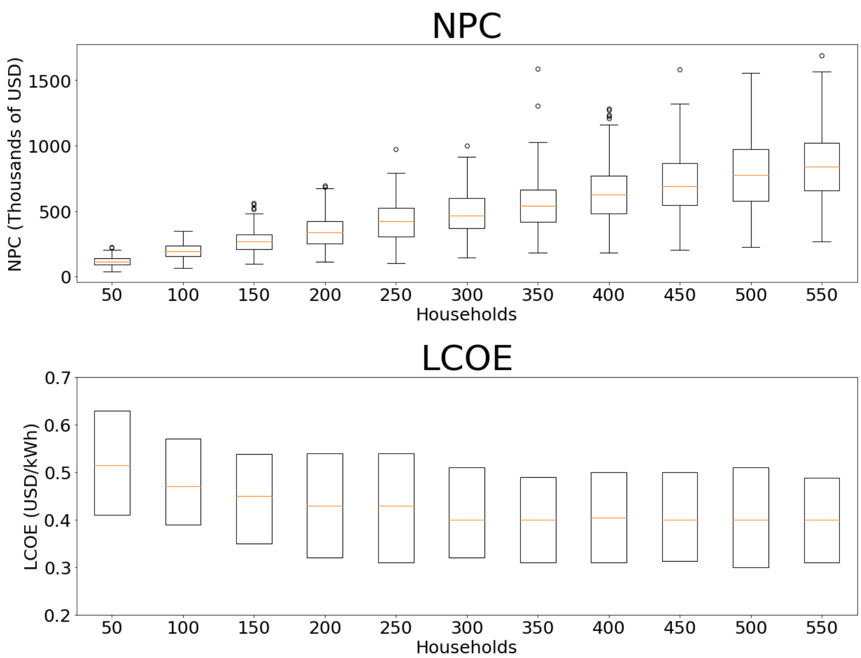

FIGURE 8.4: Box plot for the NPC and LCOE. The box contains the lower to the upper quartile of the data, they have a median line. The whiskers shows the range of the data and the points consider outliers are plot separately as circles.

.

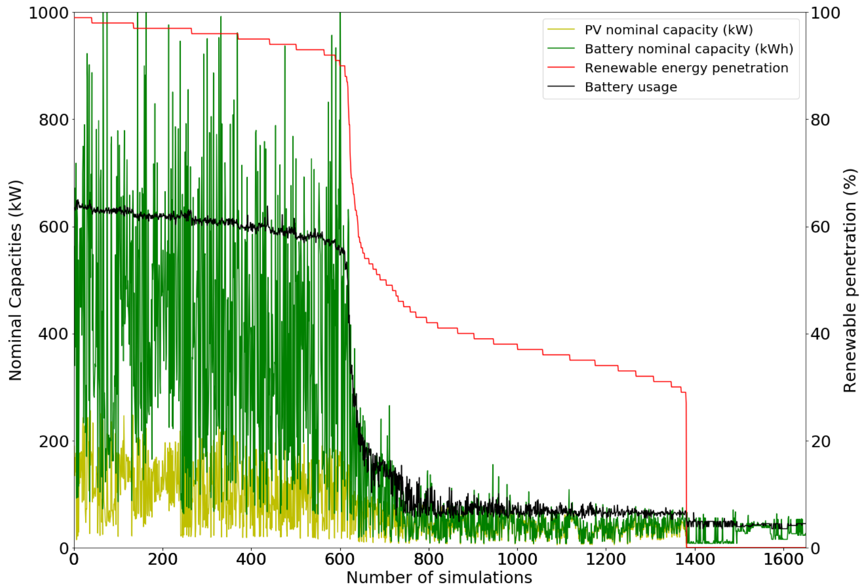

FIGURE 8.5: Installed capacities in each simulated case. The values are ordered according to renewable penetration.

TABLE 8.4: Input and output parameters for the surrogate model.

Input parameters	Output parameters
PV invesment cost	NPC
Battery invesment cost	LCOE
Depth of discharge	PV installed capacity
Battery Cycles	Battery installed capacity
Generator investment cost	
Generator efficiency	
Low heating valuer	
Fuel cost	
Toto demand in the year	
Total PV production from one unit	

It is finally worthwhile to note that the highest renewable penetration reached during the optimization process is 99 %. These instances also corresponds to the highest NPC and LCOE due to the necessity to oversize the PV and batteries. Although in those cases a diesel generator is still installed as a back-up to ensure the system reliability.

8.4.2 Surrogate models

The amount of information generated while solving each instance is important and includes, among others, the system architectures, the optimal component sizes, the dispatch strategy or the cost information. To showcase the proposed methodology, only a subset of the model outputs have been selected as dependent variables for the surrogate models: the NPC, LCOE, battery and PV installed capacity. These variables are deemed as the most relevant for the purpose of the GIS analysis, but other variables could easily be added by following the same methodology. Table 8.4 summarizes the input and output variables used for the creation of the surrogate models.

The regression results are shown in Table 8.5. In the case of the NPC, a high correlation and a relatively small MAE are achieved. Figure 8.6 shows that MVLR has significant lower performance if compared to GPR. Although it can approximate adequately values that are close to the average NPC, its performance is inferior in the low-NPC range. Some negative results are obtained for some cases, which is not acceptable. The LCOE surrogate model has a similar R^2 value, but presents lower variability (and thus no negative values), which make it a more reliable indicator for the purpose of this work. It is finally important to highlight that the highest model errors are obtained for the extreme values (i.e. the boundaries of the simulation space), which have a lower probability of occurrence.

TABLE 8.5: Surrogate model indicators.

	NPC		LCOE		PV		Battery	
Type of MLT	GPR	MVLR	GPR	MVLR	GPR	MVLR	GPR	MVLR
r2	0.99	0.86	0.98	0.81	0.92	0.76	0.86	0.58
MAE	22	82	0.015	0.05	11	22	52	115
RMSE	36	111	0.022	0.07	16	27	85	148

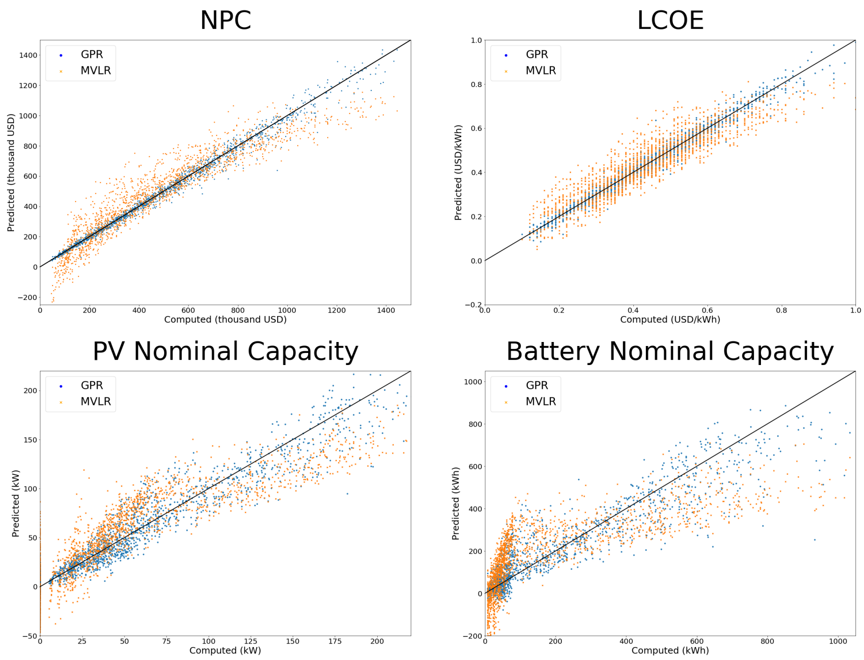

FIGURE 8.6: Predicted vs computed plots with 5-folds cross validation results.

These results indicate that GPR is a powerful tool to predict the NPC and the LCOE for a rural isolated microgrid without the need of a computationally intensive optimization for each specific case. On the other hand, it exhibits lower performance when estimating the nominal capacities of the Battery and PV systems.

These effects are further explored by means of a one-dimensional analysis: all the techno-economic parameters are kept constant except the diesel price. The fixed values correspond to the typical case of a Lithium-ion battery (battery cycles of 5500, Depth of discharge of 20 % and Unitary investment cost of 550 USD/kWh), average

PV price (1500 USD/kW) and typical diesel Genset characteristics (efficiency of 31 %, lower heating value of 9.9 kWh/l and 1480 USD/kW of investment cost). The quantity of Households is set to 300 and the fuel price changes from 0.18 to 2 USD/l.

As shown in Figure 8.7, and in agreement with the previous results, there is a good match between the computed NPC and LCOE points with the GPR functions. MVLR can predict outside the search space of the optimization process while the GPR rapidly loses its prediction capacity outside the search space. The error in the prediction of the installed capacities clearly appear in the 1-D analysis of the PV capacity regression: the rapid non-linear transitions between typical system configurations are smoothed out by the GPR surrogate models, which significantly increases the error around these points (Figure 8.7).

In the figure with different households sizes, the estimation for the PV is good as long as it does not enter in the zone with high renewable energy penetration. The quality of the GPR surrogate model could possibly be improved with more observations (i.e. optimizations), with a more limited number of independent variables or with a more advanced kernel functions or regression methods. The compromise between model accuracy and complexity is however deemed acceptable for the purpose of this work, which, considered the scale of the analysis (country or regional level), is only marginally affected by the smoothing of fast individual transitions.

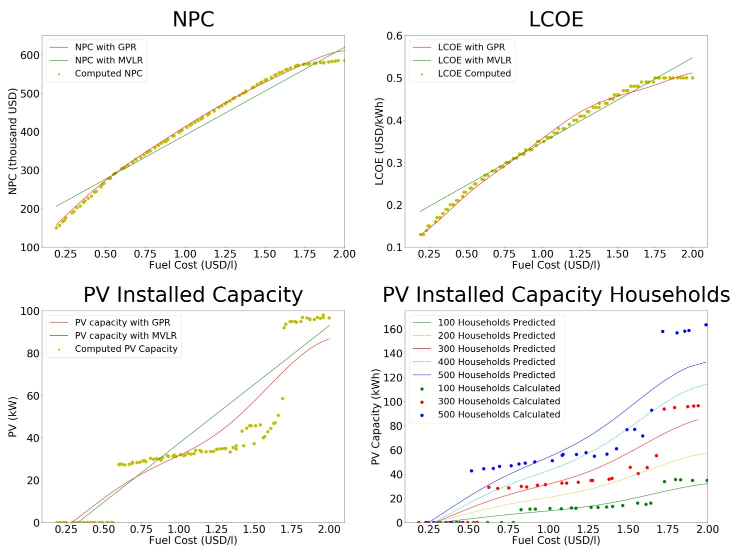

FIGURE 8.7: Computed vs predicted values for the chosen target variables

8.5 Conclusions

A methodology to derive surrogate models for energy planning purposes based on MLT is proposed in this chapter. To accomplish this, data concerning the lowlands communities in Bolivia is used to create plausible demand scenarios and a MILP sizing model is used to create a dataset of optimal size microgrids systems under different techno-economic conditions. MLR techniques are applied to train and validate surrogate models to predict the outcomes of the optimal sizing problem.

Throughout the 1650 different optimizations, hybrid microgrids proved to be a cost-optimal technology in many cases. PV was part of the optimal choice in more than 80 % of the cases, even when the price of the technology was high. This leads to a large penetration of renewable energy, which supplies it mainly during the day. Batteries are mostly used to cover peaks and day/night transitions, when the Genset is not able to provide energy due to operational constraints. The LCOE of hybrid microgrids is competitive in the rural energy market in Bolivia, ranging from 0.07 to 0.21 USD/kWh which is competitive with diesel-only microgrids. This competitiveness is achieved despite an important subsidy of diesel in Bolivia, which caps its price to 0.18 USD/l (international diesel markets are around 1 USD/l).

Overall, the surrogate models show a good capacity to predict the NPC and LCOE values of the optimized system, with a high R^2, and a low MAE and RSME. PV and battery installed capacities are less accurate because of the difficulty to replicate step-wise transitions from one typical system configuration to the other. These transitions are smoothed out, which makes the regression model unsuitable for the detailed sizing of a particular microgrid which is deemed acceptable for macroscopic analyses. The main advantage of this methodology is its adaptation capability, since it can be applied to a wide range of technologies and the continuous variation of their installed capacity. The following conclusions and lessons learned can be extracted for the surrogate model creation process:

1. The bottom-up demand profile generation described in the previous chapters is a very flexible tool and constitutes a powerful method to model not-yet electrified communities from limited socio-economic data.
2. Surrogate models are an excellent way of exploring the most cost-efficient solutions from a set of viable technologies. This is especially true when planning at a national scale where there can be thousands of decentralized systems to consider simultaneously.
3. The creation of the dataset is a computationally-expensive process. In the case of a low number analyzed systems, individual optimizations might remain the best solution.
4. The search space must be carefully defined and a high density of data samples must be ensured in the regions where is more likely that the surrogate models will be used.
5. The GPR model performed significantly better than the MVLR.
6. To deal with the observed clusters of typical system configuration, the regression could be complemented by a classification machine learning algorithm, assigning the considered setup to a typical configuration.

The proposed surrogate models proved to bring significant improvement for energy planning purposes: instead of a single simplistic configuration (characterized by a fixed LCOE and a rigid microgrid design), they allow to adapt the microgrid configuration to the specific boundary conditions of each community (diesel price, size, demand peculiarities, etc.) and thereby significantly refine the macroscopic analysis. Surrogate models offer an excellent solution to explore such multidimensional optimal deployment problems at the country level.

Chapter 9

The Bolivian pathway to 100 % electrification

Preliminary versions of the analysis proposed in this chapter have been proposed in:

Peña, J., ***Balderrama, S.****, Lombardi, F., Stevanato, N., Sahlberg, A., Howells, M., Colombo, E., & Quoilin, S. (2020). Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET). Energy for Sustainable Development, 56, 98-118.*

Balderrama, S.*, Lombardi, F., Stevanato, N., Peña, G., Colombo, E., & Quoilin, S. (2021). Surrogate models for rural energy planning: Application to Bolivian lowlands isolated communities. Energy, 121108.*

9.1 Introduction

With pressing priorities in the development agenda, policy makers in developing countries are in the difficult situation of prioritizing policy actions. Limited government and utility budgets need cost effective solutions to bring the desired development benefits of electrification, health, education and food security, among others. Energy access is a pre-requisite for economic activity and for human development and interacts in synergy with other development needs. In this context, the usage of modeling tools can inform decision making at different levels of the planning strategy. In this chapter, the methods developed throughout the wholed thesis are applied to a answer practical question: What is the most cost-effective manner to grant access to electricity to the Bolivian population? This is answered by tackling the following sub-objectives:

1. To adapt and harmonize OnSSET for the use of surrogate models.
2. To Propose a baseline scenario for the electrification of Bolivia.
3. To analyze the impact of the subsidy on diesel price in the electrification strategy in Bolivia.

9.2 OnSSET adaptation

The principal aim of this thesis is to propose a methodology that sizes and simulates many decentralized rural electrification systems at the country level in a computationally tractable manner. In the particular case of Bolivia, there are 8761 communities without access to the main grid. Since there are multiple solutions to achieve this, a system design has to be optimized for each of them. However, finding the economical optimum is a demanding task from a computational point of view: solving an optimization for each community could last days in a computer with standard capabilities. The OnSSET algorithm was therefore modified to allow the use of surrogate models based on the methodology described in the previous chapter.

To that aim, the following improvements and adaptations to the model have been implemented:

1. Adapt OnSSET to allow for a variable number of considered technologies instead of the pre-defined, hard-coded original technologies. In particular, the hybrid PV technology should be introduced instead of the diesel-only or PV-only technologies.
2. Adapt OnSSET to accept surrogate models to assess the demands based on different parameters. This allows taking into account the individual specificities of each community.
3. Include the possibility of using surrogate models for the prediction of the LCOE in the planning algorithm.
4. Include surrogate models to calculate installed PV, Batteries and genset capacity for each community.

These adaptations allow to use more advanced costs functions for microgrid systems in place of the original fixed LCOE hypothesis based on the peak demand and the capacity factor of the technology (cfr. section 7.2.4). This flexibility allows to consider hybrid microgrids tailored for the particular case of a community instead of a fixed and non-optimal design for all communities. The new sizing model also optimizes the energy flows, leading to a more accurate NPC and LCOE. This is an important feature when analyzing energy systems with different energy sources, as an un-optimal dispatch strategy could lead to a higher operation cost or energy curtailment of the renewable sources. [27].

9.3 The Bolivian Case Study

To test the proposed methodology, the cost of electrification for Bolivian communities without access to electricity is investigated. The selected technologies are grid extension, hybrid microgrids (communities from 50 to 550 households) , and PV/battery home systems (communities from 0 to 50 households). The most relevant techno-economic parameters to perform the calculations are shown in table 9.1.

To account for the steady improvements in renewable technology costs and uncertainty on the continuity of fossil fuels subsidies in microgrid deployment, various cost scenarios (fuel and capital costs) are defined. Two existing subsidy schemes in the Bolivian electricity sector are included as scenarios together with international

TABLE 9.1: Onsset technology characteristics

Parameter	Unit	Value
Lifetime of the grid	years	30
Discount rate	%	12
load moment (50 mm aluminium)	kW m	9643
Power factor grid		0.9
Grid losses	%	18.3
Distribution losses	%	0
MV max distance reach	km	50
MV line cost (33 kV)	USD/km	9000
LV line cost (0.24 kV)	USD/km	5000
Transformers (50 kVA)	USD	3500
Max nodes per transformers	nodes	300
Substation (400 kVA)	USD	10 000
Substation (1000 kVA)	USD	25 000
Additional conection cost to the grid/microgrid	USD	125
Operation and maintenance of distribution lines	%	2
Grid capacity investment cost	USD	1722
Grid electricity generation cost	USD/kWh	0.13
Diesel truck consumption	l/hour	33.7
Diesel truck volume	l	15000

diesel prices. Figure 9.1 illustrates the modeled instances and Table 9.2 describes the values used in each of them.

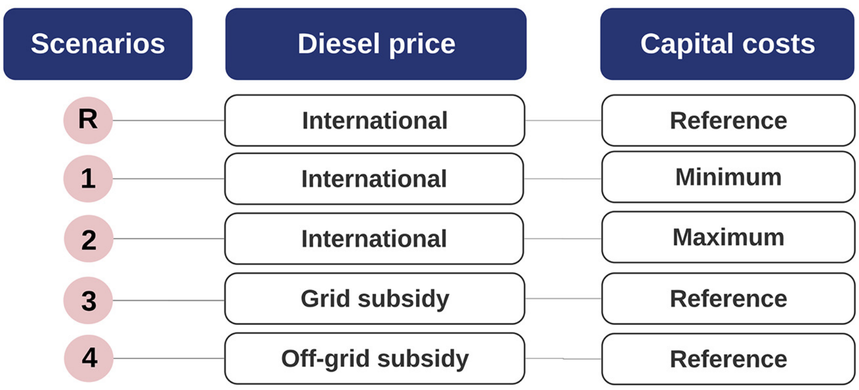

FIGURE 9.1: Cost-scenario components. Reference scenario and other four scenarios with combinations of diesel price and capital investment costs described in table 9.2

TABLE 9.2: Specific values used in the cost scenarios.

Parameter	Unit	Reference	Scenarios
Diesel cost	USD/l	0.8	Diesel grid-subsidy: 0.53 Diesel off-grid subsidy: 0.18
Battery invesment cost	USD/Wh	550	Minimun = 350 Maximun = 750
PV invesment cost	USD/kW	1500	Minimun = 1100 Maximun = 1900
Diesel genset investment cost	USD/kW	1480	Minimun = 1100 Maximun = 1900

9.4 Results

The OnSSET tools provides an algorithm to determine the initial electrification status for a given scenario. However, this was not used in the present work since the baseline electrification status can be obtained from the census database described in section 3.2. The location, size and electrification status of each Bolivian comunity is shown in Figure 3.1.

Additional models to calculate the LCOE, NPC, investment,PV and battery installed capacity for the highlands were created in addition to those presented in Chapter 8 (restricted to the lowland regions). The performance metrics for these new surrogate models were similar to those in Table 8.5.

TABLE 9.3: Results of the base scenario of the optimal electrification process.

	Newly electrified people	NPC (millons USD)	New Capacity	Average Capital investment per household (USD/hh)
Grid Extension	3 308 928	1 635	279.8	1 811
Hybrid microgrids	53 140	44	2.8	2 413
Solar home systems	493 225	236	66.4	1 326
Total	3 855 293	1 915	349	1 636

The investment cost per household varies largely depending on the electrification technology. For households electrified through grid extension, the investment cost

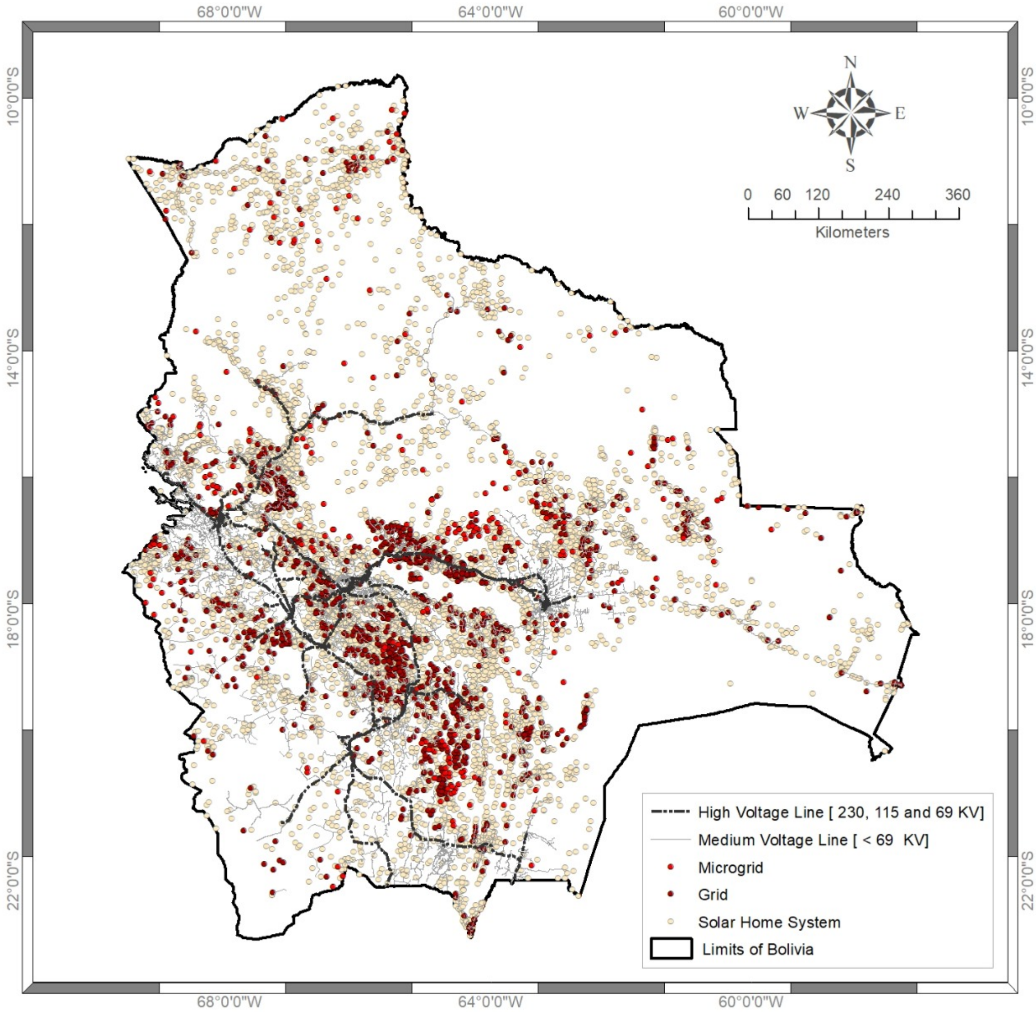

FIGURE 9.2: Cost-optimal deployment of electrification technologies for the reference scenario

increases with increasing distance to the transmission lines and decreases with increasing population density. The average cost of connecting a household with the grid amounts to 1811 USD. Table 9.3 summarizes the number of new connections per technology type, investment cost estimates and new capacity when using the OnSSET for microgrids. Through 2025, Bolivia will need to increase the grid capacity by 280 MW and the off-grid capacity by 69.2 MW (PV and gensets) in order to meet the increased residential demand and electrification targets indicated in our reference scenario. In order to meet the SDG, this additional generating capacity would have to come from renewable technologies.

The investment cost necessary for a 100 % electrification is divided into two main components (Table 9.3). Around USD 1.6 billons for the extension of the network. Without a new planning of high voltage lines, the number of people reachable through grid extension in the near future remains fairly limited. The second part of these costs are the investments necessary to reach isolated populations, the amount needed is more than 280 million dollars. The total cost is high, but it is in line with that calculated by the Vice Ministry of Electricity and Alternative Energies of around 1.99 billions of USD [52].

Figure 9.3 shows the obtained results if OnSSET classic representation of isolated energy systems is used. The techno-economic data for SHS, Diesel and PV micro-grids are detailed in annex F.4. The results shown are similar to the REF scenario (Table 9.4), in part due to the high penetration of the grid. The main difference can be seen in the capability of the surrogate models to model hybrid microgrids as part of the solution pool. This makes a significant difference, since this is the optimal configuration for all the deployed microgrids, as shown in Figure 9.4.

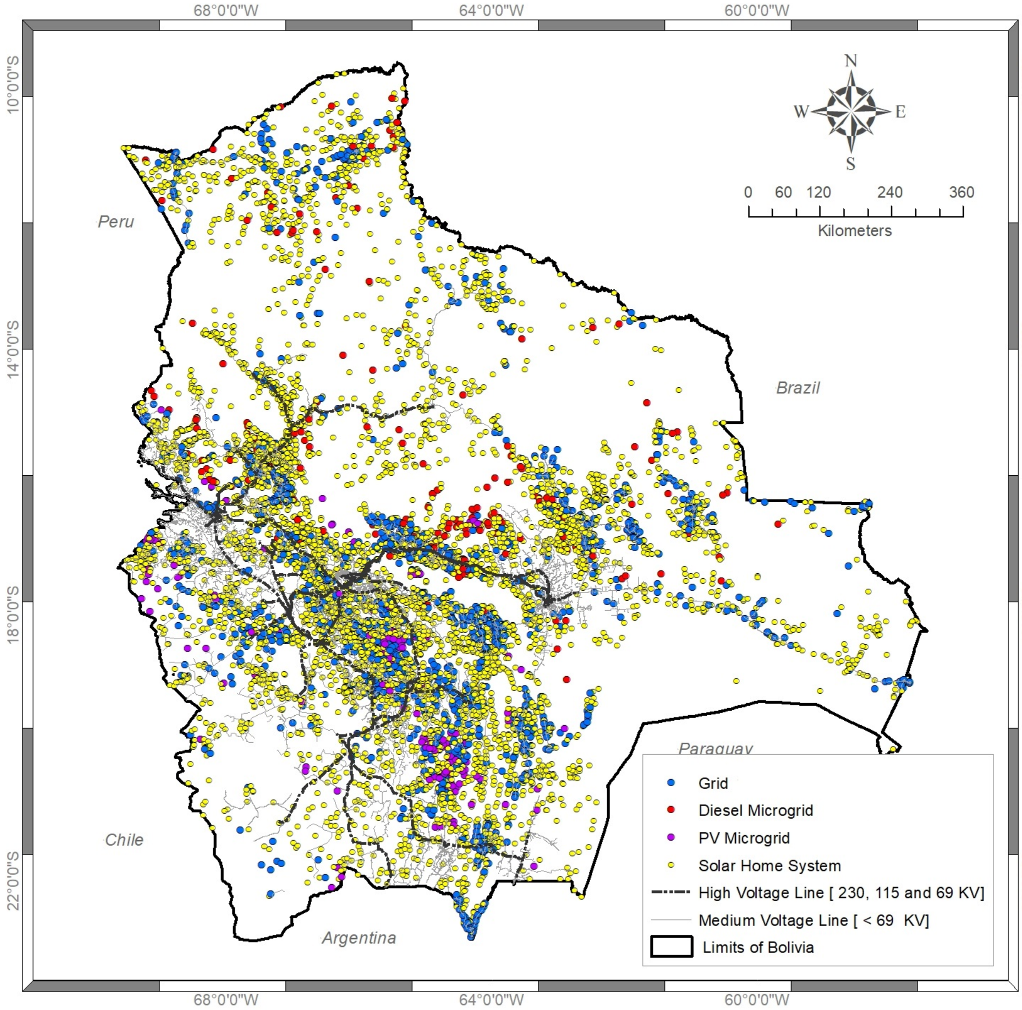

FIGURE 9.3: Cost-optimal deployment of electrification technologies for the OnSSET classic scenario.

9.5 Sensitivity analysis

Figure 9.4 compares the results for microgrids under different cost scenarios in terms of LCOE, NPC and PV installed capacity. Although the number of microgrid communities in the REF scenarios is low (221), it is in line with the Bolivian capacity to deploy them in a 5 years time span. The main impact of the cost variations on the number of electrified communities is the following: The lower investment cost of SC1 only adds 38 communities, on the other hand the higher cost of SC2 changes 58 microgrids to grid connections. The number of communities with a microgrid increases with the decrease of diesel price to 345 (SC3) and 701 (SC4) communities.

TABLE 9.4: Results of the classic OnSSET for the optimal electrification process.

	Newly electrified people	NPC (millons USD)
Grid Extension	3 306 338	1 628
PV microgrids	36 340	26
Diesel microgrids	44 014	31
Solar home system	468 601	281
Total	3 855 293	1 966

Although more communities are electrified with microgrids in the case of lower diesel prices, in SC4, all microgrids are diesel only. In SC3, the installed capacities are lower than in the REF scenario. This can have an important impact on the accomplishment of SDGs for Bolivian rural communities. It is also worthwhile to note that the change in investment cost for microgrids does not have an important impact on NPC or LCOE, but it has on the installed renewable capacity. This different considerations should be taken into account at the moment of creating subsidies for rural communities in Bolivia.

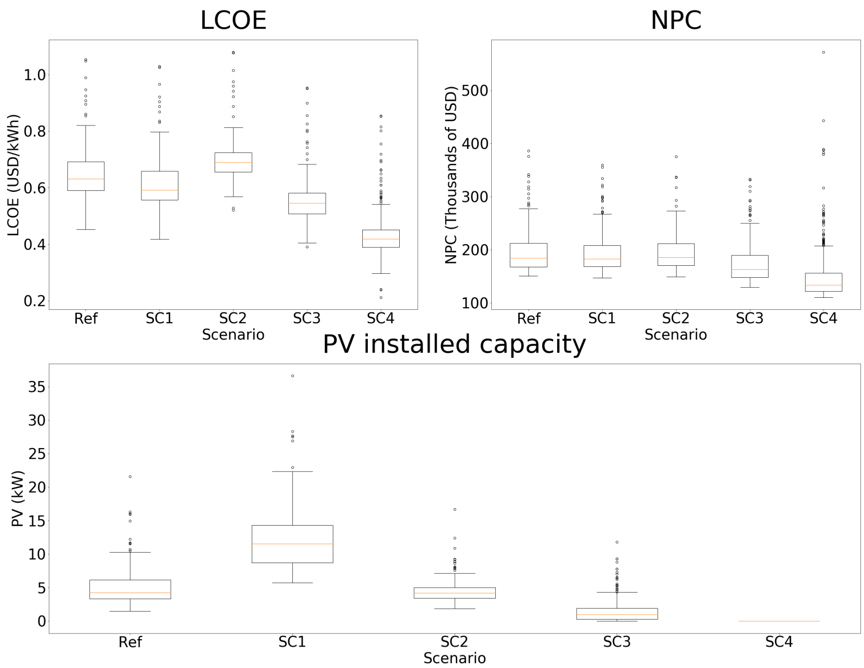

FIGURE 9.4: Summary of the sensitivity analysis results for selected communities. a. LCOE. b. NPC. c. PV installed capacity.

9.6 Conclusions

Achieving 100 % electrification involves significant investments in economic and human resources for the Bolivian state. A significant part has already been achieved, reaching almost 90 % of the population. Despite this, there are more than 7400 communities that will need a isolated solution to gain access to electricity, representing

159 798 households. Today, the investment of around 280 millions of USD to provide electricity to these communities is an affordable amount for a country like Bolivia. This is true, if its taken in account that the electric plan until the year 2025 assings 859 millons of USD to rural electrification projects for the the period 2021-2025. Among those without access to electricity, the most vulnerable population are 144 256 isolated rural families, which undoubtedly have individual photovoltaic systems as the best technical-economic option.

The most significant barrier is Bolivia's ability to carry out the number of projects necessary to close the gap. It is important to continue with the training of technical and administrative personnel capable of designing and executing this type of project with great urgency. Likewise, it is necessary to form strategic alliances between different organizations and strengthen the institutional framework, considering that the energy issue for the isolated population constitutes a technical-economic and management challenge, radically different from conventional systems, in terms of density, structure and regulation and management practices.

Results from our cost-scenario analysis reveal how sensitive the electrification results are to diesel prices. The continuation of diesel subsidies strongly reduces the economic competitiveness of local renewable energy resources. Fossil fuel subsidies are remarkably widespread in developing countries for several socioeconomic factors [130]. For a small economy, a meaningful change in subsidy schemes would consequently produce large macroeconomic impacts through economy-wide changes in sectoral relative prices and demands [131]. Therefore, it may be counterintuitive to remove fossil fuels as a measure to foster energy access. Yet, in the same way fossil fuels subsidies are used to promote affordable energy, renewable energy subsidies could be considered to compensate for this market distortion [132].

As electrification planning diversifies with the inclusion of decentralized alternatives, different affordability and financing concerns emerge. Further enabling actors should be considered by electricity planners and policymakers to address the entire range of affordability concerns for both grid and off-grid rural consumers. More importantly, better coordination among national stakeholders is needed to develop a local renewable energy industry able to mobilize public finance towards sustainable rural electrification projects.

Part IV

Conclusions and future work

Chapter 10

Conclusions and future work

Humanity has entered a crucial stage in their presence on the planet earth, as the efforts to stop climate change and generate sustainable development are synthesized in an agenda of 17 goals. From this set of objectives, SDG number 7 is related to the access to affordable and clean energy for all. It stands out from the rest for its transversality to other goals. It is for example impossible to reach quality education (SGD 4) without access to the internet, light to study at night, or computers to learn the latest technological trends in the world. Other areas that are widely affected by the availability or affordability of energy include the water sector (SDG 6) or the food sector (SDG 2).

Although no meaningful sustainable development can be achieved without access to energy, there are still more than 800 millions of people on the planet do not have access to electricity. Furthermore, as electrification rates come closer to 100 %, the grid extension stops to be a feasible technology due to economic, social or environmental constraints. Recent research suggests that decentralized systems (microgrids or SHS) have reached the needed maturity to supply reliable, clean and affordable energy to rural isolated communities. However, the potential role of these technologies is not clearly quantified, as the electrification planning involves comparing different technologies at a macro scale. This can easily lead to computational tractability problems, forcing researchers and practitioners to resort to over-simplified hypotheses to solve this complex energy planning task.

Under these circumstances, the aim of this thesis was to propose a comprehensive methodology to plan and evaluate the cost of electrification in remote rural areas. It applies a bottom-up approach, ranging from the problem or demand generation to the siting and optimal sizing of isolated systems in conjunction with a grid extension strategy to reach 100% rural electrification. It accounts for the uncertainty in demand and renewable generation through the consideration of multiple stochastic scenarios.

The structure of the thesis reflects this bottom-up approach: the first chapters are dedicated to the household-level demand assessment, followed by chapters relative to the optimal sizing of community energy systems, while the last chapters focus on country-scale optimal rural electrification strategies.

In particular, chapter 2 focuses on demand modeling techniques, with two main approaches. The first one is a top-down analysis with aggregated information and statistical trends. The second one analyses the contributions of individual system components, with a high level of detail covering different economic sectors and appliance usages. It was concluded that the bottom-up approach is the most appropriate way to estimate future energy consumption for un-electrified rural villages.

In the third chapter, the Bolivian context is introduced. Although the country has a plan to increase its renewable share and its electrification rate, both have experienced relatively low progress in recent years. They have stalled as the grid stops to be a viable solution to reach the most isolated rural communities. The tremendous Bolivian renewable energy potential has barely been taken advantage of, at the exception of few deployed microgrids in the last years. As a first step to unlock the deployment of such decentralized systems, low consumption demand profiles typical of rural households were created for the highlands and lowlands of Bolivia. The expected consumption defines a minimum energy demand for different human activities. It does not, however, include productive uses of electricity such as electricity for agriculture and manufacture. Since they do not always occur immediately after electrification [16], these new potential demands require careful assessment as they could further increase the system load. This is considered to be beyond the scope of the present thesis which focuses on universal access for inhabitants, but it constitutes an important research track for future works.

The fourth chapter presents a sizing framework for isolated systems under uncertainty, it gathers the best practices in energy modeling to create a tool that can be adapted to practitioners needs, as proven throughout the course of this thesis. It proposes different variations of the model, which can be stochastic or deterministic and can be formulated as an LP or MILP problem. Furthermore, it puts a special focus on the detailed representation of the different technologies, taking profit of the available field monitoring data.

In order to assess the capabilities of microgrids to supply energy in the Bolivian context, operational data from the "EL Espino" microgrid is analyzed in chapter 5. It was found that the microgrid control strategy is largely sub-optimal and leads to a high curtailment of renewable energy. An improved dispatch strategy that minimizes the curtailed energy is proposed; it consists of using the genset during the demand peaks. In addition, a new sizing of the microgrid is performed. In general, the battery and PV capacities were correctly sized but the system could benefit from a smaller genset that can better adapt to the actual load. Furthermore, the different sizing methods were compared, leading to the conclusion that there is a trade-off between system reliability and computational tractability. Although the difference on the NPC is small, there is an important variation in the size of the components and in the operation of the system when optimizing with various levels of complexity.

Current literature indicates that SHS is a feasible solution to electrified households in a rural context. However, as shown in chapter 6, supplying the total household demand results in unrealistically oversized systems with a high LCOE. On the other hand, allowing a significant lost load probability can affect the feasibility daily activities. To balance this trade-off, a methodology to analytically determine the permissible LLP of the system is proposed. The results show that the pareto front between system cost and LLP has a recognizable form and that a "knee point" can be defined as an acceptable compromise between these two antagonist objectives.

The state of the art in the modeling of isolated systems with GIS-based electrification tools is presented in chapter 7 together with its main limitations. In a nutshell, current methodologies either simplify the sizing process or resort to individual optimizations that are too computationally-intensive for the simultaneous optimization of all communities in a country. It is concluded that surrogate models based on machine learning techniques to estimate the design parameters of a given microgrid can be a solution to the tractability problem of GIS electrification tools.

A methodology to create and validate surrogate models is then presented in chapter 8. To accomplish this, the demand time series generated in chapter 3 and the sizing tool described in chapter 4 are combined with different solar PV outputs and techno-economic parameters. An input and output dataset is created and fed to different machine learning algorithms to create surrogate models. The validation process shows that, among the tested algorithms, Gaussian Processes is the most suitable technique to estimate the different design variables of a microgrid.

In chapter 9, the work done in this thesis is condensed by integrating the surrogate models into OnSSET. The GIS-based model is adapted to calculate different design parameters based on the new methodology and with a high level of technical detail. The enhanced version of OnSSET is run to calculate the total cost of electrification in Bolivia. The results show the importance of isolated systems to close the gap of unelectrified households in the country. An important conclusion from the sensitivity analysis is the impact of diesel price on the SDG 7: low prices ensure the affordability of the energy but negatively affect the environmental impact of rural electrification.

Throughout the different optimizations, hybrid microgrids proved to be a cost-optimal technology in many cases. PV is part of the optimal choice in more than 80 % of the cases, even when the price of the technology is high. This leads to a large penetration of renewable energy, which supplies energy mainly during the day. The batteries are mostly used to cover peaks and day/night transitions, when the diesel genset is limited by operational constraints. The LCOE of hybrid microgrids can be competitive with diesel-only microgrids (0.09 to 0.16 USD/kW). This competitiveness is achieved despite an important subsidy for diesel in Bolivia, which caps its price to 0.18 USD/l, while international diesel market prices can reach more than 1 USD/l. On the other hand, combustion engines cannot be discarded in microgrids as they increase the quality of the service. In this context, green fuels should be developed to mitigate negative impacts.

The first step to start to close the electrification gap in Bolivia is to find the communities that, if electrified would extend the access to health and education service to more families. This is in part due to the limited capabilities to construct microgrids in a short time. This strategy will increase the impact of the microgrids, as more people will see their life quality improve. After this, all the other communities that meet the requirements for a hybrid microgrid could be built according to a previously agreed plan.

From the chapter 9 results, it is clear that SHS will supply energy to most of the nonelectrified households in Bolivia. They should be able to provide as much energy as possible without compromising the economical viability of the whole electrification project. In order to improve the quality of service for SHS, it is possible to design appliances that take advantage of solar energy, although this was not considered in this thesis.

The amount of SHS and hybrid microgrids deployed to reach 100% electrification is significant. Bolivia counts with one of the biggest reservoirs of Lithium and it has made important efforts to develop its own industry of batteries. Under this context, the continuation of these endeavors could create a virtuous circle in which the high demand of batteries creates the needed money to improve them. This could lead to lower cost of energy for the final user, an a unique opportunity for the country to develop and grow its high-tech industry.

Finally, it is clear that the accomplishment of SDG7 will impose several trade-offs in which a decision must be made between antagonistic objectives such as cost, quality of service and environmental impact. Under these circumstances, the long term sustainability of the planet should be the maximum priority, without forgetting that the decision made will have a large impact in the quality of life of many generations of people living in some of the most remote locations in the world.

10.1 Future work

The thesis proposes a first of its kind methodology to tackle the problem of optimal rural electrification under uncertainty with a high level of details for decentralized systems. Due to the comprehensive scope of the methods developed during this thesis, several new research paths could be pursued to improve the obtained results. In the following lines, possible future works that could enhance the approach are discussed.

The difficulty to reach universal energy access in different contexts is clearly highlighted throughout this thesis. From a technical perspective, the most suitable technology for this objective is the extension of the current grid. However, economic or environmental constraints prohibit the full grid extension over the whole territory of a developing country. In this context, understanding and defining the concept of ”energy sufficiency” for a particular household or community has a special importance. As seen in several chapters, isolated solutions are not necessarily capable to cover high peak demands, or prolonged periods of low solar irradiation, thus resulting in a lower quality of service than the grid. At the same time, an imperfect access to electricity is more valuable than no access at all and might be sufficient to cover some, if not all, basic needs.

The concept of energy sufficiency was introduced by [31]. The authors define it as a state in which people’s basic needs for energy services are met equitably while ecological limits are respected. The focus is on energy needs for shelter, health, work, mobility and communication. It is important to note that these needs vary according to local conditions and cannot be defined in a rigid manner. According to the definition presented above, the "sufficiency" line coincides with the minimum amount of energy required by people to live a dignified life. However, this concept has mostly been conceived and explored in the Global North (with the aim of reducing consumption). It is a challenge to apply it in developing countries, where it is necessary to look for ways to increase provision and consumption in less favored rural areas.

Under this context, it seems very interesting to adapt and use the concept of energy sufficiency in future works, as a framework to identify minimal states of energy that the system must provide. It should however be clearly defined in relation to other concepts such as ”basic needs” or ”energy justice”. It is also important to take the time evolution of the energy needs into account since appliance adoption is characterized by a diffusion time and high disparities between urban and rural areas. Finally, the concept can also usefully inform energy planning when considering the trade-off between cost and non-served energy highlighted in chapter 6.

Although this thesis focuses on SDG 7, a more comprehensive assessment should be proposed focusing on the design resilient energy systems that help to meet other SDGs instead of limiting their accomplishment. There are limitations for each technology (environmental, service quality, etc) that cannot be evaluated properly with a cost-based comparison and would required to endogenize other SDGs in the model. This, however, should be performed with a high level of spatial granularity: there is the need to divide the Bolivian territory to represent the cultural richness of indigenous nations and its relation with the access to electricity.

The nexus between energy, water and food should be explored to create synergies inside rural communities. A coordinated system can lower the operation cost of a microgrid, for example by coordinating the use of irrigation pumps with the solar availability. New generation sources could also be integrated to the model, to maximize these synergies. Demand side management at the consumer level is another powerful tool to lower the energy cost, and is closely related to the tariff structures and business models. As shown in chapter 5, the real-time optimal control of such systems have the potential to transform the traditional microgrid into a "smart rural community".

The Design of isolated microgrids/SHS is not a trivial task because of the inherent uncertainty in the multiple model inputs and parameters. The two-stage optimization framework developed in chapter 4 covers different types of parametric uncertainties such as demand and renewable output. However, other sources of uncertainty should be taken into account at the design stage, including fuel prices and demand growth throughout the lifetime of the project. As seen in chapter 5, the methods to deal with this are computationally intensive, and more advanced mathematical formulations (i.e benders decomposition) might be required to decrease the solving times.

The power of surrogate models for rural energy electrification is still largely untapped. Beyond the preliminary approach proposed in chapter 9, their use should be extended to other aspects of the planning process, e.g. for the accurate evaluation of installed PV capacities. This requires improving the machine learning models to better account for the sharp transitions between typical system configurations.

In the electrification algorithm (OnSSET), communities are currently well differentiated. There is a need to enable microgrids covering more than one community, thus creating clusters that take advantage of the economies of scale. Similarly, the algorithm should be coupled with long-term planning tools at the grid level to evaluate the pathway towards 100% electrification at different points in time .

Finally, special attention should be given to communities where it is not feasible to build microgrids. These will most likely depend on SHS, which suffer from a limited capacity to cover the overall energy demand. In this context, the high curtailment levels stated in some simulation open the possibility to create rural energy communities. In this arrangement, households could share electricity between them to meet unexpected energy consumption or peaks. If designed correctly, this bottom-up approach can allow to power community services in a decentralized way and dramatically improves the quality of service [133].

Appendix A

Data and Script online repository

Open science has become pivotal to ensure transparency and reproducibility of research works. For this reason, a preference for tools and databases with an open license were given. In the case that this was not possible, software with free license to academics were chosen. Under these constraints, most of the work was done in the programming language Python. The large collection of libraries and one of the biggest communities in the world makes the use, modification or creation of research tools an easier task.

The scripts and most of the data used through the course of this thesis is uploaded to the following repository: `https://github.com/Slbalderrama/Phd_Thesis_Repository`. In the repository it is possible to find a detailed description on how to use each script. The scripts should be run without modifying them, the most important results are printed in the console. In the main page of the repository, a README file can be found, with all the information on how to set up the python environment. Each chapter has its own folders except chapters 1 and 10. From a general point of view, it is possible to find the following information on each folder:

- A README explaining the peculiarities of each folder, with a particular emphasis on the explanation on where to find the results and plots of the thesis.
- The inputs for the different scripts.
- The main scripts, they are the ones to be run in order to reproduce the results of this thesis.
- Additional scripts to pre-process the inputs.

Appendix B

Survey for households with access to electricity, el espino

B.1 General Information

Date: / /

Location:

Number of user:

Interviewed information:

Name:
Age:
Role:
Sex:
Education:
Work:
Coordinates:

1) Fill table B.1 with the following questions:

a. How many members of this segment live in this household ?.

b. which are their activities?.

c. When/where they do each activity?.

TABLE B.1: Household composition and general activities

Gender/Age	Number	Activity	When/Where
Children 0-5			
Children 5-17			
Women over 17			
Men 17-59			
Men over 59			

2) Fill table B.2 with the following questions, for each segment of the family:

a. What time do each segment wake up?

b. Which is the activities time?

c. When do they return home?

d. When do they go to sleep?

e. How seasonality affects the activities?

TABLE B.2: Household members daily activities.

Hour	Children 0-5	Children 5-14	Women over 14	Men 15-59	Men over 59
0					
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

3) Do you have some change of the time schedule of the day in the weekend? In the different seasons?

4) How many rooms there are in your household?

B.2 Habit Changes

5) How the electrification has affected your habits?

a. Wake up / go to sleep time

b. Work / School time

c. Lunch / Dinner time

d. Free activities time

B.3 Income

6) Which activity do you do as job?

a. Agriculture:

i Land extension, kind of cultivation, production (self consumption/selling/trade), months

-

-

-

-

b. Cattle breeding:

i Quantity, kind, production (self consumption/selling/trade), months

-

-

-

-

c. Seasonal work, months:

-

-

d. Self-employment (shop):

i . Which activity, months

-

-

e. Receive help from family

f. Receive subsidies

g. ....

7) How many times do you go to shop monthly/weekly? How much do you spend?

8) Which goods do you buy?

9) How often do you pay your electricity bill ?

- Daily
- Weekly
- Monthly
- Bimonthly
- Threemonthly
- Other (please specify)...............................

10) How much do you pay the bill on average every time? Bs

11) Which is the monthly average electricity consumption of your household? (Do you have a meter? Do you check consumption?)

12) Do you have any comments about the electricity supply?

B.4 Electricity use and supply

13) Which are the devices using electricity in the house? For each appliance, answer the following questions and fill Table B.3.

a. Which is the nominal power of the appliance P_{ij}?

b. How many appliances of these type do you have?

c. How much time do you use it in a day?

d. In which part of the day do you use it? [i.e. 8:00 to 10:00 & 18:00 to 21:00]

e. How much time this appliance is ON at least? (minimum time)

f. Does the seasonality affect this time schedule?

g. When do you buy it?

TABLE B.3: Household appliances.

appliance i	Nominal Power P_{ij} [W]	Number Nij	Average daily use h_{ij} [h]	Functioning window W_{f,ij} (seasonality)	Min time D_{ij} [min]	Season-ality	When did you buy it?
In lights							
Out lights							
radio							
Cellular							
fridge							
freezer							
Mixer							
hot plate							
kettle							
fan							
AC							
TV							

14) Which are the devices using electricity in the house before the electrification? For each appliance, answer the following questions and fill Table B.4.

a. Which is the nominal power of the appliance P_{ij}?

b. How many appliances of these type do you have?

c. How much time do you use it in a day?

d. In which part of the day do you use it? [i.e. 8:00 to 10:00 & 18:00 to 21:00]

e. How much time this appliance is ON at least? (minimum time)

f. Does the seasonality affect this time schedule?

TABLE B.4: Household appliances before electrification.

appliance i	Nominal Power P_{ij} [W]	Number Nij	Average daily use h_{ij} [h]	Functioning window W_{f,ij} (seasonality)	Min time D_{ij} [min]	Season-ality	When did you buy it?
In lights							
Out lights							
radio							
Cellular							
fridge							
freezer							
Mixer							
hot plate							
kettle							
fan							
AC							
TV							

15) Can you put in a priority and utility order the electric appliances? (appliances of the table before) How much are you willing to pay for these appliances?

-
-
-
-

16) Answer the following questions for each energy source and fill Table C.1.

a. Do you use this source?

b. Which are the purpose? (more than one answer is allowed)

c. Where did you get it?

d. How much does it cost?
e. How much do you get it?
f. How many times do you get it in a week?
g. How do you transport it?
h. How much time do you need to transport it?

TABLE B.5: Other energy sources.

	Used	Purpose	Supply	Price	Quantity	Frequency	Transport	Time
Biomass								
Diesel								
Gasoline								
GLP								
Kerosene								
Charcoal								

B.5 Candles use and supply

17) Do you use candles at home? Yes No

18) How many candles do you buy? How much do the candles cost?

19) Where do you buy it? How much time do you need for purchasing it?

B.6 Cooking (for cooking fuel supply see the table above)

20) Is cooking done indoors, outdoors, both?

21) With which frequency?

B.7 Heating & Cooling

22) Do you have cold during the year?

23) Which appliance do you use for heating?

24) Do you have hot during the year?

25) Which appliance do you use for cooling?

26) Which is the material of house?

B.8 Spatial Network

27) Answer the following questions for each energy source and fill Table B.6.

a. Do you have it?
b. Will you buy it?
c. How many owners of this appliance did you know when you bought it? (1)
d. How many owners of this appliance do you know? (2)
e. Why? (2, no)

TABLE B.6: The spatial network for appliances.

Appliances i	Question: a	Question: b	Question: c	Question: d	Question: e
in. lights					
out. lights					
radio					
Cellphone					
fridge					
freezer					
Mixer					
hot plate					
kettle					
fan					
air conditioner					
TV-dvd					
Other (please specifiy)					

Appendix C

Cuestionario Acceso Energético del Pueblo Atacameño de Toconao

C.1 Información General

Direcciónn:..

Fecha:......./ ..../ ....

Sexo: M F

Edad:.......

1) Posición en la familia:

a. Cabeza de Familia
b. De la familia
c. Otro (especificar)

2) Información de los miembros del hogar (Table C.1).

TABLE C.1

	Miembro familia	Sexo	Edad	Ocupación
1				
2				
3				
4				
5				
6				
7				

C.2 Uso y Suministro de Electricidad

3) Está su casa conectada a la red eléctrica local (CESPA)? SI NO

4) Tiene un medidor de corriente instalado en su casa? SI NO

5) Cuál es la potencia de su medidor? kW

6) Cuáles son los dispositivos que utilizan electricidad en la casa? (Table C.2)

TABLE C.2

Artículos	Número	Uso diario promedio	Edad	Ocupación
TV				
Radio				
Celular				
Luces interior				
Luces exteriores				
Refrigerador				
Notebook				
Plancha eléctrica				
Hervidor				

7) Cuál es el consumo promedio por mes? kWh

8) Cuál es la factura promedia del hogar? CLP

9) Cuantas horas de servicio eléctrico tienes al día? y después del atardecer?

10) Podrías estimar cuantos cortes de luz ocurren por mes? (Table C.3)

TABLE C.3

	N^o de cortes por mes
Menos 10 minutos	
10 minutos a 1 hora	
1 hora a 2 horas	
Mas 2 horas	

11) Alguna vez ha tenido algún dispositivo dañado o roto debído a las fluctuaciones de voltaje de la red eléctrica?

12) Como paga su factura?

13) Alguna vez ha tenido accidentes con electricidad en su hogar?

14) Tienes algún comentario sobre el suministro de electricidad?

C.3 Cocina

15) Leña para cocinar:

a. Cuáles son los sistemas de cocción a leña utilizados?
   - Fuego con piedras
   - Estufa de leña
   - Estufa avanzada
   - Otros (Especificar ....)

b. Dónde se coloca la cocina a leña en su casa?
   - Dentro de la casa
   - Fuera de la casa
   - Cocina común con otros hogares

c. Sobre una base semanal, cuántas comidas cocinas usando leña?

d. Cuánto tiempo se necesita para que el equipo esté listo para cocinar?

e. Alguna vez ha tenido accidentes con leña en su hogar?

16) Gas para cocinar:

a. Sobre una base semanal, cuántas comidas cocinas usando gas?

b. Alguna vez ha tenido accidentes con gas en su hogar?

C.4 Calentamiento de Espacios

17) Con que frecuencia utiliza estos combustibles para calefacción? (Table C.4)

TABLE C.4

	Gas	Leña
Diariamente		
Semanalmente		
Mensualmente		
Anualmente		

18) Calidad de calefacción (Table C.5)

TABLE C.5

	Gas	Leña
Cuantas habitaciones de la casa son calentadas?		
Por cuanto tiempo tienes la calefacción que necesitas?		
Por cuanto tiempo tienes una temperatura agradable?		
En el último año tuvo accidentes en su uso		

C.5 calentamieto de agua

19) Con qué frecuencia calienta el agua sanitaria de esta manera? (Table C.6)

TABLE C.6

	Gas	Leña	Colector solar
Diariamente			
Semanalmente			
Mensualmente			
Anualmente			

C.6 Provisión de Combustibles

20) De dónde saca la leña?

a. Dentro del Pueblo, gratis
b. Dentro del Pueblo comprándola,
c. fuera del pueblo de manera gratuita
d. Fuera del Pueblo comprándola
e. Otros (Especificar)................

21) Cómo se transporta la leña a su hogar?

a. A mano
b. Con moto
c. Con auto
d. Con camioneta
e. Con camión
f. Otros (Especificar) ....

22) Cuánto tiempo se necesita para adquirir y preparar la leña?

23) Cuánto compras? (Use el medio de transporte arriba como unidad de medida)

24) Cuánto pagas por la leña? .... CLP

26) Qué tipo de botella de gas tiene (Tamaño y Marca)?

27) Con qué frecuencia lo vuelve a llenar en promedio? Cada ....

28) Cuánto pagas? ......CLP

C.7 Otras fuentes de energia

29) Utiliza otras fuentes de energia? (Table C.7)

TABLE C.7

	si/no	Proposito principal	En que cantidad	Cual es el costo
Carbon				
Kerosene				
Velas				
Otros				

C.8 Producción de residuos domésticos

30) Tiene una granja? Si No

31) Qué tipo de cultivos tiene? (Table D.4)

TABLE C.8

	si/no	Cual es la extension aproximada?
Choclo		
Porotos		
Cereales		
Papas		
Pasto		
Vegetales		
Otros árboles frutales		

32) Consume los productos de su granja?

a. Sí, consumimos todos los productos

b. Sólo se consume parte, se vende el resto

c. Todo se vende

33) Qué animales tiene usted (si los hay)? (Table C.9)

TABLE C.9

	Numero	Como los alimentas ?
Cerdos		
Cabras		
Vacas		
Conejos		
Llamas		
Aves de corral		
Otros árboles frutales		

Appendix D

Ramp input data

D.1 Hospitals RAMP input data

TABLE D.1: Ramp input data for Hospitals.

Appliance	n	P [W]	Cycle [min]	Tot Use [min]	Start W1	End W1	Start W2	End W2
Indoor Bulb**	12	7	10	690	08:00	12:00	14:30	24:00
Outdoor Bulb**	1	13	10	690	00:00	05:30	17:30	24:00
Cellphone charger	8	2	5	300	08:00	12:00	15:00	24:00
Fridge*	3	150	30	1440	00:00	24:00	-	-
PC	2	50	10	300	08:00	12:00	17:30	24:00
Mixer	1	50	1	60	08:00	12:00	17:30	24:00

D.2 Schools RAMP input data

TABLE D.2: Ramp input data for Schools.

Appliance	n	P [W]	Cycle [min]	Tot Use [min]	Start W1	End W1	Start W2	End W2
Indoor Bulb**	8	7	10	60	17:00	18:00	-	-
Outdoor Bulb**	6	13	10	60	17:00	18:00	-	-
Phone Charger	5	2	5	180	08:30	12:30	13:30	18:00
PC	18	50	10	210	08:30	12:30	13:30	18:00
Printer	1	20	5	30	08:30	12:30	13:30	18:00
Fridge*	1	200	30	1440	00:00	24:00	-	-
TV	1	60	5	120	08:30	12:30	13:30	18:00
DVD	1	8	5	120	08:30	12:30	13:30	18:00
Stereo	1	150	5	90	08:30	12:30	13:30	18:00

D.3 LowLands community RAMP input data

TABLE D.3: Ramp input data for lowlands communities.

User Class	Appliance	n	P [W]	Cycle [min]	Tot Use [min]	Start W1	End W1	Start W2	End W2
HC	Indoor Bulb**	6	7	10	120	19:30	24:00	00:00	00:30
	Outdoor Bulb**	2	13	10	600	19:30	24:00	00:00	05:30
	TV	2	60	5	180	12:00	15:00	19:30	01:00
	DVD	1	8	5	60	12:00	15:00	19:30	01:00
	Antenna	1	8	5	120	12:00	15:00	19:30	01:00
	Cellphone	5	2	5	300	18:30	24:00	00:00	00:30
	Fridge*	1	200	30	1440	00:00	24:00	-	-
	Mixer	1	50	1	30	07:00	08:00	11:00	12:30
LC	Indoor Bulb**	2	7	10	120	19:30	24:00	00:00	00:30
	Outdoor Bulb**	1	13	10	600	19:30	24:00	00:00	05:30
	TV	1	60	5	90	12:30	14:00	19:30	00:30
	DVD	1	8	5	30	12:30	14:00	19:30	00:30
	Antenna	1	8	5	60	12:30	14:00	19:30	00:30
	Cellphone	2	2	5	300	18:00	24:00	-	-
Church	Indoor Bulb**	10	26	60	210	20:00	24:00	-	-
	Outdoor Bulb**	7	26	60	240	20:00	24:00	-	-
	Speaker	1	100	60	240	20:00	22:30	-	-
Public Lighting	Lights**	***	150	300	600	19:00	24:00	00:00	06:00

D.4 Hihglands communities RAMP input data

TABLE D.4: Ramp input data for highlands communities.

User Class	Appliance	n	P [W]	Cycle [min]	Tot Use [min]	Start W1	End W1	Start W2	End W2
HC	Indoor Bulb**	7	7	10	300	18:00	24:00	00:00	00:30
	Outdoor Bulb**	1	13	30	300	20:00	24:00	00:00	07:00
	TV	2	60	30	360	09:00	13:00	18:00	24:00
	Radio	1	7	30	240	08:00	12:00	18:00	23:00
	Cellphone	4	5	10	360	20:00	24:00	00:00	07:00
	Fridge*	1	250	30	1440	00:00	24:00	-	-
	Laptop	1	70	30	90	16:00	24:00	-	-
	Iron	1	700	1	30	10:00	20:00	-	-
LC	Indoor Bulb**	5	7	10	300	18:00	24:00	00:00	00:30
	Outdoor Bulb**	1	13	30	60	20:00	24:00	00:00	07:00
	TV	1	60	30	240	09:00	13:00	18:00	24:00
	Radio	1	7	30	240	08:00	12:00	18:00	23:00
	Cellphone	2	5	10	360	20:00	24:00	00:00	07:00
	Iron	1	700	1	30	10:00	20:00	-	-
Church	Indoor Bulb**	10	26	60	210	20:00	24:00	-	-
	Outdoor Bulb**	7	26	60	240	20:00	24:00	-	-
	Speaker	1	100	60	240	20:00	22:30	-	-
Public lighting	Lights**	***	150	300	600	19:00	24:00	00:00	06:00

* Fridge follow a specific ad-hoc cycle, is not functioning full power for 24h
** Total duration and functioning windows depending on seasonal sunrise and sunset times
*** Depending on population size

Appendix E

Analysis of the PV monitoring data

In order to determine the relevant features for the prediction of the PV performance from the monitoring data, a matrix of scatter plots (Figure E.1) and a covariance matrix (Figure E.2) are plotted. From these, it appears that the module temperature (PV temperature 2), the ambient temperature, the hour of the day and the solar irradiation should be taken into account when predicting the performance of the solar array.

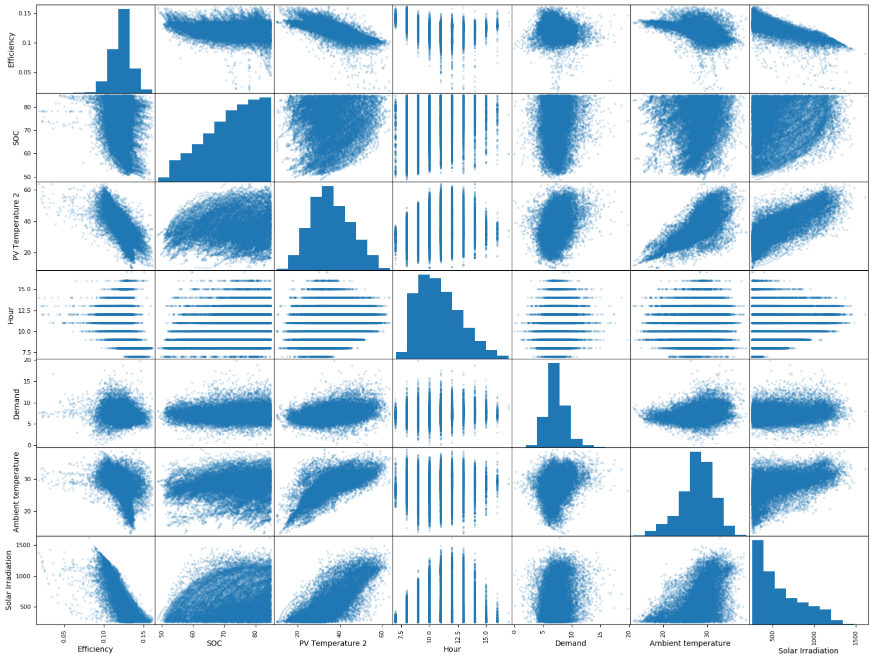

FIGURE E.1: Matrix of scatter plots for each variable

Índice	Efficiency	SOC	PV Temperature 2	Hour	Demand	Ambient temperature	Solar Irradiation
Efficiency	0.0001750298…	-0.0187130655159446	-0.0891441464582999	-0.0077417734320297	-0.00290954123148718	-0.0215924168876024	-2.78985464579991
SOC	-0.018713065…	74.2766482976215	5.36213505108898	4.27400570063379	0.494707450859872	1.71676643872333	343.35256296799
PV Temperature 2	-0.089144146…	5.36213505108898	89.8250920340669	3.39203766771127	5.45219840286553	24.9516579343256	1952.96081494307
Hour	-0.007741773…	4.27400570063379	3.39203766771127	3.66644434780292	0.692819176530804	0.661612020551953	84.7646094214399
Demand	-0.002909541…	0.494707450859872	5.45219840286553	0.692819176530804	2.85948939625868	1.65162627685584	80.893820417212
Ambient temperature	-0.021592416…	1.71676643872333	24.9516579343256	0.661612020551953	1.65162627685584	14.4768922816192	532.658473767393
Solar Irradiation	-2.789854645…	343.35256296799	1952.96081494307	84.7646094214399	80.893820417212	532.658473767393	76636.5385422866

FIGURE E.2: Covariance matrix between the PV variables

Appendix F

OnSSET input data

F.1 Geospatial datasets and assumptions

TABLE F.1: Open-source GIS data used in the model.

Dataset	Description	Type	Source
Population and electrification Census	Information of each Bolivian community.	Point vector	[134]
Administrative boundaries	Delineates the boundaries of the analysis.	Line vector	[135]
Existing grid network	Used to estimate the costs of grid extension.	Line vector	[136, 137]
Substations	Used to specify grid extension suitability.	Point vector	[138]
Roads	Used to specify grid extension suitability.	Line vector	[139]
Planned grid network	Planned extension of the national electric grid.	Line vector	[140, 141]
Travel time	Travel time to the closest town (>50 000 people).	Raster	[142]
Digital Elevation Map	Altitude in meters above sea level.	Raster	[143]
Land Cover	Land cover maps.	Raster	[144]
Poverty	socio-economic information of the population .	Polygon vector	[145]

F.2 Socio-economic parameters used in the electrification model for Bolivia

TABLE F.2: Socio-economic parameters used in the electrification model for Bolivia.

Parameter	Metric	Value 2012	Value 2025
Population total	Million persons	10 351 118	12454178
Urban population	Percent of total population	67.71%	72.80%
Electricity access	Percent of total population	82.08%	100%
Urban household size	People per household	3.84	3.84
Rural household size	People per household	3.41	3.41

F.3 Techno-economic parameters related to the grid connected technologies

TABLE F.3: Techno-economic parameters related to the grid connected technologies.

Technology type	Expected capacity in 2025, MW	% share	Investment costs, $/kW
Hydropower	2393	43%	2100
Gas combined cycle	2698	47%	1140
Solar power	160	3%	1400
Wind power	180	3%	1320
Biomass thermal	62	1%	2200
Geothermal	110	1%	590
Diesel	35	2%	5218

F.4 Techno-economic characteristics for classic OnSSET off-grid technologies

TABLE F.4: Techno-economic parameters related to off-grid technologies for OnSSET classic.

	SHS	PV microgrid	Diesel Microgrid
O&M of tramission lines (%)	0	0.02	0.02
Connection per household (USD)	0	125	125
Base to peak ratio	0.9	0.9	0.529
Teach life (years)	15	20	20
O&M costs (%)	0.02	0.02	0.02
Capital cost (USD/kW)		3500	1480

TABLE F.5: Capital cost for SHS for OnSSET classic.

Capacity (kW)	Capital Cost (USD/kW)
0.02	20000
0.05	11050
0.1	7660
0.2	5780
Others	5070

Bibliography

[1] United Nations Framework Convention on Climate Change (UNFCCC). *Adoption of the Paris Agreement – Report No. FCCC/CP/2015/L.9/Rev.1*. 2015. URL: http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

[2] IRENA. *Tracking SDG 7: The Energy Progress Report (2020)*. IEA, 2020. URL: https://www.irena.org/publications/2020/May/Tracking-SDG7-The-Energy-Progress-Report-2020.

[3] IEA. *World Energy Outlook 2018*. IEA, 2018. URL: https://www.iea.org/reports/world-energy-outlook-2018.

[4] EA, IRENA, UNSD, World Bank, WHO. *Tracking SDG 7: The Energy Progress Report*. https://trackingsdg7.esmap.org/data/files/download-documents/2021_tracking_sdg7_report.pdf(visited 03/10/2021). 2021.

[5] Pedro Ciller and Sara Lumbreras. "Electricity for all: The contribution of large-scale planning tools to the energy-access problem". In: *Renewable and Sustainable Energy Reviews* 120 (2020), p. 109624.

[6] Seán Collins et al. "Integrating short term variations of the power system into integrated energy system models: A methodological review". In: *Renewable and Sustainable Energy Reviews* 76 (2017), pp. 839–856.

[7] Rhonda LeNai Jordan. "Incorporating endogenous demand dynamics into long-term capacity expansion power system models for Developing countries". PhD thesis. Massachusetts Institute of Technology, 2013.

[8] JG Peña Balderrama et al. "Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET)". In: *Energy for Sustainable Development* 56 (2020), pp. 98–118.

[9] Dimitrios Mentis et al. "Lighting the World: the first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa". In: *Environmental Research Letters* 12.8 (2017), p. 085003.

[10] Jiangyu Dai et al. "Water-energy nexus: A review of methods and tools for macro-assessment". In: *Applied energy* 210 (2018), pp. 393–408.

[11] Sergio Balderrama et al. "Surrogate models for rural energy planning: Application to Bolivian lowlands isolated communities". In: *Energy* (2021), p. 121108.

[12] Markus Groissböck. "Are open source energy system optimization tools mature enough for serious use?" In: *Renewable and Sustainable Energy Reviews* 102 (2019), pp. 234–248.

[13] Stefan Pfenninger et al. "The importance of open data and software: Is energy research lagging behind?" In: *Energy Policy* 101 (2017), pp. 211–215.

[14] UN News Centre. "Transforming our world: The 2030 Agenda for Sustainable Development". In: *United Nations Department of Economic and Social Affairs* (2015). URL: http://www.un.org/en/development/desa/news/sustainable/un-adopts-new-global-goals.html#more-15178.

[15] Makoto Kanagawa and Toshihiko Nakata. "Assessment of access to electricity and the socio-economic impacts in rural areas of developing countries".

In: *Energy Policy* 36.6 (2008), pp. 2016–2029. ISSN: 03014215. DOI: 10.1016/j.enpol.2008.01.041.

[16] Fabio Riva et al. "Electricity access and rural development: Review of complex socio-economic dynamics and casual diagrams for more appropriate energy modelling". In: *Energy for sustainable development* 43 (2018), pp. 203–223. DOI: https://doi.org/10.1016/j.esd.2018.02.003.

[17] Nicolo' Stevanato et al. "Modeling of a Village-Scale Multi-Energy System for the Integrated Supply of Electric and Thermal Energy". In: *Applied Sciences* 10.21 (2020), p. 7445.

[18] Jhonny Villarroel-Schneider et al. "Biogas based polygeneration plant options utilizing dairy farms waste: A Bolivian case". In: *Sustainable Energy Technologies and Assessments* 37 (2020), p. 100571.

[19] Golam Rasul and Bikash Sharma. "The nexus approach to water–energy–food security: an option for adaptation to climate change". In: *Climate Policy* 16.6 (2016), pp. 682–702.

[20] Samuel Asumadu Sarkodie and Phebe Asantewaa Owusu. "Bibliometric analysis of water–energy–food nexus: Sustainability assessment of renewable energy". In: *Current Opinion in Environmental Science & Health* 13 (2020), pp. 29–34.

[21] Fabio Riva et al. "Soft-linking energy demand and optimisation models for local long-term electricity planning: An application to rural India". In: *Energy* 166 (2019), pp. 32 –46. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2018.10.067.

[22] Kirsten Ulsrud et al. "The Solar Transitions research on solar mini-grids in India: Learning from local cases of innovative socio-technical systems". In: *Energy for Sustainable Development* 15.3 (2011). Special issue on off-grid electrification in developing countries, pp. 293 –303. ISSN: 0973-0826. DOI: https://doi.org/10.1016/j.esd.2011.06.004.

[23] Bo Zhao et al. "Optimal sizing, operating strategy and operational experience of a stand-alone microgrid on Dongfushan Island". In: *Applied Energy* 113 (2014), pp. 1656 –1666. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2013.09.015.

[24] P. Diaz et al. "FAR from the grid: A rural electrification field study". In: *Renewable Energy* 35.12 (2010), pp. 2829 –2834. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2010.05.005.

[25] Toru Kobayakawa and Tara C. Kandpal. "Analysis of electricity consumption under a photovoltaic micro-grid system in India". In: *Solar Energy* 116 (2015), pp. 177 –183. ISSN: 0038-092X. DOI: https://doi.org/10.1016/j.solener.2015.04.001.

[26] Fabio Riva et al. "Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies and insights from a modelling perspective". In: *Energy strategy reviews* 20 (2018), pp. 71–89. DOI: https://doi.org/10.1016/j.esr.2018.02.006.

[27] Sergio Balderrama et al. "A two-stage linear programming optimization framework for isolated hybrid microgrids in a rural context: The case study of the âĂIJEl EspinoâĂİ community". In: *Energy* (2019), p. 116073.

[28] Theocharis D Tsoutsos and Yeoryios A Stamboulis. "The sustainable diffusion of renewable energy technologies as an example of an innovation-focused policy". In: *Technovation* 25.7 (2005), pp. 753–761.

[29] Peter S Hofman, Boelie E Elzen, and Frank W Geels. "Sociotechnical scenarios as a new policy tool to explore system innovations: Co-evolution of technology and society in the NetherlandâĂŹs electricity domain". In: *Innovation* 6.2 (2004), pp. 344–360.

[30] T. Darby, S. and Fawcett. "Energy sufficiency: an introduction (concept paper)". In: *European Council for an Energy Efficient Economy (ECEEE)* (2018). URL: https://www.energysufficiency.org/static/media/uploads/site-8/library/papers/sufficiency-introduction-final-oct2018.pdf.

[31] Sarah Darby and Tina Fawcett. "Energy sufficiency: an introduction Concept paper". In: *Energy Sufficiency project, ECEEE* (2018). https://www.energysufficiency.org/static/media/uploads/site-8/library/papers/sufficiency-introduction-final-oct2018.pdf(visited 26/08/2021).

[32] Stefano Mandelli, Marco Merlo, and Emanuela Colombo. "Novel procedure to formulate load profiles for off-grid rural areas". In: *Energy for sustainable development* 31 (2016), pp. 130–142.

[33] Lukas G Swan and V Ismet Ugursal. "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques". In: *Renewable and sustainable energy reviews* 13.8 (2009), pp. 1819–1835.

[34] Francesco Lombardi, Matteo Vincenzo Rocco, and Emanuela Colombo. "A multi-layer energy modelling methodology to assess the impact of heat-electricity integration strategies: The case of the residential cooking sector in Italy". In: *Energy* 170 (2019), pp. 1249–1260.

[35] JM Nichols et al. "A simple algorithm for generating spectrally colored, non-Gaussian signals". In: *Probabilistic Engineering Mechanics* 25.3 (2010), pp. 315–322.

[36] Thomas Schreiber and Andreas Schmitz. "Improved surrogate data for nonlinearity tests". In: *Physical Review Letters* 77.4 (1996), p. 635.

[37] Dimitris Kugiumtzis and Efthimia Bora-Senta. "Simulation of multivariate non-gaussian autoregressive time series with given autocovariance and marginals". In: *Simulation Modelling Practice and Theory* 44 (2014), pp. 42–53.

[38] Francesco Lombardi et al. "Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model". In: *Energy* 177 (2019), pp. 433–444.

[39] N. Stevanato et al. "Two-Stage Stochastic Sizing of a Rural Micro-Grid Based on Stochastic Load Generation". In: *2019 IEEE Milan PowerTech*. 2019, pp. 1–6. DOI: 10.1109/PTC.2019.8810571.

[40] Rahman Saidur, Haji Hassan Masjuki, and IA Choudhury. "Role of ambient temperature, door opening, thermostat setting position and their combined effect on refrigerator-freezer energy consumption". In: *Energy Conversion and Management* 43.6 (2002), pp. 845–854.

[41] Jasmin Geppert and Rainer Stamminger. "Analysis of effecting factors on domestic refrigeratorsâĂŹ energy consumption in use". In: *Energy Conversion and Management* 76 (2013), pp. 794–800.

[42] Shonali Pachauri et al. "Energy access for development". In: (2012).

[43] Adrian J Bradbrook and Judith G Gardam. "Placing access to energy services within a human rights framework". In: *Human Rights Quarterly* (2006), pp. 389–415.

[44] Ronald Muhumuza et al. "Energy consumption levels and technical approaches for supporting development of alternative energy technologies for rural sectors of developing countries". In: *Renewable and Sustainable Energy Reviews* 97 (2018), pp. 90–102.

[45] Kamil Kaygusuz. “Energy services and energy poverty for sustainable rural development”. In: *Renewable and sustainable energy reviews* 15.2 (2011), pp. 936–947.

[46] Reza Kowsari and Hisham Zerriffi. “Three dimensional energy profile:: A conceptual framework for assessing household energy use”. In: *Energy Policy* 39.12 (2011), pp. 7505–7517.

[47] Kate O’Sullivan, Oleg Golubchikov, and Abid Mehmood. “Uneven energy transitions: Understanding continued energy peripheralization in rural communities”. In: *Energy Policy* 138 (2020), p. 111288.

[48] Annemarije L Kooijman-van Dijk and Joy Clancy. “Impacts of electricity access to rural enterprises in Bolivia, Tanzania and Vietnam”. In: *Energy for Sustainable Development* 14.1 (2010), pp. 14–21.

[49] Kathleen Klaniecki et al. “Applying the energy cultures framework to understand energy systems in the context of rural sustainability transformation”. In: *Energy Policy* 137 (2020), p. 111092.

[50] Markus Kottek et al. “World Map of the Köppen-Geiger climate classification updated”. In: *Meteorologische Zeitschrift* 15.3 (July 2006), pp. 259–263. DOI: 10.1127/0941-2948/2006/0130. URL: http://dx.doi.org/10.1127/0941-2948/2006/0130.

[51] Instituto Nacional de Estadistica. *Estadisticas Demograficas de Bolivia*. 2018.

[52] Ministerio de Hidrocarburos y Energia. *Plan ElÃl’ctrico del Estado Plurinacional de Bolivia 2015âĂŞ2025*. 2014.

[53] *Bolivian Constitution*. http://www.sepdavi.gob.bo/cs/doc/159BoliviaConsitucion.pdf(visited 03/10/2021). 2009.

[54] Ministerio de Hidrocarburos y Energia, ComitÃl’ Nacional de Despacho de Carga, and ENDE Corporacion. *Plan optimo de Expansion del Sistema Interconectado Nacional 2012-2022, Year = 2012, note = https://observato-rioccdbolivia.files.wordpress.com/2015/08/plan-de-expan-sic3b3n-del-sin-2012-2022.pdf (visited 03/10/2021)*.

[55] Estado Plurinacional de Bolivia, Ministerio de Hidrocarburos y Energia, and Viceministerio de Electricidad y Energias Alternativas. *Plan ElÃl’ctrico del Estado Plurinacional de Bolivia 2025*. https://observatorioccdbolivia.files.wordpress.com/2015/08/peebol2025.pdf(visited 03/10/2021). 2014.

[56] Gomez, Enrique. *Evaluacion Rapida Del Sector Energia En Bolivia*. 2010.

[57] Espinasa, Ramons and Sucre, Carlos and Gutierrez, Marta and Anaya Fernando. *Dossier Energetico: Bolivia*. 2017. DOI: http://dx.doi.org/10.18235/0000586.

[58] Ministerio de Hidrocarburos y Energia. *Balance energÃl’tico nacional 2000 - 2014*. 2015.

[59] ComitÃl’ Nacional de Despacho de Carga, Ministerio de Energias, Estado Plurinacional de Bolivia. *MEMORIA 2016. Resultados de la operacion del SIN., Year = 2016, note = https://observatorioccdbolivia.files.wordpress.com/2015/08/peebol2025.pdf (visited 03/10/2021)*.

[60] Empresa Nacional De ElectricidadâĂŞCorporacion. *Memoria Anual 2015, Year = 2015, note = https://www.ende.bo/public/memorias/memoria-2015-10-final.pdf (visited 03/10/2021)*.

[61] BALANCE ENERGÃL’TICO NACIONAL 2000-2014. *Memoria Anual 2015, Year = 2015, note = https://www.ende.bo/public/memorias/memoria-2015-10-final.pdf (visited 03/10/2021)*.

[62] Autoridad de Electricidad. *Anuario Estadistico 2015, Year = 2015, note = https://www.ende.bo/public/memorias/memoria-2015-10-final.pdf (visited 03/10/2021).*
[63] C. Sanchez, E. Cardozo, and S. Quoilin. "Building innovation systems for renewable energy diffusion in developing countries: A review on systemic problems". In: 2020.
[64] Mario Pansera. "Renewable energy for rural areas of Bolivia". In: *Renewable and sustainable energy reviews* 16.9 (2012), pp. 6694–6704.
[65] M Lucano and I Fuentes. "Evaluation of the global solar radiation potential in the department of Cochabamba (Bolivia) using models of geographic information systems and satellite images". In: *Revista Boliviana de Física* 17.1562-3823 (2010), pp. 13–21.
[66] M Lucano and I Fuentes. *Atlas de radiacion solar global de Bolivia*. 2010.
[67] NE, VMEEA. *Cobertura del Servicio ElÃl'ctrico a Nivel Comunidades Censo 2012, Year = 2015, note = http://sigvmeea.minenergias.gob.bo/maps/73 (visited 13/05/2019).*
[68] Francesco Fuso Nerini et al. "A cost comparison of technology approaches for improving access to electricity services". In: *Energy* 95 (2016), pp. 255–265.
[69] Dante Ayaviri Nina, Ayaviri Nina, and Alarcón Lorenzo. "Clasificación socioeconómica de los municipios de Bolivia Socio-economic classification of the municipalities of Bolivia". In: (), pp. 29–55.
[70] Maria F. Gomez and Semida Silveira. "Rural electrification of the Brazilian Amazon âĂŞ Achievements and lessons". In: *Energy Policy* 38.10 (2010), pp. 6251 –6260. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2010.06.013.
[71] Dimitrios Mentis et al. "A GIS-based approach for electrification planningâĂŤA case study on Nigeria". In: *Energy for Sustainable Development* 29 (2015), pp. 142 –150. ISSN: 0973-0826. DOI: https://doi.org/10.1016/j.esd.2015.09.007.
[72] Francesco Fuso Nerini et al. "A cost comparison of technology approaches for improving access to electricity services". In: *Energy* 95 (2016), pp. 255 –265. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2015.11.068.
[73] James T Murphy. "Making the energy transition in rural East Africa: Is leapfrogging an alternative?" In: *Technological Forecasting and Social Change* 68.2 (2001), pp. 173–193. DOI: https://doi.org/10.1016/S0040-1625(99)00091-8.
[74] B. Pickering and R. Choudhary. "District energy system optimisation under uncertain demand: Handling data-driven stochastic profiles". In: *Applied Energy* 236.September 2018 (2019), pp. 1138–1157. ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.12.037.
[75] Xiufeng Yue et al. "A review of approaches to uncertainty assessment in energy system optimization models". In: *Energy Strategy Reviews* 21.July (2018), pp. 204–217. ISSN: 2211-467X. DOI: 10.1016/j.esr.2018.06.003.
[76] Fabio Riva et al. "Soft-linking energy demand and optimisation models for local long-term electricity planning: An application to rural India". In: *Energy* 166 (2019), pp. 32–46.
[77] Fabio Riva, Emanuela Colombo, and Carlo Piccardi. "Towards modelling diffusion mechanisms for sustainable off-grid electricity planning". In: *Energy for Sustainable Development* 52 (2019), pp. 11–25.
[78] Laia Ferrer-Martí et al. "Optimizing microwind rural electrification projects. A case study in Peru". In: *Journal of Global Optimization* 50.1 (2011), pp. 127–143.

[79] Stefano Mandelli et al. “A sizing methodology based on Levelized Cost of Supplied and Lost Energy for off-grid rural electrification systems”. In: *Renewable Energy* 89 (2016), pp. 475–488.

[80] Matthew S Orosz and Amy V Mueller. “Dynamic simulation of performance and cost of hybrid PV-CSP-LPG generator micro grids with applications to remote communities in developing countries”. In: *ASME 2015 9th International Conference on Energy Sustainability collocated with the ASME 2015 Power Conference, the ASME 2015 13th International Conference on Fuel Cell Science, Engineering and Technology, and the ASME 2015 Nuclear Forum*. American Society of Mechanical Engineers. 2015, V002T13A007–V002T13A007.

[81] Diana Neves, Carlos A Silva, and Stephen Connors. “Design and implementation of hybrid renewable energy systems on micro-communities: A review on case studies”. In: *Renewable and Sustainable Energy Reviews* 31 (2014), pp. 935–946.

[82] Sara Ghaem Sigarchian et al. “Optimum design of a hybrid PVâĂŞCSPâĂŞLPG microgrid with Particle Swarm Optimization technique”. In: *Applied Thermal Engineering* 109 (2016). Special Issue: Solar Energy Research Institute for India and the United States (SERIIUS), Concentrated Solar Power, pp. 1031 –1036. ISSN: 1359-4311. DOI: https://doi.org/10.1016/j.applthermaleng.2016.05.119.

[83] Q. Altes Buch et al. “Rule-based control and optimization of a hybrid solar microgrid for rural electrification and heat supply in sub-Saharan Africa.” In: *Proceedings of the 30th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems* 1.1 (2017), pp. 1263–1273.

[84] Claudio Brivio et al. “A novel software package for the robust design of off-grid power systems”. In: *Journal of Cleaner Production* 166 (2017), pp. 668–679.

[85] Amin Khodaei, Shay Bahramirad, and Mohammad Shahidehpour. “Microgrid planning under uncertainty”. In: *IEEE Trans. Power Syst* 30.5 (2015), pp. 2417–2425.

[86] Zhe Zhou et al. “A two-stage stochastic programming model for the optimal design of distributed energy systems”. In: *Applied Energy* 103 (2013), pp. 135–144.

[87] Li Guo et al. “Multi-objective stochastic optimal planning method for stand-alone microgrid system”. In: *IET Generation, Transmission & Distribution* 8.7 (2014), pp. 1263–1273.

[88] Jay H Lee. “Energy supply planning and supply chain optimization under uncertainty”. In: *Journal of Process Control* 24.2 (2014), pp. 323–331.

[89] Jing Qiu et al. “Planning of solar photovoltaics, battery energy storage system and gas micro turbine for coupled micro energy grids”. In: *Applied Energy* 219 (2018), pp. 361–369.

[90] Yun Yang, Shijie Zhang, and Yunhan Xiao. “Optimal design of distributed energy resource systems based on two-stage stochastic programming”. In: *Applied Thermal Engineering* 110 (2017), pp. 1358–1370.

[91] Apurva Narayan and Kumaraswamy Ponnambalam. “Risk-averse stochastic programming approach for microgrid planning under uncertainty”. In: *Renewable Energy* 101 (2017), pp. 399–408.

[92] Davide Fioriti et al. “Stochastic sizing of isolated rural mini-grids, including effects of fuel procurement and operational strategies”. In: *Electric Power Systems Research* 160 (2018), pp. 419–428.

[93] International Energy Agency. "Energy Access Outlook 2017: From Poverty to Prosperity". In: *IEA: Paris, France* (2017), p. 144.
[94] William E. Hart et al. *Pyomo–optimization modeling in python*. Second. Vol. 67. Springer Science & Business Media, 2017.
[95] Bryan Stephen Palmintier. "Incorporating operational flexibility into electric generation planning: Impacts and methods for system design and policy analysis". PhD thesis. Massachusetts Institute of Technology, 2013.
[96] AndrÃľ Malheiro et al. "Integrated sizing and scheduling of wind/PV/diesel/battery isolatedÂăsystems". In: *Renewable Energy* 83 (2015), pp. 646 –657. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2015.04.066.
[97] Rodolfo Dufo-Lopez and JosÃľ L. Bernal-Agustin. "Multi-objective design of PVâĂŞwindâĂŞdieselâĂŞhydrogenâĂŞbattery systems". In: *Renewable Energy* 33.12 (2008), pp. 2559 –2572. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2008.02.027.
[98] William F Holmgren, Clifford W Hansen, and Mark A Mikofski. "pvlib python: a python package for modeling solar energy systems". In: *The Journal of Open Source Software* 3 (2018), p. 884.
[99] Wilhelm Durisch et al. "Efficiency model for photovoltaic modules and demonstration of its application to energy yield estimation". In: *Solar energy materials and solar cells* 91.1 (2007), pp. 79–84.
[100] Martin Kittel and Wolf-Peter Schill. *Renewable Energy Targets and Unintended Storage Cycling: Implications for Energy Modeling*. 2021. arXiv: 2107.13380 [econ.GN].
[101] Ed Klotz and Alexandra M. Newman. "Practical guidelines for solving difficult mixed integer linear programs". In: *Surveys in Operations Research and Management Science* 18.1 (2013), pp. 18–32. ISSN: 1876-7354. DOI: https://doi.org/10.1016/j.sorms.2012.12.001. URL: https://www.sciencedirect.com/science/article/pii/S1876735413000020.
[102] Sylvain Quoilin et al. "Quantifying self-consumption linked to solar home battery systems: Statistical analysis and economic assessment". In: *Applied Energy* 182 (2016), pp. 58 –67. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2016.08.077.
[103] Bart Van Campen, Daniele Guidi, and Gustavo Best. "Solar photovoltaics for sustainable agriculture and rural development". In: *Rural DevelopmentâĂİ, Fao Publication*. Citeseer. 2000. URL: fao.org/uploads/media/SolarphotovoltaicforSARD.
[104] Fabian Benavente et al. "Photovoltaic/battery system sizing for rural electrification in Bolivia: Considering the suppressed demand effect". In: *Applied energy* 235 (2019), pp. 519–528.
[105] Raphael Ebenhoch et al. "Comparative levelized cost of energy analysis". In: *Energy Procedia* 80 (2015), pp. 108–122. URL: https://upcommons.upc.edu/bitstream/handle/2117/80329/1-s2.0-S1876610215021451-main.pdf.
[106] Donald E Grierson. "Pareto multi-criteria decision making". In: *Advanced Engineering Informatics* 22.3 (2008), pp. 371–384.
[107] Indraneel Das. "On characterizing the "knee" of the Pareto curve based on normal-boundary intersection". In: *Structural optimization* 18.2-3 (1999), pp. 107–115.
[108] Wenhua Li et al. "Knee point-guided multiobjective optimization algorithm for microgrid dynamic energy management". In: *Complexity* 2020 (2020).
[109] Leandro Homrich Lorentz, Ronaldo Erichsen, and Alessandro Dal'Col Lúcio. "Proposta de método para estimação de tamanho de parcela para culturas agr{\'\i}colas". In: *Revista Ceres* 59.6 (2012), pp. 772–780. URL: https:

//www.scielo.br/scielo.php?script=sci{_}arttext{\&}pid=S0034-737X2012000600006.

[110] Aaron Meurer et al. "SymPy: symbolic computing in Python". In: *PeerJ Computer Science* 3 (2017), e103.

[111] Stefan Pfenninger and Iain Staffell. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". In: *Energy* 114 (2016), pp. 1251–1265.

[112] Peter Lopion et al. "A review of current challenges and trends in energy systems modeling". In: *Renewable and sustainable energy reviews* 96 (2018), pp. 156–166.

[113] Stefan Pfenninger, Adam Hawkes, and James Keirstead. "Energy systems modeling for twenty-first century energy challenges". In: *Renewable and Sustainable Energy Reviews* 33 (2014), pp. 74–86.

[114] Douglas Ellman. "The reference electrification model: a computer model for planning rural electricity access". PhD thesis. Massachusetts Institute of Technology, 2015.

[115] Catherina Cader, Philipp Blechinger, and Paul Bertheau. "Electrification planning with focus on hybrid mini-grids–A comprehensive modelling approach for the Global South". In: *Energy Procedia* 99 (2016), pp. 269–276.

[116] JG Peña et al. "Incorporating high-resolution demand and techno-economic optimization to evaluate micro-grids into the Open Source Spatial Electrification Tool (OnSSET)". In: *Energy for Sustainable Development* 56 (2020), pp. 98–118.

[117] Dimitrios Mentis et al. "A GIS-based approach for electrification planningâĂŤA case study on Nigeria". In: *Energy for Sustainable Development* 29 (2015), pp. 142–150.

[118] Alexandros Korkovelos et al. "A geospatial assessment of small-scale hydropower potential in sub-Saharan Africa". In: *Energies* 11.11 (2018), p. 3100.

[119] Alexandros Korkovelos et al. "The role of open access data in geospatial electrification planning and the achievement of SDG7. an OnSSET-based case study for Malawi". In: *Energies* 12.7 (2019), p. 1395.

[120] Amir Mosavi et al. "State of the art of machine learning models in energy systems, a systematic review". In: *Energies* 12.7 (2019), p. 1301.

[121] Baran Yildiz, Jose I Bilbao, and Alistair B Sproul. "A review and analysis of regression and machine learning models on commercial building electricity load forecasting". In: *Renewable and Sustainable Energy Reviews* 73 (2017), pp. 1104–1122.

[122] ATD Perera et al. "Machine learning methods to assist energy system optimization". In: *Applied Energy* 243 (2019), pp. 191–205.

[123] WG Wedel et al. "Improving information gain from optimization problems using artificial neural networks". In: *THE 32ND INTERNATIONAL CONFERENCE ON EFFICIENCY, COST, OPTIMIZATION, SIMULATION AND ENVIRONMENTAL IMPACT OF ENERGY SYSTEMS*. 2019.

[124] Pedro Ciller, Fernando de Cuadra, and Sara Lumbreras. "Optimizing off-grid generation in large-scale electrification-planning problems: A direct-search approach". In: *Energies* 12.24 (2019), p. 4634.

[125] Sylvain Quoilin and Jessica Schrouff. "Assessing steady-state, multivariate experimental data using Gaussian processes: the GPExp open-source library". In: *Energies* 9.6 (2016), p. 423.

[126] Sergio Luis Balderrama Subieta et al. "Automated evaluation of levelized cost of energy of isolated micro-grids for energy planning purposes in developing countries". In: *PROCEEDINGS OF ECOS 2019* (2019).
[127] Nicolò Stevanato et al. "Long-term sizing of rural microgrids: Accounting for load evolution through multi-step investment plan and stochastic optimization". In: *Energy for Sustainable Development* 58 (2020), pp. 16–29.
[128] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
[129] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.
[130] Sandor Szabo et al. "Energy solutions in rural Africa: mapping electrification costs of distributed solar and diesel generation versus grid extension". In: *Environmental Research Letters* 6.3 (2011), p. 034002.
[131] Ian Coxhead and Corbett Grainger. "Fossil fuel subsidy reform in the developing world: who wins who loses and why?" In: *Asian Development Review* 35.2 (2018), pp. 180–203.
[132] Sandor Szabó et al. "Sustainable energy planning: Leapfrogging the energy poverty gap in Africa". In: *Renewable and Sustainable Energy Reviews* 28 (2013), pp. 500–509.
[133] Sebastian Groh et al. "Swarm electrification: Investigating a paradigm shift through the building of microgrids bottom-up". In: *Decentralized Solutions for Developing Economies*. Springer, 2015, pp. 3–22.
[134] Instituto Nacional de Estadistica and Viceministerio de Electricidad y Energias Alternativas. *Cobertura del Servicio ElÃľctrico a Nivel Comunidades Censo 2012*. http://sigvmeea.minenergias.gob.bo/maps/73(visited 16/05/2019). 2012.
[135] COSIPLAN. *GeoBolivia - Zonas administrativas de Bolivia*. visited 15/10/2021. 2015. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/51b70741-35f5-42d6-84e1-f60d1f3f5169.
[136] Autoridad de fiscalizacion y control de electricidad. *GeoBolivia- High voltage transmission lines 2013*. visited 15/10/2021. 2013. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/450c9fd2-23de-4789-b249-504c3c0d334e.
[137] Autoridad de fiscalizacion y control de electricidad. *GeoBolivia- Medium voltage transmission lines 2013*. visited 15/10/2021. 2013. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/78379ddc-86d5-473e-9948-e2dfd94a7dca.
[138] Autoridad de fiscalizacion y control de electricidad. *GeoBolivia- Substations of Bolivia 2012*. visited 15/10/2021. 2012. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/02999582-b29f-400e-934b-e52dacbd657e.
[139] Geofabrik GmbH. *OpenStreetMap project*. visited 15/10/2021. 2018. URL: http://download.geofabrik.de/.
[140] Autoridad de fiscalizacion y control de electricidad. *GeoBolivia- High voltage transmission lines 2016*. visited 15/10/2021. 2016. URL: :https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/a2f37a8a-b672-4cee-b19d-07be3337ee10.
[141] Autoridad de fiscalizacion y control de electricidad. *GeoBolivia- Medium voltage transmission lines 2016*. visited 15/10/2021. 2016. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/a2f37a8a-b672-4cee-b19d-07be3337ee10.

[142] D.J. Weiss et al. “A global map of travel time to cities to assess inequalities in accessibility in 2015”. In: *Nature* 553.7688 (2018), pp. 333–336.

[143] Tom G Farr et al. “The shuttle radar topography mission”. In: *Reviews of geophysics* 45.2 (2007).

[144] Viceministerio de Tierras - Unidad TÃl'cnica Nacional de la Informacion de la Tierra. *GeoBolivia- Mapa de cobertura y uso actual de la tierra*. visited 15/10/2021. 2010. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/62cefa0a-a3f8-4e67-af3f-946397d21943.

[145] U. Unidad de Analisis de Politicas Sociales y Economicas, I. Instituto Nacional de Estadistica, and P. Programa de las Naciones Unidas para el Desarrollo. *GeoBolivia. Indicadores Sociales (VAM-NBI), 2012.* visited 15/10/2021. 2012. URL: https://geo.gob.bo/geonetwork/srv/spa/catalog.search#/metadata/3f73d133-031c-4ed2-a7c3-e4569f03c5a6.