POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Dipartimento di Energia

Corso di Laurea Magistrale in Energy Engineering

ANALYSIS AND MODELLING OF TECHNICAL AND SOCIO-ECONOMIC ASPECTS AND UNCERTAINTIES IN OFF-GRID RURAL ENERGY PLANNING

Relatore: Prof. Emanuela COLOMBO

Co-relatori: Prof. Pierluigi LEONE
Ing. Fabio RIVA

Tesi di Laurea di:

Cecilia VICINANZA Matr. 849965

Anno Accademico 2016-2017

Acknowledgements

Vorrei ringraziare l'Ingegnere Fabio Riva, che mi ha supportata e sopportata in questo lungo progetto, la Professoressa Emanuela Colombo, che in questi cinque anni ha sempre saputo guidarmi con pazienza, e il Professor Pierluigi Leone per aver creduto nel mio lavoro. Tornando indietro nel tempo, però, è mio dovere ringraziare la Professoressa Rosanna Bruzzone, che mi ha insegnato la matematica e la fisica, e che un giorno, su un treno di ritorno da una gita in Liguria, mi ha aiutata a scoprire chi volessi diventare da grande. Vorrei ringraziare i miei compagni di viaggio: quelli di una vita (Alessandra, Alessandro, Francesca), quelli dello Schuster (Davide, Linda), quelli del Poli. In particolare devo un grazie a Chiara e Simone, che sono stati i miei migliori "Amici di Fatiche" nello sperduto quartiere di Bovisa. Devo poi ringraziare gli ASPer, alcuni di loro in particolare, con cui ho condiviso una delle esperienze più formative della mia vita. Un grazie va a Free2Change e a chi ne fa parte, per avermi permesso di mettermi in gioco. Devo poi ringraziare la Svezia e la Chalmers University of Technology per avermi fatta crescere e istruita da tutti i punti di vista e per aver fatto nascere amicizie che spero durino nel tempo. Devo infine un grazie a tutta la mia famiglia, nessuno escluso, ma soprattutto alla mia Mamma, che è la donna a cui più mi ispiro ogni giorno e in ogni mia scelta, a lei devo il carattere che mi permette di affrontare ogni sfida; al mio Papà, che da oggi posso chiamare collega e senza i cui preziosi consigli e insegnamenti non sarei arrivata dove sono oggi, a lui devo il gusto e la capacità di emozionarmi davanti al bello; al mio fratellino Giuseppe, che non smette mai di sorprendermi e da cui finisco sempre per imparare qualcosa, nonostante, forse, dovrebbe essere il contrario. Un grazie immenso va al mio ragazzo Gianmarco, che mi segue nelle mie peripezie, scolastiche e non, senza mai abbandonarmi quando sono alla ricerca di consigli, supporto e, perché no, qualche critica, o semplicemente un abbraccio. Con il suo carattere riesce sempre a smussare i miei spigoli. Tutte queste persone hanno in qualche misura contribuito alla mia crescita e alla mia formazione e se un giorno tutti coloro che ho ringraziato saranno orgogliosi di me e di chi diventerò, allora avrò compiuto la mia missione e potrò sentirmi fiera di me stessa e appagata nella vita di tutti i giorni.

The farthest Thunder that I heard
Was nearer than the Sky
And rumbles still, though torrid Noons
Have lain their missiles by-
The Lightning that preceded it
Struck no one but myself-
But I would not exchange the Bolt
For all the rest of Life-
Indebtedness to Oxygen
The Happy may repay,
But not the obligation
To Electricity-
It founds the Homes and decks the Days
And every clamor bright
Is but the gleam concomitant
Of that waylaying Light-
The Thought is quiet as Flake-
A Crash without a Sound,
How Life's reverberation
It's Explanation found-
(Emily Dickinson, Poem 1581)

Ai miei nonni,
che mi hanno resa chi sono oggi.

Extensive Summary

Abstract

Most of the additions of power capacity to allow electricity access in rural areas of DCs is forecasted to be provided by off-grid systems. The purpose of this work is to analyse and model technical and socio-economic aspects and uncertainties in off-grid rural energy planning, in order to provide a reliable and integrated approach to forecast long-term electricity demand. To accomplish this target, I relied on the Bass diffusion theory to assess the adoption of the connection to the microgrid across the social network of a fictitious rural village in Tanzania, whose characteristics were extrapolated from field data relative to a real village in the same area. I then developed a model based on Gompertz curves theory, to study the diffusion of appliances at the household level. Thanks to the obtained simulations, I could use LoadProGen platform to create daily load profiles of the same village and, eventually, I provided an example of a realistic off-grid system sizing procedure based on HOMER Pro software. The results of the study suggest that, depending on input parameters, different outputs can be found. Taking into account the diffusion of appliances, for example, as the input data vary, a variability of almost 80% can be found in year 5 of the simulations at the output (from less than 50 to more than 250 purchases for fans). Therefore, all of the uncertainties need to be taken into consideration. The configuration of the same system can change in time and the needed size of the microgrid might even double between year 10 and year 20 of the system. A software allowing to study the evolution of the off-grid system in time would be of help in the sizing procedure. Lastly, these previous two aspects should be considered at the same time due to the observed different patterns in the daily load profiles of the different years of various studied scenarios.

Introduction and Literature Review

Almost 1.1 billion people still live in a condition of energy poverty. One of the possible ways to mitigate this urgent problem is to make sure that the investments in this field are sustainable and that the planning and sizing of electricity production is properly performed. Rural areas of DCs tend as well to be affected by the lack of reliable information to allow researchers and potential investors to make reasonable estimates and forecasts. The reason why forecasting energy demand is so important is discussed by Hartvigsson [1], who developed a system dynamics model to show that the lack of power availability due to wrong forecasts may affect both the willingness of people to stay connected and the utility revenues.

The purpose of this study is to analyse which are the main drivers of the diffusion of electricity in a rural village and to model the main uncertainties that lead to the final composition of the daily load curve of a certain area, in order to obtain a more comprehensive and reliable sizing of an off-grid system.

I will organise the work by starting with a literature review, I will then present the method and tools I used and I will explain how I relied on actual field data to calibrate some parts of the model. I developed and calibrated an agent-based Bass diffusion model to simulate the grid connection spread across a fictitious rural village. I developed a second diffusion process for the electric appliances, based on Gompertz curves theory, as explained by Van Ruijven [2]. Then, I built daily load profiles through the software LoadProGen and gave them as an input to HOMER Pro, to obtain the sizing of a realistic off-grid system.

The first issue related to the agent-based model is to describe the social network across which the diffusion process will take place. In the papers by Piccardi [3] and by Riva et al. [4] three different types of network are suggested to better define rural villages conformation: the random network, based on the randomization of the choice of the next node to add to the network and of its first contacts, given a certain average degree; the Barabasi-Albert network, based on Barabasi's work [5] on scale free networks, where who has more contacts is more likely to further increase them; the social network, that is based on preferential attachment, where triangles tend to form, causing a high clustering structure.

It has now been a while researchers have started analysing development and energy poverty mitigation through the study of diffusion processes within social networks. The first realistic diffusion models where developed in the 1960s and one of the best examples of these efforts is Bass model, which was created by Frank Bass in 1969. The diffusion of any good was theorized to be dependent on advertising and word of mouth, that is external and internal influence respectively [6]. Bass model has a fundamental equation for the adoption rate at time t, which is the following:

$$f(t) = [p + qF(t)][1 - F(t)] \qquad \text{(a)}$$

where p is the probability of adoption due to external influence, q is the probability of adoption due to influence from other adopters and F is the proportion of population that has already adopted at time t.

In his book, John D. Sterman [7] interprets the parameter q in the form

$$q = ci \qquad \text{(b)}$$

where c is the contact rate and i is the adoption rate of someone influenced by others.

Van Ruijven's [2] study is based on the use of Gompertz curves theory, which is a different formulation for describing diffusion processes. He cites an article by Kemmler [8], which states that household expenditures are the main correlating factor for electricity use by a household, to be considered by dividing the population in quintiles. He also introduces the concept of ownership of the appliances and builds a relation with the household expenditure.

Rao and Ummel underline the importance of considering affordability rather than income in diffusion processes, which represents the share of expenditure to be maximally devoted to the adoption of an appliance [9].

In order to have more specific information about the load curves that a microgrid will have to manage, forecasts can be made in accordance with the purpose: load curves for intuitive system sizing are based on the estimation of the likely peak load of the system, but this might cause over- or under-estimation issues; load curves for numerical sizing are based on the use of more structured approaches in order to derive detailed load profiles.

Mandelli [10] developed a procedure, called LoadProGen and characterised by the following features: it is based on input data coming from practical experience or local surveys; it is based on a rigorous mathematical formulation; it is bottom-up.

Once the load curves are available, Rojas-Zerpa [11], in his work about energy planning, explains which are the main aspects to take into consideration when deciding which is the power system optimization tool that should be used. First, the application area: considering rural areas, models for distributed generation are the most interesting. Second, the planning horizon: it is useful, in some circumstances, to take into consideration long-term (11-20 years) planning tools. Third, the objective of decision making: economic, social, or environmental. Fourth, the technologies to be included in the planning. A software called HOMER (Hybrid Optimisation Model for Electric Renewables) is taken into consideration, developed by NREL (National Renewable Energy Laboratory, USA) [12], which can handle a large set of technologies and can perform an optimization to decide which is the cheapest configuration in terms of Net Present Cost for decentralised systems.

Materials and Methods

In Figure a, a flow diagram of what will be explained in this chapter is shown.

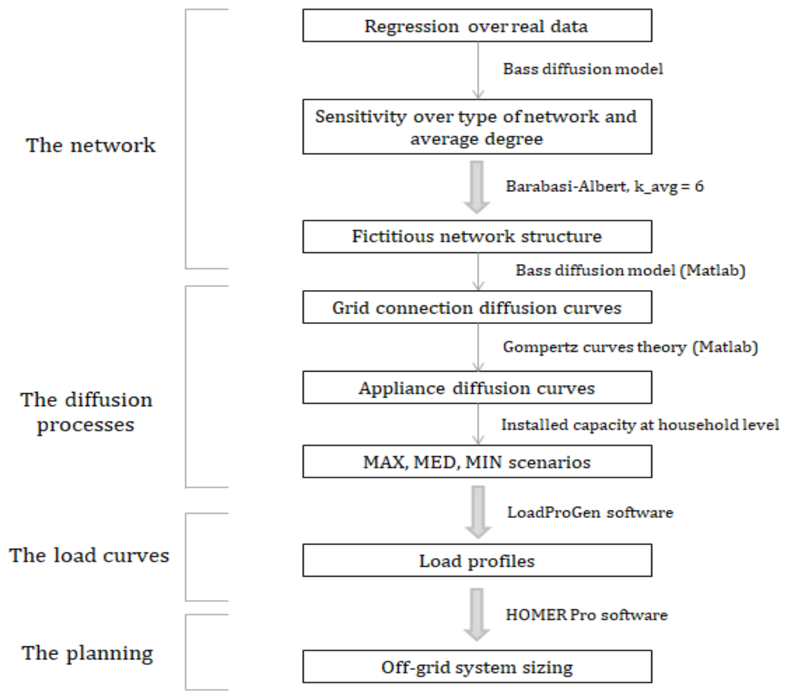


Figure a: flow diagram of the study procedure

I started the research by collecting data that were available in the region of Njombe, in Tanzania, more precisely in the village of Bulongwa. This data, which were provided by Chalmers University of Technology, were collected from the control unit of an existing microgrid, which would allow knowing at each time-step (years from 2009 to 2016) how many people were connected to the grid for the first time, while the first connection dated back to 2001. Bulongwa is a village of approximately 700 households, where the South-Central Diocese of the Evangelican Lutheran Church created a mini-hydropower facility (180 kW), which would feed a microgrid in the village. To obtain the curve of diffusion of the connection to the system, I performed a regression to produce a diagram, shown in Figure b, for the entire lifetime of the system, the red line representing the division between regression (based on the same growth rate of the available data of 2009-2010 years) and real data.

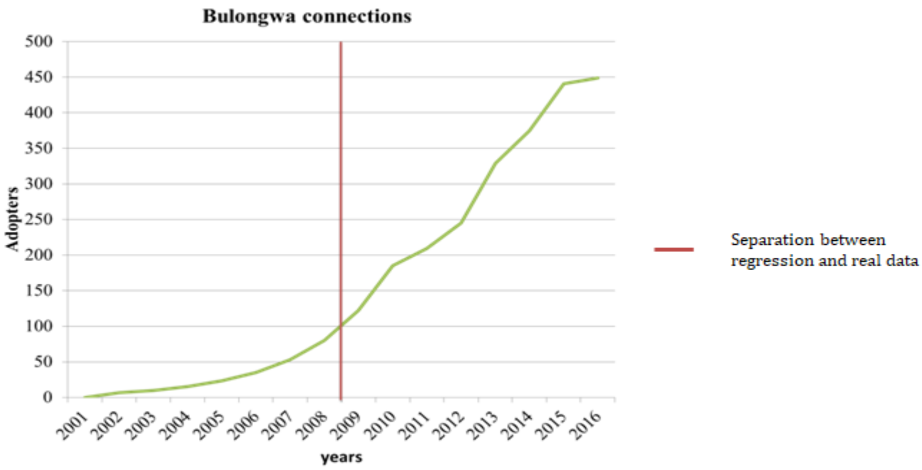


Figure b: Bulongwa grid connection diffusion from 2001 to 2016

Taking into account as reference a Bass diffusion process and using the regression equations (c-h), I could find the most likely p and q values of equation (a) for this village. These could be used as input in a Bass diffusion model of the connection to the system, which I developed on Matlab with the purpose of finding out which was the most appropriate type of social network I should use for this type of context.

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k + \varepsilon \qquad \text{(c)}$$

$$b_1 = \frac{(\sum x_2^2)(\sum x_1y)-(\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2)-(\sum x_1x_2)^2} \qquad \text{(d)}$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2y)-(\sum x_1x_2)(\sum x_1y)}{(\sum x_1^2)(\sum x_2^2)-(\sum x_1x_2)^2} \qquad \text{(e)}$$

where

$$\sum x_1y = \sum X_1Y - \frac{(\sum X_1)(\sum Y)}{N} \qquad \text{(f)}$$

$$\sum x_2y = \sum X_2Y - \frac{(\sum X_2)(\sum Y)}{N} \qquad \text{(g)}$$

$$\sum x_1x_2 = \sum X_1X_2 - \frac{(\sum X_1)(\sum X_2)}{N} \qquad \text{(h)}$$

It was possible to obtain b_1 and b_2 values that in the current case correspond to p and q values.

Thanks to the regression and a sensitivity analysis (shown in Figure c for the chosen network), using as input to Bass model the different types of network described above, a Barabasi-Albert was identified for Bulongwa, with an average degree (*k_avg*) equal to *6* , presenting the smallest standard deviation from the real process.

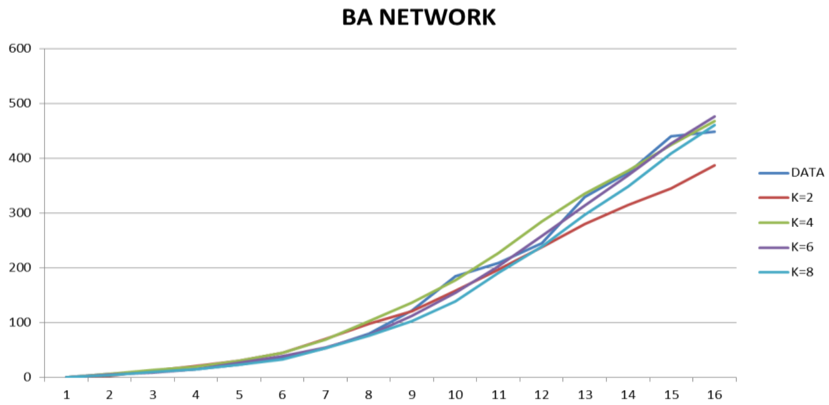


Figure c: Barabasi-Albert network sensitivity analysis and standard deviation evaluation

For the rest of the study, a fictitious village will be considered, for which the type of network and the average degree will be kept constant. Due to their endogenous characteristics, which make them really area-specific, the variety of p and q will be brought forward in the next phases of the analysis through Monte Carlo method, which would consider a uniform distribution sampling of possible values for the two parameters. I chose to simulate and study a fictitious village composed of 400 households, assumed to be located in the same geographic area of Bulongwa, but having no access to the grid at the time in which the study begins.

I decided to model the microgrid connections along the planning horizon on Matlab through a standard Bass model. The values of p and q should vary in this way:

$$p = 0.002 + (0.01 - 0.002) * rand \qquad \text{(i)}$$

and

$$q = 0.2 + (0.7 - 0.2) * rand \qquad \text{(j)}$$

Their ranges include the calibrated values found for Bulongwa. A hundred simulations were performed, each one being saved in 21 matrices, containing the diffusion process year by year.

To understand how the electrical appliances would spread across the village, I decided to develop a second part of the model, based on Gompertz curves theory, which was well explained by van Ruijven [2]. It is based on the concept of ownership, which is regulated by the following equation:

$$Ownership_{q,A,U(t)} = \alpha_A * EXP(-\beta_{A,U} * EXP(\tfrac{-\gamma_{A,U}}{1000} * PCOpc_{q,U(t)})) \qquad \text{(k)}$$

Where PCO is the expenditure per capita, β and γ are exogenous coefficients based on linear regressions from real data collection, differentiated by appliance, while α is the upper limit of appliance ownership.
Once the level of ownership is available, the expenditure available of each household for a certain appliance is first necessary. I then made the assumption that the richest purchase first. Lastly, it is in the same work that the value of ownership parameters were provided, but I decided to let them vary through Monte Carlo method starting from the Indian values of van Ruijven (+/- 20% for Africa, with sampling from uniform distribution of probability) to avoid being too area-specific.
The World Data Bank provides the yearly behaviour of per capita expenditure in Tanzania. Since the entire study by van Ruijven [2] is based on quintiles, it was decided to keep the same format. Five behaviours of the average expenditures of the 5 quintiles of the village were obtained and it was possible to evaluate the yearly level of ownership, in each quintile, of each technology, depending on a correspondent level of average expenditure of the population.
Plus, I obtained as many sets of curves of adoption of the appliances as the number of simulations (100 in the current case).
In order to build the load curves, it was possible to use a software called LoadProGen: a platform, based on Matlab, which gives as output daily load profiles, which can be given in hours, quarters of hour, minutes or seconds.
The total installed capacity (viz. the total number of forecasted electrical appliances owned by households) of the 100 simulations at year 20 was evaluated. Then, among all the 100 simulations, I selected the scenarios with the greatest, the lowest and the median number of installed appliances (viz. the 3 situations in which the ownership of the appliances among the population is the lowest, the median and the highest), namely MAX, MIN and MED scenarios. In order to make a long term forecast of the daily load profiles, it was decided to build the load curves for year 1, year 10, that is half of the lifetime, and year 20 of the planning, that is the last year of the system, for each one of the three scenarios.
Once daily load curves were obtained it was possible to start the actual off-grid system sizing, which was performed using HOMER Pro software. HOMER Pro attempts to simulate a viable system for all possible combinations of the equipment

that the user wishes to consider, while making sure a certain load profile is responded. For each proposed solution a set of techno-economic parameters is provided, which allow the user to assess the economic feasibility of the different options. Given the output, it will be the user who will make his own evaluations based on his needs and requirements.

Results and Discussion

The whole research was started from the creation of the network structure of a fictitious community, that is shown in the figure below and was obtained using a Matlab script based on Barabasi-Albert network formula for the probability for a node to have a certain degree.

$p(k) \approx k^{-q}$ (1)

where p is the probability and k the considered degree, while q is an experimentally determined parameter equal to 3 for BA networks.

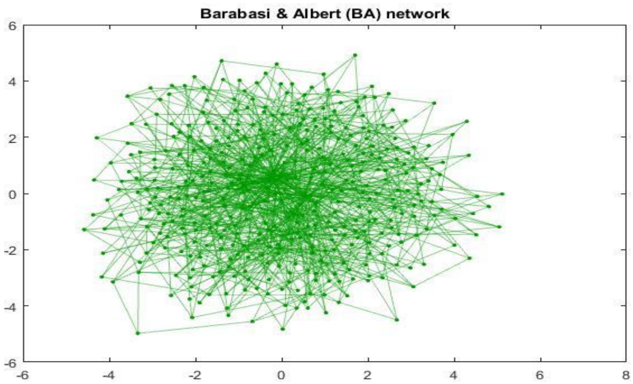


Figure d: BA network structure for fictitious village of 400 households, average degree of 6

This network of Figure d is characterized by an average degree of 6. The average eigenvector centrality, equal to 0.0025, is a measure of the influence of a node in a network and is pretty low, considering its maximum could be 1. Closeness centrality of 0.000808 suggests that the network might have some nodes who are quite far from the rest, because the overall value keeps very low. Betweenness centrality of 425.625, suggests that to go from a node to the other a long distance should be covered, making it more likely to pass through many nodes several times.

As a second step, the diffusion curves of appliances for 100 simulations were built, which are shown below for one of the chosen scenarios. Looking at the diagram of

Figure e for MAX scenario, it is possible to notice that it is only four appliances out of six that in 20 years actually get to be adopted.

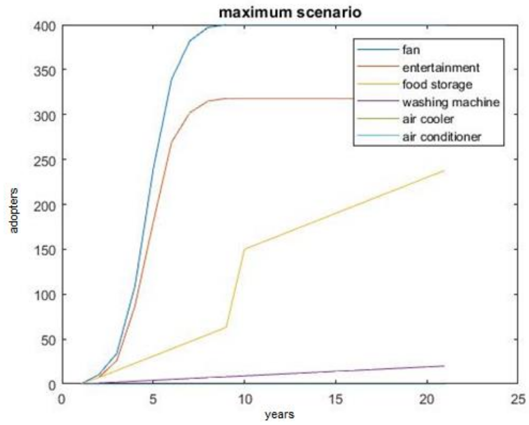


Figure e: appliance diffusion curves for MAX scenario

Some appliances will never be adopted, since their price is larger than the 80% of total expenditure (due to affordability constraints) of each and every household. The diffusion of lightbulbs is not present, because as soon as one adopts electricity it is assumed that he will also install a lightbulb.

Since MAX, MED and MIN scenarios also reflect the Bass diffusion model output, it can be useful to analyse the values of the parameters involved in the three cases, selected each time thanks to Monte Carlo method, from a uniform distribution of values.

Scenario	p	q
Maximum	0,0091	0,2425
Minimum	0,0035	0,2389
Median	0,0070	0,2013

Table a: parameters values from Monte Carlo method

Given this table, it is possible to notice that *p* values follow the order of maximum, median and minimum. This, indeed, is reflected in the diffusion processes.

In each of the scenarios, the year in which food storage adoption shoots up corresponds to the reaching of market saturation for fan and entertainment. The curves of washing machines are not S-shaped yet, because 20 years result not being enough for this technology to spread around.

I then used the software LoadProGen to obtain 250 load profiles for each considered case. The variability of MIN scenario at year 10 is larger (8% vs 5%), therefore this curve is not as representative of what really happens as the curve for MAX case. Moreover, in every scenario, year 1 is not relevant to the sizing of the grid, since it only presents very small numbers of adoption.

Year 20 curve of MED scenario is more similar to year 20 curve of MAX scenario, which suggests that the 100 simulations were closer to MAX rather than to MIN case. An example of the results for MAX scenario in shown in Figure f.

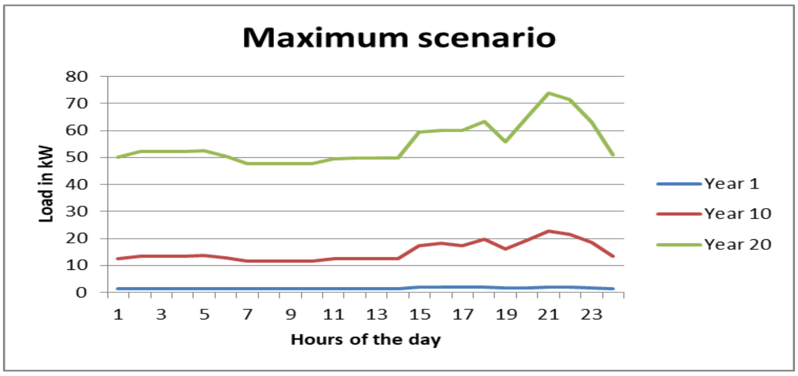


Figure f: average load curves for year 1, 10 and 20 of MAX scenario

It can be interesting to verify whether the shape of these load curves is realistic.

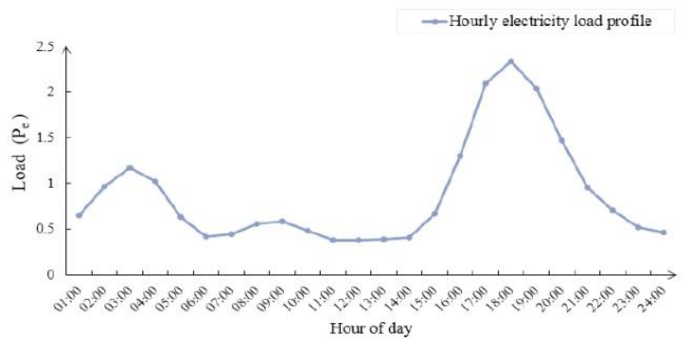


Figure g: potential rural electricity load profile reference [10]

Looking at this diagram, it can be said that the load profiles obtained for the current study have a realistic shape. It is not possible to make a comparison in terms of consumption due to the smaller considered capacity of the studied system.

The off-grid system sizing was performed thanks to the use of the software HOMER Pro and the scheme of the microgrid, that was designed taking inspiration from what literature [13] suggested, is the following.

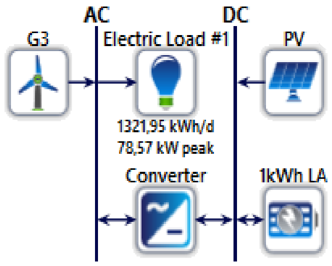


Figure h: scheme of microgrid for sizing

Data relative to solar irradiance and wind speed were taken from NASA "Surface meteorology and Solar Energy" [14] and the prices and lifetimes of technologies were also found in the literature [15]–[17]. In Table b, an example of the results of HOMER optimization processes can be seen, specifically years 1, 10 and 20 of MAX scenario.

MAX SCENARIO	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
YEAR 1	11,3	0	48	2,70	59826
YEAR 10	105	0	454	24,4	561562
YEAR 20	393	0	1622	87,7	2,05M

Table b: HOMER Pro output for MAX scenario, year 1, 10 and 20, optimized case

It can be noticed how between year 10 and year 20 of the system, the necessary installed capacity of PV increases of almost 75% and an increase can be consequently found in the net present cost values. It is also important to notice that year 1 represents not even 5% of the final configuration and it would not be suggestable to take this into account for the sizing process. Results suggest that it is always necessary to take into account the whole lifetime of the system to have a clear idea of what to expect from the future, trying to avoid cost recovery failure. In all of the situations, it can be noticed that the optimal solution always consists of an all solar solution. This optimizing algorithm only takes into account the economical optimization of costs, while neglecting, for example, that having wind in addition to solar might give greater reliability to the grid, allowing for a differentiation in the generation resources.

The drawback of this method is that it will not give a univocal outcome, but the decision will have to be a result of an analysis made by the user. Unless one wants to make adjustments very soon after starting the facility, it is suggestable to take into account a long enough period of time to project the microgrid to avoid over- and under-estimation issues.

Conclusion

The aim of this work was to analyse how the main uncertainties related to developing countries realities influence the diffusion of electrical appliances and therefore the configuration of off-grid systems. To respond this purpose, the author started the research from real data analysis and used it as the basis to develop a model in Matlab, constituted of two parts. A first one based on Bass diffusion model and a second one based on Gompertz curves theory. 100 simulations were performed, which allowed to demonstrate that the uncertain endogenous factors actually have an impact on the output of the model, mainly contributing to the speed at which the diffusion process happens. For example, as the input data vary, a variability of almost 80% can be found in year 5 of the simulations at the output (from less than 50 to more than 250 purchases for fans). Three scenarios were analysed more in depth. A long-term analysis was necessary to be able to properly design the microgrid, otherwise over- and under-estimation issues might have taken place. Between year 10 and year 20, indeed, there can even be a doubling of necessary generation capacity. The software used throughout the sizing procedure were LoadProGen and HOMER Pro. The first allows for the creation of many different load profiles at the same time, but does not take into account the evolution of the households in time. On the other hand, HOMER Pro only takes into account one load profile at a time and keeps it constant for the entire lifetime of the system it is sizing.

The patterns which can be found comparing year 20 of the various scenarios do not always repeat in the previous years, which means that, by considering only a smaller amount of time (e.g., stopping the analysis at year 10) one would probably get the long-term estimates wrong and might incur bad cost recovery failures.

It would be useful in the future to find or create a software which would allow to consider continuous changes in the load demand and in the household configuration, so to be able to size the system in one only step, by considering the 20 years evolution all at once.

Table of Contents

Abstract

Most of the additions of power capacity to allow electricity access in rural areas of DCs is forecasted to be provided by off-grid systems. The majority of rural villages is characterized by the lack of reliable data and information, which might cause the inappropriate sizing of energy solutions, leading to supply shortages or cost recovery failure. The purpose of this work is to analyse and model technical and socio-economic aspects and uncertainties in off-grid rural energy planning, in order to provide a reliable and integrated approach to forecast long-term electricity demand. To accomplish this target, (i) I implemented a Bass diffusion process to assess the adoption of grid connection across the social network of a fictitious rural village in Tanzania, whose characteristics were extrapolated from field data relative to a real village in the same area. (ii) I then developed a model based on Gompertz curves theory, to study the diffusion of appliances at the household level. Thanks to the obtained data, (iii) I could use LoadProGen software to create daily load profiles of the same village and, eventually, (iv) I could provide an example of a realistic off-grid system sizing procedure based on HOMER Pro software. The study will go through all four phases. The results of the study suggest that all of the uncertainties need to be taken into consideration to avoid over- or under-sizing issues. Indeed, changing the uncertain endogenous values through a Monte Carlo algorithm, I obtained very different speeds of technology diffusion. For example, year 5 of the simulations presents 80% variations in purchases of fans depending on input data. A software which allows to study the evolution of the off-grid system in time would be of help, since HOMER Pro only takes one load curve at a time as an input. The importance of taking into account the entire lifetime of the system is further shown by the differences in the patterns observed in the daily load profiles of the different years of various studied scenarios. In a single scenario, the necessary generation capacity can double from year 10 to year 20 and the same happens to the net present costs of the system. While, comparing different scenarios, the one presenting maximum installed capacity at year 10 does not hold the same property at year 20.

keywords: access to electricity, social network, diffusion process, grid sizing, appliance diffusion model

Sommario

La maggior parte degli interventi per aumentare la capacità produttiva di energia elettrica delle aree rurali nei paesi in via di sviluppo, si prevede si concentri su sistemi "off-grid", isolati e autonomi. La maggioranza dei villaggi rurali è caratterizzata dalla mancanza di dati e informazioni affidabili. Questo potrebbe portare a una progettazione inadeguata delle soluzioni energetiche, che, a sua volta, provocherebbe il pericolo di blackout e difficoltà nel recupero dei costi. Lo scopo di questo lavoro è analizzare e modellizzare gli aspetti tecnici e socio-economici e le incertezze che si riscontrano durante la progettazione dei sistemi off-grid nelle zone rurali, con l'obiettivo di fornire un metodo affidabile ed integrato per prevedere la domanda elettrica nel lungo termine. Come prima cosa (i) ho implementato un processo di Bass per analizzare la diffusione dell'allacciamento alla rete elettrica all'interno della rete sociale di un villaggio rurale fittizio in Tanzania, le cui caratteristiche provengono dall'estrapolazione di alcuni parametri dai dati reali di un villaggio della stessa area geografica. (ii) Ho poi sviluppato un modello, basato sulla teoria delle curve di Gompertz, per studiare la diffusione delle apparecchiature elettriche nelle case del villaggio. Utilizzando i dati così ottenuti, (iii) ho creato, grazie alla piattaforma LoadProGen, dei profili di carico giornalieri per il villaggio stesso e, infine, (iv) ho fornito un esempio realistico di progettazione di un sistema off-grid attraverso HOMER Pro. Lo studio toccherà i quattro aspetti che ho descritto. I risultati suggeriscono che, se si vogliono evitare problemi di sovrastima o sottostima della domanda, tutte le incertezze devono essere prese in considerazione. Infatti, cambiando il valore dei parametri endogeni in input con un algoritmo Monte Carlo, si ottengono velocità di diffusione delle tecnologie molto diverse. Ad esempio, all'anno 5 delle simulazioni si ottiene una variabilità dell'80% nell'adozione dei ventilatori, dipendente dai parametri in input. Inoltre, sarebbe più efficiente avere un software che, a differenza di HOMER Pro, il quale prende una curva di carico alla volta come input, permetta lo studio dell'evoluzione nel tempo della domanda. Infine, l'importanza di prendere in considerazione l'intera vita utile del sistema è ulteriormente sottolineata dalle differenze che si possono osservare confrontando le

curve di carico dei diversi anni dei vari scenari studiati. Se si considera l'evoluzione nel tempo di un singolo scenario, la capacità di generazione necessaria può arrivare a raddoppiare tra l'anno 10 e l'anno 20 e lo stesso può succedere ai costi di investimento. Invece, se si confrontano i diversi scenari nel tempo, lo scenario che all'anno 10 presenta la massima capacità installata in termini di domanda, non mantiene questa condizione se si considerano gli anni successivi.

parole chiave: accesso all'elettricità, reti sociali, processi di diffusione, pianificazione delle reti, modelli di diffusione delle apparecchiature elettriche

1 Introduction

During September 2015 UN Summit, the 17 Sustainable Development Goals were adopted by the world leaders and on the 1st of January, 2016 these actually came into force within the 2030 Agenda for Sustainable Development. One of the goals, number seven, seeks to provide "affordable and clean energy" for all. These days, almost 1.1 billion people still live in a condition of energy poverty. One of the possible ways to mitigate this urgent problem is to make sure that the investments in this field are sustainable and that the planning and dimensioning of electricity production and distribution are properly performed. An issue that comes to surface at this stage is that those places in which people suffer from energy poverty tend as well to be affected by the lack of reliable information and datasets to allow researchers and potential investors to make reasonable estimates and forecasts. The reason why forecasting energy demand is so important is discussed by several authors in their studies. Hartvigsson [1] developed a system dynamics model to show how the power supply capacity should be accurately considered based on the forecasts of electricity demand. Indeed, the lack of power availability may affect both the willingness of people to stay connected and the utility revenues. Brivio et al. [18] demonstrate that the optimal size of the components of an off-grid system, especially the capacity of the battery energy storage system of photovoltaic off-grid systems, are sensitive to the evolution pattern of load. Van Ruijven et al. [19], while developing a bottom-up model to assess trends in electrification over the next decades in DCs, demonstrate how the demand level is a significant factor when assessing the potential of mini-grid technologies. Kivaisi [20] and Cabral et al. [21], [22] highlight the need to take into account the evolution of the electricity load when planning the system, since the marginal costs of energy services vary among supply alternatives (i.e. small photovoltaic (PV) systems when the load is low, grid-extension when it is high). Fuso Nerini [23] demonstrates how the cost of the energy system for reaching different levels of energy demand to satisfy in the village of Suro Craic in the years 2010-2030 may vary from few hundreds to 8000 2010US$.

The purpose of this work is to identify and model the main drivers and complexities related to the diffusion of electricity use in a rural village and to model the main uncertainties that lead to the final composition of the daily load curve of a certain area, in order to obtain a more comprehensive and reliable sizing of an off-grid system.
The approach that was chosen to perform this research is multidisciplinary, and goes from engineering to sociology and economy. The sizing of an energy facility depends on several fundamental parameters, such as the load curves of demand, which strongly depend on socio-economic aspects, e.g. the willingness of people to connect to the grid and to adopt a certain technology. Such socio-economic aspects are strongly related to the social environment that surrounds the individuals, who cannot be considered homogeneous actors of a standard world, but are characterized by a whole bunch of complexities and subjective features that should be indeed endogenously represented in energy models [24]. My objective is to fill the gap in literature and make a connection between social networks theory, appliance diffusion process theory and off-grid systems sizing models, with the aim of reaching a more reliable design process of microgrids.
In order to catch the whole variety of characteristics, I organised the work by starting with a literature review, which covers all the topics that were examined to understand the features of the studied environment and the necessary knowledge to work in it. This will be dealt with in chapter 2. Then, in chapter 3, I will present the method and the models and I will explain how I relied on actual field data from Tanzania for calibrating the social network structure to be used across the entire work. Moreover, I will go through the steps I followed to develop a model based on Bass diffusion process, to simulate how the microgrid connection spreads across a fictitious village. To be able to obtain reliable load profiles, a second diffusion process for the electrical appliances was developed, based on the Gompertz curves theory, as has been explained by Van Ruijven [2]. Eventually, it was possible to build daily load profiles through LoadProGen and give them as an input to HOMER Pro, to obtain a possible sizing of a realistic microgrid. In chapter 4, I will present and discuss the main results of the research, and in chapter 5 I will provide several final remarks plus some suggestions for future developments.

2 Literature Review

For this work, I carried out a review of 75 papers and scientific articles. I downloaded the studies on Scopus platform and I found them using keywords like "energy in developing countries", "electricity diffusion", "diffusion models", "behaviour of consumption", "load curves", "grid sizing" and other similar ones, which would allow me to find material about all of the topics I was going to go through.

When someone is willing to plan and size the creation of an off-grid system, there are several aspects he should take into consideration. To properly size a microgrid, a forecast of the future load curves is necessary, to have an idea of which will be the peaks the system should be able to manage. The amount of energy a village uses, depends of course on how many appliances the households will use, but before that it depends on how many people actually have access to the grid. When taking into account a place with no access to electricity at year 0, it can be tricky to understand how the use of energy will diffuse across the population. Many uncertainties will come up along the process and several socio-economic and endogenous factors will contribute to the definition of the future load curves. A way to better design diffusion processes was proposed by several authors which I found in the literature. It consists of taking into account the innovation diffusion processes and the investigation of social network structures within a rural village, in terms of links among the households, which could allow the flow of information and the diffusion of knowledge about certain issues. Indeed, Peres et al. give an interesting definition of innovation diffusion as 'the process of the market penetration of new products and services that is driven by social influences, which include all interdependencies among consumers that affect various market players with or without their explicit knowledge' [25]. The demand of energy, therefore, will be strongly influenced by the network dynamics of a certain place. If the members of a given group adopt, the members of another group who is willing to differentiate by the other will tend to avoid adoption [26].

2.1 The network

The overall idea I could extrapolate from literature was that if one is trying to study diffusion across networks, it is necessary to know how to describe the network itself. In case of data availability, networks can be built by deriving the information (e.g., relationships among people) on the field and a matrix (called *network graph*) representing the real contacts across the people of the considered village would be obtained. The problem is that, often, data related to rural villages are very few. It will then be useful to have an idea of what the network looks like, or better, how it is possible to produce a realistic network, just knowing how many nodes (households) compose it.

Some of the most important ideas relative to networks theory, come from the work of Matthew O. Jackson, a major expert in the field of social network studies, who wrote a book called "Social and Economic Networks" [27]. Here, several structures are explained and defined, based on how people can come into contact, with which criteria they get to know each other and which formations get to life among the nodes (the individuals) of the network (the community). Jackson says that along life people influence and get influenced by others and these mechanisms depend on who they meet and how they relate to the met ones. People have several types of contacts, that can differ depending on the social status of people, or the possibility of being relatives, or friends, or colleagues. The author also presents the main characteristic parameters of a network, with their definitions and explanations. The concepts and the different types of centrality are addressed thoroughly, explaining the difference among degree, closeness and eigenvector centrality.

When defining networks, many researchers have proposed their own type of networks, which follow very specific formation rules.
Many studies take into account the network structures to understand the social dynamics of rural areas, some starting from econometrics, like [28], others focusing on rural areas in general [29] or on more specific areas, like India [30] or Malawi [31].

In their study, in which the knowledge of the level of poverty of others in 600 communities of Indonesia is assessed, Alatas et al. [32] say that the contacts among people depend on their relative economic well-being, but add that, as was checked through independent data, it often happens that individuals get the richness ranking wrong and get biased by their own wrong beliefs. Overall, the majority of studies confirms that better connected people (in terms of number of connections) are also better at ranking others and usually place themselves in socially strong positions, as richer, as more influential and as more educated.

Di Falco and Bulte [33], instead, while studying the types of farm management that might help mitigating the effect of weather shocks on the soil in rural areas, propose a type of network that is based on the distinction made by each individual between peers and non-peers among their contacts. Each person would be linked to his/her peer contacts, and given all links the network gets formed. This type of network is called kinship network and the main problem with it is that it can be built only through very specific surveys, containing questions that can be tricky, since it is not always clear to someone who is a peer and who is not in a context like the developing country village one. The same type of theory was already supported by Van Der Broeck and Dercon [34]. They had the idea that, for rural regions' communities like Kagera in Tanzania, kinship networks should be considered the most appropriate form of network. Still, this type of networks is argued to be very sensitive to subjective impressions and it results very difficult to verify the information collected through the surveys. Other types of network, whose formation is more objective to assess, might be preferable for this type of work.

When assessing any of these types of network, as very well explained by Jackson [27], there is always a question to answer, that is whether the links formed by the nodes are bilateral or unilateral, meaning that a person A can get influenced by/be linked to a person B, but not necessarily the other way around. Knowing if a message can flow only from A to B, or only from B to A, or both ways, can be important to understand how a technology will spread around.

An interesting work that I studied is by Van Den Bulte and Joshi [35]. While studying the diffusion of innovations, they define two different categories of consumers, starting from the difference between influentials and imitators. The first ones are more in touch with new developments than others, and have a disproportionate influence on others' adoptions [36]. The latter ones, instead, tend to prefer low-risk

innovations and are in need of guidance of people that usually have similar or higher social status. Influentials tend to be independent or slightly influenced by other influentials; imitators, on the other hand, can get influenced both by influentials and by other imitators.
In case of presence of this distinction, the network might be expected to present a conformation with some nodes that have a higher degree (number of contacts) than others, who are likely to be influentials, while more isolated nodes will more easily represent imitators.
Influential people are believed to be convincing, informed and widely socially tied [37]. For this definition Goldenberg would take inspiration by Weimann [38], who suggests that influence is a combination of "who one is", "what one knows" and "who one knows". Goldenberg adds being innovative as another typical trait for influentials. Still, it is not enough to be either innovative or socially affirmed, but who is both, in case of the diffusion of new products or innovations, like electric appliances, is more likely to become one of the first to purchase (adopt) such technologies; imitators, on the other hand, will wait for the feedbacks of previous adopters before adopting something themselves and will go forming the large group of people who will actually push the innovation towards the reaching of market saturation.

While the studies presented so far give a theoretical point of view, several authors try to give numerical values to the parameters related to network structure, which can give an idea of the orders of magnitude to be considered in rural contexts. Bandiera and Rasul, in their paper "Social networks and technology adoption in Northern Mozambique", estimate that the average degree of the social network is 4.9 contacts per person [39]. Beaman et al., instead, in the work "Can Network Theory-based Targeting Increase Technology Adoption?" say that the probability of having multiple contacts who purchased a technology increases more rapidly as the technology spreads through the network. They call "degree" the number of contacts the nodes have who have purchased something and not the number of contacts in general, that is therefore going to be a variable number [40]. Finally, it is again Bandiera and Rasul who provide Table 1, which allow us to say that a good range for k_avg (average degree) should be from 2 to 8, since their values are estimated for Mozambique that is a reality similar enough to Tanzanian one [39].

Social Networks by Adoption Status and Network Type

Mean network size (standard deviations in parentheses, 25th, 50th and 75th percentiles in brackets)	Total	Adopters	Non adopters
Number of adopters among family and friends	4.92 (5.18) [0, 4, 7]	5.87 (4.92) [3, 5, 8]	3.91 (5.28) [0, 3, 5]
Number of adopters among family	2.46 (3.36) [0, 1, 4]	2.69 (3.01) [0, 2, 4]	2.22 (3.70) [0, 0, 3]
Number of adopters among friends	2.46 (2.86) [0, 2, 4]	3.19 (3.02) [1, 3, 4]	1.69 (2.46) [0, 0, 3]
Have no adopters among family and friends	0.278 (0.449)	0.167 (0.374)	0.396 (0.491)

Table 1: Network average degree reference

It is from Peres et al. [25], instead, that it is possible to learn more about the concept of clusters, which leads to the definition of the clustering coefficient. This parameter represents a measure of the connectedness of a network. Clusters are linked to the concept of homophily, that is the tendency of similar individuals to group among each other. A network which presents high clustering is one in which if a node A is linked to n nodes, a large part of these is also connected among each other, which for example is true in reality when you have a group of friends that all know each other, forming therefore high clustering levels. One of the main characteristics of these closed social structures is that they help strengthening the role of weak ties. Once given as input into a cluster, the information will not likely leave it. The only way it will have to get out of it is through a weak tie with the outside.

The clustering coefficient can be evaluated as:

$$CC(v) = \frac{2 * N_v}{K_v * (K_v - 1)} \quad (1)$$

where v is the name of the considered node, K is its degree and N is the number of links that are present among its contacts.

Starting from the concept of clusters, it is Christine Kiss and Martin Bichler [41] who try to go further into detail in the distinctions between the two consumer categories, explaining that both imitators and influentials can have not only a positive impact on the willingness to adopt of others, but also a negative one, in case of negative feedbacks. They say that "dissatisfaction produces more negative word-of-mouth than satisfaction produces positive word-of-mouth" and propose a network structure called "scale-free", based on the presence of centrally located and extensively high degree "hubs" that of course will represent the influentials. They also, for the first time, mention the importance of coupling network theory with other models that provide, as they say, "orthogonal information" related with diffusion mechanisms and endogenous aspects of the different nodes, in order to be then able to study in detail the diffusion of innovation in specific communities.

2.2 Diffusion Process

It has now been a while researchers have started analysing development and energy poverty mitigation through the study of diffusion processes within populations. Two types of diffusion need to be studied for the purpose of this work: the diffusion of the connection to the electric grid and the diffusion of the adoption of electric appliances, which are both necessary information in order to be able to size an off-grid system properly. An analysis of the diffusion models that will then be used for the connection to the microgrid is first performed. Then, some useful aspects for appliances diffusion will be added.

The first realistic diffusion models were developed in the 1960s and one of the best examples of these efforts is Bass model, which was created by Frank Bass in 1969. For the first time the diffusion of any good was theorized to be dependent on two main aspects: advertising and word of mouth, that is external and internal influence respectively [6]. Bass model has a fundamental equation for the adoption rate at time t, which is the following:

$$f(t) = [p + qF(t)][1 - F(t)] \qquad (2)$$

where p is the probability of adoption due to external influence, q is the probability of adoption due to influence from other adopters and F is the proportion of population that has already adopted at time t.

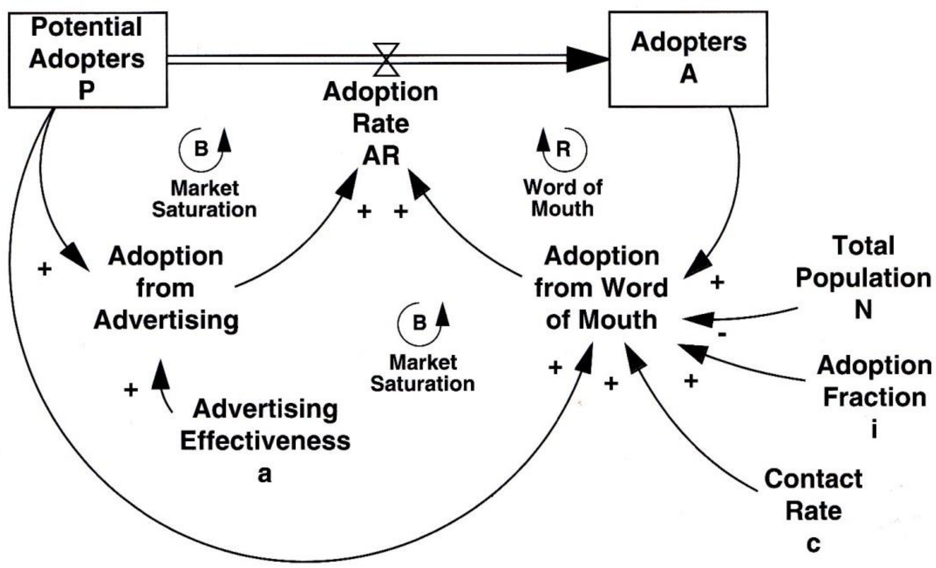


Figure 1: Bass model flow diagram

This equation, that has the strength of being very simple and intuitive, can be enriched in different ways with the target of taking into account other possible aspects that might influence the decision to adopt an innovation.

Bass Model was found to have one major drawback, it indeed assumes homogeneous individuals who behave the same and get influenced with the same probability. Influence could originate by advertising or by other people who adopted beforehand. To respond this issue, many studies have been carried out about how to merge the simple and intuitive structure of Bass Model with the more specific and individualistic approach of Agent Based Models. What an ABM does, is taking into consideration the individual characteristics that differentiate one agent from the other and which might have an influence on the outcome of the diffusion process. The importance of the different aspects can be weighted accordingly to the considered individual or context. On the other hand, implementing an ABM is definitely more demanding in terms of data availability and might bring to a larger level of uncertainty in the outputs. Many examples of attempts to take into account individual aspects can be found in literature (e.g., [42], [43], [44], [45], [46], [47], [48], [49]) and the most relevant ones will be discussed below.

One of the first to propose this type of approach is a study in which Bass himself participated. The aim of his research work was to try to include in the model a state of the agents that were taken into account, who could now be considered to be either influentials or imitators, depending on whether they were supposed to be more influenced by the external inputs or by word of mouth [50].

In the papers by Piccardi [3] and by Riva et al. [4], Bass model is compared to agent-based models, while trying to understand which might be the influence of social networks in the diffusion processes. Three different types of network are suggested to better define rural villages conformation. The three of them are analysed and compared to make a speculative analysis and assess whether or not the network conformation is of influence in the diffusion process output. The proposed network typologies are: the random network, based on the randomization of the choice of the next node to add to the network and of its first contacts, given a certain average degree; the Barabasi-Albert network, based on Barabasi's work [5] on scale free networks, which involve the idea that who has more contacts, that is a higher degree, is more likely to further increase them, acquiring even more links with respect to who had already less at each time step. Last, the social network is proposed, that is based on preferential attachment, where triangles tend to form and if node A and node B are "friends", an added node C, that is already linked to B, will be more likely to get linked to node A, rather than to a fourth node D that is not linked to anyone yet, causing a high clustering structure. The formulation and obtainment of the three network typologies will be further explained in Materials and Methods chapter.

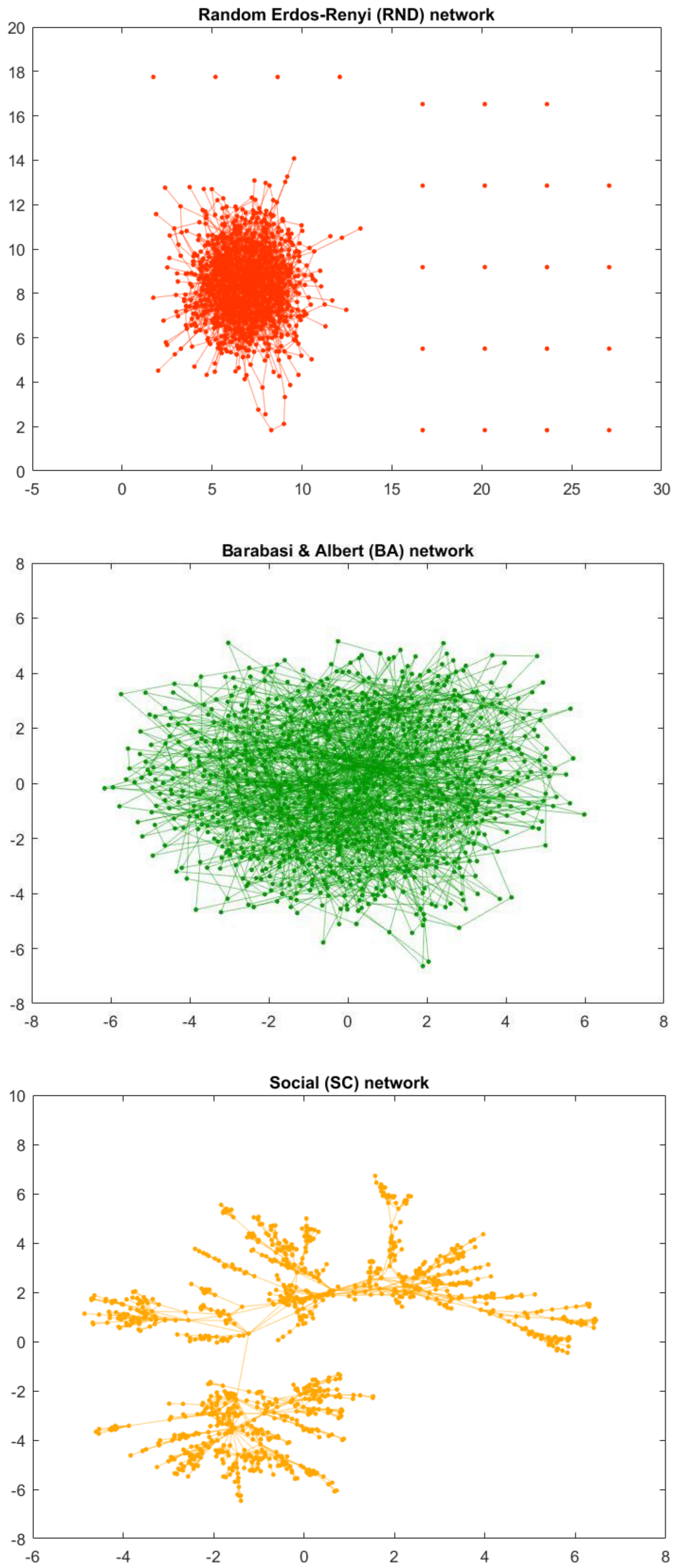


Figure 2: Random, Barabasi-Albert and Social network examples [4]

Two more things need to be highlighted from these papers. First, here as well the distinction between influentials and imitators is provided, based on the same definitions already given by the previous literature. Second, an important concept is expressed: in rural contexts, the behaviour of electricity demand often follows the diffusion of new electrical appliances and an increase in their use. By relying on diffusion models, it would be possible to simulate eventual scenarios of electrical appliances diffusion.

The reason why network theory is so strongly linked to diffusion processes in developing countries, is that people in rural villages need to learn about the technology from multiple people before they adopt themselves and the people they hear from tend to be the ones that belong to their links in the network of their community. In these cases it might be that some people represent better entry points than others at inducing cascades of information about innovations and it would be valuable to identify the ones that would maximize diffusion [40]. To do it, network structures need to be correctly calibrated.

In the beginning, researchers used to consider physical proximity as a good proxy for the connections that lead to technology diffusion. Beaman [40], instead, highlights that physical proximity does not appear to be a good proxy for social connections and cites Banerejee et al. [51], who say that in India, for example, a simple question like "if we want to spread information about a new loan product to everyone in your village, to whom do you suggest we speak?" is successful in identifying individuals with high eigenvector centrality and diffusion centrality, two parameters that allow to find the most suitable individuals to start an information cascade. The first one takes into account not only the number of links of the considered node, but also the number of links of its contacts, giving a better idea of how far information could spread; the latter instead adds the idea that the information is not going to spread more and more forever, but only takes into account a finite amount of time.

A strong supporter of peer-based networks helps the connection between network-theory and diffusion models stating that individuals want to act like their friends, individuals learn about the benefits of the technology from their friends, and individuals learn from their friends about how to use a new technology. Oster also says that peer effects are more important in early months after product distribution, which is shown by the concave behaviour of the value of information: when someone has no information about something, having some more information is very helpful, moving from having a lot of information to even more is less helpful. [52]

Once the starting scheme of the Bass model is set and its links with network theory are clear, it is possible to go back to the parameters definition.

If the purpose is to analyse the adoption of electricity and electrical appliances at the village level in developing countries, several studies tried to understand how the external influence should be considered in models [8], [53], [54]. Since most often no advertising nor marketing mechanisms are implemented in such environments, Riva et al. [4] make an example with p=0 as an assumption. The attention can therefore shift towards the definition of what is inside the parameter q and what instead could be added outside the entire formula (2), meaning that there might be some variables that imply an exogenous contribution to the adoption rate, such as income of a household or education level of people.
Once again, Bass contributed to the definition of this problem and, in a study he performed, he gives a number of possible alternatives to the initial model he had built, with the aim of including aspects such as price elasticity of electricity or income variation of households [55].
The study that most of all opens the path to the one here presented is included in a book by John D. Sterman of 2000 [7], in which he puts forward an innovative interpretation of the parameter q as the product of two sub-factors in the form

$$q = ci \qquad (3)$$

where c is the contact rate of the individual and i is the adoption rate, the two of which multiplied together give as a result the probability of adopting thanks to the process of word of mouth.

In the paper “A discrete bass model and its parameter estimation”, Daisuke Satoh estimates for several situations the values of p and q, like shown in Table 2, using two different methods [56].

The ordinary least squares procedure (OLS) involves the estimation of the parameters by taking the discrete or regression equivalent of the following differential equation, that is discretized with an ordinary forward difference equation.

$$\frac{dA(t)}{dt} = (p + \frac{q}{m}A(t))(m - A(t)) \qquad (4)$$

where the parameters are the same of equation (2), plus m, that is the total population, and $A(t)$ that is the total number of adopters at time t.
A time-interval bias is present in the OLS approach, since discrete time-series data are used to estimate a continuous-time model. To overcome the shortcomings of the OLS procedure, the nonlinear least squares estimation procedure (NLS) was designed, using the cumulative distribution function. The NLS estimation procedure overcomes the time-interval bias, but has some other problems: it may sometimes be very slow to converge or may not converge, it may be sensitive to the starting values for p, q, and m, or it may not provide a global optimum.

Procedure	p	q	m
OLS	0.01	3	100
dOLS1	-0.004693221	1.493157342	14717.17075
dOLS2	-0.007577906	0.843251539	6663.316838

Table 2: p and q values from literature

Lawrence and Lawton [57], instead, found out that $(p + q)$ ranged from 0.3 to 0.7 over several innovations diffusion processes. Thanks to this and other literature it was possible to establish a range for p going from 0 to 0.1.

It is possible to analyse on field gathered data and extrapolate empirical ranges of values for p and q parameters thanks to linear regression.
An issue that might affect the shape of the adoption curve is discussed in a paper by Bandiera and Rasul [39], who explain that having many adopters in the network, while allowing for better circulation of information, increases incentives to delay adoption strategically and free ride on the knowledge or adoptions accumulated by others.

Bonan et al. [58] as well, highlight something that should be kept in mind, which is the existence of spillover effects. People, indeed, do not only get information by their links in the network or being reached by some kind of advertisement: they might gather knowledge about something on their own, by seeing other people doing things in a certain way, even if they might not be among their known ones.

Dan Horsky [59] suggests the existence of models including the price decrease over time and also the existence of a category of “non-adopters” who cannot afford adopting.

Still, in reality, many more factors might get to influence the adoption of something. Going through past literature it was possible to find several examples of studies which tried to include cultural or financial factors inside some diffusion models, independently on whether they were using Bass model or others.
The most meaningful examples are listed in Table 3.

AUTHORS	INFLUENCING FACTORS	DEVELOPING COUNTRY
Pothitou et al. [60]	Education level, household income	NO
Peres et al. [25]	GDP per capita, health status, lifestyle	NO
Banerejee et al. [61]	Age, caste, education, language, native home, occupation	YES
Leijten et al. [53]	Monthly income, completed education, household compilation	NO
Eder et al. [62]	Health, income, education level, housing, work, food, transportation, mobile payment usage, bank usage	YES
Louw et al. [63]	Health, education	YES
Van Ruijven et al. [2]	Health, expenditure, household size, education, labour, electricity tariffs, geographical information	YES
Wilson and Dowlatabadi [64]	Income, education, absence of young children, people in ill health, elderly people presence, gender, age	NO
Sabah et al. [65]	Income, level of education, age, number of household members, interest in starting a business, house ownership	YES
Bonan et al. [58]	Household expenditure, schooling, health	YES
Vivi Alatas et al. [32]	Years of education, leadership position in the village, belonging to ethnic minorities or religious minorities, gender	YES

Bonan et al. [66]	Composition of the household, socio-economic status, health, education, income, working conditions, time allocation, savings, sources of energy, household expenditure on energy, appliances and cooking stoves, risk preferences, participation to informal groups, GPD localization	YES
Bandiera and Rasul [39]	Livestock ownership, income, food consumption	YES
Saweda et al. [67]	Household size, age of household head, highest years of education in the household, distance from the nearest farm, nearest market and nearest paved road, amount of land, value of non-productive assets	YES
Van den Broeck and Dercon [34]	Sex, age, completion of primary education, land holdings	YES
Oster and Thornton [52]	Age, grade, test scores, school fixed effects, parental education, family income	YES
Van den Bulte and Stremersch [36]	Income, wealth, education, occupation, aesthetic preferences, place of residence, lineage race	NO
Horsky [59]	Size of household, education, number of children	NO
Zhang et al. [68]	Income, profession, education, family size, social network, price of product or service	YES
Rao and Ummel [69]	Income, appliance price, affordability, reliability, race, religion, age, urban/rural, dwelling quality, vehicle ownership, household size, education, number of rooms, gender, home owning	YES
McNeil and Letschert [70]	Demography, health appliances, living standards	YES

Table 3: influencing factors found in literature

As it is possible to see in the table above, every effort made in the past to study the diffusion of some technology or novelty lead to the definition of some indicators which can all be grouped in three main categories: health, education and income, of which only income and education tend to be relevant when studying the diffusion of electric appliances.

Sopha et al. [71] proposed an innovative way to consider the decision making process of the individuals when deciding to adopt an appliance. Their theory categorizes four decision strategies: repetition, consumers will habitually consume a product that they have previously consumed; deliberation, consumers will evaluate all possible alternatives and consume the best one; imitation, consumers will choose the product that most of their social network consumes; social comparison: consumers will conduct a social comparison by comparing the product previously consumed with the

product that most of their peers consume and choose the best between those two. Also, they stated that parameterizing the ABM using survey results is a promising approach, because it provides a strong empirical foundation for the development of an agent based model [71].
While this paper looks for social characteristics of the individual to affect the decision making process, others tend to look at the subjective characteristics of each node.

Pothitou et al. explain how household income, and to a lesser extent gender, is associated with energy-saving habits and behaviours [60].
Menezes et al. [43] highlight that Bass model ignores the existence of more rigid barriers to adoption of new products by population, such as low level of income per capita and define the difference between total household earned income and fixed expenses which cannot be compressed (e.g. food, health, etc.) as the average disposable income of the households. The price of a technology needs to stand below this threshold in order for a household to be able to adopt it, that is, to become a potential adopter in the Bass diffusion model. Otherwise, the household will be a non-adopter for that time step, until its disposable income will update.

Bass model, though, is not the only type of diffusion model which allows to obtain the curve of spreading of an innovation across a group of people. Another example is provided by Gompertz curves, which involve aspects which are more related to the economic conditions of the individuals, while neglecting other parameters related to the social influence. Van Ruijven [2] suggests that household size and temperature should also be considered, plus adds a distinction between what happens in rural or urban areas and high or low income categories. Abdullah and Jeanty consider a further differentiation between private households and economic activities [65].
Other three aspects of van Ruijven's paper are really relevant. First, it gives the advice, in case of lack of data, to use the application of electricity for lighting as a proxy for electrification rates of households [2]. Second, it cites an article by Kemmler [8], which states that household expenditure is the main correlating factor for electricity use by household, to be considered by dividing the population in quintiles. Third, it highlights several clusters of appliances, represented by: space cooling

applications, food storage facilities, washing machines, entertainment and communication appliances.

Thanks to these ideas it was possible for Van Ruijven to introduce the concept of ownership of the appliances and to move from the use of Bass model, to the building of Gompertz curves. These are supposed to give a similar output as Bass model, but instead of taking into account the values of the endogenous p and q parameters, are based on a correlation with the household expenditure, which is explained by the following equation.

$$Ownership_{q,A,U(t)} = \alpha_A * EXP(-\beta_{A,U} * EXP(\tfrac{-\gamma_{A,U}}{1000} * PCOpc_{q,U(t)})) \qquad (5)$$

Where PCO is the expenditure per capita, β and γ are exogenous coefficients based on linear regressions from real data collection, differentiated by appliance, while α is the upper limit of appliance ownership.

Van Ruijven explains how to build ownership curves in its paper “Model projections for household energy use in India” [2]. In it, useful values for the parameters of the diffusion model for appliances are provided, which are the only available in the existing literature for one developing country (i.e. India).

Many others tried to use different methods to develop diffusion models of appliances.

A first example of it is by Labandeira et al., who develop the so-called model of Random Effects. The idea is that households do not demand electricity for direct consumption but rather use it to produce a series of final goods and services: the final energy good (x) can be defined as a function dependent on the electricity consumed (e) as well as the natural gas consumed (g) and the stock of household appliances (a) [72].

$$x = f(e, g, a) \qquad (6)$$

At first, consumers tend to minimize the costs of producing the energy good, then, they maximize their utility and when the price of electricity varies, households modify their stock of appliances.

Van Den Bulte and Stremersch [36] conclude that diffusion curves reflect the level of income distribution, so that networks lose their commonly given importance. This finding will be very useful for the current work in the next chapters, even if the attention will move from the GINI index to expenditure levels. Moreover, they concentrate their efforts on the estimation of the ratio *q/p* rather than on the two single parameters. Lastly, they assign much importance to one type of contagion, that is cross-cultural and social-normative one.

Assimakopoulos [73], instead, proposes an innovative way of forecasting residential energy demand through appliances usage. This method consists of applying energy demand equations to 'homogeneous' groups of consumers which are endogenously defined by using multivariate statistical techniques on data. The decisions of households are then simulated. The repartition by energy products is then estimated for each group.

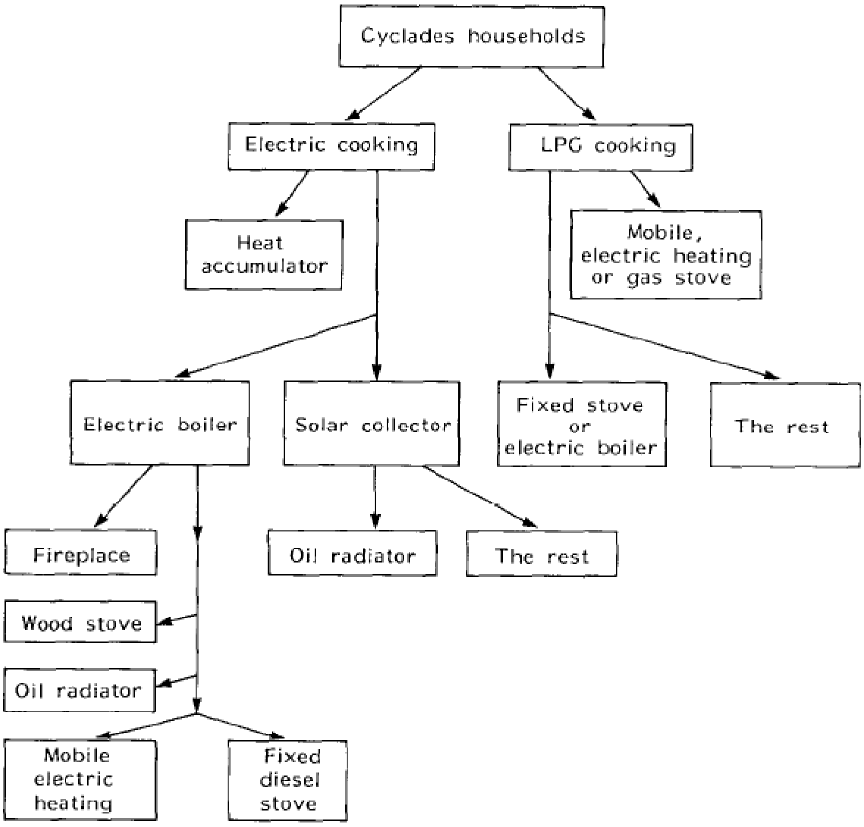


Figure 3: example of structure of energy choices for a case study in Cyclades Islands

Zhang et al. [68] create a model which includes income, profession, education, family size and social network of each segment of the network and captures the diffusion correlation between dependent technologies. The diffusion of one technology or product, indeed, may impede or improve the diffusion of another one.

Narashima D. Rao and Kevin Ummel in their paper “White goods for white people? Drivers of electric appliance growth in emerging economies” explain how ownership can be defined through parameters like market access, wealth, plus sometimes race, but most of all they underline the importance of considering affordability rather than income, which accounts for appliance price as well and can therefore change more easily over time. Each individual, depending on its own expenditure level, will have a marginal probability of owning a certain appliance. Given this, the individuals with the higher marginalities will be the ones that first adopt a certain technology, the others following in descending order. Affordability sets a threshold, which means that through affordability it is possible to choose a certain percentage value that should represent the share of expenditure to be maximally devoted to the adoption of an appliance [9].

The problem of this type of approach is that it is very context-specific and it is very difficult to use the data from a certain place as a basis for a theoretical analysis of another place. Therefore, for the current study it was necessary to find a different solution, consisting of making a ranking of the individuals based on their expenditure level.

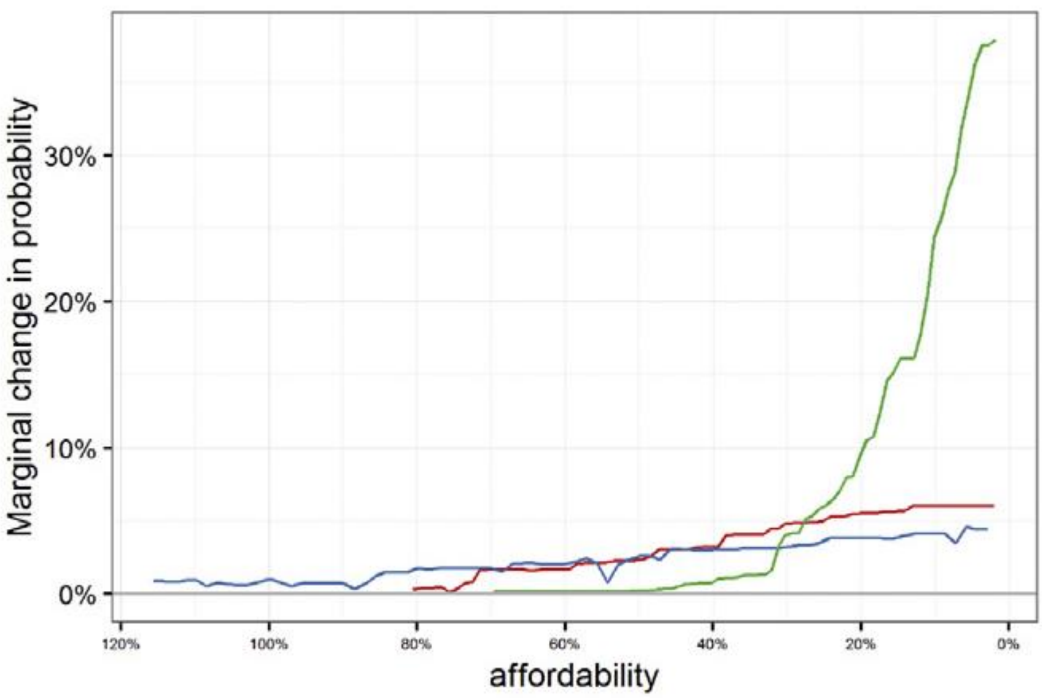


Figure 4: Marginal probability to adopt with respect to affordability level

In terms of complexities behind the forecast of electricity demand, another issue should be considered, that is the behaviour of electricity users that generate different electricity load profiles. Several studies have been carried out in order to assess the influence of consumer behaviour and habits on the elasticity of demand and on demand fluctuations.

In most developed countries energy habits strongly depend on household income and to a lesser extent gender. Positive environmental attitudes are not necessarily indicative of a high level of knowledge of environmental issues or energy saving, but monetary incentives can be a very strong influencer of people behaviour. The conclusion of Pothitou et al. is that people with high environmental motivation are less sensitive to price than average [60].

Rai and Henri state that there is the need to understand why people use energy the way they do and how individuals respond to information about the costs and benefits of energy choices. Networks shape individuals' access to information about technologies, their costs and benefits, and their propensity to adopt new patterns of behaviour. This means that it is not possible to separate the study of energy demand from the study of social interactions [74].

In order to link the diffusion processes and the obtainment of a load curve, an important economic tool is necessary, which is the creation of S-shaped curves. These are obtained as a final output of the diffusion processes. Their structure can be easily explained by looking at the different phases of adoption of a technology. At first only innovators (early adopters) purchase a technology, even if it still has no feedbacks. In fact, it will be they the ones who are going to give feedbacks to the rest of the community. After this first phase, the S-curve can either form, or it can fall back down, meaning that the technology does not spread at all.

If the innovation does spread across the network, a phase of strong growth and high adoption rate will follow, until the majority of people will own that technology and the market will reach saturation. Then the curve will flatter again and the diffusion process will be over.

It is again the paper by Riva et al. that shows the correlation between different types of network and different shapes of diffusion curves [4].

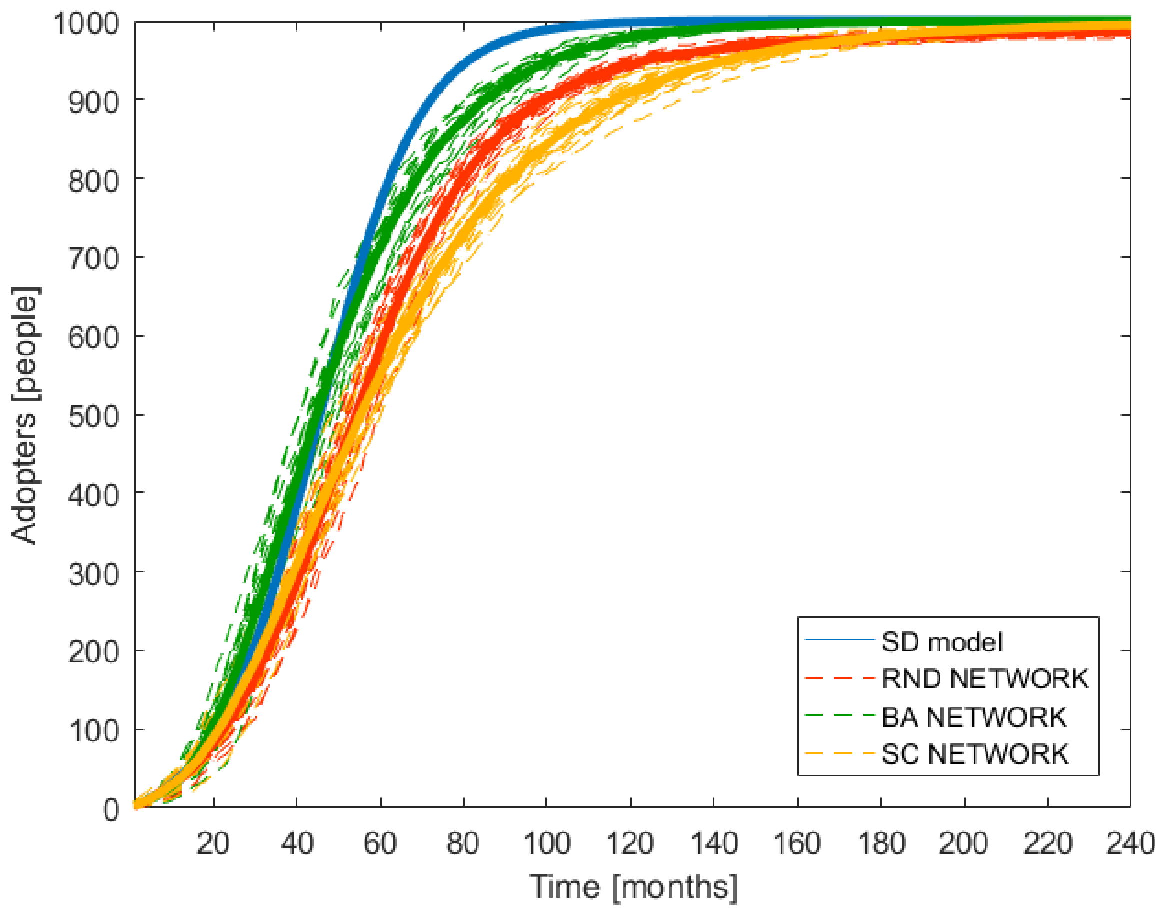


Figure 5: S-Shaped curves for k=4, A(0)=0 and different types of network

In the picture it is possible to see an example of the difference among the diffusion processes across different network structures, assuming initial adopters equal to 0 and an average degree of 4 contacts. Once the diffusion process has been assessed, it will be possible to have clear which appliance was adopted by each household. In order to be able to size an entire off-grid system, though, it is necessary to have more specific information about the load curves that the microgrid will have to manage.

2.3 Load Curves

In order to make reasonable forecasts, several methods can be used which can be distinguished based on type of sizing process to perform.

Intuitive sizing

For intuitive sizes of off-grid systems, especially in the case of home-based system and small capacities installed (i.e. in the order of Watts), load forecasting methods are based on the approximation of the reasonable peak power the power system should be able to respond to. In other words, the process can follow two possibilities. First, one can take in consideration the energy consumption related to each appliance and simply sum up the nominal power of all of those, assuming they might be switched on at the same time.

$$E_c = \sum_j^{Userclass} N_j (\sum_i^{Appliance} n_{ij} P_{ij} h_{ij}) \quad (7)$$

where E_c is the total energy consumption, N_j is the number of households per user class, n_{ij} is the number of appliances of a certain type per household, P_{ij} is the nominal power of the appliance, h_{ij} is the number of hours of operation of that appliance. Taking then into account the efficiency of the electricity generation it will be quite straight forward to obtain the amount of capacity we need to install. This first method might, nonetheless, bring to overestimations of the real load peak. The second option, instead, is to take the average load of a day and spread it across all the hours, so to make it uniform. This, on the other hand, will likely cause underestimations of the load peak.

Numerical sizing

For more detailed design processes, analysis of operation of the systems, long-term simulations, *etc.,* more structured models are needed to generate reliable load profiles.
In this case, more details for each appliance will be requested, such as the functioning windows and functioning time, which respectively represent the moment during the day in which a certain appliance might be in use and the total amount of time an appliance keeps working every day.

It is Mandelli [10] who helps making further distinctions among the load curve forecasting procedures. He explains that load forecasting can be divided in: short-term, which is used to predict loads from 1 h to a week ahead; medium-term, used to predict weekly, monthly and yearly peak loads up to 10 years ahead and is required for efficient grid operational planning; long-term, used to predict loads up to 50 years ahead and is required for grid expansion planning. A second categorization would divide the forecasting methods as top-down or bottom-up approaches, whose definition is well explained in the image below.

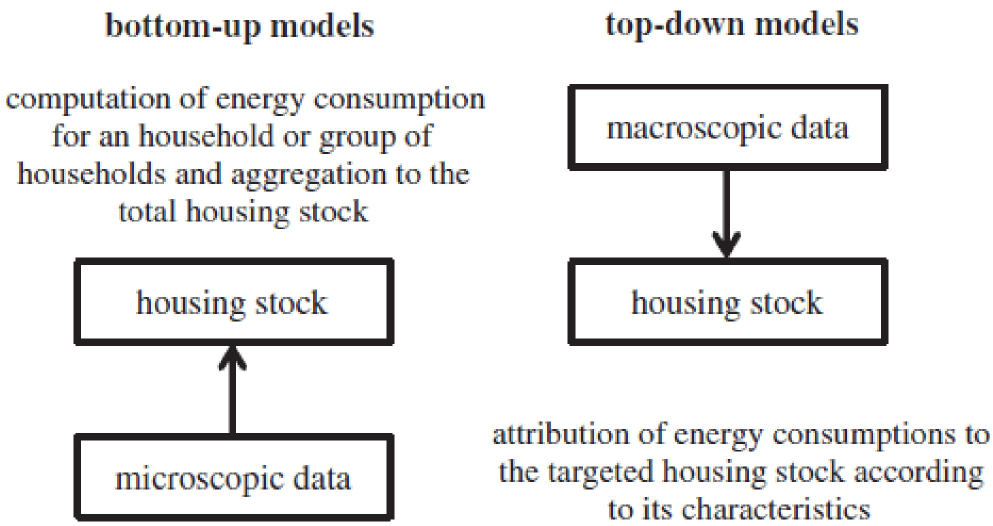


Figure 6: top-down and bottom-up forecast approaches definition

Given these definitions, it is possible to take into account what Grandjean says an ideal model should look like. [75]

▪ It has to be parametric in order to simulate various scenarios.

▪ It has to be technically explicit, i.e. the different specificities of the simulated appliances must impact the load profile results.

▪ It has to be evolutionary, i.e. new elements can be introduced so as to be simulated.

▪ It has to be aggregative, so that results can be obtained at different levels (household, city, region, etc.).

▪ All end-uses can be considered in the load profile calculations.

In the light of this reference, Mandelli developed a new procedure, called LoadProGen and characterised by the following features:

▪ It is based on input data that can be easily assumed based on practical experience on similar context conditions or by mean of local surveys.

▪ It has to be based on a rigorous mathematical formulation, which allows generating the load profile, i.e. apart input data, the designer judgments should not affect the profile shape.
▪ It has to be bottom-up, i.e. the load profile formulation has to rely on microscopic input data referring to each appliance's features within a specific type of user class.

With LoadProGen approach, each appliance of each household contributes to the load profile with its power demand. Through a stochastic approach, the switching-on times are defined and a realistic functioning of the appliances is simulated. For each user class a proper peak is obtained, thanks to a relation between load factor (ratio between actual electricity used and maximum possible theoretical usage), coincidence factor (ratio between actual power peak and possible theoretical power peak) and number of users. Therefore, the load curves will present many spikes, whose values will not be random, but will result from the features of the considered appliances. LoadProGen is based on input data that can be surveyed or assumed in rural areas. It is important to know that the main purpose of this procedure is not to forecast load profiles, but rather to formulate them in an appropriate manner to support electrification studies in rural areas. [10]

Once the load curves are available, it is eventually necessary to have a mathematical tool which allows us to plan and size the off-grid system.

2.4 Off-grid System Planning

Rojas-Zerpa [11], in his work about energy planning, explains which are the main aspects to take into consideration when deciding which is the tool that should be used to design off-grid systems. First, the application area: different models might indeed imply different planning methodologies. When considering rural areas, though, models for decentralized (or distributed) generation have recently gained a lot more interest. Second, the planning horizon: when designing an off-grid system it will be necessary to decide which is the time period we are willing to take into account. As Rojas-Zerpa says, the majority of existing literature takes into account short-term (1-4 years) or medium-term (5-10 years) periods, but it is in fact useful, in some circumstances, to take into consideration long-term (11-20 years) planning tools.

Third, the objective of decision making, which might be either economical, or social, or even environmental. Fourth and last, the technologies to be included in the planning, depending on whether the off-grid system should be made of renewables only, or some other technologies are going to be implemented as well.

It is in this context that a software called HOMER is first taken into consideration. HOMER (Hybrid Optimisation Model for Electric Renewables), developed by NREL (National Renewable Energy Laboratory, USA), appears repeatedly in the literature as a preferred tool [12]. It can handle a large set of technologies (PV, wind, hydro, fuel cells, boilers, etc.), loads (AC/DC, thermal and hydrogen), and can perform hourly simulations. HOMER is an optimisation tool that is used to decide the system configuration for decentralised systems. Its target is to find the cheapest solutions in terms of Net Present Cost, respecting the input constraints the user can give. The major drawback of this software is that it does not take into account the evolution of load curves in time and considers one only load profile for the entire lifetime of the system. It is, in fact, quite usual to find studies which take into account the long-term time horizon, but decide to use a constant load demand for the entire period of the study anyhow, which will likely bring to inaccurate results.

Once the literature review was over and a sufficient knowledge of all the interesting topics for this research was developed, it was then possible to shift to the actual building of the model and to the learning of the functioning of the different necessary tools.

3 Materials and Methods

In Figure 7, it is possible to observe all the steps I will explain in this chapter.

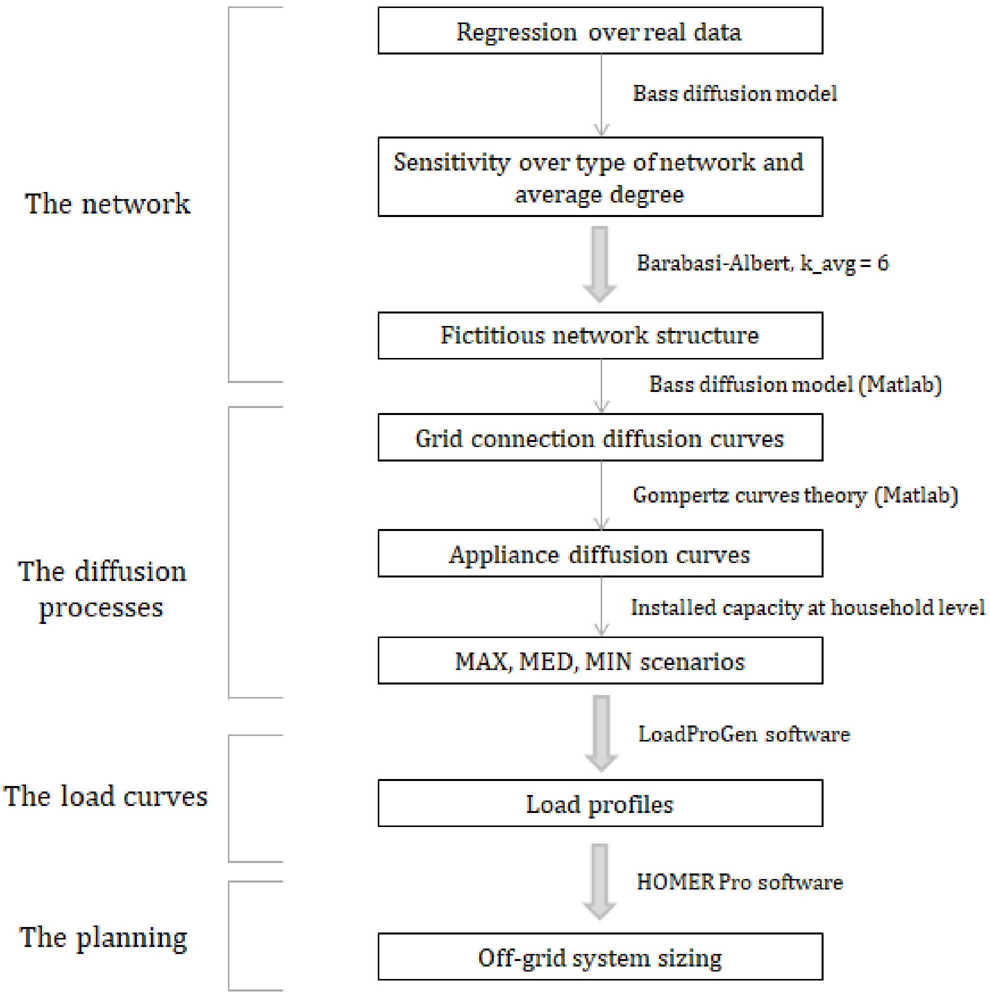


Figure 7: flow diagram of the study procedure

3.1 The Network

In order to build a realistic model, it was necessary to start from real data. Even if the aim of the work is mostly speculative, it was possible to start the research from some data that had been collected in the region of Njombe, in Tanzania, more precisely in the village of Bulongwa. These data were collected from the control unit of an existing mini-grid, which would allow knowing at each time-step (years) how many people were connected to the mini-grid for the first time.

Bulongwa is a village of approximately 700 households, where the South-Central Diocese of the Evangelican Lutheran Church landed with a project for a hospital, which was open in 1968. This same project allowed, several years later, the creation of a mini-hydropower facility (180 kW), which would feed a mini-grid in the village, which is considered by GIZ in a report from 2009 about hydropower in Tanzania [76].

Figure 8: The hospital of Bulongwa

The data, which were provided by Chalmers University of Technology, are relative to the connections to this same grid and range from year 2009 to year 2016, while the first connection dates back to 2001.

In order to obtain the S-curve of diffusion of the connection to the grid, it was possible to perform a regression which allowed obtaining a diagram for the entire lifetime of the system. Data were collected twice in the years and differed slightly; therefore, a yearly average was taken for the current study use. To obtain the values from 2001 to 2009, I decided to take into account the growth rate of the number of connections between 2009 and 2010 and to keep it constant going backwards in time. The amount of connections obtained at year 2001 was assumed to be the number of people who had first connected at year 1 of the system. The resulting curve is shown in the graph, being the period after 2009 (marked in red) the relevant one.

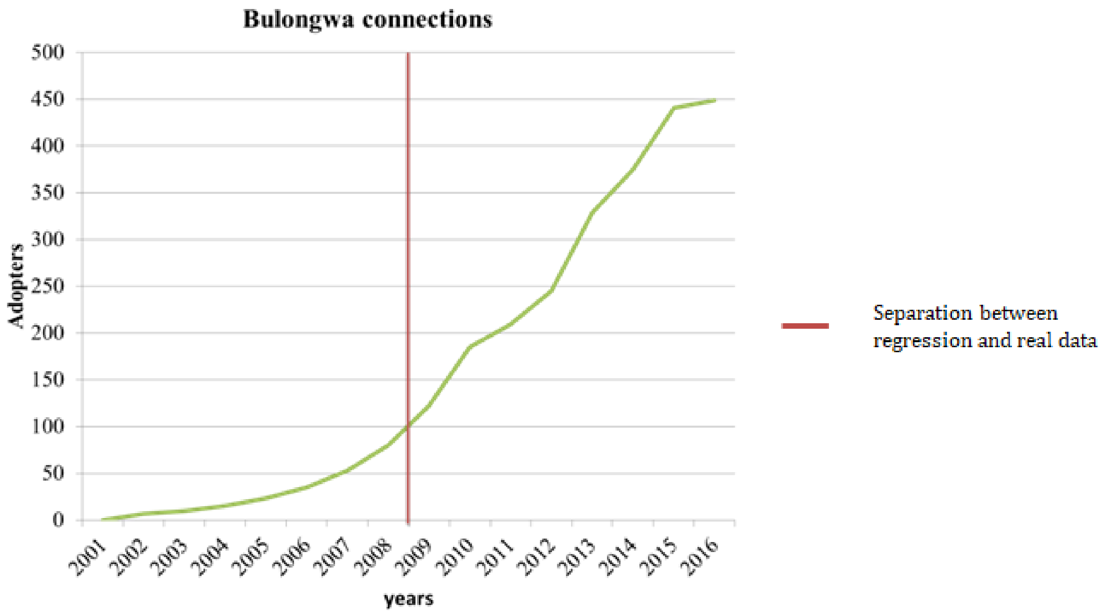


Figure 9: Bulongwa connections to the grid

The first thing that had to be included in the model was a way to produce the network in case it was not given. It is important to notice that in this case each node of the network represents a household and not an individual person, so that each node might represent more than one subject. Starting from the paper by Riva et al. [4], it was possible to define which type of network mostly resembled the one of Bulongwa, which was not given.

The procedure consisted of using the available data to extrapolate the values of p and q parameters, in order to use them in a Bass model procedure and, through a sensitivity analysis, it was then possible to find which type of network, with the right input parameters, would give as an output the diffusion curve most similar to the one of Figure 9.

As a first step, a linear regression was performed, starting from Bass formulation of its model.

Starting from equation (2), it is possible to re-write the model, multiplying everything times N, so to obtain:

$$Nf(t) = (Np + ciA(t))(N - A(t)) \qquad (8)$$

where N is the total number of individuals, or, in this case, households. This can easily be written in the following manner:

$$AR(t) = pP(t) + ci\frac{A(t)}{N}P(t) \qquad (9)$$

where $AR(t)$ is the adoption rate at time t, $P(t)$ is the number of potential adopters equal to N minus the number of actual adopters, all at time t.

Looking at this equation, it is possible to notice that the only missing information in the real data from Bulongwa are the values of p, i and c, or p, q and one between c and i, being the three dimensions dependent one on the other. Being equation (9) a linear one, it was possible to extrapolate the values of p and q through linear regression.

Given the formula

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k + \varepsilon \qquad (10)$$

and

$$b_1 = \frac{(\sum x_2^2)(\sum x_1y) - (\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2} \qquad (11)$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2y) - (\sum x_1x_2)(\sum x_1y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2} \qquad (12)$$

and knowing that

$$\sum x_1y = \sum X_1Y - \frac{(\sum X_1)(\sum Y)}{N} \qquad (13)$$

$$\sum x_2y = \sum X_2Y - \frac{(\sum X_2)(\sum Y)}{N} \qquad (14)$$

$$\sum x_1x_2 = \sum X_1X_2 - \frac{(\sum X_1)(\sum X_2)}{N} \qquad (15)$$

it was possible to obtain b_1 and b_2 values that in the current case correspond to p and q values.

Given the values of p and q, it was then possible to build in Matlab a Bass model simulation, which would allow making a sensitivity analysis around the value of c, which in the script was called *k_avg* (average degree).

First of all, based on literature, it was decided that *k_avg* should vary between 2 and 8.

Then, given the network creation process described by Riva et al. [4], it was possible to re-adapt it in order to obtain 12 different networks, built with three different methods and each of these for four different values of *k_avg* (2, 4, 6, 8).

The three types of network that were used are the random, the Barabasi-Albert and the social one, which can be obtained using the following equations for the probability of a node to have a degree k.

Random network

$$p(k) = \frac{e^{-k_{avg}} * k_{avg}^{k}}{k!} \qquad (16)$$

where p is the probability, k is the considered degree and k_{avg} stands for average degree of the network, given by the equation

$$k_{avg} = (N-1) * p \qquad (17)$$

where N is the number of nodes of the network.

Barabasi-Albert network

$$p(k) \approx k^{-q} \qquad (18)$$

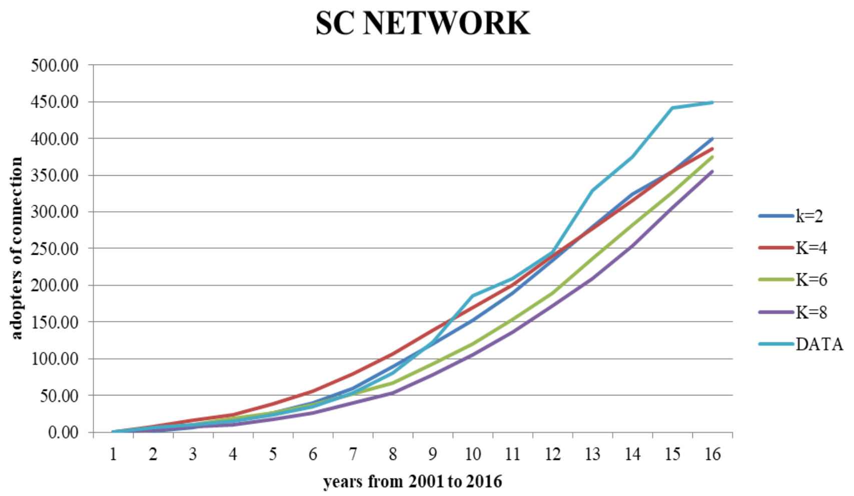
where p is again the probability and k the considered degree, while q is an experimentally determined parameter equal to 3 for BA networks.

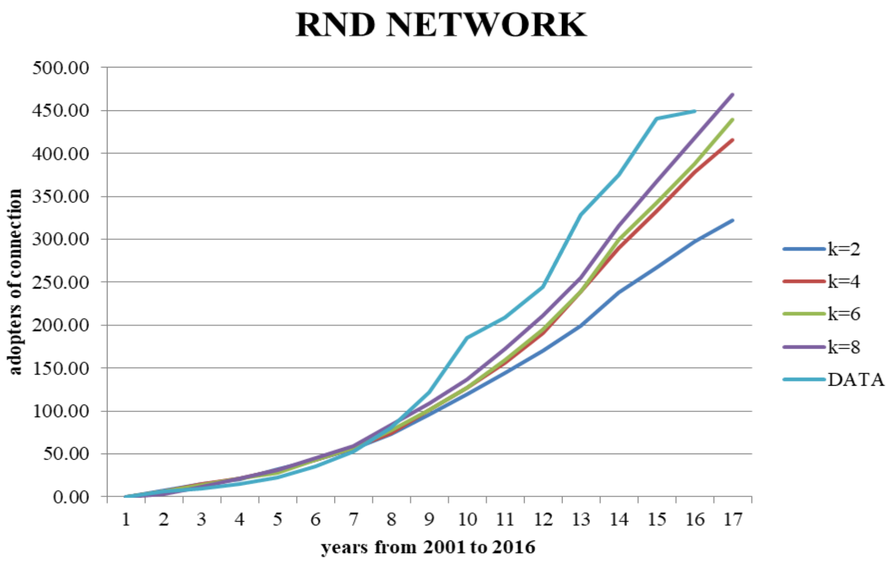
Social network

$$p(k) = \alpha * (k+\beta)^{\frac{-2}{m_{s,avg}^{-3}}} \qquad (19)$$

where p and k continue to be probability and degree, α and β are two experimental constants and $m_{s,avg}$ is the average number of nodes that at each time step gets m_r new contacts.

Once these structures were clear, twelve network graphs (3 types of networks for 4 values of *k_avg*) were built, in which each row and each column represented a household of the village and the boxes were equal to 1 if between the households there was a link and to 0 if there was not (the diagonal was therefore filled with zeros, since no loops around oneself are possible). These graphs were, one by one, given as input to the Bass model with p and q equal to the ones obtained from the regression and with a different value for *k_avg*, depending on which network was used. Twelve diffusion curves were obtained and, in order to understand which one was closer to the real data one, the standard deviation was evaluated for all the curves, taking into account only the relevant period from 2009 to 2016 (see Figure 10). The curve with the lower error was chosen and a certain type of network, the Barabasi-Albert, was therefore identified, with a *k_avg=6*, that allowed to simulate better the real process.





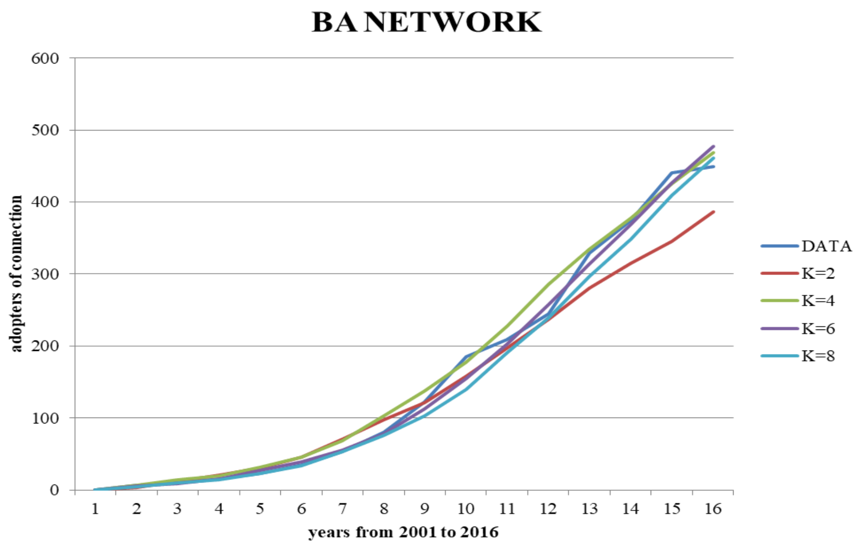


Figure 10: Social, Random and Barabasi-Albert network sensitivity and standard error evaluation

The values of *p*, *q* and *k_avg*, which were extrapolated from the regression and the simulations, were then compared to the values that can be found in literature and they happened to be in accordance with them.

3.2 The Grid Connection Diffusion Process

In general, once *k_avg* value is established, it is possible to obtain the range of *q*, or of *i*, which are both correlated with *k_avg* with the formula *q=k_avg*i.* In my case, the only obtained value will be *i*, while ranges of values of *p* and *q* were taken from literature.

This preliminary phase of the research allowed to decide that for the following simulations and studies one only type of network would have been used, with *k_avg=6*, while the variety of *p* and *q* would have been brought forward in the next phases of the analysis, in order to make the research more comprehensive. To let the parameters vary, the Monte Carlo method was used by repeated samplings from a uniform distribution around the average values found in the literature. This algorithm allows performing several simulations of the same process using every time a different (random) combination of the input parameters, assigning them the values chosen within a certain range. It therefore allows the outcome to be more complete and effective, since it would take into account several reasonable possibilities starting from several ranges of values provided by the literature. To decide which parameters should be let varying, the given data should be deeply analysed and understood.
One more necessity, instead, would be deciding where the ranges for the varying values should be taken from.

The parameters that was decided should vary are *q*, the probability of adoption thanks to word of mouth and *p*, the probability of becoming connected to the grid thanks to external influence from advertising.
The reason why some parameters were kept constant and others were not is based on the fact that *p* and *q* are endogenous factors, while the type of network and the value of *k_avg* can be evaluated through a survey. For sake of simplicity it was chosen to simulate and study a fictitious village composed of 400 households only, assumed

to be located in the same geographic area of Bulongwa, but having no access to the grid at the time in which the study begins.

At this point, one Barabasi-Albert network composed of 400 households was built, with a *k_avg=6*.
Once this was done, the diffusion process of the connection to the off-grid system could be run.

It was decided to manage the attachment to the grid through a standard Bass model. The values of p and q should be let vary through the Monte Carlo method.
p was assumed to vary among 0 and 0.01 and q among 0.2 and 0.7, ranges that were taken from literature [56], [57], therefore the simulations were made for

$$p = 0.002 + (0.01 - 0.002) * rand \qquad (20)$$

and

$$q = 0.2 + (0.7 - 0.2) * rand \qquad (21)$$

100 scenarios were therefore obtained, through a cyclic script which would create 100 possible combinations of p and q values, also letting vary several other values related to the diffusion of appliances. A cycle for the simulations was created and inside the cycles, the diffusion process typical of a Barabasi-Albert network would take place, based on the previously obtained network. In fact, several simulations can be performed also of this sub-process of diffusion and a mean result can be then taken as a final output for each cycle. The Matlab script for this diffusion process can be found in Appendix A.
In order to keep track of the values that were assigned to each variable at each simulation, several arrays were built, which would update at the beginning of each cycle in order to save the random values created at each round and be able to perform some analysis based on the Monte Carlo method at the end of the process.
Once the diffusion processes of electricity were obtained, it was possible to analyse appliances diffusion.

3.3 Appliance Diffusion Process

This second part of my model has two main objectives. The first is to provide the diffusion curves of appliances across the village, allowing me to know how many of

each appliances are purchased at each time step. The second target is to know which of the households purchase a certain technology at each time step. A first thing to notice, is that the adoption of the grid connection is assumed to be contemporary to the adoption of a lightbulb, which is considered to be a good proxy for electricity use [2], and will not appear among the studied appliances.
The clusters of appliances were chosen according to van Ruijven [2] and it is necessary to decide how to allocate the appliances across the network, how many units of each component are sold and to whom. In order to answer these questions, the current study proposes an innovative possibility, which is studying the sizing of a grid thanks to the construction of two types of tools: ownership curves, depending on household expenditure, and load curves, depending on the effect of expenditure on the level of diffusion of a technology. These tools are already widely used in economic development studies and their use is favoured by the possibility of using standard surveys that allow estimating the needed values, with fewer uncertainties, right on field.

When someone gets to know of the existence of some interesting technology, which could improve his/her life conditions, there is still one aspect, at least, that might force him/her not to adopt it, that is: he/she might not be able to afford it.
In order to know how many appliances had to be adopted at a given time t, it was decided to use a similar approach to the one suggested in the literature by van Ruijven [2]. His model is a bottom-up approach to describe the evolution of residential energy use in India, starting from the dynamics of development and per capita expenditure. The author demonstrates, thanks to validation through historic residential data, that the variation in income distribution significantly influences future projections of off-grid systems. Van Ruijven takes into account a concept called "ownership" which depends on the behaviour of per capita expenditure level along the years and depends on the following equation, already presented in the literature:

$$Ownership_{q,A,U(t)} = \alpha_A * EXP(-\beta_{A,U} * EXP(\tfrac{-\gamma_{A,U}}{1000} * PCOpc_{q,U(t)})) \qquad (5)$$

Through the formulas that had been derived by van Ruijven, it was possible to create some reference ownership curves also for the country of Tanzania.

The World Data Bank provides the yearly behaviour of per capita expenditure in Tanzania. Since the entire study by van Ruijven is based on quintiles, it was decided to keep the same format. It was therefore created, around each yearly value of average per capita expenditure, a distribution of 400 values, which were subsequently divided into 5 quintiles each. Of each of these quintiles, an average value was taken for each year and, putting all of the data together, 5 behaviours over time of the average expenditures of the 5 Tanzanian village quintiles were obtained.
Starting from the values and equations in van Ruijven's study, it was possible to obtain the yearly level of ownership, in each quintile, of each technology, depending on a correspondent level of average expenditure of the population.

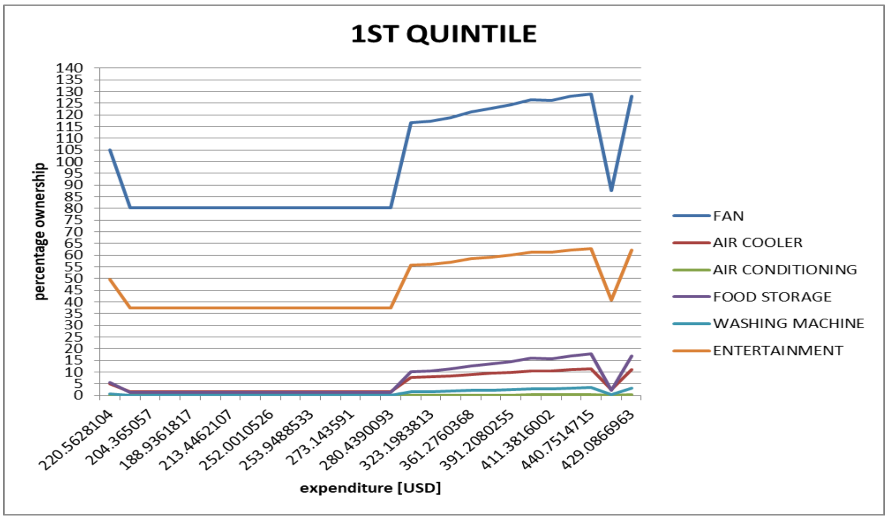


Figure 11: 1st quintile appliance diffusion

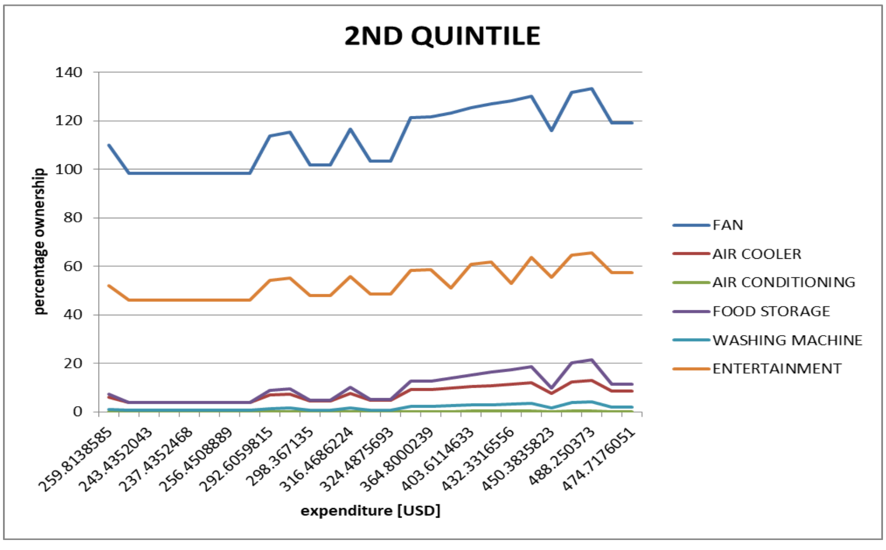


Figure 12: 2nd quintile appliance diffusion

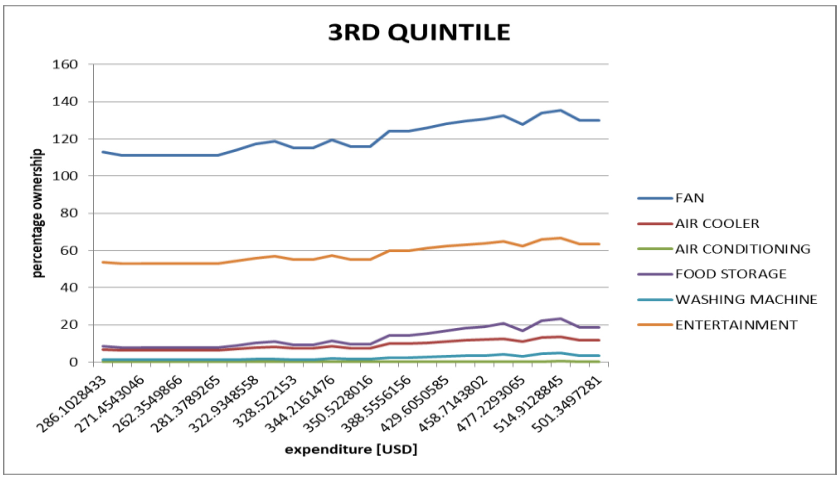


Figure 13: 3rd quintile appliance diffusion

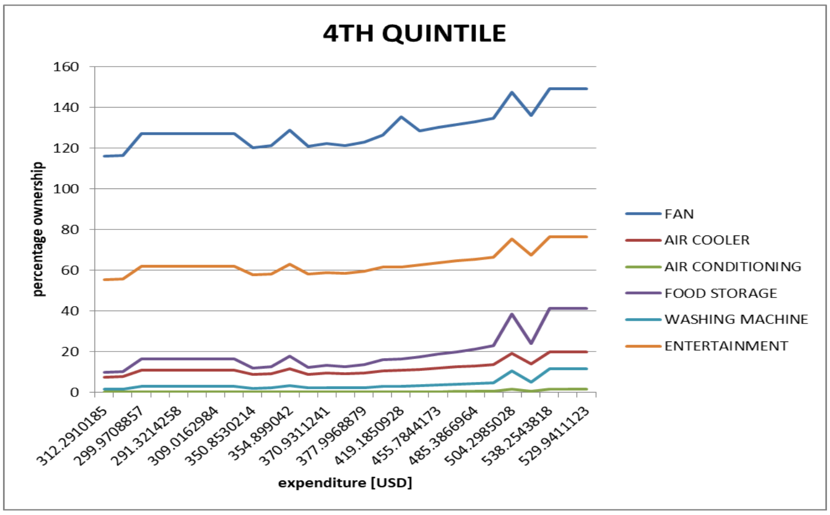


Figure 14: 4th quintile appliance diffusion

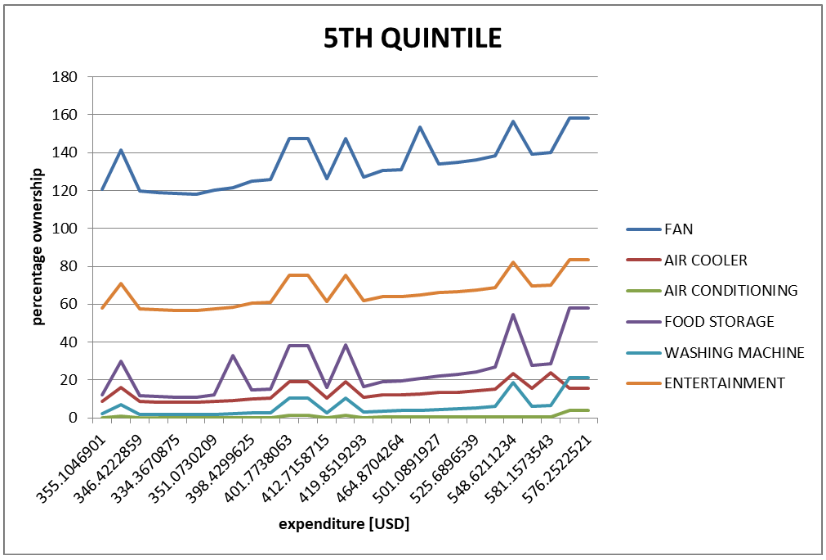


Figure 15: 5th quintile appliance diffusion

The obtained curves allowed understanding which were the appliances that were mostly adopted by the households of each quintile, given a certain economic availability. Once the reference curves of ownership were built, it was necessary to develop a method which would allow allocating a certain amount of appliances, depending on the correspondence on the ownership curves, within the population of the considered village.

The level of ownership should depend on three aspects.

- The expenditure available for a certain appliance: indeed, at each time step, only the first appliance in order of adoption would be acquired starting from the full amount of available money, while the other appliances would be acquired during the same time step only if the remaining money was enough. In the model, the order in which the families were able to adopt was decided to be based on the price of the appliances, from the cheapest to the most expensive, which was: fan, entertainment facilities (TV), food storage facility (refrigerator), washing machine, air cooler (portable), air conditioner (cooling or heating system).
- The position of a person in the social ranking. The social ranking is a list of the households based on the level of expenditure of each one. The idea is that, as long as they are connected to the grid, richer people adopt before with respect to the others because they have greater economic availability. Therefore, when trying to allocate the appliances to a certain percent amount of people of each quintile, it was possible to put households in a ranking and allocate the appliances only to the people connected to the grid who could "better" afford them.
- β and γ parameters, which were empirically evaluated in the study of van Ruijven for the Indian area. Since Africa is still a developing country it is possible to assume that the values of the parameters of appliances are similar to those of India. Yet, it would be too much of a strong assumption to state they are the same. This is why, once again, Monte Carlo method was used in order to obtain a variability of the Indian values equal to a +/- 20%. At each simulation, besides choosing a random value for p and q, random values for β

and γ parameters of each appliances were obtained as well. These values would be saved inside several arrays to keep track of their history.

In order to simulate a realistic process for the diffusion of appliances, it was decided to create a set of 400 realistic expenditures, based on the average values of expenditure per capita of Tanzania, provided by the World Bank for the year 2016. Of course in a real setting, with a real investment analysis going on, it would be necessary to collect the real data on-field. Since it would have to be an exogenous value, expenditure was assumed to update each year in order to go back to the starting value, assuming no changes in the economic status of the considered village for 20 years, which were supposed to be the lifetime of the considered off-grid system to be built.
Given the average expenditure per capita, the 400 values of expenditure were obtained and then ranked in ascending order. The households were then divided into quintiles and for each quintile the average expenditure was calculated. From these values, it was possible to obtain the level of ownership of each quintile, which represents the number of adopters that should be reached at that certain time step. The allocation of appliances would go in order of price, meaning that the cheapest technologies were allocated first at each time step and then, based on the remaining economic availability, the households could "decide" to acquire something else.
A model was therefore written which would allow the performing of all of these steps and would be a cycle going on for an arbitrary amount of time. The time step which was chosen was 1 year and the likely lifetime of the off-grid system was decided to be 20 years, therefore the cycle was built for this exact amount of time.

For each one of the 6 clusters of appliances a script like the one presented in Appendix B was created. This same script was repeated for each appliance cluster and the available expenditure would decrease going through the technologies, while at the beginning of each time-step it would get back to the initial value.

For each appliance the parameters related to the ownership curve are necessary and can be evaluated through empirical data taken from surveys, while the costs were

taken, as already explained, by a catalogue written by GIZ, just like it was done for the nominal power [77].
Along the code, it is possible to see a term called *aff.* This parameter represents the affordability and in the considered study it was set to be equal to 1.2, meaning that the cost of the appliances had to be smaller or equal to the 80% of the expenditure in order for the adoption to be possible. The concept of affordability was found in a study by Rao [69] and its meaning is that most people would not spend their entire budget of the year on something that is not of primary necessity.
Eventually, the two outputs of this second part of model would be:

- as many sets of S-curves of adoption of the appliances as the number of simulations (100 in the current case). Each set of curves was characterized by a different combination of random values of *p*, *q* and appliance parameters. Each set containing 6 S-curves of diffusion, relative to the 6 clusters of technologies available in the code.

- a set of matrices containing the record of who had purchased a certain technology and at which time step had done so.

Thanks to these outputs further research could be done, as will be next explained.

3.4 Load Curves

Once the appliance diffusion model was created and the 100 simulations were ready, in order to build the entire load curves, it was possible to use a software called LoadProGen.
It is Mandelli et al., who explain how this innovative software works in a paper from 2016 called "Novel procedure to formulate load profiles for off-grid rural areas" [10].
LoadProGen is a platform, based on Matlab, which gives as output daily load profiles, which can be given in hours, quarters of hour, minutes or seconds. In order to do this, it requires some inputs to be given. Fortunately, almost all of the needed input, was given as output by the model that was just described.
To start LoadProGen, the output scenarios of the diffusion model need to provide the distribution of appliances across the households. Depending on which types of

appliance a household has adopted along the process, it will be assigned to a category, or household type. The higher the number of categories, the higher the computational load for the software. Each one of the categories will be characterized by a specific number of appliances and for each appliance some data need to be filled in.

In the following figure an example of GUI (Guided User Interface) of LoadProGen is shown.

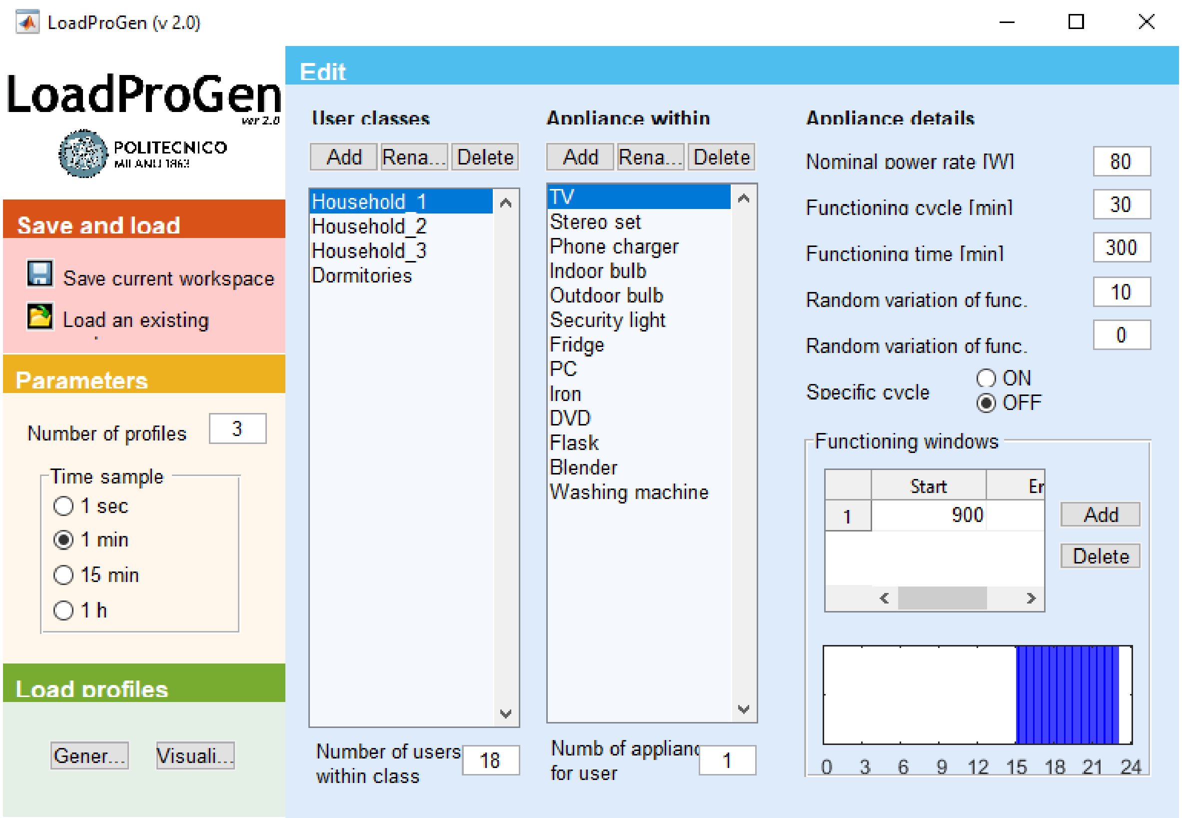


Figure 16: User interface of LoadProGen software, named GUI

In this figure, the general settings are shown and clicking on each household it is possible to see which appliances the members of that category own. Moreover, by clicking on each appliance it is possible to set its parameters, as it is possible to notice on the right hand side of the figure.

Nominal power rate values were taken from the catalogue of DC-Appliances, made by GIZ, called Photovoltaics for Productive Use Applications [77].

The *Functioning cycle* represents the amount of time the appliance goes on working in a row.

Functioning time, instead, sums up the entire working time, considering all cycles.

Random variation parameters are supposed to give a certain variability, meaning that it might happen that something is switched on or off at times it was not expected to happen.
Moreover, a specific cycle can be created for appliances which consume a different amount of power along one same cycle, but this can apply only if the time step is of the order of minutes.
Lastly, the *functioning windows* represent the times of the day at which we expect the appliance to be at work, e.g., a lightbulb will be more likely to be switched on at night time.
So far, in order to guarantee the highest level of variability possible, 100 simulations were performed. When implementing scenarios in LoadProGen it was decided to pick only some of the simulations to obtain the daily load profiles. The total installed capacity of each of the simulations at the end of the lifetime of the grid was evaluated. Then, the scenarios with the greatest, the lowest and the median capacity installed at year 20 were spotted, which will be called respectively MAX, MIN and MED. There were more than one with the same maximum load and more than one with the same lowest and median values. One random case among the others was chosen for each of the three groups of simulations and it was set up in order to be used as input in LoadProGen software.
For each of the selected simulations, it is necessary to take into account the entire lifetime progress, since otherwise it is not possible to have an idea of how the system is supposed to be growing in time.
When sizing a grid, it is possible to choose among several techniques to take care of the evolution of the system in time. Two of these methodologies can be of interest in this case.
The first possibility is to take into account only the short term. In this case a very embryonal prototype of the entire grid will be produced at year 0, where only the short term demand of the first 1 to 5 years will be taken into account in order to build a grid which satisfies it. In this type of projects, there is usually the tendency to make a long term planning of which will be the needed future analysis to be carried out in order to expand the system when and if it will be necessary. This type of approach is of course more precautionary, since it allows to delay the full investment and to be

able to loose less money in case of a less florid outcome than the hoped one. On the other hand, it will imply that at the end of the lifetime the total amount of money spent on the grid sizing will be higher, because this approach will involve the need of more than one research campaigns, one at each expansion of the system.

The second approach, instead, is definitely riskier and implies the use of a model or group of models like the ones proposed by the current study, which allow to forecast the demand along the entire lifetime of the grid to be built. In this case a larger project will be created and a larger capacity with respect to the one needed at year 0 will be built, in view of the future expected adoptions. This second approach is the one that will be assumed in the current situation.

In order to make a long term forecast of the daily load profiles of the considered fictitious village in Njombe, it was decided to consider 3 separate years for each of the selected simulations. The considered time-steps would be year 1, year 10, that is half of the lifetime, and year 20, that is the last year of the system. For each of the simulations, three Microsoft Excel files were built. In each file, the data related to one year were included. Each page contained the data related to a certain household category, but page 0, which contained general data useful for LoadProGen settings. In the following figures an example of page 0 and of one of the other pages is shown.

	A	B	C	D
1	Parameters for the generation			
2				
3	Number of load profiles	250		
4	Number of time step	1440		
5	Number of user classes	5		
6	Maximum number of appliance	5		
7	Maximum number of windows	3		
8				
9				
10	Power profiles of specific appliances			
11				

Figure 17: page 0 example

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	**Specific user class (j)**	**Number of users in class (N_j)**	**Type of electrical applianc e (i)**	**Nominal applianc e power rate [W] (P_{ij})**	**Number of applianc es in class (n_{ij})**	**functioni ng cycle [min] (d_{ij})**	**functioni ng time [min] (h_{ij})**	**Rh_{ij} [%] *[0-100]***	**Rw_{ij} [%] *[0-100]***	**Starting time Win 1 [min] *[1-1440]***	**Ending time Win 1 [min] *[1-1440]***	**Starting time Win 2 [min] *[1-1440]***	**Ending time Win 2 [min] *[1-1440]***	**Starting time Win 3 [min] *[1-1440]***	**Ending time Win 3 [min] *[1-1440]***
2	Household_6	2	fan	20	1	60	180	10	0	660	1260	0	0	0	0
3			food stor:	200	1	1440	1440	10	0	1	1440	0	0	0	0
4			lightbulb	24	1	120	360	10	0	1	360	1080	1440	0	0
5															

Figure 18: generic page of a household category example

As it is possible to see from Figure 17, page 0 contains general information which would give an idea of the needed computational effort that the procedure will require. The number of load profiles represents the number of simulations that have to be run, the time steps, instead, can be either 1440 (minutes) or 24 (hours), or others depending on which is the smallest time unit needed. The number of user classes will represent the number of pages apart from page 0 that will need to be filled in, one for each household category. Maximum number of appliances and maximum number of windows depend on how many appliances are adopted by the household who adopts most and how many time windows the appliance with the most complicated features has. Eventually, a section called "Power profiles of specific appliances" is necessary in order to specify possible power cycles of certain appliances.

Figure 18, on the other hand, is an example of a page relative to a household category. The different boxes are the same that were already explained looking at the GUI of the software and when the Excel file is loaded onto Matlab, these data will serve the exact same purposes.

Once the 9 chosen scenarios were built, 3 for each selected simulation, it was possible to run LoadProGen. The daily load profiles that can be obtained thanks to it are shown in the picture below, where the units are Watts on the ordinates and hours of the day on the abscissas.

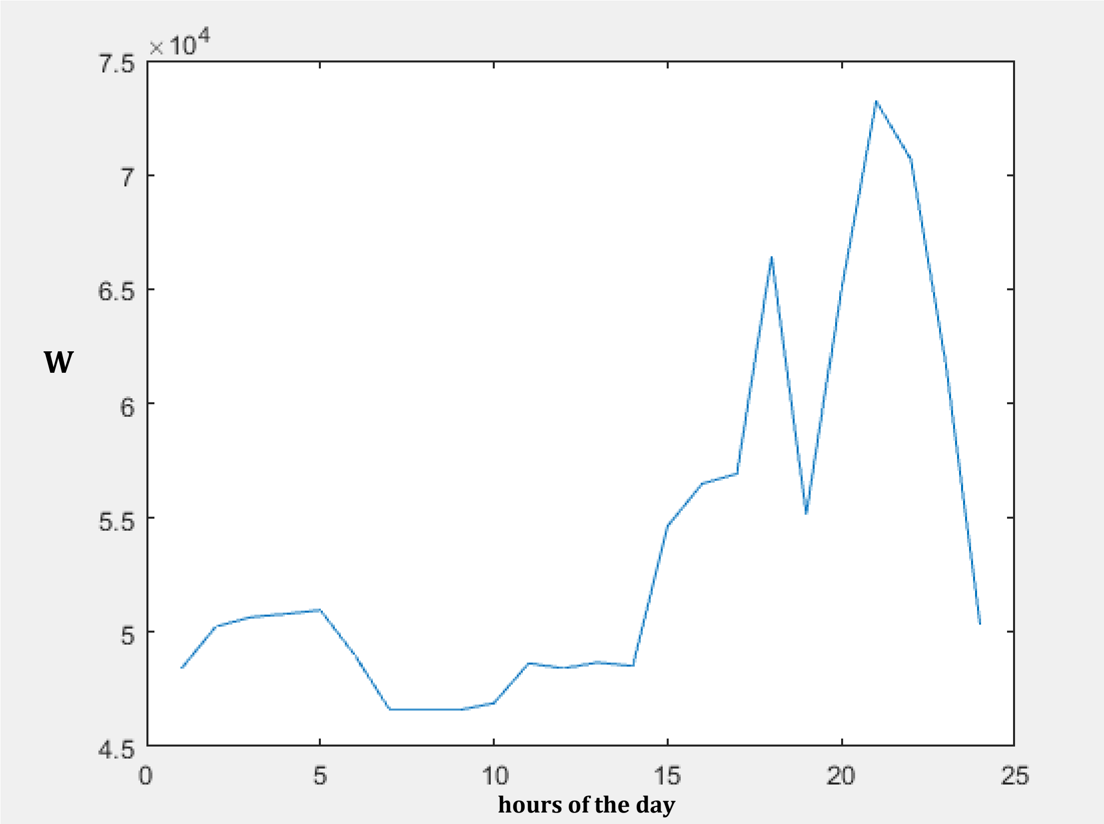


Figure 19: example of LoadProGen output for MIN scenario, year 20

Given these many load profiles like the one above, there is still one more step to perform before obtaining the sizing of a grid.

3.5 Off-grid System Sizing

HOMER Pro by HOMER Energy is an optimization software for microgrid design, originally developed by NREL (National Renewable Energy Laboratory). HOMER stands for Hybrid Optimization (Model) for Multiple Energy Resources and it attempts to simulate a viable system for all possible combinations of the equipment that the user wishes to consider.

The time frame of the software is one year, but it is possible to provide different load forecasts for each day of the year, with a time step of one minute or one hour. For each proposed solution a set of economic parameters is also provided, which allow the user to assess the economic feasibility of the different options.

Another useful tool is the Sensitivity Analysis option, which allows to run different simulations of the same system changing the value of some parameters and comparing the results. It finally allows to have access to databases with resources forecasts for wind speed, sun radiation, temperature, fuel costs, etc., allowing the user to either choose a reference or to upload his own.

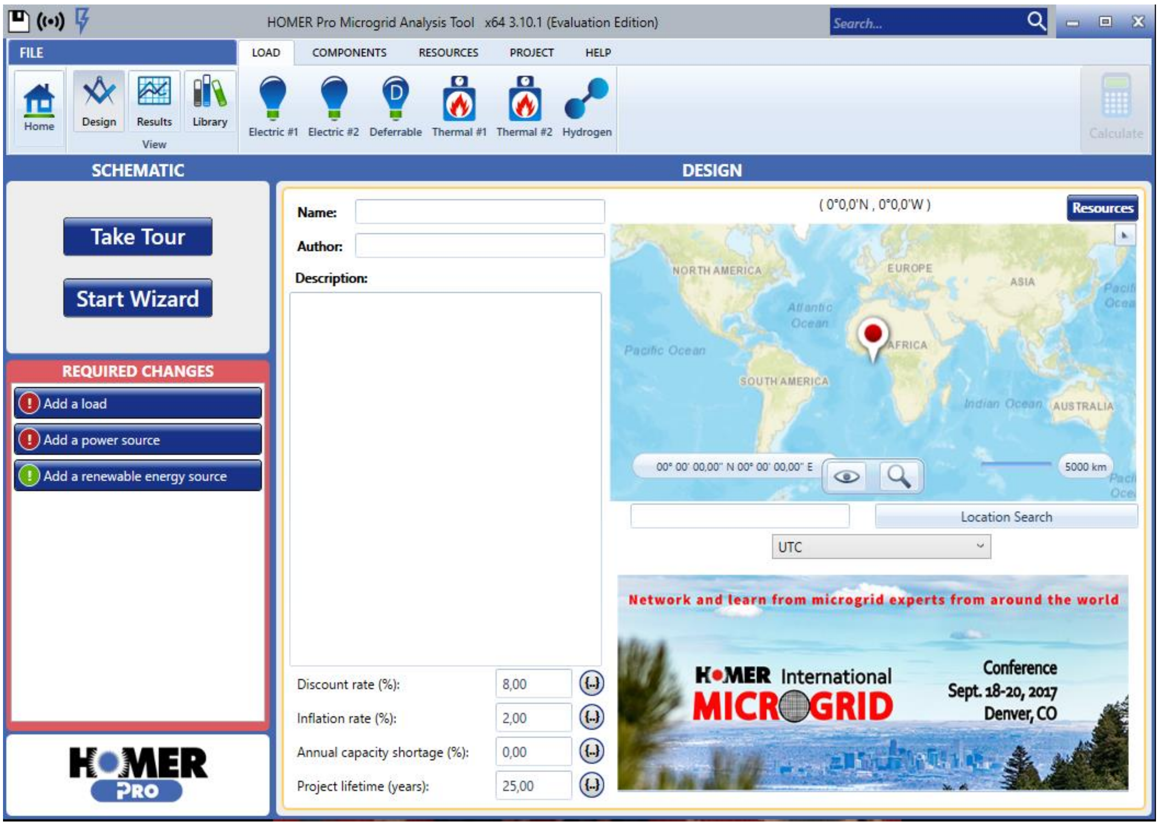


Figure 20: HOMER home page

In Figure 20, it is possible to see the home page of the software that allows for the choice of a name of the project and also asks for a location where this microgrid should be built in order for the engine to be able to look for the data related to the resources in the area. On the left hand side, below “required changes” the software collects possible suggestions to improve the on-going project and clicking on those tabs the correspondent page will open.

Once the location has been set, it is necessary to give as input a certain type of load. It will be possible either to choose one from those provided by the software as a standard, or to upload a load the user has already from his own surveys, or in this case from LoadProGen output.

Figure 21: HOMER load input page

As a second step, the load settings must be filled in. Among the required data, there will be two types of random variability. Day-to-day random variability represents the variability given by the difference in total consumption among the days of the year. It was assumed that all days would have the same load curve, but to be more precise it is appropriate to assign a certain variability when necessary. The time step variability, instead, represents the differences that might be present at the same hours in different days.

Starting from LoadProGen output, which provided 250 possible load curves for the same day, it was possible to calculate these two values in the following manner.

Time step variability

- The average value *A* of load among the 250 available at each hour of the day was obtained.
- The standard deviation *SD* was evaluated for each of the 24 distributions and the covariate was thus obtained:

$$COV = \frac{SD}{A} \qquad (22)$$

- The mean value of all covariates was considered to be the time step percent variability to be used as input in HOMER.

Day-to-day variability

- The total daily power consumption for each of the 250 scenarios was calculated
- The average and standard deviation of the results were obtained
- The covariate (equation 22) of the distribution of total consumptions was obtained and this was used as input for the day-to-day variability in HOMER.

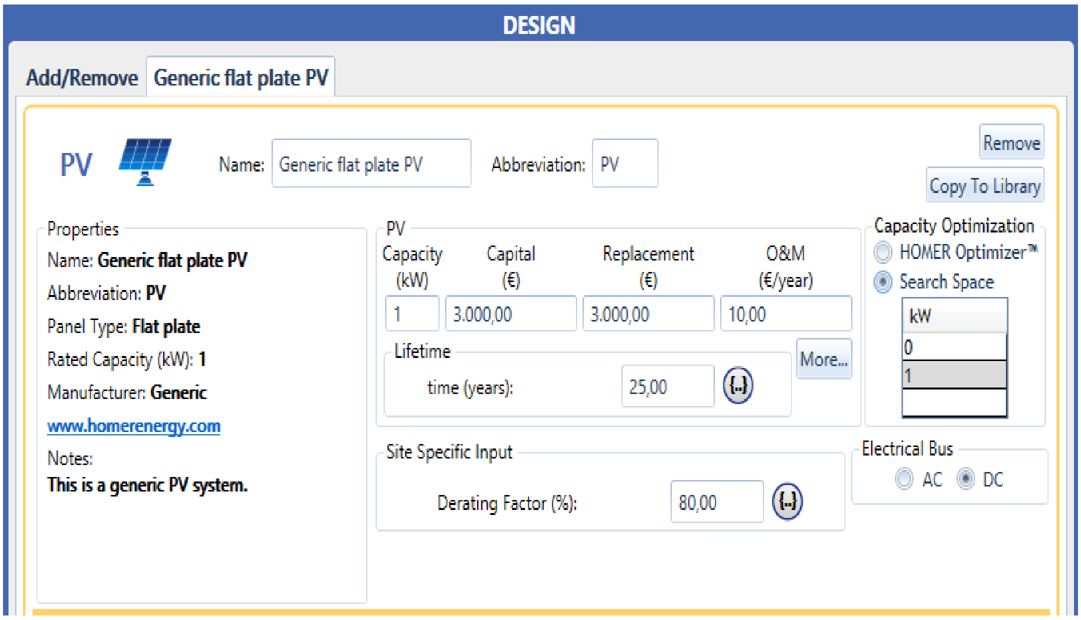


Figure 22: HOMER component data input page

Once the load is properly designed, it is necessary to choose which will be the power generators allowing to respond the demand of electricity. In Figure 22, an example of input page for a solar system is presented. In this case a generic flat plate PV was chosen, but many others are available in HOMER library. Default data are provided for 1 kW of capacity and on the right hand side it is possible to define the sizes in terms of total capacity that the system is expected to need. The simulations will run using all of the options provided by the user and the software will give as an output the optimal solution in terms of demand satisfaction and costs. This type of data input was done for solar power, wind turbines, batteries for storage and a converter to switch from AC to DC and the other way around. By selecting "HOMER Optimizer" option, the software will automatically optimize the amount of capacity needed.

Once all of the input data is completed, it is possible to click on "calculate" and if everything is consistent the software will produce several possible solutions for grid sizing as an output. Given those, it will be the user who will make his own evaluations based on his needs and requirements. It might be that a certain differentiation of resources is preferred, even if it might increase the investment costs, or otherwise it might be preferable to spend the least money possible. These and other considerations depend on the surrounding settings and change with circumstances.
A clear idea of what HOMER algorithm does is given by the following diagram.

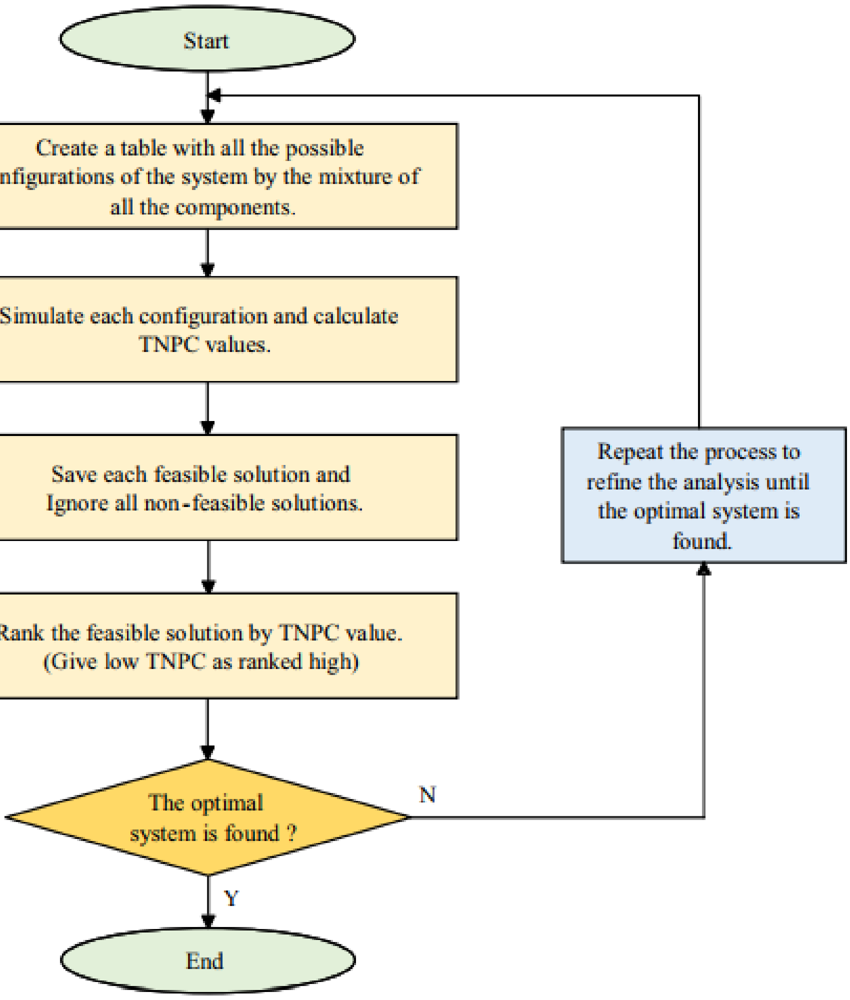


Figure 23: scheme of HOMER Pro algorithm

Once the overall mechanism is clear it is possible to go through which were the specific characteristics of my case study.

Wind and solar power were chosen to be the generation resources. To be able to make reasonable forecasts, it was necessary to have some data relative to the temperature, solar irradiance and wind speed of the area of Njombe. All of these were available from the library of HOMER and it was decided to use the same reference for all the data, that is NASA "Surface meteorology and Solar Energy" work of 2011 [14]. From this work it was possible to obtain three parameter behaviours: global horizontal radiation monthly averaged values over a 22 years period from July 1983 to June 2005, from which it was possible to obtain the monthly average solar Global Horizontal Irradiance (GHI) data; the air temperature monthly averaged values over a 22 years period from July 1983 to June 2005; the wind speed at 50 m above the surface of the earth for terrain similar to airports monthly averaged values over a 10 years period from July 1983 to June 1993.

Given this data, it was possible to decide which type of generators to take into account. The choice made, was to keep the simplest components available in the software and check for their investment costs and lifetimes referring to literature.

The wind turbine was a generic 3 kW turbine, characterized by an investment cost of 4000 €/kW of capacity [15] and by a 20 years lifetime, given that it was assumed that no component would have a longer lifetime than the whole system itself. Operation and maintenance (O&M) costs were assumed to be 120 €/year, because they were supposed to decline proportionally to how much the investment costs had decreased with respect to default values of HOMER library. The behaviour of wind speed in the area and the relationship between wind speed and power output for the current component are shown in the diagrams below.

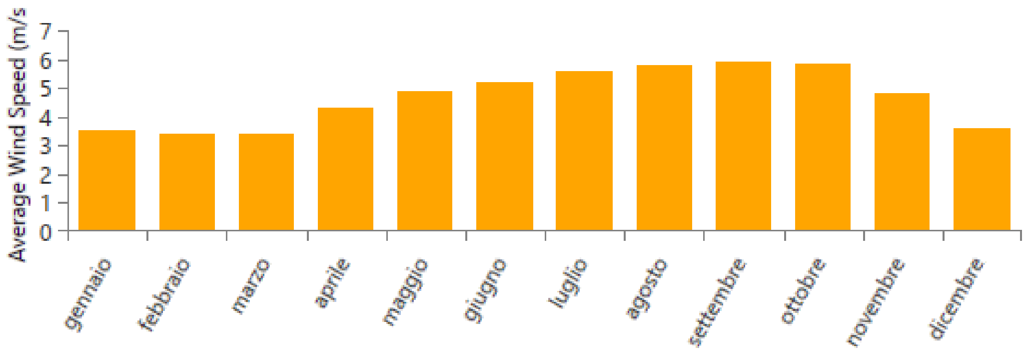


Figure 24: monthly average wind speed data

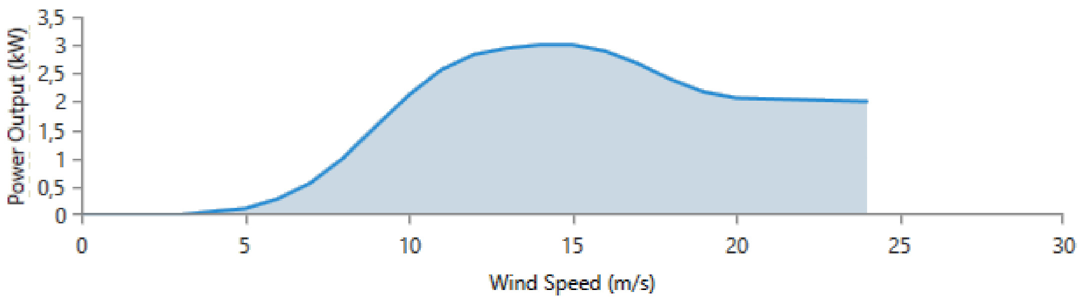


Figure 25: power output relationship with wind speed for a 3 kW wind turbine

What happens in HOMER Pro is that one gives as an input the type of components that can be installed, then selects the “Homer Optimizer” option before running the simulations and the software will decide how many pieces of each technology are needed in order to respond the forecasted load demand.

Taking into account solar technologies, instead, the chosen component was a generic flat PV with a rated capacity of 1 kW, which investment cost was 2000€/kW [17], while O&M costs would only be of about 10€/year for a lifetime of 20 years. In the figures below, the curve of GHI and the temperature behaviour in the area are shown.

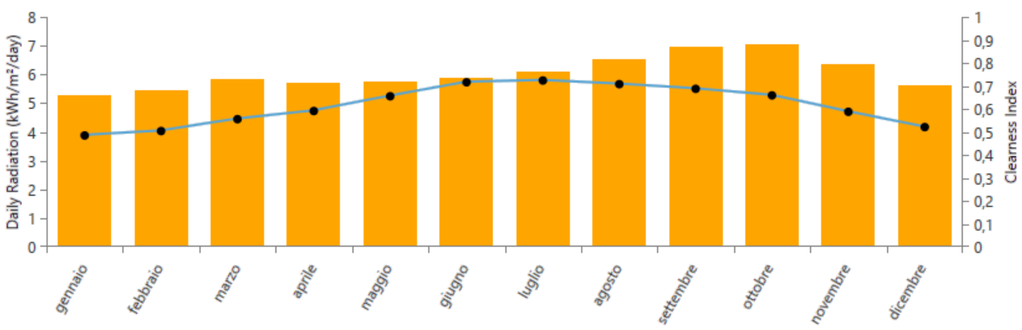


Figure 26: monthly average solar Global Horizontal Irradiance data

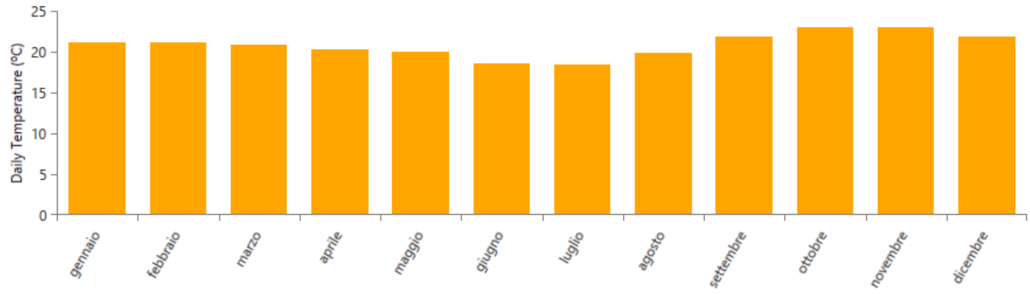


Figure 27: monthly average temperature data

In order to better understand the data above, it is possible to notice some characteristics which can be very important. In Figure 24 it is possible to see how the level of wind speed along the year is not really high. In places like northern Scotland, where wind is usually present constantly along the year, wind speeds of ~11 m/s are usually reached. In Njombe region, it is possible to observe a level of wind speed of ~5 m/s as a maximum, which though keeps practically constant between 4 and 5 m/s for the entire year. Looking at Figure 26, instead, we can notice how radiation and clearness index keep a very constant behaviour along the year, which is confirmed by the level of temperature, which keeps between 15 and 25°C for almost the entire year. If one compares the radiation behaviour in Njombe region with what happens, for example, in the south of Italy, it is possible to see how in this second location radiation keeps between 2 and 8 kWh/m^2/day, while in Njombe the reached level is never as high, but it never goes below 5 kWh/m^2/day for the entire duration of the year.

It is then possible to analyse the other two components of the grid.

As a storage facility, a very basic type of battery was chosen, that is a small 1 kWh generic lead acid battery with a nominal capacity of 1 kWh and a nominal voltage of 12 V, it would have an investment cost of 300 € and O&M costs of 10 €/year. It would last for about 10 years, after which a second battery would need to be purchased to allow the grid to continue working [16].

The last needed component in order to make this grid working and realistic is the converter, which is a generic system with a 15 years lifetime, an investment cost of 300€ and no O&M costs for the lifetime period. Its efficiency is assumed to be 95%. Once all of the components have been selected, it is necessary to understand how many of each should be installed. To make this evaluation, it is possible to take into account the load forecasts, plus there is the chance to set a certain level of efficiency of the grid which tells which is the acceptable percentage of not responded demand every year. First, the average daily load curves of the various scenarios were taken into account, then the maximum annual capacity shortage was set to be equal to 10%, since it is typical for a developing country to have some shortages during the day. Now that the main characteristics of the grid are set, it is possible to go through which where the main results of this study.

4 Results and Discussion

4.1 The Network

The whole research was started using as a reference real data from Bulongwa village, which allowed to have a realistic basis in terms of network type and average degree on which to build the rest of the process. Once the basis was set, a new fictitious village of 400 households was created, through a Matlab script, based on the previously shown formula for BA networks. This village would have its characteristics and details, including the expenditure levels of the inhabitants. The obtained network structure is shown in the figure below.

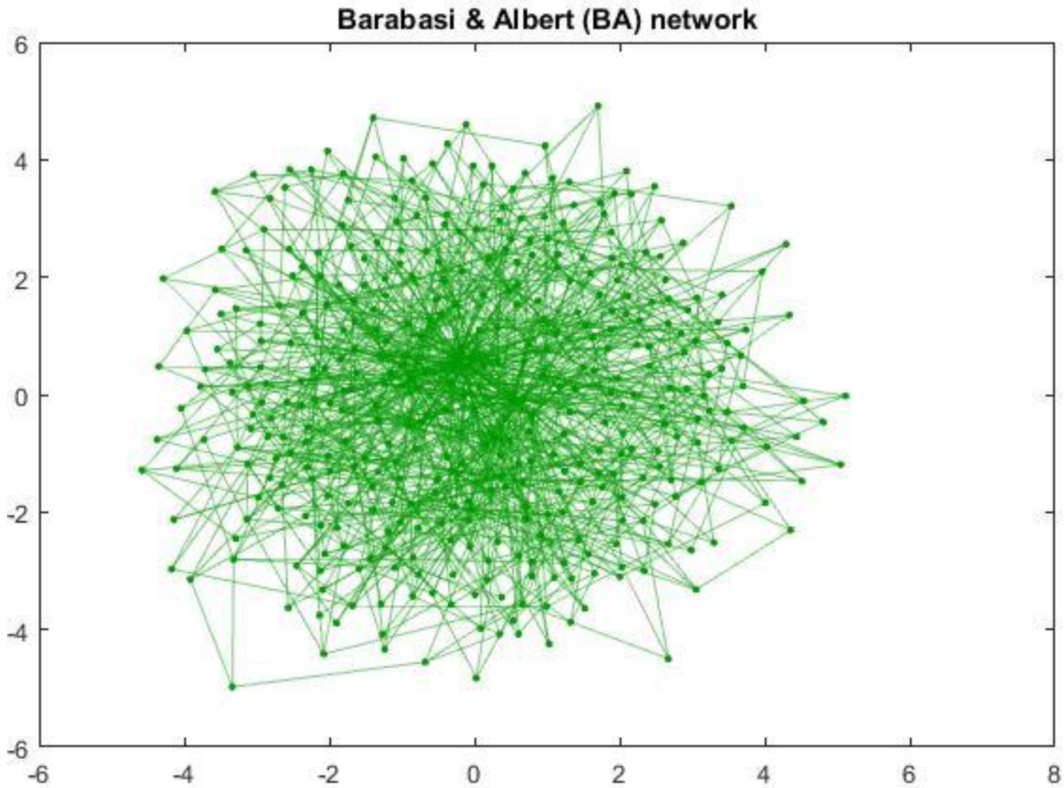


Figure 28: Network structure

From the image it is possible to recognize the main features of a Barabasi-Albert network. It is based on the concept of preferential attachment, so that it is easy to understand which are the "older" nodes and which are the new ones, based on the number of links that they have. Several assumptions can be made about how the links might happen in reality. Either by geographical proximity, or by age, or race, religion and other personal traits of each individual, but in this case it was decided to rank the households in terms of degree and richness. The ones with a higher degree (number of contacts) were assumed to be the richest. Of course this type of assumption would not be necessary in reality, because one would assess the network shape and the expenditures of the single nodes with a survey.

To better understand which type of network I will work on, it is possible to evaluate some characteristic parameters.

- *Eigenvector centrality* is a measure of the influence of a node in a network. The eigenvector centrality of a node is proportional to the sum of the centralities of the nodes it is in contact with. The average eigenvector centrality of this network is 0.0025, which is pretty low, considering that its maximal value could have been 1.
- *Closeness centrality* is inversely proportional to the sum of the length of the shortest paths between the considered node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes. The average value for this network was 0.000808, which suggests that the network might have some nodes who are quite far from the rest, because the overall value keeps very low.
- *Betweenness centrality* is based on the rule that for every pair of vertices in a connected graph, there exists at least one shortest path between the vertices. The betweenness centrality for each node is the number of these shortest paths in the entire network that passes through the node itself. The average value of it in the current network was 425.625, which, in accordance to closeness value, suggests that many paths pass through each node, as if to go from a node to the other a long distance should be covered, making it more likely to pass through many nodes several times.

4.2 Electricity Connection Curves

Once the network structure is given, it was next possible to build a Bass model on Matlab, based on a Monte Carlo algorithm, to let p and q values vary in the ranges found in literature. From this second phase of the research, it was possible to obtain the grid connection diffusion curves shown below.

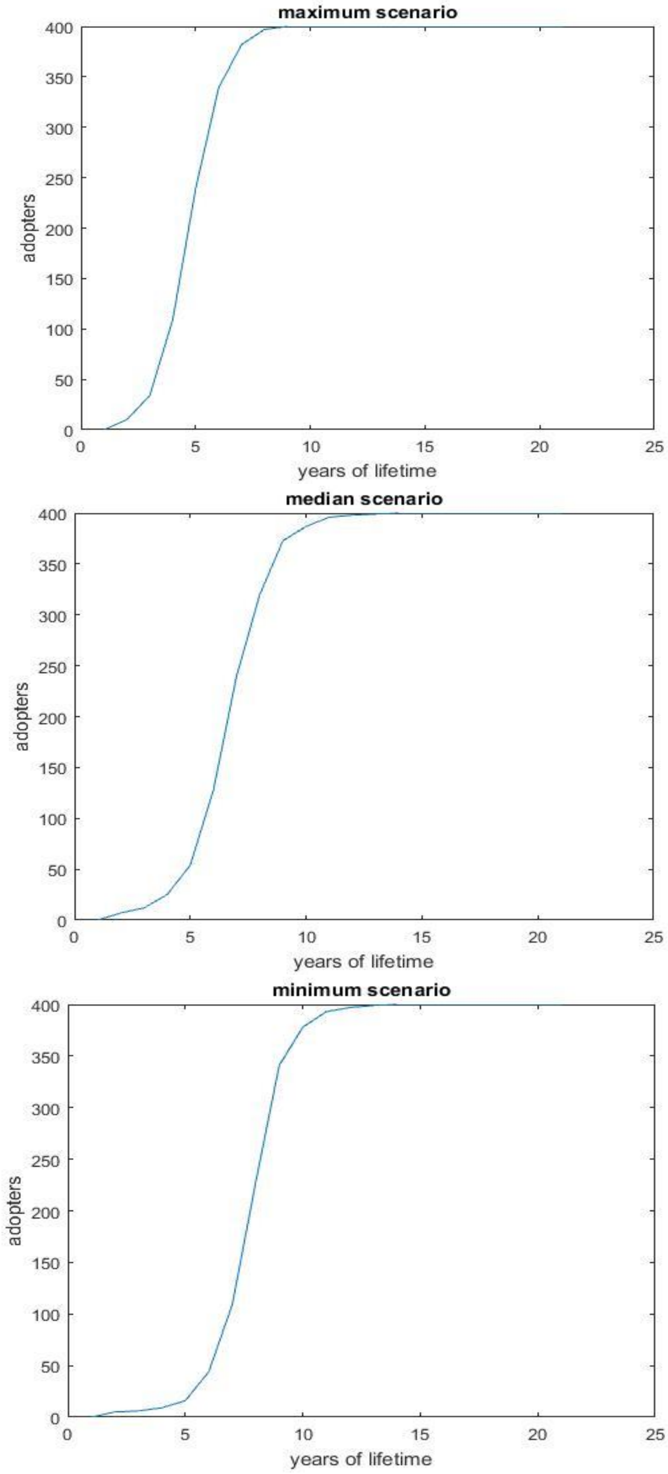


Figure 29: grid connection diffusion curves for MAX, MED and MIN scenarios

The selected scenarios were chosen on the basis of the installed capacity at the household level at year 20. The maximum, the minimum and the median values of capacity were extrapolated from next paragraph results and the three correspondent simulations chosen. From here on, for sake of simplicity, I will call the scenarios MAX, MIN and MED. It is necessary to keep in mind that the model is built in a way in which at each time step, as this diffusion process upgrades and goes on, also the diffusion process of appliances goes on. The diffusion of appliances is strongly dependent on the diffusion of electricity, since anyone who does not connect to the microgrid will not be able to purchase nor use any appliances. Therefore, every scenario, not only will have different diffusion curves of appliances, but also different grid connection diffusion curves, since they will be the output of different input parameters, which will be further analysed in the next paragraph.

4.3 Appliance Diffusion Curves

Starting with the first intermediate outcome, it is possible to take into account one appliance as a first example to see what the Monte Carlo method produced in terms of differences among the simulations. If we consider the 100 diffusion curves of the fan, what we observe is the following diagram.

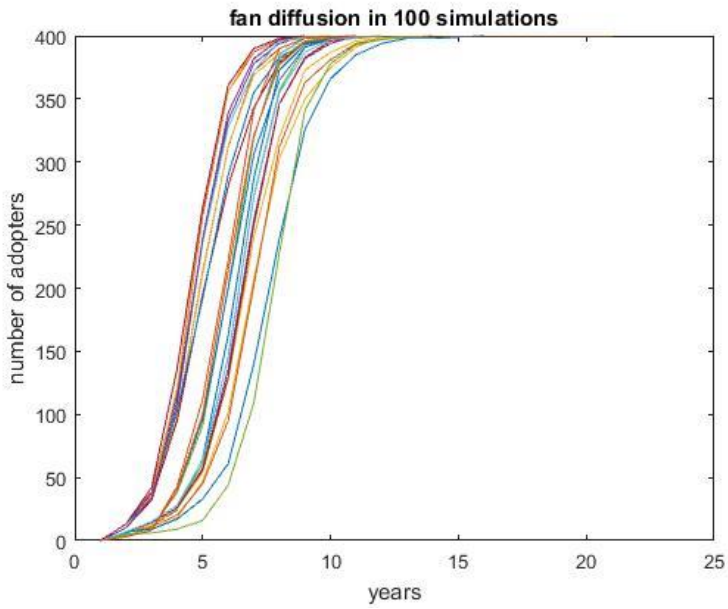


Figure 30: 100 simulations of fan diffusion curve through Monte Carlo method

Looking at this picture, it is evident that depending on the values which we attribute to the endogenous parameters (viz. p, q, β and γ) , the final output will change much. By zooming on a detail at year 5, for example, it is possible to find out that the minimum option involves ~10 adoptions at year 5, while taking into account the highest curve the adoptions would be already above 250. This situation is better shown in the following figure.

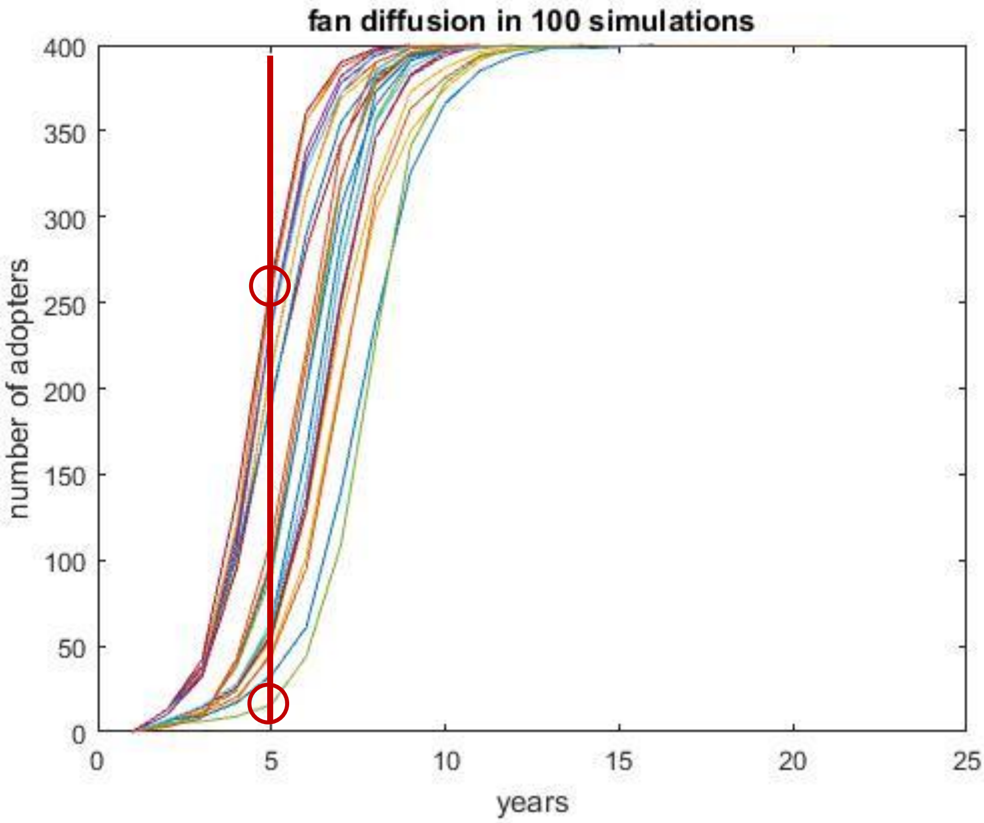


Figure 31: detail of year 5 of fan diffusion curves

Given this situation, it is easy to understand why I directly analysed only three simulations, that are then going to be used for the entire study.

The diagrams below show the S-shaped curves of diffusion of all the considered appliances across the village in the three different cases.

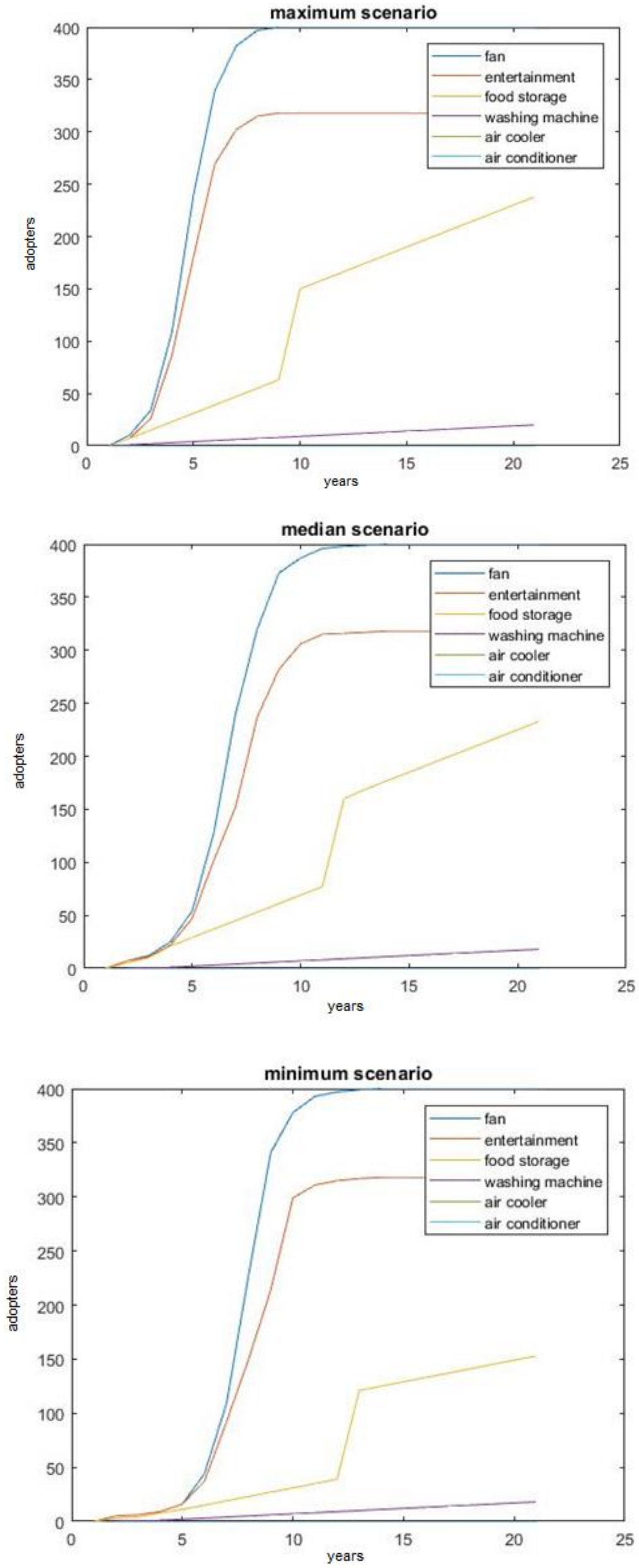


Figure 32: diffusion curves of appliances in the 3 scenarios MAX, MED and MIN

As it is possible to see from the diagrams, there are some patterns that repeat themselves in all cases. First, in the three histories, it is only four appliances out of six that in 20 years actually get to be adopted. It is important to remember the assumption of updating expenditures made in the beginning. Every year the households re-start from a certain amount of expenditure assumed to be constant over the years. This is usually not very realistic, since it is to be expected that households' conditions might improve in the years, maybe thanks to the adoption of electricity itself, but it is a good approximation for such a tiny reality. It is also due to this assumption that some appliances will never get to be adopted, since their price is larger than the 80% of total expenditure (due to affordability constraints) of each and every household.

The appliances that instead get to the market are five. First of all, lightbulbs, which are not present in the figures because as soon as one adopts electricity it is assumed that as a consequence he will also install a lightbulb, so that they were not included in the diffusion process based on Gompertz curves. Therefore, electricity and lightbulbs diffusion curves (see Figure 29) will be the same.

Fans, which are the least expensive technology, are widely adopted and everyone who adopts electricity gets to have one at the end of the process in any of the scenarios. It was also noticed that by relaxing the constraint of "uniqueness of adoption" (each household is assumed to adopt only one piece of each appliance) it happened that some households would adopt more than one fan if they had enough expenditure left to use, which did not happen with any other more expensive component. The third appliance in order of amount of adoptions is the entertainment facility, for which the data of television were taken into account, yet it would have been the same to consider radios. In this case it is not the entire amount of electricity adopters who purchases an entertainment facility, but in all of the three cases a similar amount of people gets to buy a TV.

Independently from which appliance we focus on, the difference among the three scenarios, is of course the velocity at which the market saturation is reached. To understand what is that makes the three scenarios different it can be useful to analyse which were the values of the parameters involved in the three cases. Indeed, before getting to the diffusion of appliances, at each time step a certain number of household will connect to the grid. The number of connected households will vary depending on the input parameter chosen through Monte Carlo algorithm and will strongly affect who will have the chance to purchase electrical appliances in the second part of the simulations.

Scenario	p	q
Maximum	0,0091	0,2425
Minimum	0,0035	0,2389
Median	0,0070	0,2013

Table 4: values of the parameters p and q of the three simulations

Given Table 4, it is possible to notice that the values of p actually follow the order of maximum, median and minimum. This indeed is reflected in the previous diagrams. For example, focusing on year 5, it is possible to see how in MAX scenario with respect to the other two (especially MIN one) the number of adopters of all technologies is already much higher. Looking for instance at the fan curves, it is possible to see that in MAX scenario at year 5 a number of adopters close to 150 is already present, which decreases to approximately 50 in the median case and falls to less than 50 in the minimum installed capacity scenario. The same reasoning can be done for all of the appliances. Independently from the numbers of adopters for which each S-curve will become flat, in the maximum installed capacity scenario the curve will grow faster than in the median and the minimum capacity ones.

Other parameters to be taken into account are the characteristic values of β and γ of each appliance which as well change at every new simulation. Their variability influences the level of ownership of appliances and therefore the total number of adopters at each time step. In general, it is possible to state that the higher the value of β, the lower the number of adopters of a certain technology, because it represents the value of a negative exponential. The contrary is instead true when considering γ. Their values for the current study are shown in the table below.

Scenarios	Fan		Entertainment		Food storage		Washing machine		Air cooler		Air conditioner	
Parameters	β	γ	β	γ	β	γ	β	γ	β	γ	β	γ
Maximum	1,1715	3,7315	1,3259	2,7400	4,8497	2,2014	8,3281	1,7102	4,6351	2,0202	11,1780	1,2539
Minimum	1,2097	3,6095	1,4041	2,2808	6,0771	2,7089	7,0235	1,9621	4,0476	1,9732	11,1056	1,2995
Median	1,2627	3,5601	1,7346	2,1016	4,6848	2,3717	7,9334	1,3973	3,4175	2,0310	10,8716	1,2770

Table 5: appliance parameter obtained through Monte Carlo method for the chosen scenarios

A last question that one might ask when looking at the diffusion curves of appliances is what happens to the curve of food storage when it suddenly jumps up one year to the other. The answer to this can be found thinking about how the model works. Along the code, at every time step, households adopt the cheapest technology they do not possess yet as first, then, if some expenditure share is left they can buy other things. Food storage facilities are too expensive to be bought as a second purchase, therefore, in order for people to be able to adopt a refrigerator the market of fans and entertainment facilities had to be saturated first. In each of the diagrams, therefore, the year in which food storage adoption shoots up corresponds to the following year with respect to the flattening of the other two upper curves.

Lastly, looking at the curves of washing machine in the different cases, it is possible to notice that they are not S-shaped yet. This because 20 years result not being enough for this technology to actually spread around. The adopters who appear already are early ones, or innovators, but more time is needed for the market to get to saturation.

One of the best qualities of this model is that it does not only provide the diffusion curves as they are shown above, but adds qualitative information besides the quantitative knowledge of “how many” appliances were purchased. At the end of the process it is possible to know “who” purchased which appliances as well, which is a fundamental piece of knowledge for the next steps of the research.

4.4 Daily Load Profiles

In order to obtain the daily load profiles of the fictitious off-grid system that was meant to be sized, the software LoadProGen was used. As explained in the previous section, this software takes information about the single households as an input, to then give the requested number of load profiles as an output (250 at each simulation in this case).

From the previous steps, we adopted the MIN, MED and MAX scenarios of appliances diffusion to build the load profiles. For each of these it was decided to take into account three different years: year 1, year 10 and year 20. Doing so, it was possible to obtain the "evolution" of the three scenarios in time.

LoadProGen was run nine times (viz. year 1, year 10 and year 20 for each MIN, MED and MAX scenarios) and each time it produced 250 possible daily load profiles, which were saved inside a 250x24 matrix (24 are the hours of a day). The output that was possible to obtain is shown in Figure 34 for MAX scenario, year 20.

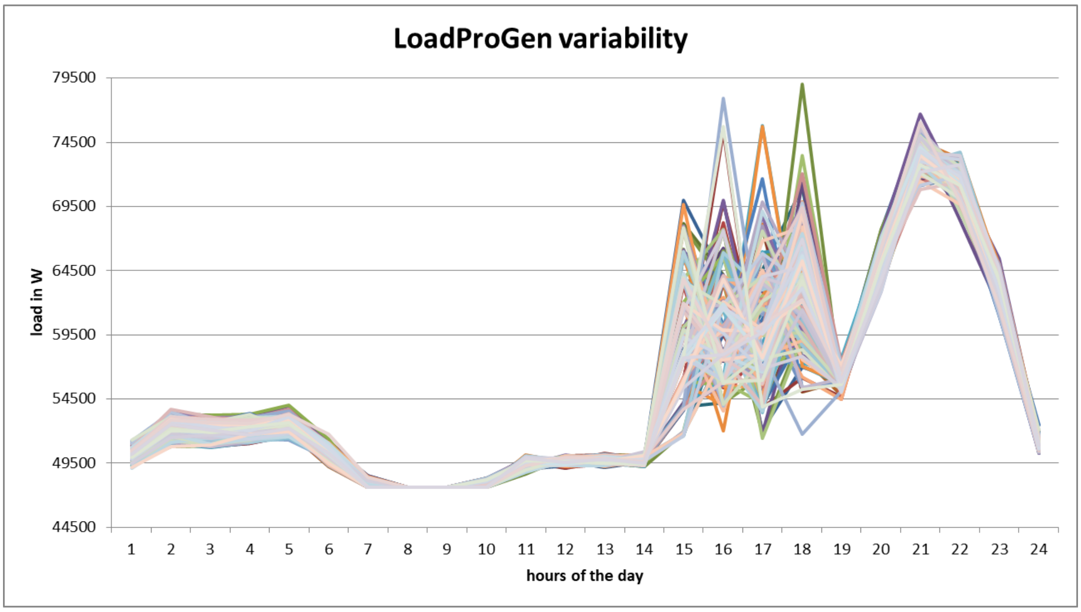


Figure 34: LoadProGen variability for MAX scenario, year 20

Thanks to this, it was possible to evaluate an average daily load profile among the 250 available and use it as an input for the sizing software. The resulting average curves for each of the 9 cases are shown in the following figures, for the three selected years of each chosen scenario.

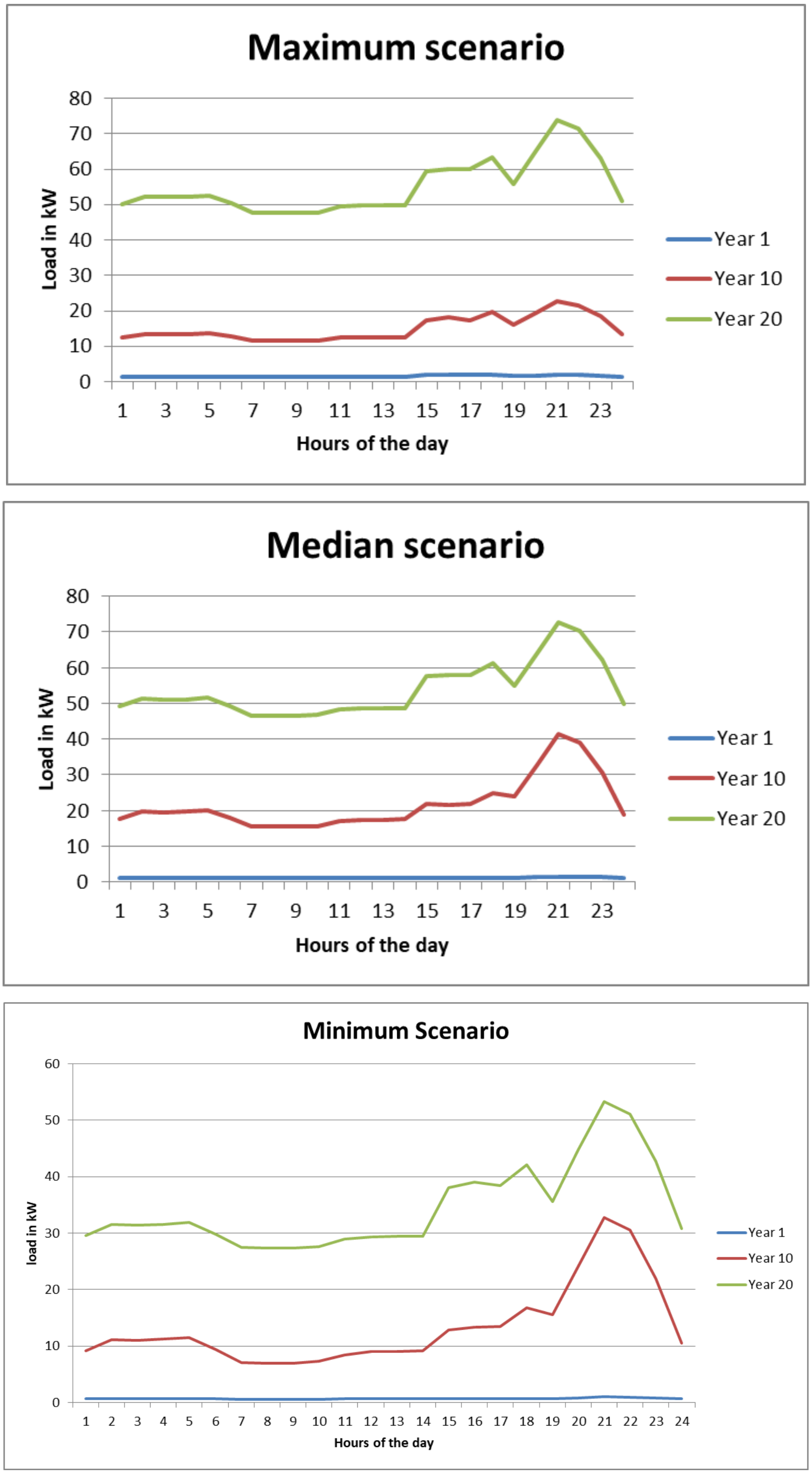


Figure 35: LoadProGen average output load curves for the 3 scenarios MAX, MED and MIN

Looking at these diagrams, it is possible to analyse maximum and minimum capacity scenarios first. While in MAX scenario a peak load of ~75 kW is reached during an average day of year 20, in MIN one the peak stops at ~55 kW. Going back in time it is noticeable that the situation reverts. The scenarios I decided to analyse were chosen based on the situation at year 20, which means that at year 10 and year 1 the levels of adoption might be different. If we consider year 10 configuration, its load curve is higher in MIN scenario than in MAX one. This suggests three possibilities, which could bring MIN scenario above MAX one at year 10:

- the number of appliances at year 10 in MIN case is larger than in MAX one, then at year 20 the situation reverts, possibly because MAX scenario has stronger word of mouth effects in the long term
- the number of highly power demanding appliances in the MIN case is larger with respect to what happens in MAX scenario in the same period, then the situation reverts at year 20
- the variability (among the 250 simulations) of MIN scenario is larger, therefore this curve is not as representative of what really happens as the curve for MAX case

By looking at Table 4 and at the diagrams reported in the previous chapters, it is possible to state that, even if q parameter is slightly larger for MAX scenario than for MIN one, the third option is the most likely and in the following tables the variabilities of the three scenarios are shown to support this theory.

Minimum Scenario	Year 1	Year 10	Year 20
Day-to-day	0	1	0
Time-step	5	8	3

Table 6: random variability for minimum installed capacity scenario

Maximum Scenario	Year 1	Year 10	Year 20
Day-to-day	0	0	0
Time-step	10	5	2

Table 7: random variability for maximum installed capacity scenario

Median Scenario	Year 1	Year 10	Year 20
Day-to-day	0	1	0
Time-step	3	5	2

Table 8: random variability for median installed capacity scenario

To understand what these variabilities represent, it is possible to take into account the formulas obtained in Materials and Methods chapter for COV parameter (equation 22), or to look at the example in Figure 34. These values were also used as HOMER Pro input in the next phase of the research.

Looking at year 1, lastly, it is evident that this time horizon is not relevant to the sizing of the grid, since it only presents very small numbers of adoptions.

If the MED scenario is eventually taken into account, it is possible to see how its year 20 curve is more similar to the year 20 curve of the MAX scenario rather than to the one of the MIN case. This suggests that the 100 simulations made were more similar to the maximum case rather than to the minimum one, given the median diagram situation.

Moreover, once again, the variability issue gives as a result a median year 10 with a greater electricity demand than the maximum scenario one. Yet, taking into account the MIN and the MED scenarios their behaviours at year 10 go back to what one would have expected since the beginning, so that MIN has a lower load curve than MED.

Once the comparison is done, it can be interesting to verify whether the shape of these load curves is realistic or not. In modern western society, daily load curves for residential households present two peaks during the day: a smaller one in the morning, when people wake up and get ready for the daily routine, and a higher one during the evening when people come back from work, cook and switch on some appliances, such as TV or radio to spend their free time at home. An example of this can be seen in the figure below, taken from Terna statistics for the day 29/09/2017 in Italy.

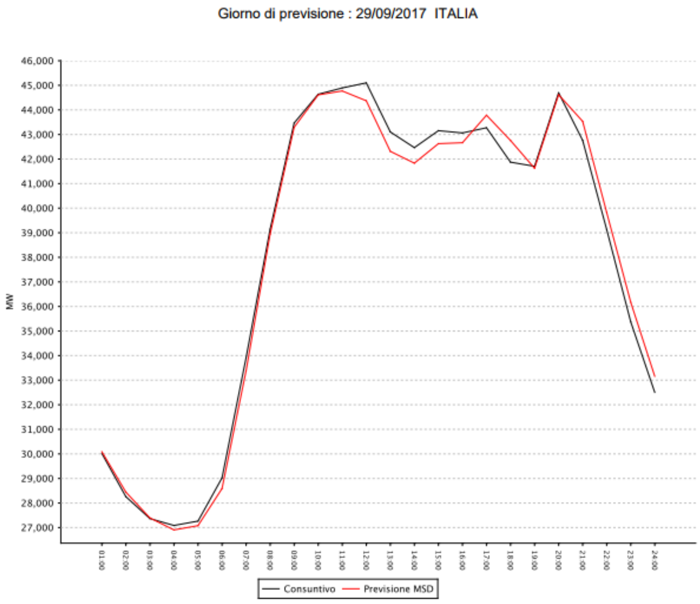


Figure 37: example of Italian daily load profile

In a rural country though, where the available facilities are assumed to be only the ones considered in the study, it is easy to understand why the morning peak seems not to be present. People are expected to wake up early in the morning, get ready and eat breakfast and then go to work. The appliances they might have inside the house include: lightbulbs, which will only be switched on at dark times; entertaining facilities, which tend to be used during free time; food storage facilities, which represent the base load for the grid, since these cannot be switched off if the purpose is to preserve food; washing machines, which might be used in the evening when people go back home and might be willing to wash something for the following days. The other appliances were not purchased by anyone in this study and might actually represent a source of variability in the obtained load curves. Given these assumptions, it is interesting to see what other studies have obtained as real load profiles for rural countries. Some examples are shown in the following images.

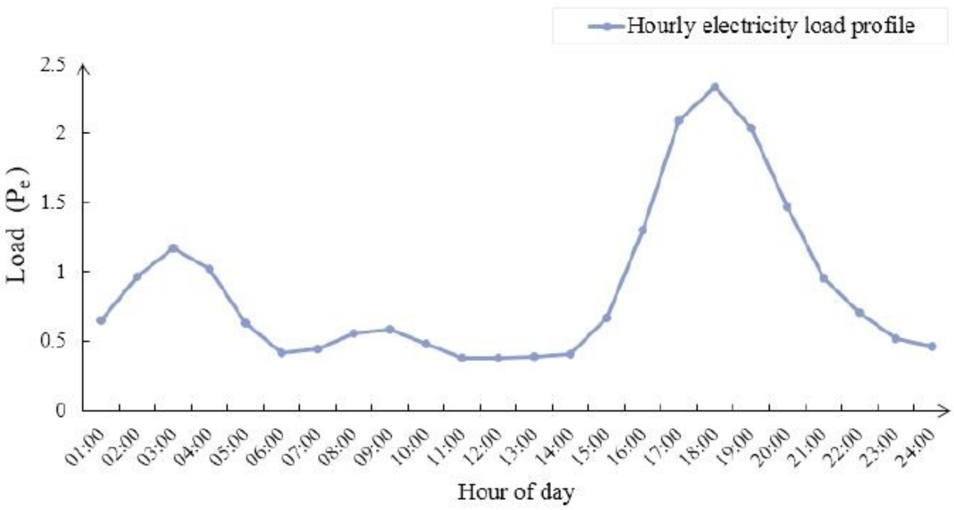


Figure 38: potential rural electricity load profile reference [10]

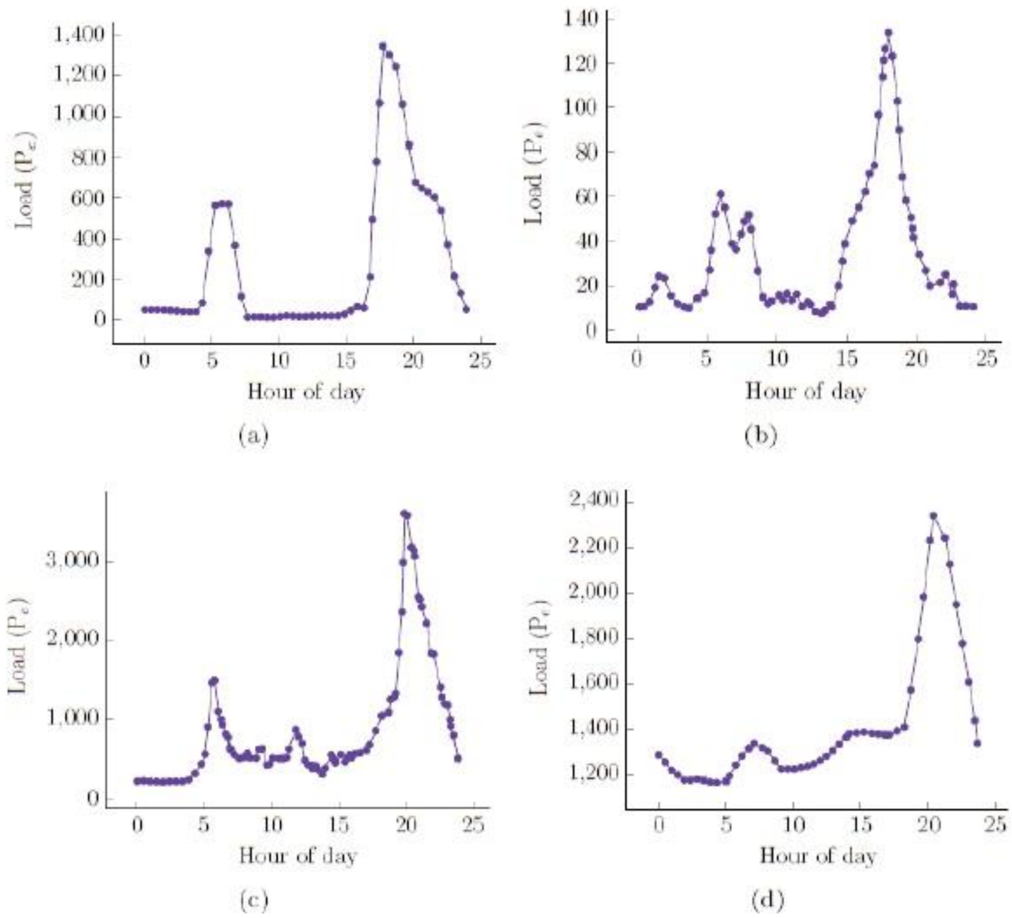


Figure 39: load profiles of different rural countries; (a) Ban Pang, Thailand; (b) Rural Western Australia; (c) Alaminos, Rural Philippines; (d) San Juanico, Rural Mexico;[78]

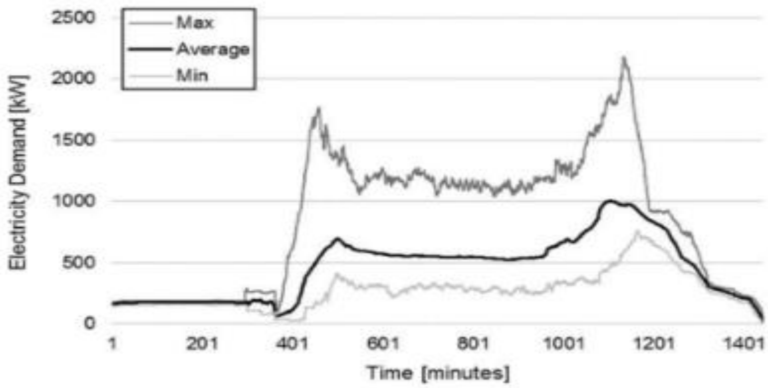


Figure 40: Load curves for thirteen Tanzanian villages, obtained through LoadProGen assuming 50% level of electricity access [79]

Looking at these diagrams, it can be said that the load profiles obtained for the current study have a realistic shape, which is especially close to the (d) case of Figure 39 and to the reference case of Figure 38. Lastly, another useful comparison is provided by Figure 40 which shows a very similar shape to the one I obtained, moreover for the same geographic area I am considering. Of course it is not possible to make a comparison in terms of consumption due to the large dimension of the examples present in literature and the smaller considered capacity of the studied system. As the communities of Sub-Saharan Africa develop and increase the number of used appliances for their daily routine, it is expected that their electricity load profiles get to have a shape that is always more similar to the one of a developed country such as Italy or any other.

4.5 Off-grid System Optimization

Once the daily load profiles are available, it is possible to start the actual process of off-grid system sizing.

The process was performed thanks to the use of the software HOMER Pro, which is available online in its trial version and comes with a large amount of features and possibilities for the user.

The first thing to do, was to decide which type of off-grid system was the most appropriate for this study case. It was decided to consider the installation of renewables only. As previously said, in Njombe area some hydropower is available and the existing grids rely on water resources. It was thought it could be interesting to assess the dimensioning of a 100% renewable grid, without the use of hydropower, to make it a more generic option, since not every village has such an availability of water nearby. Moreover, HOMER Pro takes into account constant load profiles in time, which would make it unrealistic to consider the presence of, e.g., a diesel generator, which depends on operation costs (fuel prices) and therefore is way more subject to changes in the demand over time. Avoiding fossil fuel generation allowed me to make the system planning for three different years for each scenario, forcing HOMER to simulate an evolution in time of the microgrid.

The scheme of the microgrid that was sized, taking inspiration from what literature [13] suggested, is the following.

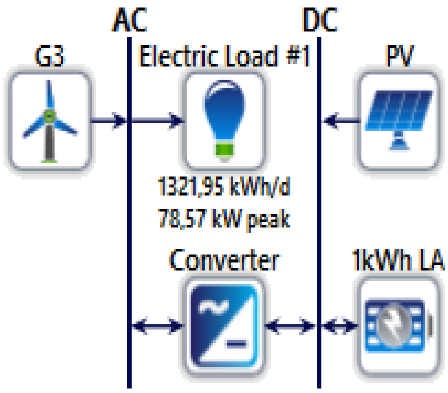


Figure 41: grid scheme for year 20 of maximum installed capacity scenario

Considering the scheme of Figure 41, it is possible to see the similarities with the schemes of other projects which were found in literature, such as the one shown below, taken from [13]. The only difference between the following figure and my case study is given by the fact that instead of using the grid as it was done in this example, I will use the batteries to stabilize the system.

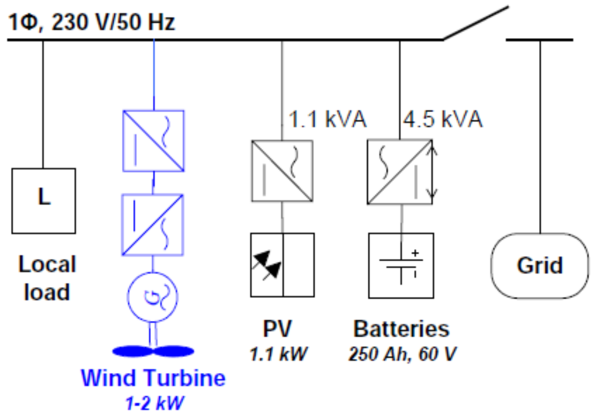


Figure 42: scheme of a microgrid project for a laboratory experiment in Athens

Once the scheme is clear and the various components have been selected, it is possible to consider the usual three scenarios. For each of these, the three years of interest (years 1, 10, 20 for MIN, MED, MAX scenarios) were analysed. The drawback of using HOMER Pro, is that its optimizing algorithm only takes into account the economical optimization of costs, while neglecting the advantages that some other features might bring to the overall system, such as, for example, the possibility to have a more differentiated generation. HOMER Pro takes into account the constraints one gives as an input and develops the least expensive solutions possible. It is then on the user to make reasonable evaluations about the obtained results.

A first thing one should decide, is whether he is fine with the definition of optimal solution HOMER proposes, because it might happen that someone is willing to optimize the system with respect to some other characteristics, even if it might mean to spend more money than what HOMER proposes. This reasoning helps understanding why two cases were included in the analysis. It was because, as it can be seen in the tables below, case 2 adds wind turbines to the generation capacity. Having wind in addition to solar might give greater reliability to the grid, for example during rainy or cloudy days, allowing for a differentiation in the generation resources. In the following tables the results can be found.

Year 1	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	4,72	0	21	1,24	25693
Case 2	3,73	3	22	1,08	39122

Table 9: year 1 sizing for minimum installed capacity scenario

Year 10	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	89,6	0	457	36,8	531343
Case 2	88	3	456	37,2	542245

Table 10: year 10 sizing for minimum installed capacity scenario

Year 20	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	244	0	1044	56,9	1,30M
Case 2	231	3	1077	85,6	1,31M

Table 11: year 20 sizing for minimum installed capacity scenario

Year 1	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	11,3	0	48	2,70	59826
Case 2	10,5	3	46	2,79	71902

Table 12: year 1 sizing for maximum installed capacity scenario

Year 10	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	105	0	454	24,4	561562
Case 2	101	3	457	35,8	574579

Table 13: year 10 sizing of maximum installed capacity scenario

Year 20	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	393	0	1622	87,7	2,05M
Case 2	381	3	1649	89,9	2,06M

Table 14: year 20 sizing of maximum installed capacity scenario

Year 1	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	7,75	0	34	1,77	41764
Case 2	6,75	3	33	1,81	53919

Table 15: year 1 sizing of median installed capacity scenario

Year 10	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	150	0	698	49,8	839548
Case 2	153	3	678	49,9	849496

Table 16: year 10 sizing of median installed capacity scenario

Year 20	PV [kW]	Wind [kW]	Battery [kWh]	Converter [kW]	NPC [€]
Case 1	367	0	1646	109	2,01M
Case 2	366	9	1634	109	2,05M

Table 17: year 20 sizing of median installed capacity scenario

In the following diagram a summary of case 1 results for each scenario is provided.

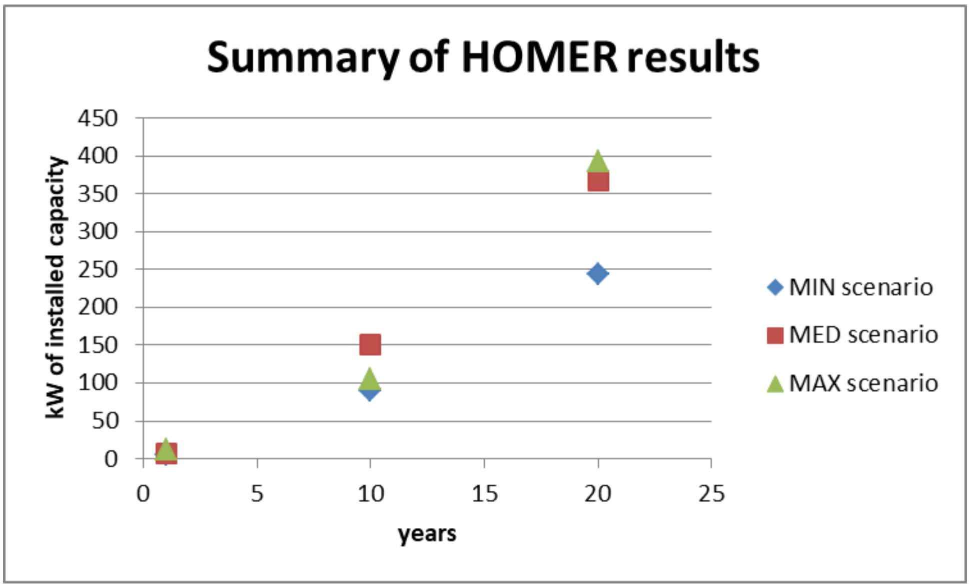


Figure 43: Summary of HOMER results for MAX, MED and MIN scenarios

Given these results, it is now possible to compare them and analyse their content more in detail. The first thing to clarify is that in each of the simulations two or three results were given as an outcome by the software, but only the two best options were reported here. For each of the tables above, HOMER had selected case 1 as the optimized case in terms of Net Present Cost (NPC).

In all the 9 situations, it is possible to recognize some patterns for the two selected cases. Analysing case 1, the optimum, it can be noticed that the output grid is always an all solar solution, with no wind turbines installed and therefore needing more batteries than the correspondent case 2.

Giving a more comprehensive look at all of the tables, it can be said that the outcomes respect the expectations one could have in the beginning. The maximum capacity scenario is expected to present a larger amount of installed components at the end of the process, while in the single scenarios, the installed generation capacity is expected to increase in time as the load profile does. These two things happen and accordingly to this output the same happens also to the net present cost of the system. The patterns that had been noticed in the previous chapters are also respected. Therefore, looking at year 10, case 1, for maximum and median scenarios, the number of installed kW of solar is greater in the median case rather than in the maximum one and the situation is going to reverse in the following years. For what concerns investment costs, MAX, MED and MIN scenario at year 20 keep the same ranking as for the installed capacity, but again the pattern changes going back to year 10.

As explained in the previous sections, when facing the issue of sizing a long-term facility with scarce information, it is necessary to make some strategic decisions about how to manage the differences between short- and long-term behaviour of the demand. Looking at year 1 of the various scenarios, it can be said that this is not really representative of how the grid is supposed to become in the following years. For example, year 1 of MAX scenario represents not even 5% of the final configuration. Therefore, unless one wants to make adjustments very soon after starting the facility, it would be suggestable to take into account a slightly longer period of time to project the off-grid system.
The problem, at this point, is that there is a huge uncertainty about which might be the values of the endogenous parameters that were let vary in the diffusion model thanks to the use of Monte Carlo method. There are two ways to handle this issue: the first is to prepare very specific questionnaires to assess which values it would be better to assign to each parameter. This process should be performed in every single village of interest and for a very large amount of households, making sure all types of these are taken into account. This process would be very time demanding and would need for a large amount of resources, both in economical and human terms. The questionnaires, indeed, would need for translation to the local language and local workers would be needed to go around administering the questionnaires to people. To do this, many guides are available, produced by many famous ONGs and

international cooperation and development organizations, usually linked with the United Nations programs. A very useful guide, for instance, is provided by the World Bank [80].

The second option, which is definitely to be considered less expensive and quicker to perform, follows exactly the steps that the author went through during this research work. The drawback of this method is that it will not give a univocal outcome, but the decision will have to be a result of a sensitivity analysis which takes into account the different possible solutions obtained as an output of the model and of the microgrid sizing software.

In this second case, the main issue will be to decide whether to project and build the microgrid taking into account the entire needed generation capacity at year 20, or to start by projecting a certain amount of capacity, keeping in mind that some updates and modifications of the system will be needed at some point of its lifetime. Indeed, it can be noticed how between year 10 and year 20 of the system in MAX scenario, the necessary installed capacity of PV in case 1 increases of almost 75% and this increase can be consequently found in the net present cost values. Moreover, in the current case, three different scenarios to choose from were available, all of them relative to the same 400 households of one only village. It will be very difficult to decide which is the scenario expected to best fit into the considered situation.

It is, of course, impossible to reach 100% certainty of which is the right decision to make, therefore a precautionary behaviour in the decision making process is definitely suggested.

In this case it can be observed in the above tables that year 10 of median and maximum scenarios, imply the installation of less than one half of what would be expected to become the mini-grid capacity in the respective years 20. Looking at the minimum installed capacity scenario, instead, it can be noticed that year 20 represents a half way with respect to the situations at year 10 and 20 of the other scenarios. It might therefore be a good choice to project the microgrid taking into account this option, in order to decrease the risk of money loss. It will then be easier to upgrade the system from ~244 kW of solar and ~1044 kWh of batteries to ~393 kW and ~1622 kWh, rather than starting from ~105 kW and ~454 kWh as requested by year 10 of the maximum installed capacity scenario forecast.

A further improvement which could help the decision process for the current study would be the identification of some parameters on which to perform a sensitivity analysis through HOMER, such as for example the irradiation values, temperature data or wind speed. In general terms, though, the output of this work can be considered realistic and consistent with the expectations. A number of parameters which were let vary allowed to have an output which is generic enough to be fitting with a range of situations.

5 Conclusion

The aim of this work was to analyse how the main complex dynamics and uncertainties related to developing countries realities influence the diffusion of electrical appliances and therefore the configuration of off-grid systems. In order to do it, it was necessary to model several technical and socio-economic aspects. It was possible to start by performing a wide literature review, which would cover all of the topics that might be needed in order to go through the research work. It was found out that socio-economic aspects had been studied and modelled thoroughly to contribute to the definition of diffusion processes, but so far in the energy sector the majority of studies would consider standard agents with no individual characteristics. The authors would tend to consider a stationary model in time, upon which to build their projects, without taking into account possible evolutions in time of the system or of the hosting community.

The work, therefore, was started by considering a fictitious realistic village in the Tanzanian region of Njombe and its social network structure was obtained, thanks to a regression on field data, in the form of a Barabasi-Albert network of 400 households with an average degree of 6. A Bass diffusion model for the connections to the off-grid system was then developed, along with a model based on Gompertz curves theory for the diffusion of electrical appliances across the village. The two of these would run at the same time, updating each other's results at every time step, over a time period of 20 years, which was assumed to be the lifetime of the off-grid system to size. In order to take into consideration the uncertainties related to the presence of many endogenous factors, it was decided to use the so called Monte Carlo method to be able to maintain a certain level of variability of the most uncertain factors, letting them vary uniformly within some ranges found in the literature, or extrapolated from real data analysis. 100 simulations were performed, which allowed to demonstrate that the endogenous factors actually have an impact on what will be the output of the model. Depending on the input parameters, at year 5 of the simulations the fan might be purchased by less than 50 to more than 250 people. To keep the study as general as possible, 3 very different simulations, among the 100 available, were chosen to be used in the following analyses. The choice was made on the basis of which was the amount of installed capacity at year 20 (last year of lifetime). For MAX (maximum installed capacity), MIN (minimum installed capacity) and MED (median installed capacity) scenarios a long-term analysis was necessary, to be able to properly design the microgrid, otherwise over- and under-estimation issues might have taken place. The software used throughout the sizing procedure were LoadProGen and HOMER Pro. LoadProGen allows for the creation of many different load profiles at the same time, but does not take into account the evolution of the household categories in time. On the other hand, HOMER Pro only takes into account one load profile at a time and keeps it constant for the entire lifetime of the system it is sizing. What was done, was running LoadProGen 3 times for each scenario, freezing the situation at year 1, 10 and 20, in order to simulate the evolution of the mini-grid along time. In order to keep this evolutionary behaviour in HOMER Pro as well, it was decided to let the software design 9 different systems (MIN, MAX and MED scenarios, each taken at year 1, 10 and 20), as if one was willing to plan an off-grid system able to answer to the specific load demand of each considered case for the entire duration of its lifetime. This allowed to make reasonable considerations about which type of approach would be more convenient when trying to size an off-grid system that is going to evolve in time in an uncertain manner. It would be useful in the future to find or create a software which would allow to consider continuous changes in the load demand and in the household configuration, so to be able to size the system in one only step, by considering the 20 years evolution all at once. Lastly, taking into account 3 different periods for each scenario, allowed to further confirm the necessity for a continuous software, since, the patterns which can be found comparing year 20 of the various scenarios do not always repeat in the previous years, which means that, by considering only a smaller amount of time (e.g., stopping the analysis at year 10) we would probably get the long-term estimates wrong and might incur bad cost recovery failures. Taking into account a single scenario, following its evolution in time, the necessary generation capacity might even double from year 10 to year 20 and accordingly to this output the same thing would happen to the net present cost of the system. Comparing different scenarios, the one presenting maximum installed capacity at year 20 might not hold the same property at year 10, which is indeed reflected in HOMER Pro output. For what concerns investment costs, MAX, MED and MIN scenarios at year 20 will keep the same ranking as for the installed capacity, but again the pattern would change going back to year 10. It would be interesting in the future to try and use this procedure in a real case study in the field, to further demonstrate the necessity of modelling the complexities and uncertainties related to energy planning in rural areas.

Appendix A

Here is presented the Matlab script for the Barabasi-Albert diffusion process:

```
rand('state',0);

%%% Barabasi & Albert (Scale-Free) MODEL
%% DIFFUSION PROCESS

ADOPTERS_BA = zeros(simulations,t+1);
ADOPTERS_BA(:,1) = A_0;

verifica_time_step_BA = zeros(N,t+1);

for k = 1:simulations

   deg = [degree(graph(A_poll_BA(:,:,k)))]';    %degree evaluates the degree of a node, while graph joins the nodes depending on the rules given inside the network matrix A_poll

%%% INITIAL "SEEDING" of A0

   states_N_iniz = zeros(1,N);

   if A_0 ~= 0
        [deg_sorted pos_sorted] = sort(deg(:),'descend');    %sort orders in ascending (default) or descending (if specified) order
        pos_intitial = pos_sorted(1:A_0)';
        states_N_iniz(1,pos_intitial)=1;
   else
        pos_intitial=[];
   end

%%% CREATE THE STRUCT of the STATES

    states_N = struct('s', [states_N_iniz; zeros(t,N)], 'type', zeros(1,N));
```

```
        %0 = POTENTIAL ADOPTER
        %1 = ADOPTER

    list=zeros(1,sum(deg));  %list is a vector of dimension equal to the
number of total links present in the network, inside list there will be 4
times number 1 if node 1 has degree equal to 4, 3 times number 2 if node
2 has degree equal to 3, and so on.
    posto=1;
    for i=1:N
        for j=1:deg(i)
            list(posto)=i;
            posto=posto+1;
        end;
    end;
    for i=1:N
        repeat=1;
        while repeat
            n=round(rand*sum(deg)+0.5);
            if states_N.type(list(n))==0
                states_N.type(list(n))=1;
                repeat=0;
            end;
        end;
    end;

    clear i
    clear j

%%% DIFFUSION PROCESS
            for m = 2 : t+1
        for j = 1 : N
            if states_N.s(m-1,j)==1
                states_N.s(m:end,j)=1;
                continue
            end
            contact_j  =  find(A_poll_BA(j,:,k));      %find  finds  the
positions in which the matrix is different from 0
contact_adopters_all_j=sum(states_N.s(m-1,contact_j));
    if states_N.type(j)==1
```

```
states_N.s(m,j)=rand<(p+q*contact_adopters_all_j);
verifica_time_step_BA(j,m)=(k==1)*(p+q*contact_adopters_all_j);
            elseif states_N.type(j)==0
                states_N.s(m,j) = rand < (p + q*contact_adopters_all_j);
verifica_time_step_BA(j,m)=(k==1)*(p+q*contact_adopters_all_j);
            end
        end
        ADOPTERS_BA(k,m) = sum(states_N.s(m,:));
    end
end
ADOPTERS_BA_mean = mean(ADOPTERS_BA);
```

Appendix B

Here is the Matlab script for the appliance diffusion process (fan example):

```
%appliance parameters variability
%fan
beta_fan=1.194*0.8+(1.194*1.2-1.194*0.8)*rand;
gamma_fan=4.144*0.8+(1.194*1.2-1.194*0.8)*rand;
b_fan(sim)=beta_fan;
g_fan(sim)=gamma_fan;

%fan parameters
p_fan=20; %power in W
fan_cost=25; %cost in $
MaxOwn_fan=1; %maximum ownership of 1 household
AvFloorSpace=8.98;  %average floor per capita in a household
AvHouseSize=4.8;    %average number of people in a household
alfa_fan=0.04*AvHouseSize*AvFloorSpace*MaxOwn_fan;  %alfa  of  ownership
curves

%fan ownership

%average expenditure of each quintile
AV_Q1_fan=mean(EXP_Q1);
AV_Q2_fan=mean(EXP_Q2);
AV_Q3_fan=mean(EXP_Q3);
AV_Q4_fan=mean(EXP_Q4);
AV_Q5_fan=mean(EXP_Q5);

%ownership level of each quintile
OWN_Q1_fan=alfa_fan*exp(-beta_fan*exp(-gamma_fan/1000.*AV_Q1_fan));
OWN_Q2_fan=alfa_fan*exp(-beta_fan*exp(-gamma_fan/1000.*AV_Q2_fan));
OWN_Q3_fan=alfa_fan*exp(-beta_fan*exp(-gamma_fan/1000.*AV_Q3_fan));
OWN_Q4_fan=alfa_fan*exp(-beta_fan*exp(-gamma_fan/1000.*AV_Q4_fan));
OWN_Q5_fan=alfa_fan*exp(-beta_fan*exp(-gamma_fan/1000.*AV_Q5_fan));

%rounded number of adopters of each quintile at the considered time-step
A_Q1_fan=round(OWN_Q1_fan*length(EXP_Q1));
A_Q2_fan=round(OWN_Q2_fan*length(EXP_Q2));
A_Q3_fan=round(OWN_Q3_fan*length(EXP_Q3));
A_Q4_fan=round(OWN_Q4_fan*length(EXP_Q4));
A_Q5_fan=round(OWN_Q5_fan*length(EXP_Q5));

O_f_Q1=0;
O_f_Q2=0;
O_f_Q3=0;
O_f_Q4=0;
O_f_Q5=0;

for i=length(EXP_Q1):-1:1
    if EXP(2,i)<MaxOwn_fan && EXP(1,i)>=aff*fan_cost && O_f_Q1<=A_Q1_fan
&& EXP(9,i)==1
        EXP(1,i)=EXP(1,i)-fan_cost;
        EXP(2,i)=1;
        O_f_Q1=O_f_Q1+1;
```

```
    end
end

for i=(length(EXP_Q1)+length(EXP_Q2)):-1:(length(EXP_Q1)+1)
    if EXP(2,i)<MaxOwn_fan && EXP(1,i)>=aff*fan_cost && O_f_Q2<=A_Q2_fan
&& EXP(9,i)==1
        EXP(1,i)=EXP(1,i)-fan_cost;
        EXP(2,i)=1;
        O_f_Q2=O_f_Q2+1;
    end
end

for                    i=(length(EXP_Q1)+length(EXP_Q2)+length(EXP_Q3)):-
1:(length(EXP_Q1)+length(EXP_Q2)+1)
    if EXP(2,i)<MaxOwn_fan && EXP(1,i)>=aff*fan_cost && O_f_Q3<=A_Q3_fan
&& EXP(9,i)==1
        EXP(1,i)=EXP(1,i)-fan_cost;
        EXP(2,i)=1;
        O_f_Q3=O_f_Q3+1;
    end
end

for     i=(length(EXP_Q1)+length(EXP_Q2)+length(EXP_Q3)+length(EXP_Q4)):-
1:(length(EXP_Q1)+length(EXP_Q2)+length(EXP_Q3)+1)
    if EXP(2,i)<MaxOwn_fan && EXP(1,i)>=aff*fan_cost && O_f_Q4<=A_Q4_fan
&& EXP(9,i)==1
        EXP(1,i)=EXP(1,i)-fan_cost;
        EXP(2,i)=1;
        O_f_Q4=O_f_Q4+1;
    end
end

for
i=(length(EXP_Q1)+length(EXP_Q2)+length(EXP_Q3)+length(EXP_Q4)+length(EXP
_Q5)):-1:(length(EXP_Q1)+length(EXP_Q2)+length(EXP_Q3)+length(EXP_Q4)+1)
    if EXP(2,i)<MaxOwn_fan && EXP(1,i)>=aff*fan_cost && O_f_Q5<=A_Q5_fan
&& EXP(9,i)==1
        EXP(1,i)=EXP(1,i)-fan_cost;
        EXP(2,i)=1;
        O_f_Q5=O_f_Q5+1;
    end
end

A_fan(t)=A_fan(t-1)+O_f_Q1+O_f_Q2+O_f_Q3+O_f_Q4+O_f_Q5;
```

List of Figures

List of Tables

Nomenclature

n	number
CC	clustering coefficient
K_v	degree of the node
N_v	number of links among the contacts of the node
$f(t)$	probability of adoption at time t
p	probability of getting influenced by external inputs
q	probability of getting influenced by word of mouth
$F(t)$	proportion of adopters at time t
c	contact rate
i	adoption rate
$Ownership_{q,A,U(t)}$	ownership level based on quintile, appliance, area
α_A	upper limit of appliance ownership
$\beta_{A,U}, \gamma_{A,U}$	exogenous appliance related coefficients based on linear regression
$PCO_{PC,q,U(t)}$	per capita expenditure by quintile and area
k_avg	average degree
m	maximum amount of adopters
$AR(t)$	number of adopters at time t
$P(t)$	potential adopters at time t
$A(t)$	adopters at time t
N	total number of households

Abbreviations Index

DCs	Developing Countries
UN	United Nations
ABM	Agent Based Model
OLS	Ordinary Least Squares
NLS	Non-linear Least Squares
PV	Photovoltaics
O&M	Operation & Maintenance
GHI	Global Horizontal Irradiance
NPC	Net Present Cost
MIN	Minimum installed capacity scenario
MED	Median installed capacity scenario
MAX	Maximum installed capacity scenario

References

[1] E. Hartvigsson, E. Ahlgren, J. Ehnberg, and S. Molander, “Rural Electrification Through Minigrids in Developing Countries: Initial Generation Capacity Effect on Cost-Recovery,” *33rd Int. Conf. Syst. Dyn. Soc.*, no. Iea, pp. 1–12, 2015.

[2] B. J. van Ruijven *et al.*, “Model projections for household energy use in India,” *Energy Policy*, vol. 39, no. 12, pp. 7747–7761, 2011.

[3] C. Piccardi, “ON THE ROLE OF NETWORK TOPOLOGY IN PRODUCT DIFFUSION,” pp. 1–6, 2011.

[4] F. Riva, E. Colombo, and C. Piccardi, “Modelling social networks in innovation diffusion processes : the case of electricity access in rural areas.”

[5] A.-L. Barabasi, “Scale-Free Networks: A Decade and Beyond,” *Science (80-.).*, vol. 325, no. 5939, pp. 412–413, 2009.

[6] F. M. Bass, “A New Product Growth for Model Consumer Durables,” *Management Science, Vol. 15, No. 5, Theory Series*, pp. 215–227, 1969.

[7] J. D. Sterman, “Business Dynamics: Systems Thinking and Modeling for a Complex World,” *Journal of the Operational Research Society*, 2000. [Online]. Available: http://www.amazon.com/dp/0071179895.

[8] A. Kemmler, “Factors influencing household access to electricity in India,” *Energy Sustain. Dev.*, vol. 11, no. 4, pp. 13–20, 2007.

[9] N. D. Rao and K. Ummel, “White goods for white people? Drivers of electric appliance growth in emerging economies,” *Energy Res. Soc. Sci.*, vol. 27, pp. 106–116, 2017.

[10] S. Mandelli, M. Merlo, and E. Colombo, “Novel procedure to formulate load profiles for off-grid rural areas,” *Energy Sustain. Dev.*, vol. 31, pp. 130–142, 2016.

[11] J. C. Rojas-Zerpa and J. M. Yusta, “Methodologies, technologies and applications for electric supply planning in rural remote areas,” *Energy Sustain. Dev.*, vol. 20, no. 1, pp. 66–76, 2014.

[12] S. C. Bhattacharyya, “Review of alternative methodologies for analysing off-grid electricity supply,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 1, pp. 677–694, 2012.

[13] D. Georgakis, S. Papathanassiou, N. Hatziargyriou, A. Engler, and C. Hardt, “Operation of a prototype microgrid system based on micro-sources equipped with fast-acting power electronics interfaces,” *PESC Rec. - IEEE Annu. Power Electron. Spec. Conf.*, vol. 4, no. February, pp. 2521–2526, 2004.

[14] Nasa, “Surface meteorology and Solar Energy,” *Solar Energy*, 2011. [Online]. Available: http://eosweb.larc.nasa.gov/sse/.

[15] C. Mone, M. Hand, M. Bolinger, J. Rand, D. Heimiller, and J. Ho, “Cost of Wind Energy Review,” *Nrel*, no. May, 2015.

[16] K.-P. Kairies, “Battery storage technology improvements and cost reductions to 2030: A Deep Dive,” *Int. Renew. Energy Agency Work.*, 2017.

[17] R. Fares, “The Price of Solar Is Declining to Unprecedented Lows,” *Sci. Am.*, 2016.

[18] C. Brivio, M. Moncecchi, S. Mandelli, and M. Merlo, “A novel software package

for the robust design of off-grid power systems," *J. Clean. Prod.*, vol. 166, 2017.
[19] B. J. van Ruijven, J. Schers, and D. P. van Vuuren, "Model-based scenarios for rural electrification in developing countries," *Energy*, vol. 38, no. 1, pp. 386–397, 2012.
[20] R. T. Kivaisi, "Installation and use of a 3 kWp PV plant at Umbuji village in Zanzibar," *Renew. Energy*, vol. 19, pp. 457–472, 2000.
[21] A. Cabraal, M. Cosgrove-Davies, and L. Schaeffer, "Best practices for photovoltaic household electrification programs," *Conf. Rec. Twenty Fifth IEEE Photovolt. Spec. Conf. 1996*, pp. 1357–1362, 1996.
[22] L. . Cabraal, A. ; Cosgrove-Davies, M.; Schaeffer, "Best Practices for Photovoltaic Household Electrification Programs - Lessons from Experiences in Selected Countries," *Washingt. DC*, 1996.
[23] F. Fuso Nerini, R. Dargaville, M. Howells, and M. Bazilian, "Estimating the cost of energy access: The case of the village of Suro Craic in Timor Leste," *Energy*, vol. 79, no. C, pp. 385–397, 2015.
[24] C. M. Macal and M. J. North, "Tutorial on agent-based modelling and simulation," *J. Simul.*, vol. 4, no. 3, pp. 151–162, 2010.
[25] R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research directions," *Int. J. Res. Mark.*, vol. 27, no. 2, pp. 91–106, 2010.
[26] J. Berger and C. Heath, "Where Consumers Diverge from Others: Identity Signaling and Product Domains," *J. Consum. Res.*, vol. 34, no. 2, pp. 121–134, 2007.
[27] M. O. Jackson, "Social and Economic Networks 1," *Network*, no. March, pp. 14–16, 2008.
[28] A. G. Chandrasekhar, "Econometrics of network formation," *Oxford Handb. Econ. Networks*, pp. 1–50, 2015.
[29] A. Maertens and C. B. Barrett, "Measuring Social Networks' Effects on Agricultural Technology Adoption," *Am. J. Agric. Econ.*, vol. 95, no. 2, pp. 353–359, 2012.
[30] A. S. Caria and M. Fafchamps, "Can Farmers Create Efficient Information Networks? Experimental Evidence from Rural India," no. January, 2015.
[31] A. Ben Yishay and a M. Mobarak, "Communicating with Farmers through Social Networks," *Mimeo (August 2013)*, no. August, 2013.
[32] V. Alatas *et al.*, "Network Structure and the Aggregation of Information :," 2013.
[33] S. Di Falco and E. Bulte, "The Impact of Kinship Networks on the Adoption of Risk-Mitigating Strategies in Ethiopia," *World Dev.*, vol. 43, pp. 100–110, 2013.
[34] K. van den Broeck and S. Dercon, "Information flows and social externalities in a Tanzanian banana growing village," *J. Dev. Stud.*, vol. 47, no. 2, pp. 231–252, 2011.
[35] C. Van den Bulte and Y. V. Joshi, "New Product Diffusion with Influentials and Imitators," *Mark. Sci.*, vol. 26, no. 3, pp. 400–421, 2007.
[36] C. Van den Bulte and S. Stremersch, "Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test," *Mark. Sci.*, vol. 23, no. 4, pp. 530–544, 2004.
[37] J. Goldenberg, S. Han, D. R. Lehmann, and J. Weon Hong, "The Role of Hubs in the Adoption Process The Role of Hubs in the Adoption," *J. Mark.*, vol. 73, no. 2,

pp. 1–13, 2009.
[38] G. Weimann, "The influentials: back to the concept of opinion leaders?," *Public Opin. Q.*, vol. 55, pp. 267–279, 1991.
[39] O. Bandiera and I. Rasul, "SOCIAL NETWORKS AND TECHNOLOGY ADOPTION IN NORTHERN MOZAMBIQUE," *Econ. J.*, vol. 116, no. 1957, pp. 869–902, 2006.
[40] L. Beaman, A. Benyishay, C. William, and A. M. Mobarak, "Can Network Theory based Targeting Increase Technology Adoption ?," pp. 1–64, 2014.
[41] C. Kiss and M. Bichler, "Identification of influencers - Measuring influence in customer networks," *Decis. Support Syst.*, vol. 46, no. 1, pp. 233–253, 2008.
[42] A. KANDLER and J. Steele, "Innovation diffusion in time and space: effects of social information and of income inequality," *Diffus. Fundam.*, vol. 11, no. 3, pp. 1–17, 2009.
[43] E. Menezes, C. M. Ogushi, G. C. Bonadia, J. C. Dall'Antonia, and G. M. De Holanda, "Socioeconomic factors influencing digital TV diffusion in Brazil," *Proc. 23rd Int. Conf. Syst. Dyn. Soc.*, p. 112, 2005.
[44] E. Kiesling, M. Günther, C. Stummer, and L. M. Wakolbinger, "Agent-based simulation of innovation diffusion: A review," *Cent. Eur. J. Oper. Res.*, vol. 20, no. 2, pp. 183–230, 2012.
[45] T. G. Conley and C. R. Udry, "Learning about a New Technology : Pineapple in Ghana Learni," *Am. Econ. Rev.*, vol. 100, no. 1, pp. 35–69, 2010.
[46] H. Gunatilake, N. Maddipati, and S. Patail, "Willingness to Pay for Good Quality, Uninterrupted Power Supply in Rural Madhya Pradesh, India," no. 13, 2012.
[47] M. Alam and S. Bhattacharyya, "Are the off-grid customers ready to pay for electricity from the decentralized renewable hybrid mini-grids? A study of willingness to pay in rural Bangladesh," *Energy*, vol. 139, pp. 433–446, 2017.
[48] E. R. Frederiks, K. Stenner, and E. V. Hobman, "Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 1385–1394, 2015.
[49] M. Filippini and S. Pachauri, "Elasticities of electricity demand in urban Indian households," *Energy Policy*, vol. 32, no. 3, pp. 429–436, 2004.
[50] F. M. Mahajan, Vijay; Muller, Eitan; Bass, "New Product Diffusion Models in Marketing: A Review and Directions fr Research," *Journal of Marketing, Vol. 54*, pp. 1–26, 1990.
[51] A. V. Banerjee, A. Chandrasekhar, E. Duflo, and M. O. Jackson, "Gossip: Identifying Central Individuals in a Social Network," *SSRN Electron. J.*, pp. 0–46, 2014.
[52] E. Oster, "Determinants of technology adoption: private value and peer effects in menstrual cup take-up," 2009.
[53] F. R. M. Leijten, J. W. Bolderdijk, K. Keizer, M. Gorsira, E. van der Werff, and L. Steg, "Factors that influence consumers??? acceptance of future energy systems: the effects of adjustment type, production level, and price," *Energy Effic.*, vol. 7, no. 6, pp. 973–985, 2014.
[54] S. Radas, "Diffusion Models in Marketing: How to Incorporate the Effect of External Influence?," *Privred. kretanja i Ekon. Polit.*, vol. 15, pp. 31–51, 2005.
[55] F. M. Bass, T. V Krishnan, and D. C. Jain, "Why the Bass Model Fits without Decision Variables," *Mark. Sci.*, vol. 33, no. 3, pp. 203–223, 1994.
[56] D. Satoh, "A discrete Bass model and its parameter estimation," *J. Oper. Res. Soc.*

Japan, vol. 44, no. 1, pp. 16–18, 2001.
[57] Lawrence and William H. Lawton, "Applications of Diffusion Models: Some Empirical Results," *New Prod. Forecast.*, 1981.
[58] J. Bonan *et al.*, "Social Interaction and Technology Adoption : Experimental Evidence from Improved Cookstoves in Mali ∗," 2017.
[59] D. Horsky, "A Diffusion Model Incorporating Product Benefits, Price, Income and Information," *Mark. Sci.*, vol. 9, no. 4, pp. 342–365, 1990.
[60] M. Pothitou, A. J. Kolios, L. Varga, and S. Gu, "A framework for targeting household energy savings through habitual behavioural change," *Int. J. Sustain. Energy*, vol. 35, no. 7, pp. 686–700, 2016.
[61] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The Diffusion of Microfinance," *Science (80-.).*, vol. 341, no. 6144, pp. 1236498–1236498, 2013.
[62] J. M. Eder, C. F. Mutsaerts, and P. Sriwannawit, "Mini-grids and renewable energy in rural Africa: How diffusion theory explains adoption of electricity in Uganda," *Energy Res. Soc. Sci.*, vol. 5, pp. 45–54, 2015.
[63] K. Louw, B. Conradie, M. Howells, and M. Dekenah, "Determinants of electricity demand for newly electrified low-income African households," *Energy Policy*, vol. 36, no. 8, pp. 2814–2820, 2008.
[64] C. Wilson and H. Dowlatabadi, "Models of Decision Making and Residential Energy Use," *Annu. Rev. Environ. Resour.*, vol. 32, no. 1, pp. 169–203, 2007.
[65] S. Abdullah and P. W. Jeanty, "Demand for Electricity Connection in Rural Areas: The Case of Kenya," no. 26, pp. 1–43, 2009.
[66] J. Bonan, S. Pareglio, and M. Tavoni, "Access to Modern Energy: a Revier of Barriers, Drivers and Impacts," *Nota di Lav.*, 2016.
[67] L. S. O. Liverpool-Tasie and A. Winter-Nelson, "Social Learning and Farm Technology in Ethiopia: Impacts by Technology, Network Type, and Poverty Status," *J. Dev. Stud.*, vol. 48, no. 10, pp. 1505–1521, 2012.
[68] D. Zhang, A. Ntoko, and J. Dong, "MATHEMATICAL MODEL OF TECHNOLOGY DIFFUSION IN DEVELOPING COUNTRIES," pp. 525–539.
[69] N. D. Rao and K. Ummel, "White goods for white people? Drivers of electric appliance growth in emerging economies," *Energy Res. Soc. Sci.*, vol. 27, pp. 106–116, 2017.
[70] M. A. Mcneil and V. E. Letschert, "Modeling diffusion of electrical appliances in the residential sector," no. August, 2010.
[71] B. M. Sopha, C. A. Klöckner, and E. G. Hertwich, "Adoption and diffusion of heating systems in Norway: Coupling agent-based modeling with empirical research," *Environ. Innov. Soc. Transitions*, vol. 8, pp. 42–61, 2013.
[72] J. M. Labeaga, X. López-otero, and X. Labandeira, "Estimation of Elasticity Price of Electricity with Incomplete Information Xavier Labandeira Estimation of Elasticity Price of Electricity with Incomplete Information," 2010.
[73] V. Assimakopoulos, "Residential energy demand modelling in developing regions. The use of multivariate statistical techniques," *Energy Econ.*, vol. 14, no. 1, pp. 57–63, 1992.
[74] V. Rai and A. D. Henry, "Agent-based modelling of consumer energy choices," *Nat. Clim. Chang.*, vol. 6, no. 6, pp. 556–562, 2016.
[75] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," *Renew. Sustain. Energy Rev.*, vol. 16, no. 9, pp.

6539–6565, 2012.
[76] GIZ, "Tanzania`s Small-Hydro Energy Market," 2009.
[77] giz, "Photovoltaics for Productive Use Applications," 2016.
[78] G. Prinsloo and R. Dobson, "Scoping exercise to determine load profile archetype reference shapes for solar co-generation models in isolated off-grid rural African villages," vol. 27, no. 3, pp. 11–27.
[79] F. Riva, L. Berti, S. Mandelli, J. Pendezza, and E. Colombo, "On-field assessment of reliable electricity access scenarios through a bottom-up approach: The case of Ninga SHPP, Tanzania," *2017 6th Int. Conf. Clean Electr. Power Renew. Energy Resour. Impact, ICCEP 2017*, pp. 340–346, 2017.
[80] WB, "Energy Policies and Multitopic Household Surveys: Guidelines for Questionare Design in Living Standards Measurement Studies," 2007.