POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in
Ingegneria Energetica

Estimating Load Profile for Off-Grid Power System Design: Model Improvement and Applications

Relatore: Prof. Emanuela Colombo

Correlatore: Ing. Stefano Mandelli

Tesi di Laurea Magistrale di:

Simone Losa Matr. 801574

Anno Accademico 2013-2014

Ringraziamenti

Molte persone hanno accompagnato questo mio percorso universitario.

Un ringraziamento dal profondo del cuore è però per le anime pazienti che mi sono sempre state vicine nella buona e nella cattiva sorte. Ai miei genitori, che mi hanno permesso di intraprendere questo percorso, e dai quali ho sempre ricevuto appoggio incondizionato e serenità nel momento delle scelte. A Sara, per essere sempre stata al mio fianco, motivandomi e sostenendomi nei momenti di difficoltà.

Un ringraziamento doveroso alla prof.ssa Emanuela Colombo, per avermi dato la possibilità di concludere il mio percorso di studi come avrei desiderato ed aver costituito il punto di riferimento per lo sviluppo di tutto l'elaborato.

Un ringraziamento particolare a Stefano, per tutto il tempo che mi ha saputo dedicare nello sviluppare e valorizzare al meglio quanto racchiuso in queste pagine.

Ringrazio i miei colleghi ed amici Alessandro, Gianluca, Lorenzo, Luca, Matteo, Michele e Walter che mi hanno accompagnato in questi 5 anni e che hanno reso più leggero il cammino.

Ringrazio tutti i miei amici e famigliari che, fuori dal Politecnico, mi hanno formato e fatto crescere con mille esperienze diverse.

Index

Abstract

The problem of access to energy, in recent years, has become increasingly important in the eyes of the international community. The link between access to energy, particularly electricity, and human development is clear, but the problem is far to be solved. This is especially true for the poorest countries, where the lack of access to electricity is often one of the main obstacles to development. The rate of electrification in developing countries is even lower when looking at rural areas: in sub-Saharan Africa, for example, only 14.1% of the rural population have access to electricity. For this reason, the electrification of rural areas in developing countries is an actual issue for NGOs.

There are two main possible solutions to provide access to electric energy in rural areas: (i) the traditional approach is to extend the national electricity grid to the rural areas, nevertheless large parts of these areas have low accessibility, low values of load demand and load factor. For these reasons, grid extension often results to be economically unfeasible; (ii) the alternative solution is to rely on decentralized and distributed generation, which often results to be the most appropriate technology option since power plants are installed close to the load, they can be sized in order to best fit with local load demand, and they can be fuelled by local resources (i.e. renewables sources). In recent years, for a variety of reasons, the second solution has been often preferred for rural electrification.

In order to dimension these off-grid systems, there are many advanced software based on numerical and analytical methods; however, all require as input the electrical load to be met, in the form of load profile (usually daily load profiles). A research on articles about projects on system sizing for off-grid electrification of rural areas has shown that often it is not given enough importance to the estimation phase of the load, and that there is not a clear and definite procedure to be followed for this phase.

At the light of the issue about lack of appropriate methods to estimate load profiles for supporting the off-grid system design process, in my thesis I worked for the improvement and application of an existing model, coded in MATLAB, for the estimation of load profiles. The aim of this model is to provide an appropriate method of estimation of the load to support the optimization and sizing process of off-grid systems for rural electrification. The algorithm is designed to work with few input data commonly required by even the simplest approaches for the estimation of energy requirements in rural areas currently used (installed equipment, times

of use, etc.). The main features of the model are: (i) the bottom-up approach; (ii) stochastic sampling of switching-on instants of electrical appliances; (iii) the implementation of parameters related to the load profiles, and of empirical correlations between these parameters; (iv) the possibility of considering the uncertainty of the input data by introducing randomization parameters.

The model has been finally applied in two cases. In the first application, it has been used to estimate load profiles of a un-electrified peripheral-urban area of Uganda; the estimated load profiles has been used as input data for a specific software which carried out the sizing of a photovoltaic system associated with batteries. In the second application, it has been possible to test the model, comparing the metered load profiles of an electrical load of a college in Bali, Cameroon, and the load profiles estimated by the algorithm for the same context.

Keywords: *Access to energy, Rural electrification, Off-grid power systems, Rural power systems, Electrical Load profiles, Load profile estimation, Stochastic Model for Load profile estimation, Bottom-up Model for Load profile estimation, Power systems sizing*

Sommario

Il problema dell'accesso all'energia, negli ultimi anni, è diventato sempre più importante agli occhi della comunità internazionale. Il nesso tra accesso all'energia, in particolare energia elettrica, e sviluppo umano è chiaro, ma il problema è lungi dall'essere risolto. Questo è vero soprattutto per i paesi più poveri, dove la mancanza di accesso all'energia elettrica è spesso uno dei principali ostacoli allo sviluppo. Il tasso di elettrificazione nei paesi in via di sviluppo è ancor più basso se si guarda alle zone rurali lontane dai grandi centri abitati: nell'area dell'Africa Sub-Sahariana, per fare un esempio, solo il 14,1% delle popolazioni rurali hanno accesso all'elettricità. Per questo motivo, l'elettrificazione delle zone rurali dei paesi in via di sviluppo è un tema di grande attualità per le ONG.

Le possibili soluzioni per fornire accesso all'energia elettrica a zone rurali sono due: (i) l'approccio tradizionale è quello di estendere la rete elettrica nazionale fino alle aree rurali, tuttavia molte di queste aree sono molto distanti dai centri urbani e hanno scarsa accessibilità, oltre che richieste energetiche modeste. Per queste ragioni, questa soluzione risulta raramente percorribile. (ii) La soluzione alternativa è quella di affidarsi alla generazione decentralizzata e distribuita, che spesso risulta essere l'opzione installando impianti locali non collegati alla rete elettrica principale. Negli ultimi anni, per una serie di motivi, questa seconda alternativa sta guadagnando consensi rispetto al passato, soprattutto con l'installazione di impianti off-main-grid basati totalmente o parzialmente su fonti rinnovabili disponibili localmente (fotovoltaico, eolico, idroelettrico, biocombustibili).

Al fine di dimensionare questi impianti off-grid, esistono molti software avanzati basati su metodi numerici e analitici, tuttavia tutti richiedono come dato di input il carico elettrico da soddisfare, sotto forma di curva di carico (solitamente giornaliera). Una ricerca tra gli articoli riguardanti progetti di dimensionamento di sistemi off-grid per l'elettrificazione di aree rurali, ha evidenziato che spesso non viene data sufficiente importanza alla fase di stima del carico, e che non esiste una procedura chiara e definita da seguire in questo ambito.

Per questo, nella mia tesi, mi sono occupato del miglioramento e della applicazione di un modello (preesistente), codificato in MATLAB, per la stima di curve di carico. L'obiettivo di tale modello è quello di fornire un metodo appropriato di stima del carico per supportare il processo di ottimizzazione e dimensionamento di impianti off-grid per l'elettrificazione di aree rurali. L'algoritmo è progettato per funzionare con dati di input comunemente richiesti anche dai più semplici approcci per la stima dei fabbisogni energetici nelle aree rurali attualmente utilizzati (apparecchi installati, orari di utilizzo, ecc.). Le principali caratteristiche del modello sono: (i) l'approccio bottom-up; (ii) il campionamento stocastico degli istanti di accensione degli apparecchi; (iii) l'implementazione di parametri relativi alla teoria delle curve di carico, tramite correlazioni empiriche tra di essi; (iv) la possibilità di considerare l'incertezza dei dati di input tramite l'introduzione di parametri di randomizzazione.

Il modello è stato infine applicato in due casi. Nel primo caso, è stato utilizzato per fornire curve di carico relative all'utenza di una zona periferica di un contesto urbano in Uganda, sulle quali basarsi per effettuare il dimensionamento di un impianto fotovoltaico associato a batterie per l'accumulo. Nel secondo caso, è stato possibile testare il modello, comparando le misurazioni effettuate sul carico elettrico di un college a Bali, in Camerun, e le curve costruite con l'algoritmo per lo stesso contesto.

Parole chiave: *Accesso all’energia; Elettrificazione rurale; Sistemi energetici Off-Grid; Sistemi energetici rurali; Curve di carico elettrico; Stima di curve di carico; Modello stocastico per stima di curve di carico; Modello Bottom-up per stima di curve di carico; Dimensionamento di sistemi energetici.*

1 Introduction, motivations and problem formulation

This thesis copes with a stochastic method for the estimation of load profiles of un-electrified rural areas. This theme is specifically developed within the frame of rural electrification in developing countries (DCs), and in particular the model is proposed as a support tool for off-grid system sizing projects.

1.1 Rural electrification and off-grid systems

The development of human society has been marked all throughout history by the role of energy resources. In the last decades, the importance of energy in the global scenario and the interconnections with the environment and society have become more evident. The need to fight both poverty through eradication of energy insufficiency and to increase access to modern energy service is actually recognized worldwide. The nexus between energy and development has been widely discussed ([1], [2]), and some indicators as HDI and EDI can help to clarify this correlation. Human Development Index (HDI), created by Indian economist Amartya Sen and published by UNDP in 1990, is a composite statistic of life expectancy, education, and income indices used to rank countries into four tiers of human development. Energy Development Index (EDI), recently developed by IEA, is a combination of four indicators, mainly related to electricity and modern fuels, which take in account different aspects of energy poverty, created specifically in order to better understand the role that energy plays in human development. The correlation between EDI and HDI indexes is evident (Figure 1.1), confirming that when access to energy improves in terms of quality and quantity (EDI raises), human development improves (HDI also raises).

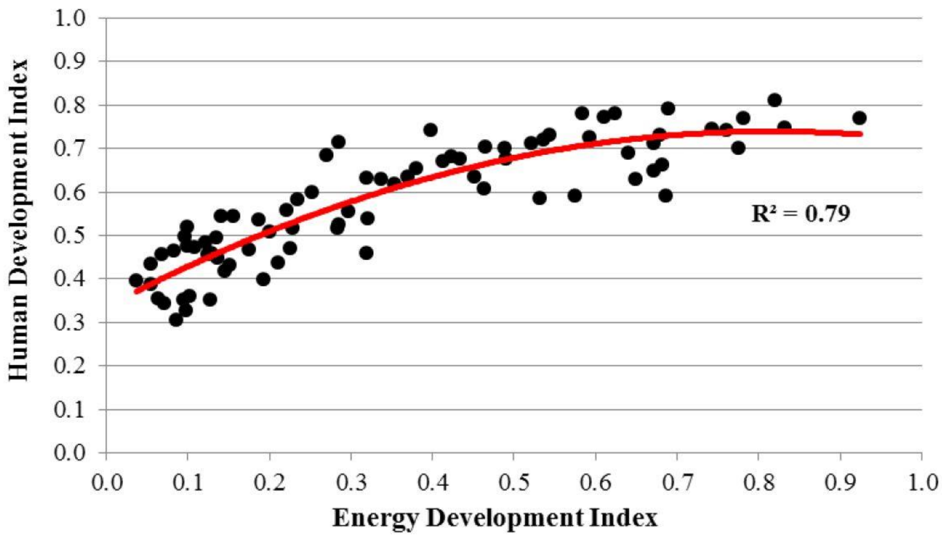


Figure 1.1: Comparison between HDI and EDI for 80 developing countries in 2012 [1]

Access to modern energy is today considered a necessary condition for human development: it is a requirement for achieving the status of essential and fundamental human rights, such as the protection of life, health, liberty, security, etc. Therefore, access to energy is essential to encourage development and to fight poverty, which is intended as a lack of opportunities and rights [3].

Dealing with the problem of access to energy means to study a dual problem: that of access to electricity (i.e. electrification) and that of access to modern fuels and efficient use of traditional biomass. In order to identify the dimension of the issue in the world, it is helpful to mention some statistics about it: today 1.3 billion people have no access to electricity, a further billion do not have access to a reliable electricity grid, and 2.6 billion people rely on traditional biomass (often inefficient and polluting appliances are used) for cooking and lighting. This situation is mainly localized in developing countries (DCs) characterized by low-income economies and low energy consumption per capita, and it becomes more evident in rural areas.

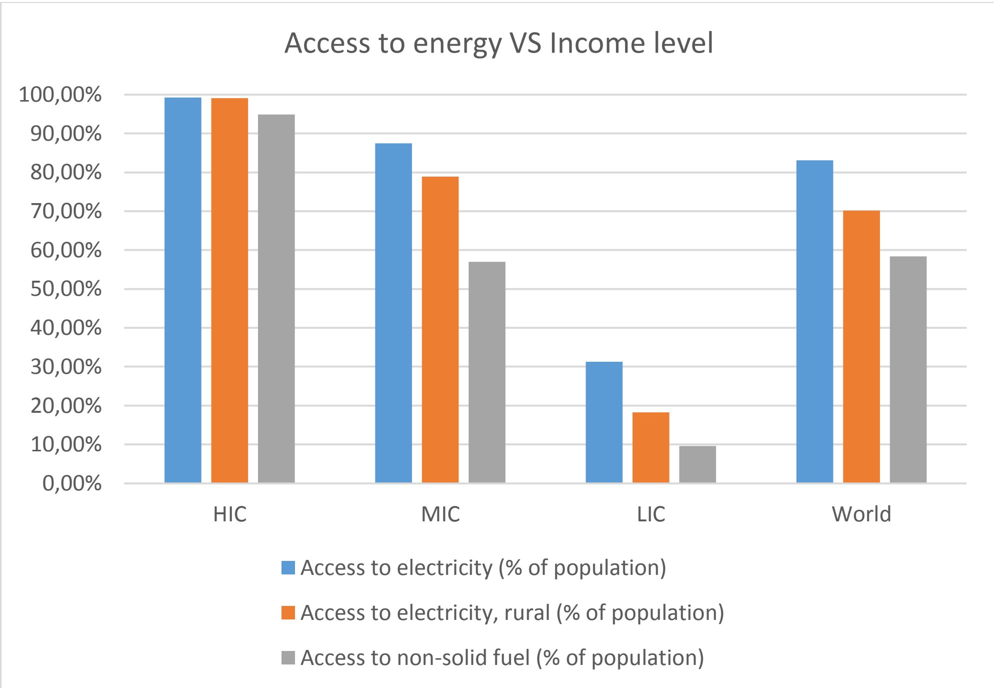


Figure 1.2: Share of population with access to electricity (total and rural) and to solid fuel in countries grouped by Income level (High Income, Medium Income and Low Income Countries)

In 2010, approximately 83% of the world population has access to electricity, but in Sub-Saharan Africa, which is one of the most critical regions in terms of electrification, only about 31.8% of the population has access to electricity, and this rate drops to 14.1% if we consider only the rural areas [4]. Although these countries are often rich in primary energy resources, these are rarely exploited at local level, moreover current energy systems are weak and characterized by low reliability. To solve the problem of access to energy, it is necessary to study and implement appropriate energy strategies that are not only effective in the short term, but that are sustainable in the long term, in the three main dimensions of sustainability: economic, environmental and social [1].

It is recognized that *access to electricity* is the main leverage for development. The World Energy Outlook 2012 foresees an increase in the electrification rate of developing countries from 76 % in 2010 to 85 % in 2030. The IEA projection states that in 2030 1 billion people will still be without electricity: Latin America will achieve universal access; developing Asia will halve the number of people affected while sub-Saharan Africa will keep a negative trend at least until 2025. Access will increase mainly in urban areas, where providing services is easier and more profitable for public utilities and private suppliers. Instead, people in rural areas continue to have poor access to electricity, because providing electric energy to small and scattered settlements is more complex.

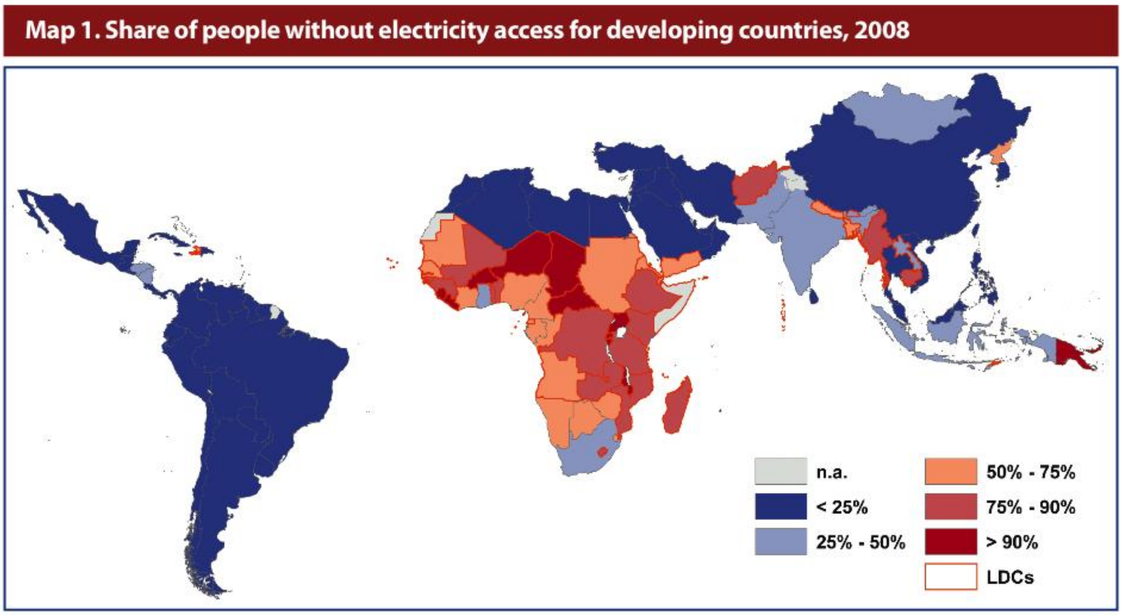


Figure 1.3: Share of people without access to electricity for developing countries (2008). Based on UNDP's classification of developing countries and the UN's classification of Developing Countries.

Concerning the electrification for rural areas, we can identify two main types of intervention:

- Increase access to national electric grids, extending the transmission and distribution grids to rural areas and raising the generation capacity of the centralized power systems.
- Invest on *off-main-grid decentralized and distributed generation (DDG)* systems, based on renewable energy technologies, which can provide cheaper electricity to rural communities, even though they require high investment costs. *Off-main-grid* refers to systems that operate detached from the main centralized grid. *Decentralized Generation*

refers to systems that are made by autonomous units where production, conversion and distribution have no interaction with other units. *Distributed Generation* refers to systems based on decentralized production and conversion units which interact through a distribution grid [1].

The traditional approach for increasing electricity access in rural areas is grid extension, in which cost of energy may be cheaper, but the cost of extending the grid to sparsely populated areas can be very high and long distance transmission systems may have high technical losses. Moreover, the customers living in these areas often have low accessibility, low values of load demand and load factor. For these reasons, grid extension often results to be economically unfeasible. In these cases Distributed Generation (DG) systems become the most appropriate technology option since they can be installed near the load, they can be sized in order to best satisfy load demand, and they can be fuelled by local renewable sources [5].

Additional reasons for which in recent years DG systems have been receiving a growing consideration are:

a) Technological improvement
   - increased performance of the small power technologies
   - development of electronic metering and control equipment
   - increased consumer demands for highly reliable power supply
b) Environmental concerns
   - growing concern as for the greenhouse gases emissions
   - public awareness of the impacts of the electric industry
   - opposition to building new transmission lines
c) Economic opportunities
   - to avoid transmission and distribution related costs
   - To tackle the current risky nature of large scale plant investments
   - to reduce power plants costs with combined heat and power generation
   - to better exploit profit margins within the competitive market
d) Political asset
   - to decrease dependence from fossil fuels
   - to increase primary source diversification
   - to reduce vulnerability of the supply chain in centralized systems
e) Social issues

- increasing public desire to promote “green technologies”
- growing interest towards energy autonomy communities and sustainability

1.2 Motivation: estimating load profiles

Today, also in the developed countries is occurring a slow shift from centralized systems to distributed systems and smart grids, with a preference for renewable resources. Consequently, in countries where electrification is still in its early stages, it would be smart to consider small scale decentralized systems mainly based on renewable energy as an alternative to the traditional centralized electric system based on fossil fuels. However it is also true that “Off-Main-Grid” decentralized technologies, in particular those partially or fully based on renewable energy, require high investment costs. Hence, the need to correctly design and size these systems, in order to optimize them to best meet the needs of users and to minimize capital costs.

Performing optimum design of off-grid plants reflects in looking for the system that best matches, at local level, energy resources with electric demand given certain technologies and context features. A common classification of the simulation and sizing techniques may be recognized, for which the techniques can be grouped in three categories [6]:

- Intuitive methods: simplified calculation of the system components size carried out without establishing any relationship between the different subsystem or taking into account the random nature of solar radiation and loads. They are mostly chosen for their simplicity in calculations, which makes them more intelligible and replicable by non-expert designer. The negative point of this approach is the results’ approximation which can lead to an over or an under sizing problem. ([7]–[10]).
- Numerical methods: several combinations of system components sizes are simulated typically on yearly basis and one or more objective functions are used to choose the best combination that addresses the load. For each time period considered, usually an hour, the energy balance of the system is calculated. Generally, they are preferred when more accurate results are required in order to optimize the energy and economic cost of the system. Numerical methods have also the advantage of allowing additional aspects of sizing to be analyzed such as the different models for the systems components. Unfortunately, these methods require long calculation time and need long and accurate data sequences. ([11]–[14]).
- Analytical methods: functional relationships between the variables of interest lead to solve the sizing problem (i.e. usually developed as a mathematical optimization problem with an

objective function subjected to one or more conditions). One of the main drawbacks of these methods is that either they are not accurate enough or they require the determination of specific coefficients for the functional relationship. On the other side, their strongest advantage is that the simulation of the different subsystem sizes is simple and relatively fast. ([15]–[17]).

Regarding the electric demand, the simplest approaches for DDG design (i.e. intuitive methods) rely on monthly or daily electricity requirements [10], while the most advanced methodologies (numerical and analytical methods) and available software (i.e. HOMER, H_2RES, etc.) rely on electric load profiles [18]–[21], which display the electric consumption as a function of time (Figure 1.4).

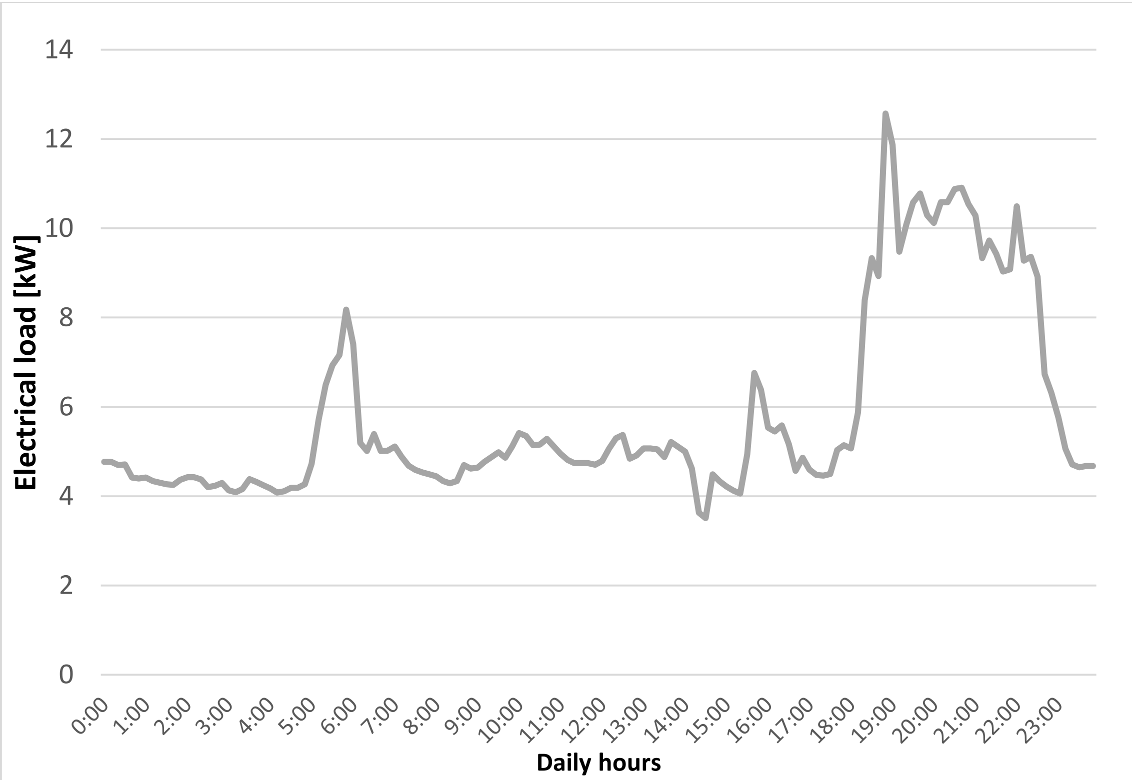


Figure 1.4: Typical electric load profile, which displays the electric consumption as a function of time.

An example of advanced software which perform a sophisticated numerical analysis is HOMER: this software is a micro-power design tool that can simulate and optimize stand-alone and grid connected power systems with any combination of technologies. Energy sources data are inputs to HOMER, as well as data on the electricity consumption of customers, in the form of annual electric load profile. In addition, the size, cost and lifetime of wind turbine, PV module, converter, battery and diesel generator are defined. Furthermore, the installation cost, design flow rate and head of hydropower source are all input to the software. HOMER performs a *simulation* process, which operates through the year one on a 1-hour basis, for each possible combination of technologies available. Identified the feasible systems (i.e. that can adequately serve the loads and satisfies any other constraints imposed by the user), it estimates their net present cost, and presents the feasible one with the lowest total net present cost as the optimal system configuration. Finally, a *sensitivity* analysis reveals how sensitive the outputs are to changes in the inputs. In the sensitivity analysis, the HOMER user enters a range of values for a single input variable. One of the primary uses of sensitivity analysis is in dealing with uncertainty.

Regardless of the methodology used to design and size the system - intuitive, numerical or analytical method - each of them requires data about the electric load as input. This means that, in order to perform an off-grid system design for an un-electrified rural area, electric demand must be estimated. The existing software, such as HOMER, perform accurate and detailed simulations on the operation of the plant in an entire year with time step up to a minute. Anyhow, since they require as input the load profile, the risk in performing the analysis in un-electrified areas, where electricity consumption are not known and must somehow be estimated, is to have an optimization software extremely precise and detailed, but which operates with input data related to the consumption profile drawn inadequately to the complexity of the optimization. Indeed, the drawing of load profile is a key point in the search of the optimal plant, because the final results of the optimization are strongly influenced by the demand to satisfy.

In order to understand what are the approaches used today for estimating user's load of non-electrified areas, a research into the literature has been performed. The research has been carried out between articles regarding the designing and dimensioning of off-grid plants for electrification of rural villages, or more generally, areas isolated from the main electricity grid. From the observations, it can be assumed that typically the problem is faced in these ways:

- A measured load profile relative to another rural area previously electrified, possibly in a similar context, is used as input load data. This method can give satisfactory results if load profiles of contexts effectively similar are chosen, but make sure this is not always easy, so there is the risk to use profiles inappropriate to the case study.
- The load profile is drawn manually, through estimations based on simple assumptions about electric appliances and user habits. Such assumptions and the related estimation methods provide adequately accuracy for intuitive methods, but not for load profile computation. Indeed they usually do not implement main features of load profiles (e.g. load factors, coincidence factors, etc.), and do not take into account the power-on random

nature of electrical appliances. Besides, when load profiles are computed according to these simple assumptions, the uncertainty given by the fact that several load profiles can occur within the same set of assumptions is not considered.

- In some few cases, even no reference is made to how the load profile is built.

Moreover, it's important to note that none of the existing system sizing software (such as HOMER) integrates a tool for estimating the load curves, and not even suggest any written procedure to perform the estimation. In the light of the findings of the research performed, it can be concluded that actually the choice of load profile is often overlooked in an off-grid plant design for un-electrified rural areas, and certainly a greater attention to this aspect could bring benefits in plant design.

1.3 Problem formulation: the LoadProGen tool

To give greater emphasis to the phase of load estimation, in my thesis I worked for the improvement and validation of a model, which estimates daily load profiles of customers in rural areas. The aim of this model is to provide a general pattern for the construction of load profiles as realistic as possible where measured data cannot be obtained, and therefore improve the designing accuracy of off-grid systems in this kind of contexts.

The model for load profile estimation is designed to work with few input data that are commonly considered in the simplest approaches for energy need estimations in rural areas (installed appliances, appliances nominal power, duration and time of functioning), also allowing to set the degree of uncertainty of the input data, coherently with the data that may be available in the context under study.

The principal features of the model are:

- It is based on a *bottom-up stochastic approach* [22]: the system load profile is obtained by a bottom-up hierarchic aggregation of coincident load profiles obtained for each end-use appliances employed by each class of user [23], in which the coincident load profiles are built through a stochastic simulation (weighted sampling without replacement). The sampling simulation set the instants of peak power time and appliances switching-on employing a probability density function made up with a uniform distribution coupled with a normal distribution (located in the peak power period).
- It takes into account the relationship between coincidence factor and the number of consumers: empirical evidence shows a relationship between group coincidence factor and

the number of consumers in a group, hence the coincidence factor of each class of user is corrected as for the number of consumers in that class by employing empirical coincidence curves [24], [25].

- It takes into account the relationship between coincidence factor and load factor: empirical evidence shows a relationship between group coincident factors and average load factors, hence the coincidence factor for each user class is computed given the average user load factor (computed from the input assumption) employing the empirical curve [24], [25].

In this way, the model accounts for the high load profile uncertainty of un-electrified areas, because each simulation generates a different possible load profile.

1.4 Thesis objective and structure

This thesis contributed to the research activities of the UNESCO Chair in Energy for Sustainable Development at Politecnico di Milano [26].

United Nations Educational, Scientific and Cultural Organization · UNESCO Chair in Energy for Sustainable Development

In this thesis I worked for the improvement and the application of an existing software, LoadProGen, coded in MATLAB, for the estimation of load profiles of rural areas. The aim of this software is to provide an appropriate method of estimation of the load to support the optimization and sizing process of off-grid systems for rural electrification.

In Chapter 2, initially the main definitions and parameters typically used to characterize load profiles are introduced. Then, theoretical concepts regarding the effects of the operation of electrical appliances and energy habits of consumers on the electrical load are explained. Finally empirical correlations, implemented in the model, that link some of the described parameters are presented.

In Chapter 3, an overview of the loads profile estimate methodologies today available is presented. Initially models used for *LF* and *RECM*, in particular used in developed countries. It is then placed greater focus on applications for off-grid rural electrification: after verification, through research in modern scientific literature, that actually do not exist models for load estimation specifically developed for use in rural areas of developing countries, the main solutions actually adopted are described. However, the research carried out has shown that the step of load estimating, in the projects of energy system sizing for electrification of un-electrified rural areas, is often overlooked.

In Chapter 4, LoadProGen, which is a software for estimating load profiles of un-electrified areas developed by the group of the UNESCO Chair in Energy for Sustainable Development of Politecnico di Milano, is introduced as regards: the general features, the required input data, and the mathematical formulation.

In Chapter 5, the main improvements developed in the software LoadProGen are shown: initially the main weaknesses of the early version of the model are summarized, and subsequently the solutions adopted in order to improve the algorithm are explained in detail. Then, the computational framework of the algorithm is presented. At the end, some examples of load profiles generated through LoadProGen are presented and compared with the results obtained using one of the simplest methods founded in literature, explained in Load profiles estimated.

In Chapter 6, two applications of the LoadProGen software tool are described. The first one aims at introducing an application of the model to perform sizing procedure of an off-grid photovoltaic system in a peripheral urban area of Uganda. The second one aims at introducing the application of the model as a load profile forecast tool, i.e. to show the capability of the model in matching metered real daily load profiles. In this case the application refers to the load profile of a college in a peri-urban area of Cameroon.

In Chapter 7, a summary of thesis contributions is given and future research directions are discussed.

2 Characteristics of consumer loads

Load profiles (or *load curves*) represent the electric consumption trends as a function of time. In the following chapter, initially the main definitions and parameters typically used to characterize load profiles are introduced. Then, theoretical concepts regarding the effects of the operation of electrical appliances and energy habits of consumers on the electrical load are explained. Finally empirical correlations, implemented in the model, that link some of the described parameters are presented.

2.1 Electrical loads

Consumers always purchase electricity as an intermediate step towards some final, non-electrical product. No one wants electric energy itself, they want the products it can provide: a cool home in summer, a warm one in winter, hot water on demand, cold beverages in the refrigerator, a lighted house in the evening, and 48 inches of dazzling colors with stereo commentary during Sunday-night soccer match. Different types of consumers purchase electricity for different reasons, and have different requirements for the amount and quality of the power they buy, but all purchase electricity as a way to provide the end-products they want. These various products are called *end-uses*, and they span a wide range, as shown in Table 2.1

Agricultural	Residential	Commercial	Industrial
Lighting	Lighting	Lighting	Lighting
Water heating	Water heating	Water heating	Water heating
Computer	Computer	Computer	Computer
Cooking	Cooking	Cooking	Filtration
Grain dryers	Clothes dryers	Inventory system	Finishing dryers

Table 2.1: Customer classes and end-use categories

Some end-uses are satisfied only by electric power (televisions, computers). In others, such as water heating, home heating or cooking, electricity is one of several possible, competing energy sources.

Each end-use is satisfied through the application of appliances or devices that convert electricity into the desired end product. The total electrical load of any group of users is the result of a simultaneous use of various electric appliances. The amount of electric load created on a power system within any end-use category depends on many factors, such as class of user, or the type of equipment installed. Let's consider for example residential lighting: people or businesses who need more lighting will tend to buy more electricity for that purpose. Also important are the types of appliances used to convert electricity to the end-use. Consumers using incandescent lighting rather than fluorescent lighting will use appreciably more electric power for otherwise similar end-uses. The schedule of demand for most end-uses varies as a function of time. In most households, demand for lighting is lowest during mid-day and highest in mid-evening, after dusk but before most of the residents have gone to bed. The demand varies not only during the hours of the day, but also during the months of the year, due to seasonal variations: the daily schedule of lighting demand usually varies slightly throughout the year, due to changes in the daily cycle of sunrise and sunset. Again, some end-uses are only seasonal: demand for space heating occurs typically only during winter, while air conditioners are turned on in summer.

2.2 Load profiles: fundamental concepts and definitions

Load behavior is dominated by individual appliance characteristics and *coincidence* – the fact that not all customers demand their peak use of electricity at precisely same time. For this reason, accurate load studies require considerable care. The following section provides the theoretical basis regarding the analysis of load profiles: fundamental concepts and main definitions are introduced.

Connected load

The *connected load* is the sum of the loads of all electrical devices connected to the composite system, i.e. the power value resulting from the sum of all nominal electric powers [W]. An household in a developed country might have a 500 watt refrigerator, a 2500 watt electric boiler, a 2500 watt washing machine, a 2000 watt dishwasher, a 1300 watt air conditioner, a 1500 watt vacuum cleaner, 30 lamps with an average load of 40 watts each, and 2500 watts of miscellaneous home entertainment, personal grooming, and other small appliances, for a total of 14000 connected watts of load. This power value corresponds to the maximum possible energy demand from this user, which could happen in the case all the devices connected are operating at the same time. Obviously, it is rare that the entire connected load in a system or at any one customer's location would be operational at one time (for example, air conditioning and heating would not be running simultaneously).

Electric load profiles

Use of the electrical devices varies as a function of time of day, day of week, and season of year. As a result, also electric load varies. A *load profile* is a graphic representation of electric load consumption as a function of time.

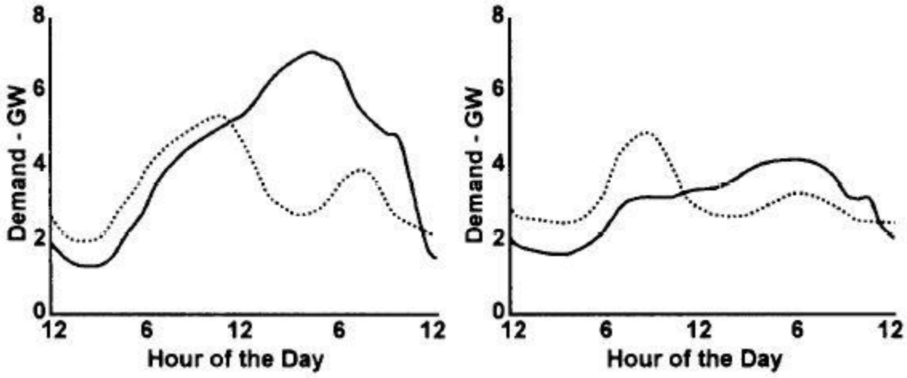


Figure 2.1: Typical summer (solid line) and winter (shaded line) peak day load profiles for a metropolitan power system in the southern US (left) and a rural system in New England (right).

Figure 2.1 shows seasonal *peak day* (i.e. the day of the season with the highest peak) load profiles for residential loads from two electric systems in the United States. In the first one, demand is highest in summer, during early evening, when probably a combination of air conditioning demand and residential activity is at a peak. In the other, peak demand occurs on winter mornings, when the electric heating demand is highest. Load profile shape - when peak load occurs and how load varies as a function of time - depends both on the connected load (appliances) and on the consumer habits in an area.

Demand and demand factor

The demand is the average load over a defined period (i.e. *demand interval*); it is calculated by dividing the energy consumed during the interval to the number of hours in the interval.

2.1

$$Demand = \frac{Energy\ during\ the\ interval\ [Wh]}{Interval\ range\ [h]}$$

Usually, demand is measured on an hourly or quarter-hour basis, but it can be measured on any interval. As the demand is a mean value of power, it does not allow to observe the load changes within the interval on which it is calculated. In this sense, the lower demand interval, the greater the accuracy with which the load profile describes the real behavior of the load, because it will lose less information relating to the maxima and minima that occur within the demand interval.

The *peak demand* (or *peak load*) is the maximum value of demand over a measurement period. For example, a period of an year contains 35040 fifteen-minute demand intervals. The maximum among these readings is the *peak fifteen-minute demand*.

The *demand factor (Df)* is expressed as the ratio between the peak demand and the connected load, i.e. between really measured peak and maximum theoretical peak; this ratio is usually considerably less than 1.

2.2

$$Df = \frac{Peak\ demand\ [W]}{Connected\ load\ [W]}$$

Load factor

The *load factor (Lf)* is the ratio between average demand and peak load in a given period. It is an indicator of which peak value level is maintained during the period under study (usually a day, a season or a year).

2.3

$$Lf = \frac{Average\ demand\ [W]}{Peak\ demand\ [W]}$$

Or:

2.4

$$Lf = \frac{Total\ energy\ during\ period\ [Wh]}{Peak\ demand\ [W] * period[h]}$$

The value of *Lf* ranges between 0 and 1: the case of *Lf* equal to 1 corresponds to a constant load profile, in which the load demand does not vary during the day; the case of *Lf* tending to zero corresponds to a null load throughout the day, but with a exception of a peak demand for an instant.

Customer classes

To improve the effectiveness of the analysis of the load, electric consumers are grouped into classes of similar demand behavior. A *user class* is any subset of customers whose distinction as a separate group helps identify or track load behavior in a way that improves the effectiveness of the analysis being performed. Examples of user classes can be: small apartments, family villas, supermarkets, schools, hospitals, etc. Usually all users belonging to the same group have load profiles with similar shapes, as they generally have similar habits, needs and types of equipment.

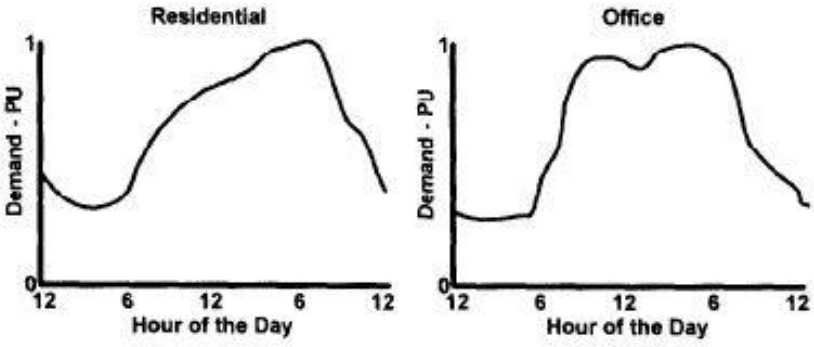


Figure 2.2: Customer classes typically display different daily load profiles. Shown here are the class summer peak-day loads from a metropolitan utility system in the southern United States.

Generally, to different classes of users correspond load profiles with different shapes (Figure 2.2), thus they do not demand their peak energy at the same time. As a result, the system peak load can be significantly less than the sum of the peaks of the various classes; this is called *inter-class coincidence* of loads. The percentage of the peak load attributable to a class is its *peak responsibility factor*.

Coincident and non-coincident load profiles

A common practice is representing the load of consumers on a class by class basis, using smooth, daily load profiles like those shown in Figure 2.3. These curves represent the average behavior or demand characteristics of customers in each class. For example, the system whose data are shown in Figure 2.3 has approximately 44000 residential customers. Its analysts will take the total residential customer class load (peaking about 290 MW) and divide it by 44000 to obtain a *typical residential load profile* for use in planning and engineering studies, a curve with a 24-hour shape identical to the total, but with a peak of 6.59 kW (1/44000 of 290 MW).

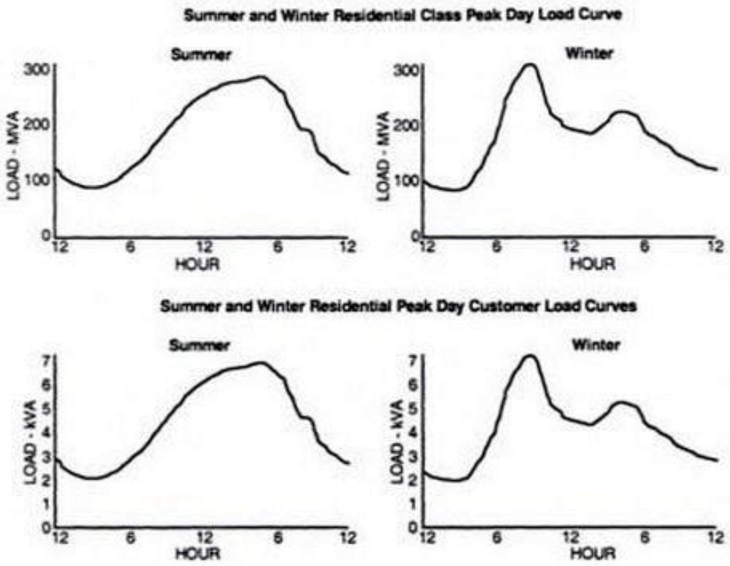


Figure 2.3: Top, summer and winter demand of the entire residential class in a southern US utility with 44000 residential consumers. Bottom, representation of individual consumer load profiles. Each is 1/44000 of the total class demand.

Actually, no residential customer in any utility's service territory has a load profile that looks anything like this averaged representation. Few concepts are important to understand why this is so, what actual load behavior looks like, and why the smooth representation is correct in some cases but not in others.

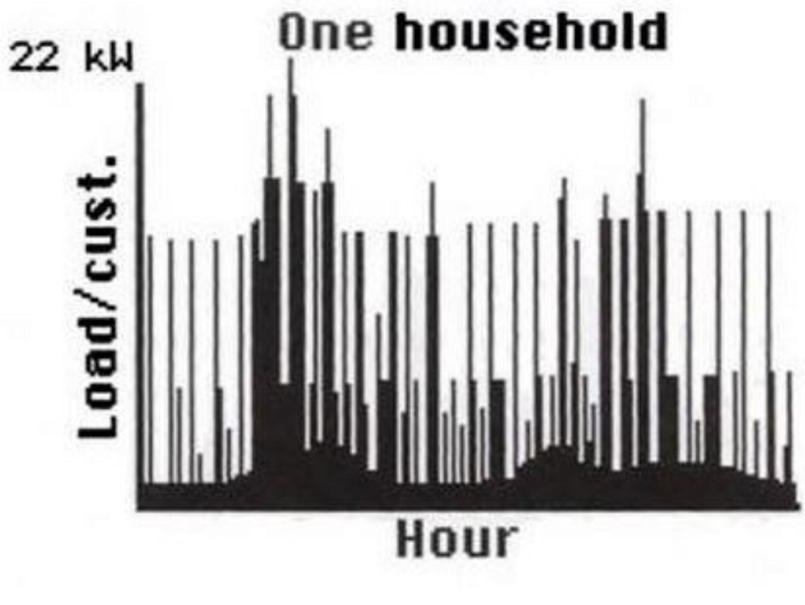


Figure 2.4: Example of actual daily load profile for an individual household, dominated by high 'needle peaks' causes by the on-off behavior of major appliances.

The load profile shown in Figure 2.4 is actually typical of what most of residential customer load looks like over a 24-hour period. Every residential customer's daily load behavior looks something like this, with sharp 'needle peaks' and erratic shifts in load as major appliances such as central heating, water heaters, washer-dryers, electric ranges and other devices switch on and off.

Appliances' duty cycle

To comprehend the reason for the erratic load behavior of a single consumer shown in Figure 2.4, we need to study the turning on-off behavior of the electric appliances connected to the system. Only a minority of electrical devices vary their load as a function of the end-use demand placed upon them. For most of them the load varies as a function of time with the form of a Boolean function, i.e. the appliance operates according to an on-off cycle in which electric load does an up and down variation between the nominal power when the device is on, and 0 when it is off. There is no curve in the sense of having a smooth variation in demand. Refrigerators, freezers, air conditioners, electric heaters, ovens, or irons all are examples of devices that vary their output as a function of time by changing their *duty cycle*, i.e. the fraction of time the device spends operating during any period.

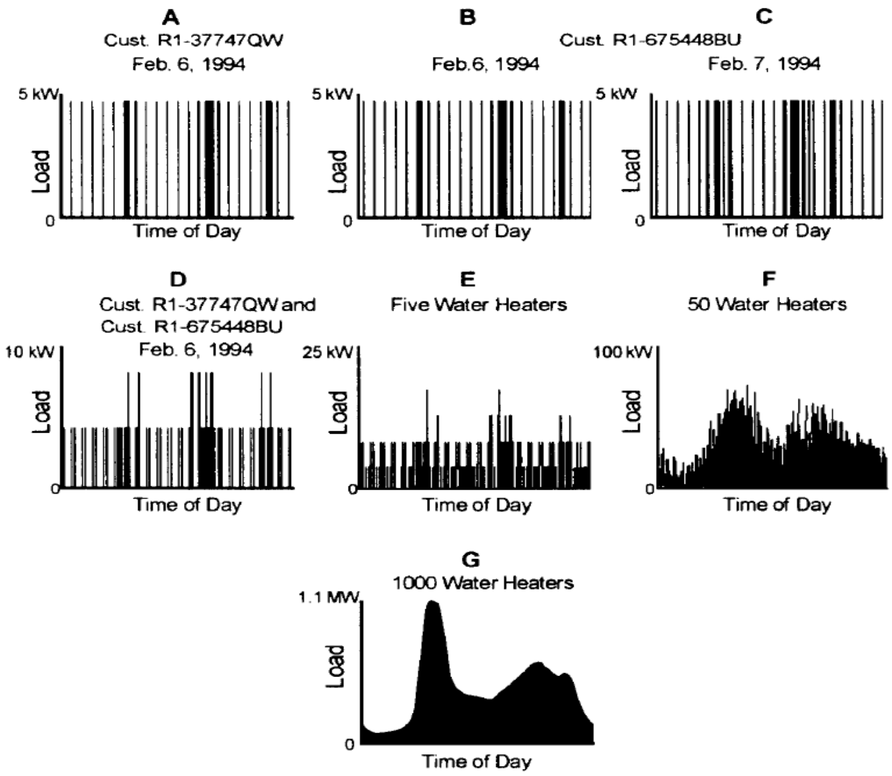


Figure 2.5: Daily load profiles for different sized groups of residential water heaters

Figure 2.5 A shows a daily load profile related to an electric boiler for domestic hot water: during most of the day, the unit does not produce any electrical load, it turns on at regular intervals for a few minutes so as to maintain the water temperature in a determined range. Only when user activity requires a greater use of hot water, the unit remains turned on for longer, increasing its duty cycle. Figure 2.5 B and C show respectively the load of another water heater in a house close to the first the same day and the day after, showing random differences and slight shifts in usage that occur from day to day. The remaining curves in Figure 2.5 D, E, F and G respectively represent the combined load profiles of 2, 5, 50 and 1000 electric water heaters. All water heaters considered exhibit the same overall on-off behavior, similar to that of Figure 2.5 A B or C, but differ slightly as to the timing of cycles. However, as an increasingly large number of water heaters are considered as a group, the erratic needle-peak behavior of the individual water heater load profile gradually disappears. The load profile representing a group's load becomes smoother as the size of the group is increased, the peak load per water heater drops, and the duration lengthens.

Coincident load behavior

Most of the loads in any home or business behave in a manner similar to the on-off, needle-peak behavior shown in Figure 2.5 A. Refrigerators and freezers, air conditioners, space heaters, water heaters, and electric ovens in homes; and pressurizers, water heaters, process and other finish heaters, and other equipment in industry; all turn on and off in a performance-regulated duty cycle manner. As a result, individual household load profiles, and many commercial and industrial site load profiles, display the blocky, on-off load behavior shown in Figure 2.6 B. As with the water heaters, when a group of similar loads (homes in this case) is considered as a single load, the load profile becomes smoother, the peak load per customer drops, and the minimum load per customer rises.

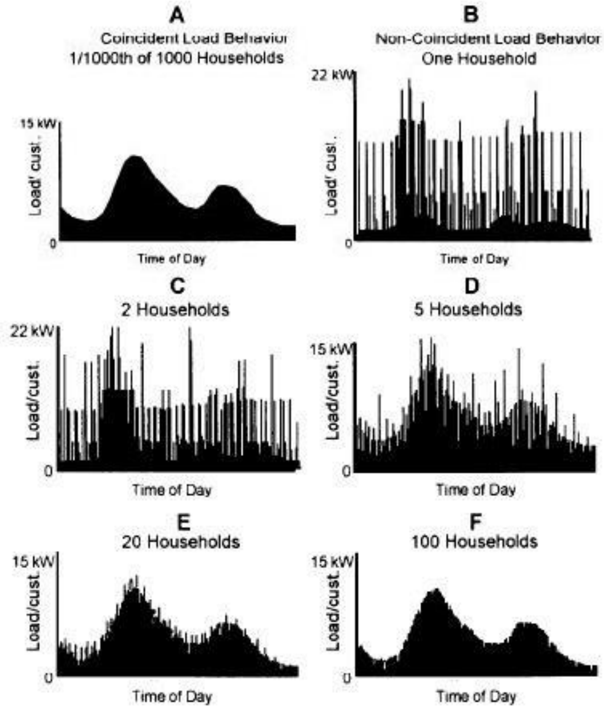


Figure 2.6: Coincident (A) and non-coincident (B) winter peak day load profiles for home in a suburban area in Florida. Curves B through F shows the gradual transformation from non-coincident to coincident behavior as group size increases. Note that the vertical scale of all six load profiles is in "load per customer" for each group.

While no single customer within the group depicted in Figure 2.6 would have an individual load profile that looked anything like Figure 2.6 A (every customer's load profile looks something like Figure 2.6 B), the smooth coincident load profile for the group has two legitimate interpretations:

- *The curve is an individual customer's contribution to system load*. On the average, each customer of this class adds this load to the system.
- *The curve is the expectation of an individual customer's load.* Every customer has a load that looks something like the on-off behavior shown in Figure 2.6 B. Figure 2.6 A gives the expectation, the probability-weighed value of daily load that one could expect from a customer of this class, selected at random.

Coincidence factor and coincidence curve

The tendency of observed peak load per customer to drop as the size of the customer group being observed is called *coincidence*. Usually, coincident load behavior is summarized for application to power distribution system engineering by the *coincidence factor* and the *coincidence curve*. Coincidence factor is defined as "the ratio of the maximum coincident total power demand of a group of consumers to the sum of the maximum power demands of the individual consumers comprising the group, both taken at the same point of supply and for the same period of time" [24].

2.5

$$Cf(n) = \frac{group\ peak\ for\ n\ customers\ [W]}{\sum(individual\ peaks)\ [W]}$$

Figure 2.7 shows a *coincidence curve*, a plot of how Cf varies as a function of customers' group size. Typically, for residential and small commercial load classes, Cf tends toward an asymptotic value of between 0.25 and 0.50 for large values of n (number of customers in the group), while for larger commercial and industrial customers, which have more stable energy consumption during the day, the value is usually higher, between 0.75 and 0.85. Coincidence behavior varies greatly from one utility to another, and among customer classes. However, the relationship between Cf and number of customers in a group will be discussed in greater detail in paragraph Relationship between coincidence factor and number of consumers.

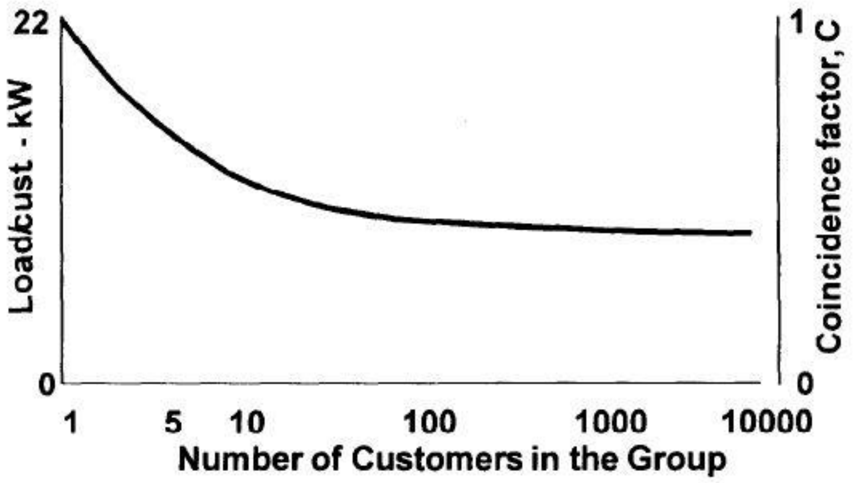


Figure 2.7: Coincidence factor (right scale) and Peak load per customer (left scale) as a function of the number of customers in a group for a residential class, from a power system in the central US.

The evaluation of the coincidence of a load has a key role in the power system design, because such equipment sizes are often determined by using coincidence curves to convert load data to estimate group peaks. For example, the "coincident peak" for a group of K customers can be estimated from individual peak load data, in this way:

2.6

$$group\ peak\ for\ K\ customers\ [W] = Cf(K) * K * average\ individual\ peak\ load\ [W]$$

The peak load for a transformer serving eight houses, each with an estimated individual peak load of 22 kW would be:

2.7

$$group\ peak\ for\ 8\ customers\ [kW] = Cf(8) * 8 * 22\ [kW]$$

A typical value of Cf(8) in a residential class is about 0.6, which returns 105.6 kW as the estimated group peak load.

2.8

$$group\ peak\ for\ 8\ customers\ [kW] = 0.6 * 8 * 22\ [kW] = 105.6\ [kW]$$

Load duration curves

A convenient way to study load behavior for some engineering purposes is to order the demand samples from greatest to smallest, rather than as a function of time, as shown in Figure 2.8. The two diagrams shown in Figure 2.8 consist of the same data samples, in a different order. Peak load, minimum load, and energy (area under the curve) are the same for both.

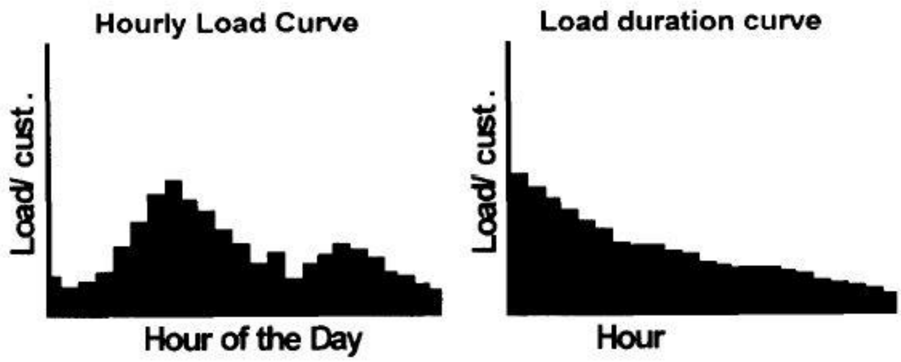


Figure 2.8: The hourly demand samples in a load profile are re-ordered from greatest to smallest values, to form a load duration curve.

Load duration curve behavior vary as a function of the system type. For example, load duration curves for small groups of customers will have a greater ratio of peak to minimum than similar curves for larger groups. Usually, load duration curves are produced on an annual basis.

2.3 Measuring load profile data

The manner in which load profile data are collected, analyzed and represented can produce a heavy effect on what the resulting load profiles look like, and can influence the perceived value of peak load and coincidence. The load profile shape can change on how the load data are measured (sampling method) and how periodically these measurements are recorded (sampling rate).

Sampling method

Sampling method refers to what quantity is measured – instantaneous load or total energy used during each period. We can distinguish two methods to sample data:

- *Demand sampling,* also called *period integration*, measures and records the total energy used during each period. At the beginning of each measurement interval, the energy meter is re-set to zero and begins counting the energy used. At the end of the period, the reading is recorded, and the counter is re-set to zero. Demand sampling always produces results that are valid within its context of measurement: hourly load data gathered by period integration for a customer will accurately reflect that customer's average energy usage on an hour-by-hour basis. Whether that sampling rate is sufficient for the study purposes is another matter.
- *Discrete sampling* (or *instantaneous sampling*) measures and records the load's value at specific periodic instances. For example, if the time-step is 15 minutes, every quarter hour the load recorder "open its eyes" to sample the instantaneous load, and then begin a waiting period of 15 minutes until the next sampling instant. This type of sampling often results in erratically data that dramatically misrepresent load profile behavior, because the recorder does not see what the load does during the waiting period: the measurement could be recorded during a short duration load peak, or during a moment when energy usage is particularly low. This type of sampling gives a good representation of the load behavior only if the sampling is more rapid then the rapid shifts in the load profile. Sampling at 15-minute interval is too much slow, because the real load can shoot back and forth from maximum to minimum several times in 15 minutes.

Summarizing, instantaneous sampling records the actual load value at specific instants every interval, while period integration averages its load measurements over the entire sample interval. There can be substantial difference in the recorded data, hence in how load profile looks like, depending on which of these two different sampling methods is used (Figure 2.9).

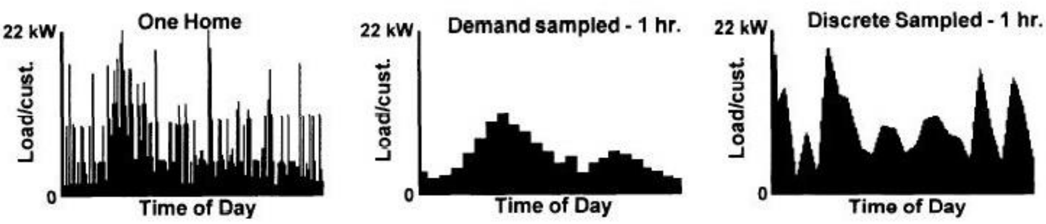


Figure 2.9: Two different load sampling methods (centre, right) applied on an hourly basis to the residential load profile from Figure 2.4 (left), produce different data.

Sampling rate

Sampling rate refers to the frequency of measurements – the number of times per hour that load data are recorded; Figure 2.10 shows Figure 2.4 load profile sampled with *period integration* on a 5, 15, 60 and 120-minute basis. Note that the resulting data displays significantly different behavior, depending on sampling rate.

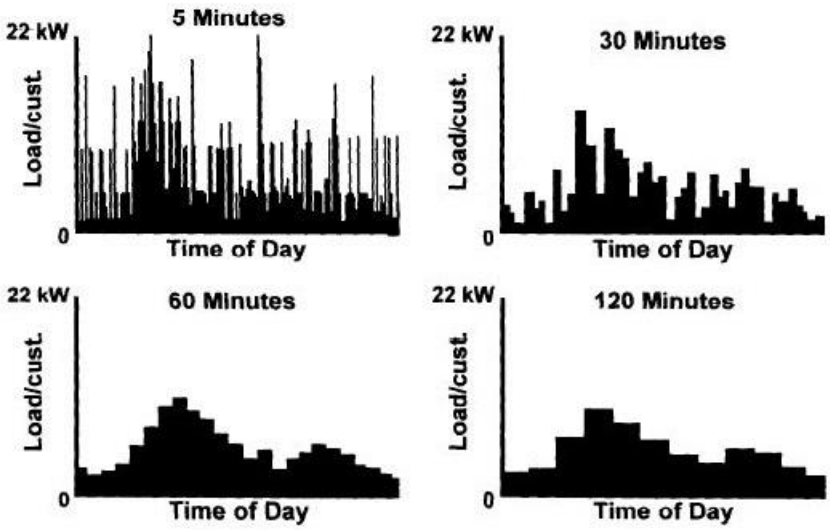


Figure 2.10: The same household load seen in Figure 2.4 sampled by period integration on a 5, 30, 60, 120-minute basis.

As the sampling is done faster, the curve shape displays more of its blocky, on-off nature: the recorded data comes closer to representing the true load profile shape peak value. But as shown, if a load is sampled by period integration applied at a slow rate, the resulting load data may look smooth, when, in fact, actual behavior is erratic, with high needle peaks.

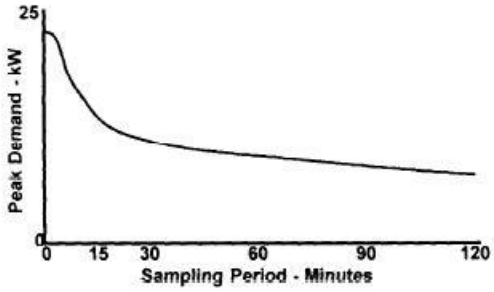


Figure 2.11: Measured peak demand of a single residential customer varies greatly depending on the intervals used to sample its load

Figure 2.11 shows peak demand for the data in Figure 2.4, plotted as a function of period integration sampling rate. The measured peak load decreases as the sampling period increases. The reason is that the sampling rate, or demand interval, defines the meaning of “peak”. Sampled at one-minute intervals, the peak is the maximum 60-second demand. Sampling on an hourly basis smoothes the needle peaks, and generates a curve whose peak is the maximum one-hour demand. A non-coincident curve (top of Fig. 20) can look like it was smoother and very “coincident” simply because it was demand-sampled at too low a sampling rate.

For what concerning the *instantaneous sampling*, it has a far different interaction with sampling rate and recording accuracy than the period integration method discussed above. Figure 2.12 shows the load for a single household (Figure 2.4) measured by instantaneous sampling on an hourly basis. One profile is the result of sampling instantaneously every hour, on the hour. The other is still sampled hourly, but a quarter past the hour. The apparent load profile shape, and peak load of these two curves are deeply different. Neither is an accurate representation of the actual load profile behavior.

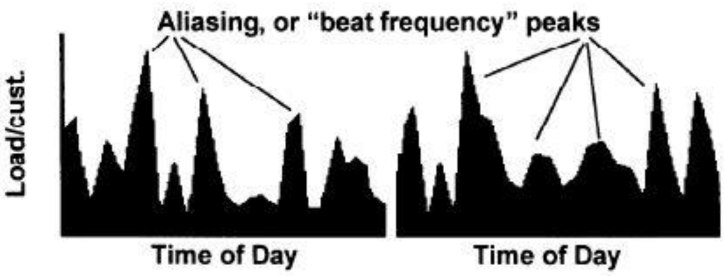


Figure 2.12: Single household load profile sampled with hourly discrete sampling. At the left, the load profile sampled discretely every hour at the beginning of the hour; at the right, the load profile sampled discretely every hour 15 minutes after the hour.

In this case, the problem is that the sampling rate is too much slow to correctly study the load, which presents the erratic on-off behavior typical for non-coincident loads. Both curves in Figure 2.12 are affected by *aliasing,* a “beat frequency” generated by interference between sampling rate and duty cycle rate of appliances. In order to avoid aliasing and to be entirely valid, this method should have a minimum sampling rate equal to the Nyquist rate, which requires that the load is sampled and recorded at least twice as often as the fastest load shift occurring in the measured load.

2.4 Coincidence factor relationships

Bary [24] demonstrated the existence in practice of two fundamental relationships between coincidence factor and:

- A change in the number of consumers in a group
- A change in their individual load factor

These relationships are of fundamental significance, as they can find applications in engineering and economic studies dealing with load forecasting for system planning, design of distribution systems, or provision of service facilities.

Major factors affecting coincidence

The degree of coincidence in the use of electric appliances by individual consumers or group of consumers is determined by many factors. *Population habits and community and business practices* exert a powerful influence upon the degree of coincidence on use of electric appliances: when people go to bed, when they get up, when and what they eat, where and how they live, where and when they work, are all aspects that condition the degree of coincidence. *Weather and climatic conditions* also have a strong influence on the coincidence in the consumers’ use of electric appliances: for example, the advent of darkness because of a storm during light-day hours may create a full coincidence of the lightings loads, or a cold spell may cause an increase of utilization of heating plants. The design of utilization equipment has a considerable influence upon the degree of coincidence: devices designed for full automatic operation under intermittent cycles usually have lower coincidence factors than those designed to accomplish the same purpose under manual operation.

Relationship between coincidence factor and number of consumers

“For consumers of equal size, with no artificial restrictions placed in the way of their service use, and with all other things equal, the degree of coincidence of the maximum power requirements

of a group of consumers decreases along a path resembling that of a rectangular hyperbola from unity for one consumer in a group to values approaching the asymptote of the rectangular hyperbola for an infinite number of consumers in a group" [24].

The equation of the rectangular hyperbola is reported below:

2.9

$$Cf(N) = Cf(\infty) + (1 - Cf(\infty))\frac{1}{N}$$

in which Cf(N) represents the coincidence factor for N customers, and Cf(∞) represents the coincidence factor for infinite customers. The results of the observations performed by Bary are shown in Figure 2.13: the observations for various type of consumers generally follow the path of the rectangular hyperbola reasonably well. Moreover, the convergence of the relationship towards the asymptote is rapid for all the cases analyzed, indicating that beyond 100 customers in a group the change in coincidence factor is relatively small, and results of sufficient engineering stability can be obtained from groups of only 30 consumers.

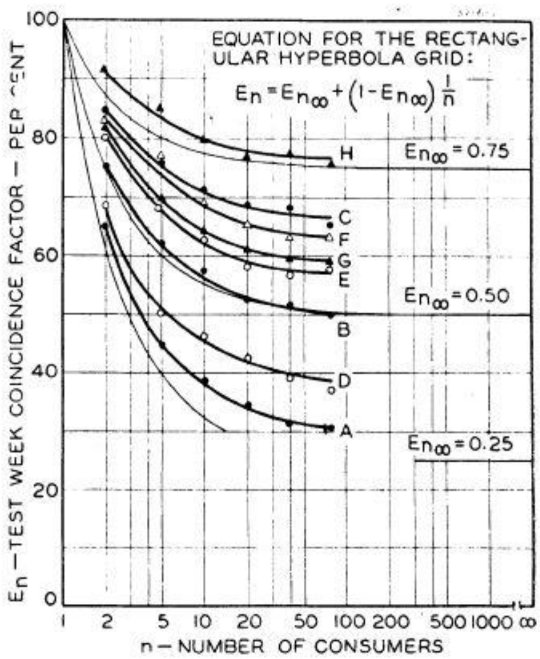


Figure 2.13: Trend of group coincidence factor as function of number of consumers in a group. Data refers to measurements recorded by period integration based on 30-minute basis, in December 1938 in Philadelphia, for various user classes. (A, B, C are relative to domestic users without electric stoves, respectively having 12%, 23%, 31% load factor; D, E are relative to domestic users with electric stoves, respectively having 10%, 16% load factor; F is relative to industrial users having 17% load factor; G, H are relative to commercial users respectively having 10%, 16% load factor)

Relationship between coincidence factor and load factor

It is obvious that a group of consumers, each having a 100% load factor, has a coincidence factor equal to unity. Furthermore, it can be demonstrated that a group of many consumers, each having a load factor approaching to zero[1], will have a coincidence factor which approach the value of one divided by the number of consumers in a group. Fixed these two extreme values, it may be expected that there will exist a relation which connect these two points. Finally, it may be demonstrated that the coincidence factor of a group of customers, each having the same load factor, can't drop below the load factor value, and obviously by definition it can't exceed the value of unity. Once set these limits, we can conclude that the relationships should lie between these limits marked AB and CB in Figure 2.14. Lines BD, whose points are equidistant from the limits, BD' and BD" descript possible different paths of the relationship.

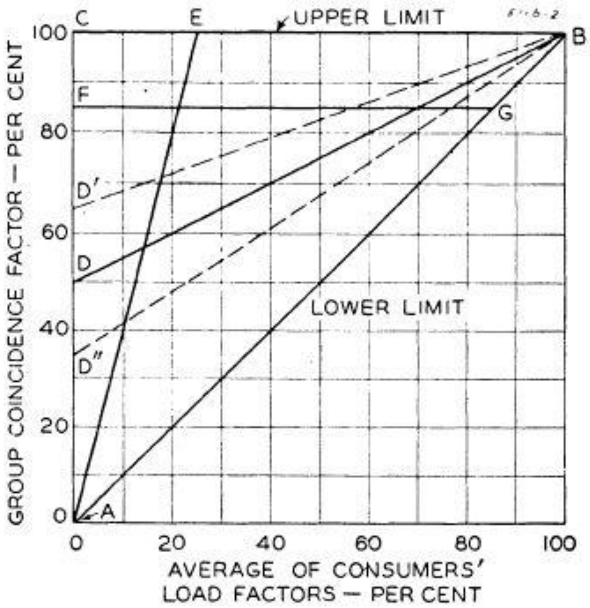


Figure 2.14: Illustration of some guidelines on chart of Cf vs Lf

Any straight line traced from the origin upward and above the lower limit line AB, such as AE, represent the path of the relationship for the theoretical condition in which the predominant use of service for a large group of consumers is confined in a limited portion of the period considered (called *high load probability period*), so that the major changes in the consumers' load factors are accomplished through changes in energy use in that specific period. The shorter this period, the steeper the slope of the line representing this case. Line AC represents the shortest period or the

[1] A consumer having a load factor approaching to zero is for example a consumer which energy demand is always null, except for a peak demand for a very brief instant of time.

full coincidence of peak demands for all customers in the group, while line AB, called “best possible line”, in addition to being the lower limit of relationship, also represents the longest period of high load probability, when the loads of all customers have the same probability to occur throughout the entire period considered. A horizontal line, such as FG, drawn from the ordinate towards the line AB represents the path of Cf-Lf relationship when the change in customers’ load factor is accomplished completely outside the high load probability period.

Bary conducted the research over a period of many years in establishing the general path of the relationship between Cf and Lf. The observations were obtained for the month of December over a period between 1936 and 1943, using *period integration* on a 30-minute basis, and are relative to an electric utility system which served metropolitan, suburban and rural territories in the eastern part of US (near Philadelphia). Consumers studied were arranged into groups, each falling into specified narrow load-factor limits, and each consisting of 30 customers, a sufficient number to produce significant results of coincidence factors. The results of the research are shown in Figure 2.15.

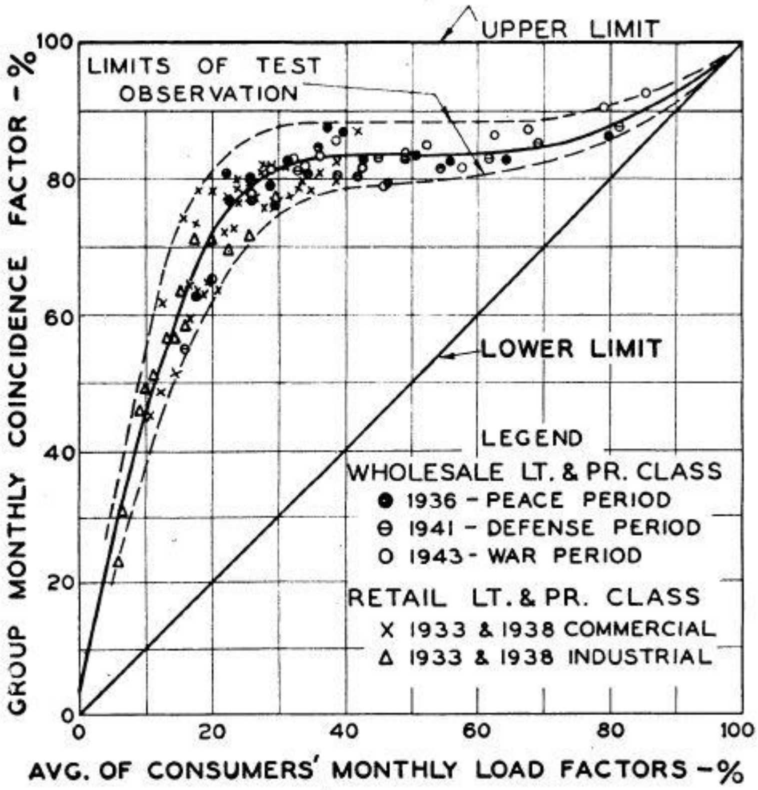


Figure 2.15: Empirical trend in the relation between coincidence factor and load factor, resulting from research conducted by C. Bary over a period of many years, for different user classes.

As the observations were available for the entire range of load factors, a complete trend of the relationship for the industrial and commercial customers could be traced. The trend of the relationship can be divided into three distinct regions. The first in which Cf increases rapidly from a value close to zero[2] with increasing load factor, the second in which contrariwise Cf remains almost constant, and the last phase in which turns up sharply to reach the “best possible line” in the ultrahigh load factor range.

- Phase one, relative to the low load factor range, indicates that the predominant energy use is confined to a certain well-defined portion of the entire period under consideration, and changes in the load factors are accomplished through changes in the energy use almost entirely in that period (as in the theoretical line AE in Figure 2.14). Hence, in this load factor region the coincidence factor rises rapidly with the increasing of load factor.
- Phase two, obtained in the medium load factor range, indicates that beyond individual consumers’ monthly load factors of 30%, a further improvement in their load factors is accomplished almost entirely outside the period of the high load probability of the group. Thus in this region the coincidence factor remain substantially constant, and its behavior with respect to load factor approximates that depicted by line FG in Figure 2.14.
- Phase three, related to the ultrahigh load factor range, appears to indicate that over an 80% monthly load factor occurs a “saturation” of time left, as the entire period under consideration becomes the period of high load probability.

Basing on the foregoing discussion it is now possible to formulate qualitatively the coincidence factor vs load factor relationship as follows:

“For a large number of consumers, of equal size, with no artificial restrictions placed in the way of their service use, and with all other things equal, the degree of coincidence of the maximum power requirements of individual consumers in a group changes with the change in their individual factors. In the low-load-factor range the coincidence factor increases rapidly with the increase of the individual consumers’ load factors. Its rate of change is substantially constant and is inversely proportional to the relative length of the high load probability of the class. In the medium range of load factors the magnitude of the group coincidence factor remains substantially fixed, until very

[2] As each group consisted of 30 consumers, the starting point of coincidence factor at load factor approaching to zero is 1/30 = 0.0333.

high load factor values are reached, when the relation between load factor and coincidence factor approaches the lower limit of the possible relationship" [24].

A mathematical formulation of the relationship seen above has been proposed by C. Bary, in which the coincidence factor has been expressed as a function of both load factor and number of consumers in the group. From theoretical studies by R.F. Hamilton [13] [14] has been found that:

2.10

$$Lf_Q = \left[a + \left(\frac{1}{Lf} - a\right) Q^{-\frac{1}{2}}\right]^{-1}$$

In which Lf_Q is the load factor for the sum of Q loads of load factor Lf each, and a is the asymptote of the curve $1/_{Lf_Q}$ vs Q shown in Figure 2.16.

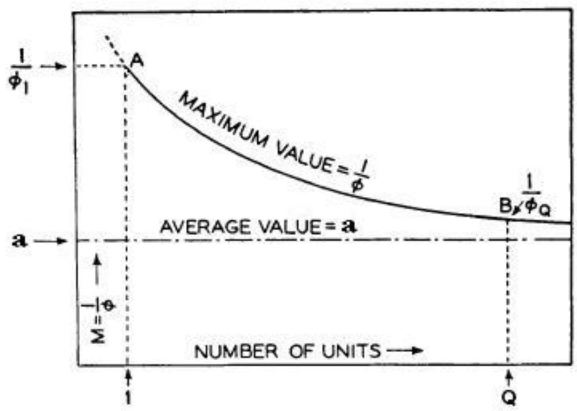


Figure 2.16: Curve of 1/Lf(Q) as a function of Q (number of consumers for which the load factor is calculated). The load factor of a group of consumers decreases as the number of consumers in the group increases in the same way that happens to the coincidence factor.

In addition, since the coincidence factor Cf_Q is the ratio of the peak for Q loads to Q times the peak of one single load, we can write, taking the peak for one load as unity:

2.11

$$Cf_Q = \frac{Peak_Q}{Peak_1} = \frac{\boldsymbol{Mean}}{Peak_1} * \frac{Peak_Q}{\boldsymbol{Mean}} = \frac{\frac{Mean}{Peak_1}}{\frac{Mean}{Peak_Q}} = \frac{Lf}{Lf_Q}$$

where $Peak_1$ is the peak demand of a single consumer, $Peak_Q$ is the peak demand of a group of Q consumers, and *Mean* is the average demand. Substituting the value of Lf_Q with the one of Eq. 2.10:

2.12

$$Cf_Q = a * Lf + (1 - a * Lf) * Q^{-\frac{1}{2}}$$

where *a* is the value of the asymptote of the curve $^1/_{Lf_Q}$ vs *Q* previously introduced (Figure 2.16), and *Lf* is the load factor of a single user. The formulation used to obtain the correct value of *a* is the following:

2.13

$$a = \frac{1}{p}\left[1 - (1 - p)^{\frac{1}{Lf}}\right]$$

where *p* is the ratio between the time in which peaks of the users may occur and the period considered, i.e. it is a parameter that gives an idea of the amplitude of the range in which peaks can occur. It seems logical to presume that the more energy the users consume the more likely is the time of his peak-load to vary. Stated in another form, when a substantial number of customers with about the same load factor have been considered, it is found that the time in which the peaks occur can be correlated with the load factor. According to empirical results from the study of Bary [28], such a correlation may be expressed:

2.14

$$p = 0{,}187 + 0{,}813 * e^{-4\left[(1-Lf)^2 + (1-Lf)^{16}\right]}$$

The empirical relationship presented cannot properly fit any context, because its shape is determined largely by specific factors as population habits, community and business practices, weather and other climatic conditions, types of appliances installed, and others. It is the result of observations made in a specific context (residential and commercial users in Philadelphia between 1936 and 1943), when surely consumers' habits were different and electrical equipment installed were different both in number and in type than the present. Therefore, applying this relationship to the present day is definitely a very strong assumption. However, due to the absence of alternative solutions and of similar studies on the load profiles in the literature, this relationship has been implemented in the model developed in this work. In addition, if this empirical relationship may not be adequate to fully electrified developed countries where users have access to various and modern electric facilities, perhaps could better adapt to consumers of just electrified rural areas, where the electrical equipment installed is likely to be more limited.

3 Approaches for load profile estimate in off-grid systems: state of the art

Load profile estimation falls within the research field of energy needs estimation. This topic can be divided into two main themes each one having a different purpose:

- Power system engineering refers to *Load Forecasting (LF)* as the domain of models able to provide data for setting the best operating and planning of power systems ([29], [30],[31]).
- Energy planning research refers to *Residential Energy Consumption Modeling (RECM)* with those models able to support energy-related policy decisions ([32]).

In this chapter an overview of the loads profile estimate methodologies available is presented. Initially models used for *LF* and *RECM*, in particular used in developed countries. It is then placed greater focus on applications for off-grid rural electrification: after verification, through research in modern scientific literature, that actually do not exist models for load estimation specifically developed for use in rural areas of developing countries, the main solutions actually adopted are described.

3.1 Load Forecasting

Electrical load forecasting is an important tool used to ensure that the energy supplied to users meets the load, including the energy lost in the system. Load forecasting is defined as the science of predicting the future load on a given system, for a specified period ahead, which can range from a quarter of an hour up to 50 years. Load forecasting can be classified according to the forecast period:

- *Short-term load forecasting*, which is used to predicts loads up to a week ahead.
- *Medium-term load forecasting*, which is used to predict weekly, monthly, and yearly peak loads up to 5 years ahead.
- *Long-term load forecasting*, which is used to predict loads as distant as 50 years ahead.

Load forecasting is not only important to provide accurate estimates for the operating of the power system but also as a basis for energy transactions and decision making in energy markets, load switching, and infrastructure development. Load forecasts are extremely important for electric utilities, energy suppliers, financial institutions, and other players of energy generation, transmission, distribution and markets. However, a more specific analysis of each category of load forecasting will allow to better understand the purposes for which such models are used.

Short-term load forecasting

Short-term load forecasting (STLF) have become increasingly important since the rise of the competitive energy markets. Many countries have recently privatized and deregulated their power systems, and electricity has been turned into a commodity to be sold and bought at market prices. STLF plays a key role in the formulation of economic, reliable, and secure operating strategies for the power system: short-term forecasts are used for unit commitment, energy transfer scheduling and load dispatch.

In STLF, the future load on a power system is predicted by extrapolating a predetermined relationship between the load and its influential variables. Determining this relationship consists in a two-stage process that requires (a) identifying the relationship between the load and the related variables, and (b) quantifying this relationship using a suitable parameter estimation technique. A prerequisite to the development of an accurate load forecasting model is understanding the characteristics of the load to be modelled [2-6]. The load supplied by a power system can be divided into three components: a *base load*, a *weather dependent load*, and a *residual load*.

The base load, which results from the business and economic conditions of the service area, accounts approximately 90% of total load. It is composed by a long-term component which reflects the economic growth of the considered area, a seasonal component that results from changes in electric demand between different seasons, a weekly load cycle which considers the consumption pattern of one day being characteristically different from the others, and finally a daily load cycle that results from the basic daily similarity of consumer activities.

The weather contributes significantly to the dynamics of a load, and so much effort has been spent to understand the relationships between weather conditions and load behavior. The effects of the weather on the load are modelled by expressing the load as a linear regression of meteorological factors such as temperature, wind speed, humidity, and daylight illumination.

The residual load component occurs in load modeling and usually accounts for a small percentage of total load, and results from irregularities in the load behavior. These abnormal demands, though quite frequent in occurrence, are very difficult to predict, as they range from public response to major television events, strikes, storms, disasters, time changes, etc.

STLF methods indicates that the most important modeling techniques used can be classified in five main categories:

- Multiple Linear Regression (e.g. [33], [34]).

- Stochastic Time Series (e.g. [35], [36]).
- General Exponential Smoothing (e.g. [37]).
- State Space Method.
- Expert Systems (i.e. Artificial Intelligence) Approach (e.g. [38], [39]).

These models are classified according to the mathematical technique used to estimate the parameters of the models, and a detailed explanation of each model category can be found here [30].

Medium-Term load forecasting

Medium-Term Load Forecasting (MTLF) is used to predict weekly, monthly and yearly peak loads up to 5 years ahead, which provides utilities information to better plan power generation expansion (or purchase), schedule maintenance activities, perform system improvements, negotiate forward contracts and develop cost efficient fuel purchasing strategies. Unfortunately, it is difficult to forecast load demand accurately over a planning period of this length. This fact is due to the uncertain nature of the forecasting process. There are a large number of influential factors that characterize and directly or indirectly affect the underlying forecasting process: all of them are uncertain and uncontrollable.

Long-Term load forecasting

To study facilities or other options in detail over a period from few years up to fifty years ahead, the utility’s planners must have a good idea of the conditions under which that equipment will function. They need a long-range plan to provide a backdrop against which they can evaluate the value of their short-range projects. Long-term load forecasting plays an important role for system planning, scheduling construction of new generating capacity and purchasing of generating units. Because it takes several years and requires a great amount of money to construct power generation capacity and transmission facilities, accurate forecasting is necessary for an electric company. An extensive overestimation of load demand will result in substantial investment for the construction of excess power facilities, while underestimation will result in customer discontentment.

By nature, Long-term electric load forecasting is a complex problem. Among other factors, its accuracy is extremely influenced by the weather as well as social behavior of the community of that load. These factors are difficult to predict for long-term load forecasting time horizon. Conversely, short-term forecasting, though affected by weather and daily social habits, the weather and social habits fluctuation for the short-term time horizon is small enough to predict load with high accuracy. However, during the last two decades many techniques have been developed to improve the long-/mid-term forecasting accuracy. They can be generally divided into two broad categories: *parametric methods* and *artificial intelligence-based methods* [31]. The *parametric methods* (examples are [40]–[42]) construct a mathematical or statistical model of load by examining qualitative relationships between the load and the load-affecting factors, and the assumed model parameters are estimated from historical data. However, because of being designed to capture the nonlinear relationships between inputs (historical load, affecting factors) and outputs (forecasted load) through an explicit formula, the mathematical complexity does not offer to the user an intuitive understanding. Recently, the artificial intelligence-based methods have been proposed as a valid substitute: the advantages of these methods over statistical methods include the consideration of robustness of load prediction system, the ability to classify in the presence of nonlinear relationships. Some examples of application of AI-based methods are [43]–[47]

3.2 Residential Energy Consumption Modeling

The world energy consumption of the residential sector accounts for approximately 30% of that consumed by all sectors. In response to climate change, high energy prices, and energy supply/demand, there is interest in understanding the detailed consumption characteristics of the residential sector in an effort to promote conservation, efficiency, technology implementation and energy source switching, such as to renewable energy. Energy consumption of other major sectors such as commercial, industrial, agriculture and transportation are better understood than the residential sector due to their more interest and expertise in reducing energy consumption, and high levels of regulation and documentation. Contrariwise, the residential sector is largely an undefined energy sink, because it includes a wide variety of users, structures, appliances.

Energy consumption models seeks to quantify residential energy requirements, as a function of input parameters, for a country, a region or a sector. Models of this nature are useful as they can guide decisions on energy policies: by quantifying the consumption and predicting the impact due to retrofits and new materials and technology, decisions can be made to support energy supply, technology incentives, new building code, or even demolition and re-construction.

Techniques used to model residential energy consumption can broadly be grouped into two categories, *top-down* and *bottom up*, with reference to the hierarchal position of data inputs (Figure 3.1). Top down models utilize the estimate of total residential sector energy consumption and other pertinent variables to attribute the energy consumption to characteristics of the entire housing sector. In contrast, bottom-up models calculate the energy consumption of individual or groups of houses and then extrapolate these results to represent the region or nation.

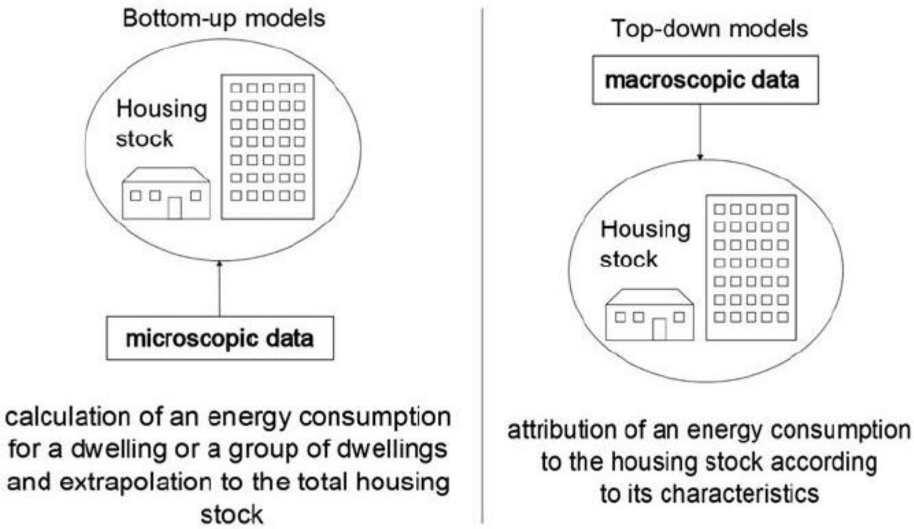


Figure 3.1: Bottom-up VS Top-down models

Top-down approach

The top down approach treats the residential sector as an energy sink and does not distinguish the energy consumption by individual end-uses. These models determine the effect on residential energy consumption due to long-term changes or transitions, for determining supply requirements. Top-down models bases their estimations on both historical energy consumption and input variables which include macroeconomic indicators, such as GDP or employment rates, technology component, such as technology development pace, housing component, such as housing contraction/demolition rates, and even climatic conditions.

Top-down models are relatively easy to develop due to their need for only aggregate data, which are widely available, and rely on historical data, but they falter when discontinuity is encountered, for example in situations of technological breakthrough: they provide good prediction capability for small deviations from the status quo. The reliance on historical data is another weak point of these model, because they cannot take into account discontinuous advances related to eventual new technologies. Furthermore, the lack of detail regarding the energy consumption of individual end-uses eliminates the capability of identifying areas for improvements, making the development of policy or incentives impossible.

Bottom-up approach

The bottom-up approach is developed to identify the contribution of each end-use over the aggregate residential energy consumption. These models compute electric demand for a few modelled households which are representative of a larger scale, and the unitary results are then extrapolated to obtain the electricity demand for the entire studied geographical scale, or the entire residential sector. The required input data may concern individual consumption of appliances and their technical properties, climate properties, dwelling geometry and characteristics, occupancy schedule and human behavior. Bottom-up models can be grouped in 2 main categories: *statistical methods (SM)* and *engineering methods (EM).*

Statistical methods

SM rely on historical dwelling energy consumption from a sample of houses and types of regression analysis, which are used to attribute the energy consumption to particular end-uses. SM models can utilize macroeconomic, energy price and income, and other regional or nation indicators, thereby gaining the strengths of simplicity of the top-down approach. Typically, the primary information source of the bottom-up SM is energy supplier billing data and other data collected by simple surveys; by disaggregating measured energy consumption among end-uses, occupant behavior can be accounted for. The reliance on historical consumption information is the principal weakness of these models, because it does not allows to consider the impact of new technologies.

Engineering methods

EM rely on information of the dwelling characteristics and end-uses themselves to calculate the energy consumption based on power ratings, use of equipment, heat transfer and thermodynamic principles, and request more detailed information compared to the SM. The high level of detail of input data requested gives the ability to model technological options, allowing taking into account even new technologies. These models explicitly calculate or simulate the energy consumption and do not rely on historical values. One of the most critical point of the EM is the assumption of occupant behavior, because it can significantly affect the estimated energy consumption. Another drawback is the high level of expertise required in the development and use of these models.

In conclusion, each approach meets a specific need for energy modeling which corresponds to its strongest attribute:

- Top-down approaches are used for supply analysis based on long-term projections of energy demand by accounting for historic response.
- Bottom-up statistical techniques are used to determine the energy demand contribution of end-uses inclusive of behavioral aspects based on data obtained from energy bills and simple surveys.
- Bottom-up engineering techniques are used to explicitly calculate energy consumption of end-uses based on detailed descriptions of a representative set of houses, and these techniques have the capability of determining the impact of new technologies.

A recent review of models to perform estimates of daily load profiles within the residential sector, has been published by Grandjean et al. [23], in which 12 domestic load profiles models are revised and described.

3.3 Load estimation for off-grid power design

The research field of the estimate of energy consumption and the forecast of load profiles is a topic of great interest. However, despite the large number of scientific papers that have addressed these themes, most of them deal with the particular case of domestic electric consumptions in Developed Countries and they are mainly devoted to support decisions on load dispatch, unit commitment, maintenance scheduling, system planning (i.e. short, medium, long-term LF methods previously introduced), or decisions as regards energy policies (i.e. RECM methods). Despite estimation of the electricity demand is absolutely necessary in the design process of off-grid power systems for rural electrification, to the author knowledge a dedicated section within the scientific literature is not present. This is even true when looking at the literature which poses focus on specific sizing methods and models rather than on their applications, but also when looking at feasibility or prefeasibility analyses. Only Celik [48] brought about the issue of load profiles in systems sizing.

In order to understand how the problem of estimating the electrical load is tackled in electrification projects of un-electrified areas, it has been performed a research into the literature for articles in which were presented case studies related to sizing and installation of off-grid systems for the electrification of such areas, which typically consist of villages or groups of houses away from the main grid (if present), and for which the grid extension would not be a cost-effective solution. In most of the projects considered, advanced sizing and simulation methods were used to design the off-grid system, and often optimizations were performed by specific software such as HOMER.

The results of my research underlined that generally the problem of estimation of load profiles in off-grid systems projects is faced in three manners:

- adapting load profiles from similar contexts (e.g. [19], [49]–[55]);
- computing the load profiles without any defined procedure, employing assumptions on electric device power rates and use behavior, or from surveys, in order to build up a load profile (e.g. [56]–[65]).
- load profiles are defined without any explanations about the origin (e.g. [66], [67]);

Load profiles taken from similar contexts

References on similar experiences are useful, although local factors have a deep influence. Basing on experiences of similar contexts means use load profiles measured in rural villages previously electrified similar to that of the case study. Usually, in these cases are chosen measured data of users or buildings that are considered representative for the case analyzed. Nfah et al. [49] considered data of 19 grid-connected households in a village in Uganda located near the capital city Kampala for their study of rural electrification for remote villages in Cameroon. Shaahid et al. [52] used the measured annual average electric energy consumption of a *typical remote village* (unspecified source) as yearly load for their study of rural electrification in Saudi Arabia. It is difficult to state whether the data used in these cases were appropriate: the profile of the electrical load is influenced by many factors, some exogenous, such as temperature and climate, and other closely related to the population, as their habits, their way of life, their social conditions, their economic level, etc. Hence, using load profiles of a rural village previously electrified in similar climatic conditions, and with the same number of users, not necessarily lead to satisfactory results in terms of estimated load. Remote rural villages may have very different energy demands compared to villages near cities, even if the living conditions and the economic level of the populations are very similar. In conclusion, it is necessary a detailed analysis of the context to be electrified. In conclusion, careful analysis of area to electrify and users to satisfy are required, because a general load profile of the "typical rural village" that can be adapted to all contexts does not exist.

Load profiles estimated

Frequently the designers of the projects, starting from the input data that include information about the applications and their use, draw the load profiles.

Sometimes, input data are compiled making assumptions about which devices will be installed, about their power rates, and about their usage pattern. For example, Kanase et al. [57] in their project of rural electrification of a remote area including a cluster of villages, classified the electrical energy demands as domestic (which includes electrical appliances such as TV, radio, fan, fluorescent lamp, etc.), agricultural (fodder cutting and threshing machines), community (schools and panchayat offices) and rural industries (small milk processing plants and milk storage), and finally the total energy demand has been somehow (unspecified) estimated. Mahmoud et al. [65] performed a techno-economic feasibility study of energy supply of a remote village in Palestine, in which they assumed the main electrical loads necessary for improving living conditions in the village: household appliances (lighting, TV, refrigerator, radio, washing machine and fan), street lighting (sodium lamps), school appliances (lighting, educational TV and lab equipment), clinic appliances (lighting, refrigerator and lab devices).

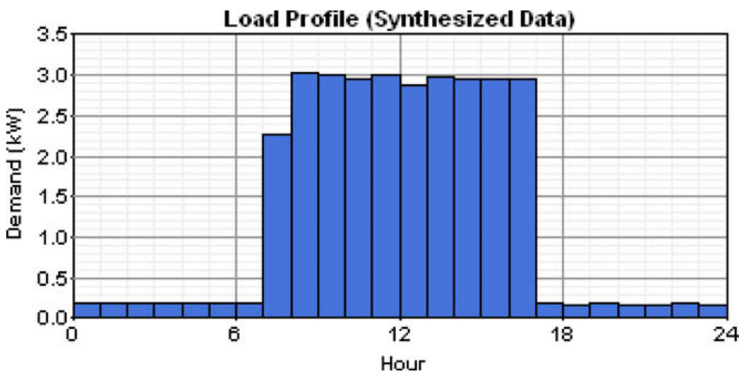


Figure 3.2: Example of estimated load profile of a health clinic in a rural area [58]*, with the assumption that the majority of the load occurs at the day time.*

Other times energy demand is calculated on the basis of data collected from surveys proposed to the population ([60], [63]): these cases highlight a greater attention, as data from surveys will be more accurate than simple estimates of designers. Camblong et al. ([59] [60]) presented a project which aim was to promote the electrification of rural regions of Senegal by the installation of micro-grids with high content of renewable energies. Surveys have been carried out in three regions of Senegal to study the needs of electrical energy of non-electrified rural villages' households. These surveys have led to the estimation of electricity needs of the concerned households.

Not always are present indications on how data (estimated or from surveys) are used to compute load profiles. It is not rare to find statements like this: "*It was assumed that the remote residential area consisted of a total of 40 houses. Each house required loads of 2 kW peak. Therefore, 40 houses* *would require a maximum of 80 kW peak”* [68]. In reality due to the coincidence of loads, the peak for a group of similar utilities can be calculated in this way:

$$group\ peak\ for\ N\ customers\ [W] = Cf(N) * N * average\ individual\ peak\ load\ [W]$$

Applying it to the case cited above, estimating a value of *Cf* for 40 residential houses of 0.5:

3.1

$$group\ peak\ for\ 40\ houses = 0.5 * 40 * 2\ [kW] = 40\ [kW]$$

a value much lower than the value estimated by the designers (80 kW).

Despite no formal specific procedures are defined in literature, two main approaches that are *unintentionally* employed can be formalized. They are briefly described in the following.

In order to develop the load profile, the definition of some input parameters is required:

- *j* refers to the *type of electrical appliances* (e.g. light, mobile charger, radio, TV);
- *k* refers to the specific *user class* (e.g. household, school, stand shop, clinics);
- N_k refers to the *number of users* within class *k*;
- n_{jk} refers to the *number of appliances i* within class *k;*
- Ft_{jk} refers to the duration of the period the appliance *i* within user class *k* is on [h] (i.e. *functioning time*);
- P_{jk} refers to the *nominal power* rate [W] of appliance *j* within class *k*;
- Fw_{jk} refers to the *functioning window(s)* of a consumer class appliance ($W_{F,jk}$), which represent the period(s) during the day when an appliance *j* of the user class *k* can be on Figure 3.3).

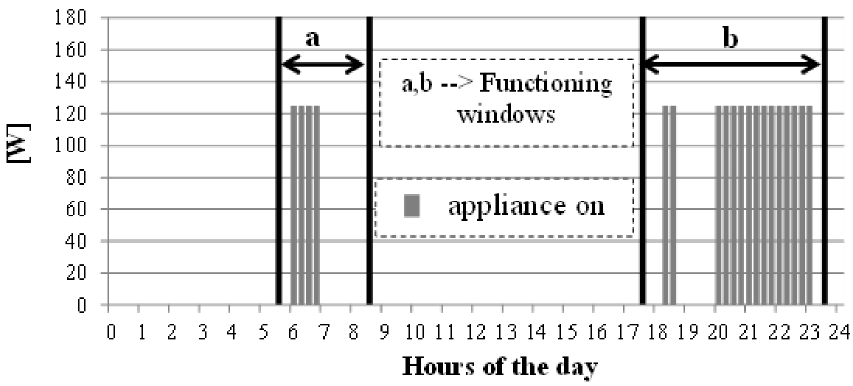


Figure 3.3: Graphical representation of functioning windows for a single appliance.

To compare the results obtained by the two different approaches, they were applied to the same set of input data, related to a peripheral urban area of Uganda (detailed data available in Appendix A). These data have been obtained from a thesis of the same research group (group of the UNESCO Chair in Energy for Sustainable Development at Politecnico di Milano) of this work [69].

The *first approach* (named *setting on-off approach*) that can be formalized is to set directly the periods of operation of electrical appliances installed. According to the previous definition of input parameters, it is based on the following condition:

3.2

$$\sum duration(Fw_{jk}) = Ft_{jk}$$

i.e. the sum of the assumed functioning windows is equal to the functioning time h_{jk}. Accordingly, all the n_{jk} appliances result to be on at the same time. Figure 3.4 shows the load profile resulting from this approach for the assumed load data of a peripheral urban area of Uganda (detailed data available in Appendix A).

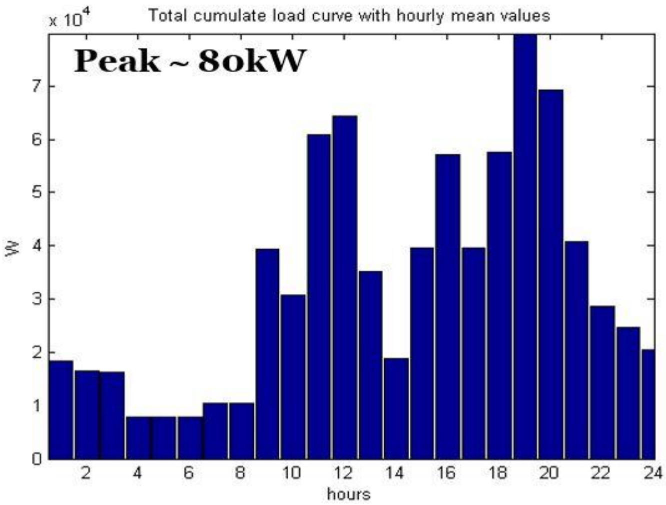


Figure 3.4: Estimated load profile for Uganda case according to the assumption of functioning windows matching the functioning time (first approach)

In the *second approach* (named *medium power approach*) analyzed, the energy consumed by appliances is “spread” over the entire duration of the functioning windows, through the definition of an average power of each appliance. In this case, firstly the energy consumption associated to each appliance *j* within each user class *k* (E_{jk}) is computed (Eq. 3.3), then the average power associated to each appliance *jk* ($P_{av,jk}$) which contributes to build up the load profile is evaluated (Eq. 3.4). Figure 3.5 shows the load profile resulting from this approach for the same case study of Uganda (detailed data available in Appendix A).

3.3

$$E_{jk} = h_{jk}[h] \; x \; P_{jk}[W]$$

3.4

$$P_{av,jk} = \frac{E_{jk}}{\sum duration(Fw_{jk})}$$

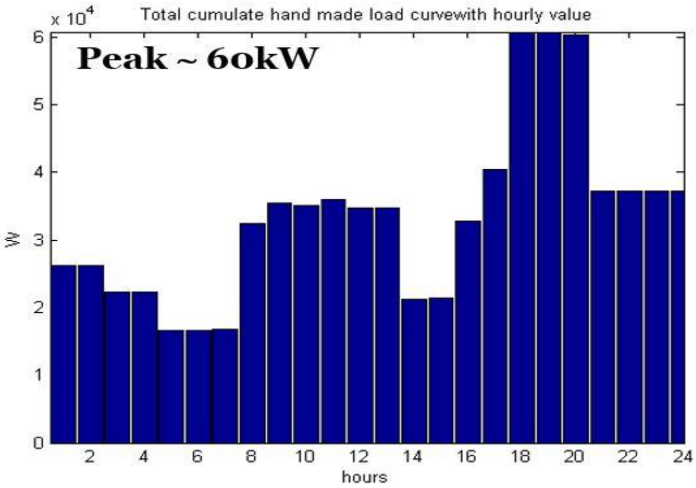


Figure 3.5: Estimated load profile for Uganda case according to the assumptions of load factors and functioning windows (second approach)

Considering the two approaches and looking at the resulting load profiles for the specific case, some considerations can be made:

- Despite both profiles account for the same daily energy consumption, the shapes are quite different. Looking at the peak loads the values as well as the time are different;
- In the first approach, it is defined how long and when the appliances are on without considering any coincidence behavior. Indeed, according to the leading equation (3.2), all appliances *jk* are on at the same time and hence the coincidence of such devices is equal to one. Thus the peak power is overestimated, and in general the profiles are not flat, but have high values and low values;
- In the second approach, input load factors are found in literature or more often are assumed. Then, the power contribution of each appliance *jk* to the load profile refers to the average power computed by "spreading" the consumed energy on the total duration of the functioning windows. In this way the coincidence factor assumes the minimum value

possible given by the input load factors and functioning windows. Hence, the power peaks are underestimated, and in general the profiles are flat.

As a consequence the two approaches are not satisfactory to perform a proper off-grid system sizing, indeed:

- No attention is devoted to appropriately compute the coincidence factor values whose results do not relate to the number of consumer considered and to the load factors. Overestimates and underestimates of power peaks can deeply affect the system sizing;
- To assume values of load factors (in the second approach) is not the most proper approach. Indeed: literature values mostly refer to developed countries context and in any case, assuming load factor values, is an harder task than assuming functioning time;
- They do not embrace uncertainty of load demand which is typical in particular for Developing Countries and rural electrification actions. In both approaches, the input data allow computing only a single load profiles, but not different stochastic profiles within the same input parameters. A random noise may be considered to add uncertainty, but the coincidence appliances behavior is not considered as well.

4 LoadProGen: a tool to perform load profile estimates for off-grid systems

At the light of the issue about lack of appropriate methods to estimate load profiles for supporting the off-grid system design process, one of the activities of the group of the UNESCO Chair in Energy for Sustainable Development of Politecnico di Milano has been the development of an algorithm for estimating the load profiles of un-electrified areas. The software, named LoadProGen (Load Profile Generator) has been implemented in Matlab. One of the tasks of this thesis was to improve the Matlab code and make substantial changes to the algorithm, in order to introduce additional functionalities and increase the execution efficiency. However, the main changes I performed on the code are summarized in the next chapter. In the following, LoadProGen is introduced as regards: the general features, the required input data, and the mathematical formulation.

4.1 General features

In setting the frame to develop the new procedure, the characteristics of an ideal method for load profiles estimate, as introduced by Grandjean et al. [23], have been taken as a reference. In their opinion, an ideal model should present the following features:

- 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 evolutive, 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, area, city, region, etc.);
- all end-uses can be considered in the load profile calculations.

At the light of this reference, the model has been developed in order to embrace the following features:

- it has to be based on input data that can be assessed or assumed by looking at the typical conditions of rural areas;
- it has to be based on a rigorous mathematical formulation which allows developing the load profile shape, 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 computation has to rely on microscopic input data referring to each appliances features within a specific type of user class;
- it has to build up the coincidence behavior of the appliances and the power peak value with regards to the existing correlation between number of users, load factor and coincidence factor;
- it has to be stochastic in order to embrace uncertainty, i.e. given the input data, the procedure output has not to be unique, but it should embrace the uncertainty given by the fact that a single correct load profile does not exist in rural areas.

Therefore, according to the classification made by Grandjean et al. [23], the LoadProGen falls within the *Bottom-up Statistical Random models.*

4.2 Input data

The model for load profile estimation is designed to work with few input data that are commonly considered in the simplest approaches for energy need estimations in rural areas, coherently with the data that may be available in the contexts under study. The first requisite required by the model is the identification of different user classes within the population to study.

The data required by the model are listed below. k refers to the specific user class (e.g. small apartment, big house, school, hospital, etc.), j refers to the type of electrical appliances (e.g. fluorescent lamps, television, radio, fridge, etc.), and i refers to the individual device

- N_k refers to the number of users within user class k (e.g. number of schools, number of households, etc.).
- n_{jk} refers to the number of appliances of type j present in each user which belongs to user class k (e.g. number of televisions in each school).
- P_{jk} refers to *nominal power rate* [W] of the appliance jk (e.g. nominal power of TV).
- Ft_{jk} refers to the *daily functioning time* [min, h], i.e. the duration of the period the appliances jk are on, in a day. Ft_{jk} is set for each type of appliance j, not for each single device i: this means that, for example, all TVs (type of appliance: TV) present in all the schools (user class: school) will have the same daily functioning time.
- Fw_{jk} refers to the *functioning window(s)*, i.e. period(s) during the day when the appliances jk can be on (Figure 3.3). Defined by a starting window time and an ending window time, which are set for each type of appliance, similarly to the Ft_{jk}. Each type of appliance can have up to three functioning windows in a day.

- d_{jk} refers to the *functioning cycle* [min], i.e. minimum continuous functioning time once appliance *jk* turns on.
- RFt_{jk} refers to the percentage random variation of the functioning time of the appliance *jk*.
- RFw_{jk} refers to the percentage random variation of functioning window(s) of the appliance *jk.*

As regards the input data, some considerations can be made:

- All the appliances are modelled with on-off functioning cycle and considering a minimum continuous functioning time d_{jk}. The functioning behavior of each appliance is considered by setting different d_{jk} values: for example, a functioning cycle *d* of 30 min can be assumed for the electric oven, while a functioning cycle of 3 hours can be considered for the phone charger. According to this operating cycle, if the *Ft* of the oven has been assumed to be 1 hour per day, it will be switched on 2 times a day for 30 minutes each. In addition, the Ft_{jk} has to be a multiple of the functioning cycle d_{jk}.
- Particular focus needs to be done for those applications where the operating cycle is not controlled directly by the user: an iron, for example, may be used for an hour per day, however it does not consume electricity for all the time during which it is used, but it will work intermittently for short instants required to keep the water in a given temperature range. Similarly operate refrigerators, electric water heater, and many other electrical appliances. According to this fact, for better results it is important that the right value of functioning cycle *d* is set for these applications: the input value of *d* for an iron should be something like 1 or 2 minutes, while its functioning time may be set to 15 minutes (assuming the iron is used 1 hour per day, so that it is *on* for a quarter of the time which it is used).
- In order to consider a degree of uncertainty in the values of Ft_{jk} and Fw_{jk}, which are usually inputs assumed by the designer, random parameters Rft_{jk} and Rfw_{jk} are introduced respectively. They set the maximum percentage of random variation to apply to Ft_{jk} and Fw_{jk}.
- Obviously, the sum of the duration of the functioning windows for an appliance Fw_{jk} is higher then its functioning time Ft_{jk}.
- Given all the input data, the total required daily energy of each user class is defined: indeed, the daily energy demand depends only on the functioning times and on the power ratings of appliances, which are all input data.

$$E_k[Wh] = \sum_j (N_k * n_{jk} * P_{jk}[W] * Ft_{jk}[h])$$

- Given all the input data, a possible *maximum power peak* of each user class is defined. In order to identify the maximum power peak of a user-class, consider all devices of all appliances of the user-class turned-on for the entire duration of their functioning windows. The peak demand of the resulting load profile, which is the theoretical maximum peak according to the functioning windows defined, is the *maximum power peak.* The daily window in which, in the resulting load profile, the maximum power peak is contained is the *peak window* (example in Figure 4.1).

An example of set of input data, relative to an un-electrified peripheral-urban area of Uganda, is shown in Appendix A.

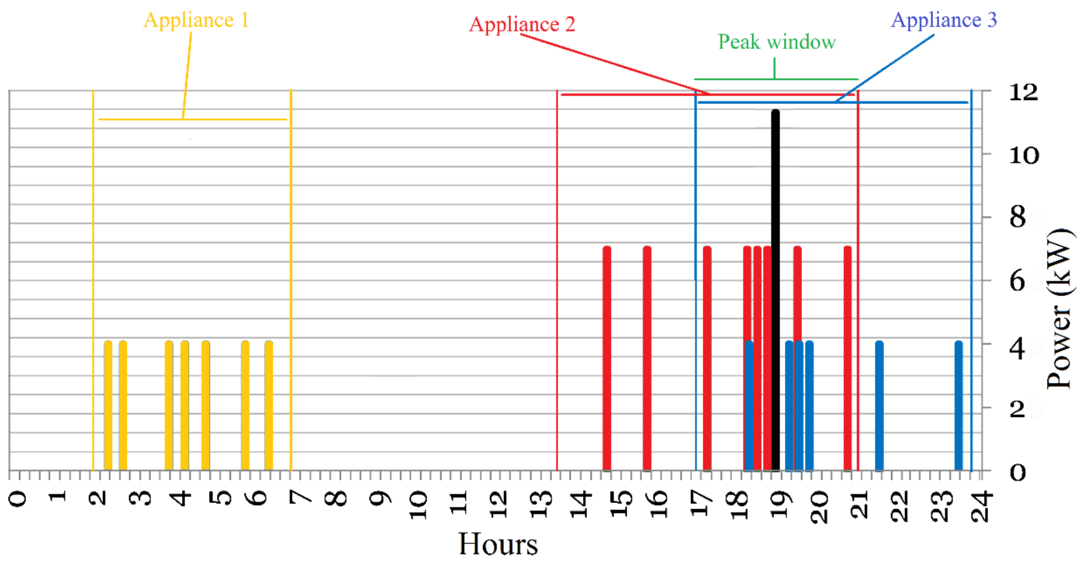


Figure 4.1: Graphical example in which is displayed a load profile of a consumer having three appliances, each represented by a color. These appliances have a single functioning window during the day (indicated by thin lines), and two of them intersect: this means that during this period (the "peak window") is possible that both devices operate in the same instant, then peak occurs (black line).

4.3 Mathematical formulation

The procedure can be formulated according to the following objective function and constraints:

Objective function

The load profile of each appliance of type *j* which belongs to user class *k* is computed by defining, in a stochastic manner, the moments *t* the appliance is switched on. These switching-on instants *t* are sampled from the vector of the daily minutes [1:1440]. Hence, having the moments *t* and the input d_{jk}, the load profile of each appliance is defined. Aggregating the daily load profiles of all the appliances of each type present in a user class, the user class daily load profile is obtained. Finally, aggregating the load profiles of all user classes we obtain the total electric load profile of the users, on a one-minute basis.

Constraints

- The functioning time Ft_{jk} is randomized and computed as follows:

  4.1

  $$Ft_{jk} = Ft_{jk} + random(Ft_{jk} * RFt_{jk})$$

  in which random(Ft_{jk} * RFt_{jk}) refers to the computation of a random value defined in the range: [-(Ft_{jk} * RFt_{jk}); +(Ft_{jk} * RFt_{jk})].
- The number of switching-on moments n_t of each appliance *jk* is defined as follows:

  4.2

  $$n_{t,jk} = \frac{Ft_{jk}}{d_{jk}}$$
- Starting from each instant *t*, the appliance is on for the following d_{jk} minutes.
- The functioning window(s) Fw_{jk}, which define the periods when moments *t* can occur, is defined as follows:

  4.3

  $$Fw_{jk} = Fw_{jk} + random(Fw_{jk} * RFw_{jk})$$

  where random(Fw_{jk} * RFw_{jk}) refers to the computation of random values defined in the range: [-(Fw_{jk} * RFw_{jk}); +(Fw_{jk} * RFw_{jk})]. According to this, the starting and ending instants of the functioning windows are modified to extend or shorten the windows length.
- A *power peak instant* is randomly chosen with uniform probability distribution within the *peak window.*
- The switching-on instants *t* are defined by random sampling within the functioning windows with two possible probability distributions: (i) for appliances which do not contribute to the peak (e.g. appliance 1 in Figure 4.1) sampling is carried out with uniform probability distribution, (ii) for appliances which contribute to the peak (e.g. appliance 2, 3 in Figure 4.1) sampling is carried out with normal probability distribution having mean value on the peak time.

- The parameters of the normal probability distribution for appliances contributing to the peak are defined in order to obtain, within each user class, a power peak value that matches with the peak power obtained via the correlation between coincidence factor, load factor and number of user. The employed correlation refers to the one presented by Bary in [24], which I introduced in Coincidence factor relationships. The coincidence factor is calculated by iteration on the load factor from Eq. 2.12, where the factors *a* and *p* are computed from Eq. 2.13 and 2.14.

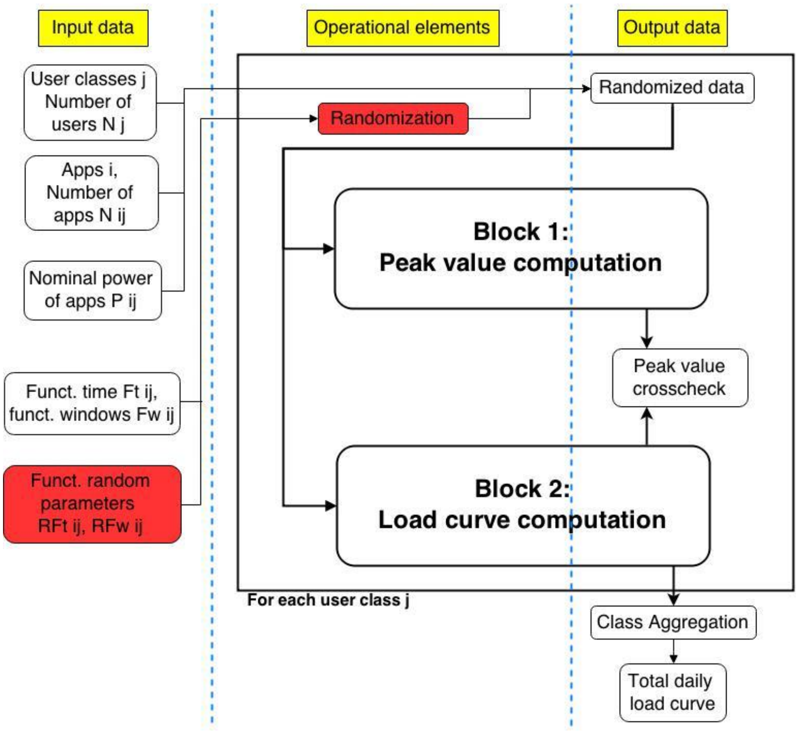
5 LoadProGen improvement: developments and implementation

In this thesis, the software LoadProGen has been deeply improved and applied in two case studies. The algorithm, in its early version, had some limitations and some steps could be developed in a better way: for example, the sampling of switching-on instants of the appliances was rigid, the Load Curve Computation block was slow and inefficient and it operated in unintuitive manner, and the management of the parameters of the normal distribution could be improved. In the following chapter, the main improvements developed in the software are shown: initially the main weaknesses of the early version of the model are summarized, and subsequently the solutions adopted in order to improve the algorithm are explained in detail. Finally, some examples of load profiles generated through LoadProGen are presented and compared with the results obtained using one of the simplest methods founded in literature, explained in Load profiles estimated.

5.1 Improvements identification

A division of the model in macro blocks has been previously presented (Figure 5.7). According to this, one can distinguish a first step which refers to the *processing of input data*, a second step which refers to the *Peak Value Computation*, and a third step which refers to the *Load Curve Computation*. The weaknesses of the early version of the model mainly concerned the first and the third step.

Weaknesses of input data management



In the early version of the algorithm, the introduction of correction factors RFt_{jk} and RFw_{jk}, in order to consider a degree of uncertainty in the values of Ft_{jk} and Fw_{jk} had been already implemented, but the software only worked if those parameters were set to null values. To understand the reason, let's have a look on how the correction factors RFt_{jk} and RFw_{jk} operates on functioning time and functioning windows of each appliance:

5.1

$$Ft_{jk} = Ft_{jk} * \left[1 + random\left(RFt_{jk}\right)\right]$$

in which *random(RFt_{jk})* is a random factor which value is defined between -RFt_{jk}/100 and +RFt_{jk}/100 (RFt_{jk} is in percentage). In the same way:

5.2

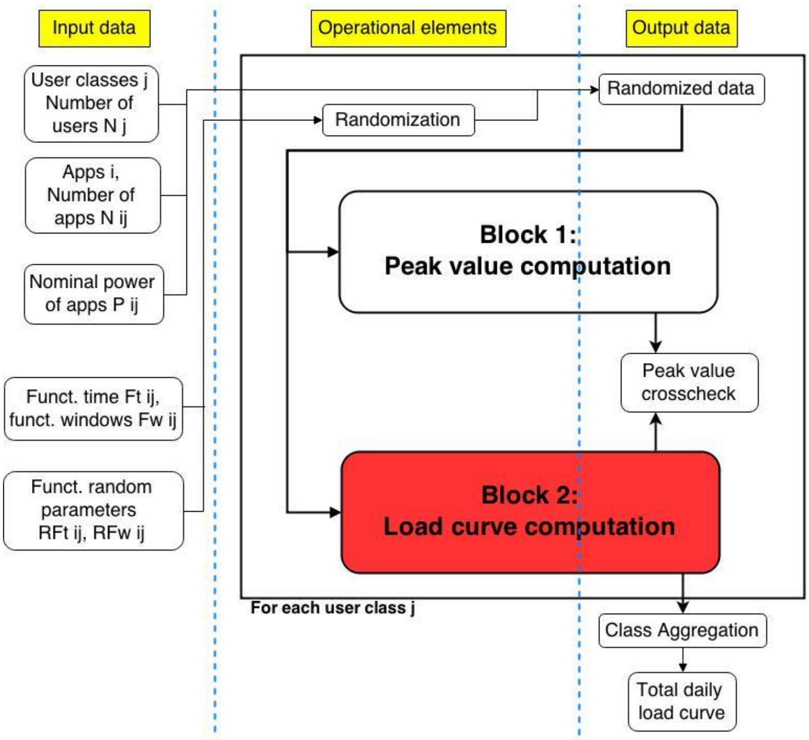
$$Fw_{jk} = Fw_{jk} * \left[1 + random\left(RFw_{jk}\right)\right]$$

According to this, the functioning time and the functioning windows of each appliance are randomly extended or shortened; this caused problems that concerned:

- The definition of the number of app starts $n_{t,jk}$. Since the number of starts of each appliance $n_{t,jk}$ is calculated as the ratio between Ft_{jk} and d_{jk} (Eq. 4.2), in order to have an integer number Ft_{jk} had to be a multiple of functioning cycle d_{jk}, which was not subject to randomization.
- The possibility that, after the computation of the new values of Ft_{jk} and Fw_{jk}, the duration of the functioning time was greater than the total duration of the functioning windows. This would not allow the algorithm, in the *Load Curve Computation* phase, to sample all the switching-on instants for all the devices of the appliance *jk*.
- The possibility that the functioning windows (if more than one) of an appliance *jk* overlapped.

For these reasons, the uncertainty of input data could not be accounted.

Weaknesses of Load Curve Computation



The algorithm, even in its early version, was bottom-up. It performed a simulation of daily operation of each device of each appliance for each user class, and aggregating all these load profiles obtained the total load profile. However, three main aspects in the Load Curve Computation presented some limitations: the first is the manner in which random selection of switching-on instants *t* was performed, the second is the manner in which the switching-on instants were handled at the peak time, and the last regarded the management of the Gaussian distribution parameters by which the samples of switching-on instants were weighted around the peak instant.

Selection of switching-on instants

The algorithm, operating on a one-minute basis, considered the daily time as a vector [1:1440]. However, for each appliance (and thus for each device *i* corresponding to that appliance *j*), could be defined up to three daily functioning windows within the vector of daily minutes, in which the switching-on of each device could occur. According to this, all the $n_{t,jk}$ switching-on instants for each single device *i* could be sampled within the vectors corresponding to the functioning windows of the appliance *jk*. However, this sampling was not performed in completely random manner: a *vector of possible switching-on instants (VPS)* was defined for each appliance. To comprehend how this vector was built, suppose to have an appliance whose functioning cycle *d* is 12 minutes, and which present two functioning windows: the first window starts at the minute

360 of the day (i.e. at 6:00 AM) and ends at the minute 480 (i.e. at 8:00 AM), and the second ranging from minute 780 (1:00 PM) to 840 (2:00 PM). In this case, the *VPS* would have been composed in this way:

$$VPS = [360\ \ 372\ \ 384\ \ 396\ \ 408\ \ 420\ \ 432\ \ 444\ \ 456\ \ 468\ \ 780\ \ 792\ \ 804\ \ 816\ \ 828]$$

The possible switching-on instants are not all those within the functioning windows, but only those spaced by a period equal to the functioning cycle *d*, starting with the first of each functioning window. Note that, in the example, the minute 480 is not considered as a possible switching-on instant: this is due to the fact that if a device was turned on at minute 480, it would stay on until minute 492, breaching the functioning window, which ends at the minute 480. The same considerations can be done for the minute 840.

The choice made by the author was due the fact that, if the algorithm had been left to perform a random sampling of the switching-on instants directly within the functioning windows, it could have run into two main problems:

- The possibility that instants too close together were sampled: since a device, after is switched on, stay on for a time equal to the functioning cycle *d*, the minimum amount of time that must elapse between two switching-on instants is *d*.
- The possibility that, despite the functioning time *Ft* had lower amplitude than the functioning windows *Fw*, the algorithm could not correctly perform a random sampling of *all* the n_t switching-on instants. To comprehend the reason, refers to Figure 5.1, which is a graphical representation of switching-on instants sampling. Let's consider an appliance j, which has a single functioning window (starting at minute 1 and ending at minute 20), a functioning time of 15 minutes and a functioning cycle of 3 minutes. This means that each device i of the appliance j should turn on 5 times within its functioning window. Above is represented the switching-on behavior according to the solution adopted in the early version of the software: since the possible switching-on instants are rigidly defined (highlighted in yellow), the algorithm can perform the sampling of 5 switching-on instants within these values. Below is represented instead what could happen if the algorithm could perform a random sampling of the switching-on instants within the functioning window: despite the functioning time has lower amplitude than the functioning window, the algorithm in this case cannot simulate the fifth switching-on, because after it has performed 4 samplings there are no more blocks of 3 free minutes inside the window.

Fw	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Fw	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Figure 5.1: Graphical representation of switching-on instants sampling. Above is represented the switching-on behavior according to the rigid solution adopted in the early version of the software. Below is represented what could happen if the algorithm could perform a random sampling of the switching-on instants randomly within the functioning window: the algorithm in this case cannot simulate the fifth switching-on, because after it has performed 4 samplings there are no more blocks of 3 free minutes inside the window.

In this way, however, the selection of the switching-on instants results rigid, because the devices which belong to the same appliance will start always in the same instants, spaced by the functioning cycle *d*.

Construction of the user-class load profile

The old model built the load profile of each user-class within two steps:

a) Initially, the switching-on behavior at the peak time was simulated. Among the applications that could have been on the peak time (i.e. the peak time was within their functioning windows, henceforth *peak appliances*), was performed a random selection of devices that would have instantly turned on exactly at the peak time, until obtain a user-class peak load value equal to that calculated into the *peak load computation* phase. Essentially, the load value at the peak time was fixed.

b) After the construction of the load profile at the peak time, the algorithm proceeded with the simulation of the switching-on of all the devices in the rest of the day, to complete the load profile. Obviously, the devices already turned on at the peak time, would be switched on once in less during the rest of the day. The management of switching-on during the rest of the day differed depending on whether the device was a peak appliance or not.

   - For the devices of *not-peak appliances* (i.e. the appliances which cannot be on at the peak time), the simulation was carried out through a simple sampling of switching-on instants, according to the technique explained in the previous paragraph *Selection of switching-on instants.*
   - For the devices of *peak appliances*, the simulation was carried out by changing the probability of sampling around the peak time. Initially, the *VPS* was filled omitting the peak time instant. Then, a vector of the same length containing the probabilistic weights with which to perform the sampling was associated to the *VPS*. This allowed to assign a specific probabilistic weight of sampling to each possible

switching-on instant, in accordance with a normal distribution mediated on the peak time: in this way, the probability that the devices switch on near the peak time increased. Initially the weights were calculated according to a normal distribution having a very high bell (i.e. having low variance). This often meant to obtain a load profile in which the load just before and after the peak time was greater than the peak load calculated in the *a)*, because the majority of the devices turned on around the peak time. If this occurred, the bell of the normal distribution was relaxed by increasing the variance (+2 at each attempt), and the simulation was executed again (*57*). This until the peak of the load profile computed was equal to the peak value calculated in *a)*. For reasons of computation speed, the algorithm could perform up to 30 attempts, after which the load profile obtained would be accepted.

A serious problem that frequently occurred was that the algorithm exhausted the limit of 30 attempts without reaching the goal. Indeed, it was sufficient that even a single device turned on just after the peak to result an electrical load higher than the load value at the peak time. In addition, the management of peak starts was unrealistic: all the devices that were selected to work at the peak time, switched on instantly exactly at the peak instant, creating a discontinuity due to a sudden increase in demand. In conclusion, the procedure appeared slow and unintuitive.

5.2 Development of the improvements

In the following section the main solutions I adopted to improve the weaknesses of the algorithm are described. In particular, major changes regards: the proper implementation of uncertainty related to *Ft* and *Fw*; the sampling of devices' switching-on instants; the construction of user-class load profile. The solutions presented are implemented in the actual version of the software.

Input data management

In order to correctly implement the option for consider the uncertainty of the input data, it was necessary to solve the problems related to the extension or shortening of functioning time and functioning windows, discussed previously. The introduction of parameters *RFt* and *RFw* should be accompanied by some adjustments without which the algorithm could not function properly.

The first critical point that requires attention is the definition of the number of starts. Consider the appliance *j* in the user-class *k*: $n_{t,jk}$, which represents the number of daily switching-on of that appliance, is calculated as follows:

5.3

$$n_{t,jk} = \frac{Ft_{jk}}{d_{jk}}$$

In order to have an integer number of app starts, Ft_{jk} has to be a multiple of functioning cycle d_{jk}, which is not subject to randomization. For this reason, after the randomization Ft_{jk} is rounded to the nearest multiple of d_{jk}.

Another possible problem was the case in which, after randomization, the Ft_{jk} was higher than the duration of the summation of the $Fw(s)_{jk}$: in this case, in the phase of load profiles computation, the software would not be able to simulate all the starts of the device. Actually, the condition of having a Ft_{jk} shorter then the $Fw(s)_{jk}$ is not sufficient to ensure that all the starts can be simulated, but you must ensure that in the $Fw(s)_{jk}$ there are a number of "blocks" of duration d_{jk} greater than or equal to $n_{t,jk}$. If this occurs, it means that there is at least a configuration of switching-on instants that allows to simulate all the $n_{t,jk}$ starts within $Fw(s)_{jk}$. Otherwise, the $Fw(s)_{jk}$ are progressively extended, as long as the condition becomes true. Moreover, if after randomization two functioning windows of appliance *j* partially overlap, these are deferred and delayed up to join in a single window (Figure 5.2).

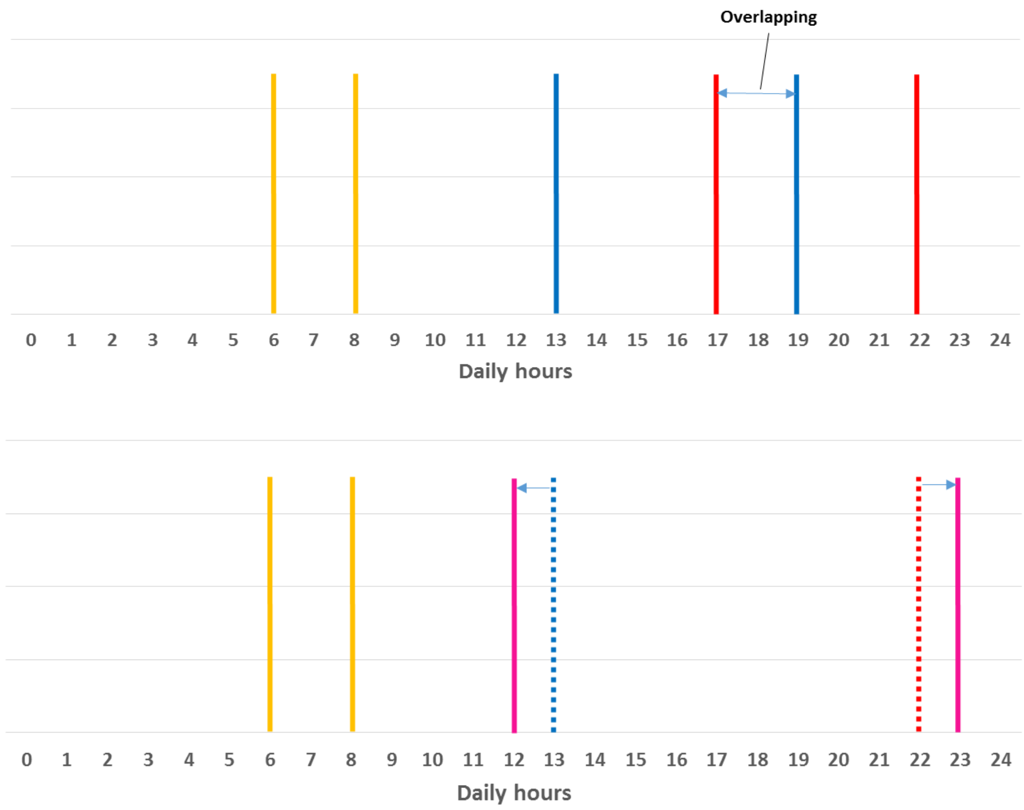


Figure 5.2: Graphs showing three different functioning windows of an appliance. In the case shown above, after randomization two functioning windows overlap (the blue and red windows). If this occurs, the algorithm anticipates and retards the two overlapping windows up to merge them into a single window, whose duration is the sum of the starting two windows.

Load Curve Computation

The *Load Curve Computation* block has been significantly changed. First, it was changed the way in which the switching-on instants are sampled, making it the most random possible. Also, actually the load profile of each user-class *k* is no longer built separating the management of peak time from the rest of the day, but it is the result of a simulation on the entire day of each device *i* of each appliance *j*. Finally, the variance of the normal distribution that drives the sampling of switching-on instants varies adaptively to the error on the peak value.

Selection of switching-on instants

The reasons that led the author of the first version of the model to develop a selection of switching-on instants in a rigid manner have already been discussed. In order to decrease the rigidity with which these instants were sampled, I had to focus in solving these problems. Consider the Figure 5.1: it represents the cases in which the sampling switching-on instants is carried out in a rigid way (above) or free way (below). In the old software, which performed the sampling always in a rigid manner, if occurred the condition that *VPS* contained more than $n_{t,jk}$ elements (i.e. the possible switching-on instants were more of the starting instants to be sampled) did not occur problems, because the possible switching-on instants were spaced cycle *d*, then all $n_{t,jk}$ starts could be simulated for all devices. Instead, if the sampling was made directly within the functioning windows, it was possible that the algorithm was not able to complete the simulation of all $n_{t,jk}$ starts, because some minutes were "wasted" in windows between a switch and the other, too short to accommodate a functioning cycle *d*. However, it appears evident that, for appliances with very long functioning windows, would not occur problems of this type.

Fw	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Figure 5.3: Graphical representation of switching-on instants sampling performed in random manner within a functioning window. This is the same case represented in Figure 5.1 (Ft=15 min; d=3 min), but in which the unique Fw has a duration of 26 minutes instead of 20 minutes. In this case, even if the algorithm has performed a sampling which is the worst possible combination in terms of minutes "wasted" (for each starts which occupies a period equal to the functioning time d, follows a window of duration (d-1) which can accommodate no functioning cycles), it has been able to simulate all the daily starts of the device.

Consider for example Figure 5.3: for applications with functioning windows long enough, with a random selection of switching-on instants directly within the *Fw,* the algorithm has no problems to perform the simulation even in the most unlucky combination, which corresponds to intersperse each functioning cycle with a window of duration (*d-1*). In light of this, it is possible to express a mathematical condition for the duration of *Fw* that allow random sampling of switching-on instants directly by the vector of *Fw,* in this way:

5.4

$$(d_{jk} - 1) * n_{t,jk} + d_{jk} * n_{t,jk} \leq size(VPS_{jk}) * d_{jk}$$

where *size*(VPS_{jk}) refers to the number of elements of the vector of possible switching-on instants constructed in the rigid way (i.e. the number of blocks of duration *d* present within the vector of *Fw*); the term at the left side of the disequation represents the time requested by the most unlucky combination of instants sampling, while the term at the right side represents the duration of functioning windows (excluding the blocks of minutes not able to contain *d*, as minutes 19 and 20 in Figure 5.1). If an appliance *j* respect the condition in Eq. 5.4, then it is possible to carry out random sampling of switching-on instants directly from the vector representing the Fw_{jk}, because the algorithm is always able to simulate all the $n_{t,jk}$ starts for each device *i*. This allows to perform for each device random switching-on instants sampling.

As regards the appliances that do not meet the condition 5.4, the task is a bit more complicated. A possible way would be to accept to perform the sampling in a rigid way for these appliances. However, the solution adopted is to perform the simulation of switching-on instants initially in a rigid manner, until the condition in Eq. 5.5 is respected, and then complete the remaining simulations in a free manner. Indeed, after each sample, the term on the left side of the disequation decreases by *d+(d-1)*, while the term on the right side decreases by *d*: then, as the switching-on instants of each device are sampled, the term on the left side will tend to decline faster than that to the right side, as long as the condition is met. After that, the algorithm can complete the simulation of remaining starts in free manner, by sampling directly the switching-on instants within the *Fw*.

5.5

$$(d_{jk} - 1) * n_{t,jk} + d_{jk} * n_{t,jk} \leq size(VPS_{jk}) * d_{jk}$$

Construction of the user-class load profile

The new algorithm, to build the curve of each user-class, does not handle the peak starts separately with respect to the starts in the rest of the day. Indeed, for each device *i* of the appliance *j* belonging to the user-class *k*, the sampling of $n_{t,jk}$ starting instants is performed compatibly with its functioning windows. The sampling is carried out as explained Selection of switching-on instants, distinguishing between appliances that meet or do not meet the condition 5.4, and then between *peak* or *not-peak appliances*.

For appliances meeting condition 5.4, free sampling of switching-on instants of each device is carried out within the vector representing the functioning windows: in this way the device can turn on in every minute included in the functioning windows.

- For the not-peak appliances, sampling is performed according to a uniform distribution, i.e. every minute included functioning windows has the same probability of being sampled as switching-on instant.
- For peak appliances, sampling is performed by assigning a specific probabilistic weight to every minute of functioning windows. To the instants of functioning windows between 150 minutes before and after the peak time are assigned weights according to a normal distribution centered on the peak time, while the weight assigned to all the others instants is equal to that of the 150^{th} element. The values of the weights are calculated according to the *probability density function* of a normal distribution: the values of this function for *x* in

the range [-2; +2] are discretized into 300 elements, and assigned to each minute within the range [peak time - 150 min; peak time + 150 min].

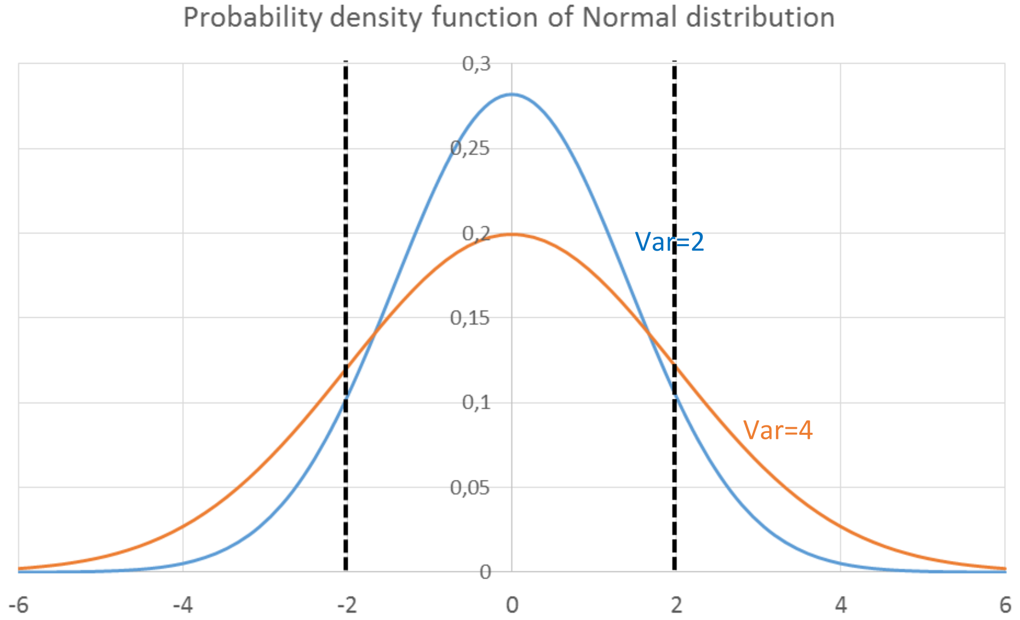


Figure 5.4: Examples of probability density function on Normal distribution. The values of the function for x ranging from -2 to +2 are discretized and applied as probability weights of sampling for Fw instants between 150 minutes before and after the peak time.

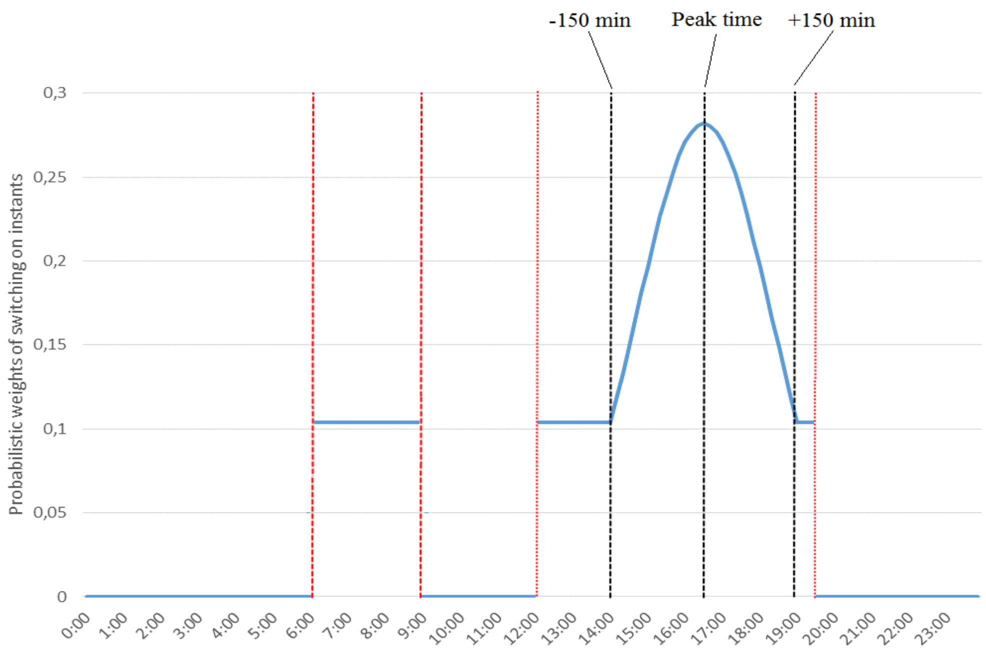


Figure 5.5: Graphical representation of the weights associated to the minutes of the functioning windows. The red dotted lines indicates the limits of the 2 daily functioning windows. The probabilistic weights of sampling within the range [peak time - 150; peak time + 150], that is between 14:00 and 19:00, are in accordance with the probability density function of a normal distribution (in this case the one having variance=2 in Figure 5.4), while to the remaining instants of the functioning windows the lowest weight is associated.

The resulting peak from the user-class load profile obtained after the simulations of all the devices of each appliance is compared with the estimated peak in the block *Peak Value Computation*: if the peak obtained is greater than that estimated through Bary's correlation, then the variance of the normal is increased in order to decrease the probability of sampling around the peak time, vice versa the variance is decreased.

For appliances that do not meet the condition 5.4, the algorithm operates in the same way, with the only difference that initially the switching-on instants are sampled in a rigid manner, as in the old version of the software. In any case, the simulation of the peak appliances is guided by the same normal distribution that is applied to the peak appliances that meet the condition 5.4.

Efficiency improvements

The software allows the user to choose, upon execution, how much load profiles generate. Indeed, since the model is stochastic, it would be of little significance to generate a single load profile. For this, it is important the execution speed of the algorithm, i.e. the time required to generate load profiles: software improvements have been performed from this point of view, thanks to a Matlab code more efficient. Although the switching-on instants sampling is performed randomly within the functioning windows of appliances, and for each simulation is required a check to verify if the condition 5.5 is respected (for appliances that do not meet the condition 5.4), the execution time is about the same as the primitive version of LoadProGen, which performed the switching-on instants sampling in a rigid way. Obviously, the execution time depends from the connected load, the number of appliances, the number of devices, and in general from the input data. To get an idea, however, the total execution time of the peripheral-urban area in Uganda (shown in Appendix A), which has a daily demand estimated about 140 kWh and an estimated peak of 70 kW, is approximately 20 seconds for reading the Excel file containing the input data, 1 minute for generation of each load profile, and 10 seconds for writing output to a new Excel file.

5.3 Computational framework

The LoadProGen software has been implemented in MATLAB, while input data and output results of the algorithm are managed through Microsoft Excel. The model follows a bottom-up approach, and operates on a single user-class: it simulates all the devices of the appliances of a user-class, and obtains the load profile for the user-class considered. If the total load is composed by more user-classes, the user-class load profiles are developed independently from each other by the algorithm, and are then aggregated to form the total load profile.

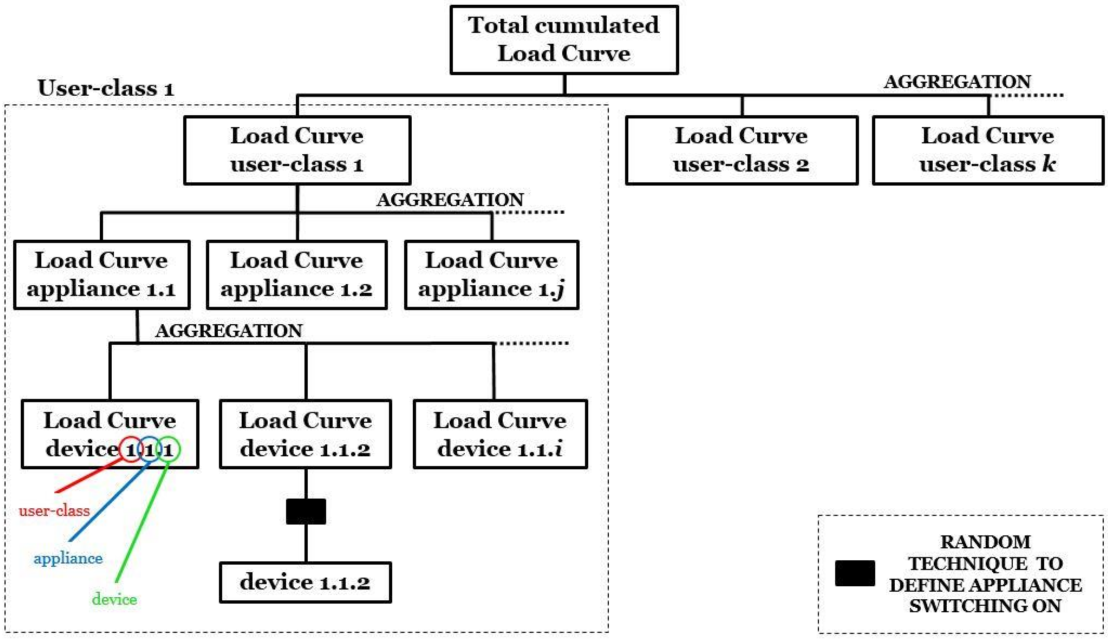


Figure 5.6: Flow chart of the bottom-up behavior of the model LoadProGen: from the load profile of the individual device, to the total load profile.

In Figure 5.7 a simplified flow chart of LoadProGen is presented.

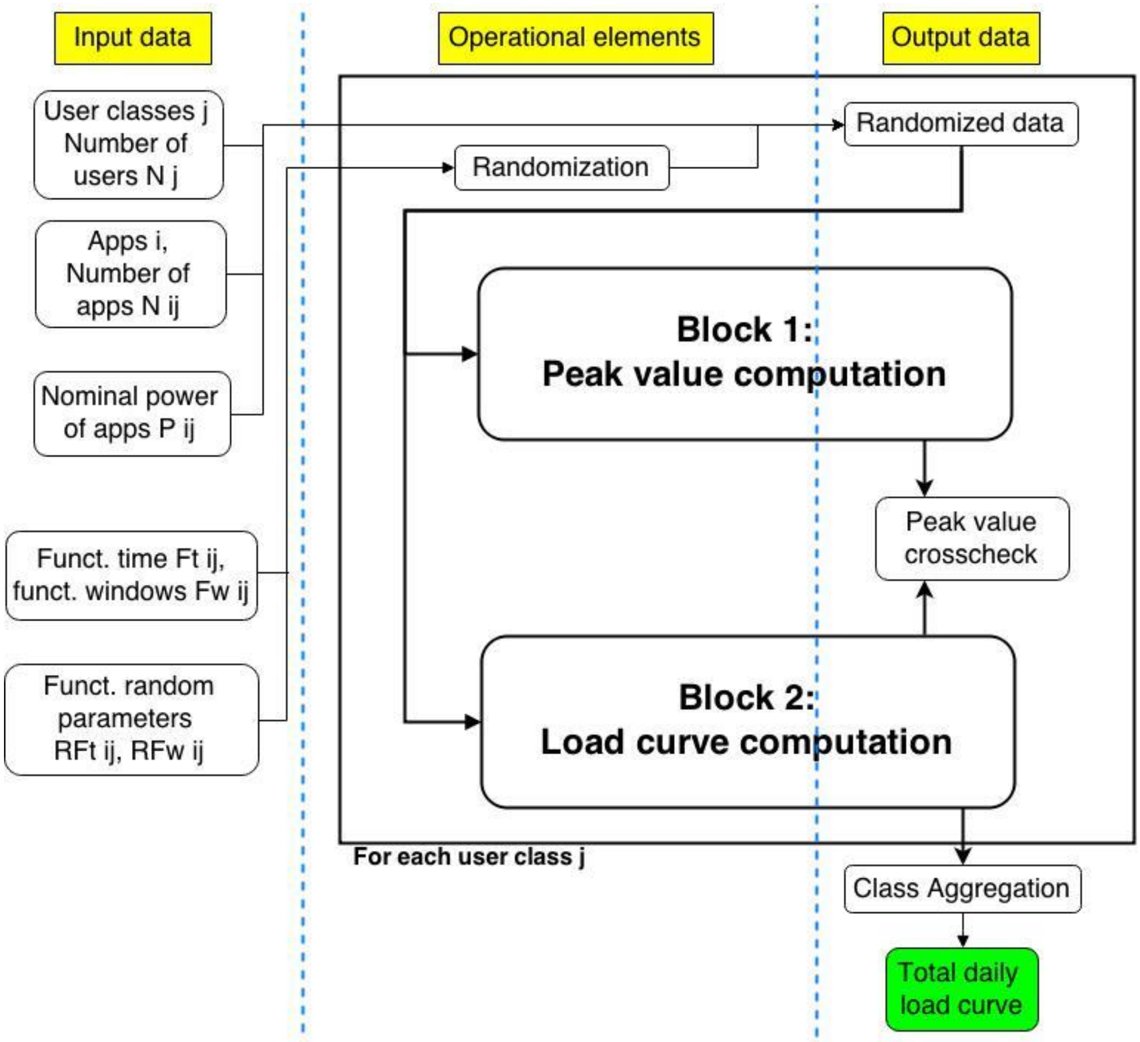


Figure 5.7: Flow chart of LoadProGen.

The algorithm LoadProGen can be divided in three sections:

- *Input data* highlights different groups of required input.
- *Operation elements* considers the different computational steps performed within the algorithm.
- *Output data* highlights different groups of outputs computed by the algorithm.

However, a more intuitive subdivision might be to divide the algorithm into sections related to the task performed. By this way, once processed the input data with their randomizations, two main blocks can be distinguished: (i) the *Peak Value Computation* and (ii) the *Load Curve Computation.*

Input data

The input data have already been discussed in Input data. In any case, it is worth remembering that some of them (*Ft* and *Fw*) are subject to a randomization according to designer parameters (*Rft* and *Rfw*), allowing to consider the uncertainty of these input data, which are often the result of estimates by the designers. The operator which use the model has to enter the input data required into an Excel file. That Excel input file is composed of many sheets as number of user-classes identified in the customers' group. In each sheet data about the respective user-class are recorded. The structure of a sheet of input file is shown in Table 5.1.

App	User Class	N_{Users}	P [W]	N_{app}	d [min]	Ft [min]	RFt [%]	RFw [%]	w1_s	w1_e	w2_s	w2_e	w3_s	w3_e
Bulb light	Family 3	15	3	8	10	360	30	30	0	120	1020	1440		
Phone Charger	Family 3	15	5	2	30	180	30	30	0	540	780	900	1020	1440
Radio	Family 3	15	5	1	15	240	30	30	360	540	1020	1440		
Security Light	Family 3	15	5	2	30	720	30	30	0	420	1020	1440		
TV (small)	Family 3	15	100	1	30	300	30	30	660	900	1020	1440		
Fridge (small)	Family 3	15	250	1	10	300	30	30	0	1440				

Table 5.1: Example of input file structure. The table refers to a single sheet, corresponding to the user-class "Family 3" from Uganda's data (Appendix A). In the table are listed all the types of appliances present in the typical customer of user-class "Family 3", and for each of them is specified the number per customer N_{app}, the nominal power P, the functioning time Ft and functioning cycle d, the random parameters RFt and RFw, and finally the start-end daily instants (in minutes, from 0 to 1440) for the functioning windows (up to three).

Operational elements and Output data

a) The algorithm elaborates the input data set by the designers in order to obtain them in the proper form to compute the load profile. The randomization process is quite complex, and the sequence requires to be described:

- Initially, the functioning time (*Ft*) of each type of appliance is randomized as described in Eq. 4.1. The input value of functioning time of each type of appliance is shortened or extended through a random value defined in the range: $[-(Ft_{jk} * Rft_{jk}); +(Ft_{jk} * Rft_{jk})]$.
- After that, *Ft* is corrected in order to have an integer number of app starts. As the number of starts n_t of each type of appliance is defined as in Eq. 4.2, the *Ft* is rounded to the nearest multiple of the functioning cycle d_{jk}.
- The functioning window(s) (*Fw*) of each type of appliance is randomized as described in Eq. 4.3. The input value of Fw_{jk} is modified through a random value defined in the range: $[-(Fw_{jk} * RFw_{jk}); +(Fw_{jk} * RFw_{jk})]$. According to this, the starting and ending instants of the functioning windows are modified to extend or shorten the windows length.
- After the randomization of *Ft* and *Fw* has been performed, there is the need to check that, for each type of appliance *jk*, the functioning window(s) are able to

"contain" all the daily starts. Otherwise, the functioning window(s) are gradually expanded until they are able to contain all the starts.

b) *Peak Value Computation*: in this block the total required energy, the peak window(s), the maximum possible power peak, and the peak time are firstly computed. Then with an iterative process the load factor and the coincidence factors are computed according with Eq. 2.12 until convergence is reached for their values. Hence the reference value of the power peak for the considered user class can be computed. Follows a more detailed description of this block:

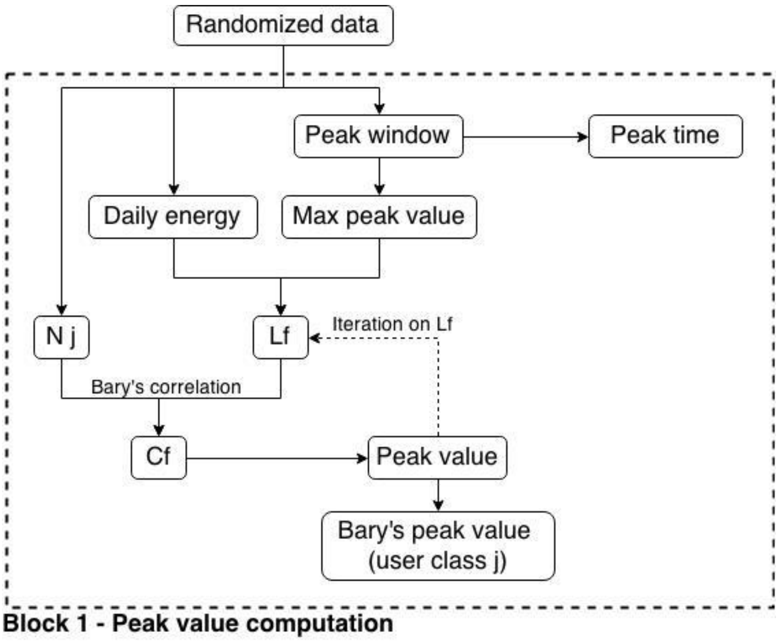


Figure 5.8: Flow chart of the block related to the "Peak Value Computation"

- Overlapping the functioning windows *Fw* of the different appliances within a class *k*, a *peak window* will result to be embraced by a number of appliances, hence defining a possible *maximum power peak* (*max_peak_k*) (Figure 4.1).
- The *peak time* is sampled from the vector of the *peak window*, which is the period during which the maximum power peak can occur.
- The daily energy demand (per user-class *k*) is calculated as follows:

5.6

$$E_k[Wh] = \sum_j \left(N_k * n_{jk} * P_{jk}[W] * Ft_{jk}[h]\right)$$

in which *(N_k * n_{jk})* refers to the total number of units of type *j* for all the users which belongs to user class *k*, and *(P_{jk} * Ft_{jk})* represent the daily energy demand for each unit.

- A starting value for the load factor *Lf(0)* is calculated in this way:

5.7

$$Lf_k(0) = \frac{E_k[Wh]}{max\_peak_k[W] * 24\ [h]}$$

- Hence, a first value of coincidence factor *Cf(0)* can be computed using the correlation by Bary (Eq. 2.12, 2.13 and 2.14).

5.8

$$Cf_k(0) = a * Lf_k(0) + \left(1 - a * Lf_{\ k}(0)\right)N_k^{-\frac{1}{2}}$$

- The peak value can be calculated:

5.9

$$peak_k(0) = Cf_{\ k}(0) * max\_peak_j[W]$$

- At this point, the iterative process begins: first the *Lf* is recalculated using the peak demand obtained by the previous iteration, then the *Cf* is computed using the correlation by Bary with the new *Lf*, and finally the new peak is calculated using the *Cf*. The process continues until the error between the peak value and that of the previous iteration drops below a tolerance value (set to 0.1%).

c) *Load Curve Computation*: In this block, for each device of each type of appliance *j*, the switching-on instants *t* are randomly sampled from the vector of the functioning window(s), through the Matlab function "*datasample*", which execute a random selection of a defined number of elements within a vector. Accordingly, the load profile for the user class *k* can be computed. However, the resulting user-class peak may not match with which calculated in the step *b)*. Therefore iterations are performed by operating on the variance of the normal probability distribution which guides the random sampling of starting instants *t* of the peak appliances.

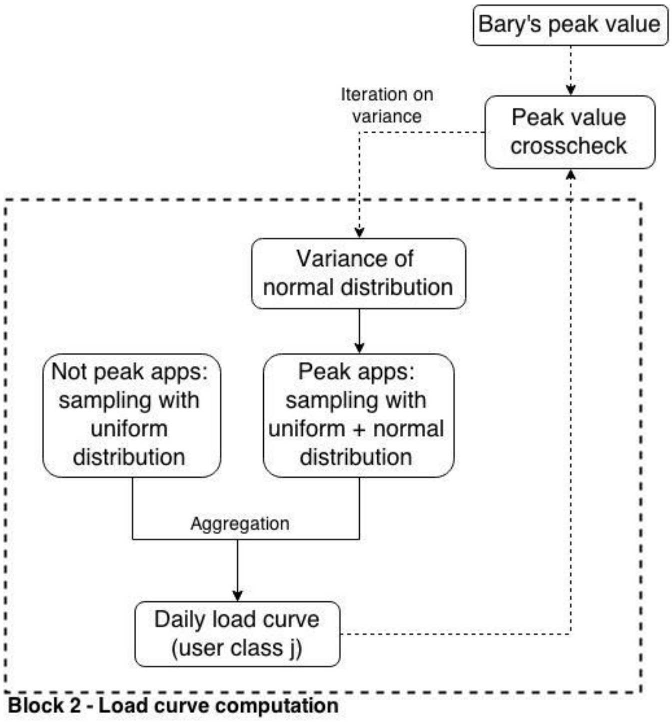


Figure 5.9: Flow chart of the block related to the "Load Curve Computation"

- For the appliances not contributing to the peak, the sampling of the switching-on instants *t* is carried out, using *datasample* function, with a uniform probability distribution from the vector representing the functioning window(s) Fw_{jk} of each appliance.
- For the appliances contributing to the peak (i.e. the *peak time* is included in their functioning windows) the sampling of the switching-on instants *t* is still carried out with a uniform probability distribution, but, in the neighborhood of the peak time (in the range around *peak time* between [-150 min; +150 min]), a sampling with a normal probability distribution is performed (Figure 5.10). At each iteration, the *variance* of the normal distribution that dominates the probability of sampling of switching-on instants *t* around the peak time is increased or decreased, depending on whether the resulting peak is respectively greater or smaller than that estimated using the correlation of Bary (in the *Peak Value Computation* block). Indeed, increasing the variance, the bell that represents the normal distribution relaxes, reducing the probability of switching-on around the peak time, and vice versa decreasing the variance. The iteration continues until the resulting user-class peak load value matches (within some tolerance) with the estimated peak value by the Bary's correlation.

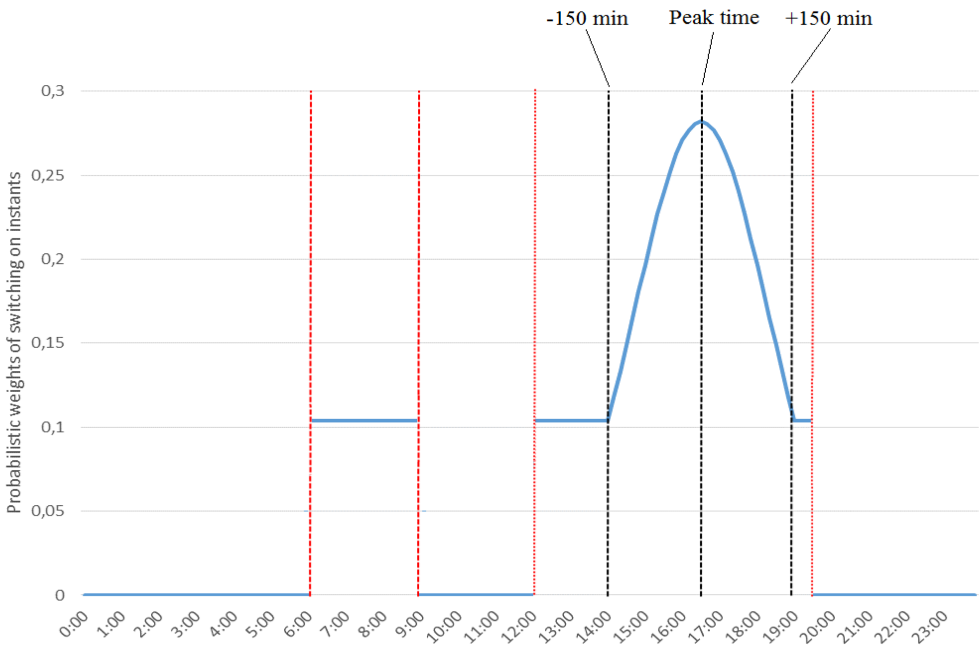


Figure 5.10: Graphical representation of the weights associated to the minutes of the functioning windows. The red dotted lines indicates the limits of the 2 daily functioning windows (6:00-9:00 and 12:00-19.30). The normal distribution affects a period that includes 150 minutes before and after the peak time of the user-class in question (16.30). The greater the variance, the lower the bell, until reaching a uniform distribution.

d) Repeating steps *b)* and *c)* for each user-class *k* and aggregating the different user-class load profiles allows computing the total load profile.

5.4 LoadProGen Vs state-of-the art approaches

In the light of the changes and improvements applied to the software, it is possible to show a sample output of LoadProGen, and make some considerations about them. Figure 5.11 and Figure 5.12 show four load profiles estimated through LoadProGen for the case study of Uganda (complete input data available in Appendix A), obtained with values of uncertainty parameters of input data (i.e. *RFt* and *RFw*) respectively equal to 0% and 30%. It can be noted as in the first case, the load profiles have a more squared shape: this is because the functioning windows are not subject to randomization, so the load may experience a discontinuity at the beginning and at the end of the windows. Obviously, the higher the value of the parameter *RFw,* the greater is the variability of the windows, the smoother are the load profiles.

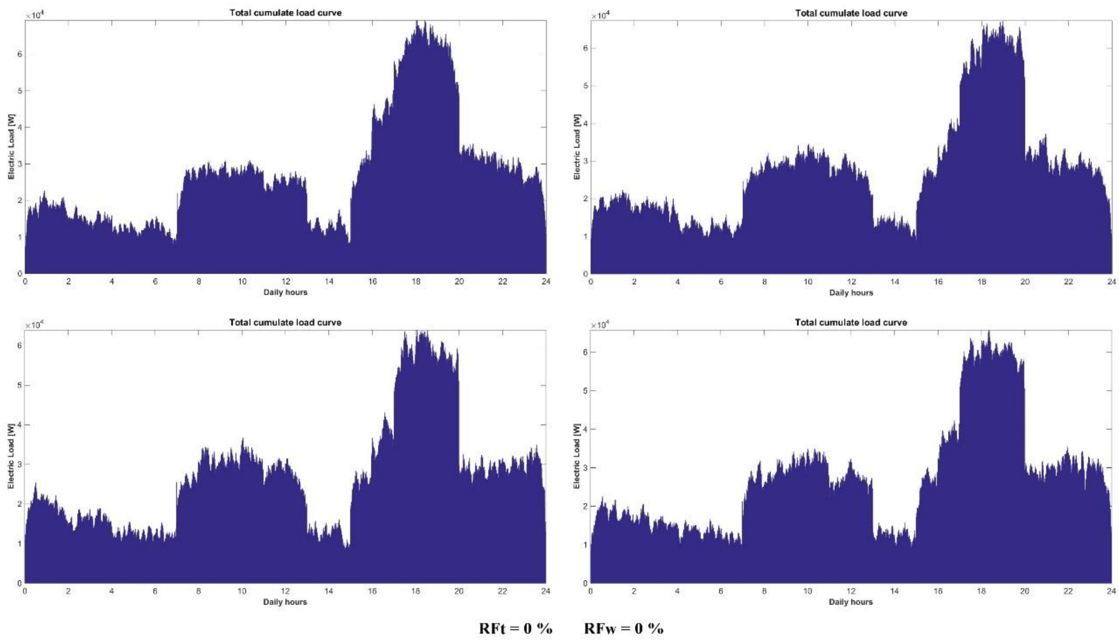


Figure 5.11: Four load profiles relative to Uganda's input data (Appendix A) estimated by LoadProGen, setting RFt=0% and RFw=0%

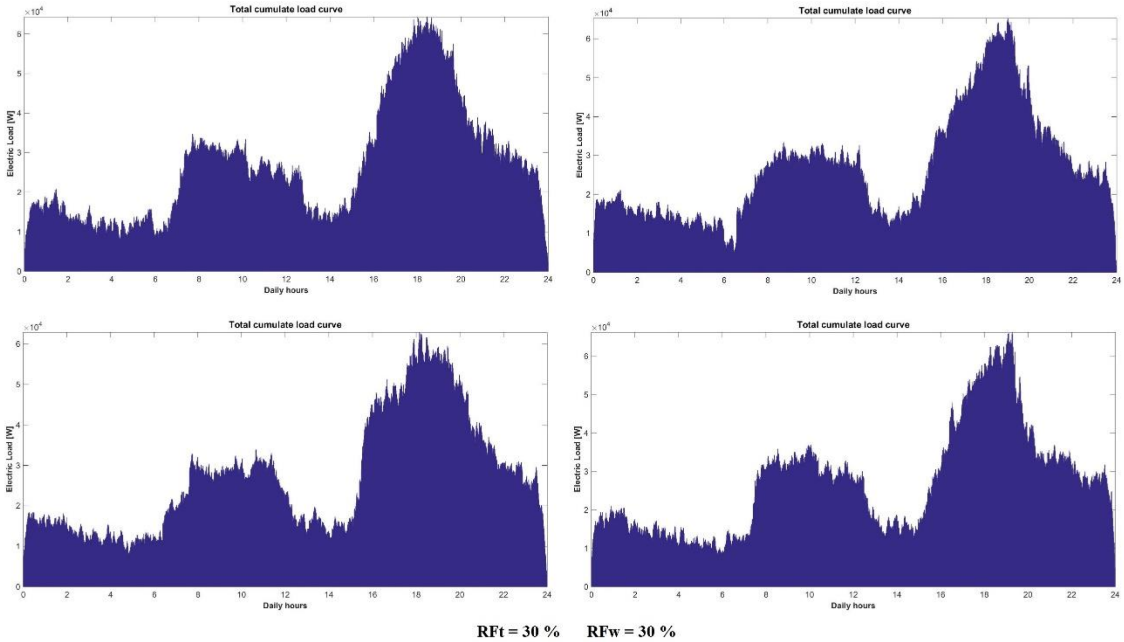


Figure 5.12: Four load profiles relative to Uganda's input data (Appendix A) estimated by LoadProGen, setting RFt=30% and RFw=30%

It can be interesting to make a comparison between the results obtained using the model and the load estimation performed with the simplest methods founded in literature, in particular the *setting on-off approach* (Figure 5.13) and the *medium power approach* (Figure 5.14) introduced in Load profiles estimated. As can be noticed, the peak value obtained with the setting on-off approach is higher than the peak value obtained using LoadProGen, because the appliances are switched on within rigid windows. Instead, the peak value obtained with the medium power approach is lower, because not defining the functioning cycles *d* of each appliance, this approach "spreads" the energy demand of each devices over the entire duration of the functioning windows. For example, if an appliance has a nominal power rate equal to 90 W, a daily functioning time of 30 minutes, and the duration of functioning windows is 1 hour, the simple approach considers a mean power of 45 W during the entire functioning windows.

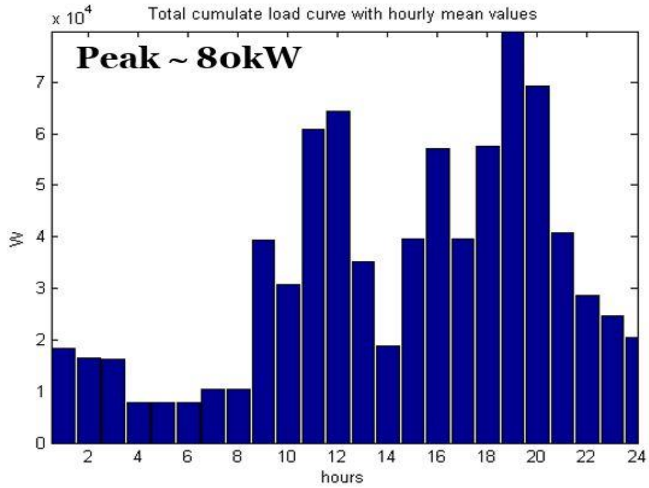


Figure 5.13: Load profile relative to Uganda's input data (Appendix A) estimated through the setting on-off approach

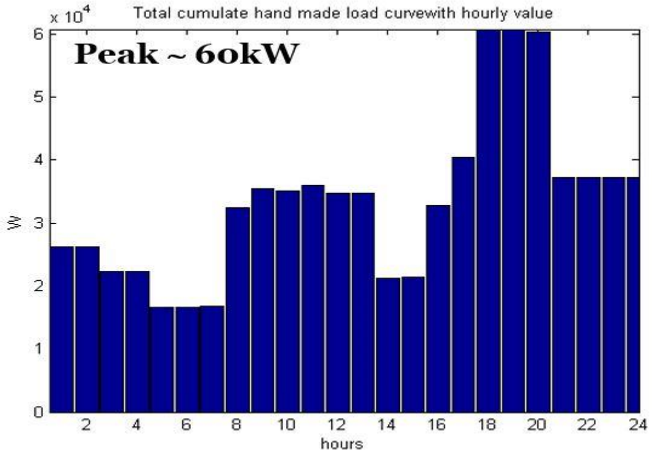


Figure 5.14: Load profile relative to Uganda's input data (Appendix A) estimated through the medium power approach

6 Applications of the software tool

In the following two applications of the LoadProGen software tool are described. The first one aims at introducing an application of the model to perform sizing procedure of an off-grid photovoltaic system in a peripheral urban area of Uganda. The second one aims at introducing the application of the model as a load profile forecast tool, i.e. to show the capability of the model in matching metered real daily load profiles. In this case the application refers to the load profile of a college in a peri-urban area of Cameroon.

6.1 Application for sizing of an off-grid PV system

The principal aim the model has been designed and developed for is to provide a defined and rigorous approach for the estimation of the electrical load profiles of unelectrified areas, in order to improve the effectiveness of off-grid power systems sizing process for the electrification. Load profiles are an input data required by many methods and software that deal with sizing of power systems, and they can deeply affect the results of the optimization and sizing process. Moreover, given a targeted context for off-grid system implementation, the expected electrical load profile is not unique, because uncertainty can be associated to the load profile. This is even truer when dealing with rural electrification, due to the lack of detailed information. Therefore the possible different load profiles related to a context can lead to different system sizing. A possible approach to this issue is to look for the system sizing which fits with the majority of the possible load profiles.

This application of the model developed aims at a first step off-grid power system sizing, which address the issue of uncertainty of load profiles. Indeed, as already highlighted, the model is based on a stochastic approach, which allows, given a set of input data, to compute different load profiles which all match with the imposed constraints, i.e. the uncertainty of load profiles for off-grid systems sizing is considered. Beside the uncertainty embraced by the model in developing different profiles for a given set of input data, further degree of uncertainty of load input data has been introduced by considering different scenarios based on different values of *RFt* and *RFw.*

This case study has been applied to the sizing of the main components of a PV Micro-Grid in a peripheral urban area of Uganda according to the following features:

- 10 scenarios of load input data have been identified. Appendix A shows the main input data which have remained constant for each scenario. The different scenarios are based on different values of *RFt* and *RFw* (Table 6.1)

Table 6.1: Definition of the 10 load profile scenarios

	RFt_{jk} [%]	RFw_{jk} [%]
Scenario 1	0	0
Scenario 2	0	10
Scenario 3	0	20
Scenario 4	0	30
Scenario 5	10	0
Scenario 6	20	0
Scenario 7	30	0
Scenario 8	10	10
Scenario 9	20	20
Scenario 10	30	30

- For each scenario, 100 load profiles have been computed using the model developed. Obviously, since the model is stochastic, curves of the same scenario will be different each other.
- For each load profile within a scenario, the main system components sizes of the PV Micro-Grid (i.e. photovoltaic nominal power and battery storage capacity) have been defined by mean of a numerical optimization approach based on life-cycle Net Present Cost and minimum Loss of Load parameter [70].
- Since for each estimated load profile within a scenario, different optimum system sizes may result, the optimum system configuration for each scenario (i.e. the optimum sizes of the PV panels and the storage) is identified as the one that fits with the majority of the employed load profiles.
- An overall best system configuration can be recognized by overlapping the results of each single scenario.

Figure 6.1 and Figure 6.2 show graphical representations for the sizing results of the Scenario 1. In particular, Figure 6.1 highlights the distribution on the search plane of the optimum system configurations: going from black to red, to yellow and white, the configurations have resulted the best ones more often. In Figure 6.2, which is the 3D representation of the previous graph, it is possible to appreciate the frequencies of different optimum systems configurations. Scenario 1 is particularly significant, because it can be noticed that even if *RFt* and *RFw* are zero (i.e. there is no uncertainty related to the load input data) both PV and storage sizes changes over a sensible range: this is due exclusively to the variable nature of the load, i.e. to the randomness of functioning pattern of electrical devices. In this scenario the best system configuration (which has been the optimal configuration for 20 load profiles on the 100 computed) has 216 kW PV array and 864 kWh storage capacity.

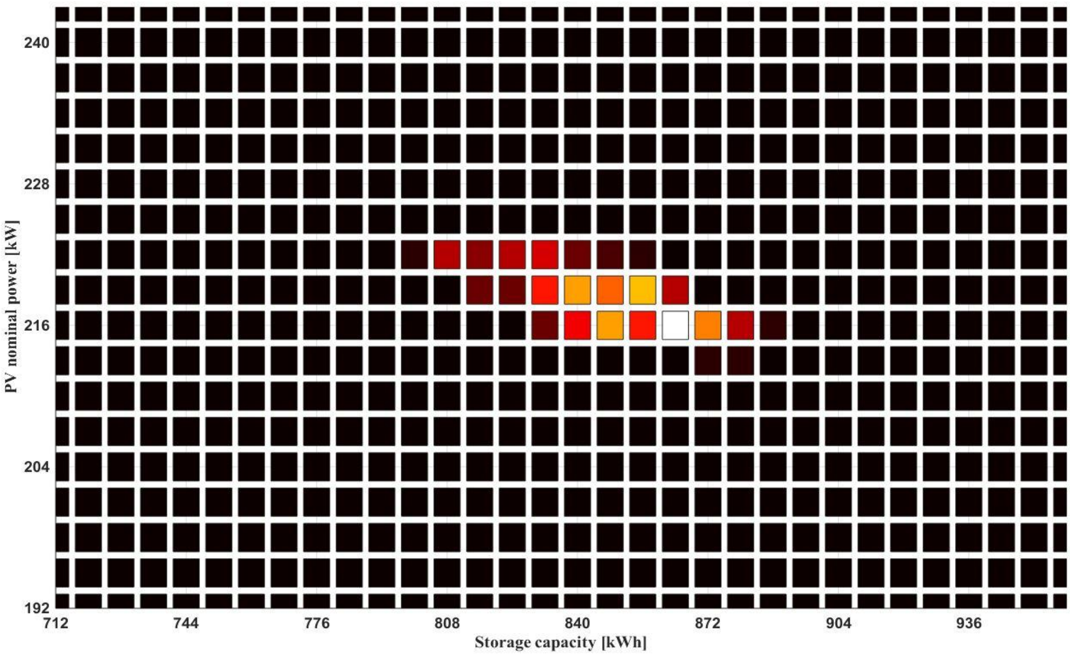


Figure 6.1: Search space and ranges of the resulting optimum PV array and storage sizes for Scenario 1

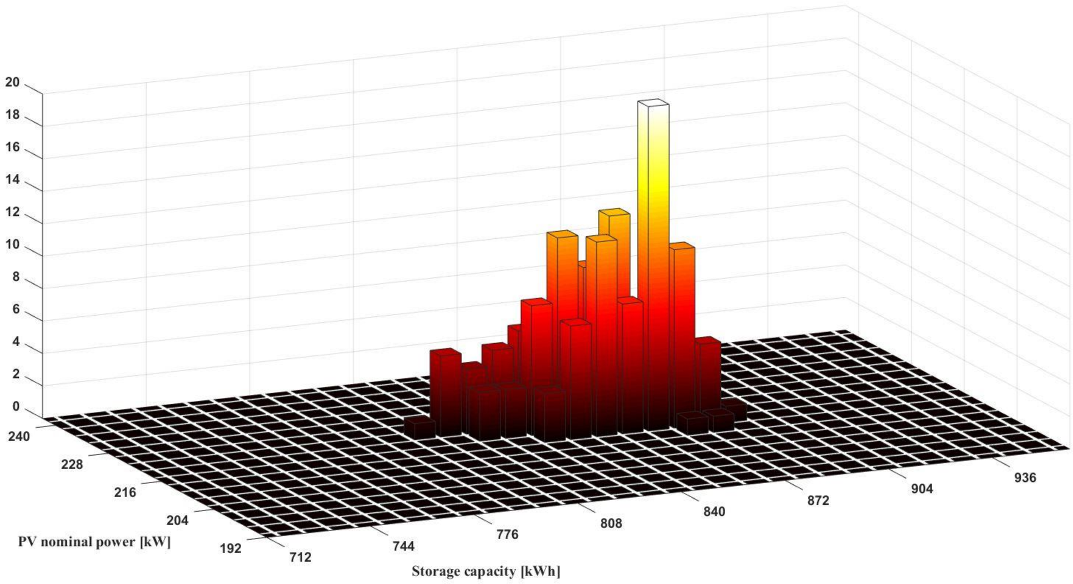


Figure 6.2: Frequency of the different optimum system configuration for Scenario 1

The same analysis has been performed for the other considered scenarios. In Scenario 10 the uncertainty in the input data has been set to the maximum value (30%), both for the functioning time either for the functioning windows. It’s interesting notice that, compared to scenario 1, the range of optimal solutions in Scenario 10 is much wider (Figure 6.3). In Table 6.2 is compared the range of variation of the optimal plants in Scenario 1 and 10.

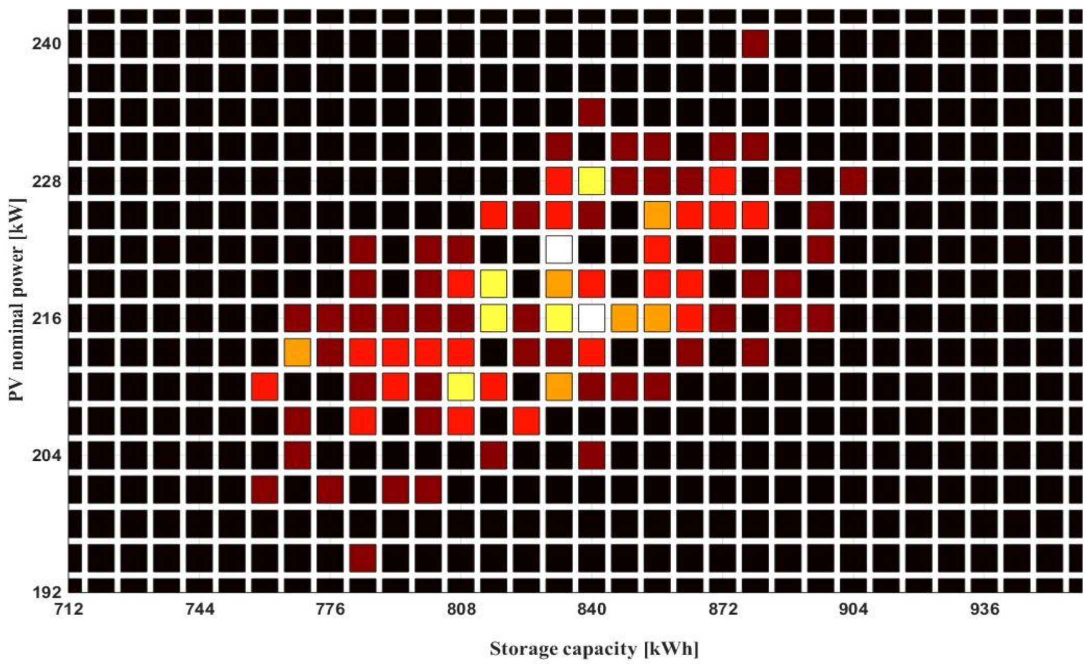


Figure 6.3: Search space and ranges of the resulting optimum PV array and storage sizes for Scenario 10

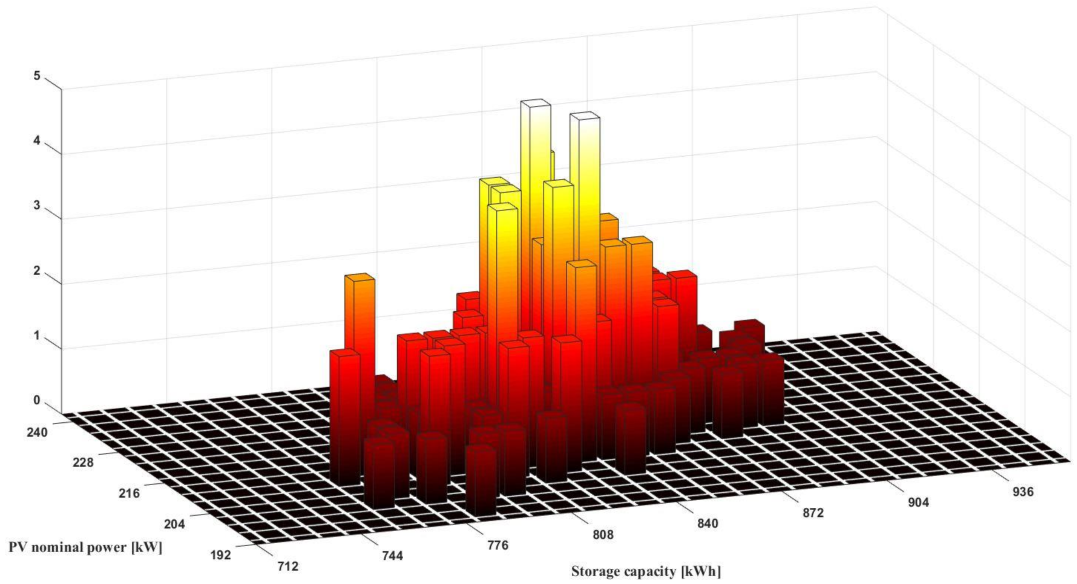
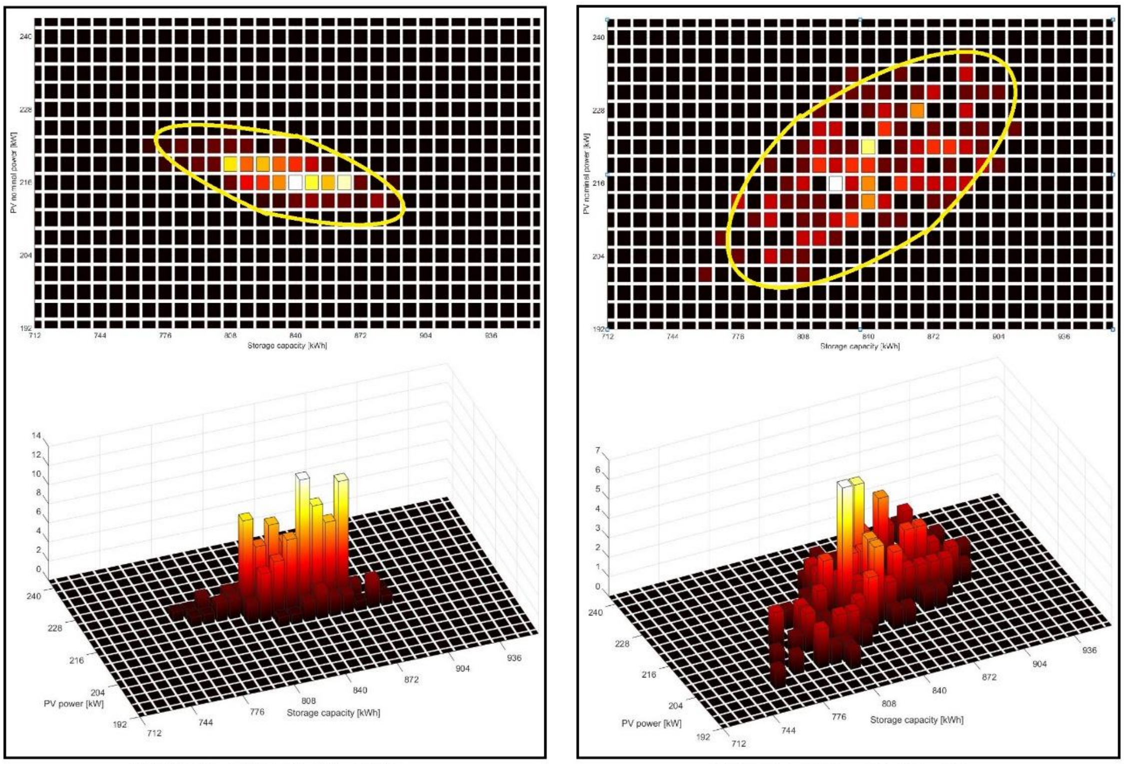


Figure 6.4: Frequency of the different optimum system configuration for Scenario 10

Table 6.2: Variation range of PV and storage of optimal configurations for Scenario 1 and 10

	Optimal configuration		**Variation Range**			
	PV [kW]	Storage [kWh]	Max PV [kW]	Min PV [kW]	Max Stor. [kWh]	Min Stor. [kWh]
Scenario 1	216	864	216	222	808	880
Scenario 10	222	832	195	240	760	904

It is also interesting to compare the results obtained for Scenarios 4 and 7: in the first case the uncertainty is zero for the functioning time and 30% for the functioning windows, conversely in the second the uncertainty is equal to 30% for *Ft* time and zero for *Fw*. This means that in Scenario 4 the daily electrical demand is always constant, but the daily windows of equipment operation vary, while in Scenario 7 the daily electrical demand can vary (because functioning time of appliances vary) but not functioning windows.



SCENARIO 4 (RFt=0% RFw=30%) **SCENARIO 7 (RFt=30% RFw=0%)**

Figure 6.5: Optimum system configurations for Scenario 4 and 7

The Figure 6.5 shows the comparison of results for these opposite scenarios. In Scenario 4 the resulting optimum plants are more balanced: there are optimal configurations having less photovoltaic and more batteries, and vice versa, but generally the trend is to obtain power plants with similar sizes. This is due to the fact that the daily energy demand does not vary, because not subject to uncertainty. The resulting optimum plants for Scenario 7 shows greater variability in absolute terms, and also can be noticed that the variability regards systems with different sizes: the optimal configurations range from small plants (with little generation capacity and energy storage), to greater systems having both major photovoltaic power and more batteries: in the first case evidently the estimated profile had a lower daily energy demand, vice versa in the second case.

The resulting optimal plants for each scenario are reported in Table 6.3.

Table 6.3: Resulting best plant configurations

	Optimal configuration		
	PV [kW]	Storage [kWh]	Frequency [%]
Scenario 1	216	864	20
Scenario 2	216	848	22
Scenario 3	216	832	11
Scenario 4	216	840	14
Scenario 5	213	856	9
Scenario 6	213	864	8
Scenario 7	216	824	7
Scenario 8	219	824	6
Scenario 9	222	816	7
Scenario 10	222	832	5
Overall	**216**	**848**	**7,2**

Finally, by aggregating the results of the 10 scenarios the same analysis can be carried out. This was done simply to show all the optimal solutions obtained for all the 1000 simulations performed, while for a realistic utilization of the software designers should set the proper values of *RFt* and *RFw* (eventually for each appliance). In Figure 6.6 and Figure 6.7 graphical representations for the sizing results of the aggregated scenarios are shown, while Table 6.3 also reports the resulting overall best system configuration, which has been the optimal system for 72 load profiles on the 1000 generated.

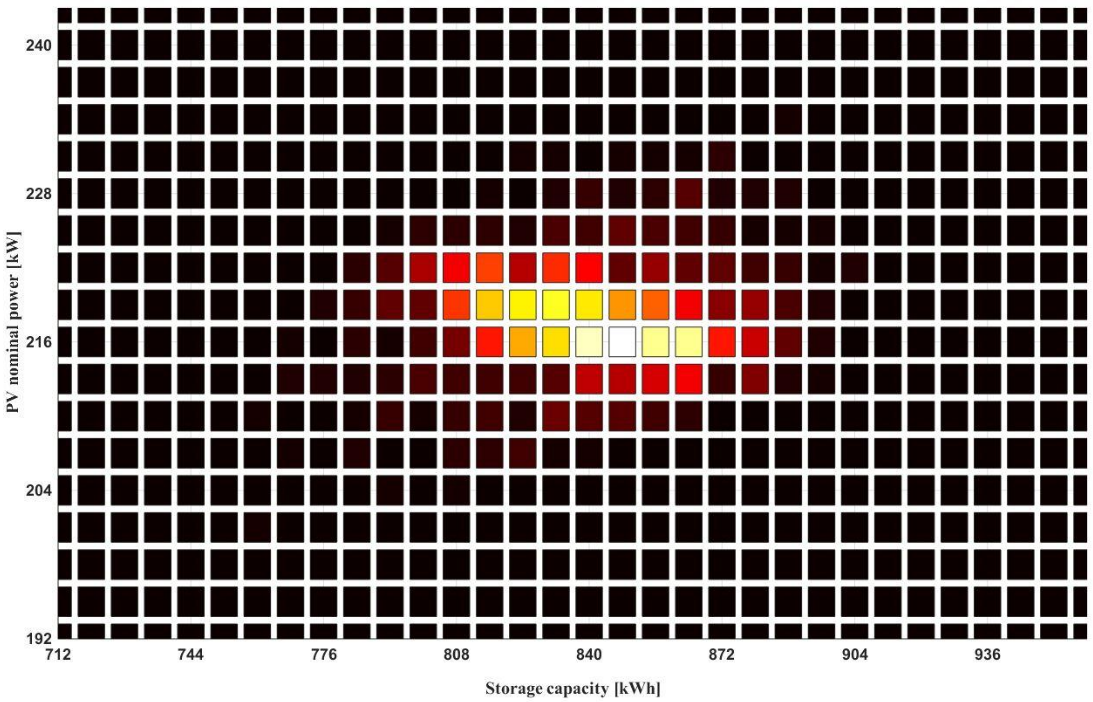


Figure 6.6: Search space and ranges of the resulting optimum PV array and storage sizes by aggregating all the scenarios

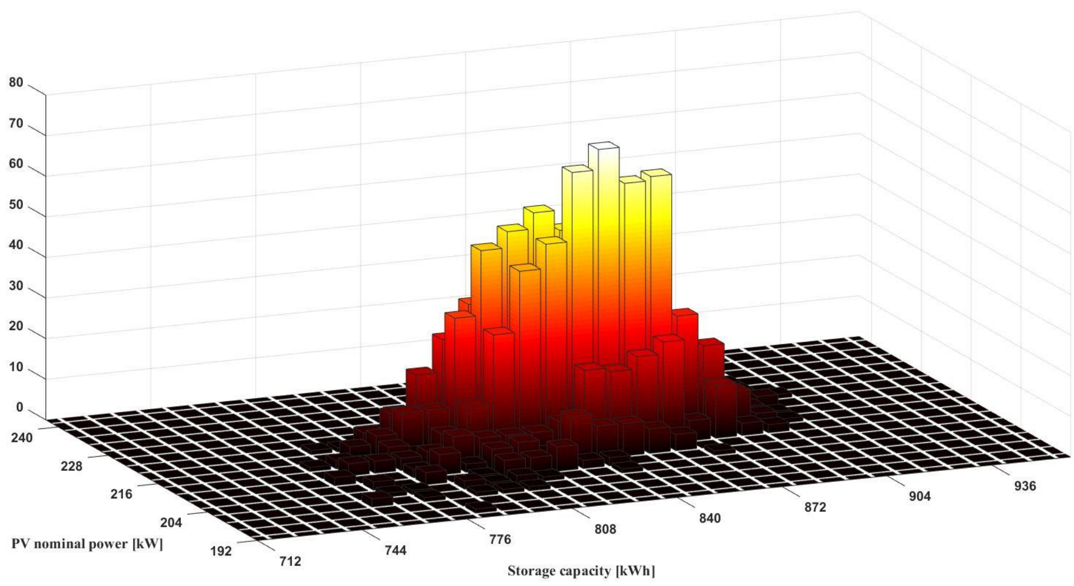


Figure 6.7: Frequency of the different optimum system configuration by aggregating all the scenarios

6.2 Load profile forecast application

In addition to the improvement of the software for the estimation of load profiles, a possible target of the thesis was the validation of this model. However, a proper validation requires to have a number of contexts for which specific data are available: (i) the input data needed by the model and (ii) a good amount of measured load profiles to check if the model's estimates are realistic. In conclusion, due to the difficulty in obtaining the necessary data, it was not possible to perform the validation.

However, with regard to a college in a peripheral urban area of Cameroon, it has been possible to obtain both the input data required by the model, either measured load profiles relative to eight days. Obviously, the metered profiles of only eight days are not enough to formalize the validation of the model, but it has been possible to show the ability of the model in matching metered daily load profiles. Furthermore, a possible procedure for a future validation of the model has been identified and applied to this case study.

Metered load profiles

The reference context is the Cameroon Presbyterian College in Bali, in Cameroon. At the moment the power supply of the college is provided by the national grid, but also backup diesel generators are locally available. As part of a study to perform an energy planning of the school aiming at identifying possible solution towards the college energy self-sufficiency, actual load profiles have been metered. In particular, eight days have been monitored by considering the local meter installed at the connection of the electric system of the school with the local distribution grid. The load was measured by *Sampling method* on a 10-minutes basis.

The load profiles relative to the eight days measured are reported in Figure 6.8. A general trend of the load can be clearly distinguished: the main peak always manifests in the evening hours, while a secondary peak occurs constantly in the early hours of the morning, when probably the customers wake up. The measured peak demand ranges between 11.5 kW and 15.4 kW, while the consumed daily energy ranges between 139.5 kWh and 161.2 kWh.

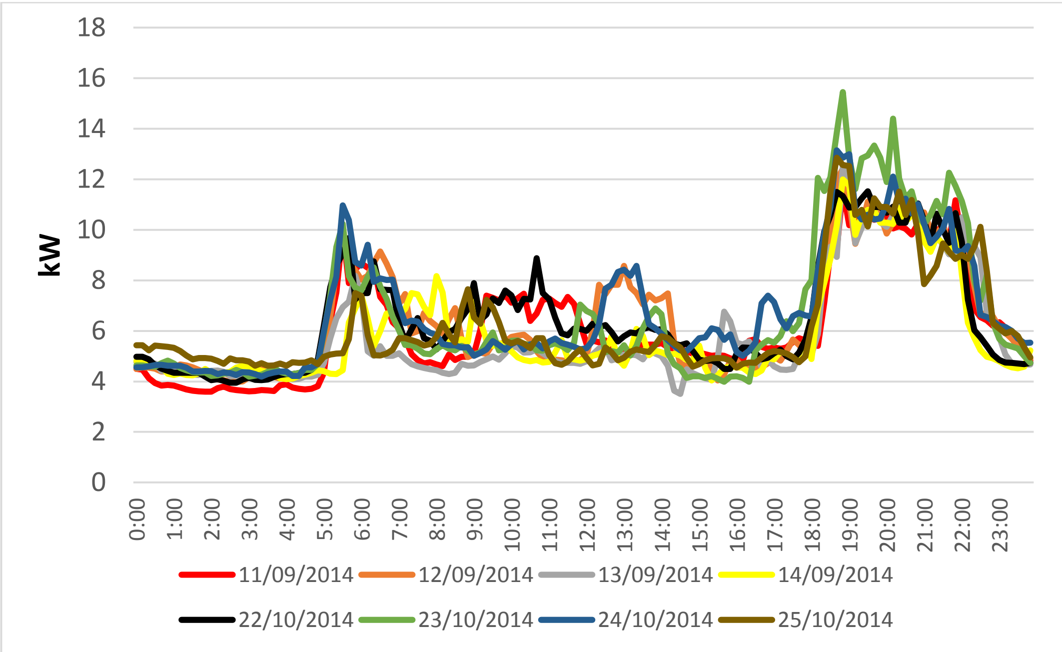


Figure 6.8: Metered load profiles for Cameroon Presbyterian College in Bali, relative to eight days between September and October 2104

The profile to be employed for the comparison with the results of the model is the average of the metered ones (Figure 6.9). The power peak is 12.6 kW, while the consumed daily energy is 151.3 kWh. Obviously, the average profile tends to lose some of the typical fluctuating trend that characterize the daily profiles.

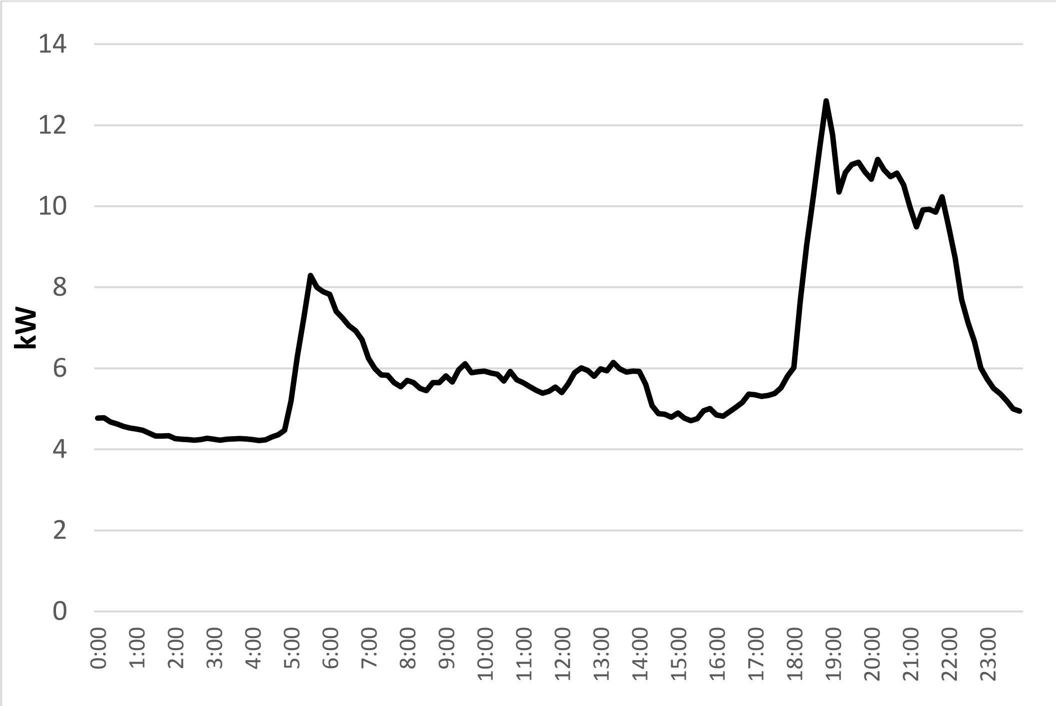


Figure 6.9: Average metered load profile for Cameroon Presbyterian College in Bali

Estimated load profiles

The collection of information was made by two students of *Politecnico di Milano*, who have proposed to the local population a questionnaire (shown in Appendix C), in which general information and energy habits were required. In this way, along with an analysis of the buildings of the territory, some different user-classes have been identified, the number of elements per class, and the electrical appliances installed and used. The detailed input data required by the procedure are shown in Appendix B

Table 6.4 shows the list of the defined user-classes and a summary of the input data in terms of consumed energy are reported (calculated on the basis of input data). However, in order to consider a certain degree of uncertainty related to the input data, both *RFt* and *RFw* have been set to 30% for each appliance of each user-class.

Table 6.4: Summary of energy consumptions for the defined user-classes (without considering the effects of RFt and RFw)

	User-class	N_{users}	$E_{class,day}$ [kWh/day]	$E_{user,day}$ [kWh/day]
1	Family_1	18	51.9	2.9
2	Family_2	14	20.0	1.4
3	Family_3	11	7.6	0.7
4	Students' Dormitories	1	10.4	10.4
5	Classrooms	1	13.8	13.8
6	Kitchen	1	4.4	4.4
7	Bakery	1	1.0	1.0
8	Dining hall	1	0.7	0.7

9	Canteen	1	1.3	1.3
10	Workshop	1	0.7	0.7
11	Dispensary	1	0.4	0.4
12	Church	1	1.7	1.7
13	Administration Office	1	7.9	7.9
14	Library	1	3.8	3.6
15	CCU	1	13.2	13.2
	Total Load		138.8	64.3

The curves estimated by the model are daily curves on a 1-minute basis. In the light of what has been explained in Sampling rate, however, it would not be correct to directly compare these curves with those measured on a 10-minute basis, because the higher the sampling period, the smoother the load profile, and the lower the peak (Figure 2.11). For this reason, the power values of each minute estimated by the algorithm have been averaged every 10 minutes, to obtain the equivalent of curves on a 10-minute basis. An example of three daily estimated load profiles for the context is shown in Figure 6.10.

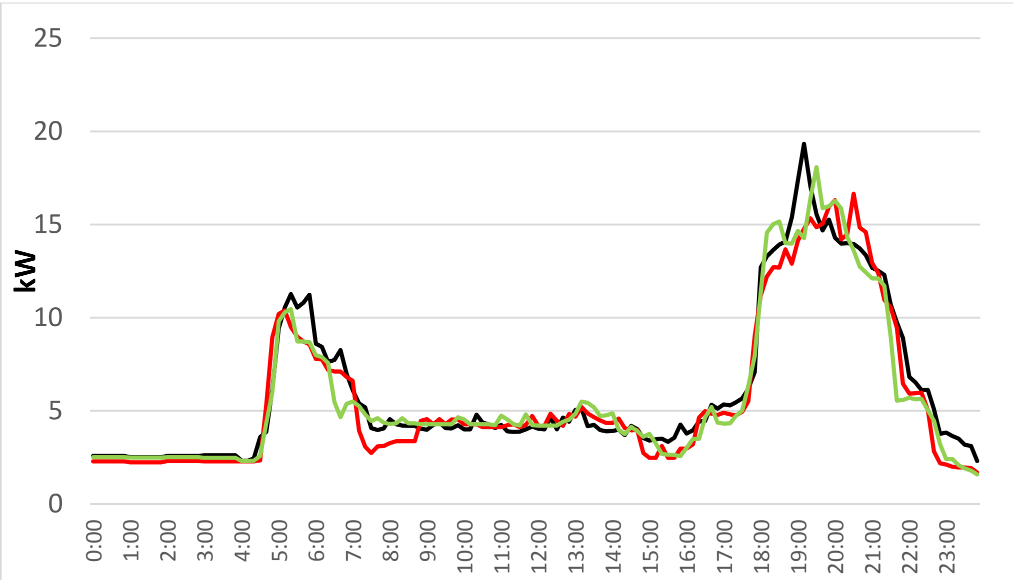


Figure 6.10: Three daily load profiles computed by the model for the Cameroon context

Due to the stochastic nature of the model, the estimated load profile to be compared with the metered one has to refer to an average profile given by the model. As a matter of fact this should represent the profile to which the model converges. Therefore, the considered profile is the average of *N* estimated profiles, when *N* is identified when both the following conditions are met:

6.1

$$\frac{\bar{y}(k)_N - \bar{y}(k)_{N+1}}{\bar{y}(k)_N} \leq 0.5\% \;\; for\ at\ least\ 80\%\ of\ time$$

and

6.2

$$\frac{\overline{std}[y(k)_N] - \overline{std}[y(k)_{N+1}]}{\overline{std}[y(k)_N]} \leq 0.5\% \;\; for\ at\ least\ 80\%\ of\ time$$

where:

- k refers to the profile daily instant, in this case the load profiles are build up according to 10min time steps, i.e. a full day is composed by 144 values.
$$k = [00{:}00 \quad 00{:}10 \quad 00{:}20 \quad \dots \quad \dots \quad \dots \quad 23.30 \quad 23{:}40 \quad 23{:}50]$$
- $\bar{y}(k)_N$ refers to the average load value of *N* generated profiles at the instant *k*.
- $\overline{std}[y(k)_N]$ refers to the average standard deviation of the load value of *N* generated profiles at the instant *k*.

According to this approach, the procedure converges by generating at least 99 profiles. The resulting average profile based on 100 profiles is shown in Figure 6.11. The power peak is 16.3 kW, while the consumed daily energy is 138.1 kWh.

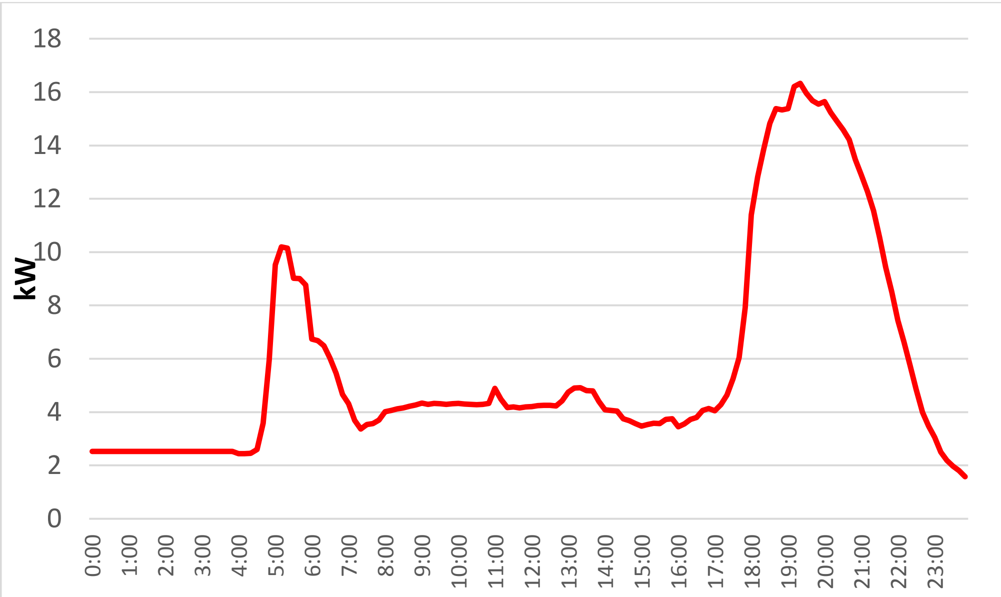


Figure 6.11: Average daily load profile based on 100 load profiles estimated by the model

Comparison (validation procedure)

Figure 6.12 shows the comparison between average metered profile and average estimated load profile. Some considerations can be made:

- The estimated profile generally tends to overestimate power peaks, and underestimate the flat part of the profile (in particular during the night).
- Power peaks occur at the same time.
- Considering the values of the estimated profile at the end of the day and at the beginning, they do not exactly match.
- The "bell" which embraces the power peak of the estimated profile, clearly remind the normal distribution profile.

Moreover, it is worthwhile to mention that the metered profile does not exactly represent the average profile of the college since it is based on only eight metered daily profiles. Nevertheless, despite this limitation, and despite the above mentioned considerations, in the author's opinion the comparison shows that the model is based on a sound approach: the input data have been easily collected or assumed at local level by survey and observation of the consumption habits, the coincidence behavior of the appliances simulated by the procedure fairly matches with the metered one, and the peak power computation is appropriately addressed both in term time and value.

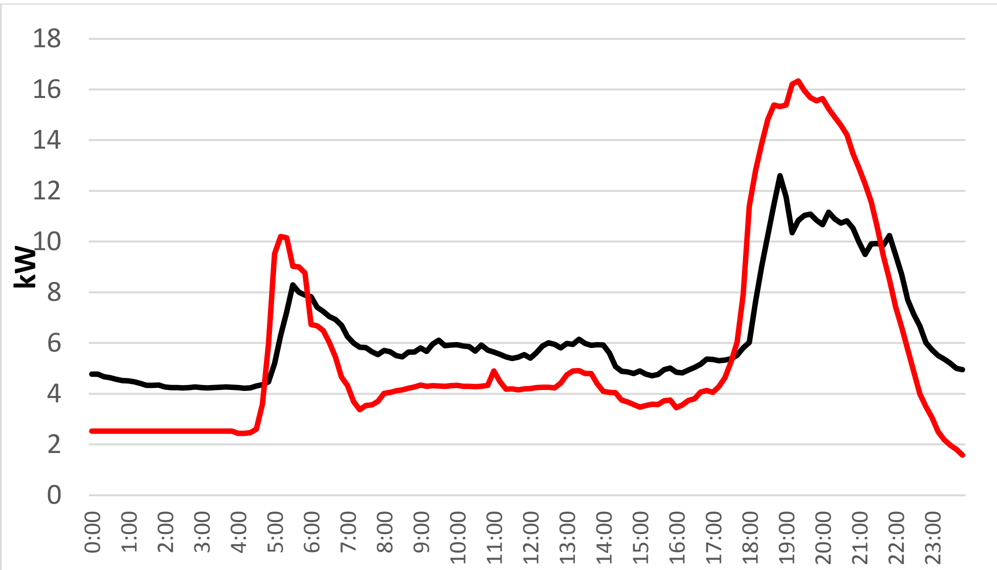


Figure 6.12: Comparison between average metered profile (black line) and average estimated load profile (red line)

Clearly further developments are required in order to improve the procedure. This can only be addressed by a number of comparisons between metered and estimated data. Moreover, although the graphical comparison is an effective tool to understand the goodness of the model, a more rigorous approach is required in order to assess the "match" between metered and estimated profiles. In this regards a set of indicators have been identified:

- Relative error of the daily energy consumption

6.3

$$\varepsilon_{daily\ energy} = \frac{E_{estimated}[Wh] - E_{measured}[Wh]}{E_{measured}[Wh]}$$

- Relative error of the power peak value:

6.4

$$\varepsilon_{peak\ power} = \frac{Peak_{estimated}[W] - Peak_{measured}[W]}{Peak_{measured}[W]}$$

- Error, in minutes, of the peak time.
- *Shape Indicator.* It is an indicator on "how much" the curve estimated is similar to the measured one: the ratio between the area subtended between the difference of the two load profiles (measured and estimated) and the area under the measured load profile, i.e. the energy difference between measured and estimated and the energy measured, for each time step *k*:

6.5

$$I_s = \sum_k \frac{|\bar{y}(k)_{metered} - y(k)_{estimated}|}{\bar{y}(k)_{metered}}$$

In Table 6.5 the indicators have been evaluated for the application to the college in Cameroon.

Table 6.5: Values of the proposed indicators for the considered application

	Value		**Error**
	Metered	**Estimated**	
Daily energy [kWh]	151.3	138.1	-8.7%
Power peak value [kW]	12.6	16.3	+29.3%
Power peak time [hh:mm]	18:50	19:20	+30 min
Shape indicator	0.32		

The average estimated daily energy is slightly lower than the average measured daily energy, but this depends exclusively on the input data: evidently the functioning time have been slightly underestimated, or maybe some appliances have been forgotten, probably appliances working at night when the estimated electrical consumption is slightly more than half of the metered one. Conversely, the estimated peak demand is higher than that measured by almost 30%, while there is a good match in time in which the peak occurs, with a discrepancy of only 30 minutes.

7 Conclusions and future work

7.1 Conclusions

In this work, an existing software for estimating the electrical load profiles of customers living in un-electrified rural areas, developed in MATLAB, has been improved and applied. The aim of this model is to provide a general pattern for the construction of load profiles as realistic as possible where measured data cannot be obtained, and therefore improve the designing accuracy of off-grid systems in this kind of contexts.

Performing optimum design of off-grid power systems reflects in looking for the system that best matches, at local level, energy resources with electric demand given certain technologies and context features. Considering the electric demand, the simplest approaches for system design (intuitive methods) rely on monthly or daily electricity requirement, while the most advanced methodologies (numerical and analytical) and available software (as HOMER) rely on electric load profiles. However, for un-electrified rural areas electric load profiles are not available, so they must be estimated. After a research of articles on projects dealing with sizing of power systems for off-grid electrification of rural areas, I have concluded that there is not a definite procedure for estimation of load profiles, and this phase is often overlooked. Usually estimations are based on simple assumptions about electric appliances and user practices. Such assumptions and the related estimation methods provide adequately accuracy for intuitive methods, but not for Load Curve Computation. Indeed they usually do not implement main features of load profiles (e.g. load factors, coincidence factors, etc.). Furthermore when load profiles are computed according to these simple assumptions, the uncertainty given by the fact that several load profiles can occur within the same set of assumptions is not considered

The model for load profile estimation is designed to work with few input data that are commonly considered in the simplest approaches for energy need estimations in rural areas (appliances nominal power, duration and time of functioning) and this is coherent with the data that may be available in the context under study. Electric load profile(s) are then computed by mean of the new model that combines the input data employing approaches and parameters of advanced energy needs estimation models, coupling:

- Bottom-up approach: the system load profile is obtained by a bottom-up hierarchic aggregation of coincident load profiles obtained for each end-use appliances employed by each class of user.

- Stochastic approach: the core feature of the model is the computation of the coincident load profile through a stochastic simulation (i.e. weighted sampling without replacement). The sampling simulation set the instants of peak power time and appliances switching on employing a probability density function made up with a uniform distribution coupled with a normal distribution (located in the peak power period).
- Relation between coincidence factor and load factor: empirical evidence shows a relationship between group coincident factors and average load factors, hence the coincidence factor for each user class is computed given the average user load factor (computed from the input assumption) employing the empirical curve.
- Coincidence curves (i.e. coincidence factor as a function of number of users): empirical evidence shows a relationship between group coincidence factor and the number of consumers in a group, hence the coincidence factor of each class of user is corrected as for the number of consumers in that class by employing empirical coincidence curves.

Such a model aims to provide a tool which is valid as a support for input data generation for the actual off-grid system design approaches and software. Furthermore, the model accounts for the high load profile uncertainty of un-electrified areas (under the given conditions several possible load profiles are indeed possible) because each simulation generates a different possible load profile.

The improvements made to the existing model have (i) introduced the possibility of considering the uncertainty of the input data, (ii) reduced the rigidity of switching-on instants sampling, increased the efficiency and execution speed of simulation, (iii) improved the management of variance of the normal distribution that leads the peak's switching-on, which now vary adaptively with respect to error.

The model has been applied for a first step sizing of the main components of a PV Micro-Grid in a peripheral urban area of Uganda. The software has provided the input load profiles requested by a sizing method based on life-cycle Net Present Cost and minimum Loss of Load parameter. In order to address the issue of input data uncertainty, many combinations of uncertainty parameters (i.e. *RFt* and *RFw*) have been considered.

Finally, it has been possible to show the ability of the model in matching metered daily load profiles related to a college in a peripheral urban area of Cameroon. The load profiles estimated by the model have been compared to the metered ones, and good match has been observed.

Furthermore, a possible procedure for a future validation of the model has been identified and applied to this case study, through the definition and calculation of some indicators.

7.2 Future works

The model for the estimation of load profiles in this work has been improved in many aspects with respect to the previous version. However, some problems are still present, and certainly further improvements can be implemented to make the model even more functional. In the following I have proposed a list of possible improvements for the model, which could be used as a starting point for future work.

- The first critical aspect of the model is the empirical relationship used in phase of *Peak Value Computation*, which relates the coincidence factor as a function of the load factor and the number of users, and with which the peak demand for each user-class is estimated. Indeed, the coefficients of the empirical correlation have been obtained from a study of Constantine Bary in the period of World War II, carried out on residential and commercial users in urban and rural areas in Philadelphia. Certainly, given the remoteness of the historical period and energy habits compared to the utilization context for which the model has been designed, it is hard to imagine that such empirical coefficients are ideal for their application load estimation of un-electrified rural areas. Obviously, in order to find new correlations that better suit to the context would require numerous measurements of electrical loads of rural areas recently electrified, but this would require the implementation of a wide project.
- Another critical point is still about the *Peak Value Computation* phase. In fact, in the iterative procedure that leads to the calculation of the peak, the n^{th} attempt's peak value is computed in this way:

  7.1

  $$peak_k(n) = Cf_k(n) * max\_peak_j [W]$$

  where $peak_k(n)$ is the n^{th} attempt's peak value of user-class k, $Cf_k(n)$ is the n^{th} attempt's coincidence factor calculated through Bary's correlation and max_peak_j is the theoretical maximum peak value (obtained turning on all the devices of all the appliances within their functioning windows). The implementation of this formula is not properly correct: in fact, the *Cf* is defined as the ratio between the peak (per customer) of a group of users and the peak of a single user (Coincidence factor and coincidence curve). In the model, therefore, assumes that the peak of a single user is the peak theoretical maximum (per customer). A viable choice would be to perform a pre-simulation of a single user for each user-class, in order to apply the correct definition of *Cf*. This would also lead to a reduction of the estimated peak, and this maybe could benefit in the light of the results obtained in the application of the model illustrated in 6.2.

- Currently, the correction of the probabilistic weights with which perform the sampling of switching-on instants can only be changed upwards: in fact, the probability of sampling around the instant of the peak (for peak appliances) follows a normal distribution, which with low variance increase the probability of appliances' switching-on around the peak, while with very high variance tends to flatten and becomes a uniform distribution, not increasing the probability of switching-on around the peak. This solution works well if the user-class peak estimated in the Peak Load Computation is greater than the peak resulting from a simulation performed with uniform distribution (or normal distribution having high variance). Instead, if the estimated peak is lower, the current model does not provide the possibility to decrease the probability of switching-on of the devices around the peak, so struggles to generate a load profile having the estimated peak.
- The model considers the uncertainty on the input data with the two factors *RFt* and *RFw*. These are the two maximum variation values for *Ft* and *Fw(s)* of appliances. Currently, for each appliance is randomly selected a value between *-RFt* and *+RFt* (in percentage) with which modify the value of *Ft*, and the same is done for the *Fw(s)*. This means that the randomly chosen factors with which correct *Ft* and *Fw* is the same for all the devices of the same appliance. A possible improvement of the model would be to introduce randomization of *Ft* and the *Fw(s)* for each device, not only for each appliance.
- It would be necessary to develop a new management of those appliances that run all day, but they consume electricity usually at regular intervals and for short periods during the day. A typical appliance which operates in this manner is the fridge: it is on during all the day, but consumes electricity for brief moments at intervals approximately constant (except when it is opened many times in a short period). Assume that the fridge turns on 24 times a day for 10 minutes each. In the current model, if a single functioning window lasts all day was set, and the random selection of switching-on instants was performed, it could happen that the majority of switching-on instants were concentrated within a certain period of the day. This would not be realistic, because the refrigerator distributes in a substantially homogeneous manner its switching-on during the day.

- Finally, it might be useful to reorganize the “for” loops in Matlab code in order to activate the “Parallel pool”, which is a Matlab function that allows to use all the computer's processor cores, and thus drastically reduce the execution time.

Appendix A

Load data assumptions for the Micro-Grid area in Soroti (Uganda)

User class	N_{us}	App Name	P [W]	N_{app}	Ft [h]	Fw_1		Fw_2		Fw_3		Tot_Fw
						h_{start}	h_{stop}					
Family_1	50	Lights	3	4	6	0	2	17	24	-	-	9
		Phone Charger	5	2	3	0	9	13	15	17	24	18
		Security Light	5	1	12	0	7	17	24	-	-	14
Family_2	15	Lights	3	4	6	0	2	17	24	-	-	9
		Phone Charger	5	2	3	0	9	13	15	17	24	18
		Security Light	5	1	12	0	7	17	24	-	-	14
		Radio	5	1	4	6	9	17	24	-	-	10
		AC-TV (small)	100	1	5	11	15	17	24	-	-	11
Family_3	15	Lights	3	8	6	0	2	17	24	-	-	9
		Phone Charger	5	2	3	0	9	13	15	17	24	18
		Radio	5	1	4	6	9	17	24	-	-	10
		Security Light	5	2	12	0	7	17	24	-	-	14
		AC-TV (small)	100	1	5	11	15	17	24	-	-	11
		Fridge (small)	250	1	5	0	24	-	-	-	-	24
Family_4	10	Lights	3	12	6	0	2	17	24	-	-	9
		Phone Charger	5	4	3	0	9	13	15	17	24	18
		Radio	5	1	4	6	9	17	24	-	-	10
		Security Light	5	4	12	0	7	17	24	-	-	14
		AC-TV (small)	100	1	5	11	15	17	24	-	-	11
		Standing Fan	55	1	6	8	24	-	-	-	-	16
		Decoder	15	1	5	11	15	17	24	-	-	11
		Fridge (small)	250	1	5	0	24	-	-	-	-	24
		Internet Router	20	1	6	0	24	-	-	-	-	24
		Laptop (small)	55	1	6	0	2	11	15	17	24	13
Family_5	5	Lights	3	16	6	0	2	17	24	-	-	9
		Phone Charger	5	4	3	0	9	13	15	17	24	18
		Radio	5	2	4	6	9	17	24	-	-	10
		Security Light	5	6	12	0	7	17	24	-	-	14
		AC-TV (big)	200	1	6	11	15	17	24	-	-	11
		Standing Fan	55	2	6	8	24	-	-	-	-	16
		Decoder	15	1	6	11	15	17	24	-	-	11
		Fridge (big)	400	1	5	0	24	-	-	-	-	24
		Internet Router	20	1	8	0	24	-	-	-	-	24
		Laptop (big)	80	2	8	0	2	11	15	17	24	13
Family_6	5	Lights	3	16	6	0	2	17	24	-	-	9
		Phone Charger	5	4	3	0	9	13	15	17	24	18
		Radio	5	2	4	6	9	17	24	-	-	10
		Security Light	5	6	12	0	7	17	24	-	-	14
		AC-TV (big)	200	1	6	11	15	17	24	-	-	11
		Standing Fan	55	2	6	8	24	-	-	-	-	16
		Decoder	15	1	6	11	15	17	24	-	-	11
		Fridge (big)	400	1	5	0	24	-	-	-	-	24
		Internet Router	20	1	8	0	24	-	-	-	-	24
		Laptop (big)	80	2	8	0	2	11	15	17	24	13
		Hair Dryer	1000	1	0.5	17	24	-	-	-	-	7
		Printer	50	1	0.5	17	24	-	-	-	-	7
		Stereo	100	1	3	17	24	-	-	-	-	7
		Water Heater	660	1	2	0	2	18	24	-	-	8
Enterprise_1	15	Fluor. Tube (small)	36	10	6	7	11	16	20	-	-	8
		Phone Charger	5	4	3	7	13	15	20	-	-	11
		Security Light	5	4	12	0	7	17	24	-	-	14
		Internet Router	20	1	10	7	20	-	-	-	-	13
		Laptop (big)	80	1	8	7	13	15	20	-	-	11
		Laptop (small)	55	5	8	7	13	15	20	-	-	11
		Printer	50	2	2	7	13	15	20	-	-	11
		Standing Fan	55	2	8	7	13	15	20	-	-	11
Enterprise_2	5	Fluor. Tube (big)	47	20	6	7	11	16	20	-	-	8

		Phone Charger	5	15	3	7	13	15	20	-	-	11
		Security Light	5	10	12	0	7	17	24	-	-	14
		Internet Router	20	1	10	7	20	-	-	-	-	13
		Laptop (big)	80	5	8	7	13	15	20	-	-	11
		Laptop (small)	55	10	8	7	13	15	20	-	-	11
		Standing Fan	55	5	8	7	13	15	20	-	-	11
		Water dispenser	550	1	3	7	13	15	20	-	-	11
		Photocopier	750	1	1	7	13	15	20	-	-	11
		Ceiling Fan	75	5	8	7	13	15	20	-	-	11
		PC	400	1	10	7	20	-	-	-	-	13
Mobile Money	5	Lights	3	2	3	8	11	16	20	-	-	7
		Phone Charger	5	3	3	8	18	-	-	-	-	10
		Standing Fan	55	1	6	10	18	-	-	-	-	8
Kiosk	10	Lights	3	2	3	8	11	16	20	-	-	7
		Phone Charger	5	1	3	8	18	-	-	-	-	10
		Standing Fan	55	1	6	10	18	-	-	-	-	8
		Fridge (small)	300	1	8	0	24	-	-	-	-	24
		Fridge (big)	500	1	8	0	24	-	-	-	-	24
Barber	2	Lights	3	5	8	8	13	15	20	-	-	10
		12V shaver	10	5	6	8	13	15	20	-	-	10
		Ceiling Fan	75	3	8	8	13	15	20	-	-	10
		UV sterilizer	50	1	2	8	13	15	20	-	-	10
Tailor	3	Lights	5	3	8	8	13	15	20	-	-	10
		Sewing machine	50	1	3	8	13	15	20	-	-	10
		Ceiling Fan	75	1	8	8	13	15	20	-	-	10
Market Place	1	Lights	3	25	3	8	11	16	20	-	-	7
		Security Light	5	25	12	0	7	17	24	-	-	14
		Fridge (small)	300	3	8	0	24	-	-	-	-	24
		Fridge (big)	500	3	8	0	24	-	-	-	-	24
		Standing Fan	55	10	8	8	13	15	20	-	-	10
		Radio	5	10	4	10	13	15	18	-	-	6
Club	3	Fluor. Tube (small)	36	10	8	0	4	17	24	-	-	11
		Fluor. Tube (big)	47	5	8	0	4	17	24	-	-	11
		Security Light	5	5	12	0	7	17	24	-	-	14
		Phone charger	5	10	8	15	24	-	-	-	-	9
		AC-TV (small)	130	2	9	0	4	15	24	-	-	13
		AC-TV (big)	200	1	9	0	4	15	24	-	-	13
		PC	400	1	9	0	4	15	24	-	-	13
		Laptop (big)	80	10	6	15	24	-	-	-	-	9
		Printer	50	1	1	15	20	-	-	-	-	5
		Pico Projector	18	1	4	0	2	20	24	-	-	6
		Amplifier	6	1	4	0	2	20	24	-	-	6
		Ceiling Fan	75	3	8	0	4	15	24	-	-	13
		Music System	178	1	8	0	4	15	24	-	-	13
		Internet Router	20	1	9	0	4	15	24	-	-	13
		Fridge (small)	300	2	8	0	24	-	-	-	-	24
		Fridge (big)	500	1	8	0	24	-	-	-	-	24
Street Lights	1	Lights (Street)	50	100	12	0	7	17	24	-	-	14
		Led strips	8	100	12	0	7	17	24	-	-	14
Primary School	1	Fluor. Tube (small)	36	10	4	8	17	-	-	-	-	9
		Phone Charger	5	7	3	8	17	-	-	-	-	9
		Security Light	5	4	12	0	7	17	24	-	-	14
Pharmacy	1	Lights	3	10	3	8	11	16	20	-	-	7
		Security Light	5	4	12	0	7	17	24	-	-	14
		Fridge (small)	300	3	8	0	24	-	-	-	-	24
		Fridge (big)	500	2	8	0	24	-	-	-	-	24
		Standing Fan	55	3	8	8	13	15	20	-	-	10

Appendix B

Load data assumptions for the Cameroon Presbyterian College in Bali

User class	N_{US}	App Name	P [W]	N_{App}	Ft [min]	Fw_1		Fw_2		Fw_3		Tot_Fw
						h_{start}	h_{stop}					
Family_1	18	TV	80	1	360	16:00	22:30	-	-	-	-	6:30
		Stereo SET	36	1	420	5:30	7:30	14:00	20:00	-	-	8:00
		Phone charger	5	3	240	0:00	6:00	22:00	24:00	-	-	8:00
		Indoor bulb	26	5	300	5:00	7:00	18:00	22:00	-	-	6:00
		Outdoor light	26	1	120	18:00	22:00	-	-	-	-	4:00
		Security light	5	1	720	0:00	6:00	18:00	24:00	-	-	12:00
		Fridge	40	1	1440	0:00	24:00	-	-	-	-	24:00
		PC	50	1	120	17:00	21:00	-	-	-	-	4:00
		Iron	800	1	6	5:30	6:00	19:00	20:30	-	-	2:00
		DVD	15	1	360	16:00	22:30	-	-	-	-	6:30
		Flask	700	1	30	5:00	5:30	-	-	-	-	0:30
		Blender	350	1	45	11:00	11:20	13:00	14:00	-	-	1:20
Family_2	14	TV	80	1	300	6:00	7:00	16:00	23:00	-	-	8:00
		Radio	5	1	240	5:00	6:30	17:00	22:00	-	-	6:30
		Stereo SET	36	1	90	20:00	22:00	-	-	-	-	2:00
		Phone charger	5	2	240	0:00	6:00	22:00	24:00	-	-	8:00
		Indoor bulb	26	4	300	5:00	8:00	17:30	23:30	-	-	9:00
		Outdoor bulb	26	1	240	18:00	22:00	-	-	-	-	4:00
		Security Light	26	1	720	0:00	6:00	18:00	24:00	-	-	12:00
		Iron	800	1	6	5:30	6:30	19:30	21:00	-	-	2:30
		DVD	15	1	300	6:00	7:00	16:00	23:00	-	-	8:00
Family_3	11	Ph. charger	5	2	180	0:00	6:00	22:00	24:00	-	-	8:00
		TV	85	1	240	16:00	22:30	-	-	-	-	6:30
		Bulb	26	3	300	5:00	6:30	18:00	22:30	-	-	6:00
		Iron	800	1	15	19:00	20:00	-	-	-	-	1:00
Dormitories	1	Bulb	26	32	120	5:00	6:30	18:00	19:00	-	-	2:30
		Tube	36	31	120	5:00	6:30	18:00	19:00	-	-	2:30
		Security light	26	21	720	0:00	6:00	18:00	24:00	-	-	12:00
Classrooms	1	Bulb	26	49	300	5:00	7:00	18:30	21:30	-	-	5:00
		Tube	36	8	300	5:00	7:00	18:30	21:30	-	-	5:00
		Safety bulbs	30	14	720	0:00	6:00	18:00	24:00	-	-	12:00
		Safety tubes	40	2	720	0:00	6:00	18:00	24:00	-	-	12:00
Kitchen	1	Bulb	26	6	690	5:30	11:00	12:00	15:00	17:00	20:00	11:30
		Radio	5	1	690	5:30	11:00	12:00	15:00	17:00	20:00	11:30
		Sharpener	50	1	1	5:30	11:00	-	-	-	-	5:30
		Fridge	53	2	1440	0:00	24:00	-	-	-	-	24:00
Bakery	1	Bulb	26	4	600	6:00	16:00	-	-	-	-	10:00
Refactory	1	Bulb	26	5	90	18:30	19:30	-	-	-	-	1:00
		Tube	36	9	90	18:30	19:30	-	-	-	-	1:00
Canteen	1	Bulb	26	1	270	8:00	9:00	14:00	15:00	17:00	19:30	4:30
		Tube	10	1	270	8:00	9:00	14:00	15:00	17:00	19:30	4:30
		Incandescent L.	40	1	270	8:00	9:00	14:00	15:00	17:00	19:30	4:30
		Fridge	40	1	1440	0:00	24:00	-	-	-	-	24:00
Workshop	1	Bulb	26	1	1440	0:00	24:00	-	-	-	-	24:00
		Radio	5	1	1440	0:00	24:00	-	-	-	-	24:00
Dispensary	1	Bulb	26	1	390	8:00	12:00	16:00	18:00	19:00	21:30	10:30
		Tube	36	1	390	8:00	12:00	16:00	18:00	19:00	21:30	10:30
Church	1	Bulb	26	8	210	6:00	7:00	19:00	21:30	-	-	3:30
		Tube	36	8	210	6:00	7:00	19:00	21:30	-	-	3:30
Ad. office	1	Bulb	26	4	540	7:30	16:30	-	-	-	-	9:00
		Tube	40	7	540	7:30	16:30	-	-	-	-	9:00
		Mini tube	18	1	420	7:30	14:30	-	-	-	-	7:00
		Electronics	32	19	402	7:30	14:30	-	-	-	-	7:00
Library	1	Tubes	40	12	420	7:00	14:00	-	-	-	-	7:00
		Photocopier	32	1	420	7:00	14:00	-	-	-	-	7:00
CCU	1	Bulb	26	4	480	8:00	16:00	-	-	-	-	8:00
		Tube	36	11	480	8:00	16:00	-	-	-	-	8:00

Laptop	55	18	480	8:00	16:00	-	-	-	-	8:00
Printer_1	550	4	10	8:00	16:00	-	-	-	-	8:00
Printer_2	510	1	30	8:00	16:00	-	-	-	-	8:00
Photocopy1	1280	1	10	8:00	16:00	-	-	-	-	8:00
Photocopy2	1300	2	5	8:00	16:00	-	-	-	-	8:00
Standby	35	1	480	8:00	16:00	-	-	-	-	8:00

Appendix C

Questionnaire for households at CPC Bali about energy consumption and supply

Instructions

- Tick the right answer □ with a X
- Fill in the space ………. when is needed
- Please specify if your answer is not listed
- Complete the empty cells where the answer is in form of a table

1. General information

House number: ……… Date: …../….. /…….

Sex: ☐ M ☐ F Age: …

1.1 Position in the family:

□ Head of the family
□ Family member
□ Other (please specify: ………………………………)

1.2 Number and sex of household members: Female … Male …

1.3 Age of household members:

	Infant (< 6)	6 - 16	16 - 30	30 - 65	above 65
female					
male					
total					

1.4 Which are the roles of the family members in CPC Bali?

□ CPC Administration (principal, vice-principal, pastor, finance clerk)
□ Teaching staff
□ Non-teaching staff
□ Student
□ Other (please specify: ……………………………………….)

2. Electricity use and supply

2.1 Is your house connected to the national electric grid (AES-SONEL)? ☐ Yes ☐ No

2.2 Which are the devices using electricity in the house?

Items	Number	Average daily use [h]
TV		
radio		
phone charger		
indoor lights (lamp, bulb)		
outdoor lights (lamp, bulb)		
fridge		
laptop		
electric iron		
others (please specify……………………………………………)		

2.3 Do you have a meter installed in your household? ☐ Yes ☐ No

2.4 Does someone in the household check the electricity consumption? ☐ Yes ☐ No

2.5 Which is the average consumption per month? ……… kWh

2.6 Which is the average bill for electricity for this household? ……. FCFA

2.7 Could you estimate how many shortages occur per month?

	n° of shortages per month
below 10 minute	
10 minutes – 1 hour	
above 1 hour	

2.8 Have you ever had any devices damaged or broken due to voltage fluctuations of the grid?

……

……

2.9 Do you have any comments about the electricity supply?

……

……

3. Firewood use and supply

Which are the main purposes of using firewood?

- ☐ Cooking
- ☐ Space heating
- ☐ Water heating (hot bath, tea, laundry, etc.)
- ☐ Others (specify: …………………………………………………………………)

□ Firewood is not used

3.1 Firewood for cooking

3.1.1. Which are the cooking systems used?

□ 3 stones fire
□ saw dust stove
□ improved stove
□ other (please specify ……………………..)

3.1.2. Where is the kitchen placed in your household?

□ inside the house
□ outside the house
□ common kitchen with other households

3.1.3. On a daily basis, how many meals do you cook using firewood? ……

3.2 Firewood for space heating

3.2.1. How often is firewood used for space heating:

□ daily
□ weekly
□ monthly
□ only during the rainy season
□ yearly

3.3 Firewood for water heating

3.3.1. How many days per week do you heat water by burning firewood?

□ 1-3
□ 3-5
□ 5-7
□ 7

3.3.2. On average, can you estimate how much water do you boil every time? …..

3.4 Firewood supply

3.4.1. Where do you take the firewood?

□ inside the campus for free
□ inside the campus by purchasing it
□ outside the campus for free (e.g. your farm)
□ outside the campus by purchasing it
□ other (please specify ……………………………………………………………)

3.4.2. How is the firewood transported to your household?

- ◻ by hands
- ◻ by motorbike
- ◻ by car
- ◻ by pick-up
- ◻ by truck
- ◻ other (please specify ……………………………………………………………………)

3.4.3. How often do you provide firewood to your household? Once every ……………………

3.4.4. H o w much do you purchase?

(Use the mean of transportation above as unit of measure) ……………………………………

3.4.5. How much do you pay for it? …………………….. FCFA

3.4.6. In your opinion, do you think the use of wood as fuel increases the ongoing deforestation (in Cameroon)?

- ◻ Definitely yes
- ◻ Yes, but only in the future if no other sources will replace firewood
- ◻ I don't think so
- ◻ I've no idea

4. Gas use and supply

4.1. Do you use gas in your household? ☐ Yes ☐ No (<u>if No skip to section 5</u>)

4.2. Which is the main purpose of using gas?

- ◻ Cooking
- ◻ Water heating
- ◻ Others (please specify ………………………………………………………………)

4.3. Which type of gas bottle do you have?

- ◻ SCTM (orange)
- ◻ CAM GAS
- ◻ TOTAL GAS
- ◻ GLOCALGAZ
- ◻ OILYBYA

4.4. Which is the size of your gas bottle?

- ◻ Extra large (35 kg)
- ◻ Medium (12.5 kg)
- ◻ Small (6 kg)

4.5. How often do you refill it on average? Every ……………

4.6. How much do you pay? ……………….. FCFA

4.7. On a weekly basis, how many meals do you cook using firewood or gas?

……… meals cooked with gas ……… meals cooked with firewood

4.8. Is there any other reason, aside the type of meal, of using firewood instead of gas?

☐ Yes ……………………………………………………………………………

☐ No

5. Other sources of energy

5.1. Do you use any other source of energy?

	Main purpose	How much	Cost
☐ Charcoal			
☐ Kerosene			
☐ Candles			
☐ Others (specify…………………………)			

6. Domestic waste production

6.1. Do you have a farm? ☐ Yes ☐ No

6.2. Which kinds of crops do you have?

	Which is approximately the extension?
☐ Corn	
☐ Banana/plantain	
☐ Beans	
☐ Cereals	
☐ Tuber (potatoes, yam, carrots, etc.)	
☐ Pineapple	
☐ Vegetables	
☐ Other fruit trees (mango, papaya, etc.)	

6.3. Do you consume the products of your farm?

□ Yes, we consume all of the products
□ Only part is consumed, the rest is sold away
□ All is sold

6.4. Which animals do you have (if any)?

	N°	How do you feed them? (choose among: animal feed, grass/bush, organic waste, mixture of animal feed and cereals)
pigs		
goats		
cows		
rabbits		
agrifowl		
countryfowl		

Thanks for your time and for the great collaboration!!!

References

[1] E. Colombo, S. Bologna, and D. Masera, *Renewable Energy for Unleashing Sustainable Development*. Springer International Pu, 2013.

[2] C. Kirubi, A. Jacobson, D. M. Kammen, and A. Mills, "Community-Based Electric Micro-Grids Can Contribute to Rural Development: Evidence from Kenya," *World Dev.*, vol. 37, no. 7, pp. 1208–1221, 2009.

[3] O. H. P. U. P. Report, "United Nations Development Program," New York, 2000.

[4] "http://databank.worldbank.org/." .

[5] C. S. Sinha and T. C. Kandpal, "Decentralised versus grid electricity for rural India," *Energy Policy*, pp. 441–448, 1991.

[6] A. Skoczek, T. Sample, and E. D. Dunlop, "The results of performance measurements of field-aged crystalline silicon photovoltaic modules," *Prog. Photovoltaics*, 2009.

[7] M. M. H. Bhuiyan and M. Ali Asgar, "Sizing of a stand-alone photovoltaic power system at Dhaka," *Renew. Energy*, vol. 28, pp. 929–938, 2003.

[8] M. Sidrach-de-Cardona and L. Mora López, "A simple model for sizing stand alone photovoltaic systems," *Sol. Energy Mater. Sol. Cells*, vol. 55, pp. 199–214, 1998.

[9] T. Khatib, a. Mohamed, K. Sopian, and M. Mahmoud, "Optimal sizing of building integrated hybrid PV/diesel generator system for zero load rejection for Malaysia," *Energy Build.*, vol. 43, no. 12, pp. 3430–3435, 2011.

[10] V. K. Sharma, A. Colangelo, and G. Spagna, "Photovoltaic technology: Basic concepts, sizing of a stand alone photovoltaic system for domestic applications and preliminary economic analysis," *Energy Convers. Manag.*, vol. 36, no. 3, pp. 161–174, Mar. 1995.

[11] A. Fragaki and T. Markvart, "Stand-alone PV system design: Results using a new sizing approach," *Renew. Energy*, vol. 33, pp. 162–167, 2008.

[12] M. Sidrach-de-Cardona and L. Mora López, "A general multivariate qualitative model for sizing stand-alone photovoltaic systems," *Sol. Energy Mater. Sol. Cells*, vol. 59, pp. 185–197, 1999.

[13] R. Posadillo and R. López Luque, "A sizing method for stand-alone PV installations with variable demand," *Renew. Energy*, vol. 33, pp. 1049–1055, 2008.

[14] a. Hadj Arab, B. Ait Driss, R. Amimeur, and E. Lorenzo, "Photovoltaic systems sizing for Algeria," *Sol. Energy*, vol. 54, no. 2, pp. 99–104, 1995.

[15] A. Mellit, "ANN-based GA for generating the sizing curve of stand-alone photovoltaic systems," *Adv. Eng. Softw.*, vol. 41, no. 5, pp. 687–693, 2010.

[16] A. Mellit, M. Benghanem, a. H. Arab, and A. Guessoum, "An adaptive artificial neural network model for sizing stand-alone photovoltaic systems: Application for isolated sites in Algeria," *Renew. Energy*, vol. 30, pp. 1501–1524, 2005.

[17] T. Markvart, A. Fragaki, and J. N. Ross, "PV system sizing using observed time series of solar radiation," *Sol. Energy*, vol. 80, pp. 46–50, 2006.

[18] T. Khatib, A. Mohamed, and K. Sopian, "A review of photovoltaic systems size optimization techniques," *Renew. Sustain. Energy Rev.*, vol. 22, no. 0, pp. 454–465, Jun. 2013.

[19] R. Sen and S. C. Bhattacharyya, "Off-grid electricity generation with renewable energy technologies in India: An application of HOMER," *Renew. Energy*, vol. 62, pp. 388–398, Feb. 2014.

[20] M. I. Howells, T. Alfstad, D. G. Victor, G. Goldstein, and U. Remme, "A model of household energy services in a low-income rural African village," *Energy Policy*, vol. 33, no. 14, pp. 1833–1851, 2005.

[21] G. Krajačić, N. Duić, and M. da G. Carvalho, "H2RES, Energy planning tool for island energy systems – The case of the Island of Mljet," *Int. J. Hydrogen Energy*, vol. 34, no. 16, pp. 7015–7026, Aug. 2009.

[22] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the UK," *Energy Build.*, vol. 37, no. 6, pp. 663–671, Jun. 2005.

[23] 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, Dec. 2012.

[24] C. Bary, "Coincidence-Factor Reationships of Electric-Service-Load Characteristics," *American Institute of Electrical Engineers, Transactions of the*, vol. 64, no. 9. pp. 623–629, 1945.

[25] R. F. Hamilton, "The summation of load curves," *Electrical Engineering*, vol. 63, no. 10. pp. 729–735, 1944.

[26] Dip. di Energia POLIMI, "UNESCO Chair in Energy for Sustainable Development," 2013. .

[27] R. F. Hamilton, "Synthetic or Equivalent Load Curves," *AIEE Trans.*, vol. 61, pp. 369–381, 1942.

[28] C. Bary, "Economic significance of load characteristics as applied to modern electric service," *53rd Annu. Meet. Assoc. Edison Illum. Co.*, 1938.

[29] S. A. Soliman and A. M. Al-Kandari, *Electrical Load Forecasting: Modeling and Model Construction*. Elsevier Science, 2010.

[30] I. S. Moghram and S. Rahman, “Analysis and evaluation of five short-term load forecasting techniques,” *IEEE Trans. Power Syst.*, vol. 4, no. 4, pp. 1484–1491, 1989.

[31] N. Jia, R. Yokoyama, Y. Zhou, and Z. Y. Gao, “A flexible long-term load forecasting approach based on new dynamic simulation theory—GSIM,” *Electr. Power Energy Syst.*, vol. 23, 2001.

[32] L. G. Swan and V. I. Ugursal, “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 8, pp. 1819–1835, Oct. 2009.

[33] N. Amral, C. S. Ozveren, and D. King, “Short term load forecasting using Multiple Linear Regression,” *2007 42nd Int. Univ. Power Eng. Conf.*, pp. 1192–1198, 2007.

[34] T. Hong, P. Wang, and H. L. Willis, “A Naïve Multiple Linear Regression Benchmark for Short Term Load Forecasting,” *Power Energy Soc. Gen. Meet. 2011 IEEE*, no. 2, pp. 1–6, 2011.

[35] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, “Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1622–1630, 2005.

[36] V. Dordonnat, S. J. Koopman, M. Ooms, a. Dessertaine, and J. Collet, “An hourly periodic state space model for modelling French national electricity load,” *Int. J. Forecast.*, vol. 24, no. 4, pp. 566–587, 2008.

[37] J. W. Taylor, “Short-term electricity demand forecasting using double seasonal exponential smoothing,” *J. Oper. Res. Soc.*, vol. 54, pp. 799–805, 2003.

[38] H. S. Hippert, C. E. Pedreira, and R. C. Souza, “Neural networks for short-term load forecasting: A review and evaluation,” *IEEE Trans.*, pp. 44–55, 2001.

[39] D. Srinivasan, “Parallel neural network-fuzzy expert system strategy for short-term load forecasting : System implementation and performance evaluation,” *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1100–1106, 1999.

[40] V. M. Vlahovic and I. M. Vujiosevic, “Long-term foreasting: a critical review of direct-trend extrapolation methods,” *Electr. Power Energy Syst*, pp. 2–8, 1987.

[41] A. M. Melamed, Y. I. Morzhin, V. F. Timchenko, E. V. Tsvetkov, and V. F. Zharkin, “Improvement of electric power system load forecasting and operational planning for control automation,” *Electr. Power Energy Syst*, pp. 218–222, 1989.

[42] Y. Minato and Y. Yokoi, “Development of a forecating method of a region’s electric power demand,” *Trans IEEE Jpn*, pp. 639–647, 1996.

[43] O. a. S. Carpinteiro, R. C. Leme, A. C. Z. de Souza, C. a. M. Pinheiro, and E. M. Moreira, “Long-term load forecasting via a hierarchical neural model with time integrators,” *Electr. Power Syst. Res.*, vol. 77, pp. 371–378, 2007.

[44] W.-J. Lee and J. Hong, "A hybrid dynamic and fuzzy time series model for mid-term power load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 1057–1062, 2015.

[45] S. Iwamoto, "Neural network theories and their applications to power and energy areas," *Trans IEEE Jpn*, pp. 698–705, 1991.

[46] S. Iwamoto and Hayashi, "Long term load forecasting using improved recurrent neural network," *Trans IEEE Jpn*, pp. 1203–1212, 1993.

[47] H. Satoh and Y. Serizawa, "Long-term electric generation system planning using fuzzy linear programming," *Trans IEEE Jpn*, pp. 147–154, 1992.

[48] A. N. Celik, "Effect of different load profiles on the loss-of-load probability of stand-alone photovoltaic systems," *Renew. Energy*, vol. 32, no. 12, pp. 2096–2115, Oct. 2007.

[49] E. M. Nfah, J. M. Ngundam, M. Vandenbergh, and J. Schmid, "Simulation of off-grid generation options for remote villages in Cameroon," *Renew. Energy*, vol. 33, no. 5, pp. 1064–1072, May 2008.

[50] S. M. Shaahid and M. a. Elhadidy, "Technical and economic assessment of grid-independent hybrid photovoltaic-diesel-battery power systems for commercial loads in desert environments," *Renew. Sustain. Energy Rev.*, vol. 11, no. 8, pp. 1794–1810, Oct. 2007.

[51] S. M. Shaahid and M. a. Elhadidy, "Economic analysis of hybrid photovoltaic-diesel-battery power systems for residential loads in hot regions-A step to clean future," *Renew. Sustain. Energy Rev.*, vol. 12, no. 2, pp. 488–503, Feb. 2008.

[52] S. M. Shaahid and I. El-Amin, "Techno-economic evaluation of off-grid hybrid photovoltaic-diesel-battery power systems for rural electrification in Saudi Arabia-A way forward for sustainable development," *Renew. Sustain. Energy Rev.*, vol. 13, no. 3, pp. 625–633, Apr. 2009.

[53] K. Kusakana, "Techno-economic analysis of off-grid hydrokinetic-based hybrid energy systems for onshore/remote area in South Africa," *Energy*, vol. 68, pp. 947–957, Apr. 2014.

[54] K. Kusakana, J. L. Munda, and a. a. Jimoh, "Feasibility study of a hybrid PV-micro hydro system for rural electrification," *IEEE AFRICON Conf.*, no. September, pp. 7–11, 2009.

[55] G. Bekele and B. Palm, "Feasibility study for a standalone solar-wind-based hybrid energy system for application in Ethiopia," *Appl. Energy*, vol. 87, no. 2, pp. 487–495, Feb. 2010.

[56] S. Kamel and C. Dahl, "The economics of hybrid power systems for sustainable desert agriculture in Egypt," *Energy*, vol. 30, no. 8 SPEC. ISS., pp. 1271–1281, Jun. 2005.

[57] a. B. Kanase-Patil, R. P. Saini, and M. P. Sharma, "Integrated renewable energy systems for off grid rural electrification of remote area," *Renew. Energy*, vol. 35, no. 6, pp. 1342–1349, Jun. 2010.

[58] A. Al-Karaghouli and L. L. Kazmerski, "Optimization and life-cycle cost of health clinic PV system for a rural area in southern Iraq using HOMER software," *Sol. Energy*, vol. 84, no. 4, pp. 710–714, Apr. 2010.

[59] G. Bekele and G. Tadesse, "Feasibility study of small Hydro/PV/Wind hybrid system for off-grid rural electrification in Ethiopia," *Appl. Energy*, vol. 97, pp. 5–15, Sep. 2012.

[60] a. Gupta, R. P. Saini, and M. P. Sharma, "Design of an Optimal Hybrid Energy System Model for Remote Rural Area Power Generation," *2007 Int. Conf. Electr. Eng.*, 2007.

[61] H. Camblong, J. Sarr, a. T. Niang, O. Curea, J. a. Alzola, E. H. Sylla, and M. Santos, "Micro-grids project, Part 1: Analysis of rural electrification with high content of renewable energy sources in Senegal," *Renew. Energy*, vol. 34, no. 10, pp. 2141–2150, 2009.

[62] J. a. Alzola, I. Vechiu, H. Camblong, M. Santos, M. Sall, and G. Sow, "Microgrids project, Part 2: Design of an electrification kit with high content of renewable energy sources in Senegal," *Renew. Energy*, vol. 34, no. 10, pp. 2151–2159, 2009.

[63] A. Gupta, R. P. Saini, and M. P. Sharma, "Steady-state modelling of hybrid energy system for off grid electrification of cluster of villages," *Renew. Energy*, vol. 35, no. 2, pp. 520–535, 2010.

[64] T. Dorji, T. Urmee, and P. Jennings, "Options for off-grid electrification in the Kingdom of Bhutan," *Renew. Energy*, vol. 45, pp. 51–58, Sep. 2012.

[65] M. M. Mahmoud and I. H. Ibrik, "Techno-economic feasibility of energy supply of remote villages in Palestine by PV-systems, diesel generators and electric grid," *Renew. Sustain. Energy Rev.*, vol. 10, pp. 128–138, 2006.

[66] D. Saheb-Koussa, M. Haddadi, and M. Belhamel, "Economic and technical study of a hybrid system (wind-photovoltaic-diesel) for rural electrification in Algeria," *Appl. Energy*, vol. 86, no. 7–8, pp. 1024–1030, Jul. 2009.

[67] S. K. Nandi and H. R. Ghosh, "Prospect of wind-PV-battery hybrid power system as an alternative to grid extension in Bangladesh," *Energy*, vol. 35, no. 7, pp. 3040–3047, Jul. 2010.

[68] K. Y. Lau, M. F. M. Yousof, S. N. M. Arshad, M. Anwari, and a. H. M. Yatim, "Performance analysis of hybrid photovoltaic/diesel energy system under Malaysian conditions," *Energy*, vol. 35, no. 8, pp. 3245–3255, Aug. 2010.

[69] C. Brivio, "Off-Main-Grid PV Systems : Appropriate Sizing Methodologies in Developing Countries," Politecnico di Milano, 2014.

[70] W. X. Shen, "Optimally sizing of solar array and battery in a standalone photovoltaic system in Malaysia," *Renew. Energy*, vol. 34, no. 1, pp. 348–352, Jan. 2009.