Traineeship Report

Distributed Power Management Smart Grid System:

Modelling of the energy consumption of a Heating, Ventilation and Air-Conditioning (HVAC) system

IES-UPM

Traineeship realized in the Institute of Solar Energy (IES) in Madrid

Student: DEVOS Pierre

Industrial Tutor: GUTIERREZ Álvaro

Academic Tutor: DELACOURT Eric

Length: 5 Months

From 1 September 2016 to 20 January 2017

Acknowledgments

This work placement was performed in the laboratory Robolabo and in the institute of solar energy (IES) of the University polytechnic of Madrid (UPM).

First and foremost, I would like to express my gratitude to my industrial tutor Pr. **Álvaro GUTIERREZ** for giving me the opportunity to have this internship in the institute of solar energy and for his advices about the objectives of my project.

Then, I would also like to express my gratitude to his colleagues, Pr. **Estefaníaa CAAMAÑO MARTI**N and Pr. **Miguel Angel EGIDO AGUILERA** for the support they gave to me with classes to help me to begin my work.

My gratitude goes to my academic tutor Pr. **Eric DELACOURT** for following me during this internship and visiting me with Pr**. Nachida BOURABAA** and my thanks go to her too.

I also extend my thanks to the other members of Robolabo: María **PERAL BOIZA,** Teresa **GÓMEZ FERNÁNDEZ, Juana GONZÁLES,** Dr. Manuel **CASTILLO CAGIGAL but particularly Dr. Eduardo MATALLANAS DE ÁVILA** for all their availability and their scientific and technical advices.

After, I would thank Pr. Jane **LAURO** for helping me in the research and obtaining this internship.

Finally, I would like to thank **Hugo REY** and **Simon LAVAUD** my ENSIAME colleagues in the institute for their support in Madrid.

Introduction

For a very long time and all around the world, the use of the energy establishes an essential need for human activities, to the point where the demand increasing continuously. It is why nowadays, the development of new means of electricity production is important.

In the future, the purpose is to substitute the fossil energy with the renewable energy. Today, the new and renewable energies play a major role in the protection of the environment and the planet. The production of these energies is make in numerous sites around the world. There are different types of renewable energies: the wind energy, the biomass energy, the geothermal energy, the hydropower and solar energy.

For my internship of the second year in the ENSIAME, I worked during five months in the institute of solar energy of Madrid. They work on solar panel in order to develop them and integrate them in the electrical grid. In fact, the consumers begin generators of power with solar panels. It is why, they work especially on the Demand Side management System to control and secure the electrical

During my traineeship, I worked in the laboratory Robolabo of the institute where a lot of professors and students work. I had at my disposal a computer with Linux, the simulator GridSim that simulate the consumption on an electrical grid and above all Matlab in order to create programs and model the HVAC system.

The purpose of my project is to improve the simulator by integrating in the loads possible a model of an HVAC system. In fact, nowadays many houses are equipped with it and it is a significant load in the electricity bill and the consumption of the network. It is why integrate this system in the simulator is important to have a better tool to study the DSM.

Therefore, in order to explain the project, this report is divide in seven chapters. The first and the second chapter is a presentation of the internship with especially explanations about the DSM, smart grid and GridSim simulator. Then, in chapter three, we will start to model the HVAC system with the multiple linear regression. The chapter four will be about the multiple polynomial regression. The chapter five will study the quadratic regression. The chapter six will be about the neural network design. Finally, in the chapter seven we will compare them in order to choose the best with the aim of integrating it into the simulator.

Chapter 1: Presentation of the internship

1.1 Internship in a foreign country: Spain

My internship took place in Madrid the capital of Spain.

Spain, officially the Kingdom of Spain is located on the Iberian Peninsula in southwestern Europe, with two archipelagos, the Balearic Islands in the Mediterranean Sea and the Canary Islands near the North African Atlantic coast. The country is bordered to the south and east by the Mediterranean Sea, to the north and northeast by France and Andorra, and to the west and northwest by Portugal and the Atlantic Ocean.

Figure 1.1: Map of Spain.

Spain is a democracy organised in the form of a parliamentary government under a constitutional monarchy. It is a middle power and a developed country with the world's fourteenth largest economy by nominal GDP and sixteenth largest by purchasing power parity. It's a member of many international organisations including European Union.

The king of Spain is Felipe VI and the head of the government is the Prime Minister with currently Mariano Rajoy.

Spain has an area of 505,990 km²; it is the second largest country in Western Europe and the European Union. Spain's capital and largest city is Madrid but there are other major urban areas as Barcelona, Valencia, Seville, Bilbao and Málaga.

1.2 Presentation of the Institute of Solar Energy (IES)

Pr. Antonio Luque is the founder and President of IES. The Institute of Solar Energy of the Polytechnic University of Madrid (IES-UPM) is a pioneer centre entirely devoted to photovoltaic conversion of solar energy since the end of 70's. The IES-UPM is nowadays a recognized centre because with a modest size, it contributed in silicon solar cells technology, with new concepts for solar cells, quantum calculation.

Figure 1.2: Institute of solar energy's building.

An example of success is the creation in 1981 of the company ISOFOTON based on the bifacial cell invented at IES-UPM. This company became during the first years of the 21st century the largest manufacturer of solar cells in Europe.

The IES-UPM was set up in part to develop solar power in Spain and to place the country at the forefront of international markets in this field. However, nowadays, with the rise of some countries in solar energy such as Japan, Germany, United States and especially China, it is essential that Spain continues its researches to remain a major actor in the design and study of solar panels. The currently director of IES-UPM is Pr. Carlos del Cañizo and for him *"Given today´s climate challenges, it is hard to dispute the need to firmly back renewable energies as the basis of a new energetic model, including photovoltaic solar energy. PV is already beginning to play an important role into the world's electricity mix, and has enormous growth potential in the coming years"*.

Figure 1.3: Solar Panel.

1.3 Research Groups

The institute comprises five R&D groups presented below.

1.3.1 Instrument and Systems Integration (ISI)

The responsible of this group is Pr. Gabriel Sala. The objective is to promote low-cost solar electricity. The lines of research are:

- New concepts in CPV (Concentrator Photovoltaics) systems and fabrication methods for concentrator optics.
- Optical and electrical instrumentation.
- Characterization methods for CPV modules and receivers.
- Reliability of CPV systems and components.

1.3.2 Photovoltaic Systems (SPV)

The responsible of this group is Dr. Luis Narvante. The objective is to assure the expected service, quality, reliability and long-term durability of PV systems in the field. The research lines are:

- The quality of large PV plants: from the energy yield simulation with low uncertainty to the quality control procedures for bankability processes.
- New market niches for PV systems: developing solutions with high potential market as PV irrigation of PV cooling.
- PV rural electrification: assuring the reliability of Solar Home Systems or PV mini grids in rural areas of developing countries.

1.3.3 Silicon and New Solar Cell Concepts (SNSCC)

The responsible of this group is Pr. Antonio Martí. The objective is to improve the development of photovoltaic energy. It consists in the reduction of costs and the increment of efficiency by different ways. The research lines are:

- Intermediate band materials and solar cells.
- Quantum mechanical calculations Nanostructures.
- Molecular beam epitaxy.

- Advanced characterization tools and device processing.
- Light management.
- Modelling and other novel solar cell concepts.
- Three terminal heterojunction bipolar transistor solar cells.
- Energy Storage: Novel system based on molten silicon latent heat.
- Solar silicon refinement.
- Defect engineering in solar silicon.
- Crystalline silicon solar cell development.

1.3.4 III-V Semiconductors

The responsible of this group is Pr. Carlos Algora. The objective of the research is the highest efficiency solar cells, namely III-V multi-junction solar cells, for both terrestrial and space application. The research lines are:

- Concentrator solar cells.
- Space solar cells.
- Flat plate photovoltaics beyond silicon.

1.3.5 Renewable Distributed Generation and Intelligent Control (GEDIRCI)

The main researcher is Pr. Álvaro Gutierrez. Nowadays, photovoltaic become profitable and can be use by a lot of residential consumers. So, recent technological and market changes have helped in an increasing penetration of photovoltaic electricity in the electrical network. It’s important to have an intelligent use of photovoltaic electricity by users and grid operators. The research lines are:

- Distributed strategies of maximum power point tracking in PV (photovoltaic) modules.
- Active demand side management with distributed PV and storage.
- PV generation and forecast in complex environments
- Solar PV potential at urban scale: from buildings to cities. Building integrated photovoltaics
- PV heating and cooling.

- Quality assurance in decentralized PV applications.
- Basic and income-generation electrical services for developing countries.
- PV Hybrid off-grid systems.

I worked in this research group, more precisely in the laboratory ROBOLABO that study on the Demand Side management and a Simulator called GridSim, which help to represent an electrical grid.

1.4 Presentation and Objectives of the Project

The GEDIRCI group in order to do researches about smart grid and the demand side management developed a simulator called GridSim. The purpose of this program is to simulate the consumption of the electrical grid where you can change different parameters. However, to do this they used the MagicBox, it is a self-contained house working with solar energy and containing various components such as dishwasher, washing machine, refrigerator, oven, kitchen hood and dryer. They used it in order to have a typical approximation of a consumption of a house. All machines quoted previously are joined into the simulator, however it's not the case for the HVAC system. So, the purpose of this project is to model it by a mathematical model.

Chapter 2: Project Background

2.1 The Electrical Line

An electric power system is a network of electrical components used to supply, transmit and use electric power. We call electrical grid (Figure 2.1) this system which gives electricity to consumers.

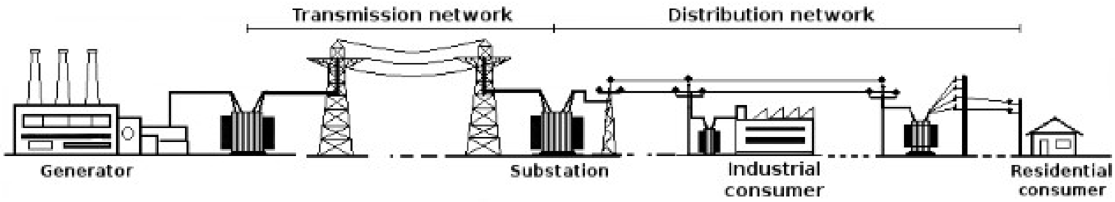


Figure 2.1: Representation of the grid.

The electrical grid can be divided into four parts:

- **Generation**: generators supplying the grid. They can be of different types like diesel generator, photovoltaic solar plants, wind turbines, nuclear power plants…
- **Transmission network**: it transfers a large amounts of electricity from generators to the substations located close to the consumers which transforms voltage from high to low, or the reverse. Every transmission line is high-voltage, three phases and alternating current (AC). The purpose of high voltage is to reduce the energy losses in long transmission. The voltage, frequency and phases depend on the regulation of each entity in charge of managing specific grid.
- **Distribution network**: the distribution network delivers the electricity from the substations to the consumer. The distribution lines are medium or low voltage and they have one or three phases. For example, in Europe the electricity lines give a single phase of 230 V.
- **Consumption**: the consumption of an electrical grid can be a complex part of an electrical grid. In fact, any device that transforms electric power into work is considered as consumption of the system. We can have different groups of consumers like industrial consumers, residential consumers...

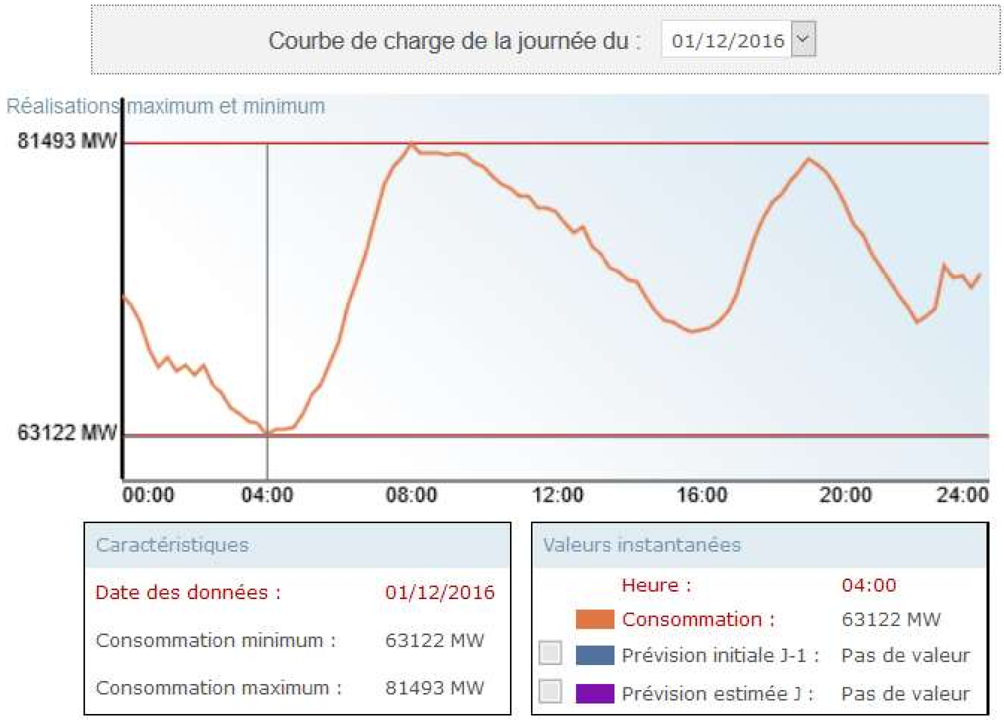


Figure 2.2: Example of the consumption of France electrical grid during a day (01/12/2016). Source: RTE France

There is only one rule for the electrical grid: the amount of active power consumed more losses must equal the active power produced. In fact, if the generated power exceeds the consumed power and losses, the voltage and frequency of the system increase and vice versa. However, these variations may cause several damages to the grid from the generation to the consumption side. Therefore, the grid generation and consumption must be always balanced in real-time. In order to have this, a synchronization between millions of devices is necessary.

This synchronisation can be done thanks simulation. Moreover, the balance between generation and consumption has been done from the generation side. The consumers consume energy regardless of the grid status and the grid must respond to consumer’s requirements. In order to adapt the power in the grid with the demand, we use the generators. When consumed power increases, the generators produce more electricity and conversely. The consumption is defined as the sum of all consumers, so it is the consumption of the grid. Figure 2.2 shows the consumption during a weekly day in France. The low consumption periods are called valleys and the high consumption periods are peeks. In fact, the consumption is not constant, for example in Figure 1.4, we have a variation between 81493 MW and 63122 MW, thus an amplitude of 18371 MW. The generators must be able to supply peaks and more important the historic peak maximal (102 100 MW in France 08/02/2012).

Données de 2015

Type de données	Capacité installée de production (MW)
Charbon	4810
Éolien en mer	10
Éolien terrestre	357
Fioul	6670
Gaz	6121
Hydraulique fil de l'eau / éclusée	10314,1
Hydraulique lacs	8213,66
Hydraulique STEP	4965
Marin	240
Nucléaire	63130
Solaire	39
Autre	62
Total	**104931,77**

Figure 2.3: Production capacity of electricity in France. Source: RTE France

Therefore, there is also a problem when the transmission and distribution line is underused. This lines duplicated to provide robustness to the grid. It is able to work even with a loss of a single line. The lines can support twice the historical maximum peak. This oversizing of the grid involves a high cost. Figure 2.3 gives the maximal production in France 104931,77 MW and the distribution for sources. If we take the valley of Figure 1.4, 63122 MW, 60 % of production facilities are not used. There is no optimization of the park of production.

There are also regional disparities (Figure 2.4 is an example for France) in the electrical grids. This was due to the generators which are at long distances from consumers. Where there is high consumption but low generation we have problems of saturation in the lines with transmission and distribution. It increases the oversizing of the electrical infrastructure.

Nowadays, there is a growing electricity demand and with it the increasing cost of raw materials and the appearance of new generation technologies. So a purpose today is to find a solution to the oversizing problem. The root of this problem is the consumption variability. That's where the Demand-Side Management comes in.

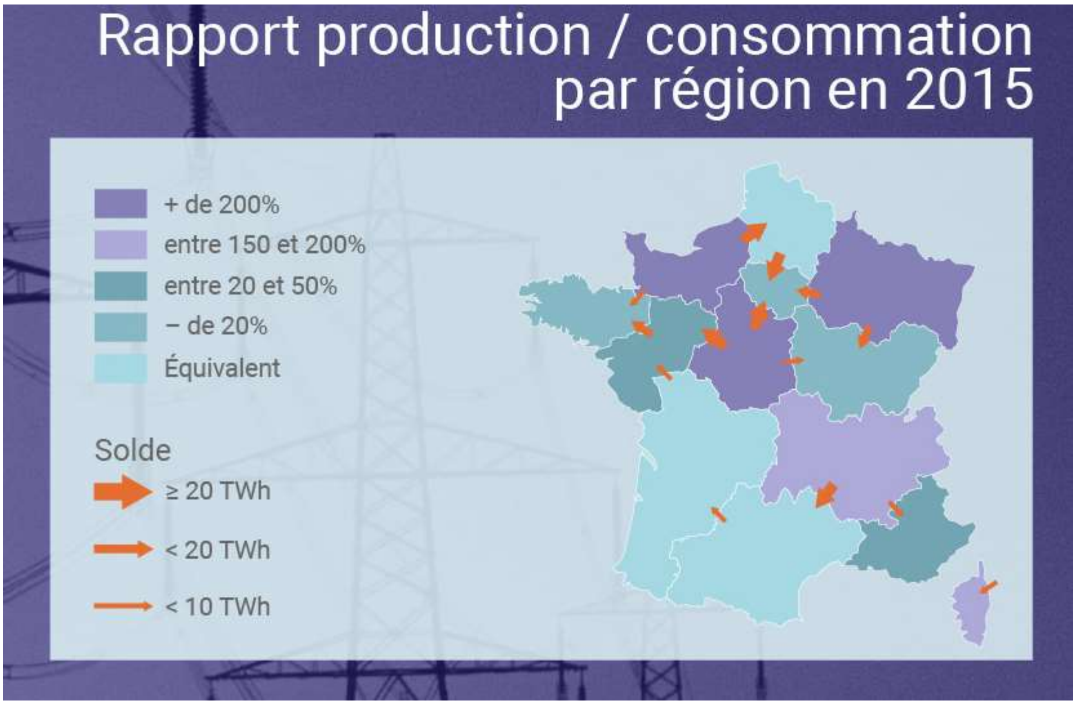


Figure 2.4: Report between production and consumption for each region in France with transmission of energy between them. Source: RTE France

2.2 The Demand-Side Management

The Demand-Side Management (DSM) is defined as actions that influence the way that consumers use electricity in order to achieve savings or higher efficiency in energy use. The advantages are:

- **Reduce the oversizing of lines**: In fact, if we reduce the consumption during peaks, the size of generation, transmission and distribution lines can be reducing. Thus, costs of installation and maintenance decreasing.
- **Improve the efficiency of the grid**: Increasing consumption in valley to have a greater use of the electricity infrastructure.
- **Improve security**: The DSM techniques help to increase monitoring and control capability of the electrical grid.
- **Integration of new generation technologies**: New generation technologies are mainly based on renewable energies. However, the generation can't be constant with solar or wind for example. So, it's more difficult to adapt generation with consumption. DSM techniques can reduce the effects of its intermittent nature.
- **Integration of new local infrastructure**: classical electrical grids have a hierarchical structure where energy is produced in power plants for consumers. However, today

Distributed Energy Resources (DERs) appeared. It is places with local generators and storage system near consumption centers. It's a new conception of the electrical grid where consumers are also generators. DSM with DERs can lead this new grid structure called Smart Grid.

However, the purpose of the DSM is to smooth the shape of the aggregated consumption. We can divide it in four parts:

- **Consumption reduction (Figure 2.5.a)**: this goal is to reduce the consumption without modify the shape of the aggregated consumption. This can be done with increasing the efficiency of appliances or the society energy awareness. It can help to reduce the oversizing of the electrical grid.
- **Increase valley consumption (Figure 2.5.b)**: we can do this with the connection of electric cars, storage systems, etc during the low consumption period. The purpose is to increase the profability and efficiency of the grid.
- **Decrease peak consumption (Figure 2.5.c)**: DSM is useful to reduce the oversizing of the electrical grid. It is possible for exemple with the load automation or blocking some consumers.

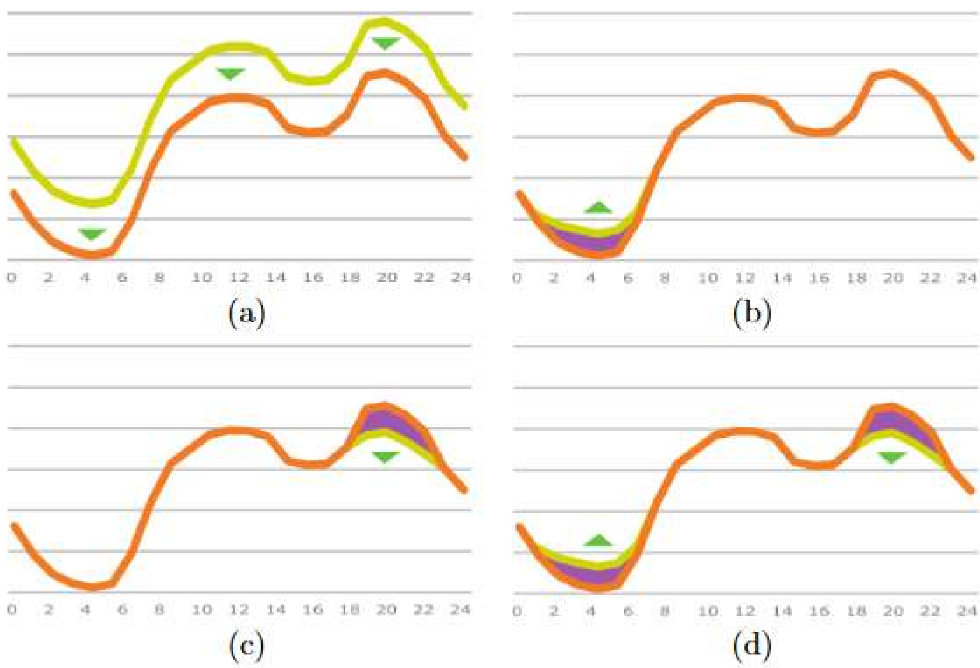


Figure 2.5: Effects of the DSM techniques in the aggregated consumption: a.consumption reduction, b.increase valley consumption, c.decrease peak consumption and d.load shifting Source: Red Eléctrica Española.

- **Load shifting (Figure 2.5.d)**: this consists of move part of the electric loads from the peak to the valley periods. It may be done thanks to a variable pricing depending on the aggregated consumption shape or the electric load automation.

The DSM can act by different mean on consumers and the electric system. In fact, there are for example encouraging consumer awareness, reduction of appliances consumption. DSM can be classified in classes in function of the degree of interaction between the consumer and the electric system:

- **Saving and efficiency programs**: These are initiatives that search to increase the efficiency of consumption. These provokes an indirect reduction on the long-term demand in terms of power consumption. For example, there are incentive campaigns for the use of energy-saving lamps, variable speed drives in electric machines.
- **Indirect control of electric loads through pricing**: these DSM programs are based on time variation of prices, in general, depending on the costs to adapt the consumption with the demand. The goal is to discourage consumption in peak hours. There are different types of regulation:
  - *Time of Use tariff (TOU)*: the time during a day is divided in blocks with different prices. Each block represents the historical average costs of producing and transporting the electric energy during the period covered by the block.
  - *Critical Peak Pricing (CPP)*: prices are increased during hours in which production costs are high, it results from difficulties to adapt the demand with supply. The purpose is to reduce consumption during peaks.
  - *Real Time Pricing (RTP)*: this is the price gives by the markets, with an average basis. So the pricing information is in real-time, it’s a way to adapt the demand to the grid status improving the use of the energy sources.
  - *Peak Time Rebates (PTR)*: there are electricity bill rebates when we don’t use power during the peak hours.
- **Direct control of electrical loads**: the operator can disconnect in this case some electrical appliances of the consumers. Nowadays, it’s not really apply for residential consumers but more in the industry.
- **Smart metering and appliances control**: these techniques comprise market programs or structures that allow consumers to participate offering electrical load reductions to adapt the demand to the grid. Greater price differentials between high-cost and low-cost periods could give greater shifts of energy usage. These techniques require to be

accompanied by the application of intelligent appliances that facilitate the implementation of DSM.

Nowadays, some of these techniques have already been developed. However, some barriers are important for the DSM. In the following, a classification of these barriers is presented:

- **Lack of infrastructure**: electrical grids are today only “energy broadcast system” without information exchange. The consumers receive only a bill with the energy consumed each month. In order to improve the DSM system, the consumers have to receive information about the cost and also the grid status. So we must develop advanced metering, control methods and Information and Communications Technologies (ICTs) for the electrical grid. The lack of these technologies in the system and international standards reduce the development of the DSM.
- **Investment recovery**: the financial aspect is the main problem to develop the DSM. The consumers should make technology investments in order to exploit the benefits of DSM: appliance automation and communication with the grid. The recovery of this investment is slow because prices of the electricity is low. For the electricity companies it’s different because there is a lack of knowledge about the costs and benefits so they don’t want to risk investing in DSM. They will not invest in a program which will give benefits only to the industry.
- **Program structure and policy**: DSM program include a lot of of participants, from big producer to small consumers. So the development of a program which includes the entire grid is difficult. This causes a solwdown in the DSM implementation. However, nowadays the situation changes with the commitment of governments and new energy policies to improve the efficiency of the grid and fight climate changes.

Smart Grids is a convergence of ICTs. It began at the end of 90’s with the use of electronic control, metering and monitoring in the electricity power system. In Vu et al. (1997) the authors proposed that the Smart Grids should be an automated system with monitoring, control, and protection devices that improves the reliability of the transmission grid by preventing wide-spread break-ups. Currently, Smart grids are an assembly of techniques to respond to the challenges of the electricity power systems. Characteristics of Smart grids are:

- **Reliability**: in an electric system, reliability is really important. In fact, a lot of companies and people are dependent on electricity so if there is a problem, economic and social effects can appear. Moreover, the increasing complexity of the grids increases the challenge of having a system reliable for all new consumers. The purpose of Smart Grid is to have a better situational awareness and operator assistance, increase resiliency against components failures and natural disasters, higher electricity quality.

- **Efficiency**: the smart grid system includes DSM and also all its benefits mentioned above. DSM can be integrated into a large concept that takes into account all the elements of the Smart Grid, which consists of an interactive electric network, for generators and consumers.
- **Flexible topology**: The DERs modify the grid topologies. They are transforming to a distributed collection of generators, consumers and storage systems spread over a region. Thus, Smart Grids are not only present in the local framework but also in transmission and distribution networks. It will be help to use resources like renewable energies.
- **Additional services**: the introduction of ICTs in the grid gives a better monitoring of the electrical power system. Smart grids allow real-time electricity pricing. This monitoring helps to prevent damages causes by grid failures.

2.3 Smart Grids

Smart grids can be bidirectional, information and energy can go to consumers from generators and vice-versa. But, it is necessary to have a communication structure with all consumers to adopt a good strategy of demand management. The operators have to give information of the grid. So, smart meters can be used. It is an advanced electricity meter that provides additional information to consumers and electrical grid operators. This information can be from different nature, from local electricity system as voltage, frequency, active and reactive power, to general grid status as instantaneous aggregated consumption, pricing, emergency signals. Although there is not a universal definition of the characteristics and benefits of smart metering, but we can list:

- **Consumption information**: one characteristic of the smart meters is digital power measurement. It has advantages for consumers but also for grid operators. From the consumer point of view, the digitizing of these data allows to know its local consumption in real-time. Studies concluded by offering consumption information to the users will help to reduce the amount of energy demand. From the grid operator point of view, the information of local consumption allows to foresee more accurately the aggregated consumption. In addition, the operators can create bills for consumers telematically, it is an economic benefit for them.

- **Pricing information**: consumers may adapt their consumption to pricing variation if they receive this information. It is an important economic advantage for residential consumers. It can incite them to use smart meters
- **Emergency signals**: grid operators can send emergency signals to the consumers. For example, to warn them of a grid failure.
- **Remote load switching**: smart meters may receive order from the grid operators. It's given to them the possibility to control remotely appliances, such as high power furnaces or lighting network.
- **Grid information network**: smart meters can be used to create a parallel information network to the electrical grid. So, thanks them grid operators and consumers can exchange information between them. This gives us the opportunity of managing the grid and the implementation of advanced control algorithms.

It's why the number of smart meters would grow during the next years in Europe. For example, it's expected in Spain that in 2018 one hundred percent of the electricity meters will be smart meters. The name of these meters is PRIME (LINKY in France). However, the communication network is not the only challenge of the smart grids. The information transmitted through this network should be managed because problems of large communications network such as congestion, robustness, computational power can appear. In order to reduce this, the grid can be reduced un multiple subgrids which self-organize to achieve a common energy goal. From the DSM point of view, it means that electricity consumption is managed locally without a central organizer. The development of a self-organized algorithm is necessary to control all the subgrids.

2.3 The MagicBox

For the study of solar energy at the institute, there is an important experimental tool since 2005 that is used: MagicBox. It is a self-contained house working with solar energy and containing various components such as: Dishwasher, washing machine, refrigerator, oven, kitchen hood, dryer. This building was originally created for the contest *Solar Decathlon 2005* and expose a profitable and ecological solar house with an attractive design: http://www.magicbox.etsit.upm.es.

Figure.2.6: Photo of MagicBox.

It combines bioclimatic design principles, photovoltaic solar energy integration, local energy storage technology and the use of ICTs to monitor and control the house power balance. MagicBox is currently use as a research laboratory to assess the effect of the combination of photovoltaic generation, load management and local storage on the electric grids. What will interest us is the heating, ventilation and air conditioning (HVAC) system.

2.4 GridSim Simulator

The simulator is programmed with C++ in order to help researchers in their work to develop the demand side management and smart grid. In fact, it simulates an electric grid with different lines consisting of nodes (consumers) whose consumption model is MagicBox. The characteristic of each lines can be changed (capacity of loads, storages and photovoltaic system of nodes). However, on a line, all nodes have the same characteristics.

In order to understand the simulator I had to modify the program in appendix 1 to get:

- The power generated by the solar panel of a node (m_vPVSignal).
- The power consumed by one node (m_vNodeSignal).
- The global power of the grid (m_vLineSignal).
- The power extracted on the grid by one node (m_vGRIDSignal).

In order to do this I added in the principal code eight lines:

```
m_vLineSignal.push_back( m_pcGrid->getPower_lines() );
m_vNodeSignal.push_back( m_pcGrid->getLines()->front()->getNodes()->front()->getNPower()->load );
m_vPVSignal.push_back(m_pcGrid->getLines()->front()->getNodes()->front()->getNPower()->pv);
m_vGRIDSignal.push_back(m_pcGrid->getLines()->front()->getNodes()->front()->getNPower()->grid);
```

Algorithm 2.1: Modifications in the code to obtain data we need, power of loads, power of a node, power of a PV system and power of the grid.

Also to plot all this characteristics:

```
m_sSimCnf->pcPlotter->setData    ( 1 , &m_vLineSignal              );
m_sSimCnf->pcPlotter->setData    ( 2 , &m_vNodeSignal              );
m_sSimCnf->pcPlotter->setData    ( 3 , &m_vPVSignal                );
m_sSimCnf->pcPlotter->setData    ( 4 , &m_vGRIDSignal              );
```

Algorithm 2.2: Modifications to plot graphs in Figure 2.7.

The completely modified program is in appendix 2 (orange rectangles are the modifications) and it gives in function of the time four graph represented in Figure 2.7:

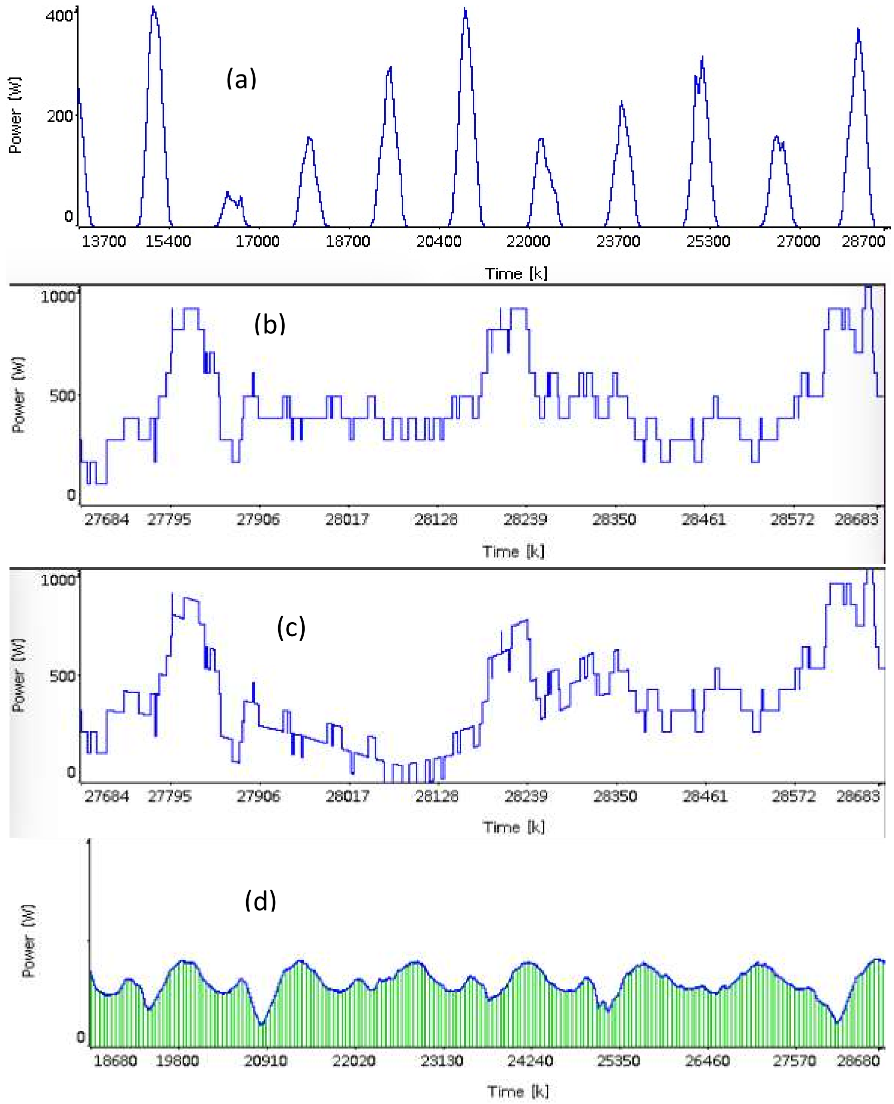


Figure 2.7: (a) Generation of solar panels in one node, (b) Power consumed by all loads in one node, (c) Power extracted by one node in the grid, (d) Global power consumed in the electrical grid.

So we can observe that a peak of photovoltaic generation (Figure 2.7.a) correspond to a valley in the global grid consumption (Figure 2.7.d). Moreover, a node consumed less on the grid (Figure 2.7.d) than the real consumption of loads (Figure 2.7.c) in the node because there is the photovoltaic generation.

Chapter 3: Multiple Linear Regression

3.1 MagicBox's Data and first representation

To achieve the project, my first idea was to search a theoretical model but Dr. Matallanas de Ávila explained to me that it would be too complex because the consumption of the HVAC system depends on many parameters. Therefore, we decided to create a mathematical model with data from the HVAC system and parameters of the MagicBox (Internal and external temperature, irradiation on the house and the power consumed by the HVAC system every minutes).

I received data from another research group of three different periods that I will use during the entire project:

Period	Name
From 12/02/2016 to 21/02/2016 From 03/03/2016 to 11/03/2016	Winter Period
From 10/06/2016 to 20/07/2016	Summer Period
From 20/10/2016 to 06/11/2016	Autumn Period

Table 3.1: Periods when data were taken.

The purpose of modelling the HVAC system is to integrate it in the Gridsim simulator, an open source software which simulates in real time the power consumed in an electrical grid. So, my first idea is to model directly the power in function of the time. In order to do this, I created a Matlab program to draw the power in function of the time (Algorithm 3.1).

```
clc;
clear;
format compact;

V1=input('For what period do you want the regression ? (Summer, Winter or Autumn)')
T1=strcat(V1,'_Data');
load(T1);

p=polyfit(Time,Phvac,20);

Y1=polyval(p,Time);

plot(Time,Phvac,Time,Y1);
set(gca,'ylim',[0 2100],'ytick',0:100:2100)
xlabel('Time (Days)')
ylabel('Power HVAC (Watt)')
```

Algorithm 3.1: Program Matlab to plot and study the power consumed by the HVAC system in function of the time.

Here, with this program, I created graphs for summer, winter and autumn periods (appendix 3). With the “input ( )“ line 5, the code asks to the user the name of the period and thanks to the function “load” line 7 it takes the data in one of three files:

- ✓ Summer_Data.dat (53 583 points).
- ✓ Winter_Data.dat (27 360 points).
- ✓ Autumn_Data.dat (24 480 points).

In it, there are five-column matrix with:

Name	Value
G	24480x1 double
Phvac	24480x1 double
Text	24480x1 double
Time	24480x1 double
Tint	24480x1 double

Figure 3.1: Content of Autumn_Data.dat.

- ✓ G, the irradiance of the sun in Watt/m².
- ✓ Phvac, the power consumed by the HVAC system in Watt.
- ✓ Text, the temperature outside of the MagicBox in °C.
- ✓ Tint, the temperature inside of the MagicBox in °C.
- ✓ Time, time in order to have a scale in days.

Each point is associated to this data for example the first point is defined by (G (1), Phvac (1), Text(1), Tint(1), Time (1)). Moreover, I added a polynomial regression (twenty degrees, line 9 in the code) in order to compare with the experimental graph Phvac(Time).

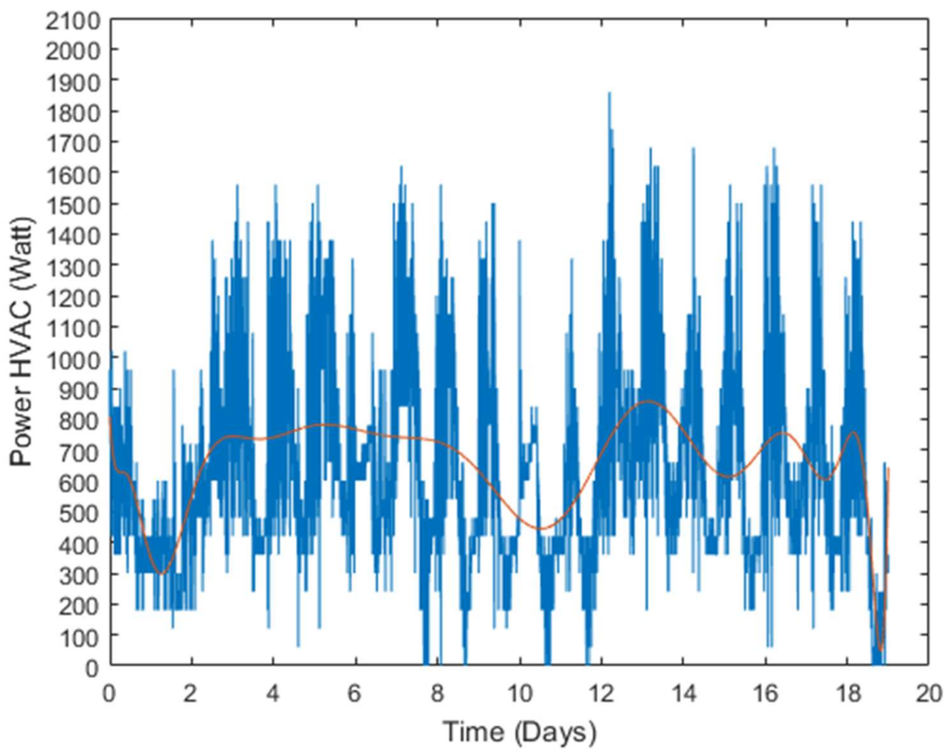


Figure 3.2: Power consumed by the HVAC system in function of time for the winter period. The experimental curve is blue and the polynomial regression in orange.

We can see in figure 3.2 for the winter period and in appendix 3 for summer and autumn that the polynomial model has large differences with the curve and there is a lot of noise in the experimental curve. However, using these variables to create a model is not very relevant because the operation of the HVAC system depends on many variables of different natures. Indeed, we can use for example the indoor and outdoor temperature at MagicBox.

3.2 Multiple Linear Regression

3.2.1 Theoretical aspect

The model is $y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \cdots + b_p x_{i,p} + \varepsilon_i\,; i = 1, \ldots, n$ with:

- y_i the variable to be explained
- $x_{i,p}$ the explanatory variable p
- $b_0 b_1 b_2 \ldots b_p$ model parameters
- ε_i the specification errors
- n is the number of variable to explained

If we want to integrate the interaction between the explanatory variable we can add synthetic variables:

For example, if k=2: $y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + b_3 (x_{i,1} * x_{i,2})$.

In order to have a better representation we use the matrix representation:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & \cdots & \cdots & x_{1,p} \\ 1 & \vdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,p} \\ 1 & \vdots & \cdots & \cdots & \cdots & \vdots \\ 1 & x_{n,1} & \cdots & \cdots & \cdots & x_{n,p} \end{pmatrix} * \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Column for the constant

$$\boxed{Y = X * b + \varepsilon}$$

Here, the multiple linear regression integrates the least squares method (LSM, figure 3.3). This method looks for the best estimation of the parameters "b" by minimizing the quantity (minimum is noted $\widehat{b}$). $S = \sum_i e_i^2$ with $e_i = Y - X\,\widehat{b}$. E the observed error (the residue) is an evaluation of the term of the error ε. Figure 3.5 is a graphical representation of the least squares method.

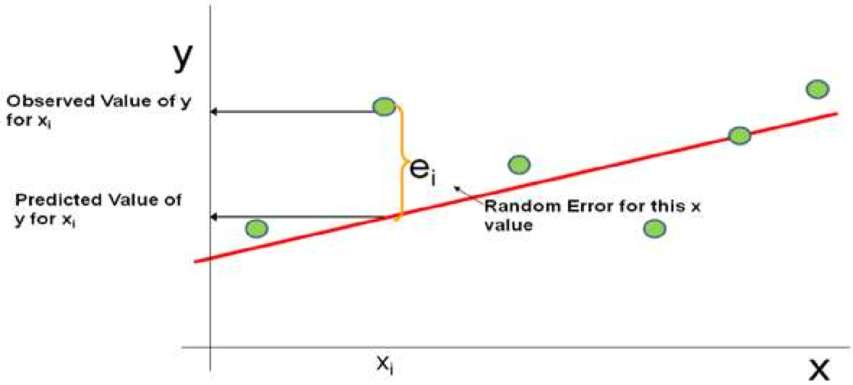


Figure 3.3: Graphical representation of the least squares method.

The Hypothesis of LSM are:

- X are observed without error (not random).
- E (ε)=0, on average the model is clearly specified.
- E (ε^2)=σ_ε^2, the variance of the error is constant .
- E (εi, εj)=0, the errors are non-correlated.
- Cov (ε,x)=0, the error is independent from explanatory variables.
- ε≡Normal(0, σ_ε).
- Rank (X'X)=p+1 or Det (X'X)≠0.
- (X'X)/n aims towards a not singular finished matrix when $n \to +\infty$.
- n > p+1, the number of observations is upper in number of parameters of the model (explicative parameters plus constant).

We must find the parameters "b" which minimize S:

$$S = \sum_i \varepsilon_i^2 = \sum_i [y_i - (b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \cdots + b_p x_{i,p})]^2$$

We solve $\frac{\partial S}{\partial a} = 0$ where there are p+1 equations without synthetic variables.

$$\text{So, } S = (Y - Xb)'(Y - Xb) = Y'Y - 2b'X'Y + b'X'X$$

$$\frac{\partial S}{\partial a} = -2(X'Y) + 2(X'X)b = 0$$

$$\boxed{\text{In the end, } \hat{b} = (X'X)^{-1}X'Y}$$

3.2.2 Modelling program

To have a model for the consumption of the HVAC I used firstly the Multiple Linear Regression with MATLAB. Therefore, I created a program (appendix 4) to link power (P, in Watt) to the others variables that we have available: internal temperature (T_{int}, in Celsius), external temperature (T_{ext}, in Celsius) and irradiance (G, in Watt/m^2). Moreover, I calculated the average error (in percentage) and coefficient of determination for each model to realize comparisons. In order to create the model, I separated the data for each seasons: fifty percent to find parameters "b" and the other fifty percent will help to check the model with the average error.

```
V1=input('Do you want a 3D or 4D regression ? (3 for 3D ; 4 for 4D)')
V2=input('For what period do you want the regression ? (Summer ; Winter ; Autumn or
S&W&A for all)')
```

```
if V1==3
```

The program asks users if they want a 3D or 4D model and the period to use

```
    T1=strcat(V2,'_Data');
    load(T1);
    X1m=Text_model;
    A1='External Temperature (°C)';
    X2m=Tint_model;
    A2='Internal Temperature (°C)';
    Ym=Phvac_model;
    A3='Power Consumed (W)';
```

Open the file with Data and choose parameters "X_model" signify that it contains the first 50% of variables

```
    X=[ones(size(X1m)) X1m X2m X1m.*X2m];
    b=inv(X'*X)*(X'*Ym)
```

Find model's parameters thanks to the equation found in the theoretical part

```
    X1=Text;
    X2=Tint;
    Y=Phvac;

    S=[];
    for i=1:length(Y)
        if Y(i)~=0
            S=[S abs(Y(i)-(b(1)+b(2)*X1(i)+b(3)*X2(i)+b(4)*X1(i)*X2(i)))/Y(i)*100];
        end
    end

    S=transpose(S);
    Mean_Error=mean(S)

    scatter3(X1,X2,Y,'filled')
    hold on
```

Calculate the average error with 100% of Data

```
    x1fit = min(X1):0.5:max(X1);
    x2fit = min(X2):0.5:max(X2);
    [X1FIT,X2FIT] = meshgrid(x1fit,x2fit);
    YFIT = b(1) + b(2)*X1FIT + b(3)*X2FIT + b(4)*X1FIT.*X2FIT;
    mesh(X1FIT,X2FIT,YFIT);
    xlabel(A1)
    ylabel(A2)
    zlabel(A3)
    view(60,30)
```

Plot a graphical representation of the model with experimental points

Algorithm 3.2: First part of the Matlab program to study the multiple linear regression with three parameters here Phvac (Text, Tint).

Here, I can test different associations of variables to find a function of the HVAC's consumption:

- Phvac in function of Text and Tint.
- Phvac in function of Text and G.
- Phvac in function of Tint and G.
- Phvac in function of Text, Tint and G.

For the last function, the end of the program annex 2 is used, it is the same principle as previously but adapted for three explanatory variables:

```
elseif V1==4

    T1=strcat(V2,'_Data');
    load(T1);

    X1m=Text_model;
    X2m=Tint_model;
    X3m=G_model;
    Ym=Phvac_model;

    X=[ones(size(X1m)) X1m X2m X3m X1m.*X2m X1m.*X3m X2m.*X3m X1m.*X2m.*X3m];
    b=inv(X'*X)*(X'*Ym)

    X1=Text;
    X2=Tint;
    X3=G;
    Y=Phvac;

    S=[];
    for i=1:length(Y)
      if Y(i)~=0
          S=[S abs(Y(i)-(b(1)+b(2)*X1(i)+b(3)*X2(i)+b(4)*X3(i)+b(5)*X1(i)*X2(i)+b(6)*X1(i)*X3(i)+b(7)*X2(i)*X3(i)+b(8)*X1(i)*X2(i)*X3(i)))];
      end
    end

    Mean_Error=mean(S)
end
```

Algorithm 3.3: Final part of the Matlab program to study the multiple linear regression with four parameters.

3.2.3 Common model for summer, winter and autumn

First, I use all data; this means a single model for all periods (summer, winter and autumn). In the table 3.2, there are the average of errors of each points in percentage.

The error is $\frac{abs(Exact\ value - Approximative\ value)}{Exact\ value} * 100$.

All 1 minute	Summer &Winter & Autumn
Phvac(Text, Tint)	63.23%
Phvac(Text, G)	91.84%
Phvac(Tint, G)	81.44%
Phvac(Text, Tint, G)	155.91%

Table 3.2: Average errors for each common models of multi-linear regression for summer & winter & autumn.

Each graphical representation with the parameters b are in appendix 5.

Thus, we can observe that in the table 3.2 the best model is the power in function of internal and external temperature with an average error of 63.23%. It is twice less if we use three variables, 155% that is a bad model. This analysis is found on the graph figure 3.4 with boxes representing all errors, not just an average. The box Phvac (Text, Tint) is the smallest four so it is the best model.

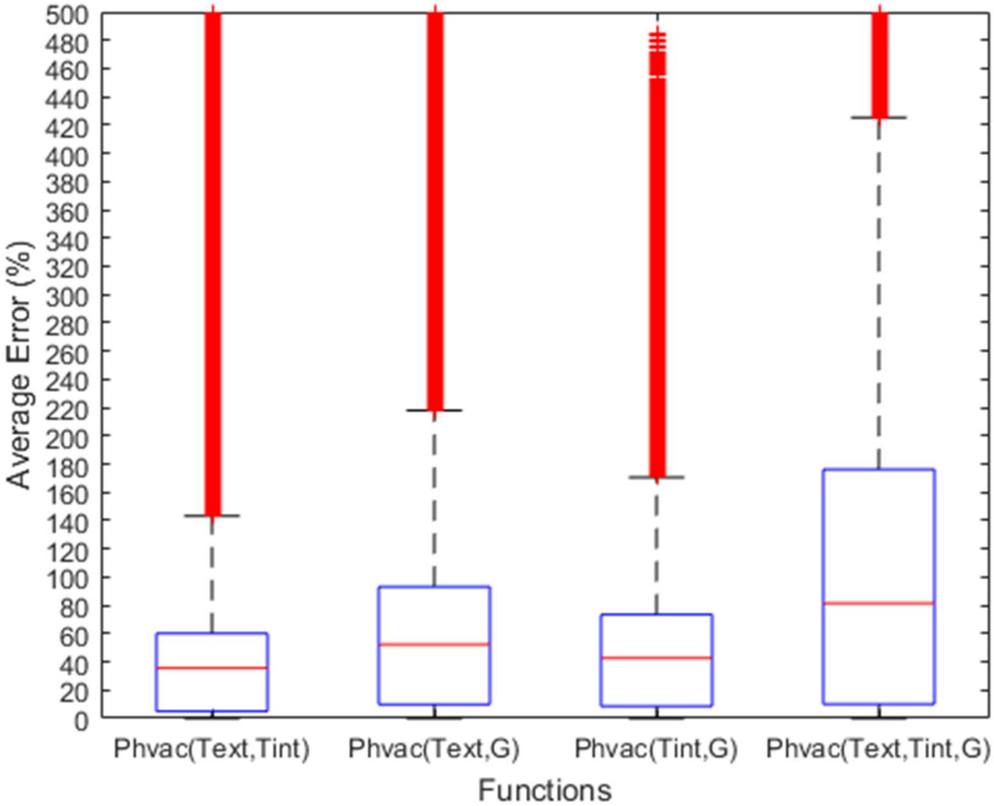


Figure 3.4: Box plots of the errors for each function of multi-linear regression present in table 3.2.

3.2.4 Model for summer, winter and autumn separately

Even if here, Phvac (Text, Tint) is the best model, 63.23% of average error is not good in general. In order to try to reduce the error, we decided to separate summer, autumn and winter:

All 1 Minute	Summer Only	Winter Only	Autumn Only
Phvac(Text,Tint)	72.71%	29.73%	32.16%
Phvac(Text,G)	96.35%	27.98%	36.08%
Phvac(Tint, G)	95.41%	33.16%	30.94%
Phvac(Text, Tint, G)	167.20%	158.60%	89.98%

Table 3.3: Average errors for each models of multi-linear regression during summer, winter and autumn separately.

We succeed to reduce about thirty points the average error for winter and autumn whereas for summer it increases about nine points. All graphical representations and parameters “b” for functions in table 3.3 are in appendix 6. From this moment, we are going to look specifically in the function Phvac (Text, Tint). As in the precedent part, I drew errors in box plots, figure 3.5).

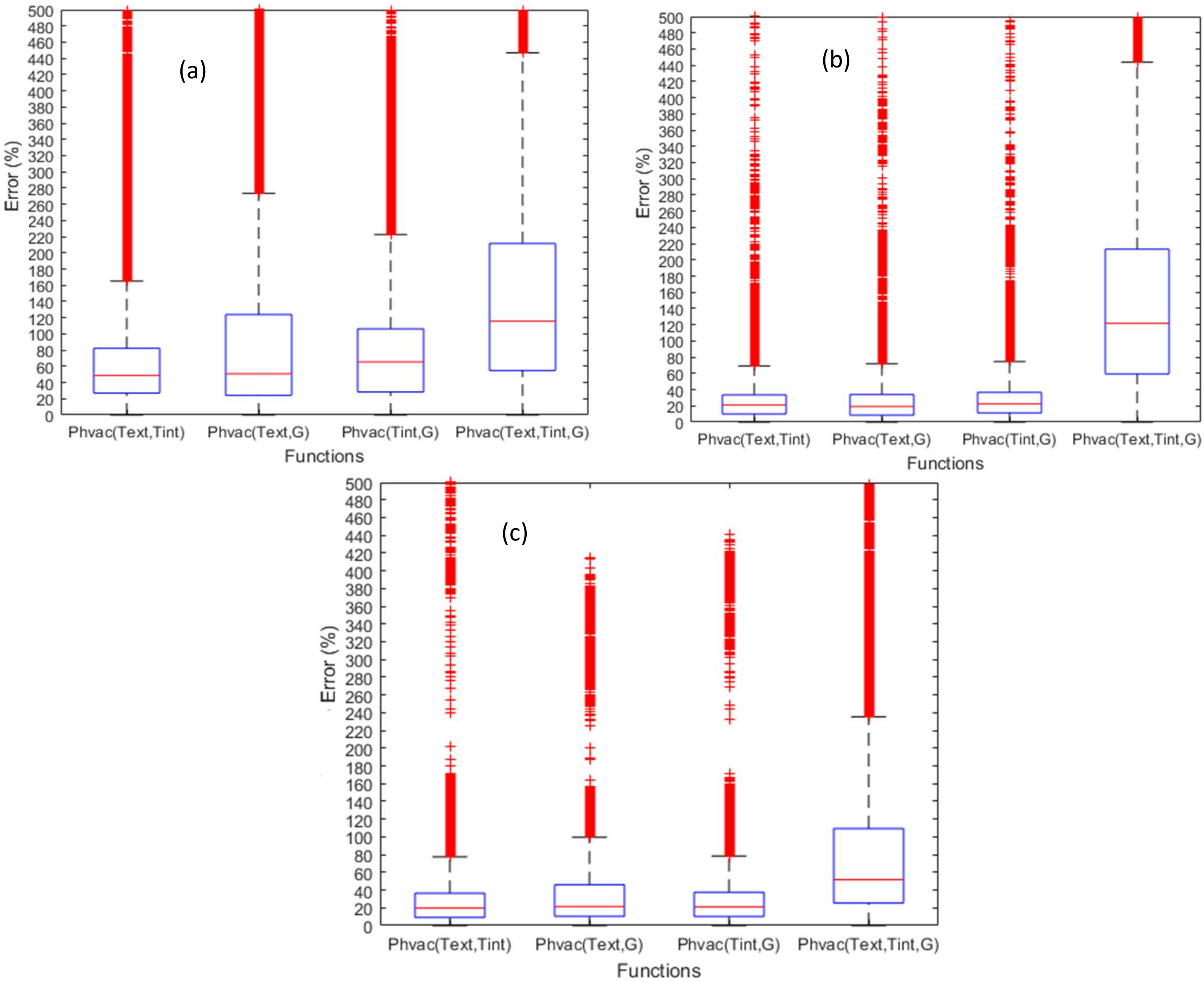


Figure 3.5: Box plots of errors for functions with multi-linear regression in table 3.3 where (a) Summer period, (b) Winter period,(c) Autumn period.

In figure 3.5.a the smallest box plot of errors is Phvac (Text, Tint) for summer. As regards winter figure 3.5.b, Phvac (Text, Tint) is also the best function but with a very weak difference (one point). Finally, in figure 3.5.c with autumn period I choose the function Phvac (Text, Tint) even if its box plot is larger than that of Phvac (Tint, G) but the difference is insignificant and it allows to have the same function for all seasons. Therefore, at the end in order to compare all models, I will use the function Phvac (Text, Tint) for summer, winter and autumn separately.

Chapter 4: Multiple polynomial regression

4.1 Theoretical aspect and function used

In order to realize this model we used the function “fit” in MATLAB for the multiple polynomial (with two explanatory variables) regression with the maximum number of degree possible in MATLAB, five. The form of the model is:

$$\sum_{i=0}^{5}\sum_{j=0}^{5} b_{ij} * X_1^i * X_2^j$$

In order to find the coefficients “b_{ij}”, the function “fit” apply also the least square method like in the chapter 3 but with a matrix X different. We decided with Dr. Matallanas to take five degrees for X_1 and X_2 after different tests showing that this polynomial gives the best results.

4.2 Modelling program

The user choses the period and the program load the associated data

```
clc;
clear;
format compact;

V1=input('For what period do you want the regression ? (Summer, Winter, Autumn))'
T1=strcat(V1,'_Data.mat');
load(T1);
```

The function “fit” creates the model choosing variables in data

```
X1m=Text_model;
A1='External Temperature (°C)';
X2m=Tint_model;
A2='Internal Temperature (°C)';
Ym=Phvac_model;
A3='Power Consumed (W)';


[fit4, gof]=fit([X1m,X2m],Ym,'poly55')
```

Plot the graph

```
plot(fit4,[Text,Tint],Phvac)
set(gca,'zlim',[-1000 1000],'ztick',-1000:200:1000)

xlabel(A1)
ylabel(A2)
zlabel(A3)
view(60,30)
```

Variables to calculate the error and the estimated points to draw the model

```
X1=Text;
X2=Tint;
Y=Phvac;
```

Obtain “b_{ij}” in “Coeffs” and create the matrix’s errors “S”

```
Coeffs=transpose(coeffvalues(fit4))
S=[];
```

Calculate the estimated points

```
for i=1:length(Y)
    if Y(i)~=0
        V=[1,X1(i),X2(i),X1(i).^2,X1(i)*X2(i),X2(i).^2,X1(i).^3,X1(i).^2*X2(i),X1(i
*X2(i).^2,X2(i).^3,X1(i).^4,X1(i).^3*X2(i),X1(i).^2*X2(i).^2,X1(i)*X2(i).^3,X2(i).^4
X1(i).^5,X1(i).^4*X2(i),X1(i).^3*X2(i).^2,X1(i).^2*X2(i).^3,X1(i)*X2(i).^4,X2(i).^5];
```

Calculate the average error

```
        Reg=(V*Coeffs);
        S=[S ((abs(Y(i)-Reg))/Y(i))*100];
    end
end

Mean_Error=mean(S)
```

Algorithm 4.1: MATLAB program for the multiple polynomial regression Phvac (Text, Tint).

I created a new MATLAB program to link the explanatory variable, external temperature, internal temperature and irradiance to HVAC's consumption by a polynomial relation. All explanations are in algorithm 4.1.

4.2.1 Common model for summer, winter and autumn

For all models, I created a table with all average error and a graph to choose the best model. Here it is not possible to do a model for Phvac (Text, Tint and G) with "fit". All graphical representations and parametres "b_{ij}" are in appendix 8.

All 1 Minute	Summer & Winter & Autumn
Phvac(Text, Tint)	60.29%
Phvac(Text, G)	57.05%
Phvac(Tint, G)	55.22%

Table 4.1: Average errors for each common model of multi-polynomial regression for summer & winter & autumn.

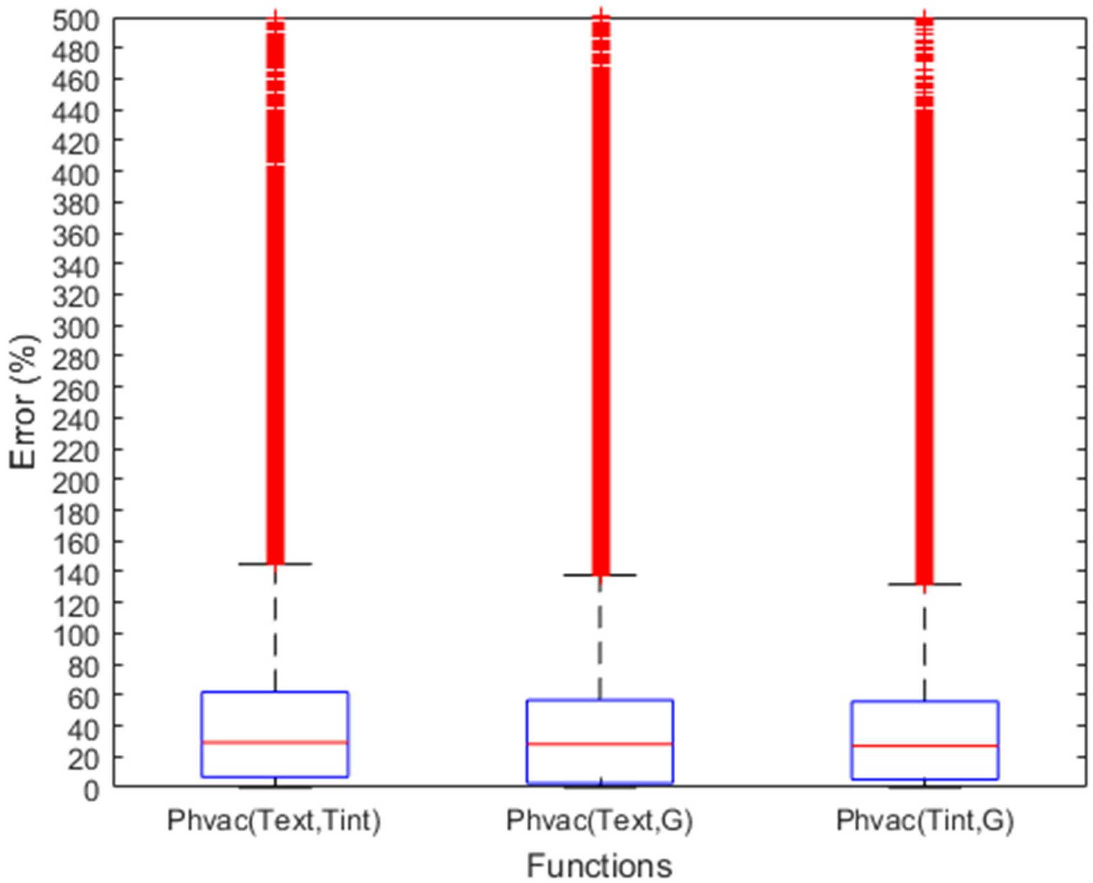


Figure 4.1: Box plots of errors for functions with multi-polynomial regression in table 4.1.

The graphs with the boxes confirms that the best model for the multi-polynomial is Phvac(Tint,G) with an averrage error of 55.22 % it's better than the multi-linear regression with 63.23 %.

4.2.2 Summer, winter and autumn considered separately.

Again, for trying to increase the precision of the models we separated the data of each period, all graphical representations and parameters "b_{ij}" are in appendix 9.

All 1 Minute	Summer Only	Winter Only	Autumn Only
Phvac(Text,Tint)	63.15%	30.17%	72.76%
Phvac(Text,G)	60.25%	29.78%	37.81%
Phvac(Tint, G)	66.90%	34.03%	33.02%

Table 4.2: Average errors for each models of multi-polynomial regression during summer, winter and autumn separately.

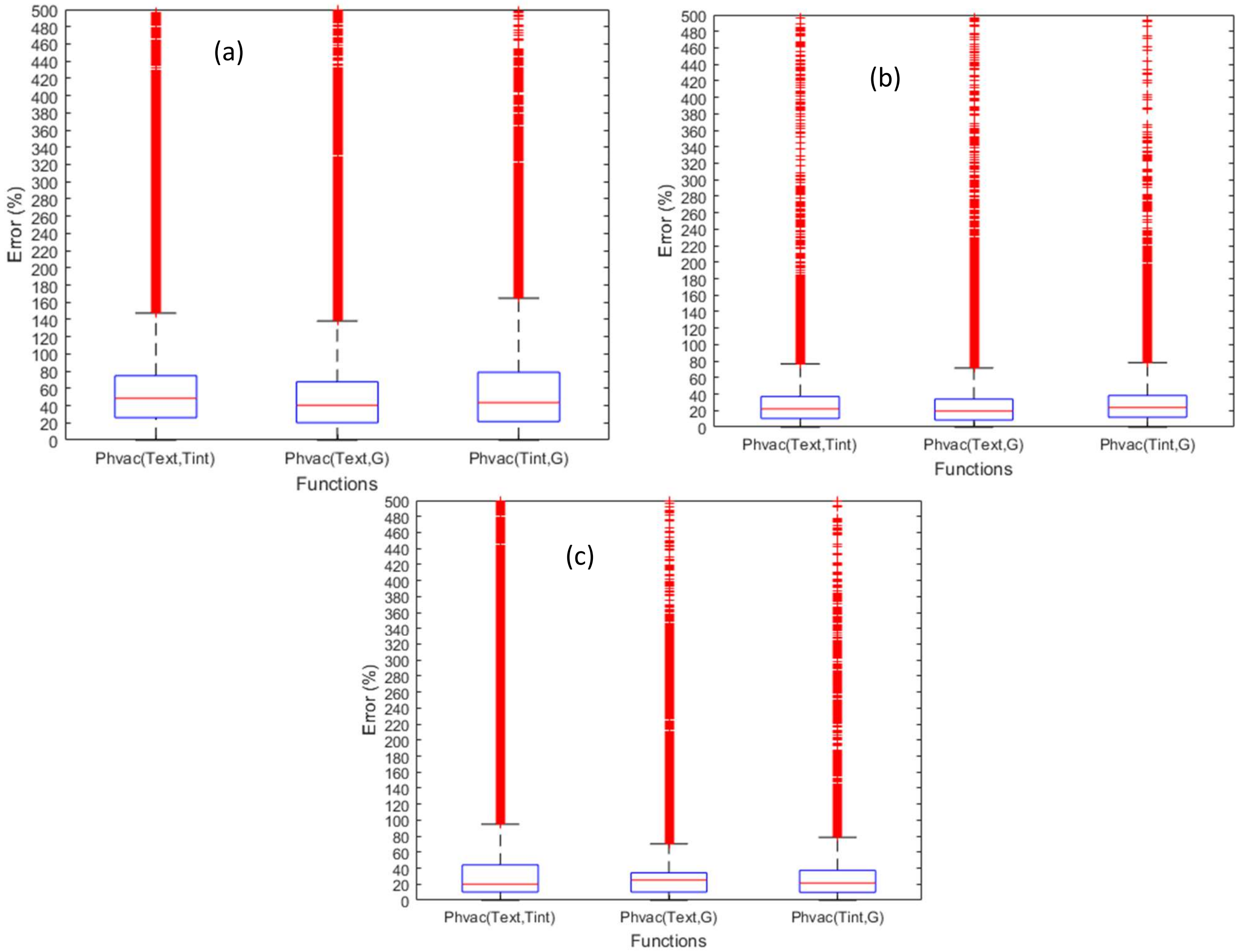


Figure 4.2: Box plots of errors for functions with multi-polynomial regression in table 4.2 where (a) Summer period, (b) Winter period, (c) Autumn period.

In figure 4.2, for each season the box of the function Phvac(Text, G) is the smallest of all. So at the end of the project we will compare the multi-polynomial regression to the ohers models with the function Phvac(Text, G).

Chapter 5: Quadratic Regression

5.1 Theoretical aspect and function used

The quadratic form is a homogeneous polynomial (a polynomial with several indefinite whose all monomials non-zero have the same number of total degree) of degree two with any number of variables. Here we used four explanatory variables: Text, Tint and G and in contrast to earlier we added the time in order to integrate the dynamism of the HVAC system. Therefore, here the quadratic form is with four variables is:

$$Y = b_1 + b_2 * X_1 * X_2 + b_3 * X_1 * X_3 + b_4 * X_1 * X_4 + b_5 * X_2 * X_3 + b_6 * X_2 * X_4 + b_7 * X_3 * X_4 + b_8 * X_1^2 + b_9 * X_2^2 + b_{10} * X_3^2 + b_{11} * X_4^2$$

5.2 Modelling program

For finding the coefficients “b_i” we integrated in a program the function “LinearModel.fit” using the least square method with each explanatory variables.

```
clc;
clear;
format compact;
```

The user choses the period and the program load the associated data

```
V1=input('For what period do you want the regression ? (Summer, Winter or Autumn))'
T1=strcat(V1,'_Data.mat');
load(T1);
```

```
mdl=LinearModel.fit([Text_model Tint_model G_model Time_model]...
Phvac_model,'quadratic')
Coeffs=mdl.Coefficients.Estimate;
```

The function “fit” creates the model and in “Coeffs” there are the parameters “b_{ij}”

```
X1=Text;
X2=Tint;
X3=G;
X4=Time;
A1='Time (Day)';
Y=Phvac;
A2='Power consumed (W)';
```

Choose variables

```
S=[];
for i=1:length(Y)
    Reg(i)=[1 X1(i) X2(i) X3(i) X4(i) X1(i)*X2(i) X1(i)*X3(i) X1(i)*X4(i) X2(i)*X3(i)...
X2(i)*X4(i) X3(i)*X4(i) X1(i).^2 X2(i).^2 X3(i).^2 X4(i).^2]*Coeffs;
    if Y(i)~=0
        if Reg(i)<0
            S=[S 0];
        else
            Error=(abs(Y(i)-Reg(i))/Y(i))*100;
            S=[S Error];
        end
    end
end
```

Calculate the errors

```
for i=1:length(Y)
    T(i)=[1 X1(i) X2(i) X3(i) X4(i) X1(i)*X2(i) X1(i)*X3(i) X1(i)*X4(i) X2(i)*X3(i) X...
(i)*X4(i) X3(i)*X4(i) X1(i).^2 X2(i).^2 X3(i).^2 X4(i).^2]*Coeffs;
end
```

Calculate the estimated points

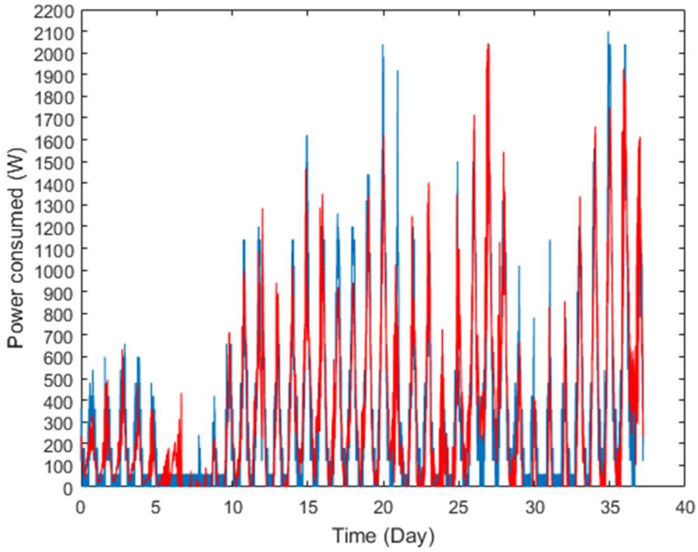
```
plot(Time,Phvac,Time,T,'r')
xlabel(A1)
ylabel(A2)
set(gca,'ylim',[0 2200],'ytick',0:100:2200)
Error=mean(S)
```

Plot the graph and give the average error

Algorithm 5.1: MATLAB program for the quadratic regression.

In this method, we added the time so we cannot have a global model for all periods because several months separate the winter period, the summer period and the autumn period. Nevertheless, we considered that the two weeks of data in winter are successive because all values are of the same order. The model is built with four explanatory variables, Text, Tint, G and Time but on the graph, we just plotted Phvac in function of the variable Time.

5.2.1 Model for summer



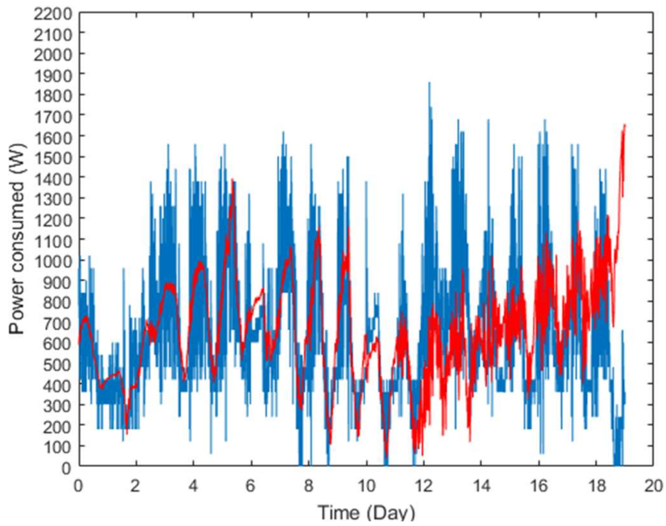
Estimated Coefficients:

	Estimate
(Intercept)	2903.9
x1	-171.29
x2	-76.541
x3	1.0303
x4	-189.85
x1:x2	1.7711
x1:x3	-0.042407
x1:x4	0.94945
x2:x3	-0.0063364
x2:x4	7.226
x3:x4	-0.022442
x1^2	3.6555
x2^2	0.13811
x3^2	9.1766e-05
x4^2	-0.34671

Figure 5.1: Phvac in function of the Time for summer created with a quadratic regression where X1=Text, X2=Tint, X3=G and X4=Time.

The average error for this model figure 5.1 is 49.35%. Moreover, graphically the estimated curve (red) gets closer strongly to the experimental curve (blue).

5.2.2 Model for winter



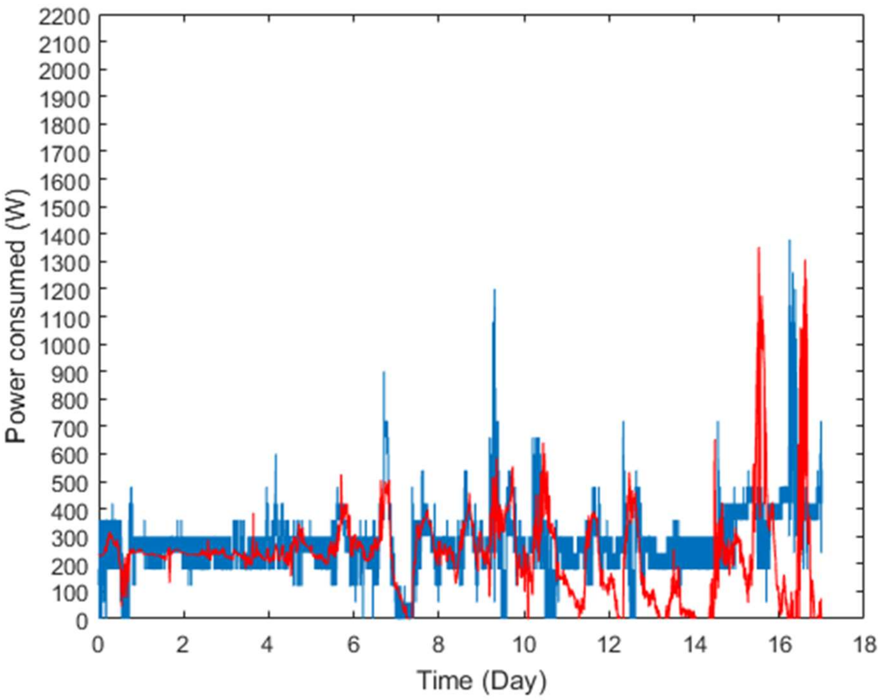
Estimated Coefficients:

	Estimate
(Intercept)	-6408.1
x1	-492.15
x2	997.36
x3	-0.19845
x4	208.5
x1:x2	30.053
x1:x3	0.016934
x1:x4	3.5601
x2:x3	0.0010651
x2:x4	-16.128
x3:x4	0.0096395
x1^2	-8.0323
x2^2	-32.883
x3^2	8.0454e-05
x4^2	4.867

Figure 5.2: Phvac in function of the Time for winter created with a quadratic regression where X1=Text, X2=Tint, X3=G and X4=Time.

The average error for this model figure 5.2 is 41.22%. Graphically the estimated curve is less accurate than the previous one for summer. However, the general form is respected.

4.2.3 Model for autumn



Estimated Coefficients:

	Estimate
(Intercept)	-5777.2
x1	176.71
x2	459.84
x3	-2.2695
x4	-203.19
x1:x2	-17.18
x1:x3	-0.12473
x1:x4	-1.1535
x2:x3	0.19467
x2:x4	11.538
x3:x4	0.060876
x1^2	5.9561
x2^2	-5.0682
x3^2	0.00018883
x4^2	-1.7749

Figure 5.2: Phvac in function of the Time for autumn created with a quadratic regression where X1=Text, X2=Tint, X3=G and X4=Time.

The average error for this model figure 5.3 is 39.60%. Graphically the model respect experimental points, even if in some places the amplitude is too low like between days zero and seven.

Chapter 6: Neural Network Design

6.1 Theoretical aspect