Technical University of Crete

Department of Electronic and Computer Engineering

Development of Novel Approaches for Smart Electric Grids Measurements Processing

Master Thesis

By

Blazakis Konstantinos

Thesis submitted for the degree of Master of Science in Electronic & Computer Engineering

Supervisory Committee
Prof. Georgios Stavrakakis (supervisor)
Prof. Konstantinos Kalaitzakis
Prof. Michalis Zervakis

Chania, Greece, January 2016

Abstract

The main issue for which we have written this master thesis is non-technical losses detection in a smart electrical system. Electricity theft produces non-technical losses and the advent of smart meters can be the best option to minimize electricity theft due to its high security, high efficiency and excellent resistance towards many of theft ideas in electromechanical meters but there still exists ways for power theft.

In the present study we modeled a small village with residential, commercial and industrial consumers using PowerWorld simulator.

By developing a smart meter reading database through Matlab for residential, commercial and industrial consumers, we apply an energy pattern for every consumer using PowerWorld Time Step Simulator tool.

Placing sum meters at the nodes of the power grid we measured the consumption of branches. The system performs a comparison between the sum meter of the branch and the smart meter readings of consumers corresponds to the branch subtracting the technical losses. If the sum meter measurement is greater than the sum of the downstream meters measurements then it is a strong indication of an illegal load or loads that consumes power.

Using Principal Component Analysis for dimensional reduction in customers' smart meter readings and applying Mean Shift clustering algorithm for first time in power theft detection, we classified the customers to legal and illegal.

In case of a problematic branch which no one of the branch customers is detected from clustering algorithms for power theft then, there is a strong indication for load or loads in the line without smart meter.

At last, we observed that Mean Shift algorithm gave us very satisfactory results for customers' patterns irregularities.

Περίληψη

Ανάπτυξη Καινοτόμων Μεθόδων Διαχείρισης Μετρήσεων σε Έξυπνα Δίκτυα Ηλεκτρικής Ενέργειας

Ο κύριος λόγος της συγγραφής αυτής της εργασίας είναι η ανίχνευση των μη τεχνικών απωλειών σε ένα έξυπνο ηλεκτρικό δίκτυο.

Ο κύρια αιτία των μη τεχνικών απωλειών είναι η κλοπή ηλεκτρικού ρεύματος και η έλευση των έξυπνων μετρητών είναι η καλύτερη μέθοδος για την ελαχιστοποίηση της ρευματοκλοπής λόγω της υψηλής ασφάλειας που παρέχουν, της υψηλής αποδοτικότητας και της άριστης αντίστασης σε περιπτώσεις ρευματοκλοπής σε σχέση με ηλεκτρομηχανικούς μετρητές ηλεκτρικής ενέργειας. Παρόλα αυτά όμως υπάρχουν μέθοδοι ρευματοκλοπής σε έξυπνους μετρητές.

Στην παρούσα εργασία μοντελοποιήσαμε ένα μικρό χωριό με οικιακούς, εμπορικούς και βιομηχανικούς καταναλωτές με την βοήθεια του λογισμικού PowerWorld.

Κατασκευάζοντας μια βάση δεδομένων για μετρήσεις ηλεκτρικής ενέργειας από έξυπνους μετρητές με την βοήθεια του λογισμικού Matlab για οικιακούς, εμπορικούς και βιομηχανικούς καταναλωτές, εισάγουμε τα δεδομένα αυτά σε κάθε καταναλωτή χρησιμοποιώντας το εργαλείο Time Step Simulator του PowerWorld .

Στην συνέχεια αφού τρέξουμε το Time Step Simulator παίρνουμε αποτελέσματα για την συνολική κατανάλωση και τις τεχνικές απώλειες των γραμμών. Ακολούθως, εφαρμόζουμε σενάρια ρευματοκλοπής σε διάφορα φορτία της γραμμής.

Το σύστημα κάνει μια σύγκριση της κατανάλωσης στον κόμβο και του αθροίσματος της κατανάλωσης στην γραμμή αφαιρώντας τις τεχνικές απώλειες. Εάν το φορτίο στον κόμβο είναι μεγαλύτερο από το άθροισμα των καταναλώσεων στην γραμμή τότε υπάρχει μεγάλη πιθανότητα για παράνομο φορτίο ή φορτία στην γραμμή.

Στη συνέχεια έχοντας εντοπίσει εκείνες τις γραμμές όπου έχουμε μη τεχνικές απώλειες κατηγοριοποιούμε τους καταναλωτές σε οικιακούς, εμπορικούς και βιομηχανικούς.

Χρησιμοποιώντας την τεχνική PCA (Principal Component Analysis) για την μείωση των διαστάσεων των δεδομένων που έχουμε πάρει για τους καταναλωτές από τους έξυπνους

μετρητές εφαρμόζουμε για πρώτη φορά στην ανίχνευση ρευματοκλοπής τον αλγόριθμο Mean Shift για να κατηγοριοποιήσουμε τους καταναλωτές σε νόμιμους και παράνομους.

Τέλος παρατηρήσαμε ότι ο αλγόριθμος Mean Shift μας δίνει πολύ ικανοποιητικά αποτελέσματα.

Acknowledgements

First of all, I would like to acknowledge my supervising Professor Stavrakakis Georgios for the assignment of this interesting thesis and for his continuous support and guidance.

Special thanks should be directed to some people whose contribution to the overall work was significant:

Antonis Tsikalakis whose advice guided me through all my research and helped me to gain insight about the overall problem, Giannis Katsigiannis for the technical support and recommendations he provided me, Vaggelis Tripolitakis for helping me in the programming part of my thesis and Plokamakis Georgios for sharing with me his experience in electrical networks.

Also, I would like to thank both professors Kalaitzakis Konstantinos and Michalis Zervakis for their participation in the three party committee of this thesis.

Last but not least, I want to thank my family to whom I owe my principles and values, whose love and caring supported me during my academic years.

Contents

1 Introduction

1.1 Problem Definition

The entire operation of the electrical network (generation, transmission and distribution) involves huge operational losses. Operational losses divided in two categories, technical losses that cannot be avoided and non technical losses due to electricity theft from illegal activities of consumers. In order to identify illegal consumers in the view of enhancing the economy of utilities, efficiency and security of the grid, a new method of analyzing electricity consumption patterns of customers and identifying illegal consumers is proposed and presented.

1.2 Dissertation Objectives and Scope

Technical losses that occur during the entire operation of the electrical network (generation, transmission and distribution) can be technically defined but non technical losses cannot be quantified completely. Substantial quantity of losses proves the involvement of Non-Technical Losses (NTL) mainly in distribution network due to electricity theft by illegal consumers.

Detection of illegal consumers is an extremely challenging problem nowadays due to the large amount of money that are not imputable to the State.

This dissertation presents a new method that uses customers' energy consumption patterns to detect illegal consumers in a smart grid environment. Initially, this dissertation conducts an extended survey on the methods implemented in pilfering electricity and technologies involved in smart energy meters. In general, utilities collect real-time energy consumption information from its customers several times every day. However, owing to the unavailability of that data, a dataset with near real-time energy consumption patterns has been developed in this work. Then, the state of the art clustering algorithm mean shift is proposed and implemented, which maps customer energy consumption patterns into legal and illegal.

1.3 Literature Review

In the recent past several techniques, were proposed to control illegal consumption or identify illegal consumers of electricity.

[1] Bandim et al. (2003). Proposed utilization of a central observer meter at secondary terminals of distribution transformer. Value of energy read by the central observer meter is compared with the sum of energy consumption values read by all energy meters in range. These two values of current are compared to estimate the total quantity of electricity that is being consumed illegally.

[2] Anand and Naveen (2003). Vigilant energy metering system (VEMS) is an advanced energy metering system that can fight electricity theft. It provides the data acquisition, transfer and data processing capabilities among the energy meters, local area stations and the base. It also facilitates load forecasting and control, identifies potential areas of theft, losses and takes measures to rectify it.

[7] Jamil et al. (2004). Proposed a microcontroller based energy meter facilitates the utility company to monitor and control the power supply to its spatially distributed consumers. This meter acts as a check meter to detect the meter tampering. However, e-metering systems can collect and process data, and can detect abnormalities in load profiles indicating electricity theft (De et al., 2003) [8] .

[9] Mano et al. (2004) Suggest proper design and implementation of rules for investigation of illegal consumers. Revenue Assurance and Audit Process is targeted at improving revenues for utility companies by reducing commercial losses by about 20% each year.

[10] Perez et al. (2005). Suggest that proper implementation of strategies in deployment and maintenance of the distribution networks can control commercial losses. In addition, strengthening the transformers at substations with higher configuration ones is also suggested. These suggestions can also be implemented in other African countries with the similar situations in order to improve the efficiency in transmission and distribution system.

[5] Pasdar and Mirzakuchaki (2007). Power line communications (PLC) presents an interesting and economical solution for automatic meter reading (AMR). If an AMR system via PLC is set in a power delivery system, a detection system for illegal electricity usage may be easily added in the existing PLC network. In the detection system, the second digitally

energy meter chip is used and the value of energy is stored. The recorded energy is compared with the value at the main kilowatthour meter. In the case of the difference between two recorded energy data, an error signal is generated and transmitted via PLC network. This paper describes a prototype of the detector system for illegal electricity usage using the power lines. The target of this study is to discover new and possible solutions for the lack of literature concerning this problem.

[3] Nagi et al. (2009). Proposed a novel approach of using genetic algorithm-support vector machines (GA-SVM) in detecting electricity theft. Load consumption data of all customers is collected and data mining techniques are used to filter and group these customers based on their consumption patterns. Customers are grouped into different classes based on the extent of the abnormality in load profile and then the customers with high probability of theft are inspected.

[4] Nizar and Dong (2009). The main objective here is to base the investigation on comparing the efficacy of the Support Vector Machine (SVM) technique with the newly emerging techniques of Extreme Learning Machine (ELM) and its OS-ELM variant as means of classification and prediction in this context. Non-technical Losses (NTL) represent a significant proportion of electricity losses in both developing and developed countries. The ELM-based approach presented here uses customer load-profile information to expose abnormal behaviour that is known to be highly correlated with NTL activities. This approach provides a method of data mining for this purpose and it involves extracting patterns of customer behaviour from historical kWh consumption data. The results yield classification classes that are used to reveal whether any significant behaviour that emerges are due to irregularities in consumption. In this paper, ELM and online sequential-ELM (OS-ELM) algorithms are both used to achieve an improved classification performance and to increase accuracy of results. A comparison of this approach with other classification techniques, such as the Support Vector Machine (SVM) algorithm, is also undertaken and the ELM performance and accuracy in NTL analysis is shown to be superior.

[6] Zeng et al. (2009). Proposed a new method to monitor the street lamp power cables and theft of cables. In this method, resonant frequency and equivalent capacitance of the cable are calculated by sending a frequency varying current signal to the street lamp power system. These values are used to detect the theft of cable, and then the location of theft.

[11] Soma Shekara Sreenadh Reddy Depuru (2010). This paper proposes an architectural design of smart meter, external control station, harmonic generator, and filter circuit. Motivation of this work is to deject illegal consumers, and conserve and effectively utilize energy. If a considerable amount of NTL is detected, harmonic generator is operated at that feeder for introducing additional harmonic component for destroying appliances of the illegal consumers.

[15] P.kadurek et al.(2010). This paper provides insight into the illegal use or abstraction of electricity in the Netherlands. The importance and the economic aspects of theft detection are presented and the current practices and experiences are discussed. The paper also proposes a novel methodology for automated detection of illegal utilization of electricity in the future distribution networks equipped with smart metering infrastructure. The necessary data requirements for smart meters and distribution substations are defined, in order to unlock this feature in distribution network. The paper also proposes the measures, which should be undertaken by the smart metering standards.

[12] Solomon Nunoo et al. (2011). The method includes receiving meter data of the measured power consumed by a customer, receiving delivered power data that includes data of the power delivered to the customer, determining a difference between the meter data and the delivered power data, determining that the difference between the meter data and the delivered power data is greater than a predetermined amount, and indicating a discrepancy if the difference between the meter data and the delivered power data is greater than a predetermined amount.

[14] Soma Shekara Sreenadh Reddy Depuru (2012). This paper designs and implements an encoding procedure to simplify and modify customer energy consumption data for quicker analysis without compromising the quality or uniqueness of the data. This paper parallelizes overall customer classification process. The parallelized algorithms have resulted in appreciable results as displayed in the results section of the paper.

[16] Shih-Che Huang et al (2013). In this paper, a state estimation based approach for distribution transformer load estimation is exploited to detect meter malfunction-tampering and provide quantitative evidences of non-technical loss (NTL). A measure of overall fit of the estimated values to the pseudo feeder bus injection measurements based on customer metering data aggregated at the distribution transformers is used to localize the electricity usage irregularity.

[13] Rong Jiang et al (2014). They discuss the background of Advanced Metering Infrastructure (AMI) and identity major security requirements that AMI should meet. Specifically, an attack tree based threat model is first presented to illustrate the energy-theft behaviors in AMI. Then, they summarize the current AMI energy-theft detection schemes into three categories, i.e., classification-based, state estimation-based and game theory-based ones, and make extensive comparisons and discussions on them.

1.4 Thesis Outline

Chapter 1: Discusses the problem definition and scope of dissertation and also provides a literature review in power theft detection techniques.

Chapter 2: Provides an insight to methods, factors and consequences of electricity theft.

Chapter 3: Illustrates the way we constructed consumers' patterns with Matlab, presents the software PowerWorld and the tool Time Step Simulator. Also, shows the electricity network model we constructed and the data mining algorithm Mean Shift that we use for power theft detection.

Chapter 4: Presents the whole procedure we followed for model development and shows the experimental results.

Chapter 5: Discusses the experimental results and further research for power theft detection.

2 Methods, Factors and Consequences of Power Theft

In general, electricity consumers may be classified as genuine consumers, partial illegal consumers, and completely illegal consumers. This chapter presents several simple and sophisticated methods used in pilfering electricity, discusses factors that influence illegal consumers to steal electricity, and refers in technical and non-technical losses in a power grid.

2.1 Methods of Stealing Electricity

The most common and simplest way of pilfering electricity is tapping energy directly from an overhead distribution feeder. The next most prevalent method of electricity theft is the manipulation of electromechanical energy meters that are used for recording and billing industrial, commercial and residential energy consumption. Though there are many techniques for tampering with such meters, some of these may include [22], [23] :

- Using mechanical objects

A customer can use some mechanical objects to prevent the revolvement of a meter, so that disk speed is reduced and the recorded energy is also reduced.

- Using a fixed magnet

A customer can use a fixed magnet like Neodymium to change the electromagnetic field of the current coils. As is well known, the recorded energy is proportional to electromagnetic field.

- Using the external phase before meter terminals

This method gives customers free energy without any record.

- Switching the energy cables at the meter connector box

In this way, the current does not pass through the current coil of the meter, so the meter does not record the energy consumption (Grounding the neutral wire).

In a smart grid environment, after installation of smart meters and related infrastructure, which are physically tamperproof, energy could still be tapped from the input terminal just before it enters the meter in comparison with the methods for electromechanical meters. Therefore, this problem should be solved by electronics and control techniques.

In addition, new software programs might be designed which help illegal consumers to hack smart meters or to manipulate meter readings.

Apart from physical damage to the meter terminals, there may be other techniques used to tamper the smart meter that are yet unknown.

2.2 Consequences of Power Theft

In daily basis operation utilities attempt to maintain good power factor, flat voltage profile and sufficient reactive power along the feeders. These operations may become difficult to perform due to dynamic and inadequate load flow information. On the other hand, illegal consumption of electricity might affect the performance of appliances connected to grid.

Primarily, electricity theft affects the utility company and then its customers. Electricity theft overloads the generation unit and decreases the frequency of the electricity network. In addition, quality of electricity supply is adversely affected, as the utility company has no estimation about the quantity of electricity to be supplied to genuine consumers as well as illegal consumers. This overload might result in overvoltage, affects the performance and even damage appliances of customers. This huge amount of NTL might trip the generation unit, which interrupts power supply to all customers .This unpredictable amount of additional load may lead to brown outs and black outs during the peak load period. In order to maintain good power factor and flat voltage profile along the feeders, sufficient reactive power has to be supplied beside the supplied electricity. Load shedding should also be done to compensate the voltage collapse during the peak load period. In addition, VAR compensation is very difficult without complete information about the total load flow because of the theft. In energy market, utility companies expect to earn their money back from the customers for the service provided, but most of this is lost due to the NTL. Electricity theft is a serious concern for utility companies as they are under threat of survival because of these incurring economic losses. Evidently, some utility companies in developing countries lose a considerable amount of their total revenue. These economic losses affect the utility company’s interest in development of the devices to control these losses. In addition, utility companies are forced to impose excessive tariff on genuine customers as they cannot afford all these losses by themselves which is absolutely not ethical to make genuine customers and utility companies responsible for the energy consumed by illegal consumers. In addition, illegal tapping of electricity raises safety concerns like electric shocks and even the death of a person who operates it. Improper handling of the distribution feeder might pose danger to the whole community, as these wires might start sparking and may cause fire during extreme weather conditions. Illegal consumers initiate tapping electricity from distribution feeders during scheduled power cuts.

2.3 Factors That Influence Illegal Consumers

Factors that influence consumers to steal electricity [23], [24] depend upon various local parameters that fall into multiple categories like social, political, financial, literacy, law, managerial, infrastructural, and economical. Of these factors, socio-economic ones influence people to a greater extent in stealing electricity.

In essence, electricity theft is proportional to the socio-economic conditions of the consumer. The most important factors are:

- Higher energy prices, unemployment or weak economic situation of a consumer.
- The belief that it is dishonest to steal something from a neighbor but not from a utility (public or large entity).
- Tax Purposes. Different tax in an electrified house.
- Some consumers might not be literate about the issues, laws and offenses related to the energy theft.
- Weak accountability and enforcement of law.
- Reasons to hide total energy consumption (e.g. Consumers who grow marijuana illegally or small-scale industries to hide overall production).

2.4 Electricity Losses

Electricity power losses [23], [24] can be classified into two categories: technical losses, and non technical losses. Electricity losses are defined as the difference between quantities of electricity delivered and quantities of electricity consumed by customers.

$$E_{loss} = E_{source} - E_{load}$$

Where:

E_{loss} . : is the amount of energy lost.

E_{source} : represents the energy that the source injects into the transmission line.

E_{load} : represents the energy consumed by the load at the other end of the transmission line.

2.4.1 Technical Losses

Technical losses in power systems are naturally occurring losses, which are caused by actions internal to the power system and consist mainly of power dissipation in electrical system components such as transmission lines, power transformers and measurement systems. The most common examples of technical losses include the power dissipated in transmission lines and transformers due to their internal electrical resistance. Technical losses are possible to compute and control, provided the power system infrastructure. Computation tools for calculating power flow, losses, and equipment status in power systems have been developed for some time. Improvements in information technology and data acquisition have also made the calculation and verification of technical losses easier. These losses are calculated based on the natural properties of components in the power system, which include resistance, reactance, capacity, voltage, and current. Loads are not included in technical losses because they are actually intended to receive as much energy as possible.

Two major sources contribute to technical losses: load losses consisting of the ohmic losses I^2R and impedance losses I^2X loss of the various system elements, and no-load losses which are independent of the actual load served by the power system. The majority of the no-load losses are due to transformer core losses resulting from excitation current flows.

2.4.2 Non Technical Losses

Non technical Losses (NTLs) refer to losses that occur independently of technical losses in power systems. NTLs are caused by actions external to the power system and also by the loads and conditions that technical losses computations fail to take into account.

NTLs relate to the customer management process and can include a number of means of consciously defrauding the utility concerned. More specifically, NTLs mainly relate to power theft in one form or another and can also be viewed as undetected customers' loads that the utilities don't know that exist. NTLs are more difficult to measure because they are often unaccounted by the system operators and thus have no recorded information. Two major sources which contribute to NTLs are: component breakdowns and electricity theft. NTLs caused by equipment breakdown are quite rare, where factors may include equipment struck by lightning, equipment damaged over time, and the elements of neglecting equipment or performing no equipment maintenance.

Even though equipment failure due to natural abuses like rain, snow and wind is rare, the equipment selected and the distribution infrastructure designed is in consideration with the local weather and natural phenomena. Reducing NTLs is crucial for distribution companies as these losses are concentrated in the LV network, their origins are spread along the whole system and are most critical at lower levels in residential, smaller commercial and light industrial sectors.

The most prominent forms of NTLs are electricity theft and non-payment, which are believed to account for most, if not at all NTLs in power systems.

3 Energy Consumption Patterns Generation, the PowerWorld Software, PCA and Mean Shift Data Mining Algorithm

This chapter explains the method we follow for constructing energy consumption data for residential, commercial and industrial customers, the model we use in PowerWorld for our experiments and the data mining algorithm we use for pattern recognition of energy data.

3.1 Generating Energy Consumption Data

With the advent of smart meters [25] and other smart grid infrastructure, it is now possible to access, collect, and analyze instant energy consumption of customers in real-time as explained. Exploiting such features offered by smart grid, customer energy consumption data is used to classify or identify illegal consumers. However, the energy consumption data required to test the proposed algorithms is unavailable owing to the privacy and confidentiality of utilities and customers. Therefore, this data has been developed using Matlab.

3.1.1 Factors Affecting Consumption Energy in the Urban Environment

Modern lifestyle and the increasing penetration of technology tend to increase the consumption of household electricity. The number and the type of electrical device used on a daily basis at home is constantly changing .Nevertheless the most intensive loads are :

- Space heating
- Water heating
- Refrigeration
- Lighting

There are many Factors affecting energy consumption at home, most essential of which are :

- Climate conditions

The weather conditions and the period during which a study of electricity consumption takes place, largely affects the results. This occurs as a load does not operate at the same frequency throughout the duration of time. For example water heating during the winter period and refrigeration during the summer period. Also fluctuations are presented in the lighting load as it changes depending on the day and the implementation of daylight saving time.

- Dwellers Income

The income of the inhabitants of a house determines consumer behavior and thus directly affects the number and type of electrical appliances of residence. For example a resident of a high income family is likely to have more than one television in comparison with families with low income.

- Dwelling characteristics

The length of time a household is inhabited whether it is permanent or seasonal residence affects electricity consumption during the course of the year. Also the size of the house and the type (apartment, building block) alters the nature and the size of loads.

- Residential Education

The educational level of the people is directly related to the use or not of some device such as the computer .Generally, educational level of residents affects their relationship with technology and therefore the type of device they own and operate.

- Characteristics of the way of life

The habits and way of life of people plays an important role in the way they use electricity. The number, the age, the professional categories of residents and the working hours decisively influence the energy consumption of the house as well as alter the period of activation and duration time of operation of most domestic appliances.

3.1.2 Consumers Categories

In our study we divide consumers into three categories: Residential, Commercial and Industrial .

3.1.2.1 Residential Consumption Pattern

The data used for the application of the method are mainly the hours staying at home and hourly probability a resident to make an action as a result the consumption of electricity.

The length of time during people are present within a dwelling decisively influences the profile loads of residence, as the majority of electric device require human presence to mobilize and supervision to remain in the working mode.

Due to the lack of recorded statistics, to determine the time periods where people are present on the house, the following scenarios were used as shown in Table 3.1.

Scenario 1	Absence from home 09:00 up to 13:00. Possibly inhabitants have part-time work in the morning.
Scenario 2	Home absence from 09:00 up to 18:00. Possibly inhabitants have a full time work.
Scenario 3	Home absence from 09:00 up to 16:00.
Scenario 4	Full home presence. Possibly infant existence under people supervision or elderly people presence.
Scenario 5	Home absence from 13:00 until 18:00. Possibly part time job in evening hours.
Scenario 6	Full absence on weekdays and partial presence at weekends. Possibly cottage near to the permanent residence.
Scenario 7	Full home absence. Presence only some days of the year for holidays. Possibly cottage far away from permanent residence.

Table 3.1 Human presence scenarios

Except the scenarios mentioned, we can simulate any scenario and various combinations of scenarios. The electrical loads that are included for residential consumption pattern are shown in Table 3.2

Load Categories	Electrical Loads
Personal hygiene	Water Heater, Hair Dryer
Preparation of food (breakfast, lunch, dinner)	Electrical oven, Microwave
Watching TV	Television
Heat, Cooling	Electrical Heating, Air Condition
Household chores	Vacuum cleaner, dishwasher, washing machine
Study	Computer
Base Load	Refrigerator, Freezer
Lighting	Electric lamps

Table 3.2 Load categories and electrical loads for residential customer

The information concerning the chances of conducting a domestic activity during the course of the day was taken from statistical data [17].The statistical data is showing the probability percentage of holding a residential activity by some residents for each of the 96 quarters of day considering that the day starts at 00:00 and ends at 24:00.

In Figure 3.1 and Figure 3.2 we can notice the probability for an inhabitant to eat dinner and breakfast during a day.

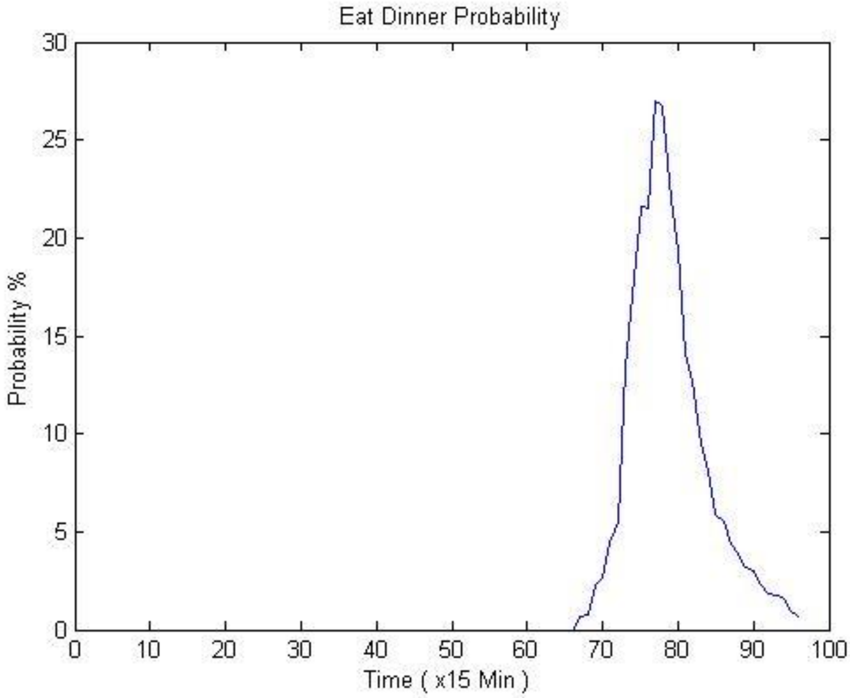

Figure 3.1 Probability for a resident to eat dinner during a day

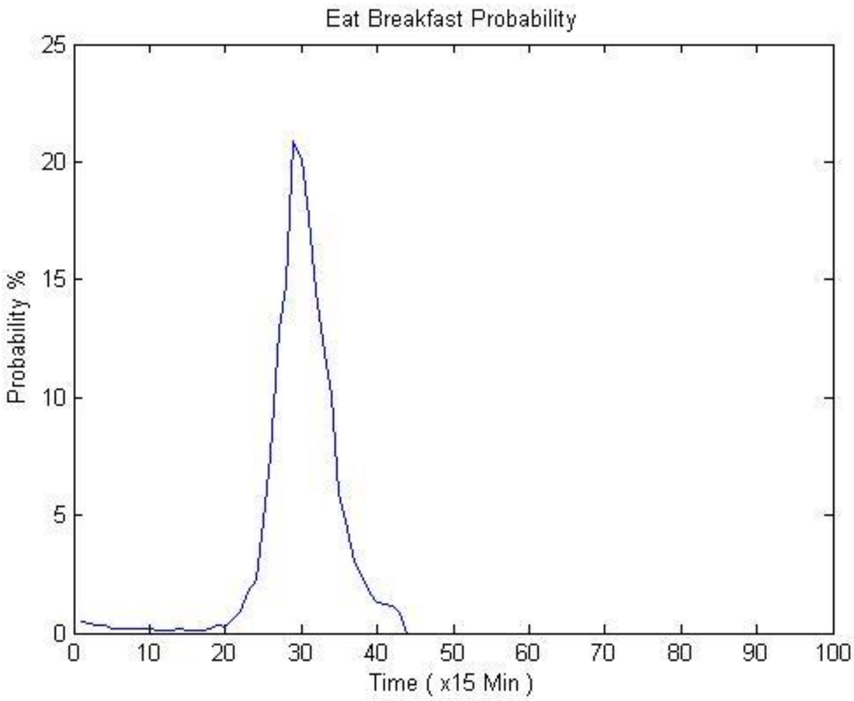

Figure 3.2 Probability for a resident to eat breakfast during a day

Having gathered the necessary data entry, for example the type of appliances and their characteristics, the probability of the activity undertaken, the duration of each activity, periods of human presence on the house and environmental conditions we proceeded to the construction of residential profile loads.

Initially, we defined scenarios as shown in Table 3.1 and we chose a scenario that will be simulated. The scenarios defined in the application as a matrix with 96 columns where a column for each quarter denoted by 1 for physical presence at home and with 0 for absence. Afterwards we introduced 96 columns with probability of conducting an activity. Due to the use of more device involves the physical presence of a person at home, the probabilities tables are multiplied by the table of scenario. With this way we ensure that a device cannot be used while there is no human presence at home.

For each load we created a table of 96 columns. Every column shows the power consumed by the specific device for each of the 96 quarters of the day.

This table is initially a zero table. Given the nominal consumption of the appliance in watts and time works every 15 minutes the table is complemented by the value of the power from column 1 until the column that corresponds to the operating time. An initial profile device created considering that the operation starts from 00:00.

Then the device consumption moved in time when it is more likely to operate. This is done using an algorithm that finds the appropriate time at which the probability process conduction related to the use of the appliance concerned is maximum. Afterwards, the algorithm shifts the device consumption when it is more likely to operate.

The method shown schematically in figures below by an example relating to the operation of the television in case of full home presence (Scenario 4). Figure 3.3 shows the distribution probability for monitoring television set from someone residents the house. Figure 3.4 presents the original profile loads of television before the time shifting, while Figure 3.5 shows the final profile loads of television that has been moved by the algorithm in time when television monitoring probability is maximum.

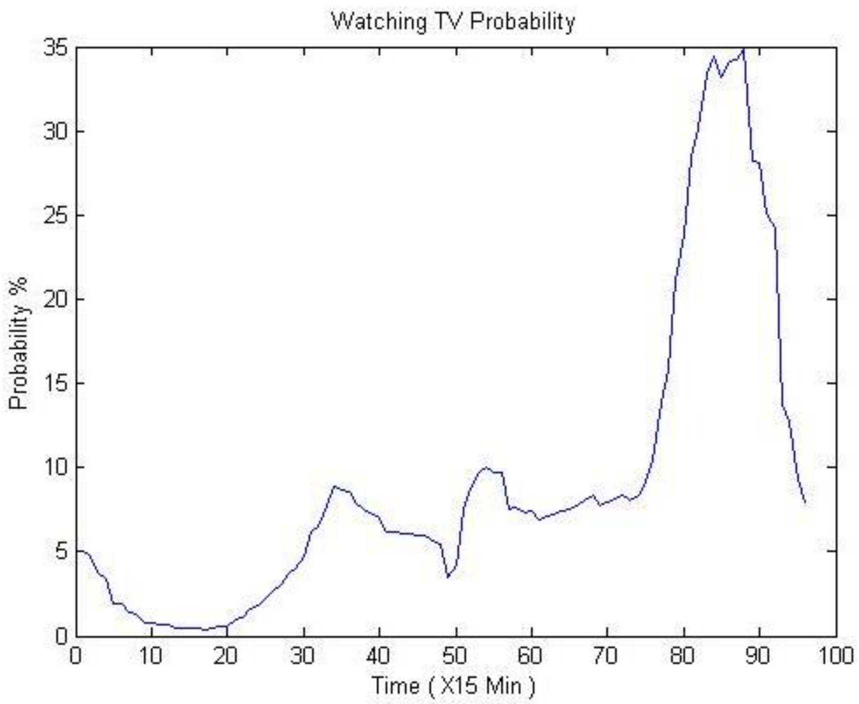

Figure 3.3 Probability for watching Tv during a day

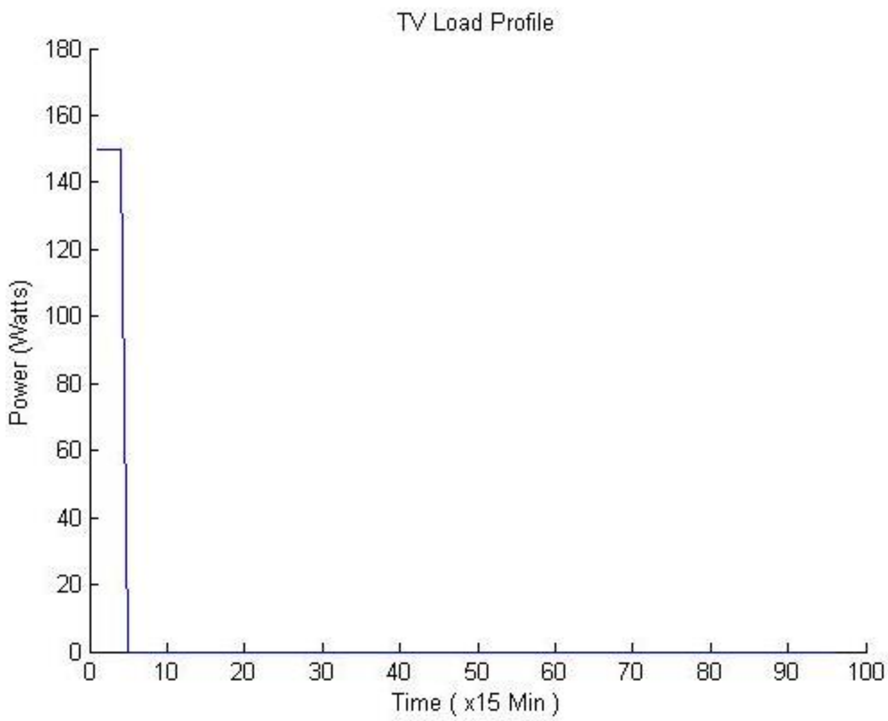

Figure 3.4 Tv load profile (Operation Time)

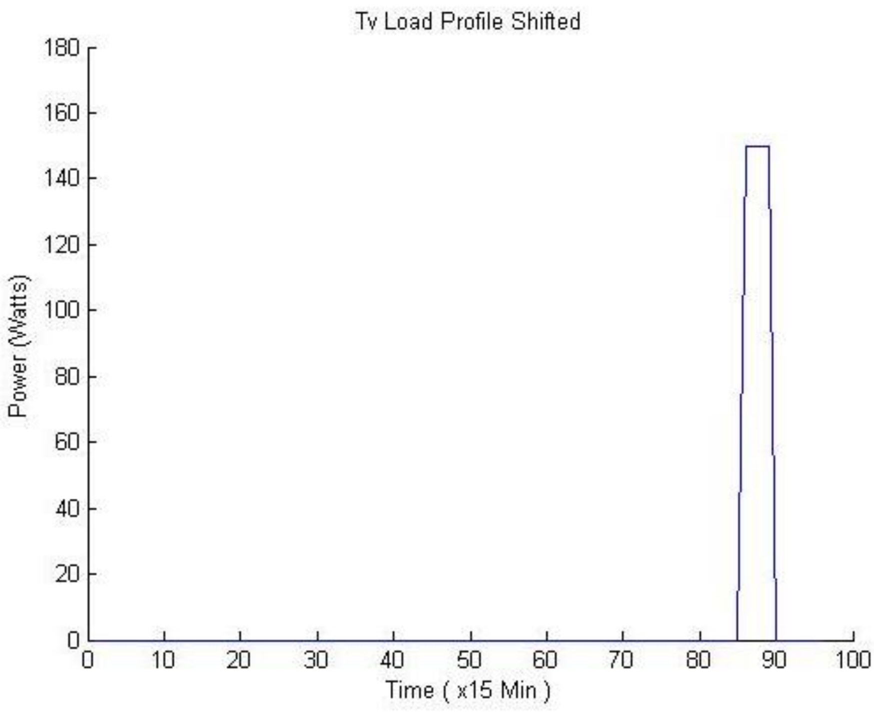

Figure 3.5 Tv load profile shifted to the maximum probability

For the loads profile that do not depend on human habits but their operation is normally defined by other factors such as base load we followed different process. Considering that the base loads are refrigerators and freezers we add to the load profile a load value that is turned on and turned off at specific time intervals with a probability as shown in Figure 3.6 and Figure 3.7.

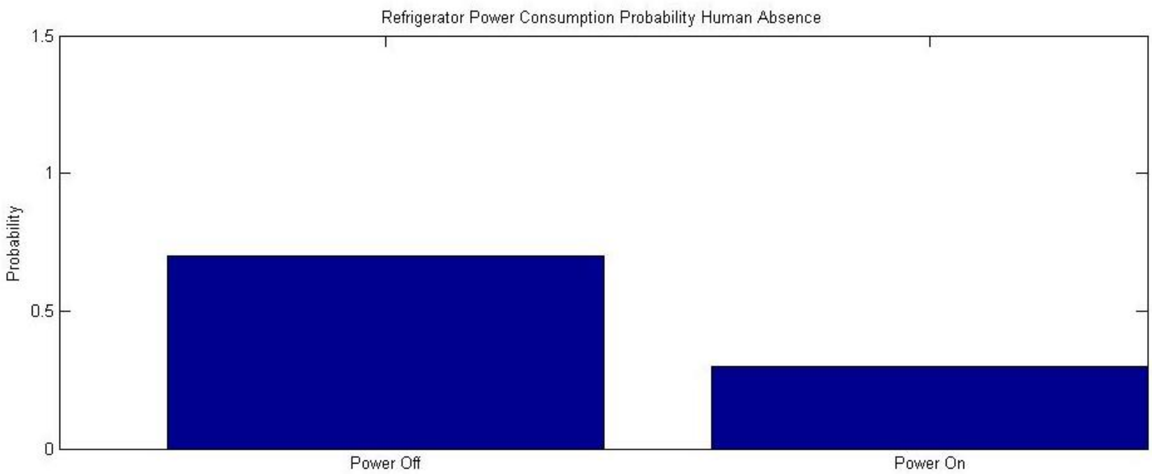

Figure 3.6 Probability for refrigerator operation with human absence from home

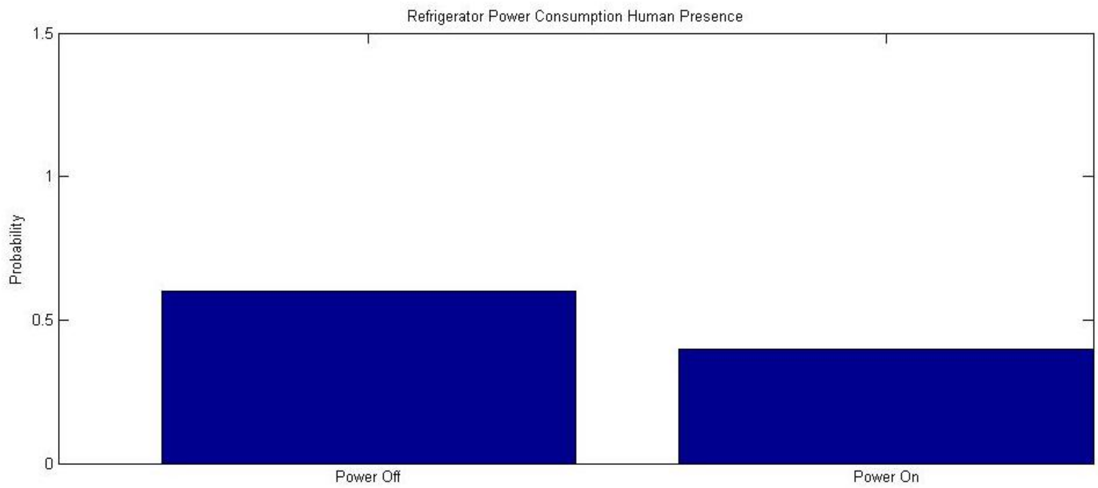

Figure 3.7 Probability for refrigerator operation with human presence at home

The method described above creates the load profile of any electrical device independently Total home load profile occurs by the superimposition of all load profiles of individual electrical appliances.

We follow the same process for all the device so we created a day profile for a single scenario as shown in Figures below.

In Figure 3.8 we can observe a day profile for scenario 2. As mentioned above in scenario 2 there exists home absence from 09:00 up to 18:00 that can be observed in Figure 3.8. In horizontal axis from 0 to 24 (00:00-06:00) and from 36 to 72 (09:00 -18:00) and only base load is operational whereas the remaining hours human activities take place.

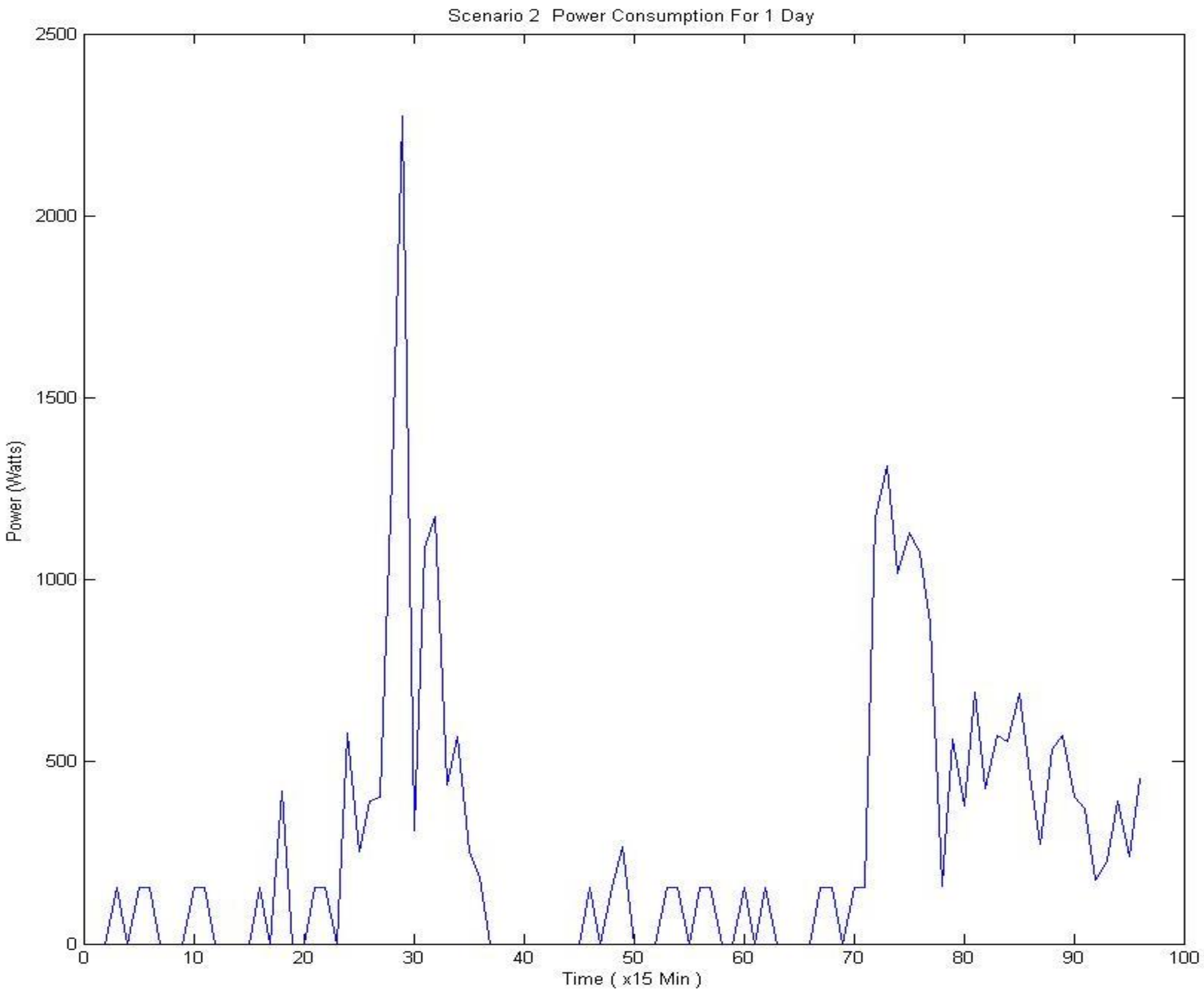

Figure 3.8 Scenario 2 power consumption for 1 day

Due to our model examine the load profile from 1st April until 30th September we repeated the day profile among all days except for the weekends by changing a little the duration of activities and the magnitude of loads.

Due to the changing of habits at weekends we construct 10 different profiles for weekends as people tend to be for outdoor activities at weekends.

Considering that residents do holidays once a year and examining a period from 1st April to 30th September we constructed a probability for the duration of vacations and the time started as shown at Figure 3.12 and Figure 3.11.

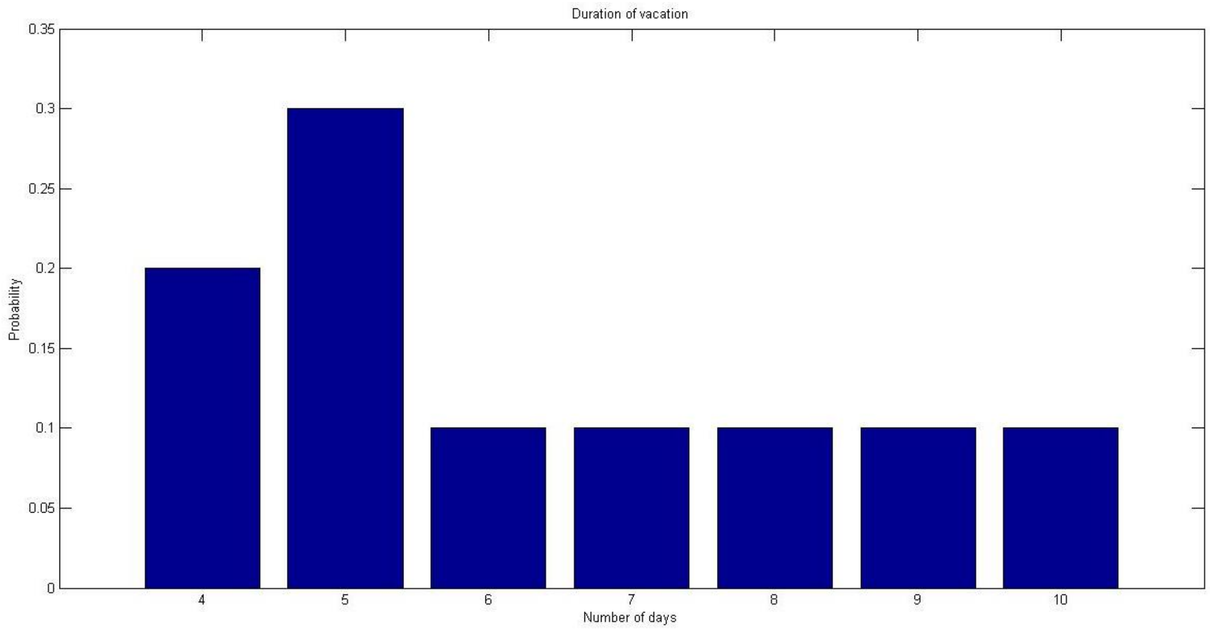

Figure 3.9 Probability for the number of vacation days

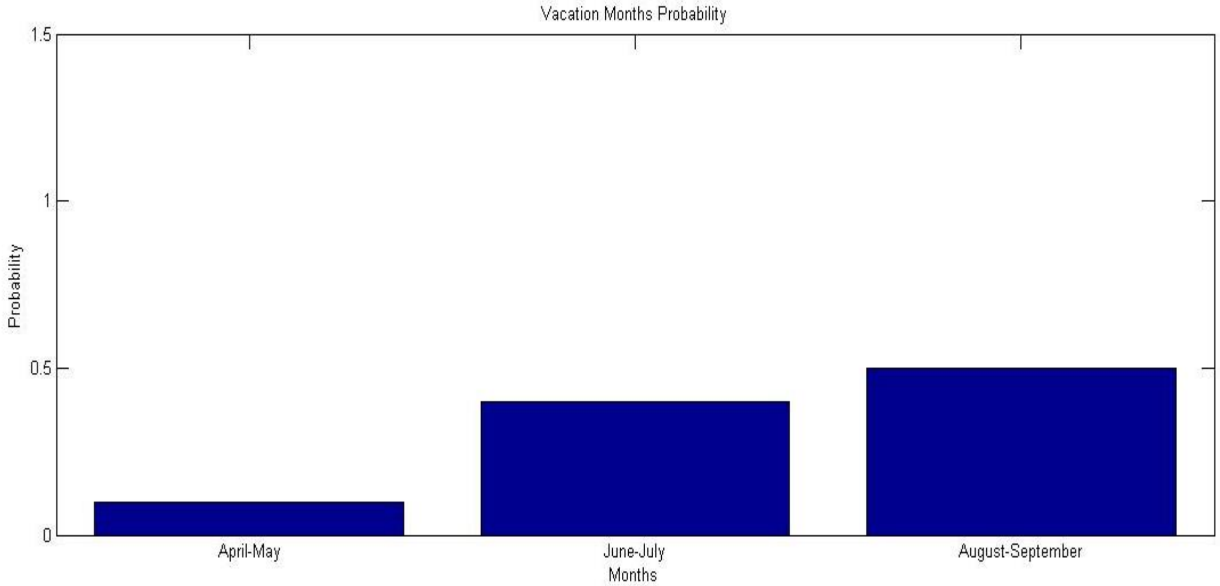

Figure 3.10 Probability for the month that a vacation begin

To include the influence of environmental conditions in the profile loads considering an increased temperature on summer months we increased the cooling loads from the final charts load as we examine a period from 1st April until 30th September.

Taking into account all the factors above we constructed a pattern for scenario 2 as shown in Figure 3.11 and Figure 3.12.

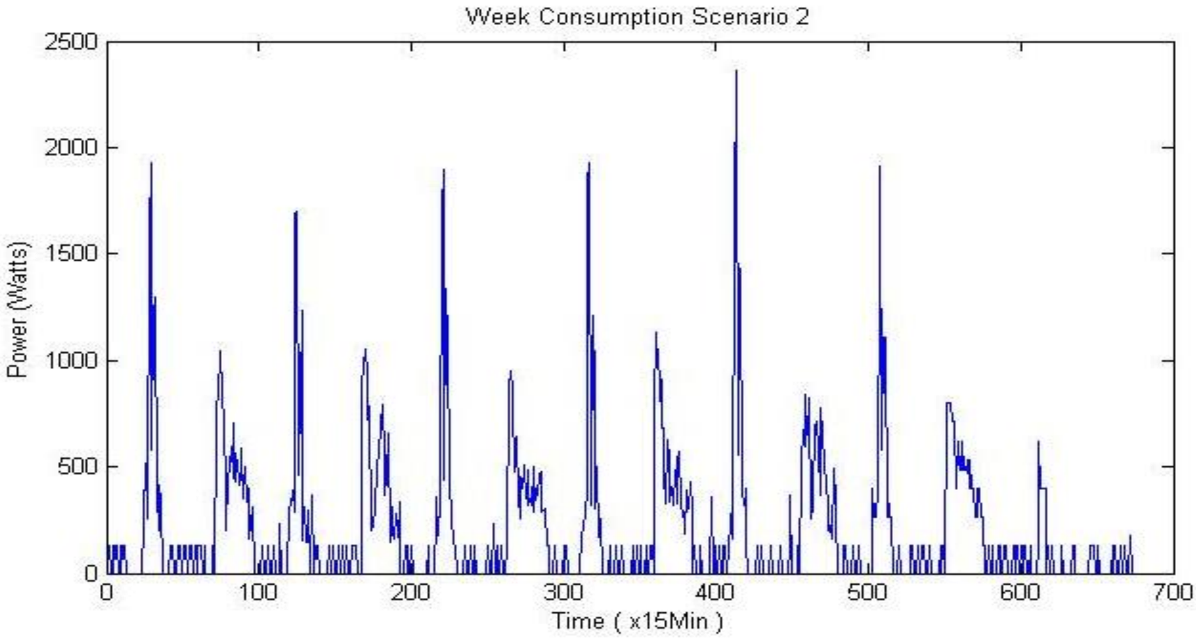

Figure 3.11 Scenario 2 power consumption for a week

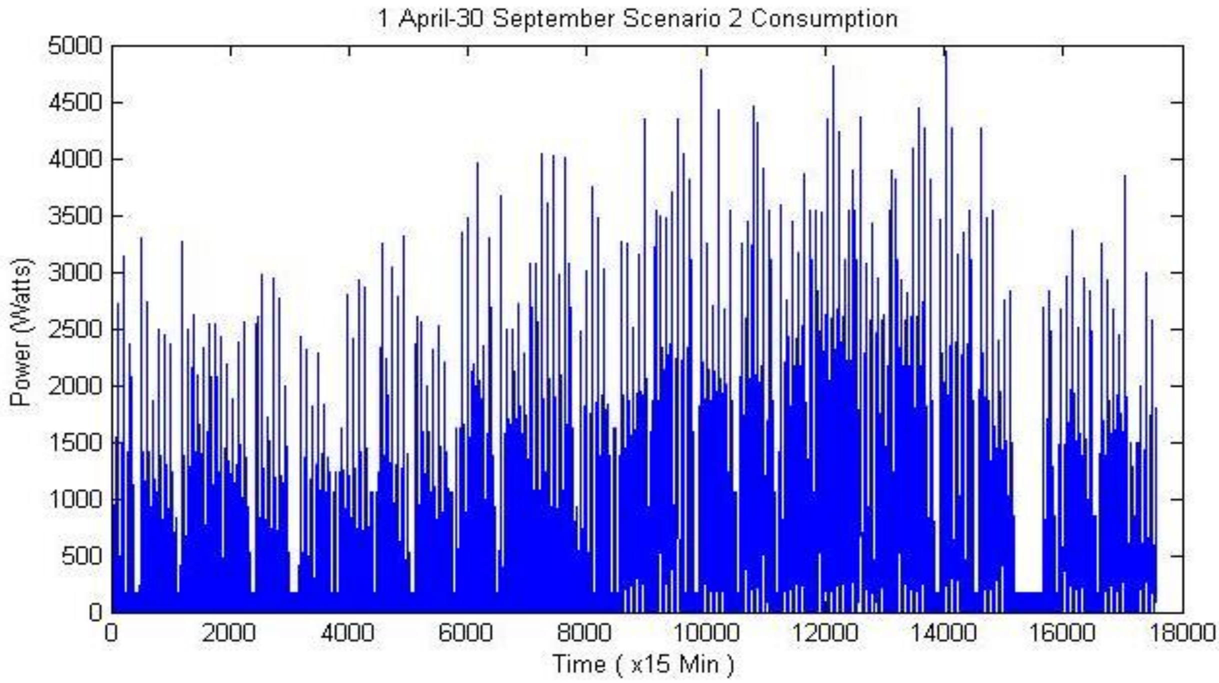

Figure 3.12 Scenario 2 power consumption from 1st April to 30th September

For every scenario we constructed 50 different patterns which differs between them in magnitude of loads (has to do with the total appliances at home).

For scenario 6 and 7 we follow different method. Especially for scenario 6 we consider that in weekdays only base load is operational and some weekends the house is inhabited with greater probability at weekends of July and August as shown in Figure 3.13.

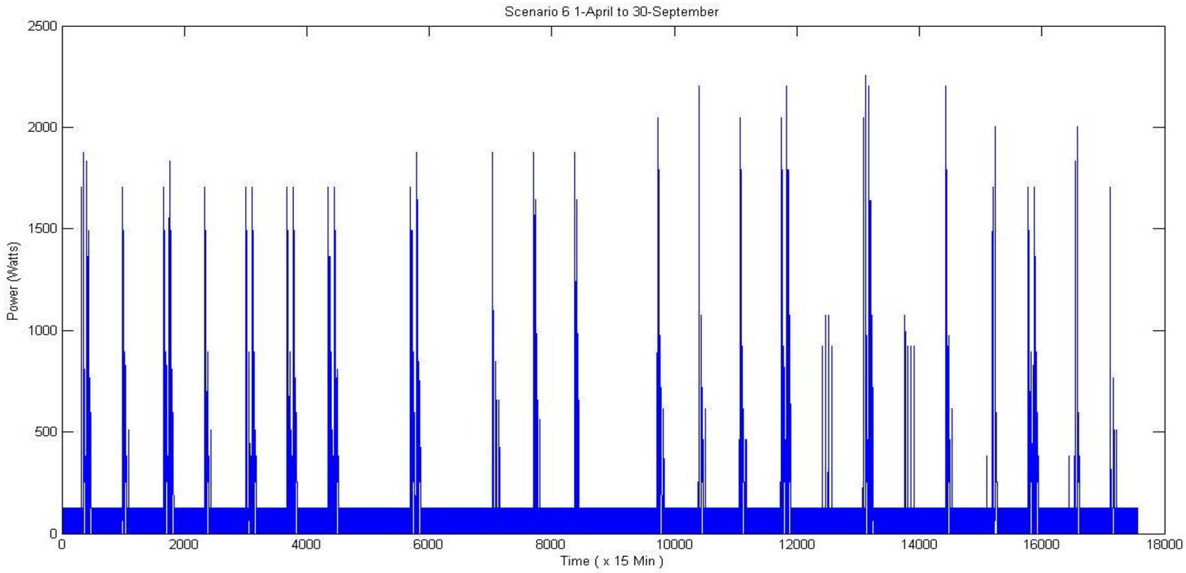

Figure 3.13 Scenario 6

For scenario 7 we consider that the base load isn't in operation. We constructed a probability table for the time periods that a resident visits the house (Figure 3.15) a probability table for the amount of house residence (Figure 3.14) and the duration of residence (Figure 3.9).

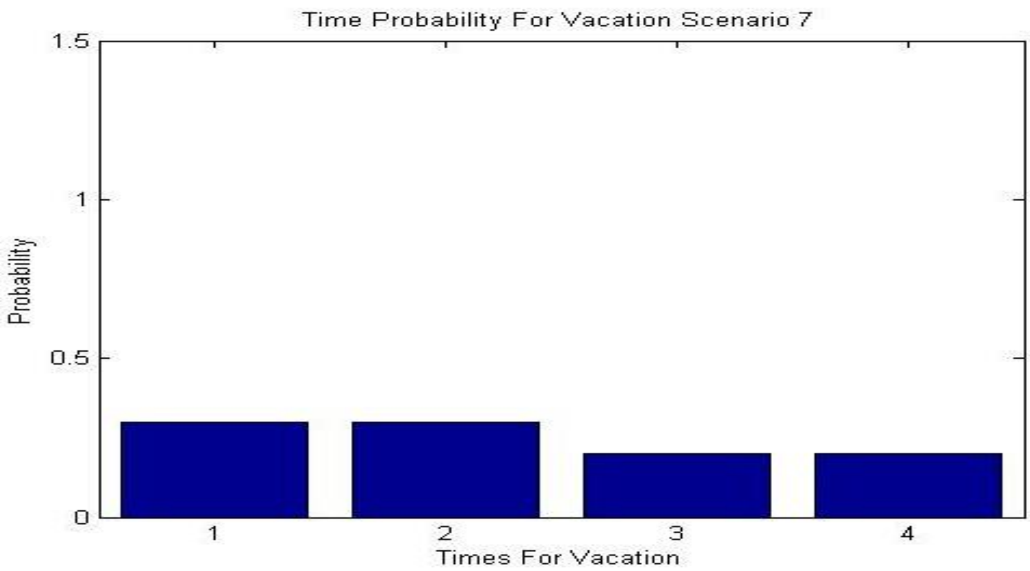

Figure 3.14 Probability for the number of vacation

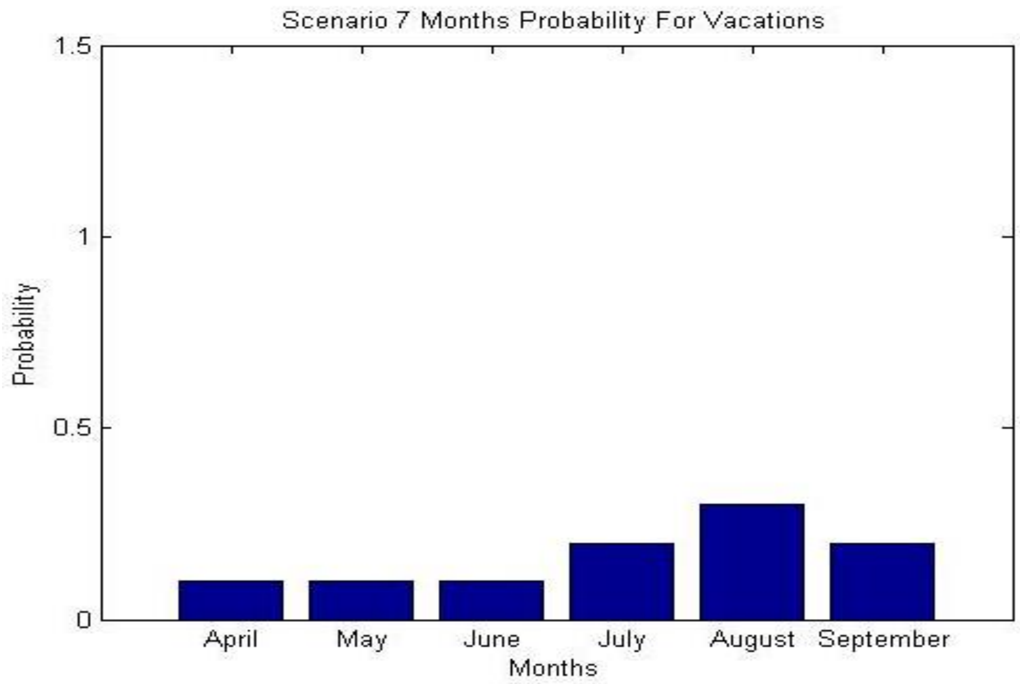

Figure .3.15 Months probability for vacation

In Figure 3.16 we can see a snapshot of scenario 7 from 1st of April to 30th of September.

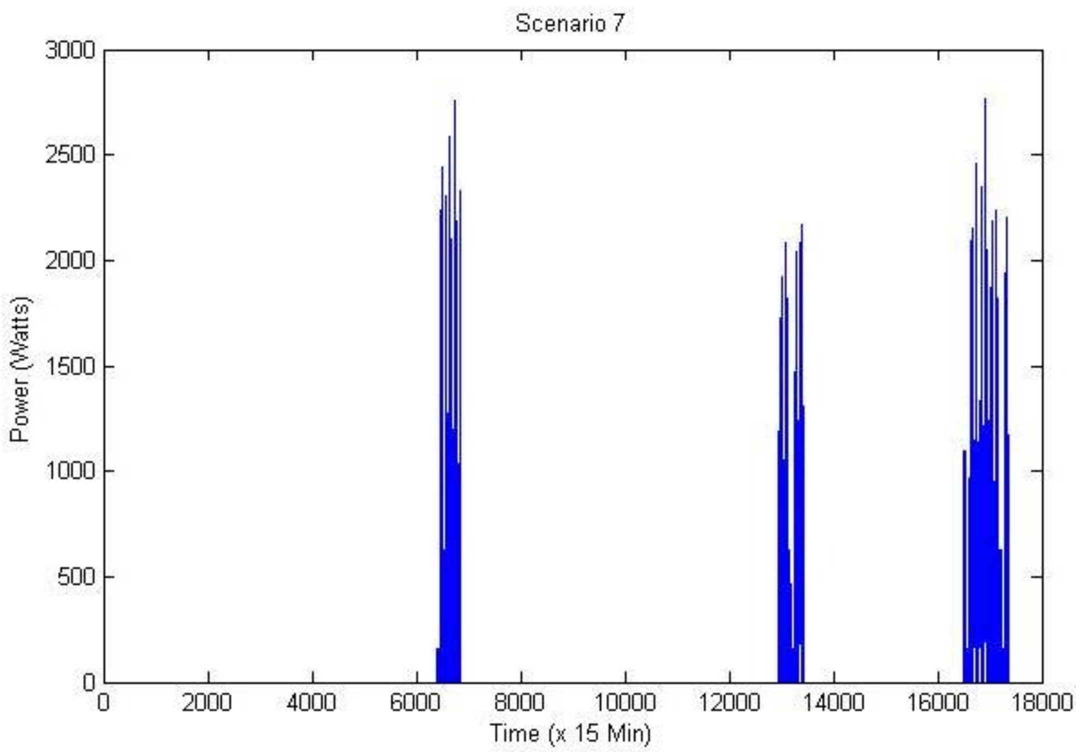

Figure 3.16 Scenario 7 power consumption from 1st April to 30th

3.1.2.2 Commercial Consumption Pattern

In our study we included 13 businesses as shown in Table 3.3.

Businesses	Number Of Businesses
Kiosk	2
Coffee shop	1
Coffeehouse	1
Souvlaki diner	1
Restaurant	1
Bakery	1
Super Market	2
Clothing store	1
Hotel	2
Church	1

Table 3.3 Businesses

The method that we follow for constructing commercial consumption pattern is almost the same as the method of residential consumption pattern. Differences exist in scenarios and the categories of electric loads. For example a bakery works from 00:00 until 18:00 (0-72 in X axis) as shown in Figure 3.17. Activities (Oven operation) starts at 03:00 until 06:00 (12-24 in X axis) and after bakery works like a shop until 18:00 (72 in X axis). We can observe that from 18:00 until 00:00 (72-96 in X axis) only the base load is in operation.

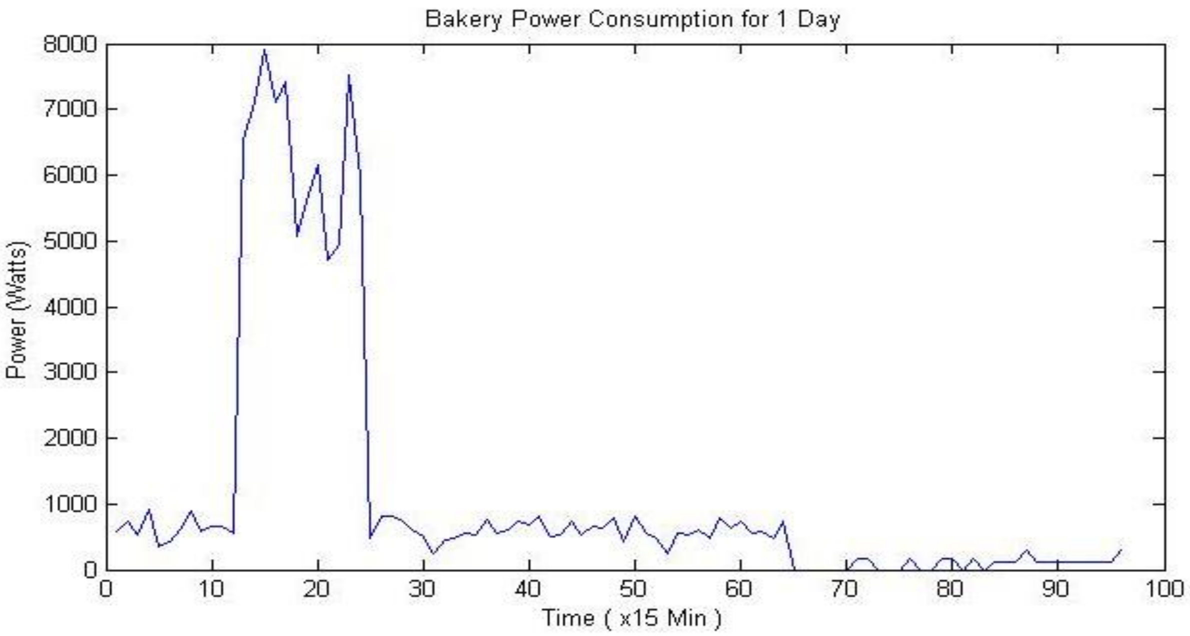

Figure 3.17 Bakery for 1 day

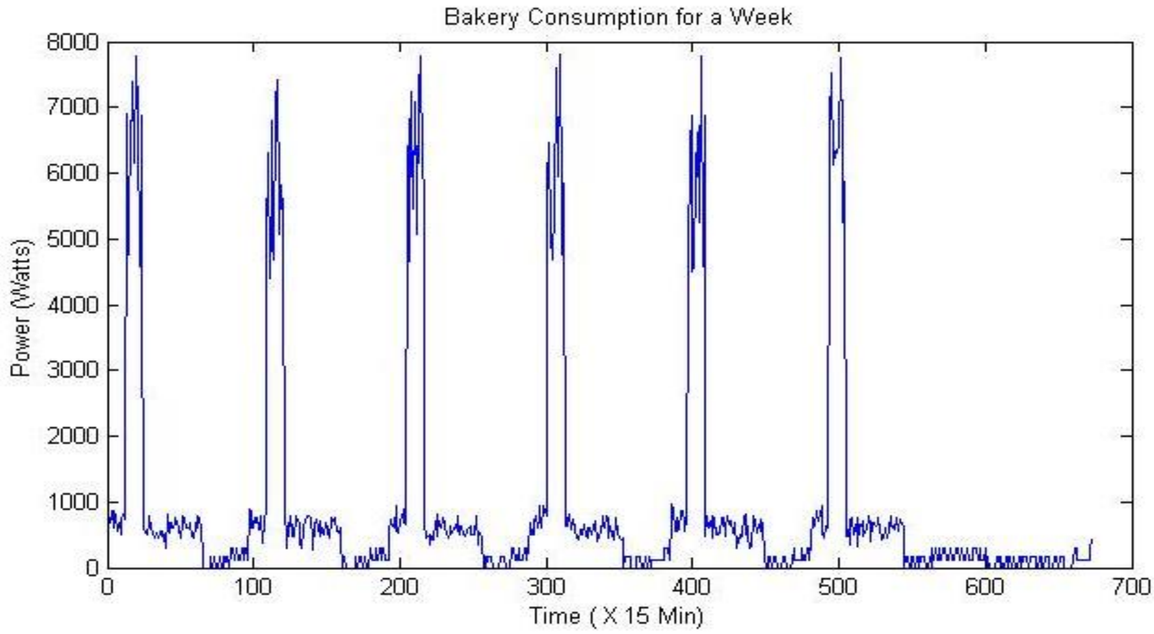

Figure 3.18 Bakery for a week

In Figure 3.18 we can see the bakery operation for a week.Bakery works for 6 days and on Sunday is closed and only base load is working.

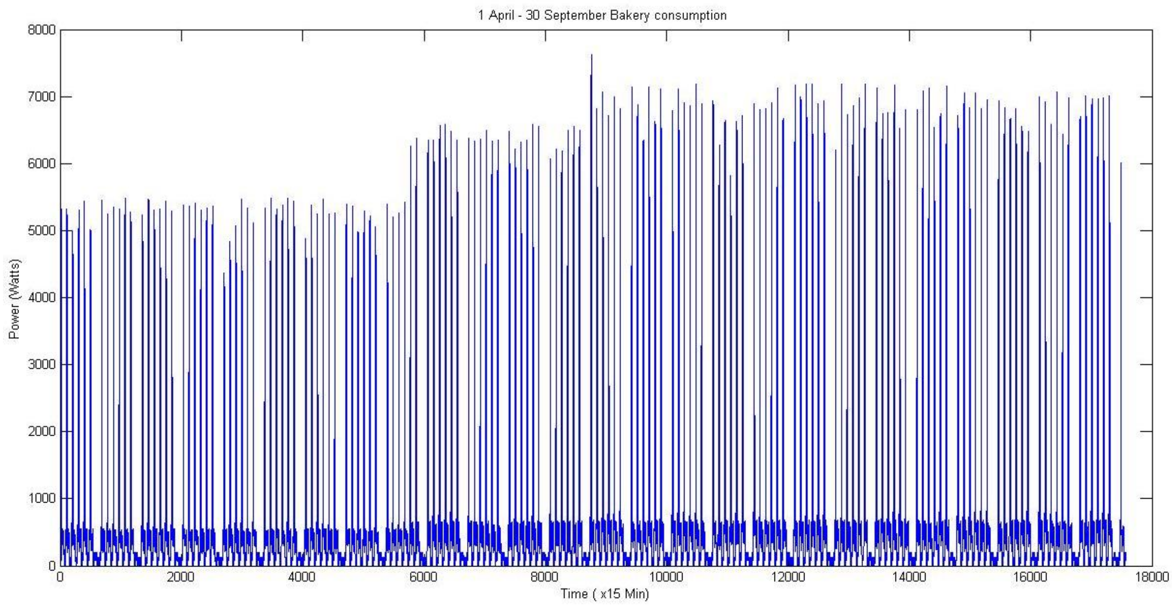

Figure 3.19 Bakery power consumption from 1st April to 30th

In Figure 3.19 we can observe the bakery operation during a period from 1stApril until 30th September. We can see that the total load from June to September is increasing due to the cooling loads that are in operation.

3.1.2.2 Industrial Consumption Pattern

In our study we included 3 different industries as shown in Table 3.4.The method that we followed for constructing industrial consumption pattern is almost the same as the method of residential and commercial consumption pattern. Differences exist in scenarios and the categories of electrical loads.

Industries	Max Power
Small industry	15 kw
Medium industry	20 kw
Large Industry	30kw

Table 3.4 Industries

In Figure 3.20 we can see a daily profile of a large industry. From 00:00 to 09:00 (0 to 36 in X axis) only base load is working (sign).Activities stop at 20:30. In Figure 3.21 we can observe that the industry is working the whole week apart from Sunday. In Figure 3.22 we present the industry profile during the whole period from 1st of April to 30th of September. We can observe that industry pattern are fairly constant over day and seasons.

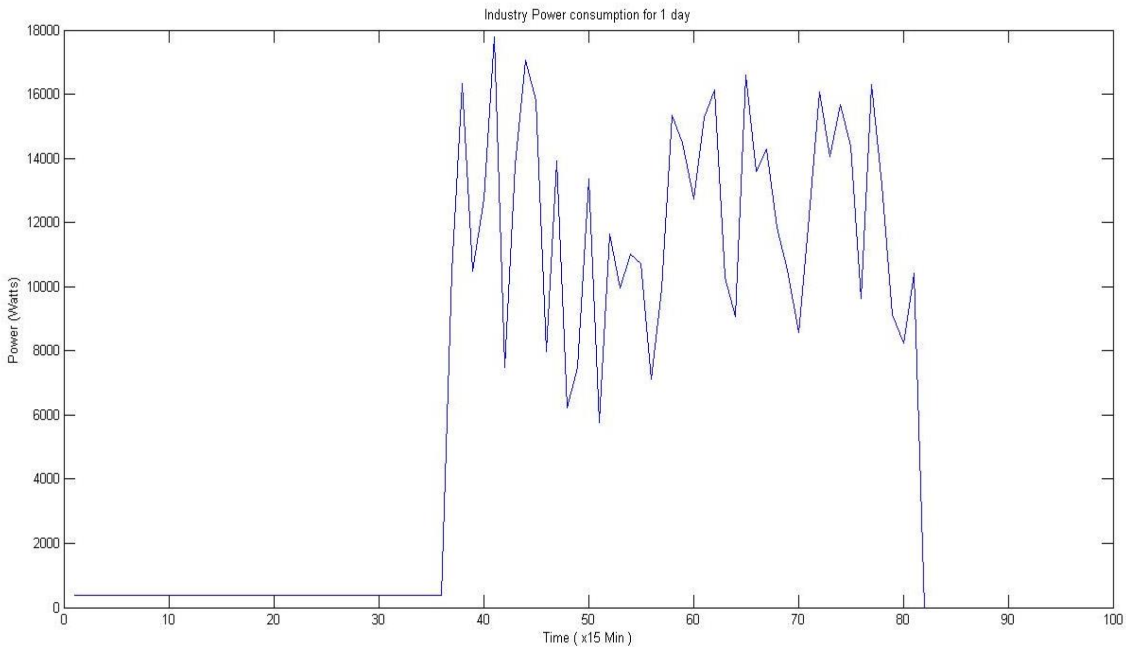

Figure 3.20 Large industry during a day

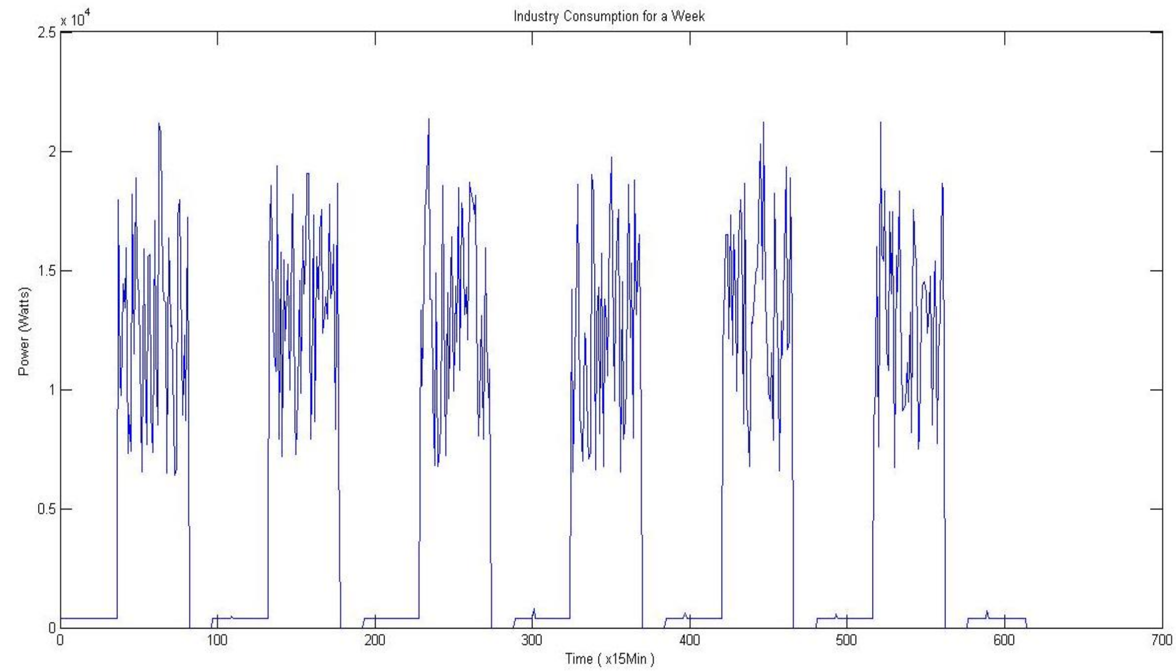

Figure 3.21 Large industry during a week

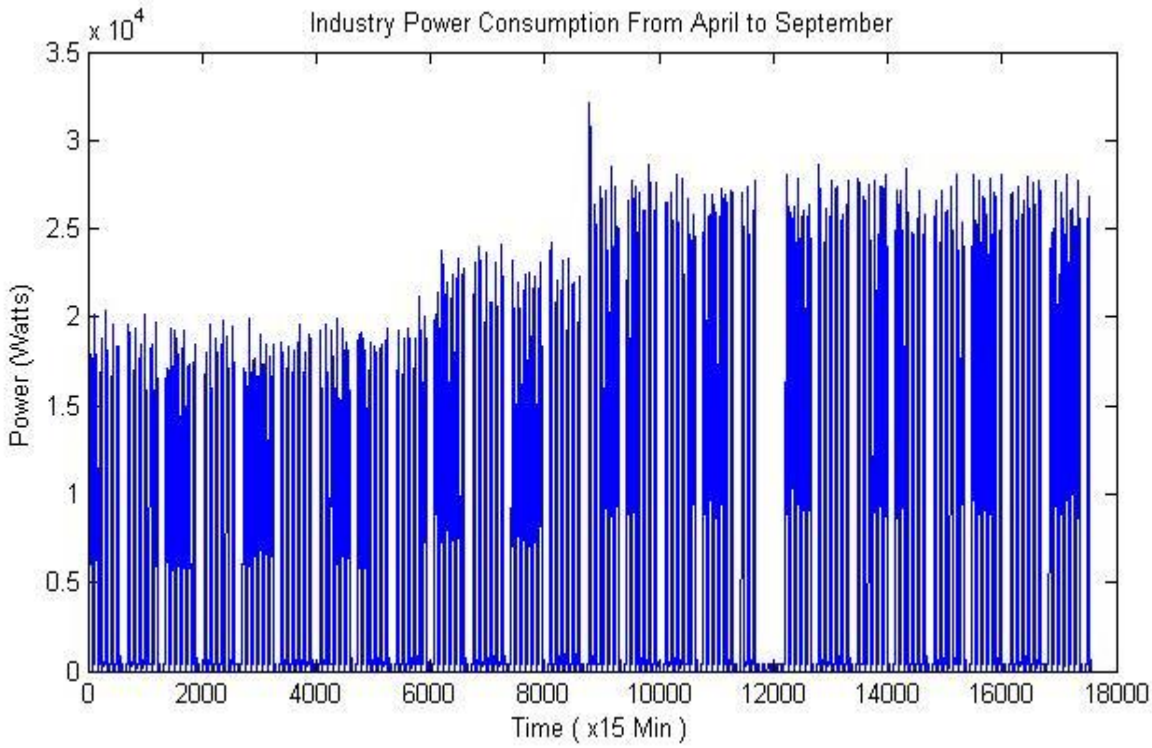

Figure 3.22 Large industry power consumption from 1st April to 30th

3.2 The PowerWorld Software

PowerWorld [26] simulator is an interactive power system simulation package designed to simulate high voltage power system operation on a time frame ranging from several minutes to several days. The software contains a highly effective power flow analysis package capable of efficiently solving systems of up to 250,000 buses.

3.2.1 Features and Components of PowerWorld

The functions and tools of the program offer infinite possibilities for in-depth research, analysis and visualization of complex and large-scale power systems. The most important features of Power World Simulator are as follows:

- Available Transfer Capability (ATC)

The Available Transfer Capability (ATC) add-on allows you to determine the maximum MW transfer possible between two parts of the power system without violating any limits. This is the same calculation commonly performed by system operators or market operators.

- Time Step Simulation Dialog

The Time Step Simulation dialog is used to control and visualize the time simulation. The top section of the form contains buttons for data input/output and buttons to control the progress of the simulation. The main part of the form has a number of Time Step Simulation pages that contain grids and options where input data can be specified and simulation results can be examined. The results of the Time Step Simulation are presented on the grids of the Results page. The Time Step Simulation allows you to specify what objects (buses, lines, generators, loads, etc.) and what object fields (bus voltage, bus LMP, gen MW, etc.) should be displayed on the results grids. This gives the user the flexibility needed to explore the relevant results, avoiding at the same time the problem of storing a massive amount of results, most of which may not be relevant. Storage is a critical aspect of the Time Step Simulation, since a set of results comparable to full a PF/OPF/SCOPF solution is generated for each timepoint.

- Distributed Computing

PowerWorld has introduced with Simulator version 15 distributed computing technology to take advantage of modern multi-processor computers and network resources to reduce the computation time of some Simulator processes. The technology is suited for processes that are easily broken into smaller, independent tasks that may be processed in parallel. Contingency Analysis, multi-scenario ATC, and multi-contingency Transient Stability Analysis are currently available.

- Geomagnetically Induced Current (GIC)

Evaluate the Risk Posed by Solar Storms with PowerWorld Simulator

PowerWorld has developed an innovative tool for analyzing the potential impact of geomagnetic disturbances (GMD), using our familiar power flow and transient stability platform. PowerWorld Simulator GIC may be the most accessible tool in the world for power system planning and operations engineers to readily assess GMD risk posed to their systems.

- Integrated Topology Processing

PowerWorld has released in its version 14.0 a new product that can reshape the power system operations and planning business processes of the electricity industry. For several decades the entire industry has lived and struggled with a two-model paradigm. The computer applications, models, and formats used in the operations and planning stages of the power industry are completely different and incompatible. This causes unnecessary complexity and innumerable problems ranging from data errors, to staff frustration, to ultimately a less secure grid. Attempts to bring these models together continue to avoid the cause of the problem requiring new formats, involved modeling, and cumbersome case conversions and mappings.

PowerWorld Integrated Topology Processing (ITP) is the only product in the industry that enables full unification of planning and operations at three levels: format, model and application environment. Based on revolutionary unified application architectures, ITP offers a simple, consistent and elegant solution to the two-model problem. ITP goes beyond matching the results of a single power flow to achieving full interoperability demonstrated by fully contingency analysis results of ISO planning and EMS tools.

- Optimal Power Flow Analysis Tool (OPF)

The PowerWorld Optimal Power Flow Analysis Tool (OPF) is an optional add-on to the base Simulator package. Simulator OPF starts with all the functionality of the original Simulator, but then adds an optimal power flow (OPF). Simulator OPF provides the ability to optimally dispatch the generation in an area or group of areas while simultaneously enforcing the transmission line and interface limits. Simulator OPF can then calculate the marginal price to supply electricity to a bus (locational marginal price or LMP), while taking into account transmission system congestion.

- Automation Server (SimAuto)

The Simulator Automation Server (SimAuto) allows you to take advantage of the power of automation to extend the functionality of PowerWorld Simulator to any external program that you write. Using Simulator Automation Server you can launch and control PowerWorld Simulator from within another application, enabling you to: access the data of a Simulator case, perform defined Simulator functions and other data manipulations, and send results back to your original application, to a Simulator auxiliary file, or to a Microsoft® Excel spreadsheet.

- Transient Stability

Powerful Dynamic Simulation with Simplicity only PowerWorld can Deliver

PowerWorld has developed with Simulator version 15 a game-changing Transient Stability application. The tool features Simulator's trademark user-friendly design, yet is as powerful as any on the market.

Some of the features that make it so revolutionary include:

- Voltage Stability Analysis: PowerWorld Simulator PVQV

PVQV is a Simulator add-on for analyzing a power system's static voltage stability margins. The information it provides can help the analyst or transmission planner determine how to strengthen the power system against the risk of voltage collapse. Simulator PVQV provides a full-featured voltage analysis tool within the easy-to-use, visual environment of PowerWorld Simulator.

3.2.2 Power Grid Model Development in PowerWorld Environment

For the purposes of this thesis we used the educational version of 40 scales that is available in the Power World website.

In the present study we modeled a village with 141 consumers which represent residential, commercial and industrial customers. More specifically our model consists of 3 industrial, 14 commercial and 124 residential customers.

In Figure 3.23 we can see that the industrial consumers and hotels are on the right and in the middle and on the left there exist both commercial and residential customers in lines 1, 2, 3 and 4. With the blue arrows are commercial customers and with black are residential.

The length of the lines (1, 2, 3, 4) are 250-300 meters as in reality. For the industrial and hotels we have separate transformers due to the increased requirement of power. Moreover at the secondary of the transformers in lines 1, 2, 3 and 4 we have installed sum meters to examine the total power that lines consumed. We did the same for the loads that have their own transformer like industrial customers and hotels.

Along a line of residential-commercial we can see that the number of customers are increasing as more and above customers are connected in reality. The generator is of 10 MW and the transformers are of 100 MVA. The slack generator in the slack bus is necessary for absorbing the additional power that generator gives or offers energy if the main generator power is not enough.

The characteristics of the distribution lines are of ACSR 50 R: 0.404 Ω/KM and X: 0.420 Ω/KM.

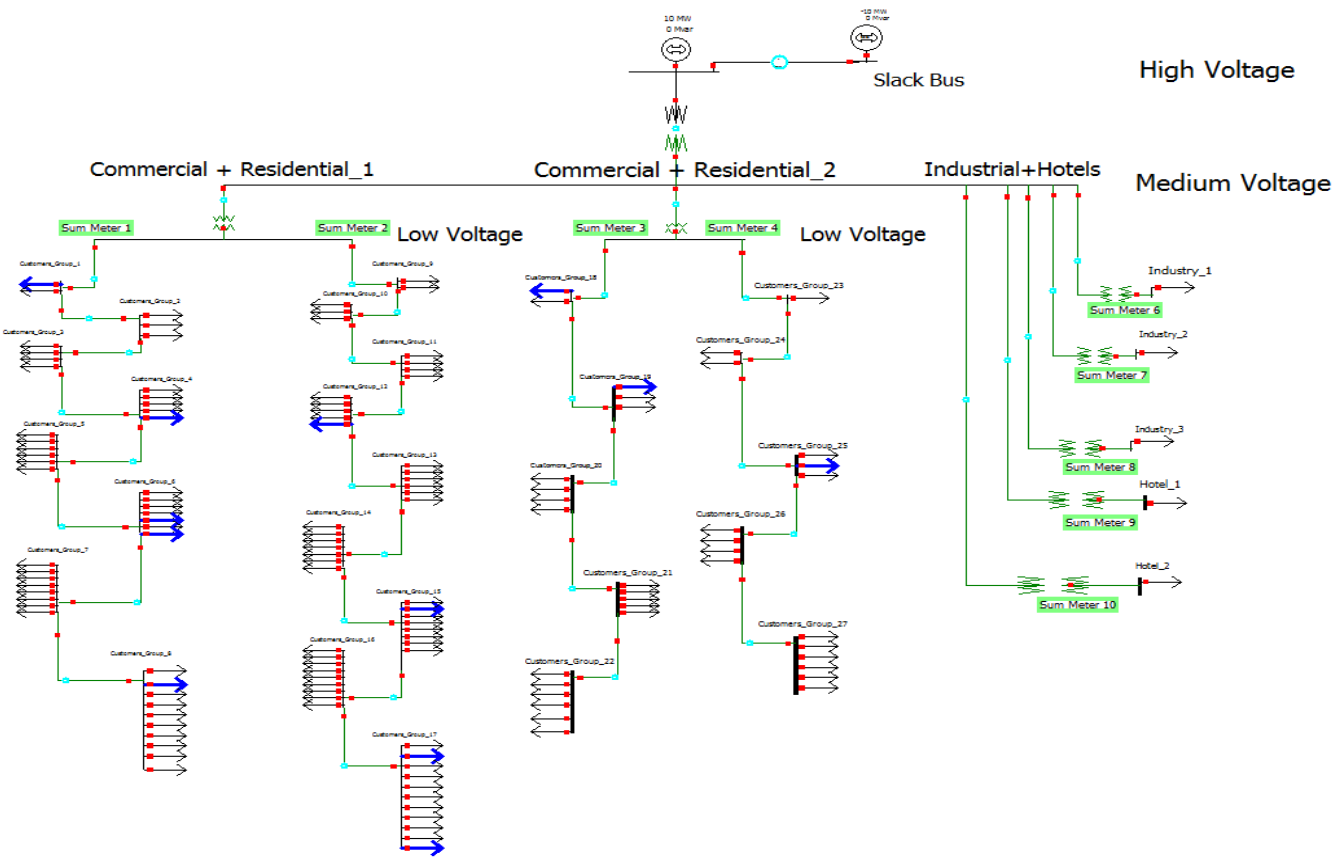

Figure 3.23 Power System Model in PowerWorld Environment

3.3 Data Mining Techniques

3.3.1 Principal Component Analysis (PCA)

[30],[19] Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables.

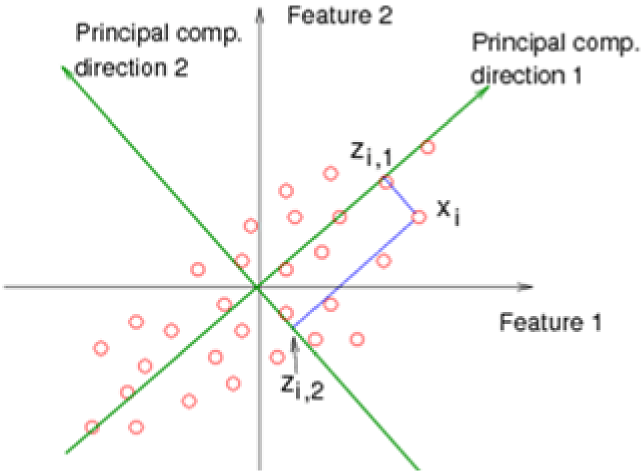

Figure 3.24

PCA [18] is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute. The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the component score).

PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a projection or "shadow" of this object when viewed from its (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

3.3.2 The Mean Shift Algorithm

[28], [29] Mean Shift is a powerful and versatile non parametric iterative algorithm which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. It can be used for lot of purposes like finding modes, clustering etc. Mean Shift was introduced by Fukunaga and Hostetler and has been extended to be applicable in other fields like Computer Vision.

Consider a set of points in two-dimensional space. Assume a circular window centered at C and having radius r as the kernel. Mean shift is a hill climbing algorithm which involves shifting this kernel iteratively to a higher density region until convergence. Every shift is defined by a mean shift vector. The mean shift vector always points toward the direction of the maximum increase in the density (Figure 3.25). At every iteration the kernel is shifted to the centroid or the mean of the points within it. The method of calculating this mean depends on the choice of the kernel. In this case if a Gaussian kernel is chosen instead of a flat kernel, then every point will first be assigned a weight which will decay exponentially as the distance from the kernel's center increases. At convergence, there will be no direction at which a shift can accommodate more points inside the kernel.

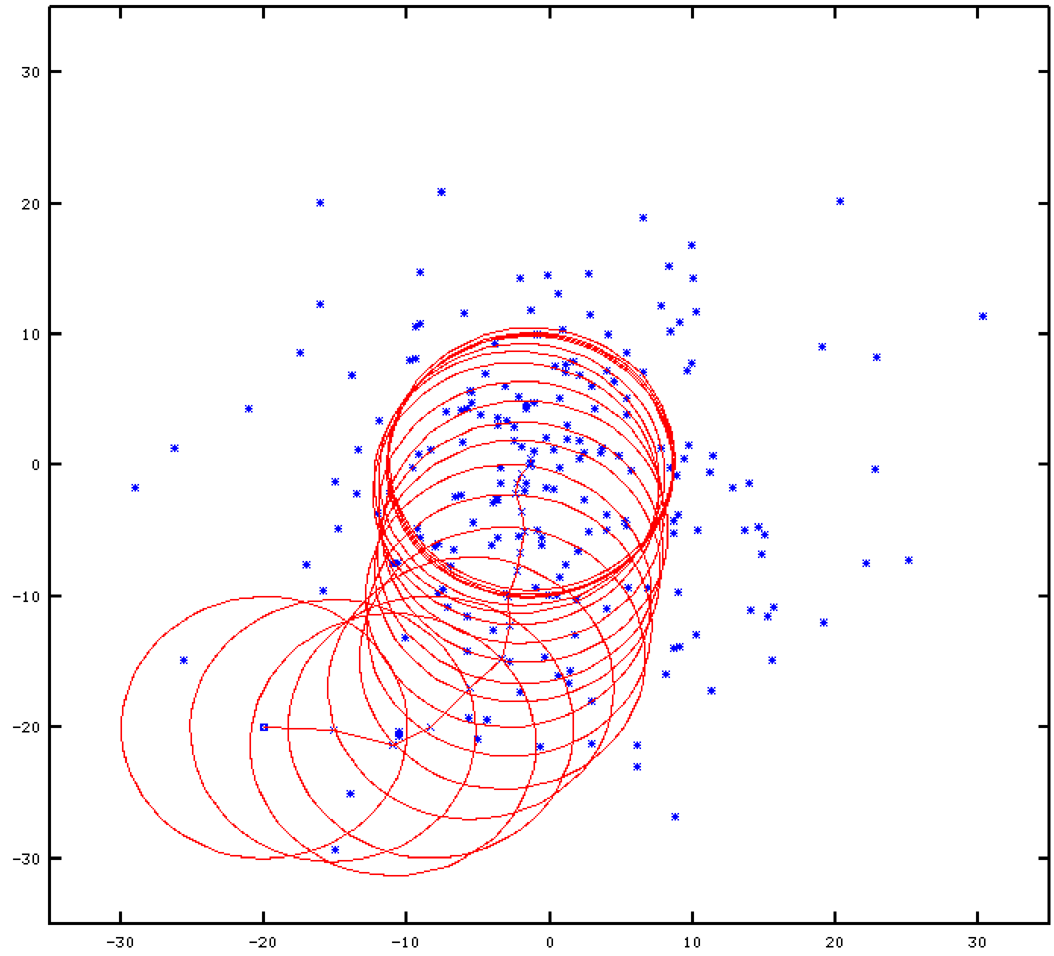

Figure 3.25

Let a kernel function $K(X_i - X)$ be given. This function determines the weight of nearby points for re-estimation of the mean. Typically a Gaussian kernel on the distance to the current estimate is used, $K(X_i - X) = e^{-c\|x_i - x\|^2}$. The weighted mean of the density in the window determined by K.

$$m(x) = \frac{\sum_{x_i EN(x)} K(x_i - x)x_i}{\sum_{x_i EN(x)} K(x_i - x)} \qquad (3.1)$$

where $N(x)$ is the neighborhood of X , a set of points for which $K(x) \neq 0$.

The difference $m(x) - x$ where $m(x)$ is the expression (3.1) is called *mean shift* .The Mean Shift algorithm now sets $x \leftarrow m(x)$, and repeats the estimation until $m(x)$ converges.

Mean shift pros:

- Does not assume any predefined shape on data clusters.
- The procedure only has one parameter, the bandwidth.
- Output doesn't depend on initializations.

Mean shift cons:

- Output does depend on bandwidth: too small and convergence is slow, too large and some clusters may be missed.
- Inappropriate window size can cause modes to be merged, or generate additional "shallow" modes.
- Computationally expensive for large feature spaces.

4 Model Development and Experimental Results

Having constructed the energy meters database with Matlab we have at our disposal 50 different residential patterns for every scenario, 10 different patterns for every industry and 10 different patterns for every business.

4.1 Applying Scenarios to Customers

Extracting scenarios from Matlab to excel and after introducing them from excel to PowerWorld we introduced a different pattern for every customer in the grid. Especially for residential customers the probabilities of scenarios are shown in Figure 4.1. We chose for scenarios 1 to 4 a bigger probability as these scenarios represent the largest part of residential population. For commercial and industrial we chose to introduce 10 different patterns with uniform probability.

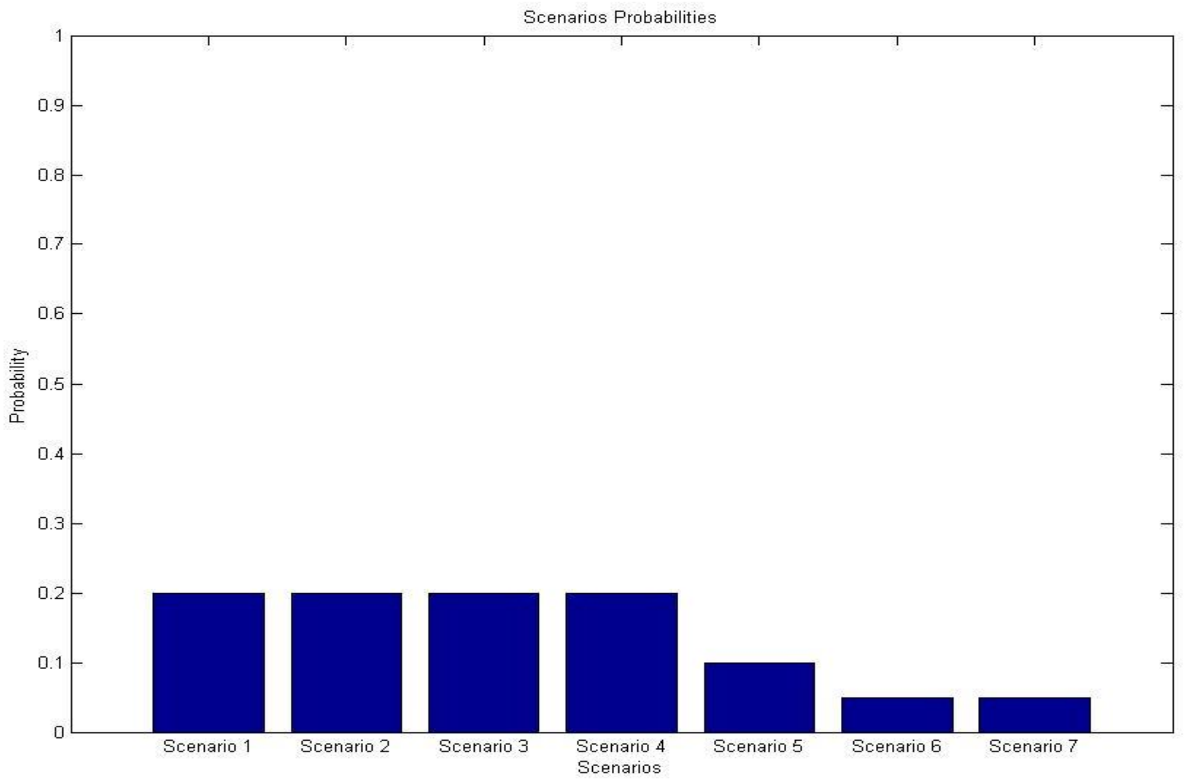

Figure 4.1

In Figure 4.2 we have introduced the smart meter readings using PowerWorld Time Step Simulator. As Time step simulator works with Mw we have converted our readings from W to Mw as shown in the Figure below. Although PowerWorld is a tool for modeling a transmission grid it works appropriately for distribution grid.

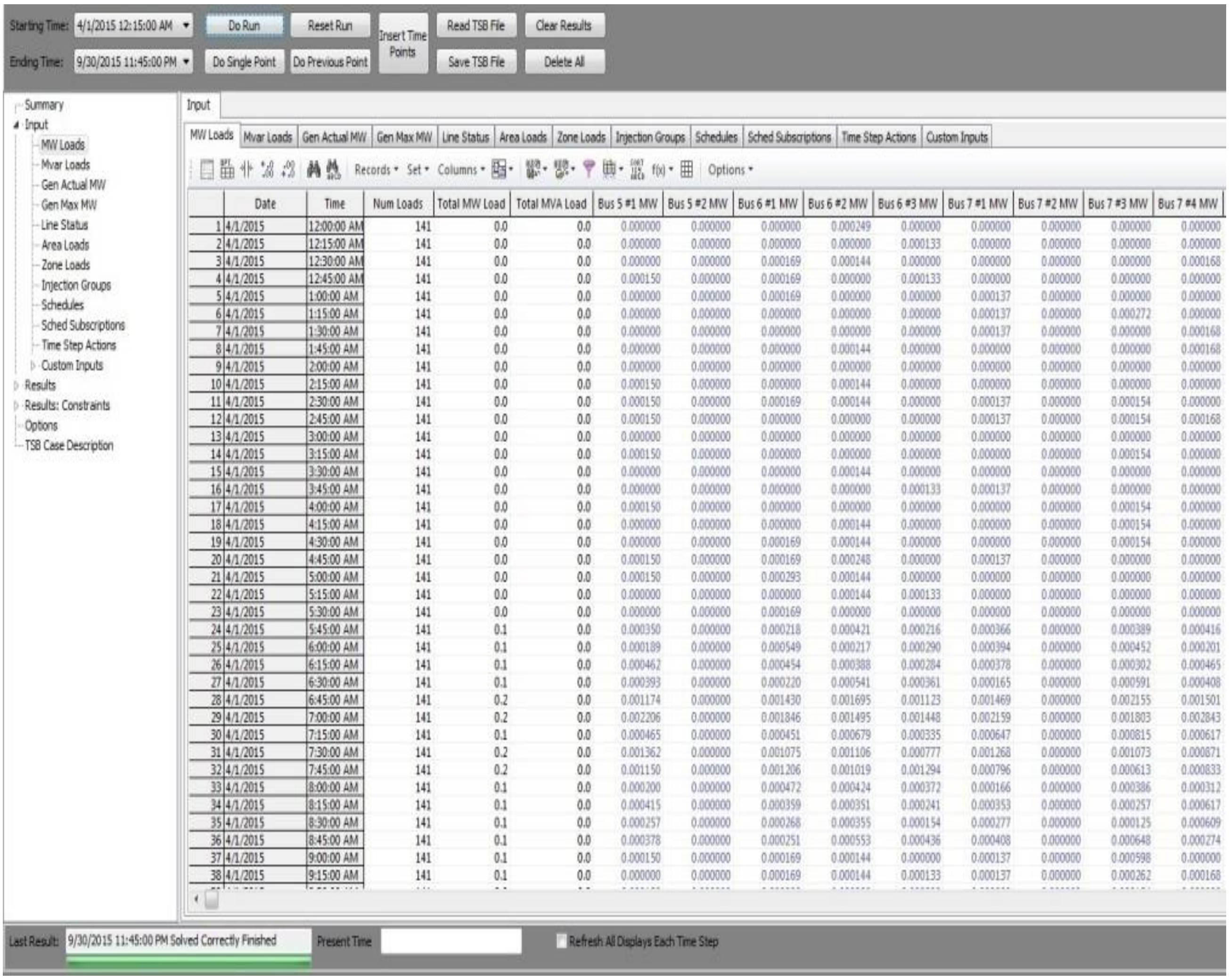

Figure 4.2 Time step Simulator

Afterwards, running the time step simulator from 1st of April to 30th of September we chose to receive results about the total energy consumed from every line and for every quarter and the power losses from each line and for every quarter respectively as shown in figure 4.3.

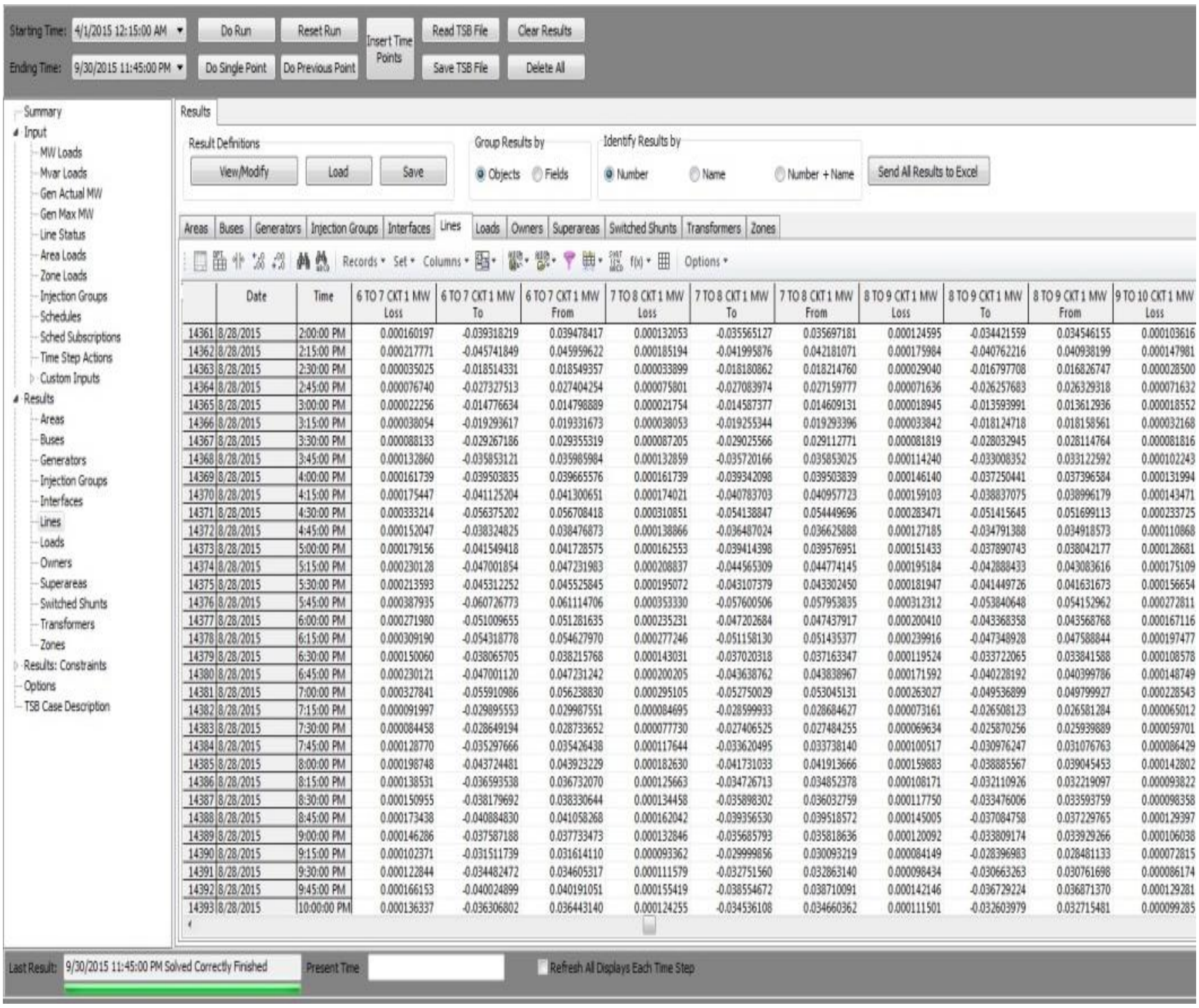

Figure 4.3 Time Step Simulator results

Subsequently we transferred the results to excel (Figure 4.4) so as to import them in Matlab for further processing.

2	Date	Time	Num Loads	Total MW Load	Total MVA Loac	Bus 5 #1 MW	Bus 5 #2 MW	Bus 6 #1 MW	Bus 6 #2 MW	Bus 6 #3 MW	Bus 7 #1 MW	Bus 7 #2 MW
3	4/1/2015	12:00:00 AM	141	0	0	0	0	0	0.000248737	0	0	0
4	4/1/2015	12:15:00 AM	141	0	0	0	0	0	0	0.000133336	0	0
5	4/1/2015	12:30:00 AM	141	0	0	0	0	0.00016925	0.000144421	0	0	0
6	4/1/2015	12:45:00 AM	141	0	0	0.000150478	0	0.00016925	0	0.000133336	0	0
7	4/1/2015	1:00:00 AM	141	0	0	0	0	0.00016925	0	0	0.000136691	0
8	4/1/2015	1:15:00 AM	141	0	0	0	0	0	0	0	0.000136691	0
9	4/1/2015	1:30:00 AM	141	0	0	0	0	0	0	0	0.000136691	0
10	4/1/2015	1:45:00 AM	141	0	0	0	0	0	0.000144421	0	0	0
11	4/1/2015	2:00:00 AM	141	0	0	0	0	0	0	0	0	0
12	4/1/2015	2:15:00 AM	141	0	0	0.000150478	0	0	0.000144421	0	0	0
13	4/1/2015	2:30:00 AM	141	0	0	0.000150478	0	0.00016925	0.000144421	0	0.000136691	0
14	4/1/2015	2:45:00 AM	141	0	0	0.000150478	0	0	0	0	0.000136691	0
15	4/1/2015	3:00:00 AM	141	0	0	0	0	0	0	0	0	0
16	4/1/2015	3:15:00 AM	141	0	0	0.000150478	0	0	0	0	0	0
17	4/1/2015	3:30:00 AM	141	0	0	0	0	0	0.000144421	0	0	0
18	4/1/2015	3:45:00 AM	141	0	0	0	0	0	0	0.000133336	0.000136691	0
19	4/1/2015	4:00:00 AM	141	0	0	0.000150478	0	0	0	0	0	0
20	4/1/2015	4:15:00 AM	141	0	0	0	0	0	0.000144421	0	0	0
21	4/1/2015	4:30:00 AM	141	0	0	0	0	0.00016925	0.000144421	0	0	0
22	4/1/2015	4:45:00 AM	141	0	0	0.000150478	0	0.00016925	0.000248457	0	0.000136691	0
23	4/1/2015	5:00:00 AM	141	0	0	0.000150478	0	0.000292946	0.000144421	0	0	0
24	4/1/2015	5:15:00 AM	141	0	0	0	0	0	0.000144421	0.000133336	0	0
25	4/1/2015	5:30:00 AM	141	0	0	0	0	0.00016925	0	0	0	0
[illegible]	[illegible]	[illegible]	141	0	0	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	0

Figure 4.4 Results in Excel

4.2 Power Theft Scenarios

Power Theft scenarios
1. Consumer with smart meter and steeling electricity fully due to power pass before smart meter.
2. Consumer with smart meter and steeling electricity partially due to power pass before smart meter.
3. Consumer with no smart meter is connected to power grid.
4. Consumer with smart meter and increased consumption of electricity due to illegal activity or power delivery to unauthorized building.

Table 4.1 Power Theft Scenarios

For our experiments we consider 2 consumers with scenario 1, 4 consumers with scenario 2, 3 consumers with scenario 3 and 2 consumers with scenario 4 as shown in Figure 4.5, Figure 4.6 and Figure 4.7.

Bus Number	ID	Power Theft Scenario
17	3	Scenario 1
21	B	Scenario 2 : 50% power theft
21	J	Scenario 3

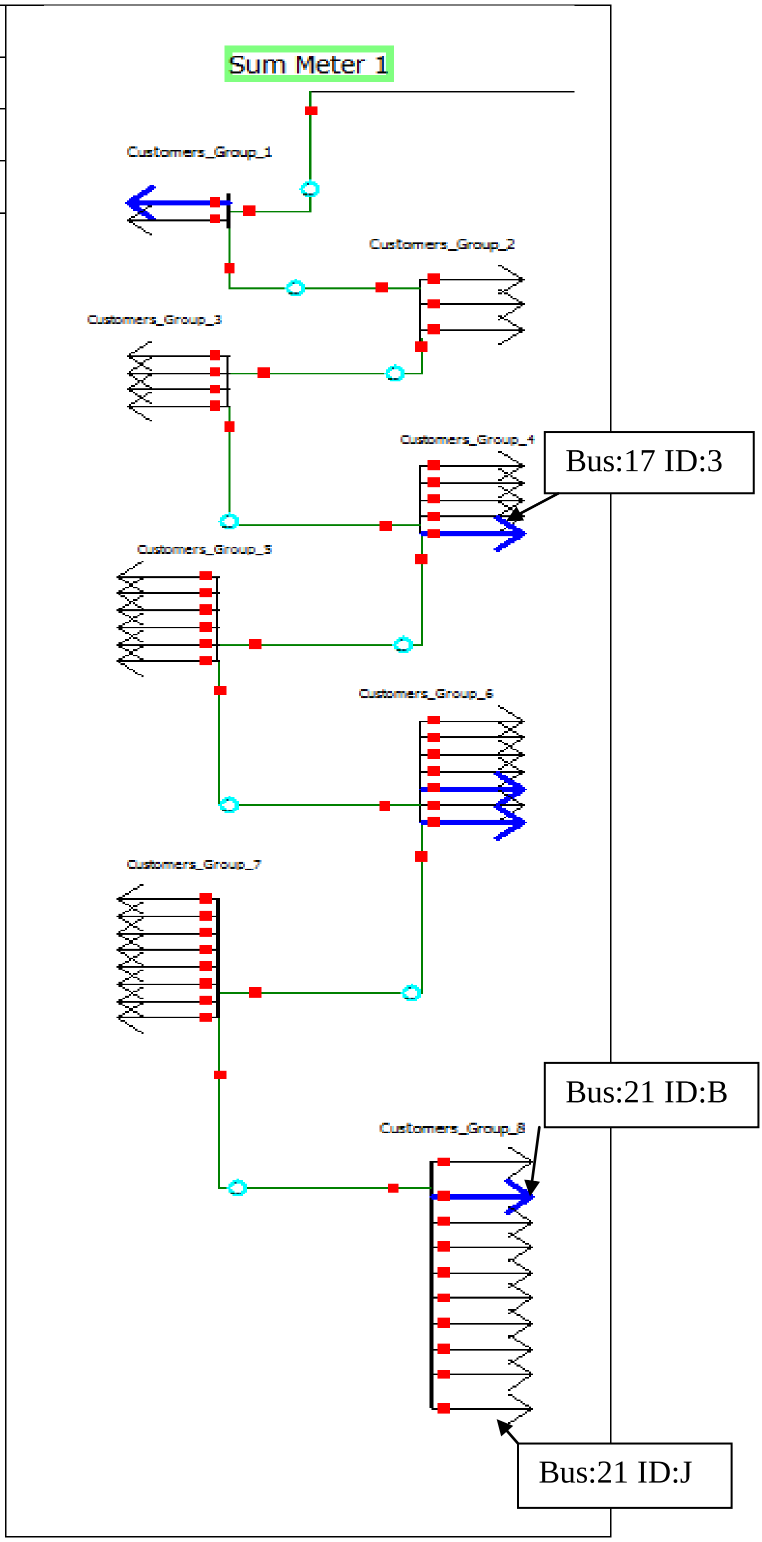

Figure 4.5 Line 1

Bus Number	ID	Power Theft Scenario
7	3	Scenario 4 : 50% overload
12	B	Scenario 2 : 60% power theft
13	B	Scenario 3
6	3	Scenario 1
9	5	Scenario 2 : 70% power theft

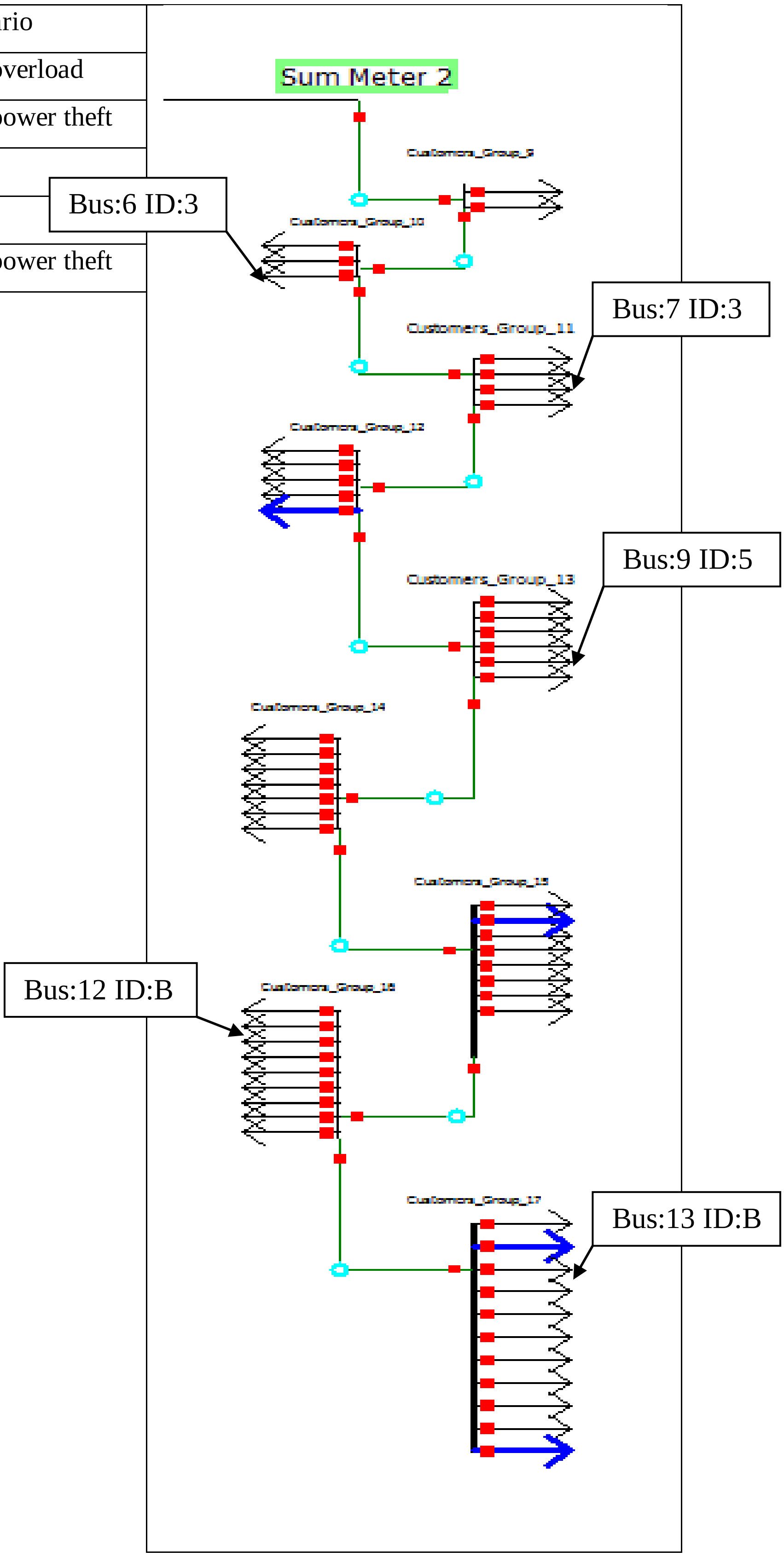

Figure 4.6 Line 2

Bus Number	ID	Power Theft Scenario
24	3	Scenario 4 : 70% overload
27	1	Scenario 2 : 80% power theft
27	6	Scenario 3

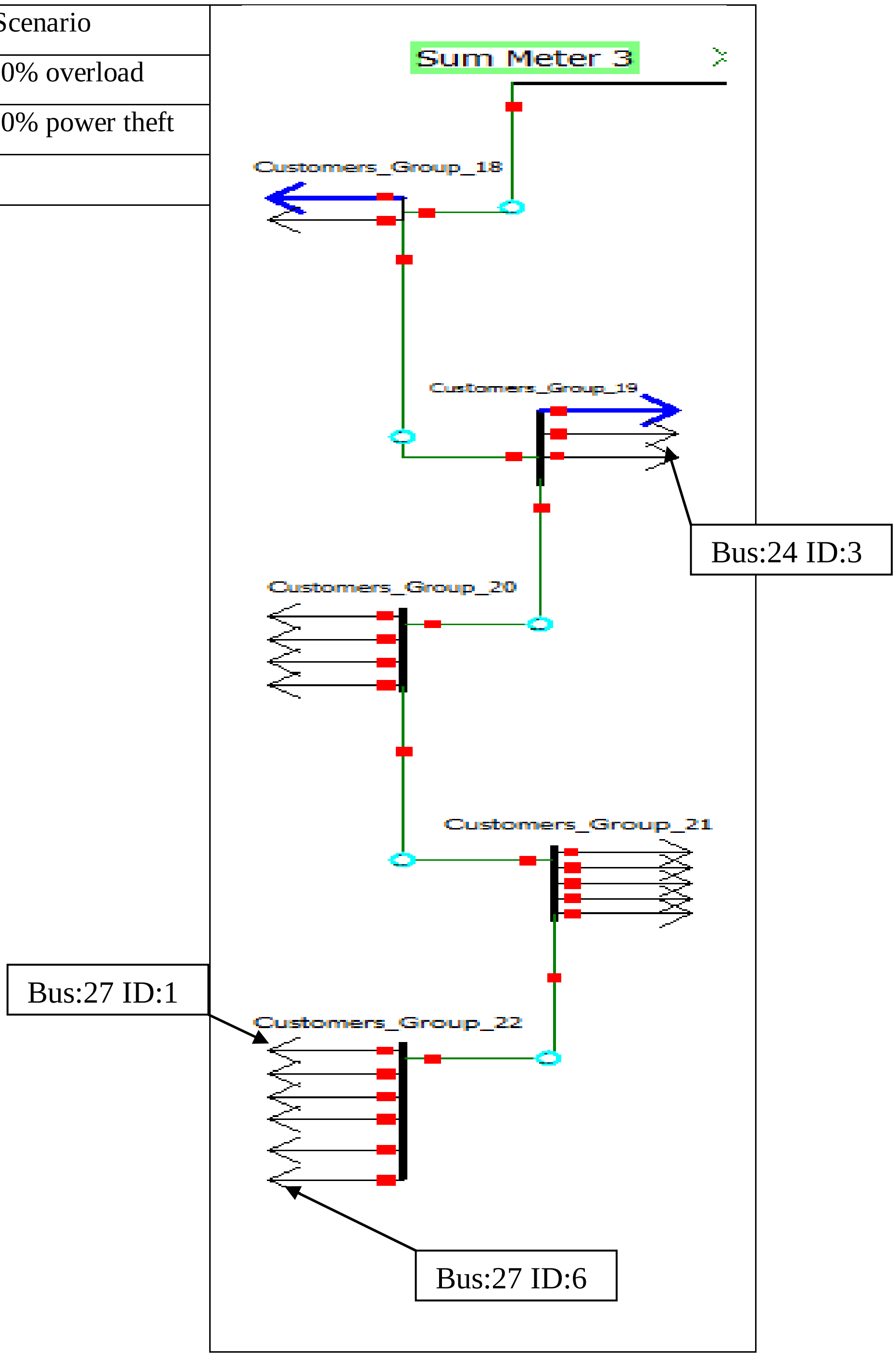

Figure 4.7

In Figure 4.8, Figure 4.9 and Figure 4.10 we see the total sum meters readings, the total losses (technical, non technical),technical losses and non technical losses.

For the technical losses we can apply the methods that refer to [27].

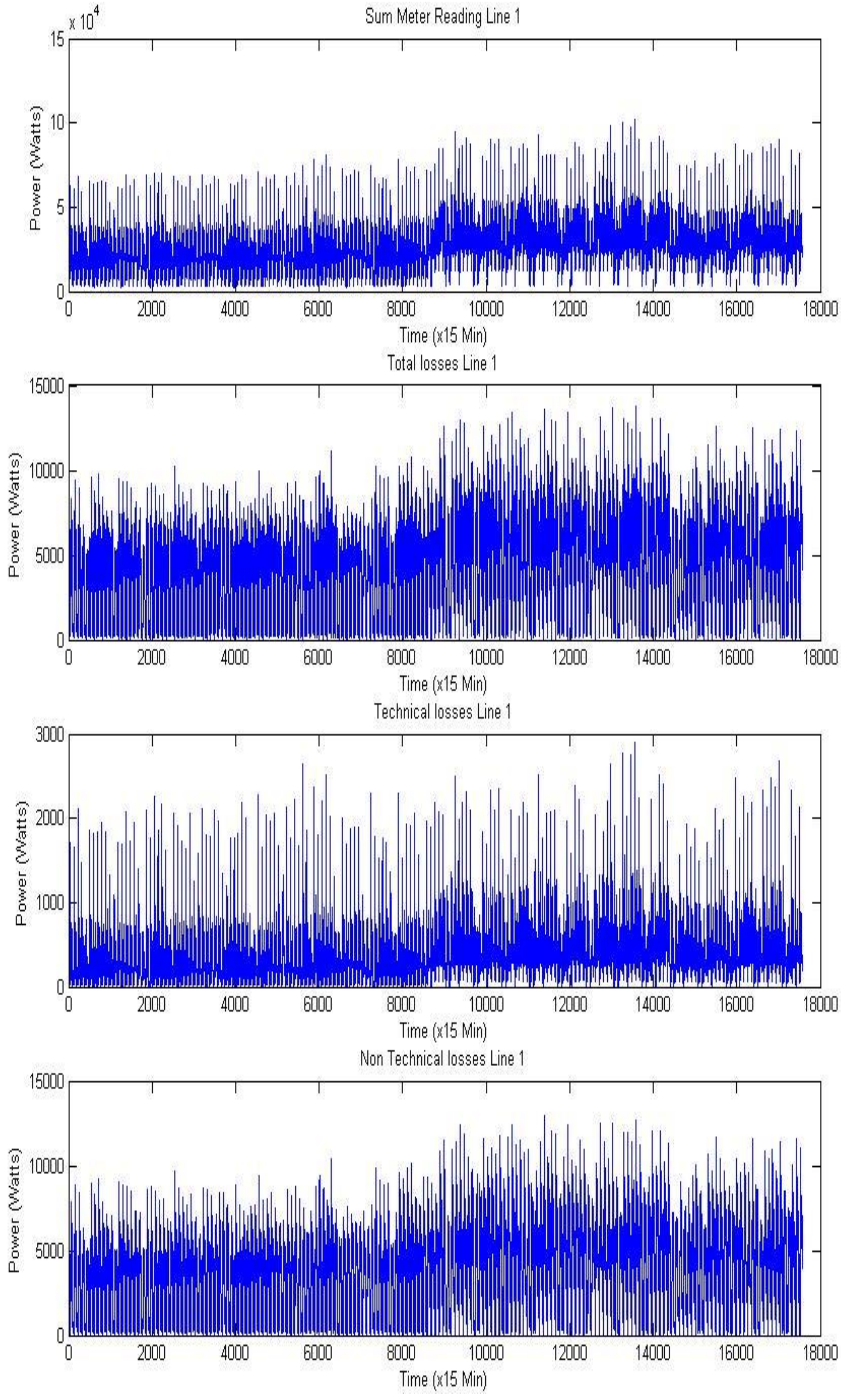
Sum Meter Reading Line 1
x 10^4
Power (Watts)
Time (x15 Min)
Total losses Line 1
Power (Watts)
Time (x15 Min)
Technical losses Line 1
Power (Watts)
Time (x15 Min)
Non Technical losses Line 1
Power (Watts)
Time (x15 Min)

Figure 4.8

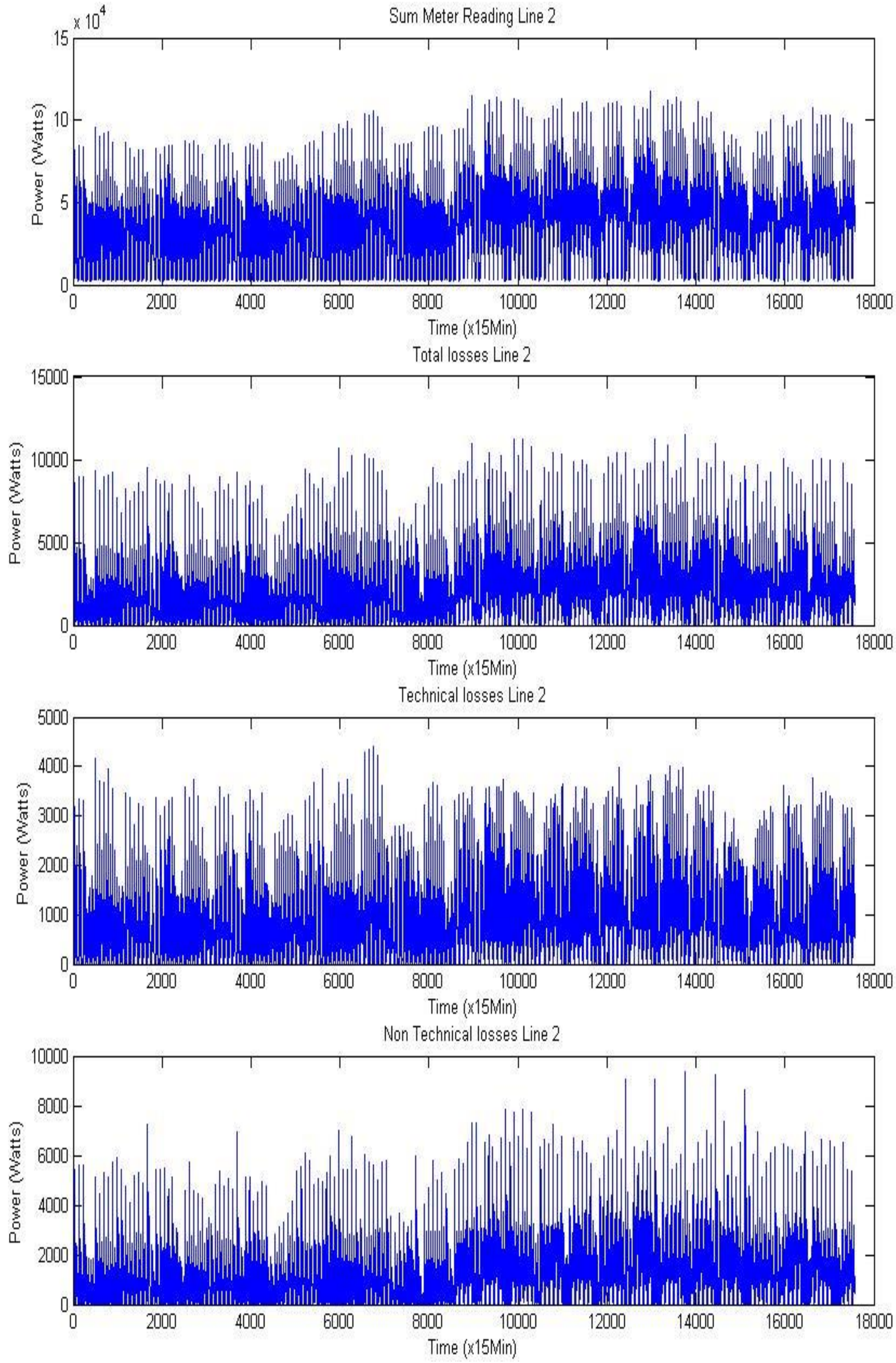
Sum Meter Reading Line 2
x 10^4
Power (Watts)
Time (x15Min)
Total losses Line 2
Power (Watts)
Time (x15Min)
Technical losses Line 2
Power (Watts)
Time (x15Min)
Non Technical losses Line 2
Power (Watts)
Time (x15Min)

Figure 4.9

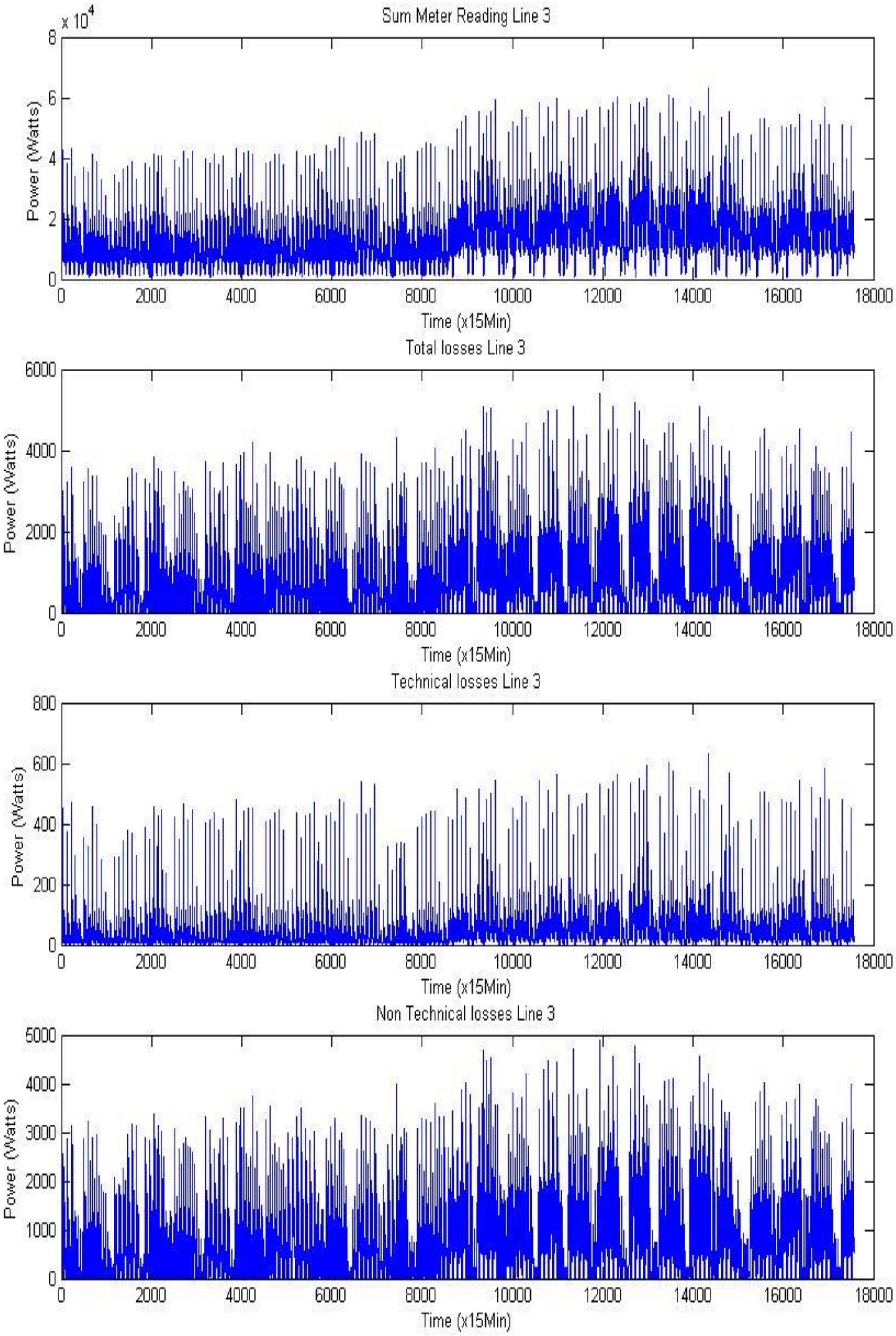
Sum Meter Reading Line 3
x 10^4
Power (Watts)
Time (x15Min)
Total losses Line 3
Power (Watts)
Time (x15Min)
Technical losses Line 3
Power (Watts)
Time (x15Min)
Non Technical losses Line 3
Power (Watts)
Time (x15Min)

Figure 4.10

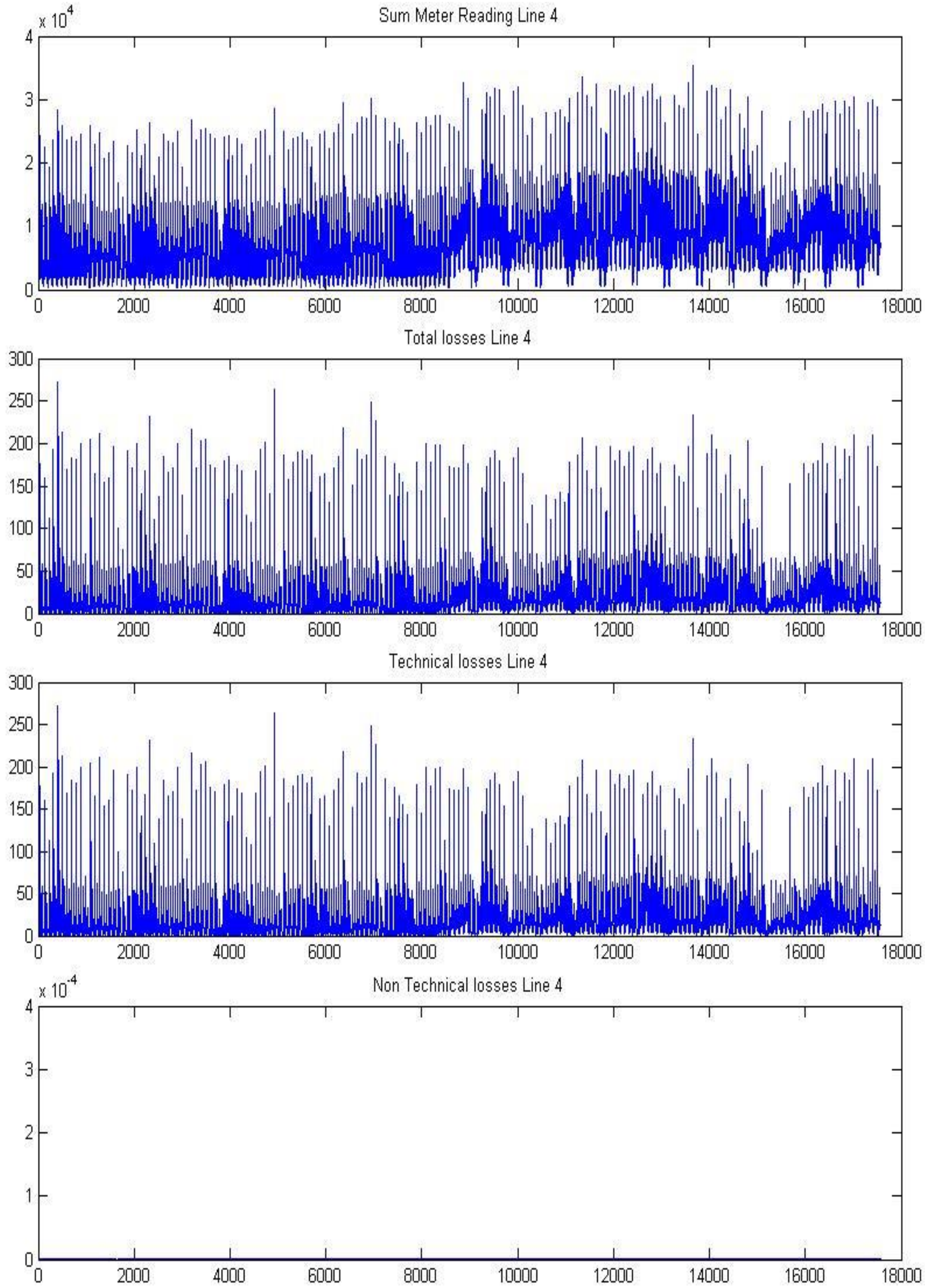

Figure 4.11

As we can observe from non technical losses diagramms we have problem in lines 1,2,3.

4.3 Customers Selection

Next step is to separate commercial residential and industrial customers.From the residential customers we throw out the customer with scenario 6, scenario 7 and those with zero consumption.

4.4 Feature Selection and Extraction

For residential consumers we keep only weekdays and we throw out weekends due to the different habits that residents have especially on weekends. In this way we enhance the clustering procedure.

4.5 Mean Normalization

Subsequently, before applying PCA its standard practice to first perform mean normalization at feature scaling so that the features has zero mean and should have comparable range of values due to the expression (4.1)

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} X_j^{(i)} \qquad (4.1)$$

We replace each $X_j^{(i)}$ with $X_{(j)} - \mu_{(j)}$.

4.6 Applying PCA and Mean Shift Algorithm

Having a matrix with residentials customers we apply pca to our data so as to minimize the dimension of the data. Applying pca we choce the appropriate number of principal components so as to keep the data variance to 95 % due to the expresion 4.2.

$$\frac{\frac{1}{m}\sum_{i=1}^{m} \| x^{(i)} - x_{approx}^{(i)} \|^2}{\frac{1}{m}\sum_{i=1}^{m} \| x^{(i)} \|^2} \leq 0.05 \qquad (4.2)$$

Then we apply mean shift algorithm by chosing the appropriate bandwith to our data.

4.7 Experimental Results

In Table 4.2 we can see on the right the clusters (modes). The numbers 64, 29, 5 corresponds to the first three inspected loads on the left column of table 4.2. The first three inspected loads are the loads with power theft scenario 2. The numbers 73 13 corresponds to the last two inspected loads. These loads are those with power theft scenario 4.

<table>
<tr><td>Inspected loads</td><td rowspan="6">results =

[1x22 double]
[1x14 double]
[1x21 double]
[1x22 double]
[5]
[29]
[64]
[73]
[13]</td></tr>
<tr><td>Bus 12#B</td></tr>
<tr><td>Bus 27 #1</td></tr>
<tr><td>Bus 7 #3</td></tr>
<tr><td>Bus 24 #3</td></tr>
<tr><td>Bus 9#5</td></tr>
</table>

Table 4.2 Results

Experimenting by changing the number of power theft scenarios we have reached the conclusion for residential customers with partial power theft up to 50% and for residential customers with up to 40 % percent overload that we have 100% detection as shown in Figure.4.12 and Figure 4.13.

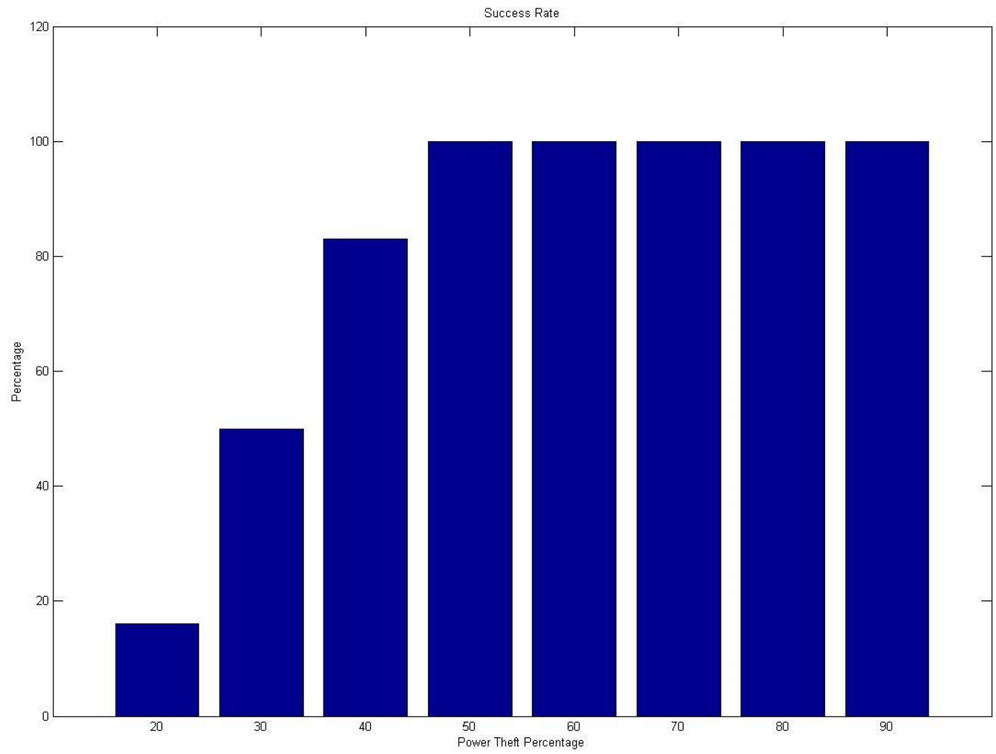

Figure 4.12

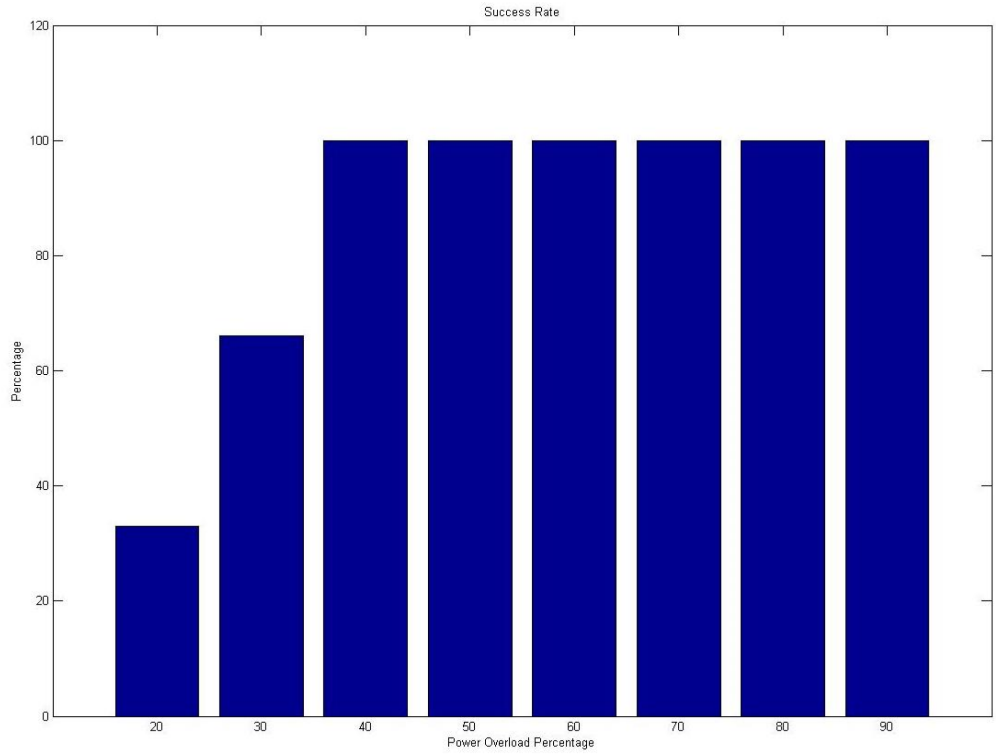

Figure 4.13

For the loads with zero consumption we check the permanent house of the residence. If the house that we examine is a permanent residence then there is a strong indication for power theft.

For scenario 6 and scenario 7 we consider that there is no incentive for consumers to steal electricity but we cannot exclude power theft so we can check if the house is a permanent residence.

For commercial customers, due to bussinesses is different from each other we propose to check them seperatery for those which are in problematic lines.

The method we follow for industrial customers and for those businesses that have their own transformer is quite simple. Due to transformers that every industry or business have we make a comparison between the sum meter in the secondary of the transformer and the indication in industry smart meter. If a discrepancy exists then the industry or business is suspected for power theft. Power losses are almost zero as the transformer is close to the smart meter of the industry.

Having checked a problematic line we can estimate the power theft percentage of illegal customes and subsequently we can compare them with total non technical losses in the line.If a discrepancy exists then there are loads in the line with no smart meter.

5 Conclusion and Further Research

Power theft is widespread phenomenon in developing countries. The transition toward future smart networks should enhance the insight into the distribution networks, increase their reliability, flexibility and reduce the transmission and distribution losses.

Energy-theft detection is a classic and difficult problem in power grid. With the development of advanced metering infrastructure in smart grid, more complicated situation in energy theft has emerged and many new technologies are adopted to try to solve this problem. The main contribution of this dissertation is the novel methodology proposal for automated detection of illegal use of electricity in the low voltage distribution networks. Smart meters are essential for this method. The theft detection should be considered as one of the important aspects of future distribution networks, as well.

In conclusion, the overall research has shown encouraging results for partial power theft up to 50% and for power overload up to 40 %. This dissertation illustrates various cases, issues and setbacks in the design, development, deployment, operation, and maintenance of electricity theft controlling devices. In addition, various factors that influence people to steal electricity are discussed. This study also illustrates the effect of NTL on quality of supply.

5.1 Further Research

- [23] Detection algorithms can be enhanced by introducing more real-world parameters and variables. Real world parameters and variables can be the economic situation of a person or illegal activity in the past.
- [23] Take into account the impact of Time Base Pricing and Distributed Generation on customer energy consumption patterns. With distributed generation the analyses of the system become more complex and new methods of power theft will arise.
- [23] Illegal consumption of electricity may be controlled by focusing on cyber security (designing stronger firewalls or enhancing firmware in view of cyber security). For example, complicating the process of meter tampering and reducing the hackings of smart meters through the establishment of certain standards.
- We can construct new scenarios such us: a commercial customer who has low tariff is offering electric power to a residence with higher tariff, customer who steel electricity from a public organization such us a church, a customer with a net metering system

maybe present false supply from distributed generation in order to sell more energy into the grid.

- Economically examine the scenario to place more than one sum meters in a line so as to localize more accurately the illegal consumers.
- Localization of customer who has not smart meter and steel electricity from the power line directly.
- Definition of the most dangerous groups for power theft.

Bibliography

[1] Bandim, C.J, Alves, J.E.R, pinro , Souza, F C,Loureiro, M.R.B, Magalhaes, C.A, Galvez, D.F, 2003.Identofication of energy theft and tampered meters using a central observer meter: a mathematical approach.In :proceedings of the IEEE PES Transmission and Distribution Conference and Exposition, Rio de Janeiro,Brazil,September,pp. 163-168.

[2] Anand,R,D,Naveen,S.A., 2003. Design and development of vigilant energy metering system (VEMS) and its applications.In : Proceedings of the Student Conference on Research and Development,New Delhi,India,August, pp 15-18.

[3] Nagi, J., Yap, K.S.,Tiong, S.K, Ahmed, S.K, Mohammad, A.M., 2009. Detection of abnormalities and electricity theft using genetic support vector machines. In: Proceedings of the IEEE Region 10 Conference TENCON, Hyderabad, India,January, pp. 1-6.

[4] Nizar,A.H, Dong, Z.Y., 2009. Identification and detection of electricity customer behavior irregularities. In : Proceedings of the IEEE/PES Power Systems Conference and Exposition, Seattle, Washington, March, pp. 1-10.

[5] Pasdar,A.,Mirzakuchaki,S.A.,2007. A solution to remote detecting of illegal electricity usage based on smart metering. Proceedings of the International Workshop on Soft Computing Applications,Oradea,Romania,August,pp 163-167.

[6] Zeng, X.,Jin,G.,Jin,W.,Xu,Y.,2009. Anti theft and monitoring system of street lamp power cables. Proceedings of the Asia-Pacific Power and Energy Engineering Conference,Wuhan,China,March,pp 1-4.

[7] Jamil,M.,Munir ,F.,Khan,A.A.,Mirza,A.,2004. Telemetering and billing system for spatially distributed electrical power clients. Proceedings of the E-Tech 2004,Karachi,Pakistan,July,pp35-40

[8] De, S, Anand, R, Naveen, A,Moinuddin,S.,2003. E-metering solution for checking energy thefts and streamlining revenue collection in India. Proceedings of the IEEE PES Transmision and Distribution Conference and Exposition, Dallas Texas, September, pp 654-658.

[9] Mano, R., Cespedes, R., Maia, D.,2004. Protecting revenue in the distribution industry: a new approach with the revenue assurance and audit process. In: Proceedings of the IEEE/PES

Transmission and Distribution Conference and Exposition, Latin America, Kema, Brazil, November, pp218-223.

[10] Perez , E.H, NKanka, B.N., Ngulumingi, C.V., Gimeno, A., Kazadi, A.B., 2005. Analysis of technical losses in distribution networks of large cities in underdeveloped African countries. Proceedings of the IEEE Power Engineering Society Inaugural Conference and Exposition in Africa, Durban,South Africa, July, pp. 92-96.

[11] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, Vijay Devabhaktuni. Electricity theft : Overview, issues, prevention and a smart meter based approach to control theft.

[12] Solomon Nunoo, Joseph c.Attachie. A methodology for the design of an electricity theft monitoring system.

[13] Rong Jiang, Rongxing Lu, Ye Wang, Jun Luo, Changxiang Shen, and Xuemin (Sherman) Shen. Energy-Theft Detection Issues for Advanced Metering Infrastructure in Smart Grid.PP 105-120.

[14] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, and Vijay Devabhaktumi EECS Department.Enhanced Encoding Technique for Identifying Abnormal Energy Usage Pattern.

[15] P.Kadurek, Student member, IEEE, J Blom, J.F.G Cobben, W.L. Kling,Member, IEEE. Theft detection and smart metering practices and expectations in the Netherlands.

[16] Shih-Che Huang, Yuan-Liang Lo, and Chan-Nan Lu,Fellow, IEEE. Non-Technical loss Detection Using State Estimation and Analysis of Variance.

[17] Ministry of Internal Affairs and Communication of Japan, ‘Survey on time use and leisure activities’, 2006.

[18] Jon Shlens, A TUTORIAL ON PRINCIPAL COMPONENT ANALYSIS.

[20] https://en.wikipedia.org/wiki/Principal_component_analysis.

[21] Αναστασία Ματωνάκη, Κωνσταντίνος Καλυμνιός. Διπλωματική εργασία με τίτλο: Μεθοδολογία δημιουργίας τυπικών διαγραμμάτων φορτίου για οικιακούς καταναλωτές.

[22] M. Anas, N. Javaid, A. Mahmood, S. M. Raza, U. Qasim, Z. A. Khan. Minimizing Electricity Theft using Smart Meters in AMI.

[23] Soma Shekara Sreenadh Reddy Depuru Dissertation entitled: Modeling, Detection, and Prevention of Electricity Theft for Enhanced Performance and Security of Power Grid.

[24] JAWAD NAGI Dissertaton entitled: AN INTELLIGENT SYSTEM FOR DETECTION OF NON-TECHNICAL LOSSES IN TENAGA NASIONAL BERHAD (TNB) MALAYSIA LOW VOLTAGE DISTRIBUTION NETWORK.

[25] Θαλασσινού Μαριάντζελα Dissertation entitled: Smart Meters.

[26] http://www.powerworld.com/

[27] Maxime Guymard Master thesis entitled : Modeling of Technical Losses in the Senegalese Transmission and Distribution Grids and Determination of Non-technical Losses.

[28] https://en.wikipedia.org/wiki/Mean_shift

[29] Cheng, Yizong (August 1995). "Mean Shift, Mode Seeking, and Clustering".IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE) 17 (8): 790–799.

[30] https://en.wikipedia.org/wiki/Principal_component_analysis