Masters Program in Geospatial Technologies

Analysis and Visualization of Energy use for University Campus

Nourhan Khalifa Abdelkareem

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

UNIVERSIDADE NOVA ISEGI
UNIVERSITAT JAUME·I
Westfälische Wilhelms-Universität Münster

Analysis and Visualization of Energy use for University Campus

Dissertation supervised by

Sven Casteleyn, PhD

Professor, Institute of New Imaging Technologies,
Universitat Jaume I,
Castellón, Spain

Dissertation co-supervised by

Sara Ribeiro, MSc

Researcher, NOVA Information Management School,
Universidade Nova de Lisboa,
Lisbon, Portugal.

Dissertation co-supervised by

Jim Jones, MSc

Research Associate, Institute for Geoinformatics (IFGI),
Westfälische Wilhelms-Universität,
Muenster, Germany.

March 2015

To my deceased Father

ACKNOWLEDGMENTS

First of all, my ultimate gratitude goes to Almighty Allah.

I would like to express my deepest gratitude to my supervisor Dr. Sven Casteleyn for all his support, prompt replies, and valuable discussions. Also, I would like to extend my gratitude to my supervisors Jim Jones and Sara Ribeiro for their vital comments, ideas, help and corrections. And for their encouragement for me all the time.

Also I would like to thank Prof. Jorge Matue for his expert advice and great help during my thesis. Also the staff of the Geotech research group especially Dr. Joaquin Torres, Ana Sanches and Joan Avariento for being available for help.

I am pleased to thank Dr. Joaquin Huerta, Dr. Christoph Brox and Prof. Marco Painho for their support and help during the masters' duration for me and for all student.

I am also thankful to Dori and Karsten, for always being available to help us all the time. I am grateful to the opportunity I had from the European Commission, I had so much experience by living with students from different countries. I am thankful to all my friends in Spain and Germany especially Aida for helping me and making my stays in Spain easier and unforgettable. Also I want to thank all my friends back home for their endless help and care.

Last but obviously not least, I would like to thank my beloved family Fatma, Radwa, Reham, Hussein and Hameed for their support, love and caring. And Dr. Mahmoud Eissa for motivating me to complete this master.

ABSTRACT

The reduction of greenhouse gas emissions is one of the big global challenges for the next decades due to its severe impact on the atmosphere that leads to a change in the climate and other environmental factors. One of the main sources of greenhouse gas is energy consumption, therefore a number of initiatives and calls for awareness and sustainability in energy use are issued among different types of institutional and organizations. The European Council adopted in 2007 energy and climate change objectives for 20% improvement until 2020. All European countries are required to use energy with more efficiency. Several steps could be conducted for energy reduction: understanding the buildings behavior through time, revealing the factors that influence the consumption, applying the right measurement for reduction and sustainability, visualizing the hidden connection between our daily habits impacts on the natural world and promoting to more sustainable life. Researchers have suggested that feedback visualization can effectively encourage conservation with energy reduction rate of 18%.
Furthermore, researchers have contributed to the identification process of a set of factors which are very likely to influence consumption. Such as occupancy level, occupants behavior, environmental conditions, building thermal envelope, climate zones, etc.

Nowadays, the amount of energy consumption at the university campuses are huge and it needs great effort to meet the reduction requested by European Council as well as the cost reduction. Thus, the present study was performed on the university buildings as a use case to:

a. Investigate the most dynamic influence factors on energy consumption in campus;
b. Implement prediction model for electricity consumption using different techniques, such as the traditional regression way and the alternative machine learning techniques; and
c. Assist energy management by providing a real time energy feedback and visualization in campus for more awareness and better decision making.

This methodology is implemented to the use case of University Jaume I (UJI), located in Castellon, Spain.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

KEYWORDS

Greenhouse Gas Emission
Building Efficiency
Building Conservation
Energy Consumption Prediction
Data Mining
Artificial Neural Network
Stepwise Multiple Regression
Web GIS
Geographical Information Systems
Energy Feed back

ACRONYMS

Acronym	Definition
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
API	Application Programming Interface
BSON	Binary JavaScript Object Notation
EDA	Exploratory data analysis
GHG	Greenhouse Gas
GIS	Geographic Information System
HVAC	Heating, ventilation, air-conditioning load
ICT	Information and Communication Technology
INIT	Institute of New Imaging Technologies
IPCC	Intergovernmental panel on climate change
JSON	JavaScript Object Notation
PV	Photovoltaic
UJI	University Jaume I
UNFCCC	United Nations Framework Convention on Climate Change

Chapter One: INTRODUCTION

1.1. BACKGROUND

The reduction of greenhouse gas emissions (GHG) is one of the big global challenges for the next decades due to the severe impact of those gases on the atmosphere. Each of them can remain in the atmosphere for different amounts of time leading to a change in the climate and other environmental factors. The climate change is considered one of the emerging problems as its impact affects the world environmentally and economically (IPCC 2013). One of the reasons of increasing the GHG is energy consumption, therefore a number of initiatives and calls for awareness and sustainability in energy use are issued among different types of institutions and organizations. In 2007 the European Council adopted energy and climate change objectives for 20% improvement until 2020 and all European countries are required to use energy in a more efficient way[1]. For achieving this, a number of governmental laws were set to work on sustainability and energy among all sectors (private, governmental and educational).

As buildings are contributing about 40% of total amount of energy consumption in Europe[2], Figure 1 shows the final energy consumption by sector in European Union, in 2009, the urge need for consumption reduction is becoming the main goal for many institutions to meet the 2050 global goal[3]. Regarding this, all educational places, especially 'universities' are participating in spreading the sustainability behavior among students and universities' staff, running several efficiency[4] measurement, policies and laws for energy conservation.

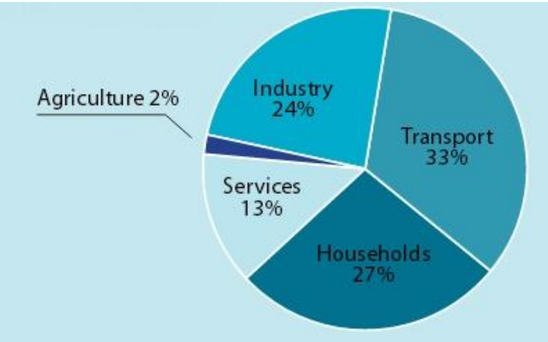

Figure 1: Total energy consumption by sector in EU 2009[5]

[1] The European Council 'Brussels, 10.1.2007 concluded that as part of a global and comprehensive agreement, the EU should reduce greenhouse gas emissions by at least 20% compared to 1990 levels, 'Limiting Global Climate Change to 2 degrees Celsius The way ahead for 2020 and beyond'.

[2] http://ec.europa.eu/energy/en/topics/energy-efficiency/buildings

[3] The European Council 'Brussels, 25.5.2011 COM(2011) 112, Communication from The Commission to The European Parliament, The Council, The European Economic And Social Commitees And The Committee of The Regions', 'A Roadmap for moving to a competitive low carbon economy in 2050'.

[4] 'energy efficiency' means using less energy while maintaining the equivalent level of activity or service.

[5] http://ec.europa.eu/energy/

For these reasons, researchers in engineering and environmental sciences conducted a number of studies lately, in order to investigate and develop different ways for analyzing energy and saving methodologies for the buildings, which is the most cost effective way for energy enhancement, trying to reach a high level of sustainability and cost reduction in most places (Fischer, 2008; EPRI ,2009; Ayres et al., 2009; Marszal et al., 2011). Understanding and monitoring the building behavior at different temporal situations and leveraging the impact of other factors on the building consumption are the main keys for any future enhancement.

Several steps should be taken in any organization for more energy efficiency and reduction[6]. Those steps include:

- Monitoring with regular analysis of measured data,
- Defining the factors that influence the consumption,
- Applying sustainability and reduction measurements in the place; and
- Improving the occupants' behavior and awareness of sustainability by showing them the hidden connection between their daily habits impacts on the natural world with different visualization methods, and promoting to a more sustainable life.

As the old rule says "you can't manage what you don't measure", the need to regularly sense the environment condition and measure the daily consumption of building's resources from water, gas and electricity are now possible with the new spread technology of smart meters and the low cost of sensors. Smart meters are small devices that allow us to measure the consumption on intervals e.g. every 15 minutes. All these technologies make it easier to retrieve, monitor and organize a huge amount of raw data that represent the building interaction with other environmental factors, which may contribute to the management and sustainable use of energy.

The energy consumption in a location or room is influenced by several factors. Some factors remain stable through time (like walls, windows, fixed equipment) while others can vary all the time (like outside temperature, sun or wind incidence, people getting into and out of the room and connecting temporarily other equipment) (Sailor et al., 1997; Pettersen, 1994). Studies on the electricity consumption characterization have used several variables, some of these variables are related to the geographic location where climatic zones are varying (Matsuura, 1995). In order to analyze the energy use in a location, these factors should be investigated and modeled, especially the dynamic factors that are varying spatially and temporally. Also, predicting the energy demand is one of the major techniques for better resources management and decision making.

Furthermore, researchers have performed several researches related to the changing behavior of end-users through energy data visualization and they estimated the possibility of building reductions of more than 20% as a result of data visualization (EPRI 2009). Visualization of energy consumption is widely considered as important means to assist not only end-users but also energy managers in facilities management for reducing energy consumption.

[6] http://ec.europa.eu/energy/en/studies?field_associated_topic_tid=All

Nowadays, the amount of energy consumption at the universities campuses is colossal and it needs great effort to meet the reduction request by the European council. Monitoring, analyzing, predicting and visualizing the energy use on a real time helps in future building planning and decision making, since it can provide useful information about the most efficient buildings, or predict energy use in different conditions for similar places.

1.2. PROBLEM STATEMENT

The importance of universities in promoting sustainable development has been increased lately as a growing number of sustainable and environmental development initiatives[7]. Policies[8] have been issued and followed by so many universities around Europe. For addressing the energy reduction, the university established the project (UJI Energia)[9] for monitoring total campus consumption and cost, by deploying several smart meters in the whole campus, replacing inefficient machines and using different sustainable resources for energy.

The project of the Universitat Jaume I (UJI)'s Smart Campus, in Spain, is devoted to create a system to query and visualize information from several university sources, including information about the personnel, locations and services, in a unified and homogenous way[10] (Benedito-Bordonau et al., 2013).

In addition to all this effort and work, the part concerning the energy data need to be addressed and will rely on the many electricity meters deployed in the UJI's campus. Apart from the meter visualization of the consumption measured, the added value of highlighting variation in the consumption in a real time can be incorporated, thus reducing the need for a specific and complex system to be developed or bought for energy management. Also the factors that have influence on energy usage for each building such as outside environmental factors (temperature, wind, etc.) are not investigated yet and there is a need of comprehensive study and analysis in campus. The resulted raw data from monitoring the energy consumption is not sufficient for understanding or managing the campus energy. Analyses and discovery of the data is the only way for: revealing the buildings behavior by identifying the nature of the phenomenon that is represented by the data, discovering the pattern in the energy use and finding the influencing factors.

Therefore, for following the researchers' recommendations and methodologies for energy reduction, the parts of energy resources management and planning in UJI campus need to be improved. This could be done by analysis and prediction of energy consumption, in addition to visualization of energy variation in a real time rather than weekly or monthly monitoring for better management.

By looking to the importance of energy modeling and prediction in planning, we found that several techniques were applied for energy prediction. One of them is the

[7] The European Council of Brussels, 10.1.2007 concluded that as part of a global and comprehensive agreement, the EU should reduce greenhouse gas emissions by at least 20% compared to 1990 levels, 'Limiting Global Climate Change to 2 degrees Celsius The way ahead for 2020 and beyond'.
[8] http://ec.europa.eu/energy/en/topics/energy-efficiency
[9] http://www.energia.uji.es/
[10] http://smart.uji.es/

traditional statistics method such as regression analysis (Lam et al., 1997). In the past, regression models showed a good result in predicting energy consumption and it was the only adopted method with strong theories as being a best linear unbiased estimator. Now, the alternative approaches like neural network has not been so popular in energy prediction modeling literature compared with regression methods, although those alternative techniques showed useful models in other fields such as risk management (Nasri, 2010).

1.3. OBJECTIVES

The aim of this thesis is to assist the university management in energy reduction goal by analyzing, modeling and visualizing the energy data. This overall knowledge of factors and models could be used in the arrangement of classes' schedules or operations in a specific building for future plan. In addition to providing useful visualization of information in order to support the management of energy resources inside the university, it is also important to increase the awareness by spatially providing real time visual feedback.

As the interest is in energy variation, only dynamic factors that can be measured will be taken into account in this analysis.

Thus, the specific objectives of the thesis are to:

- Give an insight to the energy consumption data by analyzing the change during the time, recognizing the pattern and energy waste;
- Investigate the most influence factors on energy consumption in campus;
- Implement a prediction model for electricity consumption using different techniques, such as the traditional regression way and the alternative machine learning techniques; and
- Implement a prototype of a real time energy feedback and visualization web application for more awareness and better decision making.

1.4. RESEARCH QUESTION AND HYPOTHESIS

The primarily assumption is that the environmental factors such as temperature and humidity have effect on the campus building's consumption level. These factors are varying during all seasons especially for types of buildings where the occupancy level is dynamically changing over time.

From the previous context this question is considered to be answered during the analysis.

- Are seasonality change and dynamic activities considered as a main driver for the energy consumption in the campus?
- What is the suitable model to predict the energy consumption in the university?

- How to improve the energy information provision in campus?

1.5. GENERAL METHODOLOGY

For answering the previous questions and purpose of the thesis, our methodology is generally composed of three main phases. Namely: analysis of energy consumption, modeling and predicting the energy use, and visualization of energy variation and main indicators in campus.

To achieve these phases we will undertake four steps: data collection and preparation, exploratory data analysis, exploring different techniques for energy prediction, and visualization of energy consumption. These steps are described in the following sub-sections.

1.5.1 DATA COLLECTION AND PREPARATION

With the new system for energy monitoring in the campus, energy data are available in detailed interval. This new system 'Smart metering system' automatically measures and records the consumption at regular and short interval such as 15 minutes. The system has a benefit over the old regular ones where they used to collect manually the data at weekly or monthly duration. This old approach makes it difficult to see the energy pattern or how much is consumed at specific hour of the day, or on different days of the week.

As we are studying a few number of buildings, so more detailed data are required for better analysis. Thus, data about buildings' historic consumption, factors that might influence the consumption such as daily or hourly activities in the building and environmental data (temperature, wind speed, humidity) are also required.
Generally in this step all these data will be gathered and preprocessed for removing any out of range values (outliers) and interpolating any missing data.

1.5.2 EXPLORATORY DATA ANALYSIS

After data collection and preparation, the analysis will be performed using the statistical method exploratory data analysis (EDA) for the collected data. This analysis is performed to understand the pattern and the relation between the phenomena under study (consumption) and other factors. Different type of graphs will be used to explore the data such as scatter plots. This type of graph can reveal the relationship between two variables, also energy profile plot type that shows how much energy is being used in a specific time period or duration. During this part of the work, a descriptive analysis and explanation to the energy use pattern in campus will be introduced.

1.5.3 EXPLORING DIFFERENT TECHNIQUES FOR ENERGY PREDICTION

In this part of the research, the factors that we concluded from the previous step that have some effect on the consumption will be investigated in order to determine which one of them has a major influence and could be used as a predictor for the prediction model. We have used two different techniques to predict the energy consumption in campus: stepwise regression and neural network models.

1.5.4 VISUALIZATION OF ENERGY CONSUMPTION

The last step is designing and implementing a web application for real time energy visualization in campus. The purpose of the application is to visualize the main key performance indicators in the campus for managers and end-users, and showing the variation of energy during the whole day. This application will be designed with requirements identified in the literature survey for better information and visual representation.

1.6. ORGANIZATION OF THE THESIS

The thesis is comprised of six chapters in addition to the appendixes and list of references, as follows.

- **Chapter One**: Introduction providing a quick insight to the thesis, mentioning the problem, objective, hypothesis and methodology.

- **Chapter Two**: The theoretical framework, consisting of concept and definitions with review about previous and related work using different methods.

- **Chapter Three:** A brief description about the university and the ongoing monitoring system, data and limitation of the analysis**.**

- **Chapter Four:** A detailed analysis of the energy use and modeling approach**.**

- **Chapter Five:** Contains the energy feedback visualization.

- **Chapter Six:** Discussion of the results and future work.

Chapter Two: Theoretical Framework

2.1. CONCEPTS AND DEFINITIONS

2.1.1 Regression analysis

Regression analysis is considered a tool for investigating the relationship between variables in order to determine the effect of one variable upon another (Alan, 1992). The relationship could be easily identified by assembling the data and employing regression to estimate the effect by statistical way 'degree of confidence'.

This type of analysis is widely used for prediction and forecasting. Having one dependent variable and another set of independent variables and a relationship between them should be drawn, with the objective of knowing which variable is more effective among the others on the dependent variable. In this case, regression analysis is used.

A simple linear regression model could be represented by this equation (Eq. 1):

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad \text{.........} \text{ Eq. 1}$$

Where y is the dependent variable or the variable of interest. This variable is driven by other variable x (independent variable). The relation between them is linear and (i) is the index of observation of (x, y) pair. The parameters β_0 and β_1 represent the y-intercept and the slope of the relationship, respectively.

2.1.1.1 Multiple Regression

Multiple regression is a technique that allows factors to enter the analysis separately, one after another so that the effect of each can be estimated. The advantage of this method over the simple regression is due to the omitted variable bias from the simple regression. Usually, it is recommended to use the multiple regression even in the cases of there is only one variable of interest (Alan, 1992).

2.1.2 Artificial Neural network (ANN)

An Artificial Neural Network (ANN) is a computational method that mimics the learning processes of the human brain. This method is inspired by the biological concept of neurons where animals are able to react with their environment (external or internal) using their nervous system. A good artificial model is the one that is capable to simulate the nervous system and response in a similar way. The structure of a neural network consists of input and output and by the learned knowledge that the system tries to map between them (Haykin, 1999).

According to Haykin (1999), the ANN advantage comes from its wide parallel distributed structure and its ability to learn and generalize. Generalization refers to the ability to solve and produce a reasonable output for inputs not encountered during training.

The ANN properties are non-linearity and input-output mapping. They can be described as follows (Haykin, 1999):

- Nonlinearity: the ANNs are made of an interconnection of nonlinear neurons.
- Input-Output Mapping: ANNs can use two different learning paradigms: supervised or unsupervised. Both start with the network assigning weights randomly chosen. Later the learning mechanism starts, in the supervised technique the network is provided by the desired output but the unsupervised techniques let the network responsible for making sense of the input without any help:

    a. Supervised training: the input/output are both provided, the network then starts to process the inputs and compares its results against the outputs. Resulting errors are propagated back through the system for adjustment. The process repeatedly occurs as the weights are continually tweaked. In this technique, it is ideal to have enough data so that it can be divided to different sets [training, testing] for monitoring if the network is perfectly learning well. Usually, if the result is not sufficient, the design of the network [number of layers, input, output, the neurons, transfer function and training function] should be modified.
    b. Unsupervised, or Adaptive Training: in this case the network is provided by inputs only. Later the system decide which features will use to group the input sample, this is called self-organizing (SO) or adaption. In this sense, adaption means to continually learn on their own as new variables or situations are encountered.

Figure 2 presents the architecture of a simple neural network.

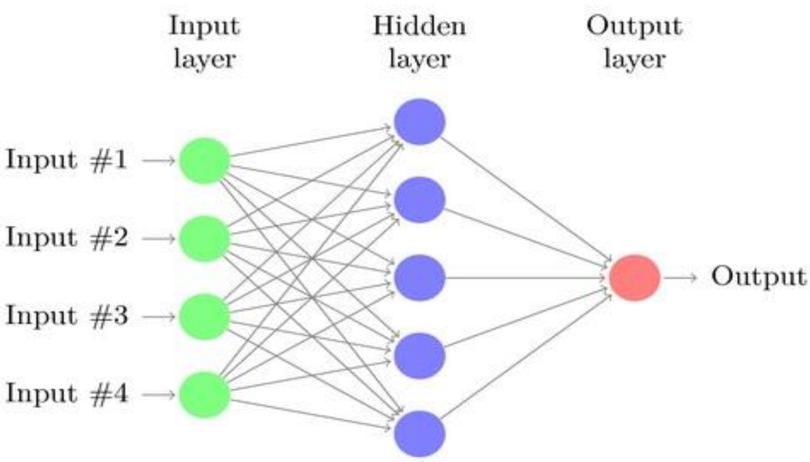

Figure 2: Simple neural network architecture

2.2. RELATED WORK

Prior to analyze the energy use inside the campus, a review of previous work to identify the influence factors and identifying patterns in buildings was necessary. The university campus is considered as a mixed use buildings environment which varies between offices, educational and even residential and commercial. Several researches relied on data analysis methods for getting useful information and knowledge from measured data by following different methodologies (Olofsson et al., 2009; Djurica and Novakovicb, 2012).

The previous researches have contributed to the identification process of a set of factors which are very likely to influence consumption. It is difficult to say that these factors are general since energy consumption depends on many criteria for building's location and environments. However, some of them will be considered as candidates through the analysis process trying to prove the hypothesis of the thesis. The next step after finding the derived factors for the consumption is the use of those factors to predict the energy using two different approaches.

A search of literature reveals that different methods were adopted to develop factors which influence energy consumption such as Regression Analysis and Neural Network. Both are described in the next sections.

2.2.1 Regression Analysis

Regression analysis is a type of statistical analysis which is widely used in building engineering for spotting the correlation between building energy consumption and its influencing factors like occupancy patterns, building envelope characteristics and heating, ventilation, air-conditioning load (HVAC) system. This technique is also used to analyze overall buildings using patterns and how these factors affect energy consumption.

Corgnati et al. (2008) has performed an analysis of 120 schools in Italy to assess the heating consumption. They collected data from meters that were deployed in schools and organized the data in index mechanism. A normalization among all data was performed, followed by statistics analyses. They collected data about climatic zones, building characteristics and monthly consumption, later a regression method was used to find the relation between the heating consumption and the change in the temperature by seasons (heating degree days and cooling degree days).

In a similar way but with different scale regarding cities, Chen et al. (2010) have analyzed the energy consumption characteristics in summer season in residential areas in different cities. Such analysis showed a wide variation in temperature and different influencing climate parameters. They used statistical analysis, specifically regression analysis to correlate different independent variables. The result of this work shows that location, building units, construction year, indoor thermal, utilization of space coolers and water heaters and outside temperature contribute to the residential energy consumption in summer.

Zmeureanu and Fazio (1991) performed regression analysis of 68 office buildings data collected. One of the conclusions is that buildings' age has an effect on consumption. Djurica and Novakovicb (2012) also performed principle component regression analysis (PCR) and Partial least square regression on low energy office buildings using data that include occupancy level, inside temperature, water and control signals. They tried to find the correlation between those variables and heating, electricity and fan energy use. The result varies depending on the months. For example, the heating is influenced by the occupancy level operation parameter and not by the outside temperature, whereas the electricity use is influenced by the indoor temperature and occupancy level. Additionally, it was found that the regression model should be updated on monthly basis according to the seasonality change.

Regarding prediction of building energy, regression analysis was widely used, not only to predict energy consumption but also for different factors inside the buildings like occupancy rate and indoor temperature. Lam et al. (1997) modelled high-rise air conditioned office buildings in Hong Kong. They collected data and found that 28 parameters are correlated with the building load and HVAC system. Both linear and non-linear multiple regression techniques were applied to develop a prediction model for annual energy. A number of those parameters were eliminated and only 12 showed high sensitive coefficient to the annual energy use prediction.

Caldera et al. (2008) used statistical analysis for energy demand assessment. Data were collected for 50 buildings with different construction years, trying to correlate the building geometry and thermo-physical parameters, like the volume and the shape, the construction criteria, like glazed surfaces, and the environmental factors, like temperature.

Catalina et al. (2008) undertook a research for developing a regression model to predict monthly heating demand. The model was based on data collected from 16 cities in France. Data comprised buildings envelop characteristics and buildings' specifications like the construction year, windows and floors area and climate data. The analysis showed that there is a strong relation between the building shape and the consumption. Also, the internal thermal characteristics have a strong effect on consumption. The resulting model was efficient with a maximum deviation between the simulated data of 1.2-5.2 %.

2.2.2 ARTIFICIAL NEURAL NETWORK

In their experiment, Datta et al. (2000) used ANN to predict the electricity demand in a supermarket, in order to control the load and avoid any penalty due to the exceeding of contract's load rate. They made the experiment with half an hour rate, considering the time of the day as one factor in addition to the environmental factors. Using the feed-forward network and training the model with real data from the store, they tested 7 networks with different combinations of input variables such as Day, Time, External Humidity and Temperature, Internal Humidity and Temperature. The results indicated that using a short-term data set may be adequate to accurately predict half hourly electrical demand.

With similar technique feed forward, Yalcintas and Akkurt (2005) modeled the power consumption of the central chiller plant, whereas the input included dry bulb temperature, wet bulb temperature, dew point temperature, relative humidity percentage, wind speed and wind direction. They were able to successfully predict the chiller plant power consumption depending on meteorological parameters. Ekici and Aksoy (2009) performed a back propagation technique for neural network in order to predict heating energy consumption in buildings with three input and one output values. They trained the network with buildings' transparency ratio, orientation and insulation thickness. The author mentions that the advantage of using neural network over other methods is its capability of performing complex modeling with high prediction accuracy for heating loads with average of 94.8-98.5%.

With a similar back propagation method, Kalogirou and Bojic (2000) used ANN to predict the energy consumption in passive solar buildings. Several use cases were simulated, depending on variables such as walls insulation and seasons. The network was trained using those variables as input and the consumption as an output. The neural network was successfully used to model the thermal behavior, having a R value of 0.9991. The advantage of the neural model is the adaptation to new input values which were not in the same training set. The authors also revealed as another advantage of using the neural network the fact of its acceptable accuracy in prediction, instead of using other expensive modeling programs for passive solar buildings.

Olofsson and Alexandersson (2001) investigated the ability to predict the annual heating energy from limited measured data (2 to 5 weeks). The parameters used for prediction were temperature (indoor and outdoor) and heating energy. They chose the model parameters by performing Principal Component Analysis (PCA). The model result was accurate in predicting single family buildings with deviation error up to 4%.

Chapter Three: STUDY AREA AND DATA

This chapter contains a detailed description of the university and its facilities with the major conservation measurement, which has been done to decrease our carbon footprint. It also includes the description of the data that has been collected and stored.

3.1 THE ROLE OF THE UNIVERSITY IN SUSTAINABILITY

The University Jaume I (UJI) has followed several ways towards sustainability enhancement in the campus, considering the fact that it is composed of a group of diverse buildings with significantly high consumption, representing a small city with different buildings' use (offices, cafeterias, shops, laboratories, sport centre, classrooms, etc.). Therefore, the call for promoting energy sustainability and conservation measurement has been guided in a number of agreements and declarations in Europe, and it has been followed to some extent in the university. The work was promoted on different scales such as the academic programs, which is taught through courses like environmental studies or by rising the awareness through data provision using the university website. Bearing in mind that all higher institutions can affect many teachers, students, leaders and decision makers of tomorrow either in class room or in community work, the green model of the university can be later extended to community scale.

All the work conducted by educational institutions in this area is being recognized, encouraged and funded by the European Union, governments and councils[11].

3.2 UNIVERSITY CAMPUS

The University Riu Sec campus is composed of several buildings and with a total built area of 219257.69 m^2 [12]. Depending on their purpose, building types are: administrative, educational, laboratory, and workshop, residential and sport facilities. Figure 3 categorizes all university's buildings.

[11] http://www.eeef.eu/
[12] http://www.uji.es/CA/serveis/otop/edificis/sctotal.html

Figure 3: Categories of university buildings

The university took some actions to reduce the energy usage in the campus, as follows:

- Replacement of inefficient cooling/heating machinery;
- Installation of photovoltaic system;
- Installation of sensors inside classrooms for "light" and "air conditioning" control;
- Strict control of temperature in classrooms;
- Reduction of air conditioning schedules;
- Lighting control schedules (interior and exterior).

Those actions in conjunction with others led to an effective reduction in the total consumption. Figure 4 displays the consumption trend during the last five years.

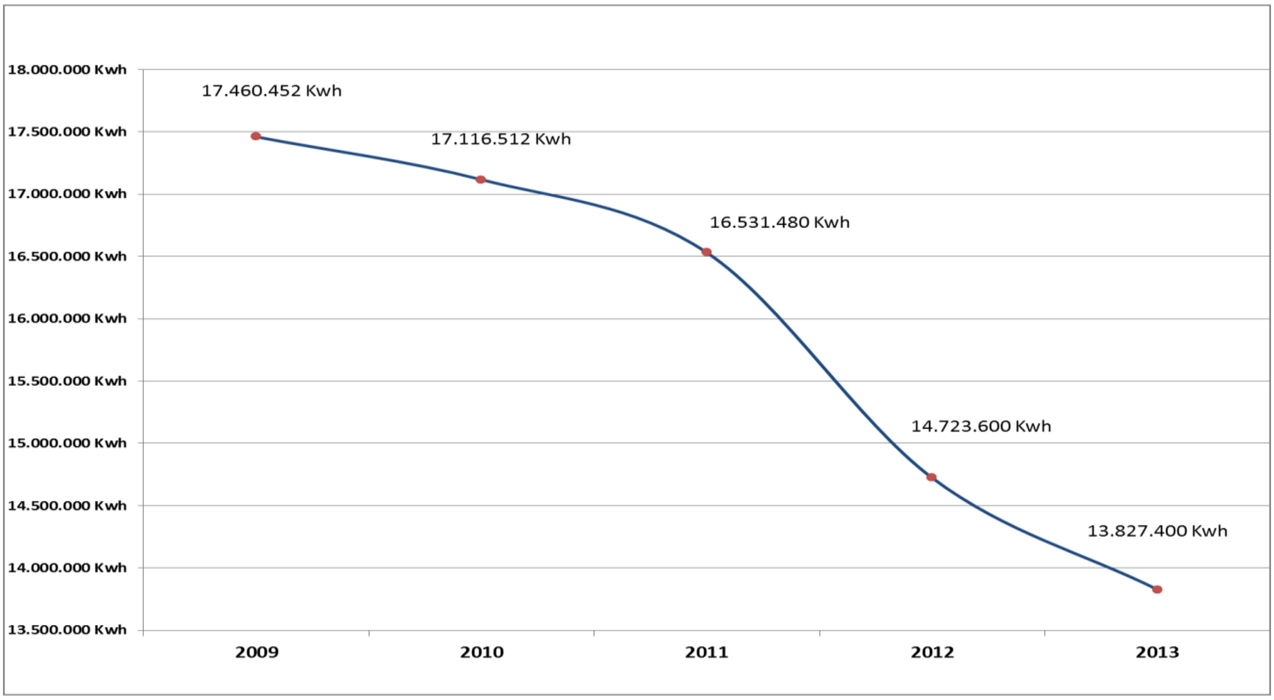

Figure 4: UJI's Energy consumption between 2009 and 2013

In addition to all these measures, an analysis of energy use in the university is highly recommended with further use of ICT (Information and Communication Technologies) in energy monitoring and information provision in campus. The first step was adopted by the management by deploying a system for energy monitoring. The next step will be analyzing those data.

3.3 DATA DESCRIPTION AND COLLECTION

For the purpose of the thesis we needed to collect data for the analysis and modeling from different sources. Data comprise the following:

- Data for energy consumption (historical data);
- Data for building specifications (area, type);
- Data for the historical (temperature, humidity, wind speed) measurements;
- Data about building occupancy (classes' schedule) and university calendar.

3.3.1 Data Source

Figure 5 shows the source data, provided by the university administration system, the campus energy monitoring system and other sources. Figure 5 also illustrates data collection process. The data contents include building information, weather data, campus activities (Classes), and school calendar. Weather data include temperature, humidity and wind speed.

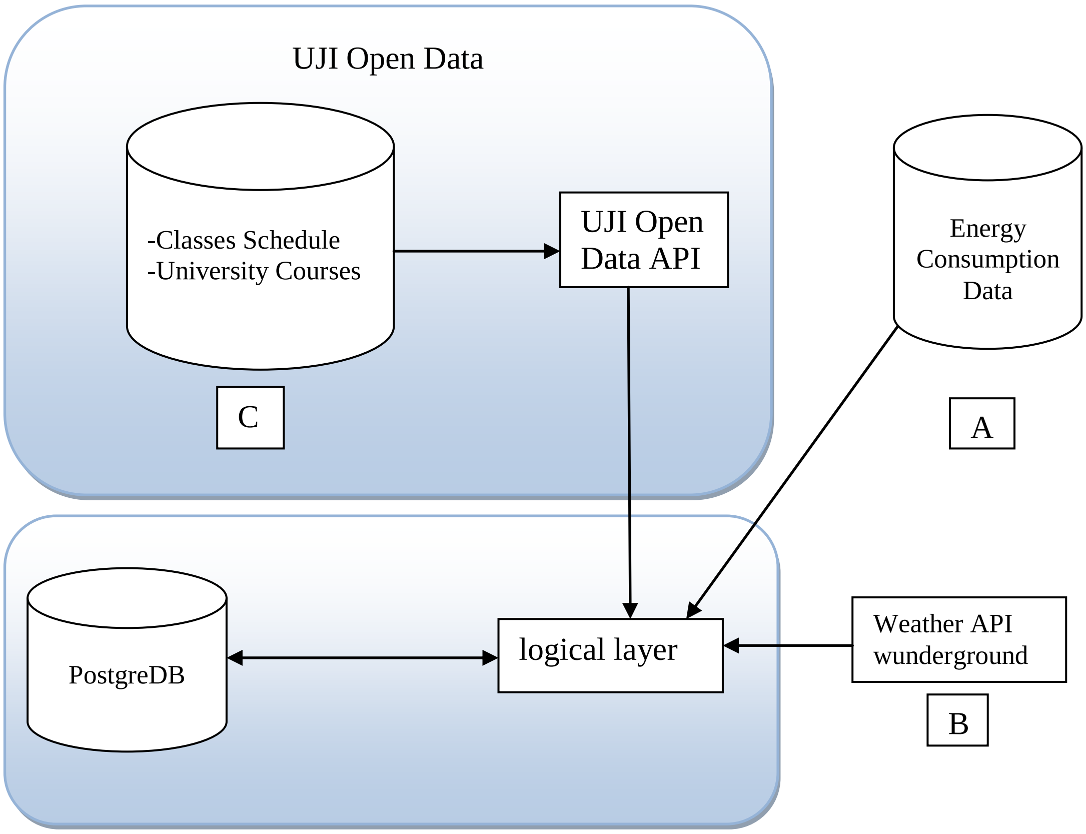

Figure 5: Diagram for data sources for analysis and collection process.

Figure 5 comprises 3 groups of data, A, B and C, described as follows.

3.3.1.1 Energy Consumption Data

Set A is composed of the energy consumption data that are collected during the monitoring of the consumption in university using the smart meters which are deployed in several buildings. The smart meter is recording the consumption of energy in different interval scales, and it communicates this information to the utility in two-way communication channel. Smart meters usually include real-time or near real-time sensors, in this case the interval is 15 minutes. The data are then stored in MongoDB.

MongoDB[13] is a document-oriented database classified as NoSQL. It has a different format called BSON[14] (Binary JavaScript Object Notation). Each record in this database is stored as a document, the document schema is similar to JSON[15] (JavaScript Object Notation) style objects. Figure 6 illustrates an example of one event document that holds one of the meters' readings, it is composed of field-and-value pairs and has the following structure.

[13] https://www.mongodb.org/
[14] http://bsonspec.org/
[15] http://json.org/

```
db.events.findOne({"_id" : ObjectId("5433caaa0cf26c324a7d993a")})
{
	"_id" : ObjectId("5433caaa0cf26c324a7d993a"),
	"ref" : "72689676-DC3E-4602-90FF-1B07D311E599",
	"sts" : ISODate("2014-10-07T10:45:00Z"),
	"rts" : ISODate("2014-10-07T10:45:48.394Z"),
	"c" : {
		"value" : 445,
		"type" : "LONG",
		"quality" : "GOOD"
	},
	"r" : {
		"value" : 445,
		"type" : "LONG",
		"quality" : "GOOD"
	}
}
```

Figure 6: Example of one document for one meter reading

This example of document includes the measured value of energy consumption and related sensor data on a specific timestamp for one meter.

Because this system is a new one, this Database does not include any of the historic data. Since we wanted to collect a long duration of data for analysis. Thus the historic data for the duration prior to the new metering system installation were provided by the university's management in Excel sheets format. These Excel data are saved later in PostgreDB. The new system MongoDB is only used in the new developed tool for real time energy visualization, the communication between the monitoring system and the new developed web application was facilitated by developing a Java program.

Previous to analysis, a preprocessing step was applied on these data, in order to remove any outlier and filling the missing data by interpolation method.

3.3.1.2 Environmental Data

Data for historic outside temperature, humidity and wind-speed were collected from Wunderground API[16] (Application Programming Interface) for one year (2014). This API is providing current and historic weather information with different data format, whereas data observations are every 30 minutes. For collecting these data and saving it in PostgreDB we had to develop a Java program. The data format was in JSON-style, Figure 7 below is a sample of one observation:

[16] http://www.wunderground.com/weather/api/

```
"observations": [
                {
                "date": {
                "pretty": "1:00 AM CEST on October
02, 2014",
                "year": "2014",
                "mon": "10",
                "mday": "02",
                "hour": "00",
                "min": "00",
                "tzname": "Europe/Madrid"
                },
                "tempm": "19.0",
                "tempi": "66.2",
                "dewptm": "18.0",
                "dewpti": "64.4",
                "hum": "94",
                "wspdm": "3.7",
                "wspdi": "2.3",
                .
                .
                }]
```

Figure 7: Example of API Weather data in JSON format

3.3.1.3 Building Occupancy Data

Set C is an open data platform provided by the university to users through an RESTful Service, it contains data for academic teachers guide, classes schedule and courses information. This API allows data retrieving in different format such as JSON, RDF, TURTLE and CSV.

The data for classes schedule were collected and saved in PostgreDB also using developed Java program. This program retrieves all courses and locations where these courses are running with their start and end time. Figure 8 shows a sample of collected data in JSON format.

```
"content": [
    {
      "grupo": "A",
      "curso": "2014",
      "inicio": "2014-09-24 15:00:00.0",
      "subgrupo": "PR1",
      "_id": "1581744",
      "semestre": "1",
      "asignatura": "AE1040",
      "fin": "2014-09-24 16:00:00.0",
      "aula": "JB1001AA"
    },
        .
        .
        .
]
```

Figure 8: Sample of classes schedule data provided by UJI Service

1.3.1.4 Building Specifications Data

The data regarding the buildings were collected from the university website[17], and contain each building surface and floor area. Since the consumption per area is considered one of the important key performance indicators, those data are very important in order to visualize the building efficiency. These data were stored in the campus Geodatabase. The UJI Geodatabase contains geographic layers and stand alone tables for all campus buildings, interior spaces, points of interest (parking areas, waste containers, restaurants and shops, etc.) (Benedito-Bordinau et al., 2013).

3.4 DATA LIMITATION AND ISSUES

During the collection and analysis of the data, we observed some limitations and issues that may cause some impact in the work. Thus, we considered important to list them as follows. We also present some of the solutions we adopted in order to overcome those issues.

1. The monitoring system changed several times and that caused loss of data for the durations between installing each system and the other.
2. Another reason for missing data was the meters malfunction; in order to overcome this fact, we had to use interpolation method to fill the missing data.
3. Some meters are covering a wide area which could include two or three buildings; that fact makes the analysis difficult as the buildings operation is different in the campus. If there were more meters, the energy tracking and assessment for individual buildings would be improved, this case was not included in the analysis.
4. The meters exact locations is not known, so we had to map the meters by its covered area (building) in a Geographic Information System (GIS). In this case each meter's ID was associated to one building. This step was important for the implementation of the visualization web application.
5. During the past years, meters have been continuously installed, giving a difference in data temporal resolution. Some buildings have measured data starting from 2012 and some from 2014. Data used in the present study only comprise a year from all the collected data (2014).
6. As one of the thesis goal is to gather all possible data about buildings and building's events for more analysis and assessment in the future, the data collection from different resources with different format, as explained in the previous section, consumed a lot of time.
7. Unfortunately, data for the building's envelope, like the thermal characteristics, were not available, thus we could not take into account those variables in the analysis.

[17] http://www.uji.es/CA/serveis/otop/edificis/

Chapter Four: DATA ANALYSIS

In this chapter, we analyze the collected data to find out if there is any waste in the energy besides investigating the influence factors on the consumption. Later, we use these factors to predict the energy consumption. As an initial step, an exploratory analysis is done in order to give more insight to the data.

4.1. EXPLORATORY DATA ANALYSIS

Behrens (1997) has described the exploratory data analysis (EDA) as a tradition way in data analysis that stems from John Tukey's early work in 1960s (Tukey, 1969). Exploratory analysis can be characterized as:

- An understanding of the data that answer back this main question “What is going on here?"; and
- An emphasis on graphic representations of the data.

Consequently, in addition to the goal of describing of what is going on, a pattern discovery is also revealed by this analysis. Behrens (1997) mentioned that EDA often linked to detective work, as the role of the data analyst is to listen to the data in as many ways as possible until a plausible story of the data is apparent. EDA has many approaches for graphical representation, such as scatter plot, frequency distribution, histogram plot, etc.
In order to make the analysis more efficient, we classified the campus buildings in three main categories according to their main use and operational hours:

1. Classes/Teaching buildings.
2. Administration/companies buildings.
3. Facilities buildings.

Figure 9 represents the total consumption during the year. Analyzing the general pattern in the university, it is noted one month off-peak, August, when the university is closed, and another peak in September due to the beginning of the semester.

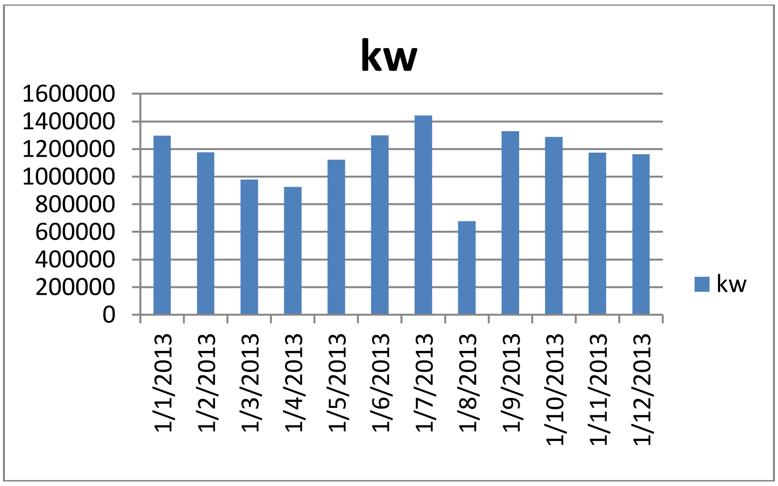

Figure 9: Total energy consumption per month for year 2013

Due to the limitations of the data that were mentioned before in chapter 3, we will analyze one building that has data available for one year, 2014. The chosen building is the Faculty of Humanities and Social Sciences, which belongs to the category of ' classes/teaching buildings'. The main activities occur during the official academic semester, and the building contains classrooms ('Aula'), laboratories and teacher's offices. The one year data set was divided to two sub data sets, from November to March and from April to October. As the period from November to march has different energy consumption source, the university is using Gas for heating and this source of data is not available. Therefore, this period was excluded from the analysis.

Figure 10 shows the intensity of the consumption of 'October'. As an example we could notice the pattern which is related to the hour of the day, the consumption starts to increase around 8:00 AM and starts to drop down around 8:00 PM. This pattern is due to the regular operation time and building type, the classes operate from 8:00 AM till 8:00 PM. From the same graph we could see the deviation between the week day and the weekend days.

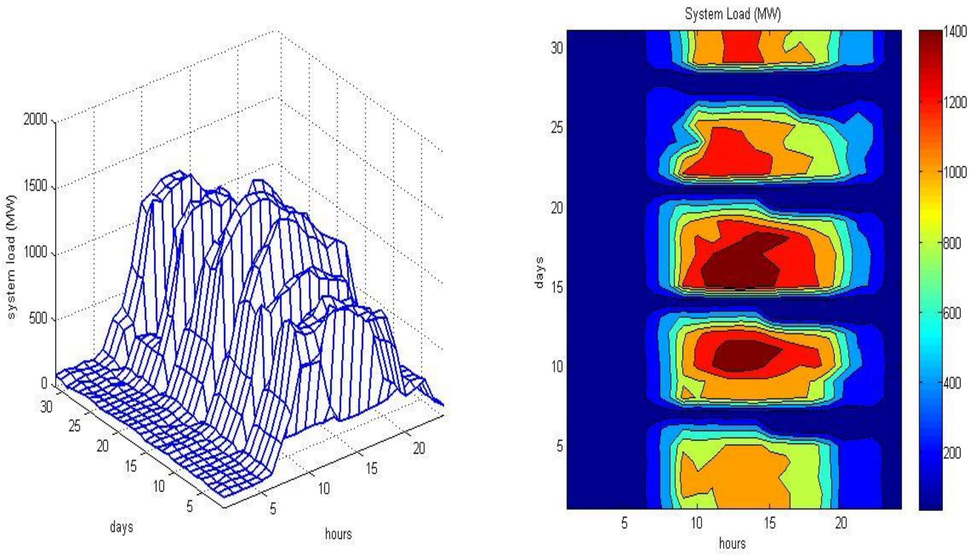

Figure 10: 3D model for energy load of 'October' 2014, Faculty of Humanities and Social Sciences.

4.1.1 ENERGY USE PROFILE FOR BUILDINGS

Graphs that presented in next two figures can show how much energy is being used every hour of the day and also for every day in the week, revealing a typical consumption which could be useful for finding energy waste. The small interval details of the measured data could lead to the discovery of any wasting of energy, which can be seen as an advantage. The main important thing is to be able to relate this waste of energy to the activity inside the building or the operation time.

For example, if the energy profile shows that there is energy being used in a period during the day/night when usually there are no occupants, this could be seen as an indication of energy waste, and it should be further investigated. For understanding the campus behavior during the day/week, we analyzed the typical hourly consumption for a whole day. In addition to the monthly profile, a standard deviation function was calculated with mean and 95% confidence interval of the daily profile (Figure 11).

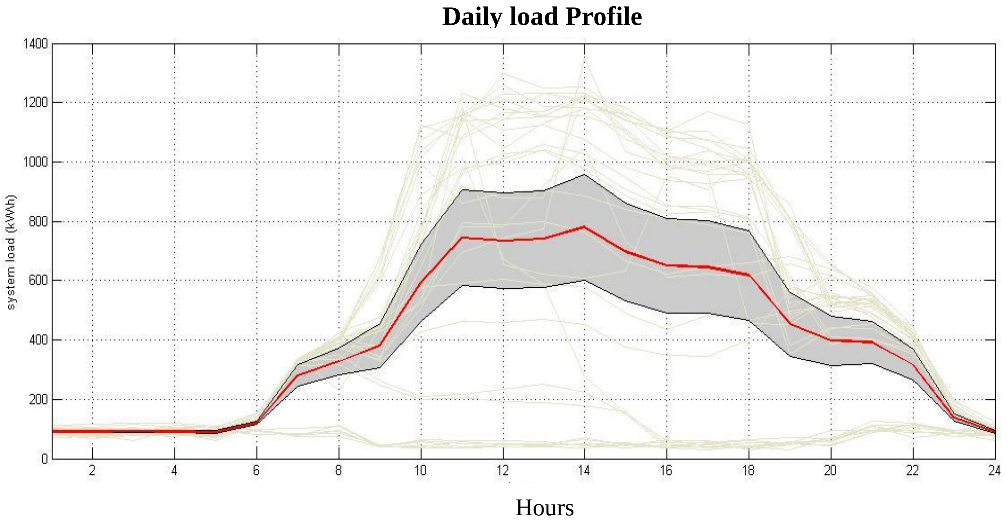

Figure 11: Daily Energy Profile of 'October' 2014, Faculty of Humanities and Social Sciences.

As stated before, the data interval makes possible to discover the pattern in energy use. The building has a broad peak, lasting from 11:00 AM to 6:00 PM approximately, during working days. The load is relatively flat at night and early morning, with an increase starting around 6:00 AM. The night-time loads drop at their minimum around midnight. The level of activity on campus is relatively low at night, yet the electricity energy is still high after 8 pm. There are certainly loads at night that cannot be avoided, but this night load can present opportunity for savings, for example reduction of lightening, use of different energy suppliers that are more efficient and also considering power saving in personal computers.

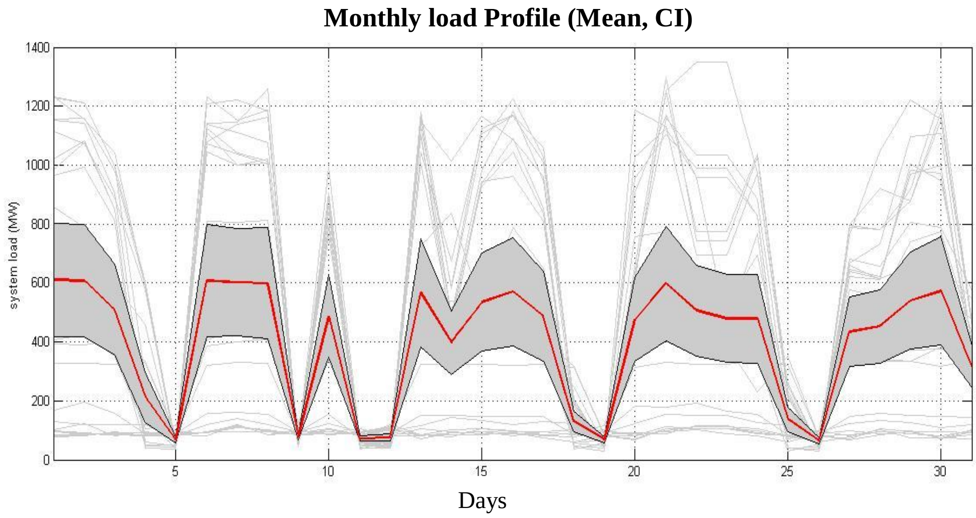

Figure 12: Monthly energy profile of 'October' 2014, Faculty of Humanities and Social Sciences.

The monthly profile (Figure 12) clearly shows the pattern during the working days and the weekend, the same discussed pattern for daily profile is repeated during the working days and the peak drop down sharply during the weekend.

From the two profiles, we could notice that the hour of the day and the day type of the week have an effect on the consumption. Generally, the operation time factor is one of the main drivers for the daily pattern. The consumption for each building during the night is almost the same. We can assume that this residual value is used for the operating machines at night, in the laboratories and servers' rooms.

4.1.2 Analysis of consumption regarding environmental factors

We also collected data about the outside temperature and humidity level with the wind speed since we wanted to investigate the relation between the consumption and the outside environmental factors.

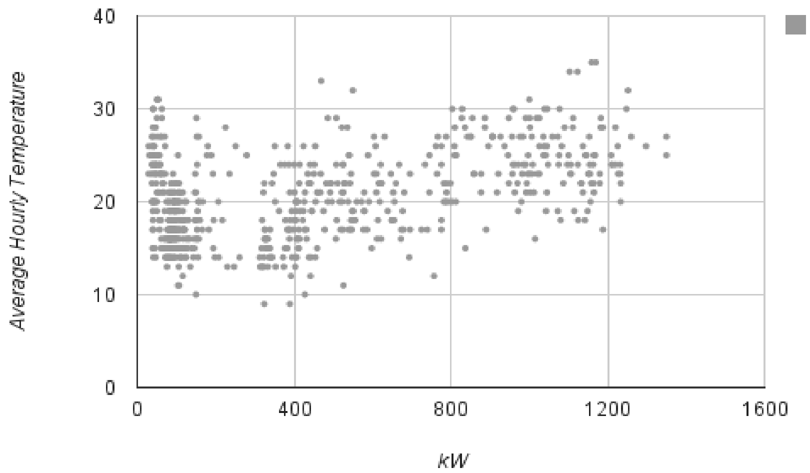

Figure 13: Correlation between hourly consumption (kW) and average hourly temperature for 'October' 2014

Figure 13 shows correlation of consumption with mean daily temperature for month 'October', with a correlation coefficient, **R = 0.59**. The goodness of fit is indicated by R Square, which would be equal to 1 in the case of a perfect fit. In this case, the value of R Square is **0.3481**, which represents a small goodness of fit.

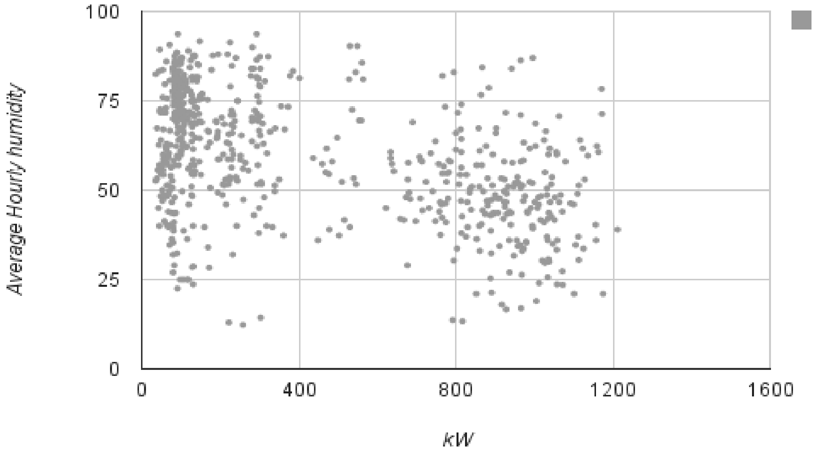

Figure 14: Correlation between hourly consumption (kW) and average hourly humidity for 'October' 2014

Figure 14 shows a negative correlation between the consumption and humidity as **R = -0.589**. This negative correlation indicates that the consumption is increasing when humidity level is decreasing. From the previous two figures, we can conclude that generally the outside environmental factors could be consider to have some effect on the energy consumption (Table 1).

Factor	Correlation coefficient R
Temperature	0.59
Humidity	-0.589

Table 1: Environmental factors correlation coefficient for 'October' 2014

4.1.3 Relation between the energy use and the occupancy level

Data regarding the schedule classes for both semesters were collected from the university web service and stored in a database. There are three types of classrooms in the university, Aula (Class), Laboratory and Seminar. Each one of them is behaving differently according to their size, type and subject. For example, laboratories for Chemistry school have different patterns from laboratories for Economy school. Figures 15, 16, 17 and 18 show the energy consumed, according to the type of classes and also the day type. By the observation of these Figures, it is noteworthy that the consumption is increasing with the number of classes and laboratories' occupancy level in general. The electricity level is depending on the building occupancy rate which is represented by the classes' number. The last graph clearly shows the relation between the day type (weekend, weekday, vacation) and the consumption: the consumption of energy mainly corresponds to the presence of occupants in the buildings.

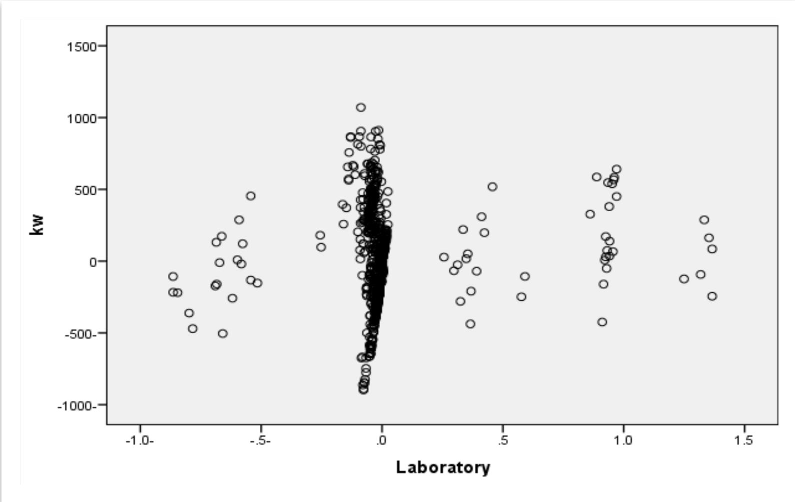

Figure 15: Correlation between classes type (laboratories) and the consumption (kW)

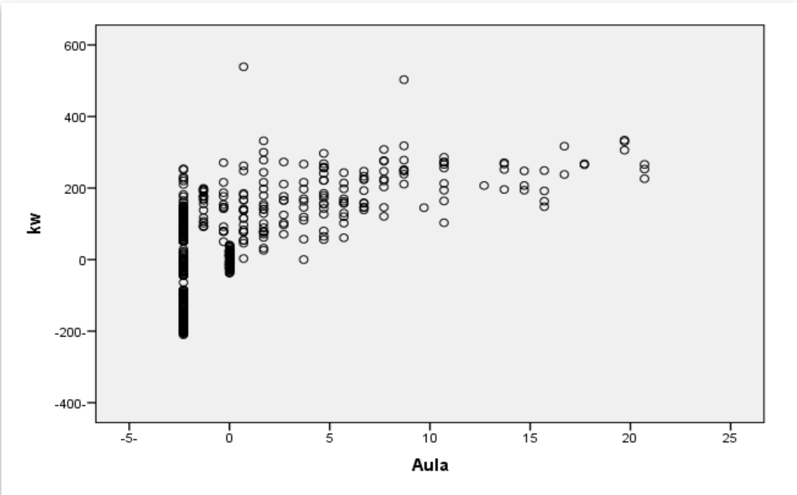

Figure 16: Correlation between classes type (Aula) and the consumption (kW)

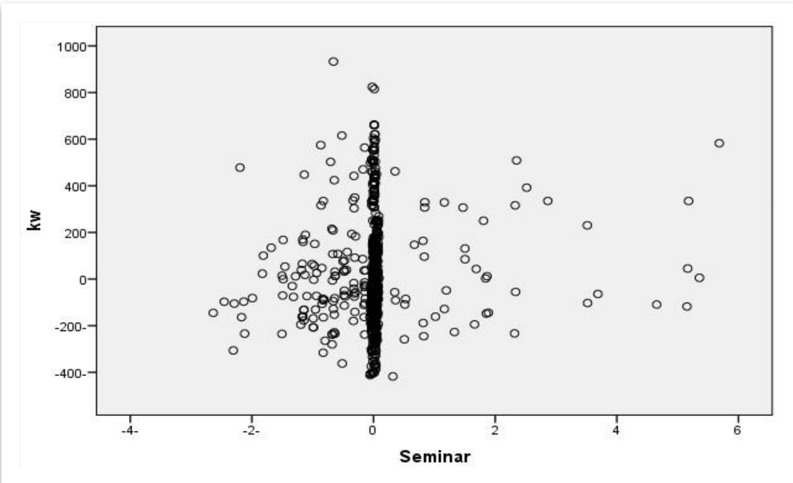

Figure 17 : Correlation between classes' type (seminar) and the consumption (kW)

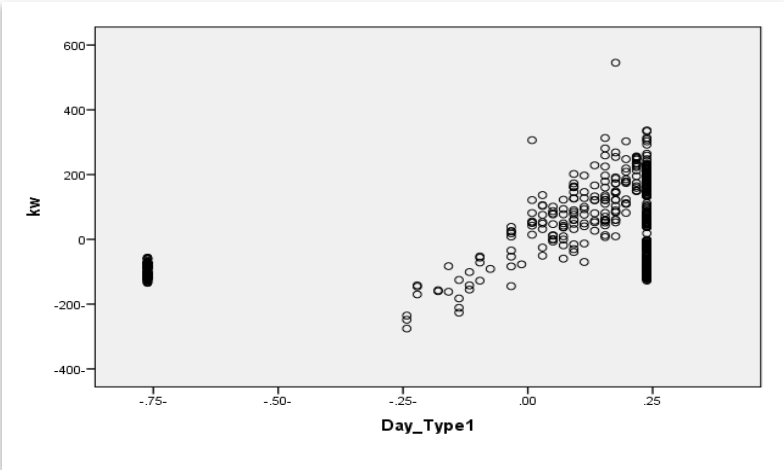

Figure 18: Correlation between day type and the consumption (kW)

Table 2 shows the correlation coefficients from the four occupancy level factors. From the four analyzed factors, it is important to note that the higher correlation is provided by the day type (**0.726**) and the lowest one is the class type 'Seminar' (**0.248**).

Factor	**Correlation coefficient R**
Class type 'Aula'	0.658
Class type 'Laboratory'	0.681
Class type 'Seminar'	0.248
Day type	0.726

Table 2: Occupancy level factors correlation coefficient

4.2. ENERGY MODELING

There are many factors that have a relation to the high energy consumption. The following goal is to know which one of them could be used as a predictor to build our prediction model. Several techniques can be used in order to build a prediction model. Generally, these models can be divided into two main categories: regression models and ANN models. Both models were built. Their description is presented in the following sections.

4.2.1 STEPWISE REGRESSION

Stepwise multiple regression can be used to answer the question 'what is the best combination of factors (independent variables) to predict the consumption (dependent predicted variable)?'. Stepwise regression takes all the predictors variables and uses them in the equation, one at a time, for each step and based on statistical methods. The predictor that contributes the most to the prediction, given its correlation coefficient value, R, is entered to the equation first. This step will be repeated if there are more predictors with high correlation. When no additional values exist which could add meaningful statistical relation to the dependent variable, the analysis stops, leaving the variables with low significance outside the equation.

From the above analysis, different independent variables can be used as a predictor for the consumption model.

For better modeling, we need to find the most significant factors among all of these independent variables. In this case, a stepwise regression technique was performed using the following factors:

- Outside temperature;
- Humidity;
- Wind speed;
- Number of classes;
- Number of Seminars;
- Number of laboratories; and
- Day Type [vacation, weekend, weekday].

The analysis was performed by statistical software SPSS. The dependent variable is the energy consumption (kW). It is important to mention that the data are normally distributed and it fit the condition for regression analysis. Table 3 shows the parameters that were entered to the model.

4.2.1.1 Stepwise Multiple Regression for the duration from April to October

Model	Variables Entered	Variables Removed	Method
1	Humidity		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Aula		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Temperature		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	Day Type		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
5	Laboratory		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
6	Wind speed		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

Dependent Variable: kW

Table 3: Variables that were entered/removed from stepwise model

A stepwise multiple regression was conducted to evaluate whether all the variables were necessary to predict the consumption. The stepwise regression works by entering each variable at a time. Variables that entered into the model can be removed at a later step if they are no longer contributing a statistically significant amount of prediction. Each step results in a model, and each successive step modifies the older model and replaces it with a newer one. Each model is tested for statistical significance. The variable for classes' type 'Seminar' was excluded from the analysis as it shows lower correlation with the consumption for this duration.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	57323894.39	1	57323894.39	492.352	.000[b]
	Residual	597745114.5	5134	116428.733		
	Total	655069008.9	5135			
2	Regression	85474667.22	2	42737333.61	385.135	.000[c]
	Residual	569594341.7	5133	110967.142		
	Total	655069008.9	5135			
3	Regression	104044916.3	3	34681638.78	323.010	.000[d]
	Residual	551024092.6	5132	107370.244		
	Total	655069008.9	5135			
4	Regression	118125291.1	4	29531322.78	282.199	.000[e]
	Residual	536943717.8	5131	104646.992		
	Total	655069008.9	5135			
5	Regression	120377309.6	5	24075461.91	230.988	.000[f]
	Residual	534691699.4	5130	104228.401		
	Total	655069008.9	5135			
6	Regression	121087775.1	6	20181295.85	193.846	.000[g]
	Residual	533981233.8	5129	104110.204		
	Total	655069008.9	5135			

a. Dependent Variable : kW
b. Predictors : (constant) , humidity
c. Predictors : (constant) , humidity , Aula
d. Predictors : (constant) , humidity , Aula , temp
e. Predictors : (constant) , humidity , Aula , temp , Day Type
f. Predictors : (constant) , humidity , Aula , temp , Day Type , Laboratory
g. Predictors : (constant) , humidity , Aula , temp , Day Type , Laboratory , wind

Table 4: ANOVA (Analysis of variance) for all 6 predictors

The 6 Analyses of Variance (ANOVA) that are presented in Table 4 correspond to the 6 models. Examining the last column of the output shown in Table 4, it is noteworthy that the final model was built in total six steps; each step resulted in a statistically significant model.

The column 'Degree of Freedom' (df) shows us that one variable was added during each step. We can also deduce that no variables were removed from the model, since for each step the count of predictors in the model is increasing from 1 to 6. This latter deduction is verified by Table 3 that tracks the variables that have been entered and removed. As can be seen, variables Humidity, Aula (classrooms), Temperature, Day type, Laboratory, Wind speed have been entered on Steps 1 through 6, respectively, without any variables having been removed on any step.

Finally we conclude from Table 4 that, by using the 6 variables together in the model, we will get the lowest mean square. This result indicates that all those 6 variables significantly influence the dependent variable (Energy consumption - **kW**). Therefore, it is recommended to use this combination of predictors to predict the consumption.
Table 5 presents the model summary which gives details of the overall correlation between the variables left in the model and the dependent variable (**kW**).

Model Summary[g]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.296[a]	.088	.087	341.217	
2	.361[b]	.130	.130	333.117	
3	.399[c]	.159	.158	327.674	
4	.425[d]	.180	.180	323.492	
5	.429[e]	.184	.183	322.844	
6	.430[f]	.185	.184	322.661	.193

a. Predictors : (constant) , humidity
b. Predictors : (constant) , humidity , Aula
c. Predictors : (constant) , humidity , Aula , temp
d. Predictors : (constant) , humidity , Aula , temp , Day Type
e. Predictors : (constant) , humidity , Aula , temp , Day Type , Laboratory
f. Predictors : (constant) , humidity , Aula , temp , Day Type , Laboratory , wind
g. Dependent Variable : kW

Table 5: Model summary for R and R Square values

Table 5 shows the model summary with values of correlation coefficient and coefficient of determination for all the models that were created during the whole process of stepwise regression. This table presents the R Square and adjusted R Square values for each step.

By examining Table 5, we can see in the footnote that humidity variable was entered into the model and the resulted R Square was (**0.088**). Next step, when Aula variable entered the model the R Square with both predictors has increased to (**0.130**) with gained value for R Square of (**0.042**). We can notice that R Square changed from one step to another and by the time we arrived to the last step, the R Square has value of (**0.185**).

Finally, we conclude that the last model, containing 6 variables, presents the highest multiple correlation coefficient ' R = **0.430** ' and coefficient of determination R Square. R represents the strength of the correlation between any two variables, showing a positive or negative relationship. Also, the standard error of estimation decreased when we used the 6 variables in the model.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	663.980	15.845		41.904	.000		
	humidety	-5.433-	.245	-.296-	-22.189-	.000	1.000	1.000
2	(Constant)	624.964	15.662		39.903	.000		
	humidety	-5.108-	.240	-.278-	-21.288-	.000	.993	1.007
	Aula	30.903	1.940	.208	15.928	.000	.993	1.007
3	(Constant)	178.466	37.283		4.787	.000		
	humidety	-3.254-	.275	-.177-	-11.837-	.000	.732	1.367
	Aula	35.708	1.943	.240	18.376	.000	.958	1.044
	temp	14.529	1.105	.198	13.151	.000	.726	1.377
4	(Constant)	82.704	37.722		2.192	.028		
	humidety	-3.283-	.271	-.179-	-12.098-	.000	.732	1.367
	Aula	32.796	1.935	.221	16.951	.000	.942	1.062
	temp	14.333	1.091	.195	13.140	.000	.726	1.377
	Day_Type1	130.672	11.265	.148	11.600	.000	.983	1.017
5	(Constant)	85.575	37.651		2.273	.023		
	humidety	-3.328-	.271	-.181-	-12.279-	.000	.731	1.368
	Aula	44.928	3.247	.302	13.839	.000	.333	3.003
	temp	14.329	1.089	.195	13.162	.000	.726	1.377
	Day_Type1	130.646	11.243	.148	11.621	.000	.983	1.017
	Laboratory	-47.340-	10.184	-.101-	-4.648-	.000	.339	2.952
6	(Constant)	111.131	38.881		2.858	.004		
	humidety	-3.631-	.295	-.198-	-12.322-	.000	.617	1.620
	Aula	44.497	3.249	.300	13.696	.000	.332	3.011
	temp	14.809	1.103	.201	13.421	.000	.706	1.416
	Day_Type1	131.229	11.238	.149	11.677	.000	.983	1.018
	Laboratory	-45.551-	10.202	-.097-	-4.465-	.000	.337	2.966
	wind	-2.935-	1.124	-.039-	-2.612-	.009	.715	1.398

Table 6: Coefficient table for all models built on combination of the six predictors.

Table 6 provides the results of the calculated coefficients. It is important to note that the standardized regression coefficients are readjusted at each step to reflect the additional variables in the model. The table shows the linear regression equation coefficients for the various model variables. From those coefficient values, a prediction equation can be easily derived for each model. Model 6 comprises all the predictors' values entered to the stepwise model. "B" values are the coefficients for each variable, i.e., they are the value which the variable's data should be multiplied by in the final linear equation. "Constant" is the equivalent interception in the equation. The Significance ("Sig." in Table 6) values should be **0.05** or below to be significant at **95** percent. A value of **0.000** means the value is too small for three decimal place representation.

The prediction equation should be presented as follows (Eq. 2):

y = constant + (v1 x coeff1) + (v2 x coeff2) + ...)………..Eq. 2

For the studied prediction model, Eq. 2 can be written as follows:

Predicted kW = 111.131 + (-3.631 humidity + 44.497 aula + 14.809 temp + 131.229 day type + (-45.551 laboratory) + (-2.935 wind)).

Humidity, Aula (classrooms), Temperature, Day type, Laboratory and Wind speed were used in a stepwise multiple regression analysis to predict electricity consumption (kW) with data for the whole duration of the study (7 months). As can be seen, all correlations except one for class type "Seminar" were statistically significant. The prediction model contained the six predictors and was reached in six steps with no variables removed. The model was accounted for approximately 19% of the variance of kW (R Square = **0.185**, Adjusted R Square = **0.184**). The consumption variable kW was primarily predicted by Humidity, and to a lesser extent by higher levels of Aula (classrooms), Temperature, Day type, Laboratory and Wind speed.

The standardized regression coefficients of the predictors together with their correlations with kW (Table 6) shows Humidity received the strongest weight in the model followed by Aula and Temperature.

4.2.1.2 Regression for each month

In order to understand the influencing factors for each month of the year, the same previous steps were applied separately for the study period, between April and October. Factors vary between each month (Tables 7 to 13; Figures 19 to 25, respectively to months April to October).

Model Summary[a,d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.658[b]	.433	.432	121.942	
2	.726[c]	.527	.526	111.432	.979

a. month= 4
b. Predictors: (constant), Aula
c. Predictors: (constant), Aula, Day Type
d. Dependent Variable: kW

Table 7: (April) Model summary for R and R Square values for each independent variable.

As described before in the previous section, Table 7 indicates that, for the duration of month April, the most significant predictors are Aula and Day type. Examining this table we can see that Aula variable was entered into the model and the resulted R Square was (**0.433**). Next step when Day type variable entered the model, the R Square with both predictors has increased to (**0.527**) and the estimated standard error decreased.
We can conclude that for month April, these are the most significant predictors with a model percentage reach to approximately **53%** of the variance of kW.
The ANOVA, Coefficients and excluded variables tables for month April are depicted in Appendixes A.

Figure 19 shows the correlation between each predictor (Aula, Day Type) for this month with consumption.

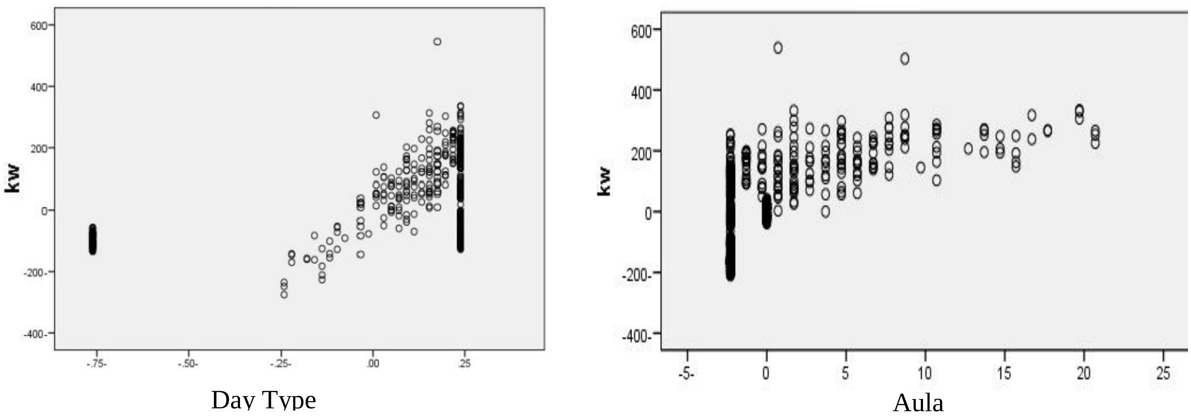

Figure 19: (April) Correlation between classes type (Aula), day type and the consumption (kW)

It is noted that the environmental factors are not significantly affecting the consumption during this month, and only factors related with occupancy are highly affecting the energy consumption. This could be due to the moderate temperature at this month.

Table 8 presents the model summary for the month of May.

Model Summary[a,g]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.430[b]	.185	.184	262.095	
2	.559[c]	.312	.310	240.981	
3	.639[d]	.409	.406	223.529	
4	.646[e]	.417	.414	222.089	
5	.648[f]	.420	.416	221.657	.453

a. month= 5
b. Predictors: (constant), Aula
c. Predictors: (constant), Aula, temp
d. Predictors: (constant), Aula, temp, Day Type
e. Predictors: (constant), Aula, temp, Day Type, wind
f. Predictors: (constant), Aula, temp, Day Type, Seminar
g. Dependent Variable: kW

Table 8: (May) Model summary for R and R Square values for each independent variable.

For May, four predictors are more significant than the rest of variables. These variables are Aula, Temperature, Day Type, Seminar. The lowest standard error and the highest R Square is defined by the combination of the 4 factors together (**0.420)**. We can say that these variables are contains **42%** of variance of kW (dependent variable). It also noticed that Class type Seminar has effect on the consumption at this month only.
In Appendix B ANOVA, Coefficients and excluded variables tables for month May are attached.

Figure 20 displays the correlation between the dependent variable kW and the predictors for this month Wind, Temperature, Seminar, Aula, Day Type respectively.

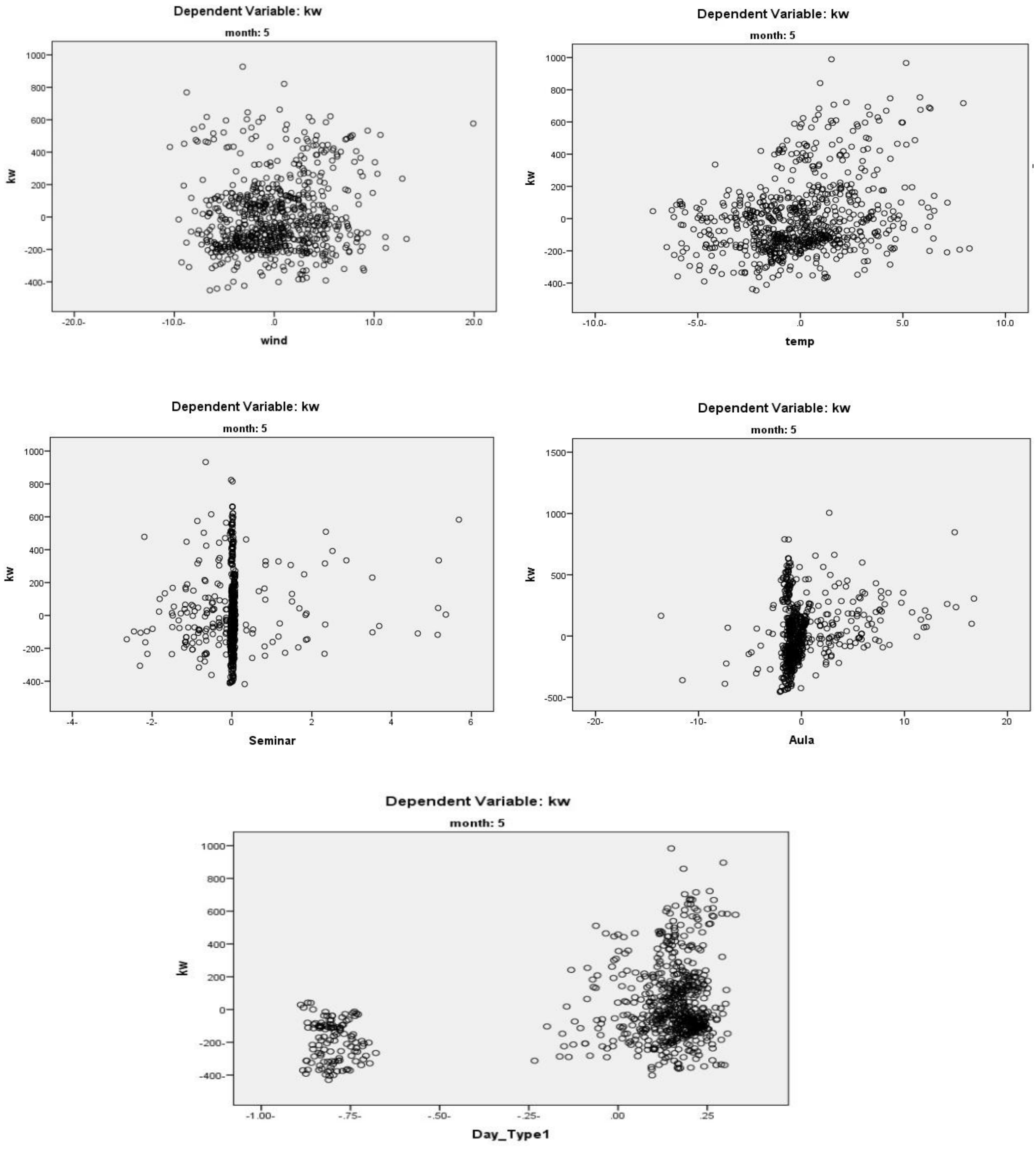

Figure 20: (May) Correlation between Wind, Temp, classes type (Seminar), classes type (Aula), Day type and the consumption (kW)

Table 9 portrays the modelling results for the month of June.

Model Summary[a,f]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.443[b]	.196	.195	345.378	
2	.547[c]	.299	.297	322.800	
3	.571[d]	.326	.324	316.620	
4	.589[e]	.347	.344	311.891	.265

a. month= 6
b. Predictors: (constant), temp
c. Predictors: (constant), temp, Day Type
d. Predictors: (constant), temp, Day Type, wind
e. Predictors: (constant), temp, Day Type, wind, humidity
f. Dependent Variable: kW

Table 9: (June) Model summary for R and R Square values for each independent variable.

In June, the environmental factors are statistically significant: temperature, wind, humidity and day type. This combination of predictors is estimated with **34%** of variance of kilowatt. In Appendix C ANOVA, Coefficients and excluded variables tables for month June are attached.

Figure 21 shows respectively the partial plot for the predictors of month of June (temperature, humidity, wind , day type).

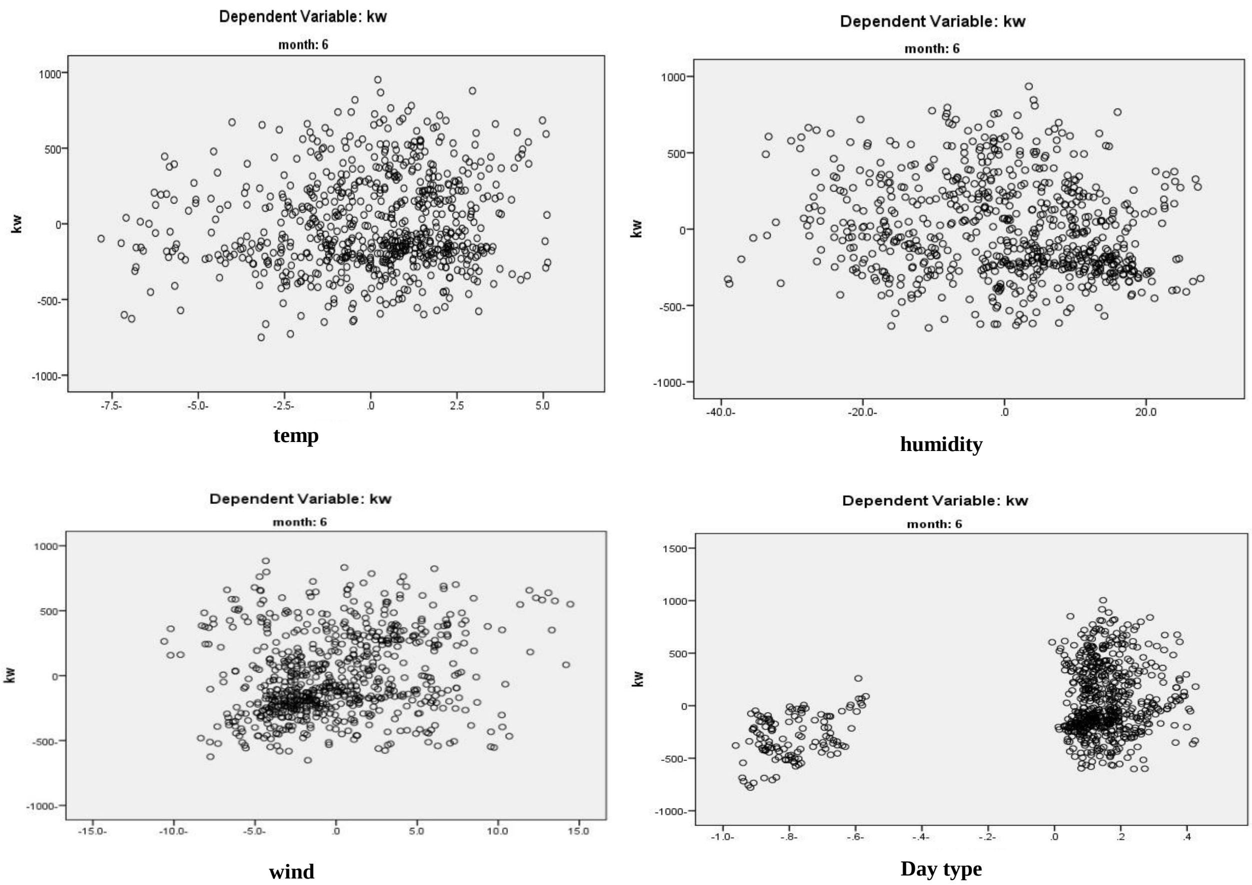

Figure 21: (June) Correlation between temp, humidity, wind, day type and the consumption (kW)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.492[b]	.242	.241	324.100	
2	.506[c]	.256	.254	321.231	.210

a. month= 7
b. Predictors : (constant) , humidity
c. Predictors : (constant) , humidity, Day Type
d. Dependent Variable : kW

Table 10: (July) Model summary for R and R Square values for each independent variable.

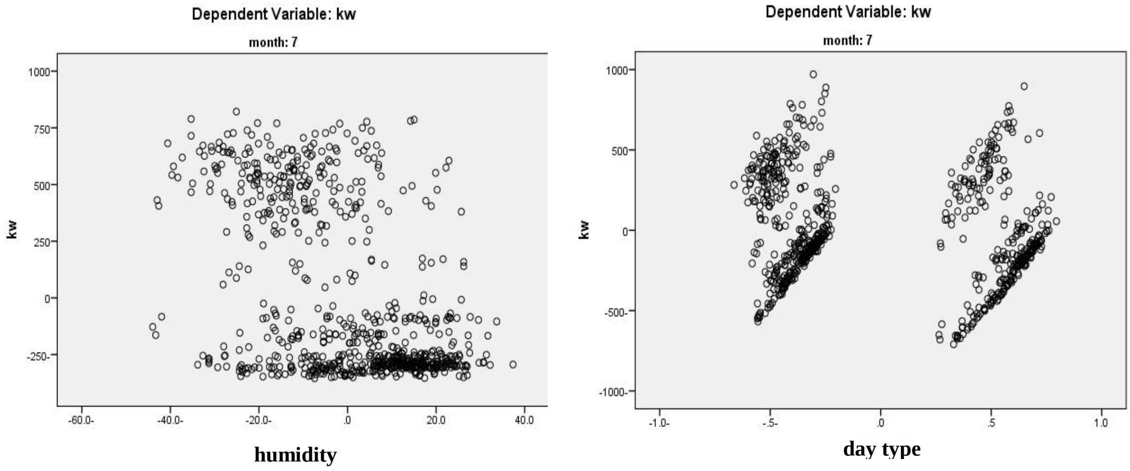

Figure 22: (July) Correlation between day type, humidity and the consumption (kW)

The month July is only affected with two variables, humidity and day type, the weighted combination of these variables are explaining approximately 25% of the variance of kW (Table 10). It is worth noticed that although this month has a high average temperature but it has no effect on the consumption due to the low occupancy level at this month "end of the semester" (Figure 22).
The related tables for ANOVA, Coefficients and excluded variables for month July are in Appendix D.

Table 11 explains the model summary of August, and the Figure 23 shows the correlation graphs with its predictors' variables.

Model Summary[a,d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.166[b]	.027	.026	152.299	
2	.270[c]	.073	.071	148.793	.199

a. month = 8
b. Predictors: (Constant), wind
c. Predictors: (Constant), wind, temp
d. Dependent Variable: kw

Table 11: (August) Model summary for R and R Square values for each independent variable.

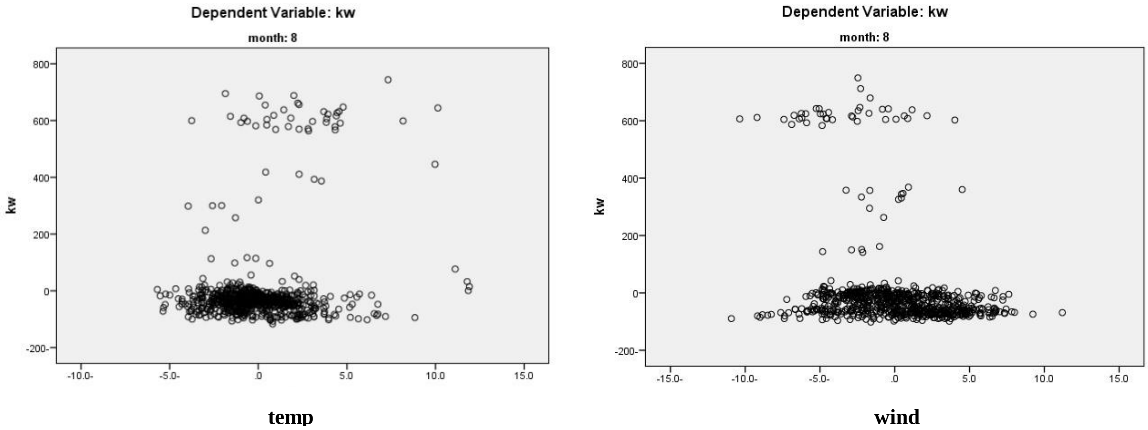

Figure 23: (August) Correlation between temp, wind and the consumption (kW)

We could notice the low correlation between the consumption and the temperature at this month. Also the fact that no other predictors is influencing the consumption during this month, is due to the vacation. All the university is closed.

Model Summary[a,f]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.474[b]	.225	.224	415.682	
2	.604[c]	.365	.363	376.541	
3	.679[d]	.461	.459	347.201	
4	.681[e]	.464	.461	346.389	.544

a. month = 9
b. Predictors: (Constant), humidety
c. Predictors: (Constant), humidety, Aula
d. Predictors: (Constant), humidety, Aula, Day_Type1
e. Predictors: (Constant), humidety, Aula, Day_Type1, Laboratory
f. Dependent Variable: kw

a. month= 9
b. Predictors : (constant) , humidity
c. Predictors : (constant) , humidity, Aula
d. Predictors : (constant) , humidity, Aula , day type
e. Predictors : (constant) , humidity, Aula , day type, laboratory
f. Dependent Variable : kW

Table 12: (September) Model summary for R and R square values for each independent variable.

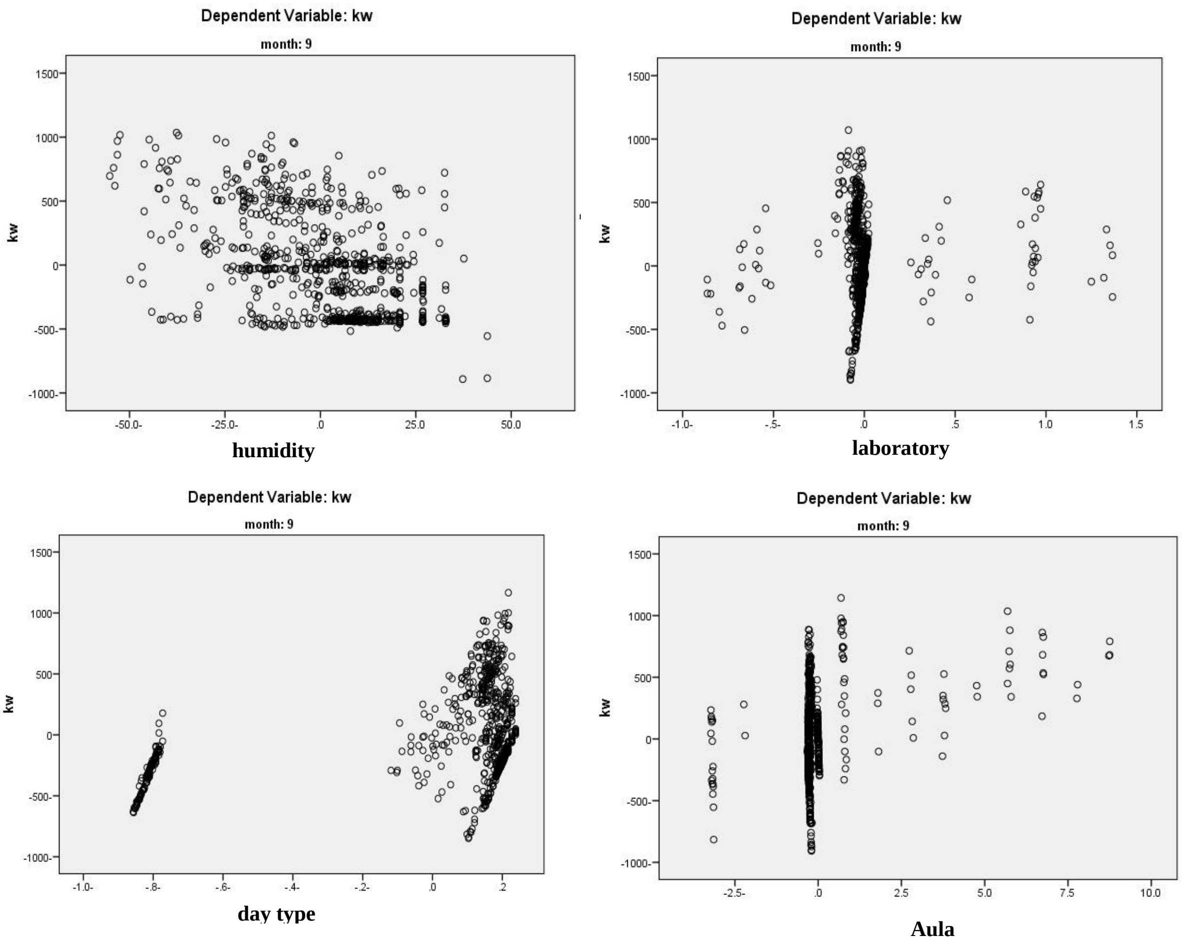

Figure 24: (September) Correlation between humidity, class type (aula and laboratory), day type and the consumption (kW)

In September, as it is the beginning of the semester, the intensity of classes is high. Also the temperature and humidity is relatively high, thus the model with humidity, class type (Laboratory and Aula) and day type as a predictors (Table 12), has the highest R Square and it is approximately 46% of the variance. Correlations between the variables and the energy consumption are presented in Figure 24.

Table 13 presents the model summary for the month of October.

Model Summary[a,g]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.474[b]	.224	.223	332.478	
2	.624[c]	.389	.388	295.170	
3	.705[d]	.496	.494	268.232	
4	.700[e]	.501	.490	267.240	
5	.710[f]	.504	.501	266.580	.726

a. month = 10

b. Predictors: (Constant), temp

c. Predictors: (Constant), temp, Aula

d. Predictors: (Constant), temp, Aula, Day_Type1

e. Predictors: (Constant), temp, Aula, Day_Type1, Laboratory

f. Predictors: (Constant), temp, Aula, Day_Type1, Laboratory, wind

g. Dependent Variable: kw

Table 13: (October) Model summary for R and R Square values for each independent variable.

Examining this table and the below graphs, we deduce that temperature, Aula, laboratory, wind, and day type variables were entered into the model separately and the resulted R Square for model number 5 was the highest with (**0.504**).

Finally, we conclude that the last model, containing the 5 variables, presents the highest multiple correlation coefficient ' R = **0.710** ' and coefficient of determination R Square. R represents the strength of the correlation between any two variables.

By looking to the fact that in October and September is considered the beginning of the semester, thus the occupancy level is one of the main driver for this month.

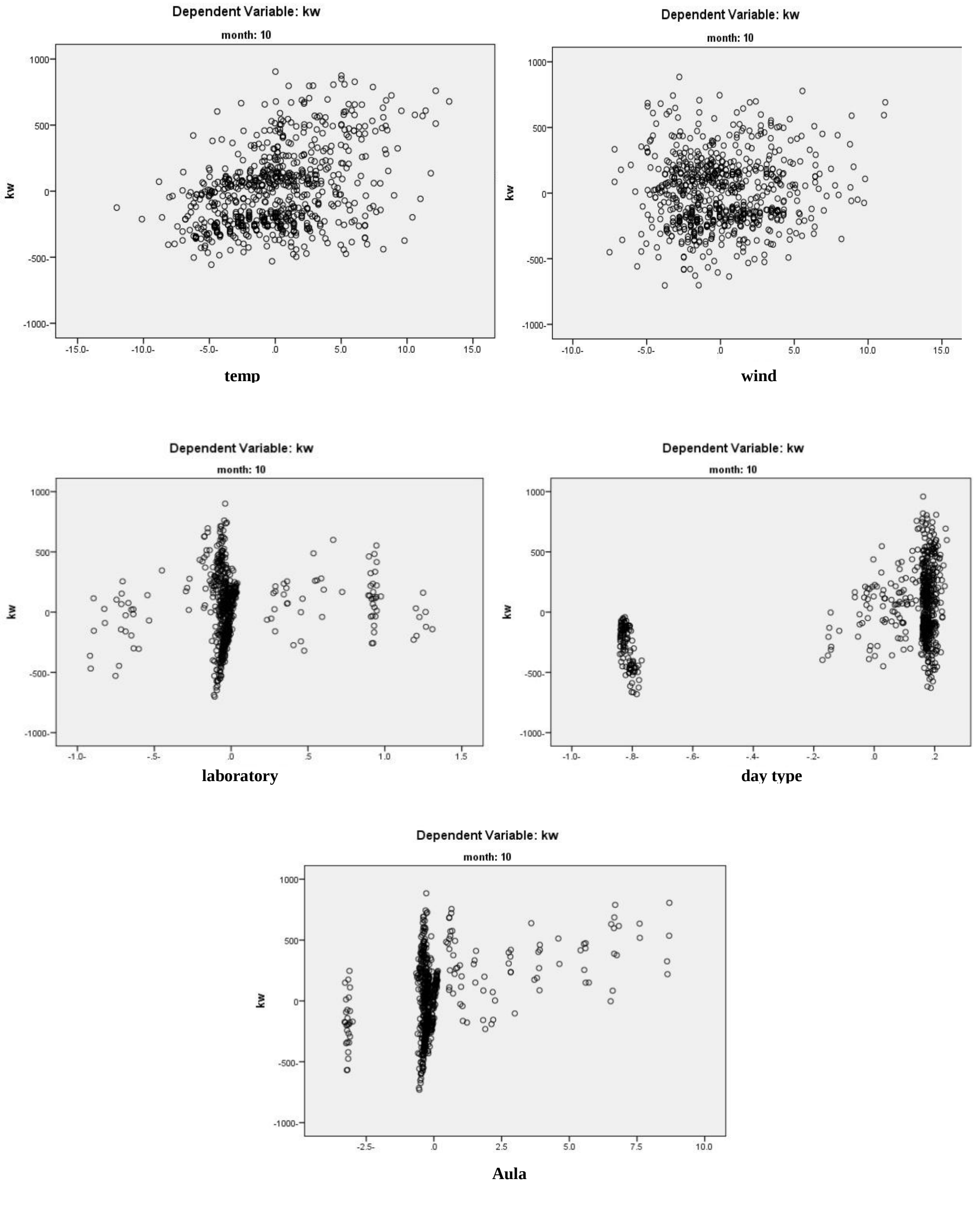

Figure 25: (October) Correlation between temp, wind, class type (laboratory), day type, class type (Aula) and the consumption (kW)

The ANOVA, Coefficients and excluded variables tables for month August, September, October are depicted in Appendixes E,F,G respectively.

4.2.2 Neural Network

In this section, we will build a prediction model using the neural network technique. The created ANN will later be compared with the results from the regression analyses undertook in the previous sections, and the best prediction model among them will be used. For the ANN study, a feed forward neural network model will be used.

Figure 26 shows a simple design of a neural network feed forward, which consists of three main layers: input, hidden and output. The input layer accepts patterns from the environment, while the output layer is holding the response to the environment. One or more hidden layers connect the input layer and output layer together.

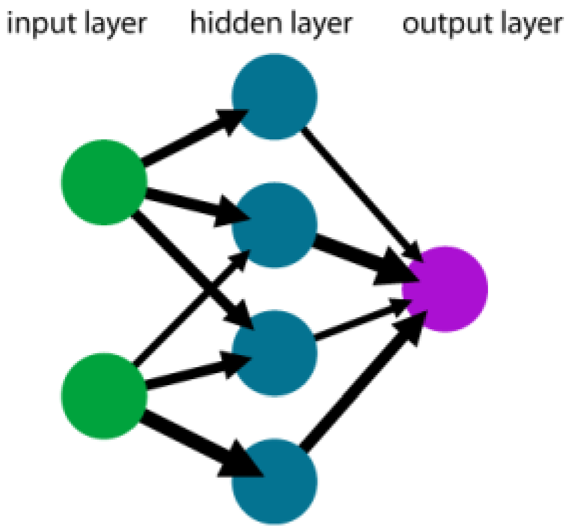

Figure 26: "Neural network example", (Source: Wikimedia)

The ANN is considered as a black box where the accuracy of the output is the only indicator of the quality of the process. ANN learn how to map between input and output through the process of iterative training. An advantage of ANN is the updating of the learned knowledge over time which makes the system inherently adaptive.

In this model we will use the back propagation technique. Back propagation is a training algorithm designed to minimize the mean square error in mapping between the output and the current value.

4.2.2.2 Data set, input and output

As a first step, data preprocessing were performed on the collected data. The interval data for smart meters reading for energy consumption was 15 minutes, while for the temperature, humidity and wind observations were for 30 minutes. Both data set were aggregated, averaged respectively to one hour interval. The data preprocessing consisted of the following steps:

- The timestamp was separated to Day, Month, Year, and Hour;
- The day of the week was represented with sequence numbers from [1 - Monday till 7 – Sunday, sequentially].

- The data related with the classes were also aggregated to one hour interval for each type of the class (corresponding to the same period interval as the reading from the meters);
- The data about the university calendar were gathered and classified into three classes: Weekend, Weekday, and Vacation, respectively 0, 1, and 2.

Table 14 presents an example of the input data.

month	date	hr	temp	wind	humidity	dayofweek	Aula	Laboratory	Seminar	Day Type	kw
4	1	8	7.666666667	2.333333333	88.66666667	2	4	1	0	1	835
4	1	9	9.666666667	1.566666667	83.33333333	2	18	4	2	1	818
4	1	10	13.66666667	1.933333333	69	2	7	0	0	1	863
4	1	11	16	0	67.66666667	2	23	2	3	1	1198
4	1	12	16.66666667	1.566666667	61.33333333	2	3	3	0	1	1131
4	1	13	18	4.633333333	56	2	11	6	3	1	1105
4	1	14	17.66666667	6.133333333	57.33333333	2	0	0	0	1	1025
4	1	15	18.33333333	5.4	53.33333333	2	8	4	1	1	1028
4	1	16	18	11.93333333	53.66666667	2	2	0	0	1	946
4	1	17	17	9.6	58.33333333	2	7	2	0	1	916
4	1	18	17	8.466666667	53.66666667	2	3	0	0	1	645
4	1	19	16	7.666666667	69.33333333	2	2	0	0	1	269
4	1	20	15	6.166666667	76	2	1	0	0	1	291
4	1	21	14.66666667	3.833333333	76	2	0	0	0	1	329
4	1	22	14	3.2	85.75	2	0	0	0	1	135

Table 14: Sample of ANN input data.

The data set for 7 months (April to October) contains **5137** instant points for **13** element. The main structure of the network comprises **12** input parameters and **1** output.
Data was split into three data sets, Table 15 shows the data sets size for the neural network.

Process	Instant	Data size
Training	3596	70% of the data used for training
Testing	770	15% of the data used for testing
Validation	770	15% of the data used for validation

Table 15: Percentage of training and testing data for ANN

It is important to mention that only seven months of data were analyzed, since there was a large amount of missing data from the remaining months (November to March). This fact was due to the use of a different measurement system in those months. Additionally, some of the energy consumed between November and March, in the heating system, was provided by gas. This second fact prevented the use of the data in the prediction model, since gas consumption was not periodically measured by smart meters, it could not be included in the model.

In order to begin modeling, we need to choose the right specification for the network, which includes the number of hidden layers, the number of neurons in each layer and the input and output. There are many general techniques for assisting in choosing those parameters, but it is hard to say that these techniques are accurate all the time.

There is no way to determine the best number of hidden units before starting the training process and do some experimentation and monitoring the error each time. Usually if there are few hidden layers, then the error resulting from the high training and generalization will increase. This is due to the under fitting and high statistical bias. In the opposite case, it could still have a generalization error due to over fitting error (German et al., 1992).

First we tried with one hidden layer, and based on the training results, we increased this number. Table 16 contains the specification of the undertaken tests.

Networks	input	Number of hidden layer	Number of nodes	output	R value
1	6	1	6	1	0.48
2	6	1	10	1	0.5
3	12	1	9	1	0.88
4	12	1	6	1	**0.9721**
5	12	2	18	1	0.82
6	12	2	10	1	0.85

Table 16: Training networks specification.

Figure 27 shows the network architecture implemented with minimum error.

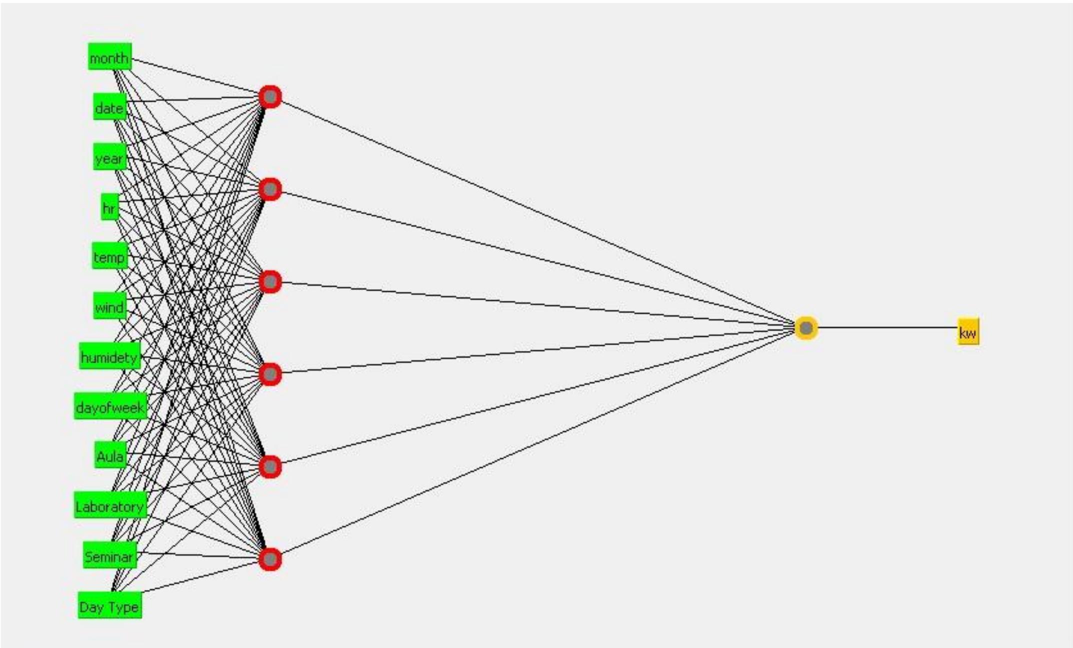

Figure 27: Implemented Neural network model.

In our model of multi-layer perceptron (MLP), the neurons on the input layer are the input variables that represent the outside conditions. In this case, all the 12 variables are representing the outside environment. The middle hidden layer is connected to the input by weighted input. The hidden layer later processes those weighted values by summing them and then performing a nonlinear transfer function. Those functions allow the ANN to map nonlinear relationship between the input and output, the hidden layer and the output is connected by the weighted values computed by the hidden layer.

According to our training process the network **4** showed better results in prediction with R= 0.9721. Thus, this configuration of the network is the most suitable.

The result is shown in Figure 28.

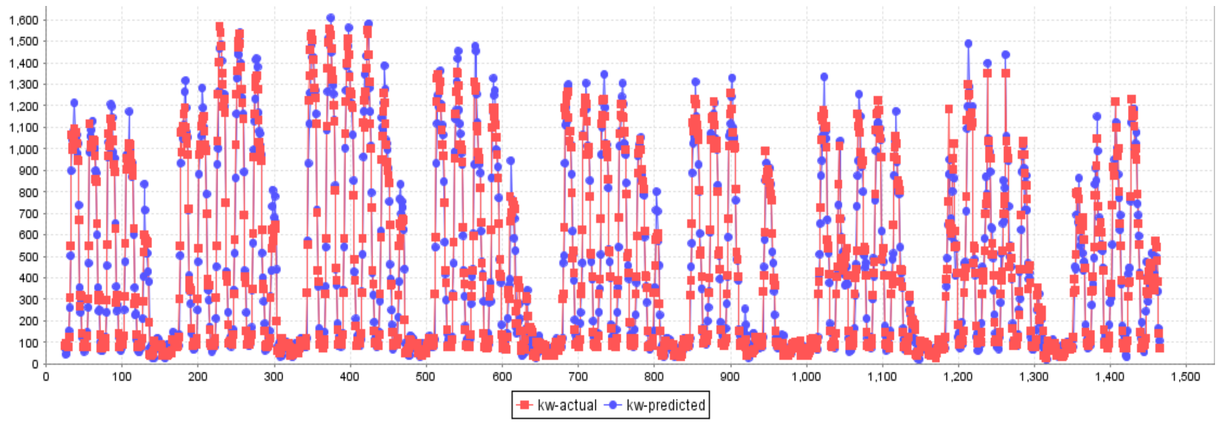

Figure 28: Predicted value and current values for ANN model.

For the network 4, the value of mean absolute percentage error is **3.9363** (MAPE is defined as the difference between the expected output and the calculated output by ANN).

4.3. RESULT AND DISSCUSSION

From the previous analyses, it can be concluded that the consumption in the university is following a daily and weekly pattern. The regular daily pattern is represented by the daily operational hour, whereas the peak of the consumption starts at 8:00 AM and lasts till 8:00 PM. It can also be concluded that weekends and vacations have different profiles than working days.

The analyses showed a correlation between the environmental factors such as temperature, humidity, wind-speed and the energy consumption. However, when we performed monthly analysis, we noticed that the effect of these factors is varying from one month to another due to the change in seasonality. Another factor is the type of the classes that operate inside the buildings. For example, the class type “Aula” and “Laboratories” are affecting the consumption more than “Seminar”.

When the stepwise regression analysis was performed on the whole duration 7 months, the result shows: humidity, aula (classrooms), temperature, day type, laboratory and wind speed variables were used in a stepwise multiple regression analysis to predict electricity consumption (kW) with data for the whole duration of the study 7 months. The correlations of the variables are shown in Table 1 and 2. As can be seen, all correlations except one for class type Seminar were statistically significant. The prediction model contained the six predictors and was reached in six steps with no variables removed. The model was accounted for approximately **19%** of the variance of kW (R Square = **0.185**, Adjusted R Square = **0.184**). The consumption variable kW was primarily predicted by humidity, and to a lesser extent by higher levels of aula (classrooms), temperature, day type, laboratory and wind speed.

The standardized regression coefficients of the predictors together with their correlations with kW in Table 6 shows humidity received the strongest weight in the model followed by aula and temperature.

In April, the most significant factors are the number of classes and the day type which represent the occupancy effect. However, the outside environmental conditions have no influence on the consumption at this month, in the opposite to June when the temperature, humidity, wind, and day type are significantly affecting the consumption. Table 17 presents a summary of the factors that substantially influence in each month period.

Month	Predictors	Percentage
April	Number of classes, Day type	53%
May	Number of classes & Seminars, Wind speed, Outside Temperature, Day type	42%
June	Outside Temperature, Wind, Humidity , Day type	345
July	Humidity, Day type	25%
August	Outside Temperature	27%
September	Humidity, Classes, Day type , Laboratories	46%
October	Outside Temperature, Classes, day type, laboratories, wind speed	50%

Table 17: Influencing factors for each month

The results provided by the stepwise regression analysis indicate that in order to predict the consumption, the combination of all 6 variables under study is the best way to model the phenomena. This analysis could be done with different temporal resolutions. The analysis undertaken only with the monthly classification could be insufficient to have a profound understanding of the consumption.

Although the prediction model for the whole 7 months has accounted for approximately **19%** of the variance of kW, it is noticed that the percentage is different from one month to another according to the predictors influence. That means we could use different model for each month for better prediction.

We also validated this result by building a neural network model for prediction of energy consumption. The model was trained and validated using data from 7 months and the result was acceptable with R= 0.9721 and percentage of error **3.9363**.

The regression model showed a significant result per month compared to the resulted percentage of the whole duration (7 months). The ANN model showed better result for the whole duration.

Chapter Five: Visualization and Exploration

In this chapter a further step is performed for improving the energy facilities management in campus. A real time energy visualization web application for buildings location in campus was developed.

5.1 INTRODUCTION

The readily available and easily accessible information visualization in a real-time feedback, that becomes available through the use of ICT, is resulting in a huge reduction of energy consumption in several places. Since the human ability to interpret image information is better, thus using visualization to communicate information and data is a very effective way and easier to assimilate than text.

According to Fischer et al. (2008), one of the most effective policies is to promote the design of technologies that rely on information visualization as a way of changing behavior toward the energy consumption. Designing such informative tool should follow some rules: it should be based on a given frequent and actual consumption and a longer period information with the ability of normative comparison. EPRI research (2009) suggests that a feedback system can effectively encourage conservation with a reduction rate of 18%.

Basically, feedback technologies consist of enhanced monthly usage with various information such as energy tips and comparison with peers. Ayres et al. (2009) show how significantly from a large experiment filed a peers comparison could create a highly net cost and carbon saving. Different features also used in designing systems are timeliness, data display, interactive contents and sociability. All those features combined with a wide range of technologies that varies in complexity and accessibility such as computers, mobile and portable devices or even public displays helps in reduction and sustainability promotion.

In addition to the use of energy feedback visualization in promoting sustainability, it is also considered as a way for information management to help decision making. Several successful projects used the advantage of easily accessible technology and the wide spread of sensors to manage the energy through applications.

For monitoring the energy consumption, the university has deployed a smart metering system in the campus. This system is allowing the management to retrieve data about buildings consumption whenever needed. The management is providing the aggregated whole campus consumption data through the university website[18]. The part concerning the energy data could be enhanced by a visualization technique which will rely on the many electricity meters deployed in the UJI's campus. Apart from the meter visualization of the consumption measured, the added value of highlighting variation in the consumption can be incorporated, thus reducing the need for a specific and complex system to be developed or bought for energy management.

The University Jaume I (UJI) has followed several ways towards sustainability enhancement in the campus. One of this projects is UJI's Smart Campus, which is

[18] http://www.energia.uji.es/

devoted to create a system to query and visualize information from several university sources, including information about the personnel, locations, services and energy data, in a unified and homogenous way.

A real time feedback that could reflect our behavior on the campus will help in raising the awareness for GHG reduction and assist the management in energy reduction goal by providing a quick real time view on the main campus key performance indicators for energy resources.

Thus, a Web GIS application were developed to represent the main key performance indicator of energy, providing real time feedback for consumption and highlighting the variation through time.

5.2 System Analysis and Design:

5.2.1 Analysis

In this step we defined the requirements, considering the nature of the roles that this Dashboard has to cover. The management already has their system for monitoring the energy, these data is saved in a database. It is only accessed whenever they need to have a look on the consumption for any specific reason.

The data is provided in a numeric list view like Excel sheet, the user can select the meters he needs to view, and then the list of readings appears. With such system it is difficult to have an overview of all campus buildings, the need to visualize the whole campus performance and each buildings' performance separately could be a help to the managers.

Generally dashboard should display visually the most important information which fits entirely on a single computer screen (information dashboard design, 2006), summarize events that running in your place. The main goal is to define exactly what it is necessary to display and the main indicators in the organization that it is essential to keep tracking it.

There are many ways that data can be collected and displayed, it is only depending on the amount of data that needs to be displayed and how it should be displayed.
Energy dashboards can display a wide range of information, including whole building energy consumption and various metrics, such as kilowatts of electricity consumption per area. Usually line, bar and pie charts visualization components are used in dashboard, in addition to maps.

By analyzing the management current system and the smart campus platform, a dashboard web application can be integrated easily with these systems. The main functionality will be as followed:

- Visualizing the current campus energy performance during the whole time, giving indicators about this performance level according to the expected value for each indicator.
- Allowing the user to get information about energy efficiency for any individual building in campus by using interactive map.

- Providing the user with the ability to investigate several buildings performance and make comparisons between them.
- Providing the user with information about each building level of efficiency, this measured according to the variation level of consumption than the expected one.

The system operation is very simple, consisting of a client-server architecture. The server side is already provided by the INIT, the next step will be to create the new services and hosting it on this server. The architecture that will be used is three-tier architecture Figure 29, this architecture was developed by John Donovan and consists of:

- Data tier: Where the data are stored and could be retrieved anytime by logic layer.
- Logic tier: Also called Business tier, it is the code/script that is executed to perform specific functionality of the application like performing calculations, coordinates, processes and communicate between the other two tiers.
- Presentation tier: is the user interface that is responsible for performing the tasks and understandable presentation to the user.

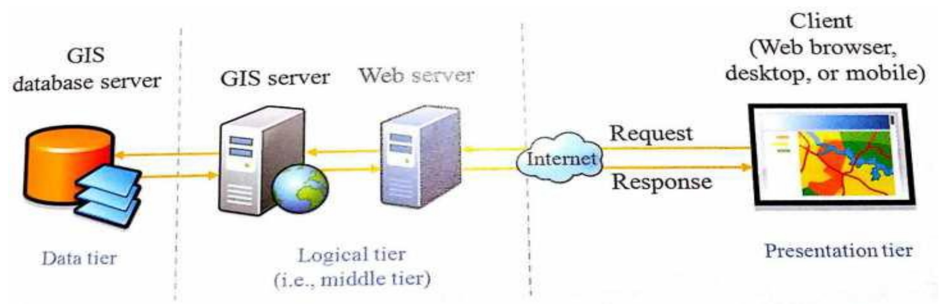

Figure 29 : Three-tier architecture

In this case the data tier contains the Geodatabase that holds all the campus maps and layers. Also, it will contain the database for energy consumption. The logical tier has the ArcGIS server that host the smart campus services this server will be used to publish the needed layers for maps visualization. The presentation tier is the dashboard interface and interaction with the user.

5.2.2 Design

The dashboard view should contain a map with several charts and indicators. By following the previous work and literature, different types of chart and visualization will be implemented in the dashboard. Those charts are selected according to the data and the purpose of visualization. As this tool's goal is information provision, thus the next charts will be used.

Line charts: This kind of charts is used to plot values as a series of points and then connect those point with lines. This type is useful in demonstrating the pattern, trend and fluctuations behind the data. Thus it will be used to visualize the current consumption data series for the day, for the main three indicators (Gas, Water, and Electricity).

Bar charts: This is used to compare variables which are related categorically. This will be useful in comparing buildings consumption for longer scale.

Gauge charts: The gauges are used to show the progress toward specific goal or value. This will be used to visualize the maximum target for the consumption that not allowed to excess, it is powerful in tracking single metric.

Circular Heatmap: These charts are usually used to visualize pattern in cyclic data; in this case, it will be used to visualize the weekly consumption for specific building showing the intensity of the consumption.

An initial prototype user interface (Figure 30) was designed with the main components, the map where the buildings will be displayed on campus base map, the charts and the main three indicators in the system electricity, gas and water.

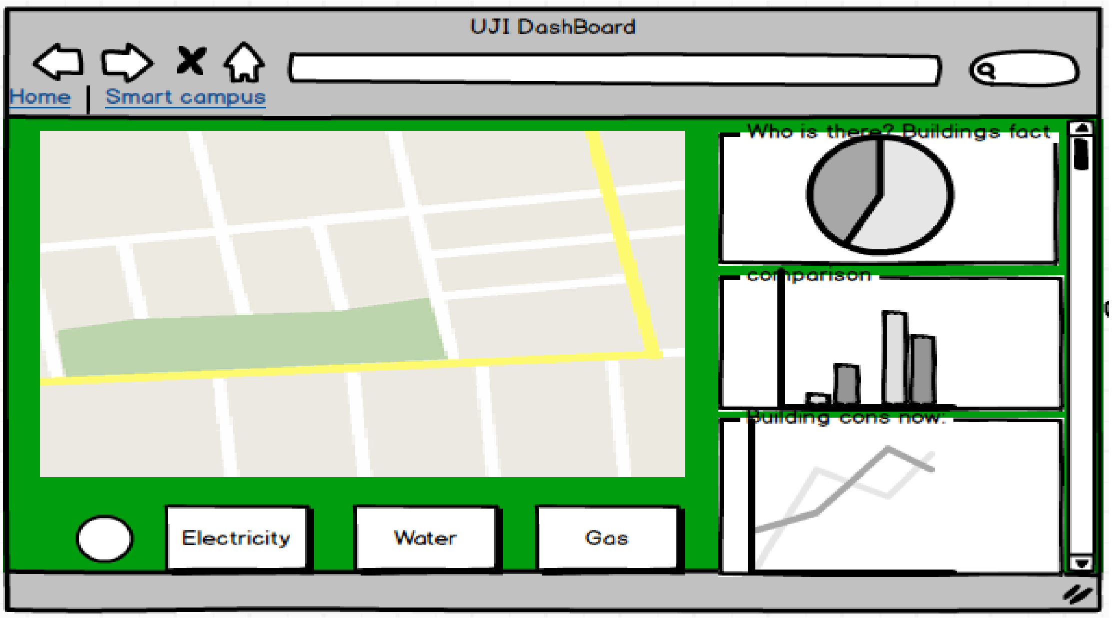

Figure 30: Initial prototype for user interface

5.3 Implementation

In this step we will introduce the main technologies that were used to develop the dashboard (Figure 31) .

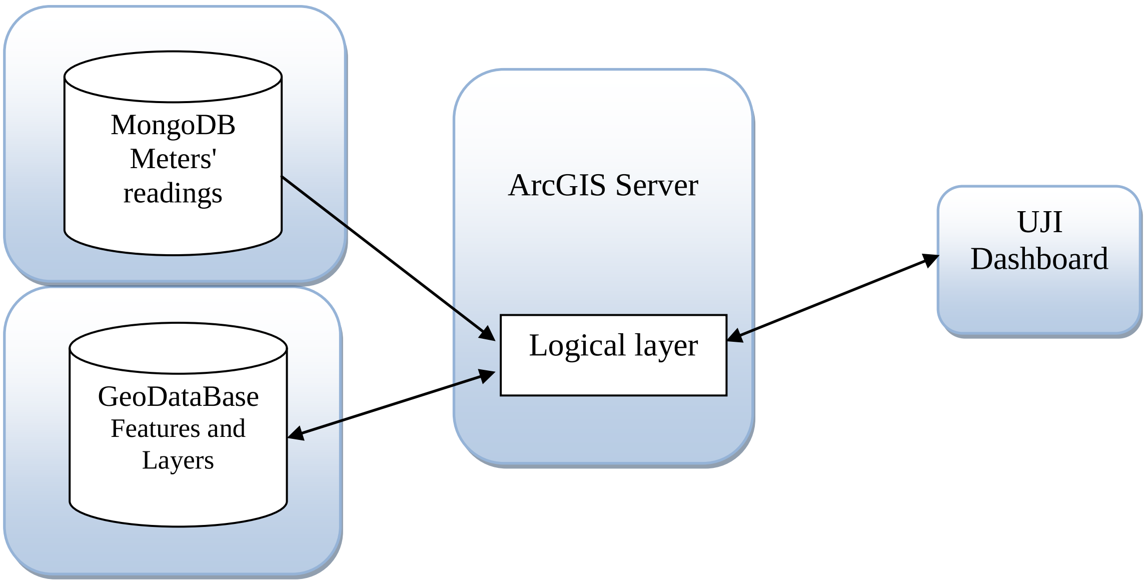

Figure 31: System architecture for the dashboard

The GIS server in UJI contains maps and services that will be handled to the application. We need to define exactly what data will be needed and how they will be treated in the system. The data is distributed on different servers, one of the main part is the connection between the energy management system and the application, the data for meters reading is stored in different servers in a database.

5.3.1 Technologies:

As shown on the previous figure the server is ArcGIS server, thus we used ArcGIS JavaScript API[19] to implement the client side. This API has implemented functions that handle the interaction with maps.

Javascript with HTML and CSS were the main core in the development of the user interface. Bootstrap[20] frame work was used to design the interface, since this application could be accessed from any mobile or tablet. This framework is used to implement a responsive applications.

We used D3[21] library for creating the charts. This library provide reach interactive graphs. In the backend, the main functionality was handled by JavaEE. For data connection in the backend side, we also used MongoDB Java Driver [22].

---

[19] https://developers.arcgis.com/javascript/
[20] http://getbootstrap.com/
[21] http://d3js.org/
[22] https://github.com/mongodb/mongo-java-driver/releases

5.3.2 Data

The data needed for the system as it shown in the Figure 31 is from two different sources. The Geodatabase, holds the information of each building, this information is needed for the visualization such as building total area. Thus we published campus building layers as a feature service, this service will be updated regularly with the latest consumption. We also published web map service to be used for visualizing the consumption for each building normalized by the building area kW/m^2.

The other source for data is MongoDB where consumption data are stored. Each record in this database is stored as document, the document schema is similar to JSON format. For every meter in the system, each instant is represented by a document in the database. Figure 32 shows example of one instant.

```
db.events.findOne({"_id" : ObjectId("5433caaa0cf26c324a7d993a")})
{
        "_id" : ObjectId("5433caaa0cf26c324a7d993a"),
        "ref" : "72689676-DC3E-4602-90FF-1B07D311E599",
        "sts" : ISODate("2014-10-07T10:45:00Z"),
        "rts" : ISODate("2014-10-07T10:45:48.394Z"),
        "c" : {
                "value" : 445,
                "type" : "LONG",
                "quality" : "GOOD"
        },
        "r" : {
                "value" : 445,
                "type" : "LONG",
                "quality" : "GOOD"
        }
}
```

Figure 32: Example of one document from the database

Retrieving the data from the monitoring system and visualizing it, is the main functionality in the application. In some cases we could have more than one sensor for the building. Thus the data is aggregated by the sensors' id that cover each building.

5.3.3 System Main Functionality implementation

The main functions which implemented can be described as follows:

1. **Visualize the current status for the campus**

The system has a routine for every 15 min, this routine is done by:

- Retrieving all sensors reading for the whole day and aggregate data for each building. Also retrieve the data for the main campus meter that has reading for the total campus.
- Update the published feature service for building with the current corresponding readings that will be reflected by the building colors.

- Compare between the latest reading and the expected reading (Compare the current value with the same last week) for this day to update and highlight the performance indicator.
- Calculate the total CO_2 emission from the latest readings of energy consumption.
- Update the charts with the current data.

2. Visualize specific building status

The system provide the user with the ability to select the building of interest and visualize its latest consumption, this done by:

- Once the user selects the building, the corresponding data will be reflected on the charts and the gauge.
- A popup window will appear with more information about the current consumption and the expected one.

Figure 33 presents a screenshot from the implemented dashboard showing the functionality of the system. The system by default displays all campus energy status, the line chart is showing the real time consumption for the main meter of campus for Gas, Water and Electricity. The heat map reflect the pattern of the consumption through the week. The gauges are indicators for the level of consumption compared with last week values and the three key performances are reflecting only the total campus status (increase or decrease) then the expected value. The gauge is changing by building indicating the building performance.

Figure 33: Dashboard interface showing the main indicators in campus

3. Compare two buildings' data

The system provide the user with the ability to compare between two buildings:

- The user can select two buildings from the dropdown list, then select the comparison scale week, month, and year.
- Then the system will aggregate the corresponding data and visualize it with bar chart.

Figure 34 shows the compare window, three sets of bar charts for each indicator.

Figure 34: Comparison between two selected buildings

4. Show The Trend in consumption for the whole campus

Figure 35: The trend of the consumption- Yearly

Chapter Six: Results and Discussion

6.1 Result and Discussion

In this thesis, several steps were performed for understanding the energy consumption pattern and assisting the university management in their reduction goal. Investigations were carried out on one building from the duration of April to October. This study could be applied on the other buildings in the university. The study performed using actual building data for the consumption with different statistical analysis techniques to understand the pattern of the consumption, find the most influencing factors on energy and implement a model for energy prediction.

Further step is to design and implement a prototype for real time visualization web application to assist the campus management.

The work started with the hypotheses that the environmental factors such as temperature and humidity have effect on the campus building's consumption level. These factors are varying during all seasons especially for types of buildings where the occupancy level is dynamically changing over time.

As a first step data were collected from different sources and stored in a Database. After that an exploratory data analysis was performed as initial step to understand the data and reveal any relation between the variables, the result shows that:

- The campus consumption is following a daily, weekly and monthly pattern, the regular daily pattern is represented by the daily operational hour whereas the peak of the consumption starts at 8:00 AM and lasts till 8:00 PM. It can also be concluded that weekends and vacations have different profiles than working days. From this, we consider the day type as one of the factors that affect the consumption.
- The analyses showed a correlation between the environmental factors such as temperature, humidity and wind-speed and the energy consumption. It was also found that the occupancy level represented by the number of classes has also an effect on the consumption. The highest correlation coefficient was for the factors Day type and Laboratory classes.

Second step was done by performing a stepwise multiple regression analysis on the whole duration from April to October using the previous correlated variables. We wanted to find exactly which one of the factors that significantly affecting the consumption and could be used as a predictor.

The results are:

- The 6 variables (humidity, aula (classrooms), temperature, day type, laboratory and wind speed) were entered in a stepwise multiple regression analysis to predict electricity consumption (kW). We found that these 6 variables were statistically significant and are affecting the phenomena under study. They were considered as predictors for the energy consumption. The model was accounted for approximately 19% of the variance of kW.

- The standardized regression coefficients of the predictors together with their correlations with **kW** shows Humidity received the strongest weight in the model followed by Aula and temperature.

- However, when we performed the analysis on each month, we noticed that the effect of these factors is varying from one month to another due to the change in seasonality. For each month we have different combinations of predictors that forms better prediction model for each month See Table 17 . For example, in September as it is the beginning of the semester and the intensity of classes is high also the temperature and humidity is relatively high, thus the model with humidity, class type (laboratory and Aula) and day type as a predictors, has the highest R Square and it is approximately 46% of the variance of kW.

- In April the most significant factors are the number of classes and the day type which represent the occupancy effect. However, the outside environmental conditions have no influence on the consumption at this month, in the opposite to June when the temperature, humidity, wind, and day type are significantly affecting the consumption.

- For using the regression model for prediction it is recommend to be updated on monthly basis a according to the seasonality change.

The overall analysis showed that the combination of factors of environmental and dynamic occupancy level are explaining more than 50% percentage of the consumption. This indicate that these factors are consider as one of the main drivers for the consumption.

We also explored the use of neural network in predicting the energy consumption. The three layered feed-forward neural networks were trained using the actual measured data collected from the building. The data duration was for 7 months. The model shows a great result in the prediction that reach R= 0.972. The model implemented with 12 variables as input and one as output. The hidden layer number was 1 with 6 neurons. The trained variables are the above mentioned 6 variables in addition to the using of the hour of the day, day of the week and months.

Although that the data is normally distributed and it fit the condition for using the regression analysis, the regression result in prediction models compared with the neural network result is low. Thus, in our case ANN model could fit better in predicting the campus consumption.

Further step was performed to enhance the energy management by implementing a WebGIS Tool (dashboard) to visualize the main indicators for the campus and makes it easier to track the campus performance in a real time. The tool provide the users with the latest and up-to-date status of energy efficiency for each building depend on real time monitoring system measurements.

6.2 RECOMMENDATIONS

During the research time, it was noticed from the analysis that we could enhance the energy efficiency and reduction by using some measurements. Recommendations for energy reduction:

(1) In April although we saw that the classes are the most influencing factor, it is recommended to increase the intensity of the classes specially the type laboratories in this month as the outside temperature and humidity is not effective factors during the whole month, with this way we are consuming less in the months that require intensity use of cooling system.

(2) One of the measurements that university already took is the strict control of the inside temperature, we can use this rule to reduce the consumption by decreasing one temperature degree or more, this could be calculated by the management according to the measured thermal comfort in the building.

(3) Some of the residual values during the night is due to the laboratories and the machines running at night. As we need to keep them operated we can consider using different source of energy suppliers during the day, such as photovoltaic energy source.

Recommendations for further analysis:

(1) For further accurate analysis and result, data about the building's envelope and specification e.g. installed capacity for space (cooling and ventilation), windows/doors/glass area, and thermal specification should be gathered.

(2) Between the moths of November till March, as the data is not available, we could not come up with any recommendation regarding the consumption reduction as this has different energy characteristics due to seasonality change. Therefore, it is recommended to study this duration in the future to have a full understanding of energy pattern.

(3) The existence of sub meters in buildings will make the analysis more 'spatial'. For example, if we have sub-meters in one building that will provide a better image for energy use, we could analyze the behavior of specific group of occupants and the efficiency of specific machines in one place, we also could define the most inefficient places in the campus it could be buildings or specific classes/Laboratories that consume more than the normal rate.

(4) For future work it is recommended to establish a database or service for all buildings specification and characteristics for better energy assessment. Also a service for universities calendar such as yearly vacations and events. Collection of all this data will be useful to create more applications for future energy monitor and management.

6.3 Future Work

The implemented dashboard could be enhanced by several ways. Adding a reporting capabilities with the dashboard will enhance the decision making process. This step could not be performed due to the time limitation. Also improving the dashboard visualization components. Another feature could be an assist is to use the Esri Geoevent processor extension to improve the application of monitoring the energy sensors in real time.

REFERENCES

Alan, S., (1992). An introduction to regression analysis, Course Lecture, The Low school, University of Chicago.

Ayres, I., S. Raseman, and A. Shih (2009). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage.

Behrens, J. (1997). Principles and procedures of exploratory data, Psychological Methods, 2, 2, 131-160., http://cll.stanford.edu/~langley/cogsys/behrens97pm.pdf .

Benedito-Bordonau, M., Gargallo, D., Avariento, J., Sanchis, A., Gould, M., Huerta, J. (2013). UJI Smart Campus: Un ejemplo de integración de recursos en la Universitat Jaume I de Castelló. IV Jornadas Ibéricas de Infraestructuras de Datos Espaciales.

Corgnati, S.P., Corrado, V., and Filippi, M. (2008). A method for heating consumption assessment in existing buildings: A field survey concerning 120 Italian schools. Energy and Buildings, 40, 5, 91-98, doi:10.1016/j.enbuild.2007.05.011.

Caldera, M., Corgnati, S.P., Filippi, M. (2008). Energy demand for space heating through a statistical approach: application to residential buildings. Energy and Buildings, 40, (10), 1972-1983.

Catalina, T., Virgone, J., Blanco, E. (2008). Development and validation of regression models to predict monthly heating demand for residential buildings.

Chen, S., Yoshino, H., Li, N. (2010). Statistical analyses on summer energy consumption characteristics of residential buildings in some cities of China. Energy and Buildings, 42, (1), 136-146.

Datta, D., Tassou, S.A., Marriott, D.(2000). Application of neural networks for the prediction of the energy consumption in a supermarket. In: Proc. CLIMA Conf., Brussels, Belgium, p. 98, 1997.

Djurica, N., Novakovicb, V.(2012) . Identifying important variables of energy use in low energy office building by using multivariate analysis. Energy and Buildings, 45, 91-98, doi:10.1016/j.enbuild.2011.10.031.

EPRI (2009). Residential Electricity Use Feedback: A Research Synthesis and Economic Framework. EPRI, Palo Alto, CA: 2009. 1016844.
http://my.epri.com/portal/server.pt?Product_id=000000000001016844.

Ekici, B.B., Aksoy, U.T.(2009). Predicting of building energy consumption by using artificial neural networks. Advances in engineering software, 40, 5, 356-362.

Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency 1 (1): 79-104.

German, S., Bienenstock, E., Doursat, R. (1992). Neural networks and the bias/variance dilemma. Neural Computation - NECO , vol. 4, no. 1, pp. 1-58, 1992, DOI: 10.1162/neco.1992.4.1.1

Haykin, S.(1999). Neural networks a Comprehensive Foundation. Upper Saddle River, N.J.: Prentice Hall.

IPCC (2013), Climate Change, The Physical Science Basis, Working Group I Contribution to the IPCC Fifth Assessment Report, ISBN 978-1-107-05799-1 hardback, ISBN 978-1-107-66182-0 paperback.

Kalogirou, S.A., Bojic,M. (2000). Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy, 25, 5, 479-491.

Lam, J.C., Hui, S.C.M., Chan, A.L.S. (1997). Regression analysis of high-rise fully air-conditioned office buildings. Energy and Buildings, 26, (2), 189-197.

Matsuura, K. (1995). Effects of Climate Change on Building Energy Consumption in Cities. Theoretical and Applied Climatology, Volume 51, Issue 1-2, pp 105-117 .

Marszal, A.J., Heiselberg, P., Bourrelle, J.S., Musall, E., Voss, K., Sartori, I., Napolitano, A. (2011). Zero energy building- A review of definitions and calculation methodologies. Energy and Buildings, 43, 4, 971-979, doi:10.1016/j.enbuild.2010.12.022 .

Nasri, N. (2010). Application of Artificial Neural Networks (ANNs) in Prediction Models in Risk Management. World Applied Sciences Journal, 10, 12, 1493-1500, ISSN 1818-4952.

Olofsson, T., Andersson, S. (2001). Long-term energy demand predictions based on short-term measured data. Energy and Buildings, 33, (2), 85-91, doi:10.1016/S0378-7788(00)00068-2.

Olofsson, T., Andersson, S., Sjogren, J. (2009). Building energy parameter investigations based on multivariate analysis. Energy and Buildings, 41,71-80, doi:10.1016/j.enbuild.2008.07.012 .

Pettersen, T.D.(1994). Variation of energy consumption in dwellings due to climate, building and inhabitants. Energy and Buildings. 21, 209-2018.

Sailor, D.J., Munoz, J.R.(1997). Sensitivity of electricity and natural gas consumption to climate in the USA-methodology and results for eight states, Energy, Vol. 22, p. 987-998.

Tukey, J.W.(1969). Analyzing data: sanctification or detective work?, American Psychologist, 24, 2, 83-91, DOI: 10.1037/h0027108.

Wikimedia, "Neural network example," http://commons.wikimedia.org/wiki/File:Neural_network_example.svg

Yalcintas,M., Akkurt, S.(2005). Artificial neural networks applications in building energy predictions and a case study for tropical climates. international journal of energy research, Int. J. Energy Res, 29, 10,891–901.

Zmeureanu, R., Fazio,P.(1991) . Analysis of the energy performance of office buildings in Montreal in 1988. Energy and Buildings, 17, 1, 63-74.

APPENDIX A: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'APRIL'

A1: ANOVA Table for month of April

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	8158135.840	1	8158135.840	548.635	.000[c]
	Residual	10676568.910	718	14869.873		
	Total	18834704.750	719			
2	Regression	9931719.503	2	4965859.751	399.924	.000[d]
	Residual	8902985.247	717	12416.995		
	Total	18834704.750	719			

a. month = 4

b. Dependent Variable: kw

c. Predictors: (Constant), Aula

d. Predictors: (Constant), Aula, Day_Type

A2: Coefficients Table for month of April

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	171.476	4.962		34.560	.000		
	Aula	25.387	1.084	.658	23.423	.000	1.000	1.000
2	(Constant)	74.625	9.286		8.036	.000		
	Aula	22.729	1.015	.589	22.391	.000	.952	1.050
	Day_Type	127.170	10.641	.315	11.951	.000	.952	1.050

a. month = 4

b. Dependent Variable: kw

A3: Excluded Variables Table for month of April

Excluded Variables[a,b]

Model		Beta In	T	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
1	temp	-.001-[c]	-.021-	.984	-.001-	.996	1.004	.996
	humidity	.016[c]	.557	.578	.021	.998	1.002	.998
	wind	.029[c]	1.040	.298	.039	.999	1.001	.999
	Laboratory	.074[c]	1.452	.147	.054	.301	3.320	.301
	Seminar	-.021-[c]	-.573-	.567	-.021-	.578	1.730	.578
	Day_Type	.315[c]	11.951	.000	.408	.952	1.050	.952
2	temp	.032[d]	1.232	.219	.046	.985	1.015	.942
	humidity	-.013-[d]	-.494-	.621	-.018-	.989	1.011	.944
	wind	.030[d]	1.165	.245	.043	.999	1.001	.951
	Laboratory	.071[d]	1.517	.130	.057	.301	3.320	.297
	Seminar	-.014-[d]	-.422-	.673	-.016-	.578	1.731	.559

a. month = 4

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), Aula

d. Predictors in the Model: (Constant), Aula, Day_Type

APPENDIX B: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'MAY'

B1: ANOVA Table for month of May

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	11568753.465	1	11568753.465	168.411	.000[c]
	Residual	50970631.997	742	68693.574		
	Total	62539385.462	743			
2	Regression	19508223.731	2	9754111.866	167.967	.000[d]
	Residual	43031161.731	741	58071.743		
	Total	62539385.462	743			
3	Regression	25565226.188	3	8521742.063	170.554	.000[e]
	Residual	36974159.274	740	49965.080		
	Total	62539385.462	743			
4	Regression	26089254.261	4	6522313.565	132.235	.000[f]
	Residual	36450131.202	739	49323.588		
	Total	62539385.462	743			
5	Regression	26280125.842	5	5256025.168	106.978	.000[g]
	Residual	36259259.620	738	49131.788		
	Total	62539385.462	743			

a. month = 5

b. Dependent Variable: kw

c. Predictors: (Constant), Aula

d. Predictors: (Constant), Aula, temp

e. Predictors: (Constant), Aula, temp, Day_Type

f. Predictors: (Constant), Aula, temp, Day_Type, wind

g. Predictors: (Constant), Aula, temp, Day_Type, wind, Seminar

B2: Coefficients Table for month of May

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	265.890	10.369		25.642	.000		
	Aula	33.334	2.569	.430	12.977	.000	1.000	1.000
2	(Constant)	-361.835-	54.525		-6.636-	.000		
	Aula	29.188	2.388	.377	12.222	.000	.978	1.023
	Temp	32.397	2.771	.360	11.693	.000	.978	1.023
3	(Constant)	-588.284-	54.599		-10.775-	.000		
	Aula	24.667	2.253	.318	10.949	.000	.945	1.058
	Temp	33.625	2.572	.374	13.071	.000	.976	1.024
	Day_Type	249.530	22.664	.317	11.010	.000	.967	1.035
4	(Constant)	-509.340-	59.408		-8.574-	.000		
	Aula	25.193	2.244	.325	11.226	.000	.941	1.063
	Temp	27.575	3.159	.307	8.729	.000	.639	1.565
	Day_Type	240.085	22.703	.305	10.575	.000	.951	1.052
	Wind	6.356	1.950	.114	3.259	.001	.649	1.540
5	(Constant)	-505.579-	59.323		-8.522-	.000		
	Aula	21.571	2.897	.278	7.446	.000	.562	1.779
	Temp	27.364	3.154	.304	8.675	.000	.638	1.567
	Day_Type	240.478	22.660	.305	10.612	.000	.951	1.052
	Wind	6.417	1.946	.115	3.297	.001	.649	1.541
	Seminar	21.718	11.019	.073	1.971	.049	.580	1.723

a. month = 5

b. Dependent Variable: kw

B3: Excluded Variables Table for month of May

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
1	temp	.360[c]	11.693	.000	.395	.978	1.023	.978
	humidity	-.258-[c]	-8.080-	.000	-.285-	.993	1.007	.993
	wind	.314[c]	10.079	.000	.347	.998	1.002	.998
	Laboratory	.056[c]	.923	.356	.034	.296	3.374	.296
	Seminar	.080[c]	1.852	.064	.068	.581	1.721	.581
	Day_Type	.300[c]	9.429	.000	.327	.968	1.033	.968
2	humidity	-.046-[d]	-1.153-	.249	-.042-	.579	1.728	.570
	wind	.161[d]	4.334	.000	.157	.660	1.515	.647
	Laboratory	.009[d]	.168	.867	.006	.295	3.392	.295
	Seminar	.066[d]	1.654	.098	.061	.581	1.723	.576
	Day_Type	.317[d]	11.010	.000	.375	.967	1.035	.945
3	humidity	-.088-[e]	-2.355-	.019	-.086-	.573	1.745	.570
	wind	.114[e]	3.259	.001	.119	.649	1.540	.639
	Laboratory	.012[e]	.229	.819	.008	.295	3.392	.293
	Seminar	.071[e]	1.906	.057	.070	.580	1.723	.563
4	humidity	-.068-[f]	-1.807-	.071	-.066-	.554	1.804	.475
	Laboratory	.002[f]	.037	.970	.001	.294	3.404	.291
	Seminar	.073[f]	1.971	.049	.072	.580	1.723	.562
5	humidity	-.066-[g]	-1.759-	.079	-.065-	.554	1.805	.475
	Laboratory	-.056-[g]	-.958-	.339	-.035-	.234	4.279	.234

a. month = 5

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), Aula

d. Predictors in the Model: (Constant), Aula, temp

e. Predictors in the Model: (Constant), Aula, temp, Day_Type

f. Predictors in the Model: (Constant), Aula, temp, Day_Type, wind

g. Predictors in the Model: (Constant), Aula, temp, Day_Type, wind, Seminar

APPENDIX C: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'JUNE'

C1: ANOVA Table for month of June

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	20919233.859	1	20919233.859	175.371	.000[c]
	Residual	85647229.128	718	119285.834		
	Total	106566462.987	719			
2	Regression	31855383.597	2	15927691.799	152.858	.000[d]
	Residual	74711079.390	717	104199.553		
	Total	106566462.987	719			
3	Regression	34788880.544	3	11596293.515	115.676	.000[e]
	Residual	71777582.444	716	100248.020		
	Total	106566462.987	719			
4	Regression	37014123.526	4	9253530.881	95.127	.000[f]
	Residual	69552339.462	715	97275.999		
	Total	106566462.987	719			

a. month = 6

b. Dependent Variable: kw

c. Predictors: (Constant), temp

d. Predictors: (Constant), temp, Day_Type

e. Predictors: (Constant), temp, Day_Type, wind

f. Predictors: (Constant), temp, Day_Type, wind, humidity

C2: Coefficients Table for month of June

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-660.825-	81.019		-8.156-	.000		
	Temp	44.028	3.325	.443	13.243	.000	1.000	1.000
2	(Constant)	-842.184-	77.764		-10.830-	.000		
	Temp	40.021	3.132	.403	12.779	.000	.984	1.016
	Day_Type1	333.307	32.535	.323	10.245	.000	.984	1.016
3	(Constant)	-743.668-	78.420		-9.483-	.000		
	Temp	31.449	3.457	.316	9.098	.000	.777	1.286
	Day_Type1	344.056	31.974	.333	10.761	.000	.981	1.020
	Wind	15.019	2.776	.187	5.409	.000	.790	1.266
4	(Constant)	-138.003-	148.335		-.930-	.353		
	Temp	15.380	4.783	.155	3.215	.001	.394	2.538
	Day_Type1	373.781	32.103	.362	11.643	.000	.944	1.059
	Wind	15.114	2.735	.188	5.526	.000	.790	1.266
	humidety	-4.190-	.876	-.216-	-4.783-	.000	.449	2.228

a. month = 6

b. Dependent Variable: kw

C3: Excluded Variables Table for month of June

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	humidity	-.115-[c]	-2.356-	.019	-.088-	.466	2.144	.466
	wind	.164[c]	4.410	.000	.163	.793	1.261	.793
	Day_Type	.323[c]	10.245	.000	.357	.984	1.016	.984
2	humidity	-.214-[d]	-4.648-	.000	-.171-	.449	2.228	.443
	wind	.187[d]	5.409	.000	.198	.790	1.266	.777
3	humidity	-.216-[e]	-4.783-	.000	-.176-	.449	2.228	.394

a. month = 6

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), temp

d. Predictors in the Model: (Constant), temp, Day_Type1

e. Predictors in the Model: (Constant), temp, Day_Type1, wind

APPENDIX D: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'JULY'

D1: ANOVA Table for month of July

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24885327.515	1	24885327.515	236.911	.000[c]
	Residual	77940394.376	742	105040.963		
	Total	102825721.891	743			
2	Regression	26362290.361	2	13181145.181	127.737	.000[d]
	Residual	76463431.530	741	103189.516		
	Total	102825721.891	743			

a. month = 7

b. Dependent Variable: kw

c. Predictors: (Constant), humidity

d. Predictors: (Constant), humidity, Day_Type

D2: Coefficients Table for month of July

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1041.741	44.073		23.637	.000		
	humidity	-10.737-	.698	-.492-	-15.392-	.000	1.000	1.000
2	(Constant)	1117.461	48.050		23.256	.000		
	humidity	-11.343-	.710	-.520-	-15.983-	.000	.949	1.054
	Day_Type	-92.682-	24.498	-.123-	-3.783-	.000	.949	1.054

a. month = 7

b. Dependent Variable: kw

D3: Excluded Variables Table for month of July

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
1	temp	.070[c]	1.507	.132	.055	.475	2.104	.475
	wind	-.062-[c]	-1.754-	.080	-.064-	.807	1.238	.807
	Day_Type	-.123-[c]	-3.783-	.000	-.138-	.949	1.054	.949
2	temp	.028[d]	.582	.561	.021	.445	2.247	.422
	wind	-.061-[d]	-1.744-	.082	-.064-	.807	1.238	.775

a. month = 7

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), humidity

d. Predictors in the Model: (Constant), humidity, Day_Type

APPENDIX E: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'AUGUST'

E1: ANOVA Table for month of August

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	486644.451	1	486644.451	20.981	.000[c]
	Residual	17210628.160	742	23194.917		
	Total	17697272.612	743			
2	Regression	1292079.082	2	646039.541	29.181	.000[d]
	Residual	16405193.530	741	22139.263		
	Total	17697272.612	743			

a. month = 8

b. Dependent Variable: kw

c. Predictors: (Constant), wind

d. Predictors: (Constant), wind, temp

E2: Coefficients Table for month of August

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	129.398	10.213		12.670	.000		
	Wind	-6.050-	1.321	-.166-	-4.580-	.000	1.000	1.000
2	(Constant)	-202.258-	55.884		-3.619-	.000		
	Wind	-11.606-	1.585	-.318-	-7.320-	.000	.662	1.510
	Temp	13.802	2.288	.262	6.032	.000	.662	1.510

a. month = 8

b. Dependent Variable: kw

E3: Excluded Variables Table for month of August

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	temp	.262[c]	6.032	.000	.216	.662	1.510	.662
	humidity	-.136-[c]	-3.154-	.002	-.115-	.695	1.439	.695
2	humidity	.037[d]	.679	.497	.025	.428	2.339	.408

a. month = 8

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), wind

d. Predictors in the Model: (Constant), wind, temp

APPENDIX F: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'SEPTEMBER'

F1: ANOVA Table for month of September

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	36013632.680	1	36013632.680	208.422	.000[c]
	Residual	124064470.120	718	172791.741		
	Total	160078102.800	719			
2	Regression	58419818.281	2	29209909.141	206.019	.000[d]
	Residual	101658284.519	717	141782.824		
	Total	160078102.800	719			
3	Regression	73765143.408	3	24588381.136	203.970	.000[e]
	Residual	86312959.392	716	120548.826		
	Total	160078102.800	719			
4	Regression	74288616.508	4	18572154.127	154.787	.000[f]
	Residual	85789486.292	715	119985.296		
	Total	160078102.800	719			

a. month = 9

b. Dependent Variable: kw

c. Predictors: (Constant), humidity

d. Predictors: (Constant), humidity, Aula

e. Predictors: (Constant), humidity, Aula, Day_Type

f. Predictors: (Constant), humidity, Aula, Day_Type, Laboratory

F2: Coefficients Table for month of September

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1324.275	59.521		22.249	.000		
	humidity	-12.272-	.850	-.474-	-14.437-	.000	1.000	1.000
2	(Constant)	1253.792	54.207		23.130	.000		
	humidity	-11.908-	.771	-.460-	-15.454-	.000	.999	1.001
	Aula	112.339	8.936	.374	12.571	.000	.999	1.001
3	(Constant)	877.599	60.084		14.606	.000		
	humidity	-11.162-	.714	-.431-	-15.642-	.000	.990	1.010
	Aula	101.748	8.293	.339	12.269	.000	.986	1.014
	Day_Type	396.107	35.108	.313	11.283	.000	.978	1.022
4	(Constant)	868.707	60.094		14.456	.000		
	humidity	-11.038-	.714	-.427-	-15.451-	.000	.983	1.017
	Aula	91.257	9.679	.304	9.428	.000	.720	1.388
	Day_Type	393.009	35.057	.311	11.211	.000	.976	1.024
	Laboratory	115.094	55.102	.068	2.089	.037	.718	1.393

a. month = 9

b. Dependent Variable: kw

F3: Excluded Variables Table for month of September

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	temp	-.047-[c]	-.881-	.379	-.033-	.372	2.685	.372
	wind	.016[c]	.431	.666	.016	.739	1.353	.739
	Aula	.374[c]	12.571	.000	.425	.999	1.001	.999
	Laboratory	.257[c]	8.133	.000	.291	.991	1.009	.991
	Seminar	.217[c]	6.798	.000	.246	1.000	1.000	1.000
	Day_Type	.352[c]	11.601	.000	.398	.991	1.009	.991
2	temp	.061[d]	1.234	.217	.046	.361	2.769	.361
	wind	.063[d]	1.817	.070	.068	.731	1.368	.731
	Laboratory	.083[d]	2.367	.018	.088	.719	1.391	.719
	Seminar	.027[d]	.767	.443	.029	.722	1.385	.721
	Day_Type	.313[d]	11.283	.000	.389	.978	1.022	.978
3	temp	.028[e]	.617	.537	.023	.360	2.780	.360
	wind	.054[e]	1.683	.093	.063	.730	1.369	.730
	Laboratory	.068[e]	2.089	.037	.078	.718	1.393	.718
	Seminar	.020[e]	.626	.532	.023	.722	1.385	.716
4	temp	.029[f]	.624	.533	.023	.360	2.780	.360
	wind	.047[f]	1.465	.143	.055	.722	1.386	.704
	Seminar	-.014-[f]	-.385-	.701	-.014-	.567	1.765	.563

a. month = 9

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), humidity

d. Predictors in the Model: (Constant), humidity, Aula

e. Predictors in the Model: (Constant), humidity, Aula, Day_Type

f. Predictors in the Model: (Constant), humidity, Aula, Day_Type, Laboratory

APPENDIX G: ANOVA, COEFFICIENTS AND EXCLUDED TABLES FOR MONTH OF 'OCTOBER'

G1: ANOVA Table for month of October

ANOVA[a,b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	23716991.082	1	23716991.082	214.553	.000[c]
	Residual	82021888.755	742	110541.629		
	Total	105738879.837	743			
2	Regression	41179053.822	2	20589526.911	236.321	.000[d]
	Residual	64559826.016	741	87125.271		
	Total	105738879.837	743			
3	Regression	52497163.274	3	17499054.425	243.217	.000[e]
	Residual	53241716.563	740	71948.266		
	Total	105738879.837	743			
4	Regression	52961505.363	4	13240376.341	185.395	.000[f]
	Residual	52777374.475	739	71417.286		
	Total	105738879.837	743			
5	Regression	53292899.436	5	10658579.887	149.984	.000[g]
	Residual	52445980.401	738	71065.014		
	Total	105738879.837	743			

a. month = 10

b. Dependent Variable: kw

c. Predictors: (Constant), temp

d. Predictors: (Constant), temp, Aula

e. Predictors: (Constant), temp, Aula, Day_Type

f. Predictors: (Constant), temp, Aula, Day_Type, Laboratory

g. Predictors: (Constant), temp, Aula, Day_Type, Laboratory, wind

G2: Coefficients Table for month of October

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-368.970-	53.768		-6.862-	.000		
	Temp	38.271	2.613	.474	14.648	.000	1.000	1.000
2	(Constant)	-330.870-	47.810		-6.920-	.000		
	Temp	34.028	2.339	.421	14.549	.000	.984	1.017
	Aula	87.989	6.215	.410	14.157	.000	.984	1.017
3	(Constant)	-616.240-	49.044		-12.565-	.000		
	Temp	34.360	2.126	.425	16.165	.000	.983	1.017
	Aula	78.427	5.699	.365	13.761	.000	.966	1.035
	Day_Type	338.389	26.980	.330	12.542	.000	.982	1.018
4	(Constant)	-606.239-	49.020		-12.367-	.000		
	Temp	33.860	2.127	.419	15.920	.000	.975	1.026
	Aula	69.820	6.606	.325	10.570	.000	.714	1.401
	Day_Type	333.824	26.940	.326	12.392	.000	.978	1.023
	Laboratory	91.088	35.723	.079	2.550	.011	.712	1.404
5	(Constant)	-577.047-	50.733		-11.374-	.000		
	Temp	30.861	2.536	.382	12.171	.000	.683	1.465
	Aula	69.973	6.590	.326	10.619	.000	.714	1.401
	Day_Type	336.147	26.895	.328	12.499	.000	.976	1.024
	Laboratory	92.620	35.642	.080	2.599	.010	.712	1.405
	Wind	6.979	3.232	.067	2.159	.031	.697	1.435

a. month = 10

b. Dependent Variable: kw

G3: Excluded Variables Table for month of October

Excluded Variables[a,b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
1	humidity	-.108-[c]	-2.186-	.029	-.080-	.423	2.362	.423
	wind	.034[c]	.890	.374	.033	.699	1.431	.699
	Aula	.410[c]	14.157	.000	.461	.984	1.017	.984
	Laboratory	.289[c]	9.349	.000	.325	.979	1.021	.979
	Seminar	.280[c]	9.113	.000	.317	.997	1.003	.997
	Day_Type	.379[c]	12.960	.000	.430	1.000	1.000	1.000
2	humidity	-.068-[d]	-1.546-	.122	-.057-	.422	2.372	.422
	wind	.049[d]	1.431	.153	.053	.698	1.432	.688
	Laboratory	.104[d]	3.078	.002	.112	.715	1.398	.715
	Seminar	.051[d]	1.394	.164	.051	.626	1.598	.618
	Day_Type	.330[d]	12.542	.000	.419	.982	1.018	.966
3	humidity	-.021-[e]	-.517-	.605	-.019-	.418	2.393	.418
	wind	.065[e]	2.100	.036	.077	.697	1.435	.688
	Laboratory	.079[e]	2.550	.011	.093	.712	1.404	.712
	Seminar	.046[e]	1.401	.162	.051	.626	1.598	.611
4	humidity	-.018-[f]	-.438-	.661	-.016-	.417	2.396	.417
	wind	.067[f]	2.159	.031	.079	.697	1.435	.683
	Seminar	.015[f]	.417	.677	.015	.525	1.905	.525
5	humidity	.001[g]	.024	.981	.001	.398	2.511	.390
	Seminar	.016[g]	.447	.655	.016	.525	1.906	.525

a. month = 10

b. Dependent Variable: kw

c. Predictors in the Model: (Constant), temp

d. Predictors in the Model: (Constant), temp, Aula

e. Predictors in the Model: (Constant), temp, Aula, Day_Type

f. Predictors in the Model: (Constant), temp, Aula, Day_Type, Laboratory

g. Predictors in the Model: (Constant), temp, Aula, Day_Type, Laboratory, wind

Masters
Program
in **Geospatial**
Technologies

Nourhan Khalifa Abdelkareem

Supported by:
Education and Culture
ERASMUS MUNDUS