A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

DATA-DRIVEN MODELING OF SMART BUILDING ENERGY MANAGEMENT

RAGHDA AHMED ABDELKERIM SALAMA

Work project carried out under the supervision of:
Prof.Qiwei Han

20-05-2021

Abstract

Buildings consume approximately 40% of energy in total, which contributes negatively to the environment. Building Energy Management Systems (BEMS) have been used to monitor energy consumption and increase usage efficiency. In this study, the components and importance of BEMS are emphasized. The data from the management system of the Chamchuri 5 building in Chulalongkorn University, Thailand, were used as a template for data-driven modeling for energy usage in smart buildings to analyze the patterns of energy consumption. Using multilevel modeling on the Chamchuri 5 building, the main factors that consume energy on a macro and micro level are analyzed. Energy variation between zones and floors was spotted.

Keywords (Business and Data Analytics, Machine Learning, Energy Efficiency, Energy Management, Smart Buildings)

Acknowledgment

I would like to acknowledge and show my appreciation to everyone who supported me. I would like to thank my supervisor Qiwei Han for guiding me and providing me with educational support. My dear family and friends who have been there for me through hard times. A special thanks to my beloved husband and son. Finally, I would like to dedicate this dissertation to my father who was always pushing me to push through challenges and seek opportunities may his soul rest in peace.

Funding and conflict of interest

The author received no financial support for the research, authorship, and/or publication of this work. The author declares no conflict of interests.

1. INTRODUCTION

Energy consumption in buildings is one of the major sources of CO2 emissions (IEA 2021), (Cao et al., 2016). It is important to determine the energy patterns in buildings to be able to provide suggestions in order to lower their consumption and increase energy efficiencies. Nowadays, with technological advancements, analysts have been able to collect and study data about energy usage through building energy systems and offer energy solutions and recommendations. In this thesis, energy consumption problems in buildings will be discussed followed by the importance of building energy management systems (BEMS) and their components. Furthermore, a case study of the Chamchuri campus building in Thailand will be analyzed to reveal the energy drivers in the building using multilevel modeling.

Energy management and efficiency plans are triggered by many reasons. To mention a few, cost reduction, the change towards sustainable cultures and values, and reducing carbon footprint. In Thailand, Chulalongkorn University is highly committed to the community and sustainability. As reported in their sustainability report 2018-2020, 72 million dollars are spent on sustainable projects and research. The university offers more than 1300 courses to promote sustainability. Its mission includes many environmental and CSR-oriented plans one of these initiatives is to ensure a sustainable building design by providing a BEMS in all the buildings on campus and installing solar roofs as a source of renewable energy (Chulalongkorn University Sustainability Report, 2018-2020). Additionally, the efforts of the university in sustainability have been recognized, the university has been awarded many energy awards (CUBEMS, 2021).

2. LITERATURE REVIEW

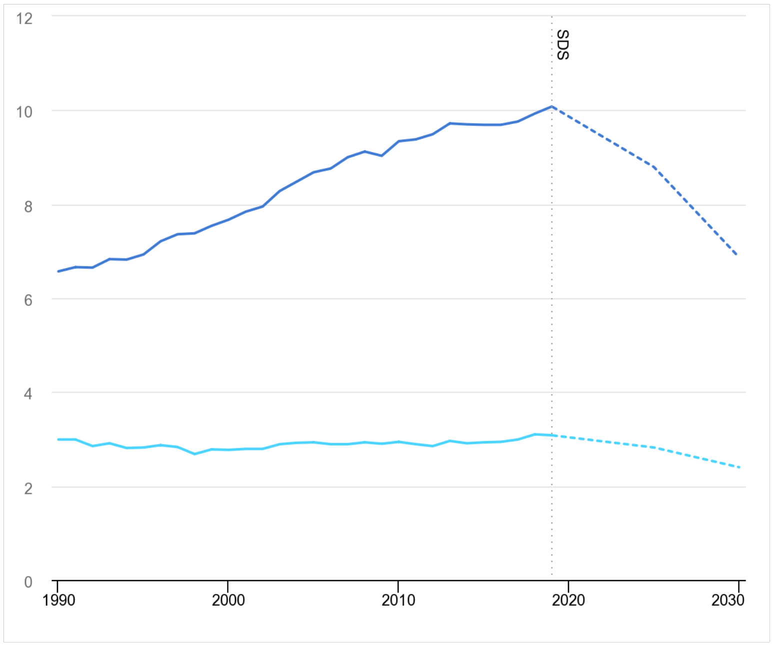
Energy consumption in buildings is one of the major sources of CO2 emissions. It has been reported that energy consumed by these buildings makes up for 28% of the global carbon emissions

(IEA 2021). In addition, they use up 40% of the total energy consumed in the US and EU (Cao et al., 2016). Moreover, energy consumption is heightened by various factors. Population growth constitutes a major energy consumer. This growth has been caused by an annual increase of floor area which amounts to 2.5%, which is outpacing the observed 0.5% to 1% declining rate in energy used per square meter as reported in 2010. Therefore, a minimum of 2.5% reduction in energy is required to balance the constant outgrowth in population per year (IEA 2021).

Other factors include the surge in demand for rapid changes in the lifestyle, which comes with a rise in technological advancements needed to enhance the living standards resulting in more energy being consumed. Additionally, the continuous climate change is causing an over-utilization of HVAC systems to maintain a comfortable ambiance. Furthermore, the use of fossil fuel energy sources has seen an annual growth rate of 0.7% since 2010 (Cao et al., 2016).

In 2019, the International Energy Agency (IEA) recorded a staggering 10 gigatons of CO2, which is the highest level of carbon emissions from buildings reported to date (Figure 1). IEA analysts have interpreted this surge to be caused by the extreme weather events that year consequent to climate change. As an example, mid-2019 witnessed the second record for the hottest year. This record was a result of the pre-warned El-Nino phenomenon which is known for causing extreme shifts in the location of warm ocean waters, characterized by much warmer waters in central and eastern parts of the basin and cooler than normal in the western tropical Pacific. As the waters became extremely hot, consequently, the usage of air conditioning systems increased to reach tolerable indoor temperatures and unfortunately, this increase surpassed the efforts made to reduce energy usage (Cao et al., 2016).

Figure 1: Carbon emissions level in gigatons from year 1990 to 2030 (IEA 2021)



Carbon emissions are in constant growth which affects the environment severely. In efforts to mitigate these impacts, many countries started to impose laws and sanctions to restrict the level of CO2 emissions. Moreover, there is an observed increase in awareness campaigns raised to inform the population of the huge impact of pollution caused by energy consumption in hopes of raising more environment-conscious behaviors. As an example, corporate social responsibility concepts are now widely used in many firms; clearly, there is a trend moving towards greener sources of energy, and energy optimization.

Undoubtedly, these overall increases in environmental consciousness and the availability of data enhanced the understanding of the main causes for energy consumption. Powerful entities with the ability to put changes in effect have started to act accordingly, Europe is currently implementing a project to gradually renovate their existing buildings which constitute the largest percentage of buildings in Europe by focusing on fronts as well as their heating systems. Following in their steps, China is investing in energy efficiency solutions in new buildings as part of their plans for expanding in construction and urbanization (Cao et al., 2016).

Potential conscious building solutions could focus on the aforementioned HVAC systems which consume a large amount of energy whether the purpose is for cooling or heating. That is where architectural design could come into play. Several passive techniques could be implemented depending on the climate conditions for the studied areas, their implementation can be used to considerably decrease the need for AC. Some of these passive solutions are building orientations, altering wall thicknesses, and thermal insulation systems which impact the thermal transmittances expressed by U-values of walls and roofs and reduce the heat exchange leading to better energy efficiencies and savings. Additionally, one could put into account the heat gain and loss of the materials used, the presence of traditional construction methods like earth construction, proper design for shading devices, windows' size and orientation, natural ventilation systems & skylights, etc. Moreover, the integration of building energy management systems, which will be explained later, is used in sustainable and smart buildings.

Further, the idea of zero energy building (ZEB) is becoming more attractive as an innovative approach in building design. These buildings use renewable energy sources to supply the building with the required energy (Torcellini et al., 2006). More countries are encouraging the implementation of this concept in buildings' construction as it is a combination of the traditionally used passive techniques in construction and renewable energy technologies. For example, the Chamchuri building has a solar energy roof that contributes to the energy supply for the building (CUBEMS, 2021).

2.1. Building Energy Management Systems (BEMS)

BEMS have been evolving over many decades, driven by the concept of energy efficiency. They started with very limited capacities in the 1970s, then transcended to computer-based systems, with the rise in technological advancements and the spread of computers. Some of the major manufacturers of BEMS are Siemens, Toshiba, Hitachi, and GE (Asare-Bediako et al., 2012). In the 2000s, BEMS evolved to a combined hardware and software system, while nowadays, compact chips (Asare-Bediako et al., 2012) and cloud networks are being used (Memoori, 2017).

2.1.1. Importance Of BEMS

BEMS are used in the equipment in residential and commercial buildings equipment to manage and control the energy usage of mechanical and electrical systems while ensuring the required standards of indoor air quality and performance are maintained. Building managers use BEMS to collect data about heating, cooling, light, vertical transportation, security systems, IT networks, etc. (Yang et al., 2017). This pivotal data paved the road for many applications such as "Supervisory Control and Data Acquisition (SCADA) with Energy Management System (EMS) functionalities, dispatcher training simulator (DTS) and optimal power flow (OPF)" (Asare-Bediako et al., 2012).

One form of monitoring is the supervision of the energy level used in each building area. These systems can be utilized for a localized adjustment of temperature in certain areas of the building through HVAC control instead of a centralized temperature system. Remote monitoring is also possible with the ability to control the BEMS to mitigate the risks that arise from emergency events by sending alarms in case of anomalies or a security failure. Such features give smart buildings an edge not only for being energy and cost effective but also for the provision of security and comfort as well that is lacked in other buildings.

These systems also allow access to stored data, which enables the analysis of the data history of the building to use it as a benchmark to compare with other buildings or even to forecast future energy patterns. Nevertheless, the implementation of BEMS requires continuous maintenance to ensure adequate performance and quality of the retrieved data. More importantly, a thorough understanding of the system and the combination of the available information for decision making without compromising the performance is crucial. Calibration and maintenance are important to determine the percentage of error in the readings, to be able to evaluate the quality of the data, and decide whether these data could be used for analysis. For example, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) used some guidelines for the calibration of devices and models. There is a level of accepted error, above this threshold, the error is considered to be too high. For instance, a mean of bias error (MBE)can be calculated. A positive value of MBE indicates an overestimation of the readings while a negative value indicates an underestimation of the readings (Ruiz & Bandera, 2017).

The type of BEMS to use depends on the budget of the project, its size, and purpose. The system is usually more effective and easier to implement when it is embedded during the early design stage of a building rather than applying it to an existing building. Although the system has a high initial cost, it is balanced out energy savings that reduce the operating costs of a building. Many studies worked on estimating the value of cost and energy saving, however, due to varied results an exact reduction rate remains undetermined (Climate Technology Centre & Network,2016).

2.1.2. BEMS Components

BEMS are a combination of computers, networks, processors, and sensors. The internet of things maintains communications between all the components in the building as well as decision-making through data mining(Gaber et al., 2019). Taking sensors as an example, they are distributed throughout the building to perform many essential functions. From light and voltage sensors to fire and smoke detectors installed to actuate alarms in case of any abnormal readings to ensure the security of the building and its inhabitants. This collection of data are then grouped and sent to another part of the BEMS (Asare-Bediako et al., 2012). Measuring devices then provide detailed information and time series data for different sources of energy consumption and transmit these data.

Figure 2: BEMS components (Asare-Bediako et al., 2012)

The data retrieval from the BEMS is combined with data mining. Data mining allows a better understanding of the data collected, we can deduct a model that best fits those datasets and even predict future patterns (Chen et al., 2015). It enables the visualization of patterns, peaks, extract hidden patterns, make classification and clustering using statistical methods and machine learning which then help us initiate the actuators in the system. Each dataset has its level of complexity (Chen et al., 2015). Furthermore, smart appliances can be integrated into the system to enhance energy efficiency and allow smoother control. Enabled information and communication technology will then pave the way for the collection of the previously mentioned data from the sensors, meters, and appliances to start monitoring and controlling various processes. Finally, the choice of the manufacturer and the management system is dependent on the needed specifications, some systems only present the information, while others can perform automated control actions, and energy forecasting (Asare-Bediako et al., 2012).

3. METHODOLOGY

This section will cover the methodology used in the analysis of the research question "Data-driven Modeling of Smart Building Energy Management." Using the Chamchuri 5 building in Chulalongkorn University, Bangkok, Thailand as a case study. The nature of the longitudinal clustered nature of the available dataset will be explained as well as a description of the analysis methods, and the multilevel model used.

3.1. Longitudinal Data

The available data for the building are granular, longitudinal data. Longitudinal data are extensively used in countless fields; namely, social, medical, financial studies, and many more domains. Unlike cross-sectional data that provide data for a single point in time, longitudinal data represent the different readings over time for a certain variable by recurrently assessing it. Such as these time intervals can be minutes, days, months; additionally, they can be the same between different people or even variable intervals.

This type of data are also useful in tracking alterations within a specified time period to further analyze the data. These alterations are known as "growth models" or "growth curve analysis". The reason for the name "growth" is that the change was assumed to be increasing over time while in fact change can decrease or have other forms (Singer & Willett, 2009).

However, analyzing these types of data can be challenging. Over the years, statistical concepts were being used in data analysis; where linear models are used to examine and interpret datasets, nevertheless sometimes the available data are too complex and are better represented by other non-linear functions to comprehend numerous underlying patterns. In addition, missing data can be problematic(Fitzmaurice et al., 2008).

Moreover, exploiting the various statistical models to analyze these datasets enable solving several research questions as they allow the examination of changes within a person known as level-1 analysis and similarly track the changes among individuals over time known as level-2 (Singer & Willett, 2009). For instance, level-1 can cover the variation in energy consumption in a certain zone over time. In this case, it is more descriptive for each zone; on the contrary, level-2 looks at the difference between the zones over time, it shows the connection between the predictors and whether the zones have different patterns among them. Each level studies a specific outcome with a set of predictors. Leading to considering (Singer & Willett, 2009). These two levels together can be combined into a multilevel model.

3.2. Clustering

Data mining consists of several techniques such as classification, clustering, decision tree, and neural networking to name a few (Finch et al., 2016). One of the techniques that will be used in this case study analysis is clustering. Whereas, clustering is quite different from classification, as clustering consists of forming groups for the data while classification is predetermined (Finch et al., 2016). For instance, clustering consists of assembling data into distinct groups or "clusters," each cluster is different from the other while data within one cluster are similar (Finch et al., 2016). In the case study in question, the data will be clustered into subsets such that different zones are nested within floors nested in a building. Consequently, in this case, the used dataset is considered to be multilevel in structure (Finch et al., 2016). Moreover, clustering makes data search and retrieval easier and more efficient especially when dealing with large datasets (Finch et al., 2016)(Rao, 2014).

The interclass correlation (ICC) formula demonstrated below which varies from 0 to 1 is can be used to calculate the correlation strength between individual data within the same cluster. Where the value of ICC is directly proportional to the strength of correlation between data.

Equation 1: Interclass correlation formula, (Finch et al., 2016)

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2}$$

Furthermore, the ICC shows to which extent the nested data can influence the variable being studied. Thus, a higher ICC value draws attention to the importance of multilevel modeling during the study. Therefore, simple linear regression models are not convenient for the analysis of this data structure instead, multilayered modeling will then be more efficient (Finch et al., 2016).

3.3. Multilevel Modeling

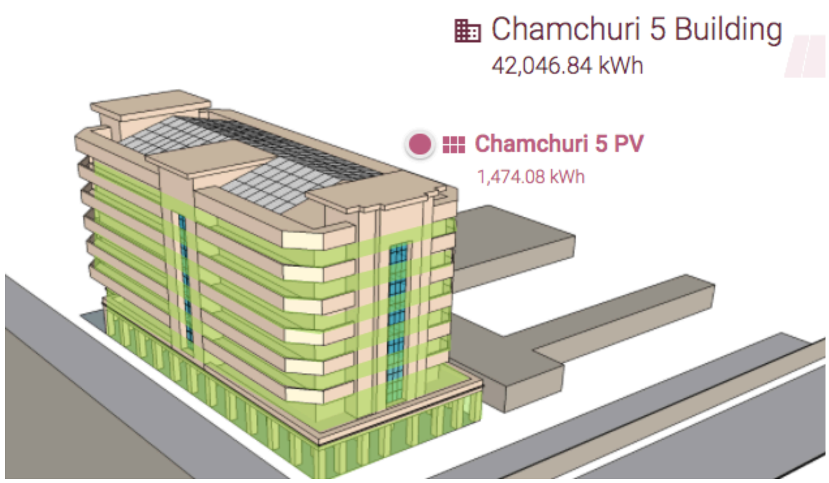
After understanding the type of data, we will be dealing with, now we will move to the type of modeling that is used in the analysis. Multilevel analysis is one of the methods used to analyze multiple levels of nested data. Careful analysis is important in this case study to avoid errors in the analysis as the variance between the zones is different from the variance between the floors (Snijders & Bosker, 1999). Multilevel modeling is often referred to as "hierarchical linear models, mixed models, or random coefficient models" (Snijders & Bosker, 1999). This model is typically used in the analysis of the nested dataset to understand the energy consumption patterns and the main sources of energy loads.

This model is based on the analysis of variance ANOVA and regression, which include random and fixed effects (Snijders & Bosker, 1999). They could be linear or any nonlinear function. Moreover, this model allows analysis for the macro-level, micro-level, and cross-level interaction. This cross-level interaction represents the relationship between the different levels; for instance, it illustrates the relationship between the individuals in the zones in level-1 and the floors in level-2(Finch et al., 2016).

4. CASE STUDY

In Chulalongkorn university, the Chamchuri 5 building has an area of 11,700 square meters and consists of seven floors (Pipattanasomporn et al., 2020). The first and second floors have a similar architectural design that stands out from the rest of the floors in the building. Each story is divided into several zones that serve different functions in the building as shown in (Table1) below.

Figure 3: Chamchuri 5 building layout,(CUBEMS,2021)



Moreover, the adoption of a building energy management system (BEMS) in the Chamchuri 5 building allowed the generation of large amounts of granular data. Air conditioning, light and plugs loads in KW have been collected through a BEMS that is installed in the building (Figure 2) in order to study the energy consumption patterns. For instance, energy load peaks, detection of anomalies, zones that consume large amounts of energy, areas of improvement that can help in cost savings and energy efficiency. In addition, internal environmental data such as temperature (Celsius), humidity (%), and light (lux) are similarly available for each zone to add a deeper understanding of the energy trends. This dataset has a particularly granular level. To illustrate, the data are available for each zone on each floor at a one-minute interval for a period of 549 days starting from 1st of July 2018 until 31st of December 2019 in a csv file format. Data are available for each AC unit, lighting, and plug loads in addition to the environmental data.(Pipattanasomporn et al., 2020)

Table 1: Functions of zones in the Chamchuri 5 building, (CUBEMS, 2021).

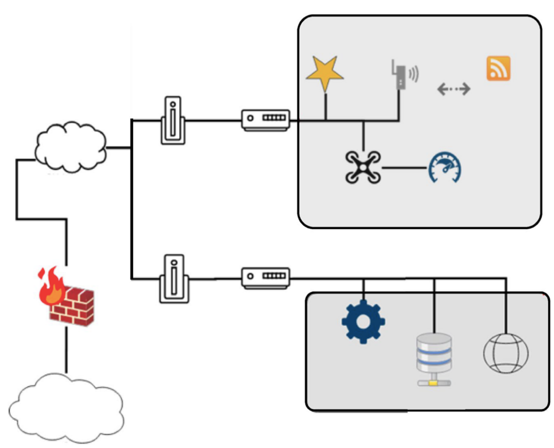
ZONES' FUNCTIONS					
ZONE / FLOOR	1	2	3	4	5
FIRST	Sitting Hall	Electrical engine room	Central stair hall	Around the building	-
SECOND	Registration and processing office	Admin department, registration & processing office	Central stair hall	Procurement and finance department	-
THIRD	University resource management center	Finance department	Central stair hall	Planning and information work	Analyze the project & admin activities
FOURTH	Infrastructure department	Planning, design, and physical system information department	Central stair hall	Accounting department	Building department
FIFTH	Mission group civil servants and employees	Mission, benefits, and personnel relations	Central stair hall	HR development department	HR development department
SIXTH	Office of strategy management	Office of research administration	Central stair hall	Office of academic admin	Office of academic admin
SEVENTH	Registration line inspect & procurement parcels	Legal library & procurement organization	Central hall	Office 3	Legal & legal center

4.1. BEMS Description

The overall Chulalongkorn University-BEMS system includes 21 energy monitoring units (EMU), 30 digital meters, 24 multi-sensors, 7 gateways, and a CU-BEMS server. Whereas, EMUs, multi-sensors, gateways, and the server have been developed in-house (Pipattanasomporn et al., 2020).

- EMU: used to measure the power loads and transfer the readings in watts through "Ethernet LAN with Modbus protocol" (Pipattanasomporn et al., 2020). The EMUs are designed to store data in case of connection problems to avoid any data loss. In addition, all the devices are calibrated before usage. The EMU readings have been estimated to have an average of 1% error.
- Digital meters: basic Siemens meters are being used to measure large AC compressors' loads and communicate the readings through Modbus TCP.
- Multi sensors: these sensors were made at the university to measure the following environmental data; temperature in Celsius, humidity in percentage, and the light in lux using a wireless network to communicate the readings.
- Gateway: the gateways have been developed at the university as well to collect the readings each minute from the multi-sensors.

Figure 4: Chamchuri 5 building BEMS components,(Pipattanasomporn et al., 2020)



As is normally found in data records, the datasets have some missing readings. The data of the period from the 15th of September 2018 to the 5th of March 2019 were missing due to the maintenance of the system, this period was ignored during the calculation of data availability. In most of the sensors, data are available for 95% of the time or more (Pipattanasomporn et al., 2020). In this case study, for simplification, only the data from the 6th of March 2019 onwards are used to avoid the calibration period and have more robust results. Furthermore, the elevator loads as well as the emergency exit signs' loads, were not recorded, they are assumed to be 1 to 2% of the total building load (Pipattanasomporn et al., 2020).

4.2. Data Analysis & Results

The datasets have been used to study the energy consumption of the building on both a macro and micro level and understand the behavior of the consumption, the patterns for the building, floors, and zones as well as identifying the main contributors of energy usage. The files are in a csv format, each floor has a separate file.

As previously mentioned, high granular data from the 6th of March 2019 were used, resulting in a total of 301 days; it is a one-minute interval data. Therefore, we have 433,440 minutes, which are equivalent to the number of rows for each floor. The number of columns varies from one floor to another as it depends on the number of AC units, lighting, plugs, etc. First, the average power consumption has been calculated for each floor, then for the light, AC and plug separately as well as the total for each floor which was added as separate columns in the data frame for each floor, their values could be seen in (Table 2). These calculations help to get an insight into the main factors of energy consumption. Afterward, the energy consumption values in megawatt-hour (MegaWh) are added together to give the total value of building consumption.

4.2.1. Macro Results

As seen in (Table 2) and (Figure 3), the first floor has the highest values in the AC, light, and plug loads resulting in the highest total energy consumption. It is evidently clear that the main source for the high consumption of the first floor is the lighting, it represents 49.66% of the first-floor consumption. This might be due to the fact that it contains the main hall of the building where everyone passes through to access the rest of the building, moreover, it has the surrounding of the building, therefore, the light might be continuously turned on to accommodate usage. The second highest consumer of energy is the seventh floor followed by the second, fifth, third, and fourth floors. While the sixth floor has the least energy consumption.

Table 2: Energy consumption values per floor in megawatt hour (MWh) and energy consumption percentage % with respect to the minimum values

Floor	1	2	3	4	5	6	7	Total Building
Total Energy MWh **% difference**	731.9 491%	207.9 68%	155.3 25%	153.9 24%	159.8 29%	123.9 0%	220.1 78%	1752.9
Light MWh **% difference**	363.4 1067%	31.2 0%	43.9 41%	31.1 0%	40.8 31%	39.2 26%	42.9 38%	592.6
AC MWh **% difference**	238.8 216%	164.8 118%	100.7 33%	107.4 42%	106.6 41%	75.5 0%	171.7 127%	965.4
Plug MWh **% difference**	129.6 2238%	12 116%	10.6 91%	15.4 177%	12.5 125%	92 66%	5.5 0%	194.8

Even though many floors share the same architectural plans, yet they have different consumption patterns, this might be due to numerous factors one of them is the diverse functions of the different zones. However, the exact reason for the difference in consumption is unknown due to the lack of information availability provided by the university. For instance, the first floor has almost 5 times the values of the sixth floor in the total consumption (Table 3). Similarly, the sixth floor has the least AC consumption. Moreover, the first floor has the highest plug consumption, this might be due to the fact that it has the electric engine room. Besides, the second floor has the minimum light consumption. In addition, the second floor has the least light values while the seventh floor has the least plug values. Unlike the first floor, the rest of the floors have high AC consumption varying from 61% to 78% of their energy consumption. As a result, almost half of the energy consumed by the Chamchuri 5 building is AC. This can be interpreted by the nature of Bangkok's weather that requires AC to achieve the comfort zone. For a university campus, it is quite important to ensure such comfort for the employees to ensure a suitable working environment contributing to higher productivity and ability to concentrate. Moreover, it is noticed that the plug loads represent the least percentage of the total energy consumption for each floor as well as the building as a whole. To illustrate, it represents 11% of the energy of the Chamchuri building. In the same way for the rest of the floors, the plugs' load varies from 3% to 18% (Figure 4).

Figure 5: AC, plug & light repartition in each floor

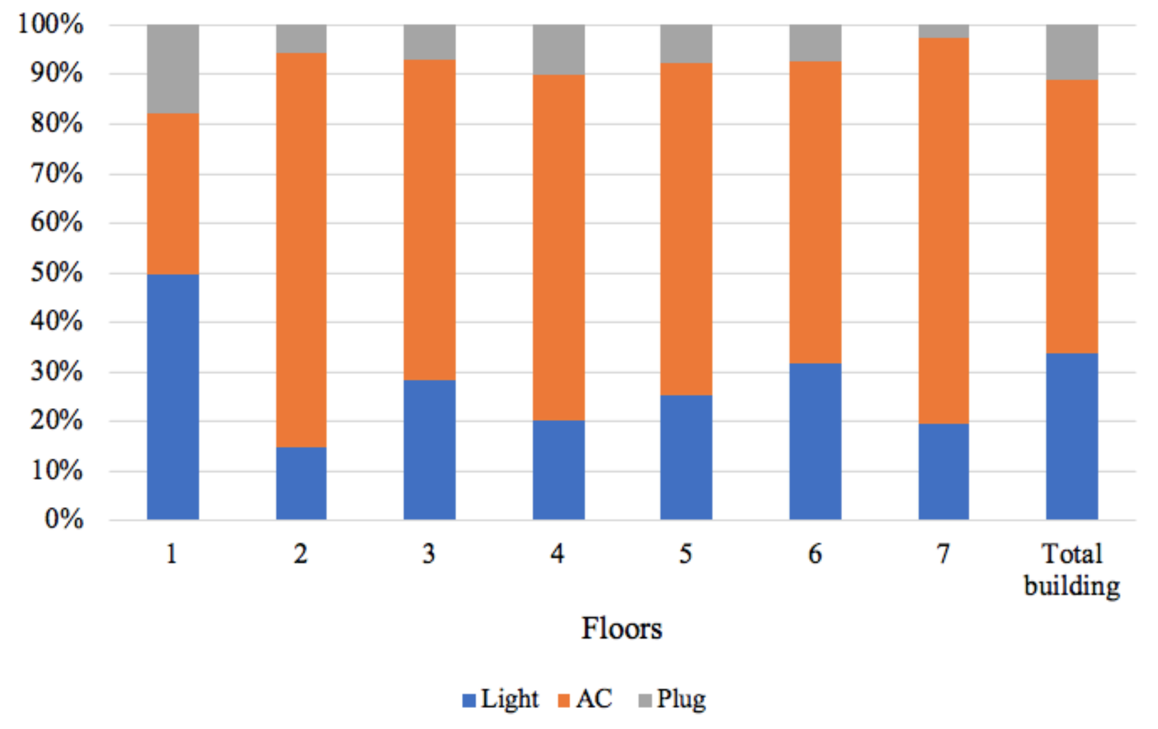
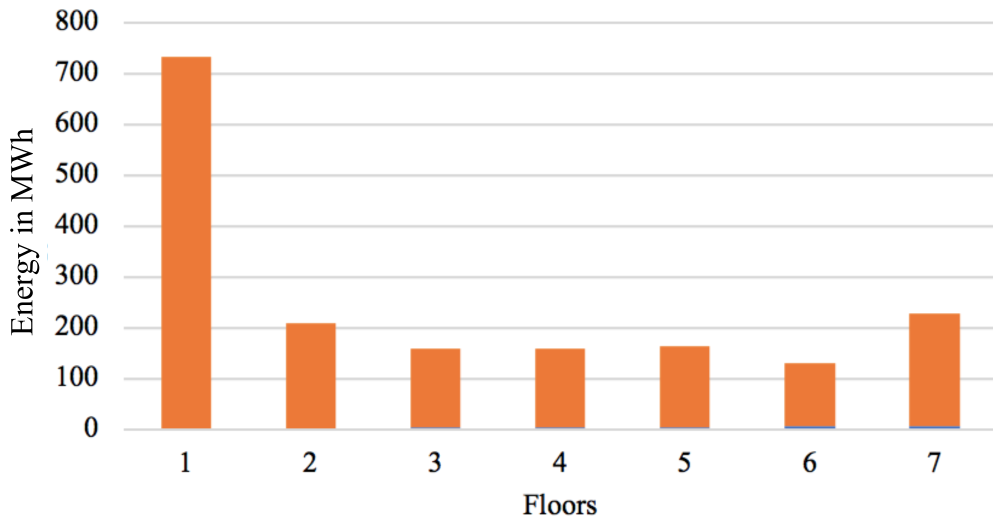


Figure 6: Total energy consumption in each floor of the Chamchuri 5 building in MWh



4.2.2. Micro Results

Finally, going more in depth in the analysis by studying each zone. The below table shows the microanalysis of the total energy consumption of each zone in each floor in MWh.

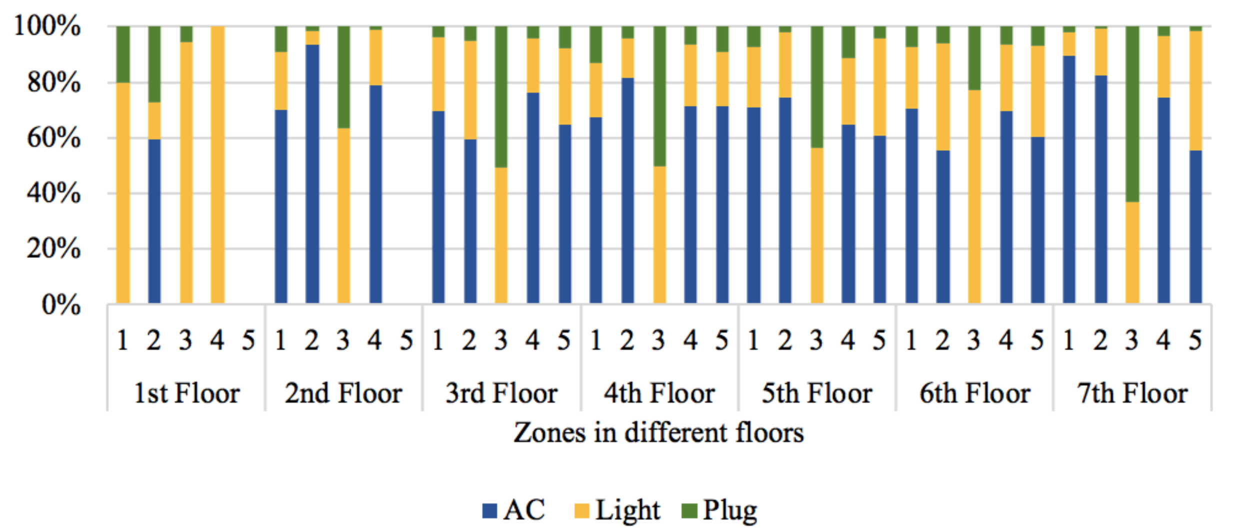
Table 3: Energy consumption values per zone in MWh

Zones \ Floors	1	2	3	4	5	6	7
1	58.560	90.566	47.577	52.829	52.688	34.208	70.547
2	400.226	91.105	43.791	30.787	37.191	37.043	56.268
3	157.674	6.087	6.641	4.872	4.608	6.471	2.029
4	115.416	20.190	37.712	46.753	43.820	32.533	59.503
5	-	-	19.553	18.693	21.528	13.621	31.750

The first and second floors only have four zones while the rest have five. The values presented in (Table 3) show that zone one is the main energy consumer for all the floors except for the first and second floors. This could be proportional to the zone's relatively large area and the nature of the zone function. While the second zone is the main one for the first and second floors.

The least consuming zone is zone 4 for the first floor. While the least one for the rest of the floors is zone 3, I believe this is due to the fact that zone 3 represents the staircase and is not air-conditioned. As mentioned before the air conditioning system is the main source of energy consumption in the building; thus, eliminating it from zone 3 it reduced its consumption significantly. In addition, the staircase doesn't need appliances like an office that need computers, printers, etc., for instance. As a result, the energy loads decrease. Moreover, zone 2 on the first floor has the highest energy consumption in the whole building this might be due to its function, which is an electrical engine room.

Figure 7 Energy repartition for each zones 1,2,3,4,5 in the different floors



The AC accounts for more than 50% of the energy consumption per zone whenever for the zones that contain AC. This is expected since the HVAC electric loads are known to be the biggest contributor. However, it can be seen that there is a considerable variation between the AC percentages between different zones in the different floors, an energy pattern couldn't be deducted from the result. For instance, in zone 2 floor 2, the AC consumption exceeds 90%, while for zone 2 floor 3, the AC consumption is around 60%. BEMS can act according to these variations to help in managing the energy consumption of the building. Moreover, the plug consumption seems to be less than the light consumption for most of the zones, which correlates with the macro analysis. Furthermore, zone 4 on the first floor has a significantly high percentage of energy consumed in lighting compared to the other zones. To illustrate, this zone represents the surrounding of the Chamchuri 5 building thus, it doesn't require any air conditioning system nor plugs. Therefore, 100% of its energy consumed is light., which is explained by the nature of the function of this zone.

4.3. Multilevel Modeling

Multilevel modeling is one of the techniques used to deal with this type of longitudinal, clustered data. As previously mentioned, the dataset consists of three levels: level-1 unit the zones are nested in level-2 the floors, and finally nested in level-3, which is the building itself. Hence, the energy loads could be affected by the different zones. This could be due to the fact that each zone serves a different purpose, the orientation of the zone in the building affects the heat absorption of the walls, the size of the windows, the different appliances in the zone, the behavior of the occupants in the space, etc.

To apply the multilevel model, first data must be cleaned and restructured. The readings for the power loads are recorded every minute, however, this interval can be small compared to the total time that is used for this study, 301 days. Therefore, for better visualization, the daily power loads are calculated by computing the algebraic average readings for each day. Then, the multi-level modeling is executed based on the daily loads.

Moreover, the data contain missing readings, thus, the data need to be cleaned. The missing data can be categorized into two cases. The first one, if the readings are known at the previous and preceding time intervals, i.e., if the reading at minute 3 is missing and the readings at minutes 2 and 4 are available, then an estimation at minute 3 is obtained via linear interpolation between the available readings. This assumption can be valid since the effect of the interpolation on daily consumption is insignificant. The second case happens when two consecutive readings or more are missing. In that case, the whole day is eliminated from the study since an estimation might be misleading and will not present the actual energy consumption. Furthermore, if the data are missing for only one of the load categories, and it belongs to the second case, the day is removed for the floor. The cleaning of the data was performed by using if conditions for the whole dataset. The number of removed days for each floor is 13, 8, 29, 109, 90, 9, and 21 from floors 1 to 7 respectively. Around 30% of the days were eliminated for floors 4 and 5, while less than 10% of the days for the rest of the floors. Afterward, a table was made to group the seven csv files together and restructured to represent the different levels: the building, floor 1, 2,..., 7, zone 11,12,…57. Each zone is represented with two digits, the first one represents the zone number, while the second one represents the floor number. For example, zone 12 denotes zone one on floor two. In this model, two levels will be used as the third level has only one building.

In Matlab, the function lme was used to fit a linear regression model for the variables in the dataset or table. First off, a model is defined to predict the energy loads in which regressors may include zone-level features and, floor-level features. The syntax of the Matlab function that represents the two-level model is as follows:

- lme= fitlme(tbl,formula)

  lm_group_AC = fitlme(tbl,AC ~ Day +Zones+ (Floor|Zones)')

  lm_group_Light = fitlme(tbl,Light ~ Day +Zones+ (Floor|Zones)')

  lm_group_Plug = fitlme(tbl,Plug ~ Day +Zones+ (Floor|Zones)')

  lm_group_Total = fitlme(tbl,Total ~ Day +Zones+ (Floor|Zones)')

Where the dataset is presented in a table named “tbl”, “Total” is the total power load per day, and “Day” is the time interval used. Total, AC, Plug, and Light are used each separately in a model as the dependent variable. “Zones”, and “Floor” represent the different clusters. Zones and Day are interpreted in the models as independent variables. The formula is written using the Wilkinson notation (Wilkinson,1973). The nesting structure is incorporated in the model, the higher level is written first then the lower level in the syntax. Additionally, cross interaction between variables could be included in the model if needed, this represents the case when variables impact each other. Multiple models could be used in the analysis, comparisons could be performed between them to identify the best fit for the data. This assessment could be done by comparing the values of AIC and BIC of the used models. Smaller values for the AIC and BIC refer to a better model, for example, it helps whether to include a variable or not in a model (Finch et al., 2016). For the Chamchuri 5 building, four models were used to model the AC, Plug, Light, and total energy respectively. The results of these models are shown in table 4.

Table 4 shows the results for the fixed effects. It includes the name of the variables and the estimate’s value for each of the four models. Moreover, the star notation is added to show the significance of the estimate. These values will be useful to interpret the model. For instance, it highlights the zones that have high consumption patterns, one zone can consume a lot of AC while another one can have the light as its main source of consumption. Therefore, the energy optimization’s solutions are different from one zone to another. To illustrate, the p-values for each variable are an indicator of whether it is significant or not in the performed study. The null hypothesis states that there is no correlation between the random effect and the response. A p-value that is smaller than alpha indicates a statistically significant variable and vice versa. As seen in table 4 with a 95% confidence level, in the light model, zone 31 estimate’s value is equal to

14.114***. This result demonstrates that zone 31 consumes a lot of light compared to zone 11 which is due to its functionality, as zone 31 is the central hall. Light sensors could be installed in this area to reduce the consumption of light. Another example, in the AC model, the p-values for zone 3 in all the floors (Zone 31, 32,33,34,35,36,37) are bigger than alpha. Thus, these zones are not significant. Which correlates with the previous results from the microanalysis, which states that zone 3 represents the stairs and has no AC. For the same model, the remaining variables are significant. In all of the light and plug models' estimates, their p-values are smaller than alpha which indicates that they are significant. Therefore, we can reject the null hypothesis. The estimates' values are relative to zone 1 on floor 1. Therefore, a positive value means that this variable has a higher energy consumption and vice versa. For example, the AC estimate for zone 21 is equal to 33.191***, because this zone has the highest AC loads and has much larger loads compared to zone 11. While some other zones have estimates' values that are quite negative, this might be due to their function, as some of them are offices, they might consume less energy than the reference zone 11. Besides, the higher the value of "Day," the lower the response value will be for all the models except for the light. This could be explained by the fact that as we move further in time, we approach winter thus the energy loads resulting from AC and plugs decrease while more light loads are needed.

Finally, looking at the results of the random effects in (Appendix 1), we can deduce that the patterns of the energy consumption somehow are correlated between the floors, as the correlations' values are relatively high.

Table 4 Fixed effects results from Matlab

Name	AC Estimate	Light estimate	Plug estimate	Total estimate
Intercept	0.90082***	6.3909***	1.7098***	9.0016***
Zone_21	33.191***	0.75038***	13.454***	47.395***
Zone_31	-1.0213e-14	14.114***	-0.43935***	13.675***
Zone_41	-1.0671e-14	9.556***	-1.6496***	7.9064***
Zone_12	8.8234***	-3.9096***	-0.50134***	4.4125***
Zone_22	11.861***	-5.8891***	-1.4724***	4.4994***
Zone_32	0.0028623	-5.9787***	-1.3418***	-7.3176***
Zone_42	2.2196***	-5.9517***	-1.6201***	-5.3522***
Zone_13	4.7283***	-4.7922***	-1.3966***	-1.4604***
Zone_23	3.7433***	-4.4139***	-1.3349***	-2.0055***
Zone_33	0.028547	-6.0667***	-1.1827***	-7.2208***
Zone_43	4.1344***	-5.5185***	-1.4233***	-2.8074***
Zone_53	1.8328***	-5.7752***	-1.4419***	-5.3843***
Zone_14	5.0801***	-5.0341***	-0.5857***	-0.53974
Zone_24	3.798***	-5.8998***	-1.4399***	-3.5417***
Zone_34	0.23763	-6.1512***	-1.2071***	-7.1207***
Zone_44	4.7388***	-5.0776***	-1.2246***	-1.5634***
Zone_54	2.0958***	-6.0525***	-1.3828***	-5.3395***
Zone_15	4.9481***	-4.912***	-1.119***	-1.0829*
Zone_25	3.8716***	-5.31***	-1.5179***	-2.9564***
Zone_35	0.15924	-6.1815***	-1.3602***	-7.3825***
Zone_45	3.8581***	-5.0978***	-0.96689***	-2.2066***
Zone_55	1.8841***	-5.5264***	-1.5128***	-5.1551***
Zone_16	4.1211***	-5.1071***	-1.2143***	-2.2002***
Zone_26	3.4824***	-3.9019***	-1.2537***	-1.6733***
Zone_36	-0.21569	-5.5823***	-1.3957***	-7.1937***
Zone_46	3.8736***	-5.0683***	-1.2927***	-2.4875***
Zone_56	1.2683***	-5.6712***	-1.4986***	-5.9015***
Zone_17	8.7438***	-5.6905***	-1.4675***	1.5858***
Zone_27	6.436***	-5.2265***	-1.5783***	-0.36886
Zone_37	0.031303	-6.4107***	-1.4664***	-7.8458***
Zone_47	6.1528***	-4.6553***	-1.3768***	0.12065
Zone_57	2.4451***	-4.6087***	-1.5728***	-3.7364***
Day	-0.0060166***	0.00081564***	-0.00040228***	-0.0056032***

$* \rho - value < 0.05$, $** \rho - value < 0.01$, $*** \rho - value < 0.001$

5. RECOMMENDATIONS & CONCLUSION

There exist some limitations in this study, some of these are the lack of some knowledge and data about the building. For instance, the study is performed only on the data from March to December 2019 due to missing data, and even in this period still, there are some missing data. Therefore, better data availability will improve future analysis and reduce results bias. Additionally, the time intervals of the missing data are not the same on all the floors. To solve this issue, more sensors could be added to replace the other ones during maintenance. Moreover, data including supplementary metrics could be used as well for a deeper analysis. For example, data collected through additional sensors can help to analyze hidden patterns and correlations between variables affecting the response. In addition, knowledge regarding the insulation of the building, area of the spaces in the building, the exact appliances used, lighting system, shading devices, type of AC that is used in the building, and occupants' behavior and number. For example, the number of occupants in each zone can vary with time, this information can be added to the model. The aforementioned data and knowledge can provide a better understanding of energy usage and can enable analysts to find solutions for energy optimization. For example, insulation material could be proposed, a control system can be implemented, energy efficiency solutions can be recommended, etc. Moreover, for future work studying the energy supplies could be helpful as it can have an impact on the energy consumption patterns of the building. Further, the energy harvested from the solar PV panels installed can be better utilized and stored by the information obtained from the modeling.

6. REFRENCES

Asare-Bediako, B., Kling, W. L., & Ribeiro, P. F. (2012). Home energy management systems: Evolution, trends and frameworks. *Proceedings of the Universities Power Engineering Conference*. https://doi.org/10.1109/UPEC.2012.6398441

Cao, X., Dai, X., & Liu, J. (2016). Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and Buildings*, *128*. https://doi.org/10.1016/j.enbuild.2016.06.089

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V, & Rong, X. (2015). Data mining for the internet of things: Literature review and challenges. In *International Journal of Distributed Sensor Networks* (Vol. 2015). https://doi.org/10.1155/2015/431047

Finch, W. H., Bolin, J. E., & Kelley, K. (2016). Multilevel Modeling Using R. In *Multilevel Modeling Using R*. https://doi.org/10.1201/b17096

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). Longitudinal data analysis. In *Longitudinal Data Analysis*. https://doi.org/10.4324/9780203929483-16

Gaber, M. M., Aneiba, A., Basurra, S., Batty, O., Elmisery, A. M., Kovalchuk, Y., & Rehman, M. H. U. (2019). Internet of Things and data mining: From applications to techniques and systems. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 9, Issue 3). https://doi.org/10.1002/widm.1292

Pipattanasomporn, M., Chitalia, G., Songsiri, J., Aswakul, C., Pora, W., Suwankawin, S., Audomvongseree, K., & Hoonchareon, N. (2020). CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-020-00582-3

Rao, I. K. R. (2014). Data Mining and Clustering Techniques Data Mining Techniques.

ResearchGates, *January*.

Ruiz, G. R., & Bandera, C. F. (2017). Validation of calibrated energy models: Common errors. *Energies*, *10*(10). https://doi.org/10.3390/en10101587

Singer, J. D., & Willett, J. B. (2009). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. In *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Snijders, T. A. B., & Bosker, R. J. (1999). An Introduction to Basic and Advanced Multilevel Modeling-Sage Publications Ltd (1999).pdf. In *Book*.

Torcellini, P., Pless, S., Deru, M., & Crawley, D. (2006). Zero Energy Buildings: A Critical Look at the Definition. *ACEEE Summer Study Pacific Grove*.

Yang, T., Clements-Croome, D., & Marson, M. (2017). Building Energy Management Systems. In *Encyclopedia of Sustainable Technologies*. https://doi.org/10.1016/B978-0-12-409548-9.10199-X

7. APPENDIX

Appendix 1: Correlations Matrix for the random effects

AC							
Floor	**Intercept**	**2**	**3**	**4**	**5**	**6**	**7**
Intercept	1						
2	0.93604	1					
3	-0.97563	-0.91323	1				
4	-0.99873	-0.93486	0.97439	1			
5	-0.97672	-0.91426	0.95292	0.97549	1		
6	-0.99489	-0.93126	0.97064	0.99362	0.97173	1	
7	-0.97652	-0.91407	0.95273	0.97529	0.9538	0.97153	1

Light							
Floor	**Intercept**	**2**	**3**	**4**	**5**	**6**	**7**
Intercept	1						
2	0.88339	1					
3	-0.99956	-0.883	1				
4	-0.99659	-0.88037	0.99615	1			
5	-0.99995	-0.88335	0.99952	0.99654	1		
6	-0.99516	-0.87911	0.99473	0.99177	0.99511	1	
7	-0.99266	-0.8769	0.99223	0.98927	0.99261	0.98786	1

Plug							
Floor	**Intercept**	**2**	**3**	**4**	**5**	**6**	**7**
Intercept	1						
2	0.754	1					
3	-0.9991	-0.75332	1				
4	-0.99677	-0.75157	0.99587	1			
5	-0.99614	-0.75109	0.99524	0.99293	1		
6	-0.98169	-0.7402	0.9808	0.97852	0.97791	1	
7	-0.98859	-0.7454	0.9877	0.9854	0.98478	0.97049	1

Total							
Floor	**Intercept**	**2**	**3**	**4**	**5**	**6**	**7**
Intercept	1						
2	-0.94011	1					
3	-0.99905	0.93922	1				
4	-0.99725	0.93752	0.9963	1			
5	-0.99124	0.93187	0.9903	0.98851	1		
6	-0.99818	0.9384	0.99723	0.99544	0.98944	1	
7	-0.99912	0.93928	0.99817	0.99637	0.99036	0.9973	1

Appendix 2: Energy repartition for each zone in the different floors