Modelling high-education buildings: A top-down modelling approach to disaggregate end-uses

Nuno Miguel Martins Senra Almeida

Thesis to obtain the Master of Science Degree in

Mechanical Engineering

Supervisors: Dr. Diana Pereira Neves

Prof. Carlos Augusto Santos Silva

Examination Committee

Chairperson: Prof. Edgar Caetano Fernandes

Supervisor: Dr. Diana Pereira Neves

Member of the Committee: Dr. Ricardo Manuel Anacleto Gomes

December 2021

ACKNOWLEDGEMENTS

I wish to thank my supervisors for time and effort they provided in guiding me in this thesis, Professor Carlos Silva and in particular Dr Diana Neves for always being available to help me developing the work.

I also want to express my gratitude to the Campus Sustentável project for providing important data for this work.

Finally, I want to thank my family and friends for supporting me and keeping me focused and motivated.

ABSTRACT

Countries are making an effort to reach environmental goals to fight climate change. Buildings play a big part in the countries' energy consumption, thus steps are being taken to reduce it through energy efficiency measures. To do so, it is important to understand the energy demand in buildings and its influencing parameters, having become the focus of many research studies. Universities can and should lead this pursuit of sustainability. However, having access to complex data regarding energy consumption is frequently a challenge.

The purpose of this work is to develop a model to forecast end-uses energy demand in university buildings, using data commonly available in universities, such as a total electricity load and equipment's and activity periods data, and try to find a relation between energy consumption and the type of activities of the building.

Two buildings in the Alameda campus of IST served as case study: the Civil Building for model validation, and South Tower to test the extend of the model performance. Results show better performance for the Civil building than the South Tower, with an error of 12.57% and 22.36%, respectively. However, they show potential to be improved. The relation between energy consumption and types of areas in the building, seems to be greatly influenced by the existence of technical rooms, and other spaces such as offices and labs. The building's electricity intensity for university buildings is found to range from 64.9 to 131.7 kWh/m^2. *HVAC* is the most consuming end-use (24.62-70.53 kWh/m^2), followed by *lighting* (16.97-30.53 kWh/m^2) and *electronics* (15.26-20.24 kWh/m^2). Other equipment like pumps, elevators and lab equipment have a low significance (8.05-9.96 kWh/m^2).

Key-words: Electricity demand; Energy in buildings; Modelling; Genetic Algorithm Optimization (GA); Building Energy Intensity

RESUMO

Vários países têm se focado em combater as alterações climáticas. Para este efeito, estes países têm se comprometido a atingir diversas metas ambientais no que respeita ao consumo energético e às emissões de poluentes, nomeadamente CO_2. Para alcançar estas metas é necessário reduzir o consumo total e investir na evolução da energia renovável. Apesar destas medidas dependerem da ação dos governos para setores como o dos transportes e da indústria, o setor dos edifícios pode acelerar este processo. Este setor contribui em grande parte para o consumo total de cada país e por isso é crucial reduzir o seu consumo através de medidas de eficiência energética. Mas primeiro é necessário compreender o consumo energético nos edifícios e os parâmetros que o influenciam, através de estudos. Sendo universidades locais de investigação e inovação científica, estas podem e devem ser líderes nesta procura de sustentabilidade. No entanto, para tal acontecer é necessário ter acesso a dados relativos ao consumo energético, algo que nem sempre é possível.

Assim, o objetivo deste trabalho é desenvolver um modelo de desagregação energética em universidades nos diversos consumos finais, que utilize informação usualmente acessível a várias universidades. Estes dados incluem consumo elétrico total e dados relativo a equipamentos e aos períodos de atividade, como o calendário escolar. Ademais, uma análise será feita de modo a obter uma relação entre o consumo energético e os tipos de atividades que ocorrem nos edifícios.

Para este trabalho dois edifícios do campus da Alameda do IST serviram de caso de estudo. O edifício de Civil serviu para validar o modelo, enquanto a Torre Sul serviu de teste de implementação do modelo noutros edifícios, de modo a analisar a performance do modelo. Resultados do caso de estudo mostraram que o modelo se comportou melhor para o edifício de Civil do que para a Torre Sul, apresentando erros relativos de 12,57% e 22,36%, respetivamente, contudo os resultados apontam para possíveis melhorias. A relação entre o consumo energético e o tipo de espaços por edifício aparenta ser fortemente influenciada pela existência de salas técnicas e também por outros espaços, nomeadamente gabinetes e laboratórios. Os resultados mostram que a intensidade energética de edifício para edifícios universitários varia entre 64,9 e 131,7 kWh/m^2. O principal consumo final é o AVAC, variando entre 24,62 e 70,53 kWh/m^2, seguido da iluminação (16,97-30,53 kWh/m^2) e de equipamentos eletrónicos (15,26-20,24 kWh/m^2). Finalmente, outros equipamentos como bombas de drenagem, elevadores e equipamento de laboratório têm pouco impacto no consumo total do edifício (8,05-9,96 kWh/m^2).

Palavras-chave: Consumo de Eletricidade; Energia em Edifícios; Modelação; Algoritmos Genéticos (GA); Intensidade Energética de Edifício

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 Motivation

The demand for electricity worldwide has been increasing each year having reached approximately 24740 TWh in 2018 [1]. The transportation and industry sectors have been responsible for most of the electricity use, however with the increasing investment on HVAC systems to provide thermal comfort, the electricity demand in residential and service buildings has increased considerably in the past years. From 1990 to 2000, the building consumption related to cooling increased by 14.6% each year, and it is expected that this consumption will increase at a rate of 3.4% a year from 2000 to 2030 [2].

Buildings account, on average, for around 40% of the energy use in Europe and one third of the CO_2 emissions (Figure 1 and Figure 2, respectively).

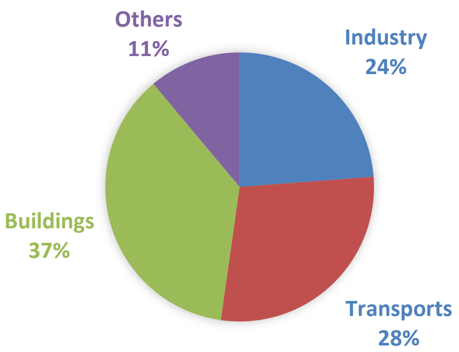

Figure 1 - Total energy consumption per sector in Europe (2019) [3]

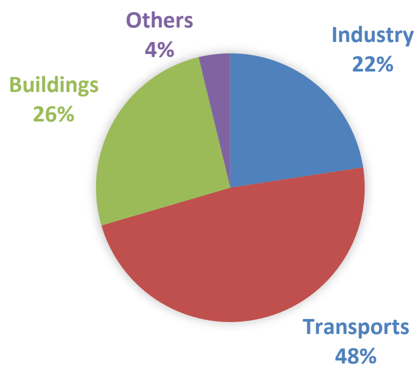

Figure 2 - CO_2 emission per sector in Europe (2019) [4]

It is believed that more than 50% of the energy consumption could be reduced by implementing measures to improve energy efficiency [5]. In terms of final energy consumption in buildings, natural gas has the highest share (36%), followed by electricity (32%). In Portugal, however, the energy demand in buildings is only 30% of the total national energy consumption [2]. This is due to the more moderate climate in the Iberian Peninsula that reduces the need for cooling and specially heating. This leads to a higher share of the energy demand in buildings being attributed to electricity (41% in residential and 77% in service buildings) [5].

Following the Council of Ministers approval of the “ECO.AP 2030“ [6], Public Administration entities have pledged to reduce the energy consumption of their buildings by 40% until 2030. Many universities have committed to reduce the energy consumption in their campi [7]. Knowing the energy demand of each building is important in order to analyse the current situation and find more efficient energy systems.

Modelling energy demand has, therefore, become of crucial importance. However, for most buildings this is still a difficult process due to the lack of public access to monitoring or detailed information regarding either its occupancy and their behaviour, as well as specific information about the energy systems installed in the buildings or their constructive solutions. Thus, it is of high importance to understand better and forecast how energy is used, especially on services buildings where electricity demand is higher, implying a high environmental impact.

1.2 Objectives

This study aims to address this problem by developing a top-down modelling framework to assess buildings' end-uses electricity demand by analysing the hourly load of total electricity demand, equipment use by type of space and activity period, developing an optimization model to estimate the main end-uses (*lighting*, *HVAC*, *electronics*, and *Others*). This information will allow institutions to analyse their campi buildings in more detail, comparing and benchmarking with other high education institutions, by an electricity intensity index by type of end-use, being able to address possible causes of inefficiency on electricity demand, and readjust if needed, without harming or influencing end-users' comfort.

By convenience of accessing data and for model validation purposes, this study will use IST as case study.

This study has the following goals:

- Analyse the hourly electricity demand of a higher education building;
- Characterize equipment use by type of area and activity period;
- Explore correlations between different electricity demand influencing parameters;
- Disaggregate the electricity demand by end-use;
- Report a range of end-uses electricity demand index per area unit (kWh/m2) according to each type of building.

1.3 Structure of the thesis

This study is organized as follows:

- Section 2: A comprehensive literature review about the different modelling methodology approaches and forecast optimization models is presented;
- Section 3: the methodology regarding how the correlation between different influencing parameters of electricity demand in buildings is described followed by how the model was developed;
- Section 4: a description of the case study characteristics and data collection is reported;
- Section 5: the model validation with IST case-study is performed, followed by the analysis of the model implementation to other buildings;
- Section 6: the final statements and conclusions are made.

2 Literature review

2.1 Modelling electricity demand in buildings

Modelling electricity demand in buildings has been the focus of many studies due to the large impact to the total energy demand and CO_2 emissions worldwide.

Specifically, for services' buildings there is a lot of energy overconsumption due to the lack of personal responsibility from the occupants. Studies show that implementing feedback systems of the energy consumption in residential buildings lead to energy reduction since the occupants have the incentive to reduce their bills [8,9]. However, these incentives do not exist for most service buildings occupants leading to poor equipment usage, such as leaving the lights or electrical equipment on after being used. During the period from 2000 to 2018 the energy consumption in the residential sector has decreased by 4.5%, while in the services sector the energy consumption has increased by 20.2%. In 2018 service buildings accounted for 34.5% of the energy demand in buildings and this value is expected to increase [10].

These studies have different purposes, such as forecasting electricity demand, analysing the effects of retrofitting measures or studying the parameters that influence energy consumption in buildings. To do so, these studies rely on various types of data, such as total and end-use load profiles, occupancy schedules, weather data. However, to effectively model electricity demand in buildings, there is the need for higher granularity of data, normally with hourly time-steps, which is not easy to access. A study intending to size an electric generation system found challenging to generate load profiles since the local utility company would only supply data from the previous month [11].

Smart meters can provide hourly or sub-hourly data on electrical load, allowing for more detailed information of a building or house. However, overall monitoring is not enough to totally understand consumption in a building. In order to have a more detailed and customized energy efficiency measures, either machine learning algorithms or a higher number of smart meters would be needed, which requires a big investment making it unviable for a larger institution [12].

Thus, it is important to understand which are the influencing parameters in the building's energy consumption. There are many different parameters that can influence the energy consumption in a building, such as climatic factors (temperature, solar radiation, geographical location), construction solutions (insulation, architecture) or activity related parameters. For example, more moderate climates have lower HVAC needs, while locations further from the Equator line have higher lighting demand due to less daylight time. F. Amara et al [13] found that temperature is one of the most important factors that influence energy consumption. R. Mena et al [14] studied the energy consumption on a solar energy research centre and found that temperature and solar radiation were both influencing factors.

Occupancy and schedules also impact the energy consumption. Not only in the more obvious way, more people in a room, the bigger the electricity demand will be, but also the schedule of occupancy. The type of activity in the building is also an influencing factor. An office will consume more electricity on computers and other electronic equipment, while gyms will spend more energy on HVAC. X. Gui et al [7] studied the impact that changing the academic calendar could have on the energy demand. They found that by changing from a semester to trimester system could reduce the electricity consumption.

The building's characteristics and construction materials also have an impact on building consumption. The walls' materials could present more or less insulation, leading to different thermal gains. The same for the window area and material, type of glazing, or the orientation of the building, all these affect the thermal gains of the building. Therefore, there are many building characteristics that influence the envelope gains or losses, and consequently, the energy demand of the building.

B. Arregi & R. Garay [15] analysed the energy consumption of three different buildings before and after retrofitting measures. The buildings studied were a university in the UK, a school in Sweden and an hospital in Spain. They found that the university showed a clear relation between temperature and consumption before retrofitting. That relation disappeared afterwards, and the consumption became constant throughout the seasons. The other buildings did not show the same relation between temperature and energy demand. The hospital showed the least difference between energy demand during the day and night, due to the high equipment usage during the night.

J. C. Wang [16] conducted a study on energy consumption for education buildings of different education levels in Taiwan. Energy indicators per unit area were calculated and it was found that the higher the education level, the higher the indicators were. There was also a clear relation between the consumption and the temperature, since the months of higher temperature had higher energy consumption.

These studies show the importance of understanding the parameters that can affect building consumption, but also comprehending the type of activities that occur in the building being studied, since this influences the role of these parameters in the energy demand.

However, this information is not always available. While meteorological data is easily accessible for numerous places in the world, this is not the case for occupancy, schedules or even building characteristics for example. Several studies try to develop prediction and forecasting models of this influencing factors to overcome this problem.

Agnetis et al [17] try to develop models to predict the occupancy in buildings by using information from sensors and wireless sensor networks. K. Song et al [18] conducted a study with the goal of developing a tool that optimizes the university courses timetables to reduce the energy demand. This optimization would allow for an inexpensive energy efficiency improvement that would not require students to change their behaviours. Another study tried to find a correlation between occupation and energy demand, by detecting internet connections/disconnections and logging them throughout a period of three weeks [19].

All these studies highlight the importance of having access to information on energy consumption, load profiles, occupancy schedules and others to allow for research. Although there are various ways to get around the lack of more detailed information, it always requires assumptions and approximations. Therefore, a lot of work has been put on developing models that can forecast the energy consumption in buildings with information more easily accessible, such as weather data.

2.2 Optimization of prediction algorithms

There are many different models that try to predict energy consumption, and they are grouped into different methods. Stochastic time series models analyse time series data in order to separate it into different components, trend, seasonality and residual. The more used models from this family are the ARIMA, ARMA and autoregressive models. These models provide good results for linear systems, and when there is a large amount of data to analyse. However, they struggle for non-linear systems [20].

Regression models, like linear regression (LR) and multiple linear regression (MLR), try to find the relation between energy consumption and one or more parameters. Just as for stochastic models, they provide good results for linear systems, and they allow for a clear understanding of the relation between the independent and dependent variables. S. Asadi et al [21] developed a MLR model to understand the relation between consumption and 17 building characteristics for different building shapes. This allowed for a clear observation of the influence of each one of those parameters in the electricity demand.

These models are simple and easy to implement, however they still present issues for nonlinear systems. With the development of computers, artificial intelligence (AI) methods, such as artificial neural networks (ANN), support vector regression (SVR) and genetic algorithms (GA) have been the focus of many studies, since they can tackle this issue.

ANNs are one of the most used models in building consumption forecasting [22]. They are based on the way the human brain works, and try to replicate its functioning to solve problems, by creating a network of neurons that transmit information from one another. These neurons are divided into different layers, input, output, and hidden layers. There is one input and one output layer, but there could be more than one hidden layer. The input layer received the input data, while the output layer provided the results of the evaluation. The hidden layers are responsible for predicting the results based on the input. This is done through training, using previous data to understand the relation between input and output. ANNs also have drawbacks, as the training requires a significant amount of data, which is often lacking, and it is time consuming to train and to run the model.

D. Hawkins et al [23] studied the effects of adding different parameters to an ANN that predicts energy consumption. They found that the error reduced when building parameters were added to the ANN. H. Dagdougui et al [24] developed an ANN to forecast energy consumption in district buildings and analysed the influence of having different model design parameters, like the number of neurons, layers or inputs.

Support vector machine (SVM) is a machine learning algorithm based on statistical learning theory. It separates sets of data into two different groups. Usually, this data cannot be separated linearly so the SVM projects this data into a hyperplane of higher dimension, by using kernel functions. There are different kernel functions, and many studies about them, due to their importance in the performance of the model [25]. SVR is a variant of SVM used for regression problems, where the linear relation between the separated data is determined. SVR is a good model for pattern recognition, and like other AI models, it provides good results for nonlinear problems. However, it requires a large computation time, even larger than ANNs. As mentioned, the kernel functions are very important to the model, the wrong choice can lead to very poor results [26].

J. Massana et al [27] studied the influence of taking into account different parameters for three different models, a MLR, an ANN and a SVR. Generally, the ANN and the SVR provided significantly better results than the MLR, however they also have much larger computation times. Reducing the parameters used to only outdoor temperature and occupancy attributes improved the computation time significantly, while also producing results with low errors.

Genetic algorithm (GA) is a heuristic model inspired by Darwin's theory of evolution. It generates many different solutions, and the best ones survive to the next generation, so that the optimal solution, or at least a good solution, can be obtained. It is one of the most used models for optimization problems, especially for nonlinear optimization. Genetic algorithms can solve various types of optimization problems, regardless of their linearity, stationarity or if there is random noise. The model generates many different solutions, that are independent of one another, so many different possibilities can be evaluated, with each parameter in analysis having a wide range of possible exploration from the solutions. However, like the other AI methods, it has a higher computation time than stochastic methods. It is also important to select the appropriate parameters, specifically, the fitness function, the crossover and mutation rates and the selection criteria [28].

H. B. Gunay et al [29] used a genetic algorithm to help disaggregate the total loads of an academic office building into three different end-uses: lighting and plug loads, air handling units (AHUs), and chillers. The GA was used to determine the parameters that related each end-use load with information from equipment like fans and pumps, for example, when they are on or off, the air pressure or the temperature. They applied different constraints to the usage of each end-use, such as, the electricity used by chillers should increase with temperature, or the electricity demand from pumps is higher when the air pressure increases.

F. Rosso et al [30] conducted a study on different possible retrofit measures and their impacts on energy consumption as well as cost of implementing those measures, where a genetic algorithm was used to determine the best options from different perspectives. The initial investment of the measures, the energy cost, the energy demand and the CO_2 emissions were evaluated for each possible combination of measures, so that the building's stakeholder could choose the best solution taking into account their own priorities.

Energy indicators are a good way to analyse the consumption in a building, and there are some studies that focus on this. As mentioned in the previous section J. C. Wang. [16] determined the energy indicators of different educational buildings in Taiwan and collected information through other research papers about these indicators for different educational buildings from various countries. B. Howard et al [31] estimated the energy indicators in Manhattan for different types of buildings, such as offices, residential, schools and others.

3 Methodology

To create a model that is able to disaggregate the hourly electricity loads into the different end-uses, a genetic algorithm optimization was developed.

First, data was collected and processed to remove any errors. Afterwards, this data was analysed to find correlations between electricity demand and its influencing parameters. Then, the model was developed, followed by the validation of the results. Finally, the model is tested for another building, and then results from both buildings are compared.

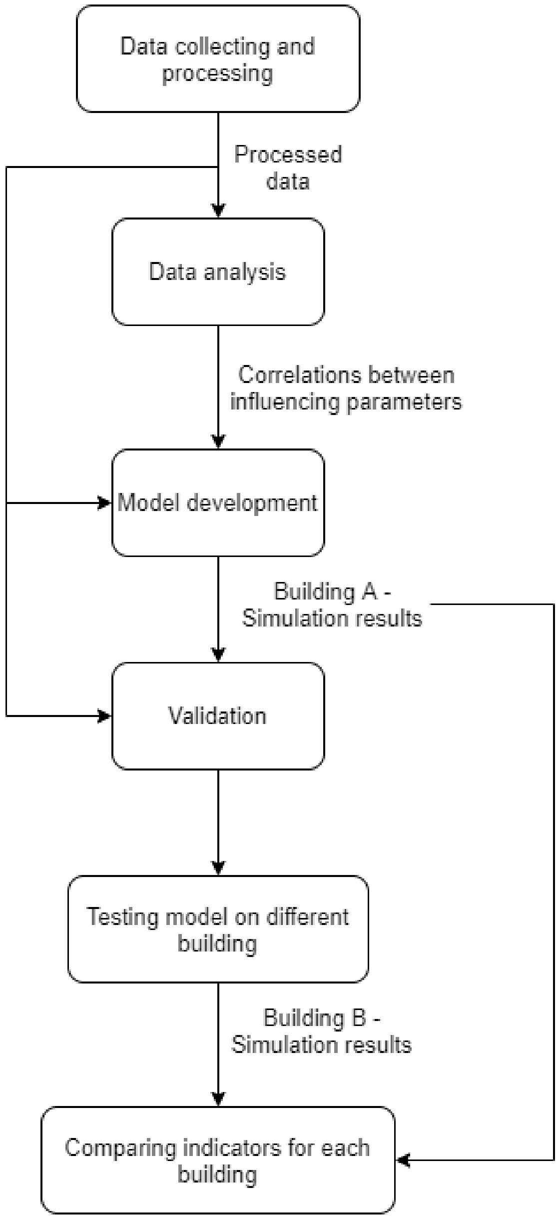

Figure 3 - Simplified view of the methodology

Nowadays, it is increasingly frequent for universities to have a sustainability and/or energy efficiency strategy in practice, which can range from collecting monthly data on energy demand through energy bill's inspection, auditing equipment or running real-time buildings monitoring [32]. Thus, the type of data that can be accessed is very different, ranging between the total electricity consumption of the buildings, the end-use share of that consumption or energy demand profiles be it daily, hourly or sub-hourly.

With this in mind, this work had the objective of developing a model based on frequent available data, so that the model could be widely implemented in other universities.

In this work, the total electricity demand and the end-use shares were used, as well as the hourly electricity consumption profile. Outdoor temperature data was used since it is easily accessible information in open-data resources. Data regarding equipment, such as the number of equipment and their nominal power, was also used, as many universities should have access to this type of information. For classifying energy demand, four main end-uses were defined: *Lighting*, *Electronics*, *HVAC* and *Others*.

Although this work does not use occupancy data, due to the lack of accurate information, there are ways of qualitatively characterizing and taking it into account. In gross terms, occupancy is high during the day, and it is usually low if not zero during the night. Occupancy in the weekends is much lower than during the week, and it changes depending on the time of year. The occupancy changes over time and so does the energy consumption. During summer and winter, when temperatures are harsher, HVAC should increase. In the winter, lighting demand should increase, since there are less hours of daylight.

Despite not having access to more complex occupancy data, taking these different periods of time into consideration as well as their impacts on energy consumption can improve the development of the model, by segmenting it into activity periods. Therefore, each day is divided into day (08:00-19:00) and night-time (00:00-07:00 and 20:00-24:00). The year was separated into different regimes considering the type of activity occurring during that time. So, 8 regimes were created and are shown in Table 1.

Table 1 - Regimes segmentation and characterization

$R1$	Sundays on periods of exams/outside of school period (Periods of time after exams and before the beginning of the next semester, when there are still administrative and research activities, but no classes)
$R2$	Saturdays on periods of exams/outside of school period
$R3$	Reduced activity (Two weeks in august). Most of common equipment, and HVAC are turned off.
$R4$	Sundays during school period
$R5$	Saturdays during school period
$R6$	Weekdays during periods of exams/outside of school period
$R7$	Weekdays during school period
$R8$	Other periods of reduced activity (Mardi Gras, Easter break and last week of the year.)

However, in order to be possible to best tune and validate the model, for the case study of Civil building, more information was available, such as: the hourly profile for the *HVAC*, which was used to compare the results of the model with and without access to this extra information.

More details on the data mentioned in this section will be explained in chapter 4.

3.1 Correlation of electricity demand influencing parameters

To overcome the lack of information regarding occupancy, the relation between outdoor temperature and energy consumption was explored in order to find its correlation, specifically with *HVAC* demand. Outdoor temperature is a commonly used parameter in several studies on energy forecasting, and it is also easily accessible.

To establish the relation between *HVAC* consumption and temperature, the correlation between these two variables was calculated. To prevent finding a weak correlation in average yearly terms and knowing that *HVAC* would also depend on the activity rate or occupancy profiles, the *HVAC* load profile was separated into the 8 different regimes established before, and according to the time of the year, these regimes were attributed to heating or cooling seasons, in order to distinguish the patterns of functioning between the two seasons, that might be different.

For establishing the correlation, only the loads during the day (08:00-19:00) were taken into consideration, since during the night the low consumption is due to the very low occupancy rather than the lower temperatures. So, to prevent false values of correlations, the night-time periods were excluded.

Thus, the average of the *HVAC* loads of each day was calculated, as well as the daily average temperatures for the corresponding day. Then using the Matlab function "corrcoef", the correlation coefficients for each regime were calculated, as shown in equation 1. The function "corrcoef" finds the correlation between two vectors by interpolating the data in a similar way to a linear regression.

$$Corrcoef\left(\overline{HVAC_{R,Season}}, \overline{Temperature_{R,Season}}\right) = Correlation_{R,Season}, \quad (1)$$

$$, R = \{1,2, \dots ,8\}, Season = \{Heating, Cooling\},$$

3.2 Prediction model

In order to predict the hourly end-uses demand profiles, a genetic optimization algorithm was chosen. Genetic algorithm is a heuristic optimization method that is based on the theory of natural selection, where from an initial population, will form other generations by crossing over, mutating and iterating through multiple generations [28].

It is a method where the fittest members of the population survive and carry on their genes to the following generation. The method consists of generating various possible solutions for the problem in hand, being each one of these solutions one member of the total population of solutions. Each member will have simulated values for each one of its variables, which are their genes. With this randomness associated with the generation of values, there will be members of the population that are better solutions for the problem than others. With this in consideration, for this algorithm to follow the process of natural selection, the process of reproduction is implemented. This is where the fittest members of the population, that is, the solutions that are closer to the optimal one, have offspring. These offspring are created by two methods, crossover and mutation:

- Crossover results from when the first members of the population, the parents, create one or more offspring. Some genes from one of the parents will be carried over to the offspring, while the other missing genes will come from the other parent;

- After creating the new offspring, some of them will suffer a process of mutation, where some of their genes will be altered, that is, the value will be changed.

These two processes allow for the creation of a new generation, where some of the offspring will be better solutions than any of the first members of the population. The population will then consist of the parents and offspring. Finally, after this the entire population is evaluated. The fittest members of the population are the ones that survive, so only the better solutions carry on for the next generations, repeating this process again. The number of surviving members is equal to the initial number. So, if there were initially 100 members, and after crossover and mutation, there were 200 members, only the best 100 members of the 200 will be selected to continue the process. This process of reproduction and selection repeats itself throughout many generations, until an optimal solution is found or the stopping criteria is achieved. The algorithm developed in this study will be described in the remainder of section 3.2 and is schematized in Figure 5.

3.2.1 Members of population

For the model developed each member of the population (nPop) will have the simulated values of the hourly energy consumption. Each member will only simulate the hourly profile for one day, instead of simulating the profile for the entire year. There are two main reasons for this decision. The first one is a more practical reason. It would take significantly longer to run the algorithm if it optimized a profile of 8760 hours, which is the number of hours in a year. It is less time consuming to optimize each day individually and iterate throughout every day of the year. The second reason is that it allows for more independent results, that is, any day with atypical consumption will not interfere with the simulation of the other days.

This way, the model will try to find the best solution for each day, delivering the hourly profile for that day. These hourly profiles will be separated into the 4 different end-uses. So, each member will have 4 different hourly values, one for each end-use electricity consumption for each hour and will do this for 24 hours. Therefore, each member of the population will resemble a matrix of dimension 24 by 4. Each line will have the consumption values for one specific hour and each column will have the hourly profile of a specific end-use for that day. The first column will have the values for *lighting*, the second for *electronics*, the third is for *HVAC* and finally the last column is for the *Others* end-use. So, for example the value in the first line and column of the matrix corresponds to the simulated energy consumption in *lighting* from midnight to one in the morning. The matrix is represented in Figure 4.

$$
\begin{pmatrix}
nPop_{h1,Lightin} & nPop_{h1,Electro} & nPop_{h1,HVAC} & nPop_{h1,Others} \\
nPop_{h2,Lightin} & nPop_{h2,Electro} & nPop_{h2,HVAC} & nPop_{h2,Others} \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
nPop_{h24,Lighti} & nPop_{h24,Electr} & nPop_{h24,HVAC} & nPop_{h24,Others}
\end{pmatrix}
$$

Figure 4 - Matrix representation

This way, by summing the 4 columns in the same line the total electricity consumption for that hour is obtained (equation 2), while by summing the 24 values in one column, the consumption in one day for one end-use is calculated (equation 3).

$$\sum_{e=1}^{4} nPop_{h,e} = Total\ Hourly\ Energy_h, \qquad h = \{1,2,\dots,24\} \tag{2}$$

$$\sum_{h=1}^{24} nPop_{h,e} = Daily\ End-use\ Energy_e, \qquad e = \{1,2,3,4\} \tag{3}$$

3.2.2 Boundaries

To assure the model finds the solutions most suitable to reality, it is important to implement information obtained from the analysed data that will approximate the simulations to reality. This is done with constraints and boundaries, that guide the optimization to improve the results, by reducing the final error and to make sure the end-use shares are close to reality.

However, before the constraints it was necessary to determine the boundaries for each member of the population. This way, each simulated hourly value is limited between a minimum and a maximum value. These bounds were estimated for each end-use and for every regime, day and night. These boundaries, therefore, are the minimum and maximum energy consumed by each end-use in each hour. Having constraints might help the model simulate profiles closer to reality, though by adding to many constraints, the model can become very restricted, and have little freedom for optimizing and achieving an optimal solution. So, by adding boundaries that add information to the model will reduce the number of constraints required, as further explained in section 3.2.4.

Other reason for implementing boundaries in the model, is that it allows for the reduction of the final error. Since these boundaries were obtained by analysis of data regarding end-use consumption in the Civil building, these ranges of values will help the initial estimates for each hourly consumption be within a reasonable distance of the real values. This way, the initial errors will be smaller and, therefore, so will the final errors.

To determine the boundaries the equipment data mentioned in the beginning of the chapter and explained in section 4.1.3 was analysed. For each of the regimes in the data, the consumption of every equipment of a specific end-use was summed, to calculate the total consumption of that end-use for the regime. To calculate the average nominal power of the end-use, this total consumption was divided by the average working time of the equipment. This process is demonstrated in equations 4-6.

$$\sum_{N}^{Number\ of\ equipments} Electricity\ consumption_{equipment} = Energy_{Total,R,End-use}\ \ kWh/year \tag{4}$$

$$\frac{\sum Hours}{Number\ of\ equipments} = \overline{Equipment\ Work\ Time}_{\,R,End-use}\ \ h \tag{5}$$

$$\frac{Energy_{Total,R,End-use}}{\overline{Equipment\ Work\ Time}_{\,R,End-use}} = \overline{Power}_{N,R,End-use}\ kW, \tag{6}$$

$$, R = \{1,2,\dots,8\}, \qquad End-use = \{Lighting, Electronics, HVAC, Others\}$$

Since the hourly consumption of an equipment, when it is running at constant power, is equal in value to that power, the bounds values are equal to the average nominal power estimated.

The energy spent while on standby was also a part of the data, so the same process was done with this data. The standby power was then added to the nominal power, as reported by equation 7.

$$\overline{Power_{N,R,End-use}} + \overline{Power_{Standby,R,End-use}} = \overline{Power_{Total,R,End-use}}\ kW, \tag{7}$$

$$, R = \{1,2,\dots,8\}, \qquad End-use = \{Lighting, Electronics, HVAC, Others\}$$

These values were used for the maximum values of the bounds. The minimum values were obtained from the permanent consumption for each end-use. This permanent consumption is the minimum consumption of one end-use, due to standby and residual loads, and it was available in the equipment data for each regime. So, at each hour, the minimum value possible for each end-use is equal to the permanent consumption.

$$MinBound_{R,End-use} = Permanent\ Consumption_{R,End-use}, \tag{8}$$

$$, R = \{1,2,\ldots,8\}, \qquad End-use = \{Lighting, Electronics, HVAC, Others\}$$

As mentioned before in this chapter, the results from the model using the bounds obtained from the equipment data were compared with results using the hourly *HVAC* profile, to see which performed better. To do so, the *HVAC* profile was used to obtain different bounds for *HVAC* than the ones obtained by the equipment data.

The hourly *HVAC* profiles for each of the regimes, separated into day and night, and the two seasons, mentioned in section 3.1 were analysed. For each of these regimes, the boundaries were calculated by averaging the maximum and minimum values of *HVAC* consumption for each day and night, throughout the entire regime. So, the maximum boundary during the day for R1 is the average of each day in this regime's highest consumption, as shown by equation 9.

$$Maximum\ Boundary_{R,Season,HVAC} = \frac{\sum max\ (HVAC_{R,d,h,Season})}{Number\ of\ Days_{R,Season}}, \tag{9}$$

$$, h = \{8,9,\ldots,19\}, \; R = \{1,2,\ldots,8\}, \; d = \{1,2,\ldots, Number\ of\ Days\}, \; Season = \{Heating, Cooling\}$$

This allows to determine more precise bounds than the estimates, since it uses the values from a *HVAC* profile, instead of using the equipment data. It also allows to divide the regimes even further. Unlike the equipment data, the *HVAC* profile allows to implement information regarding temperature. HVAC is used to help reach thermal comfort for the occupants. Therefore, when temperatures are more rigorous, the HVAC demand increases. So, it is expected that during the cooling season, the consumption increases the higher the temperature, and the inverse happens for the heating season.

This way, by separating each week regime (R6 and R7) into two new regimes can introduce this extra information regarding temperature. The average daily temperature ($MeanT$) was determined for each regime and season. Then, each day was separated according to that temperature. If the average temperature ($DailyMeanT$) of a day was higher than the average of the regime, then that day was selected to one of the new regimes, if it was lower, then it went to the other one. This way, the days with harsher temperatures would be in one regime, and the more moderate days would be in another.

$$DailyMeanT_{d,R,Season=} \frac{\sum Temperature_{d,R,Season}}{12\ hours} \tag{10}$$

$$MeanT_{R,Season} = \frac{\sum DailyMeanT_{d,R,Season}}{Number\ of\ Days_{R,Season}}, \tag{11}$$

$$R = \{6,7\}, \quad Season = \{Heating, Cooling\}, \quad d = \{1,2,\ldots, Number\ of\ Days\}$$

3.2.3 Constraints

To implement further information into the model, constraints were used.

Different constraints were used in order to add different details and information regarding the 4 end-uses. These constraints were implemented in the form of inequalities, as explained as follows.

3.2.3.1 Constraints regarding activity patterns

The ratio between the minimum and maximum consumption of *HVAC* was calculated for both day and night and for each regime. The boundaries obtained from the *HVAC* profile were used for this process. For each regime, the minimum boundary during the day was divided by the maximum during that day to obtain the min/max ratio (Equation 12). This was also done for the night-time and for both seasons.

$$\frac{MinBound_{R,Daytime,Season}}{MaxBound_{R,Daytime,Season}} = HVACRatio_{R,Daytime,Season}, \tag{12}$$

$$, R = \{1,2,\dots,8\}, \quad Daytime = \{Day, Night\}, \quad Season = \{Cooling, Heating\}$$

In the simulation, the model imposes that the ratio of the simulated values must be equal or higher than the previously calculated ones ($HVACRatio_{Real}$), as shown in equation 13.

$$\left(\frac{HVAC_{min}}{HVAC_{max}}\right)_{Sim_{R,Daytime,Season}} \geq HVACRatio_{Real_{R,Daytime,Season}}, \tag{13}$$

$$, R = \{1,2,\dots,8\}, \ Daytime = \{Day, Night\}, \ Season = \{Cooling, Heating\}$$

Despite not knowing how the consumption changes from hour to hour, it is known that the consumption during the day is higher than during the night. So, the first hour of the daytime must consume more electricity than the last hour of the night, and the same applies for the last hour of the daytime and the first hour of the night. Thus, the model checks if this is complied.

$$Total\ Consumption_{Sim}(08{:}00) > Total\ Consumption_{Sim}(07{:}00) \tag{14}$$

$$Total\ Consumption_{Sim}(19{:}00) > Total\ Consumption_{Sim}(20{:}00) \tag{15}$$

Finally, to assure the end-use shares will be close to reality, some data regarding end-use shares should be implemented. However, since the algorithm runs each day at a time, it is impossible to check if the end-use shares in the end are correct. So, to add this information to the model, the end-use shares will be evaluated only for the weekdays during school and exam periods (R6 and R7). Since the weekdays are the more numerous days of the year, they will have the highest contribution to the total energy consumption in the year. Therefore, they will have the closest patterns to the yearly average consumption. So, it is assumed that these regimes will have relatively similar end-use shares to the yearly ones.

To allow for some variation, instead of checking if the end-use shares are the same as the real ones obtained from the equipment data, the model will check if the simulated shares are in a range of values close to the real ones, as shown in equation 16. To calculate the ranges, the real shares were deviated with a value δ ranging from around 5 to 10%. When the end-use share was smaller, δ was also smaller (closer to 5%) so the end-use share range did not deviate too far from the real value. On the other hand, when the end-use share was larger so was δ.

$$Energy\ Share_{Sim,End-use} \in \left[Energy\ Share_{Real,End-use} - \delta; Energy\ Share_{Real,End-use} + \delta\right], \qquad (16)$$

$$End - use = \{Lighting, Electronics, HVAC, Others\}$$

3.2.3.2 Inter end-use constraints

Although during the night the occupancy is usually zero, and the consumption is low, there are situations where this not the case, i.e., when there are people at night in the building, either students in study spaces, or cleaning staff, the lights will be turned on and so some electronic equipment or HVAC. Thus, this relation between usage of end-uses was also implemented in the model. When the generated loads for *lighting* are above a certain value (150% of the lower bound of the corresponding end-use), the model checks if the consumption for *electronics and HVAC* are also above a certain value. It was considered that consumption below 150% of the lower bound was consumption due to standby and other residual loads. So, this separates the periods of time when the only consumption is on standby, so when occupancy is zero, and when there is more consumption due to usage of equipment when there are occupants in the building. This will assure that when lights are turned on, other equipment will also be turned on.

$$\overline{Light_N} > 1.5 * Light_{N_{Min}} \wedge \overline{Eletr_N} > 1.5 * Eletr_{N_{Min}} \wedge \overline{HVAC_N} > 1.5 * HVAC_{N_{Min}}, \qquad (17)$$

$$, N = Night,\ Min = Minimum$$

3.2.3.3 Constraints regarding HVAC consumption and temperature

Finally, constraints that relate *HVAC* consumption with the outdoor temperature were also added. This was also implemented since outdoor temperature data is easily accessible for practically any city in the world and for various periods of time. The correlations obtained in section 3.1 were used for this purpose.

The results presented in section 4.2.1 show more correlation between *HVAC* consumption and temperature during regimes of normal activity (R6 and R7), in comparison to other regimes. It also shows that on those two regimes the *HVAC* consumption increases with temperature in the cooling season but decreases during the heating season.

So, to implement this information, the model needs to check if these relations are observed in the simulation. The algorithm checks if the correlation between the generated *HVAC* loads during the day and the average temperature ($TAverage$) for that day are higher than a base value ($BaseCorr$) during regimes of activity. For the other regimes, it checks if the correlation is lower than the base value for these regimes. The algorithm also checks if the correlation is negative during heating season and positive during cooling season, for the regimes of activity.

$$\left|Corr(HVAC, TAverage)_{Sim,D}\right| > BaseCorr_R, \qquad R = 6,7 \tag{18}$$

$$\left|Corr(HVAC, TAverage)_{Sim,D}\right| < BaseCorr_R, \qquad R = 1,2,3,4,5,8 \tag{19}$$

$$Corr(HVAC, TAverage)_{Sim,D} > 0 \quad R = 6,7, \qquad Season = Cooling \tag{20}$$

$$Corr(HVAC, TAverage)_{Sim,D} < 0 \quad R = 6,7, \qquad Season = Heating \tag{21}$$

$$D = Day, \qquad Sim = Simulation$$

The value of $BaseCorr$ was determined by analysing the results and it will be explained in section 4.2.1.

3.2.4 Evaluation and creating offspring

To determine the best members (nPop) for each iteration, first it was necessary to define what parameters of evaluation should be used. Two parameters were used:

- the error (CCost);
- the count of constraints not complied (CCount).

The error is the mean absolute error for the entire day, comparing the total consumption ($Total\ Hourly\ Energy$) for each hour (summing the end-uses) with the real value for that hour ($Real\ Hourly\ Energy$), for the corresponding building. The absolute value of this difference is the hourly error, and CCost is the average of these hourly errors, thus the lower the error the better the solution is.

$$|Total\ Hourly\ Energy_h - Real\ Hourly\ Energy_h| = Hourly\ Error_h, \qquad h = \{1,2,\dots,24\} \tag{22}$$

$$CCost = \frac{\sum Hourly\ Error_h}{24} \tag{23}$$

The other parameter used is the CCount. This parameter counts the number of constraints mentioned in the previous section, that the member of population does not comply with. As stated before, the constraints were used to assure the results would be as close as possible to reality, and that the model would be able to accurately forecast the end-uses electricity demand.

Since the model simulates 4 different end-uses profiles, if it only sorted the members by the error, this could lead to situations where the error could be close to 0, however failing to deliver end-use load disaggregation that would make sense. For example, there could be a situation where the simulation predicted the total load, but the only load occurring would be *lighting*.

As such, complying with constraints is the priority of the model, followed by achieving the minimum error. Therefore, one nPop that has an error of 20 but complies with all constraints is a better solution than one nPop with an error of 0 but that does not comply with all constraints.

However, since the model prioritizes the CCount of each population member, when there are many constraints, this leads to a loss of importance of the error of each nPop. For example, if there are 20 constraints, on an initial guess, the best member of the population may comply with four or five, but it still does not comply with 15 constraints, so its CCount is 15. Each iteration will try to find a member that has a lower CCount than before, at the cost of a higher error. This cycle repeats itself until either the best solution complies with all constraints or it could still not comply with all of them. When the number of constraints is low, this does not cause problems. However, when the solution has to verify many constraints, the cycle of finding the lowest CCount but increasing the error occurs for many iterations. So, even if the CCount reaches zero at some point, this will be at a high number of iterations, so there will not be many more iterations to reduce the error. This way, the performance is severely lowered with the increase of constraints in the model.

3.2.4.1 Selection

The selection method used to choose the parents that would be used for the crossover was the roulette wheel selection. This method gives different probabilities depending on the fitness of each member of the population. The fitter the member is, the higher the chance of being chosen as a parent. For this selection the fitness parameter used was the CCost. This way, the lower the CCost of a member, the more likely it is for that member to become a parent.

Although the CCount is a priority, it would not work very well as a fitness parameter as it would not differentiate a lot the different members. Since the CCount is equal to the number of constraints not fulfilled, the parameter would only have a few values to compare with, and if most members had a CCount equal to zero, there would not be any differentiation between them.

3.2.4.2 Crossover

The uniform crossover was used in this algorithm. This method creates offspring by randomly choosing each gene from one of the parents. The number of offspring will be the same as the number of original members.

3.2.4.3 Mutation

Mutation changes some genes from some of the newly created members, by replacing those genes with a different value. This value is obtained from a normal distribution with a 2% probability.

3.2.5 Sorting and survival of the best

After the reproduction, the parents and the offspring are joined, duplicating the population size, and are then sorted from best to worst. Each member was sorted first by the lowest values of CCount and then by CCost. Then, for each of these groups with equal CCount, each member was sorted according to their CCost. After this sorting, the worst members are eliminated and the population size reverts to the original one. At the end of each iteration the mean relative error (MRE) for the best member of the population is calculated.

$$MRE_h = \frac{|Total\ Hourly\ Energy_h - Real\ Hourly\ Energy_h|}{Real\ Hourly\ Energy_h}, \tag{24}$$

$$MRE = \frac{\sum MRE_h}{24}, \tag{25}$$

$$h = \{1,2,\dots,24\}$$

This entire process, from selection to the sorting of the best, occurs until the number of iterations, previously decided, is achieved. This number was decided based on testing simulations to see when the error started to stabilize. Although the error could be improved with more iterations, the improvement would not be very significant, and it would come at the cost of a longer running time.

However, if the MRE of an iteration was lower than 5%, the cycle stops (stopping criteria) at that iteration and the best member of the population becomes the solution for that day. This was implemented to reduce the time needed to simulate the entire year while keeping a reasonable error. ASHRAE guideline 14 [33] suggests two measures of error to assess the accuracy of simulations relative to the real data. These are the normalized mean bias error (NMBE) and the coefficient of variation of the root mean square error (CVRMSE). The guideline states that, for hourly predictions, the limit of NMBE is 10% and for CVRMSE is 30%. So, a MRE of 5% was used as an acceptable value.

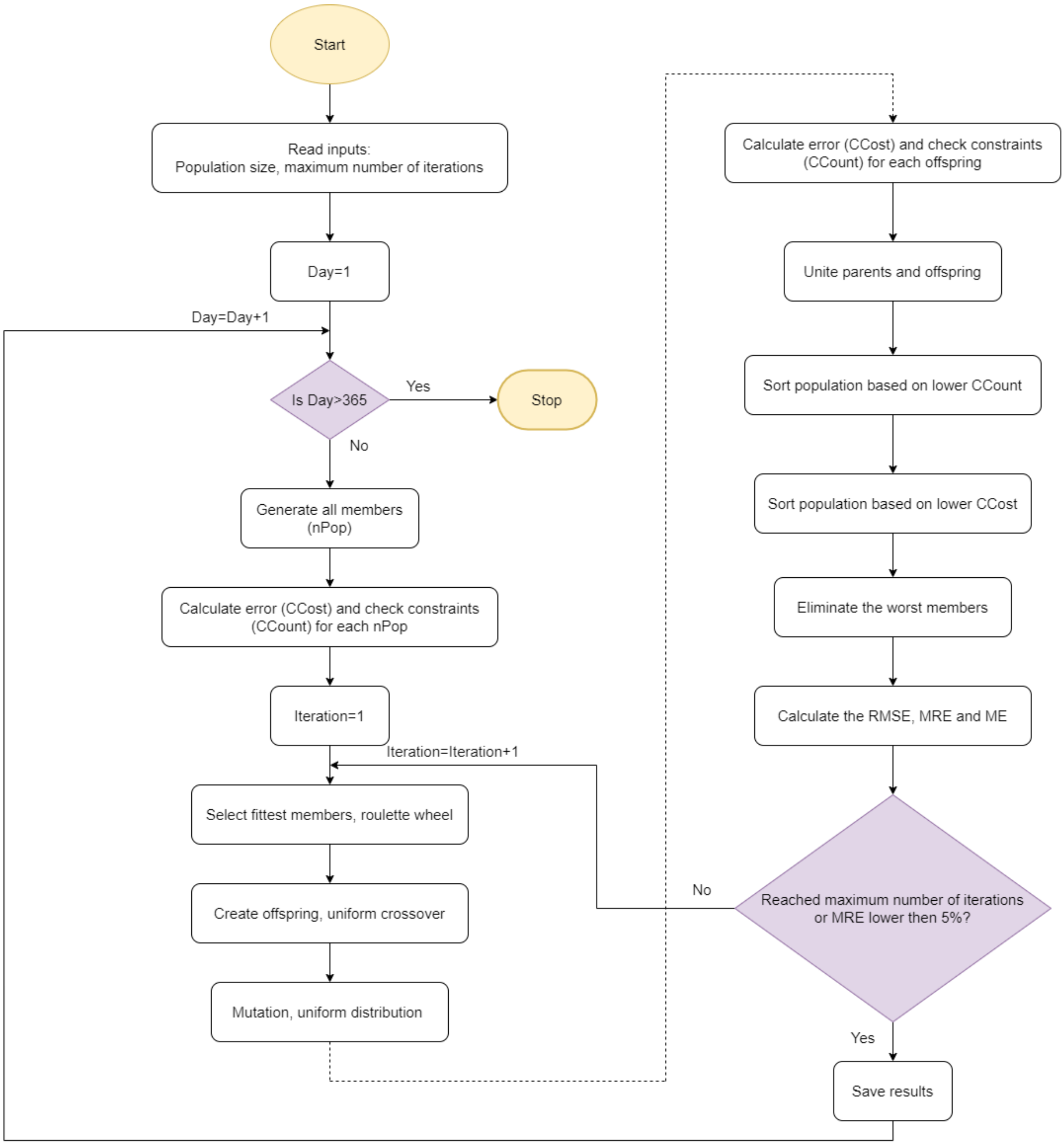

Figure 5 - Genetic Algorithm representation

4 Case-study description

The Alameda campus of the Instituto Superior Técnico was chosen to be the case study due to the accessibility of data and the knowledge of the buildings and activities of the campus. The campus has 24 buildings and is located in Lisbon, Portugal, which has a moderate climate.

Most data available regards five buildings of campus. Despite being around only a fifth of all buildings, these 5 buildings account for 57% of the total electricity consumption of the campus. This way, each one of these buildings is a representative case study for analysis.

The Civil building was chosen for developing and validating the model, since there is more information available. Once the model is validated it is ready to be tested in other buildings. Since there was the same amount of data available for the other four buildings, the South Tower was chosen for testing due to the better prior knowledge of this building's layout.

4.1 Data collection and processing

The first step was to characterize buildings electricity demand, by collecting and processing different types of data regarding the buildings to be studied. This information was collected from various sources, with different usage and relevance in the development of the model, such as readings from smart-meters, audits and/or simulations from different models. These details are further explained in the following subsections. All the data is related to buildings on the Alameda campus of IST.

4.1.1 Total electricity hourly loads

The total hourly loads, measured by smart meters, were provided by the "Campus Sustentável" project [32], for the year 2019, for five buildings of Alameda Campus: *North Tower*, *South Tower, Central Building, Civil Building* and *Mathematics Building*. As they were provided in Ah (Ampere hour), equation 26 was used to convert these loads into kWh.

$$Electric\ consumption\ (kWh) = \frac{Electric\ load\ (Ah) * Voltage\ (V) * Power\ Factor\ (PF_{Building})}{1000}, \tag{26}$$

$$Building = \{Central, Civil, Mathematics, North\ Tower, South\ Tower\}, \qquad Voltage = 230V$$

After the conversion, some data points were missing due to the meter malfunctioning during certain hours. Therefore, it was necessary to interpolate these missing values, which was done in two ways. When there was an occasional missing point, the average between the values from the hour prior and after was calculated and replaced the missing value. However, this method could not be employed for every missing value since there were situations where they span some hours or even days in a row. So, in those situations, the missing values were replaced by the value of the same day and hour of the week prior or after. This assures, or at least makes it more likely, that the missing values will be replaced by a value close to what it was in reality, given weekdays and weekends activity.

Regarding other possible outliers in the data, there were some peak values, which were reoccurring and seemed to follow a pattern of consumption throughout the year. During the weekdays, those peaks were higher in winter and summer, when the climate is more extreme, so they seemed to indicate some kind of seasonality. Therefore, it was assumed these values were not outliers and were kept in the dataset.

4.1.2 End-use hourly monitoring

In order to be able to validate the end-use prediction model and further understand end-uses patterns, hourly end-use data was needed. However, real monitored data was only available for *HVAC* for the Civil Building for the 2017 and 2018, though, simulated hourly profiles previously developed in Energy Plus software [34] of other end-uses, such as, lighting and equipment were also provided. However, real *HVAC* profiles report to 2017 and 2018, while total electricity data is from 2019. So, it was necessary to extrapolate *HVAC* 2017 profile data to match it to the consumption profile for 2019.

This was done by first dividing the hourly *HVAC* consumption in 2017 by the total hourly consumption in the same year, to find the hourly *HVAC* share at each hour. This way the pattern of HVAC consumption could be extrapolated for the year of 2019, by multiplying each percentage by the total load in 2019. Assuming that the *HVAC* behaviour did not change significantly from 2017 to 2019.

This *HVAC* profile was divided into the 8 regimes to match with the data from 2019.

4.1.3 Equipment data per end-use and area

Energy consumption data was collected and characterized by the "Campus Sustentável" team, through an audit, for the 24 buildings on Alameda campus, by end-use and type of space, as systematized in Table 2. The energy consumption data detailed the total electricity consumption in 2019, separated by each end-use and type of space, for every building. The area that each room/space takes in each building was also made available.

There were 8 different end-uses, however, for the purpose of this thesis, the end-uses were grouped into 4 major ones according to their similarities, as also shown in Table 2.

Table 2 - End-Uses and corresponding equipment

End-Use	***Equipment***	***Final end-uses***
Lighting	Light bulbs	Lighting
Computer Equipment	CPUs, printers, servers	Electronics
Plug In Equipment	Washing machines, TV screens, projectors	
Catering	Ovens, fridges, vending machines	
Local HVAC	Local split unit, fans	HVAC
Central HVAC	Chillers, boilers, VRV units	
Common systems	Elevators, drain pumps	Others
Investigation Specific Equipment (ISE)	Equipment for lab/workshop use	

Equipment data was also collected for the building of Civil only. It detailed type of equipment, number, nominal power, and average functioning time of each equipment. The total yearly consumption, standby consumption, and permanent consumption (minimum value of consumption) for each end-use was available. This data was obtained through an audit [35], where all electricity consuming equipment in the building was registered. This data was also divided into different regimes, in a similar way to the division mentioned in the previous section. However, these regimes were slightly different and were divided as shown in Table 3.

Table 3 - Regimes in audit

AR1	Weekdays
AR2	Holidays
AR3	Saturdays
AR4	Sundays
AR5	Period of reduced activity

In order to merge this data into the previously mentioned regimes, it was necessary to aggregate this data, as shown in Table 4.

Table 4 - Grouping of regimes

HVAC profile regimes	R1, R4	R2, R5	R6, R7	R3	R8
Audit regimes	AR4	AR3	AR1	AR5	AR2

All this data was separated into the different end-uses mentioned before, and for each one, the permanent energy for each regime was also available, that is, the minimum energy consumption for each end-use on a specific regime.

With the reported data it is possible to estimate the end-uses' electricity demand and calculate its share, regarding the total energy consumption in a building. Further, two other analyses can be done:

- obtain the end-use share for each type of space, allowing to understand how energy is spent in each different type of room;
- obtain the electricity consumption share per type of room, regarding the total building electricity consumption. This would help understanding which rooms/spaces are most important or will have a greater impact when forecasting energy consumption in university buildings.

4.1.4 Weather data

The weather data in the city of Lisbon was obtained from Renewable Ninja [36], namely hourly profiles for outdoor temperature, solar radiation and wind speed, for the year of 2019.

4.2 Civil Building

The Civil building has a gross floor area of 25183 m^2 and eight floors, three of them being underground. Floor -1 has amphitheatres for classes and some reunion rooms, while floor -2 also has meeting rooms and some small labs. The library takes space on both floors. Floor -3 has a few storage and technical rooms and a few reunion rooms. The ground floor has a bar, a cafeteria and a few classrooms/study spaces. The first floor has many classrooms while the second floor has the teachers' offices. There are also some offices on the third floor alongside some small workshops. The top floor has just a few archive rooms and technical rooms.

Figure 6 - Side view of the Civil building

Figure 7 shows the area allocation in the building of Civil for each important space. These spaces are considered important, since these are the spaces that consume the most electricity in the entire building.

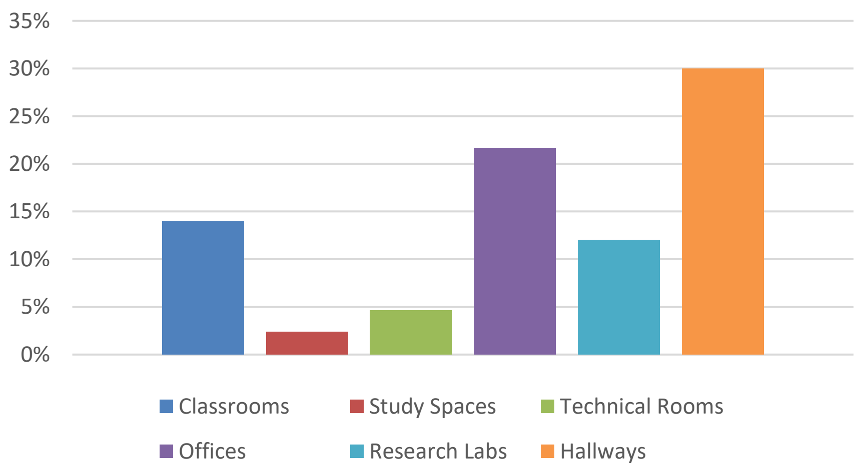

Figure 7 - Area allocation – Civil Building

Figure 8 reports the electricity demand per type of space. As it can be seen, the space that consumes the most electricity, by far, are the technical rooms, followed by offices and hallways. Classrooms, study spaces and research labs also consume a somewhat significant amount of energy, while all the other spaces only contribute to around 8% of the total consumption in the building, so these spaces are not worthwhile analysing.

To better understand how each room affects the electricity consumption in the building, the end-use shares of each type of space were calculated. This way it is possible to know how important a specific end-use is to each of these rooms. The results are shown in Table 5, and in red is highlighted the end-use with largest demand per type of space.

Table 5 - End-use share per type of space - Civil

	Lighting (%)	*Electronics (%)*	*HVAC (%)*	*Others (%)*
Classrooms	56,19	21,18	22,63	0
Study Spaces	34,87	58,94	6,2	0
Technical Rooms	0,07	0,02	81,3	18,6
Offices	21,62	44,71	33,68	0
Research Labs	73,37	23,11	3,51	0
Hallways	94,08	3,35	2,58	0

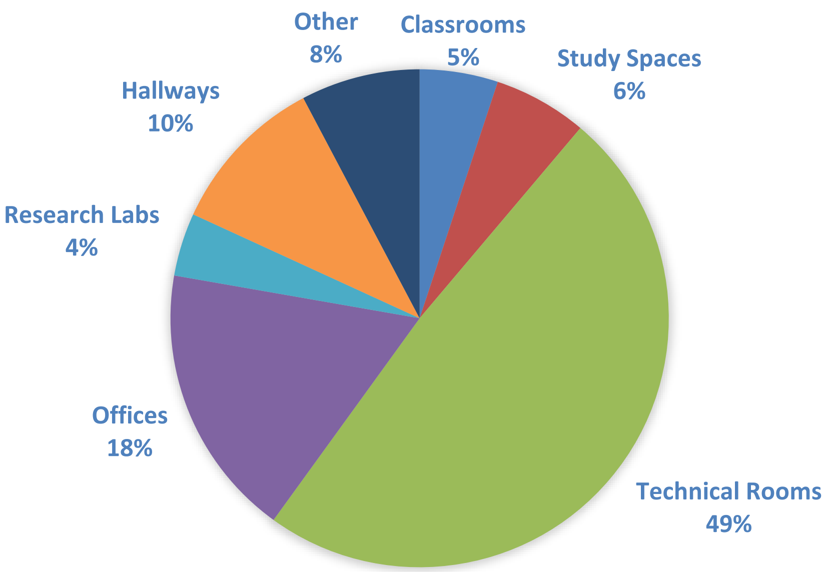

Figure 8 - Electricity consumption share per space - Civil

As it can be seen, hallways predominantly spend energy on *lighting*, while technical rooms mostly consume *HVAC*. This makes sense considering the type of spaces they are. Hallways do not have many equipment, so most consumption is related to *lighting*, while technical rooms are divisions that have equipment related to *HVAC*, like pumps, fans or other parts of *HVAC* equipment.

Comparing classrooms and study spaces there are some noticeable differences. Despite being used for similar purposes, most classrooms do not have many electronic equipment apart from a projector and some wall plugs, which are not used very often. So, most consumption regards *lighting* and *HVAC*. However, in study spaces, students can spend several hours in the same room, so they charge their laptops or phones. Also, some study spaces have desktop computers available for students to use. The same logic applies to offices, since every office has a computer per occupant, and many offices are used by more than one teacher.

Research labs stand out a bit in this analysis. The high *lighting* share is to be expected, but not that high. This value should be more balanced with higher consumption of *HVAC* and *others*, due to some lab equipment. However, the lab equipment consumption might be accounted for in the technical room.

The data present in section 4.1.3 also allowed for calculating the end-use shares (Figure 9) in Civil, which were used to determine the range of values to be used in the end-use share constraints (mentioned in section 3.2.3), shown in Table 6.

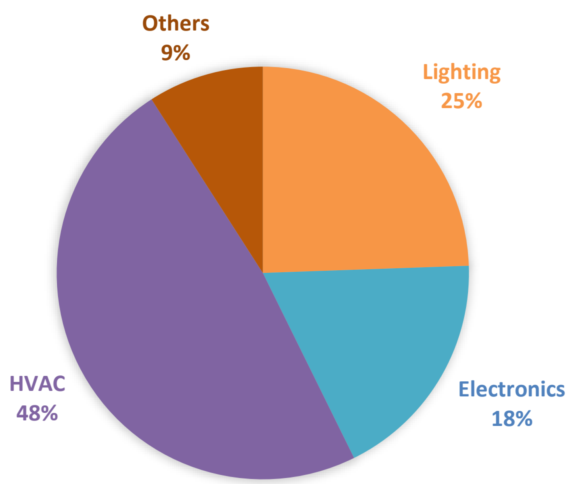

Figure 9 - End-use share - Civil

Table 6 - Model end-use share range - Civil

	Lighting (%)	*Electronics (%)*	*HVAC (%)*	*Others (%)*
Real share	25	18	48	9
Range of acceptable values	20-30	14-22	40-55	3-15

4.2.1 Correlation results

As mentioned in section 3.1, the correlation between *HVAC* consumption and outdoor temperature was calculated, as is reported in Table 7.

Table 7 - Correlations between HVAC and outdoor temperature per regime

	R1	*R2*	*R3*	*R4*	*R5*	*R6*	*R7*	*R8*
Heating	-0.0127	-0.1604	-	-0.1109	-0.1973	-0.5176	-0.2691	0.8779
Cooling	-0.3137	0.0485	-0.1699	0.4049	0.1045	0.2268	0.4656	0.2469

Checking Table 7, the following conclusions can be drawn:

- The correlations are, usually, higher, in absolute value, during the regimes of activity (R6 and R7) (20-50%) than during weekends and other regimes of reduced activity, apart from R8 (0-30%);
- There is a positive relation between *HVAC* consumption and temperature during cooling season, and a negative one in heating season. This relation is evident during the week (R6 and R7) but not so much on other regimes;

The first point is explained by the reduced occupancy during weekends, which significantly reduces the importance of thermal comfort, and thus on *HVAC* functioning. Therefore, during these regimes, the *HVAC* consumption does not have a high dependence on the temperature.

The second point shows that the *HVAC* consumption increases with the rise of temperature in cooling seasons and decreases with the increase of temperature during heating seasons. Although this is observable during the week, this linearity does not appear during regimes of reduced activity. This is due to the same reasons as explained for the previous point.

With these results in mind, the model was updated so that for each day, the algorithm checks if the correlation between the generated *HVAC* loads during the day and the average temperature for that day, is high during regimes of activity, and low for the other ones, like it was mentioned in section

3.2.3. Therefore, the *BaseCorr* for normal weekdays (R6 and R7) is 20% while for the other regimes it is 30%.

The results for the regime of holidays (R8) were not taken in consideration, since they are very different from what was expected from regimes of reduced activity. These results can be explained by the low number of days for that regime, which resulted in unreliable results. So, for the model, this regime used the same *BaseCorr* as the other regimes of reduced activity.

4.2.2 Boundaries calculation

The boundaries obtained by using equations 4-6, mentioned in section 3.2.2 are shown in Figure 10-Figure 14.

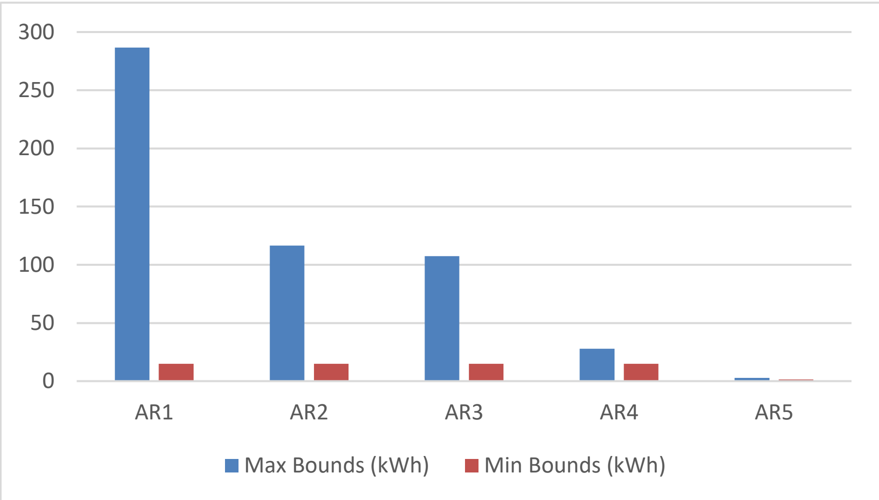

Figure 10 - Lighting bounds from audit

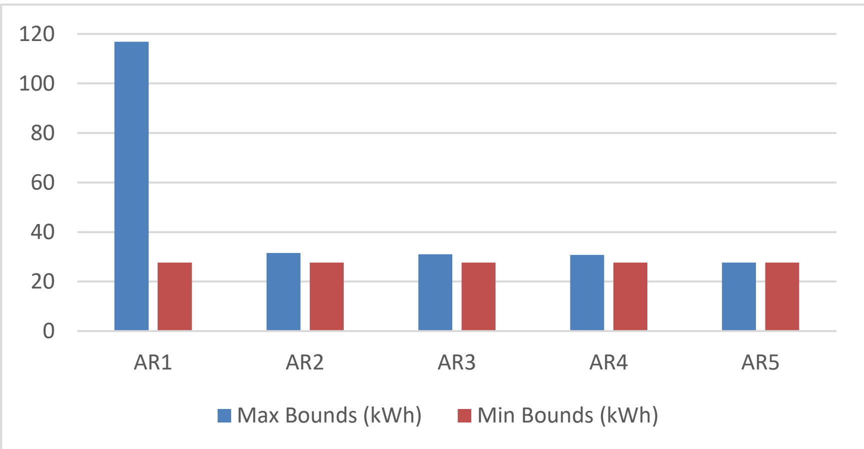

Figure 11 - Electronics bounds from audit

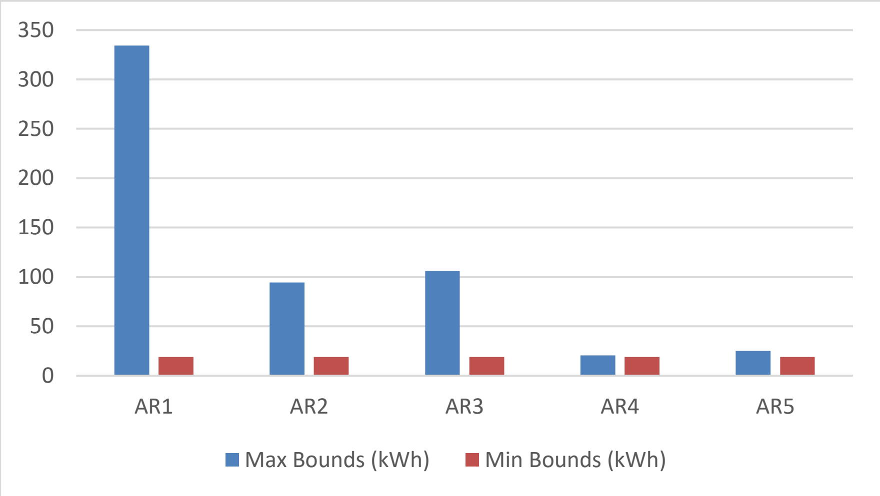

Figure 12 - HVAC bounds (cooling season) from audit

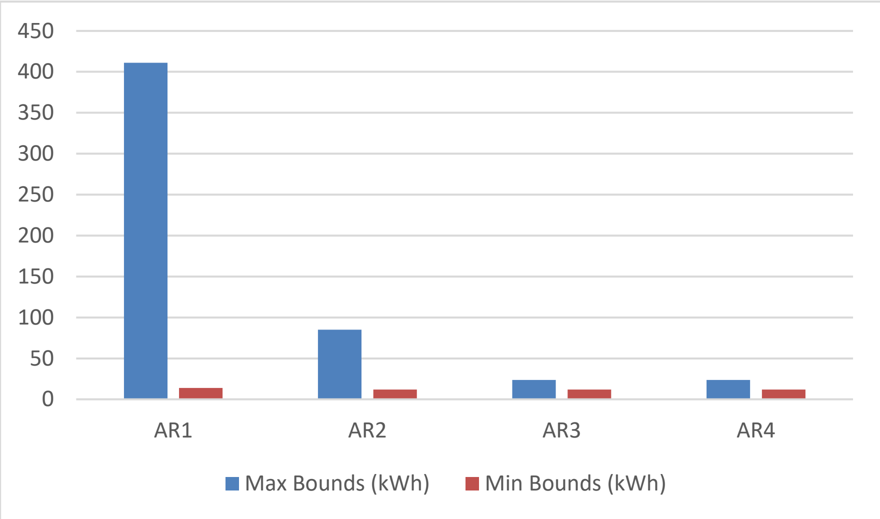

Figure 13 - HVAC bounds (heating season) from audit

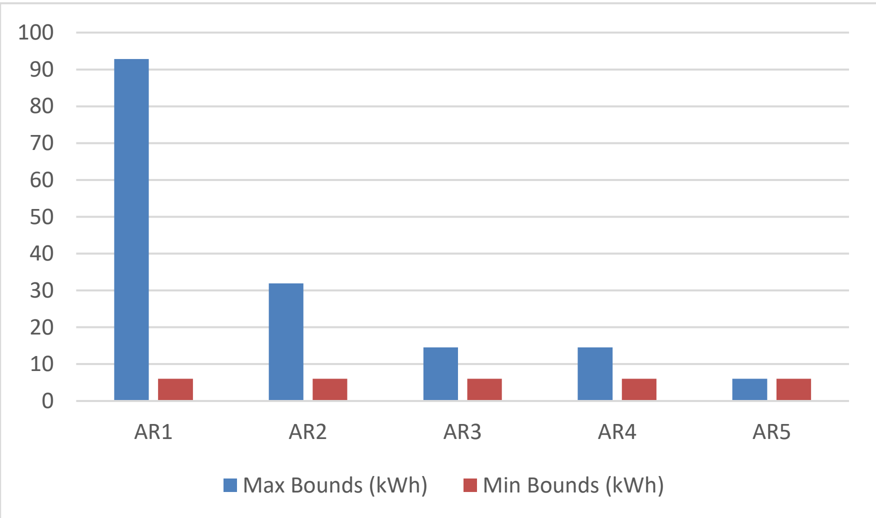

Figure 14 - Others bounds from audit

When implementing these boundaries in the model, the max bounds were increased slightly, to reduce the errors in some days where a peak of consumption could occur. So, each bound was rounded up to the next multiple of 5. For example, if the max bound is 22.7 the bound implemented was 25, if it was 37.3 it was rounded up to 40.

Another aspect to consider is that the max bounds for *HVAC* during weekdays were too high, so they were limited to a maximum of 300 kWh in the model, as seen in Figure 15, instead of the values in Figure 12 and Figure 13. Since the regime of vacations in august (AR5) only requires cooling, there are no bounds for heating during that regime.

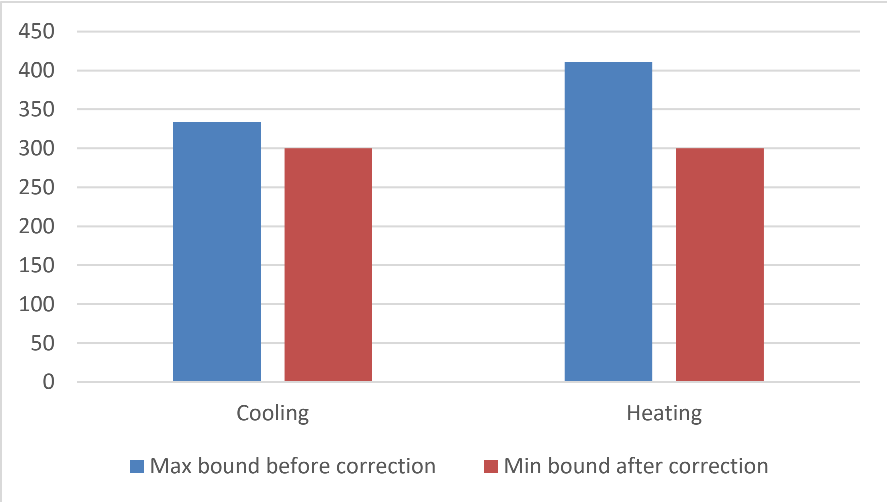

Figure 15 - HVAC bounds before and after correction

4.2.3 HVAC ratios

As mentioned in section 3.2.3 the ratio between minimum and maximum hourly *HVAC* consumption was calculated to add more information about the pattern of *HVAC* consumption to the model. The ratios for each regime, separated by season and, by day and night, are presented in Figure 16 and Figure 17.

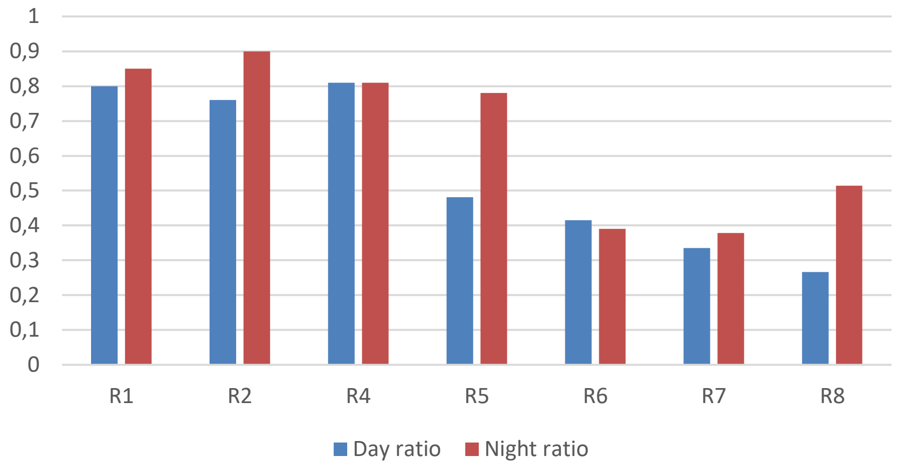

Figure 16 - HVAC ratios – Heating season

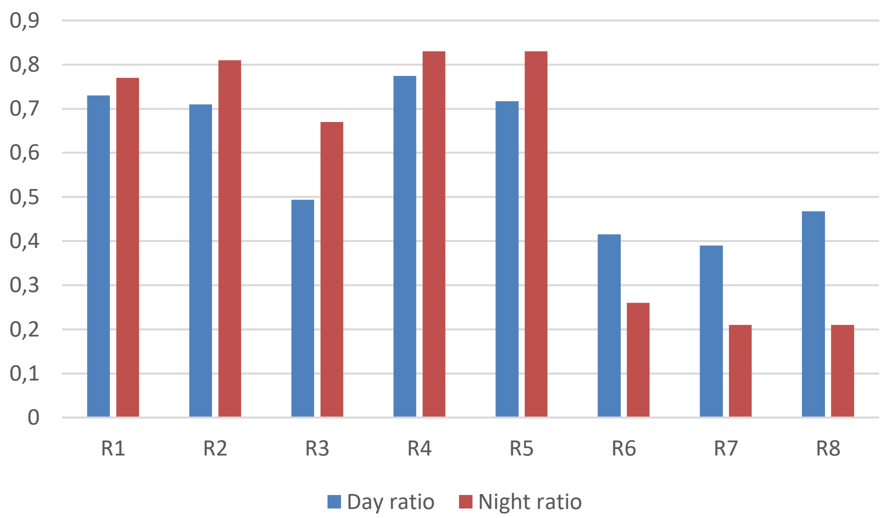

Figure 17 - HVAC ratios – Cooling season

From the results, it can be seen that, generally, the ratio is higher (more than 70%), during the weekends, in comparison with the ratios during weekdays (around 40% or less). This difference can be explained by the reduced usage of *HVAC* equipment during the weekends, so the consumption does not fluctuate significantly, most of the consumption is related to equipment on standby and residual loads form pumps and fans. On the other hand, during the week, there are different levels of occupancy, therefore during the day, this consumption can vary significantly. It also explains why the ratio is higher during the night rather than during the day, again due to the reduced activity during the night. The ratios during the night in R6 and R7 are similar or even lower than during the day, which might be due to some transition hours, where systems are being turned on so the maximum HVAC consumption is much higher than the minimum.

4.3 South Tower

The South Tower is a building with a gross floor area of 11104 m^2, it has sixteen floors, and three of them are underground. These floors have a mix of various types of rooms, which area allocation is shown in Figure 19.

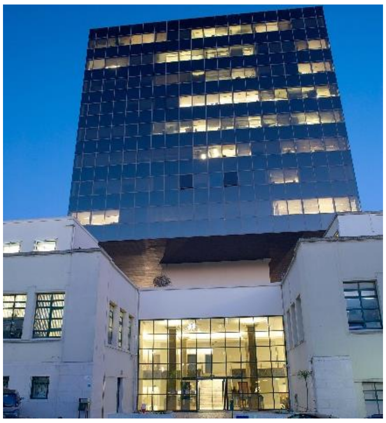

Figure 18 - Frontal view of the South Tower

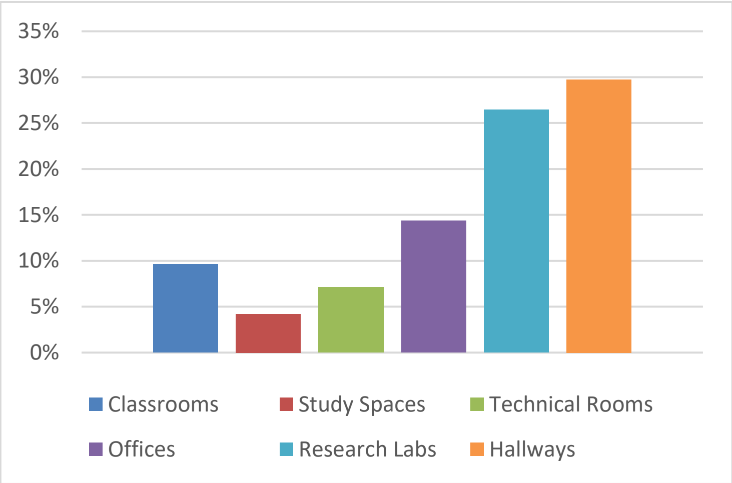

Figure 19 - Area allocation - South Tower

Just as for Civil Building, more than 90% of the total electricity consumption in the building is related to the six types of spaces stated before, as it can be seen in Figure 20.

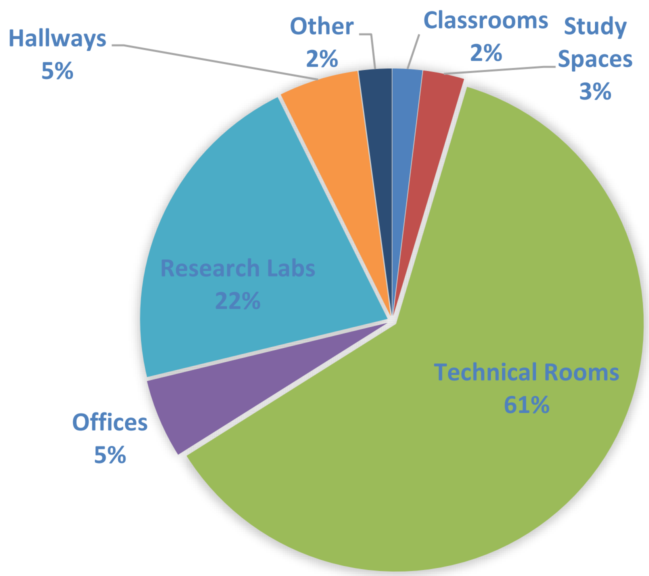

Figure 20 - Electricity consumption share per space - South Tower

To verify if the end-use share per type of space seen for Civil is similar to the South Tower, the same analysis was made. Table 8 shows that the end-use shares are consistent throughout the buildings. The biggest difference is the bigger consumption in research labs regarding *HVAC* and *Others*. This allows to conclude that the pattern of consumption in each of these spaces is similar between the two buildings.

Table 8 - End-use share per type of space – South Tower

	Lighting	*Electronics*	*HVAC*	*Others*
Classrooms	81.14	4.78	14.08	0
Study Spaces	43.14	55.7	1.16	0
Technical Rooms	0.07	0.31	96.64	2.98
Offices	26.04	65.01	8.95	0
Research Labs	40.7	14.96	23.99	20.36
Hallways	70.42	1.89	22.29	5.41

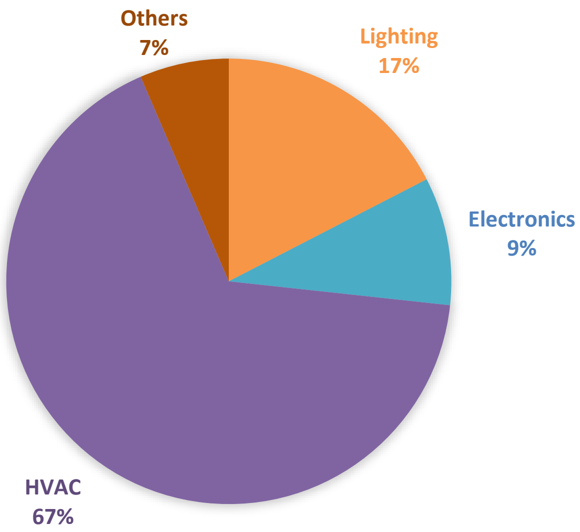

Figure 21 - End-use share - South Tower

Just as for Civil Building, to run the model the end-use shares for the South Tower are needed. These shares are shown in Figure 21, and so the range of shares for the South Tower simulation are presented in Table 9.

Table 9 - Model end-use share range - South Tower

	Lighting (%)	*Electronics (%)*	*HVAC (%)*	*Others (%)*
Real share	17	9	67	7
Range of acceptable values	12-22	5-15	60-70	1-12

4.3.1 Boundaries calculation

For the purpose of testing the model for the South Tower, new boundaries were calculated. While for the Civil Building, boundaries data was obtained from an audit, similar data was not available for the South Tower. Hence, to calculate the boundaries, some assumptions were made. It was assumed that the pattern of consumption of the different regimes would be similar for every end-use, and that the average time of operation of equipment was also the same. Since the use of equipment is highly related to occupancy, these assumptions are reasonable due to the similarity of occupancy and activity patterns, like types of activities and classes schedules, between buildings.

This way, the same process used for calculating the Civil building boundaries could be used, with a simple modification, since only the total consumption for each end-use was available, and it was not separated by different regimes. So, by finding the share of energy consumption of each regime for Civil building and then applying the result to the total consumption for the South Tower, boundaries for every regime were obtained, as exemplified by equations 27-30.

$$\frac{Energy_{Total,R,End-use,Civil}}{Energy_{Total,End-use,Civil}} = Regime\ Share_{R,End-use} \tag{27}$$

$$Regime\ Share_{R,End-use} * Total\ Consumption_{End-use,South\ Tower} = Energy_{Total,R,End-use,South\ Tower}\ \ kWh/year \tag{28}$$

$$\frac{\sum Hours}{Number\ of\ equipments} = \overline{Equipment\ Work\ Time}_{\ R,End-use}\ \ h \tag{29}$$

$$\frac{Energy_{Total,R,End-use,South\ Tower}}{\overline{Equipment\ Work\ Time}} = \overline{Power}_{R,End-use}\ , \tag{30}$$

$$, R = \{1,2,\ldots,8\}, \qquad End-use = \{Lighting, Electronics, HVAC, Others\}$$

5 Results

5.1 Model validation

To validate the model, the civil Building was used, and two simulations were made. The first one used the boundaries derived from the real *HVAC* profile. The second one used the *HVAC* boundaries obtained from the audit file. This was done to compare the errors between the two simulations, to understand how important it is to have access to real-time data, and how viable it is to apply this model to other buildings where normally there is less available data. Table 10 shows the input parameters for both simulations.

Table 10 - Input Parameters for simulations 1 and 2

	Members of population	*Number of iterations per day*
Simulation 1 – Bounds from real HVAC profile	4000	150
Simulation 2 – Bounds from audit	3750	140

To reduce the running time of the algorithm, the input parameters of the second simulation were reduced. This reduction in parameters does not affect the results significantly, since the error, usually, stabilizes before 140 iterations.

The results of the two simulations are presented in Figure 22 and Figure 23.

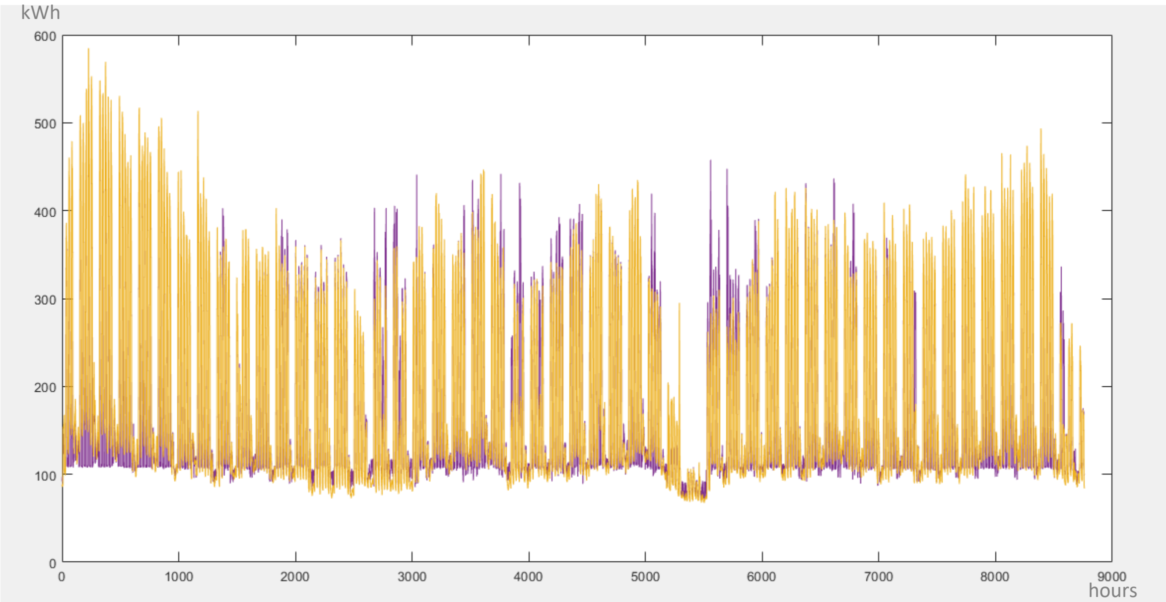

Figure 22 - Total hourly loads comparison: Simulation 1 (purple) vs Real loads (yellow)

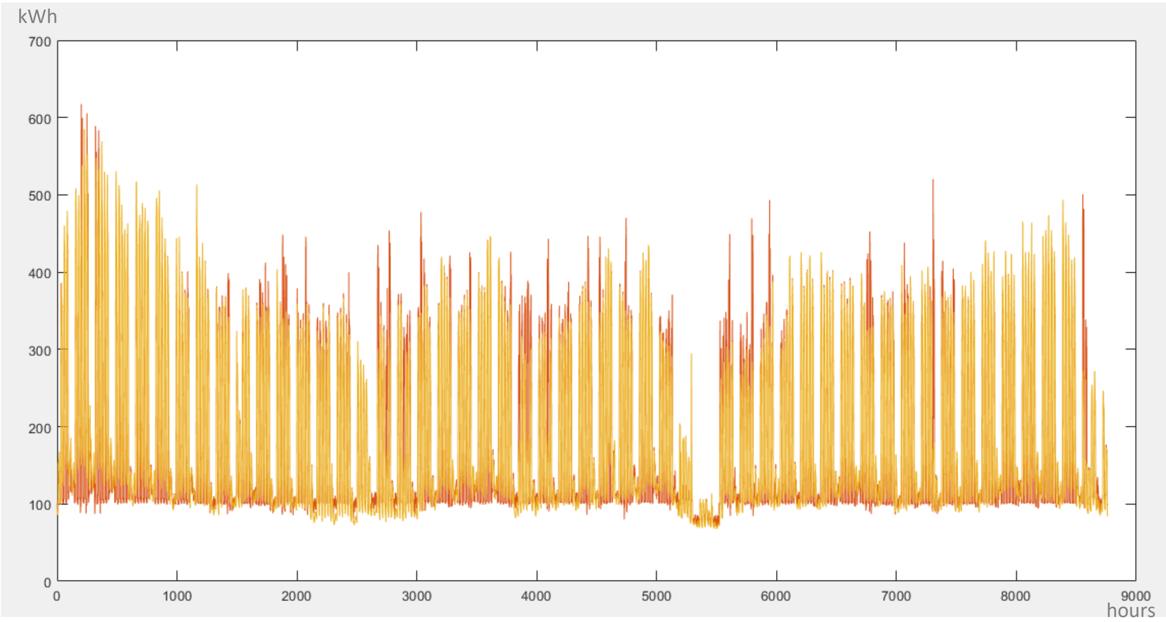

Figure 23 - Total hourly loads comparison: Simulation 2 (orange) vs Real loads (yellow)

Table 11 - Comparison of errors between simulation 1 and 2

	RMSE	*MRE*
Simulation 1	24.55	9.77%
Simulation 2	30.91	12.57%

From Figure 22 and Figure 23, it is observable that, despite some occasional days where the errors are high, the model achieves satisfiable errors for most of the days. It is also noticeable that there is a pattern of lower errors on weekends. That was expected since the loads are considerably lower on these days, therefore the errors also are lower. Also, during these regimes there is less activity and less consumption variability, so, the real consumption does not fluctuate a lot. Therefore, having a smaller range of values for the boundaries leads to an initially smaller margin of error, so the errors are lower.

There seems to be more overestimation of the electricity consumption in more moderate months, that is, during spring and autumn. This could be due to *HVAC*, since there is no separation into more moderate and more rigorous months, the model might be overestimating the HVAC consumption in those months, when in reality, *HVAC* demands are lower. It could also derive from a reduced demand from *lighting* when compared with winter. During winter the days are shorter, so, the *lighting* demand increases. But this demand decreases in the transitional seasons and is lower during summer, where some overestimation can also be observed.

Table 11 presents the random mean square error (RMSE) and the mean relative error (MRE) for both simulations. The errors are smaller in the first simulation, with the use of the *HVAC* profile. This was expected since the information obtained from the profile allowed not only for a closer interval of bounds, but also for more diverse regimes, unlike what was possible to derive from the equipment excel.

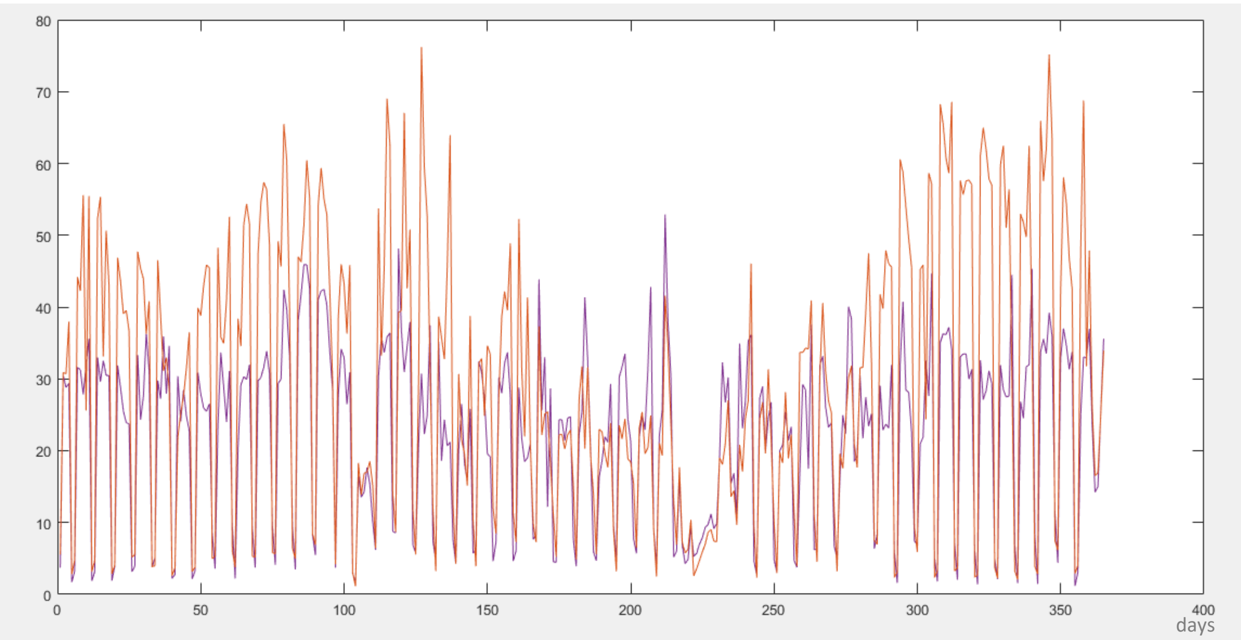

Figure 24 - HVAC daily RMSE comparison: Simulation 1 (purple) vs Simulation 2 (orange)

Looking at Figure 24, there is a clear overestimation of the *HVAC* consumption in the second simulation. This is mostly due to the reduced data available that limited the number of regimes possible to create and led to higher bounds. However, this can also be explained by one of the applied constraints. To ensure the proper allocation of each end-use, the algorithm checked if the end-use shares of that day were similar to the ones determined from the available data.

The end-use shares were relatively close to those values, showing that despite better performance from simulation 1, both simulations predicted the end-use shares well. So, even if real profile data is not available, as long as the boundaries can be well estimated, the model gives good results. This can be seen in Table 12.

Table 12 - End-use shares comparison – Civil building

End-Use Shares	*Lighting*	*Electronics*	*HVAC*	*Others*
Audit data	25%	18%	48%	9%
Simulation 1	28.9%	24.2%	34.1%	12.9%
Simulation 2	26.1%	23.5%	37.9%	12.4%

However, while the *HVAC* share calculated in section 4.2 was 48%, the *HVAC* share obtained from the real *HVAC* profile was around 28%. This also explains why there was an overestimation, since the *HVAC* shares obtained from the two datasets were different. The constraint forced the model to prefer days where *HVAC* consumption was higher than what it should be, considering the *HVAC* profile.

Although real load profiles for every end-use were not available, for the validation of *lighting* and *electronics* end-uses, available simulated profiles in Energy Plus were used, as mentioned in section 4.1.2. Despite these profiles not being very accurate it is possible to extract information from the comparison of both profiles, modelled (GA) and simulated (Energy Plus).

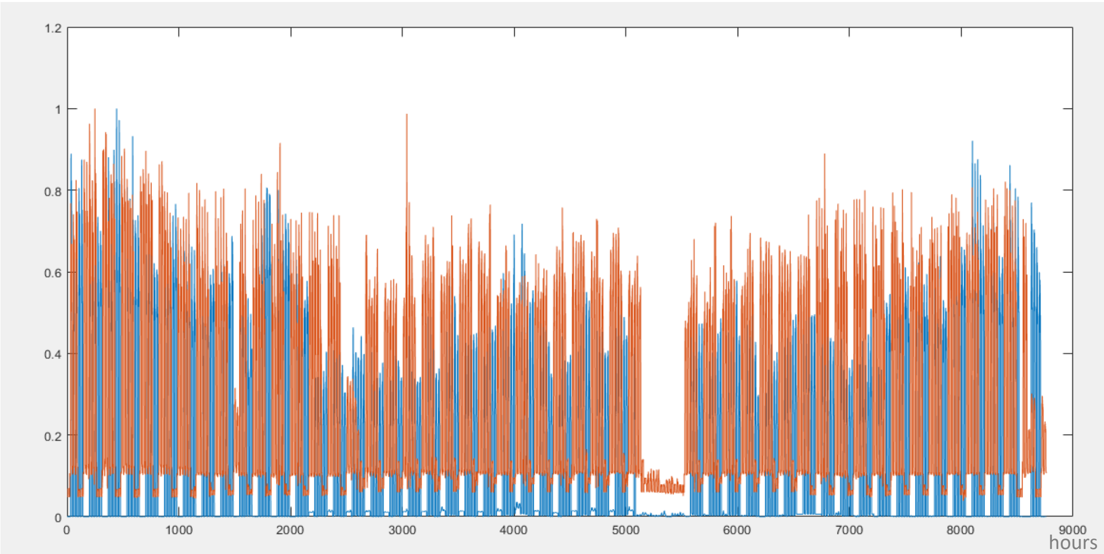

Figure 25 - Comparison between normalized HVAC profile simulated in Energy Plus (blue) and from simulation 2 (orange)

Even though there is some load variation throughout the year, it is noticeable the lack of attention to the regimes of reduced activity in Figure 25. During the night, on weekends and on some vacation regimes like in August, the *HVAC* consumption is assumed to be zero. This is a simplified assumption, not only because there is always some residual consumption, but also because there is some activity, even if reduced, during those regimes. There are people in the building in those regimes, so some *HVAC* equipment will be turned on. Despite the simplicity regarding those regimes, the assumptions used for simulating the weekly loads provide variability and some seasonality can be observed, so, these assumptions overcome the lack of more complex regime separation. However, this is not the case for other end-uses.

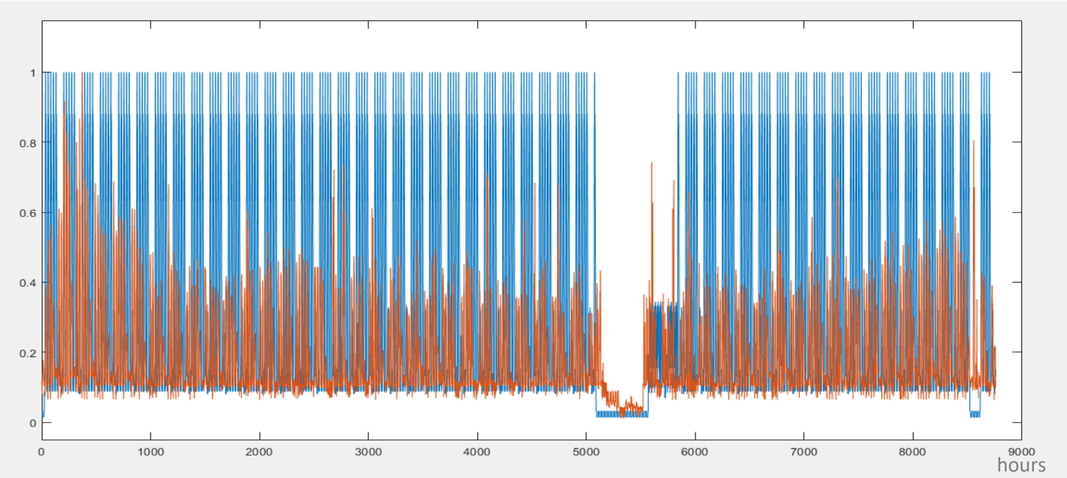

Figure 26 - Comparison between normalized lighting profile simulated in Energy Plus (blue) and from simulation 2 (orange)

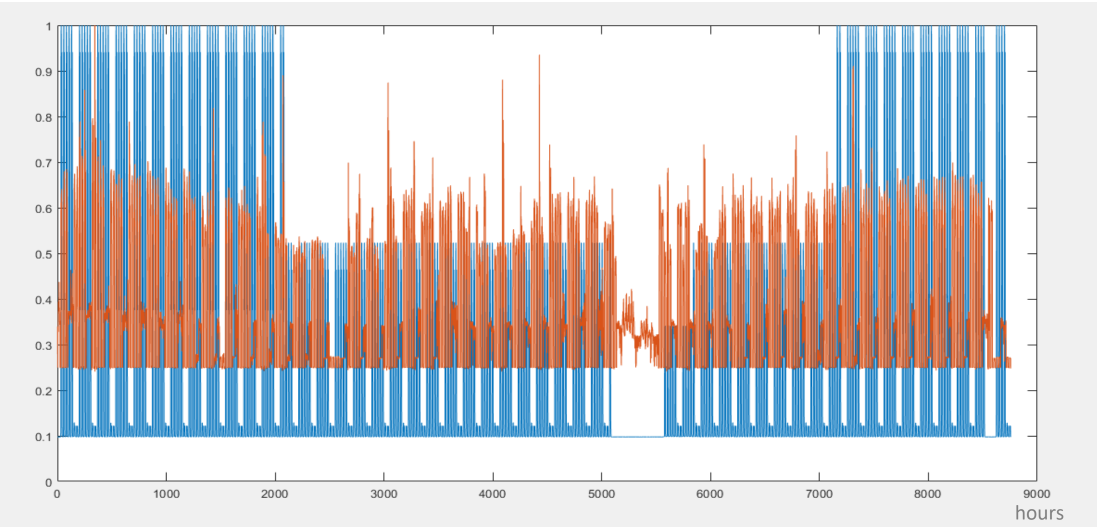

Figure 27 - Comparison between normalized electronics profile simulated in Energy Plus (blue) and from simulation 2 (orange)

As it can be seen in Figure 26 and Figure 27 , Energy Plus simulates *lighting* and *electronics* consumption in a very simplified way. For *lighting*, apart from the weeks of reduced activity in August, the consumption repeats itself in a cyclic way. The consumption during the week is always constant during the day, and the same applies for the night-time and weekends. It is assumed that when there is activity, all the lights in the building are turned on, or at least, that the same lights are always turned on. Like for *HVAC*, there is no room for consumption variation during the night, it is always the same. But unlike *HVAC*, there is no variation during the day either. There is no separation of regimes, there simply are periods of activity and inactivity, as well as the regime of august and one week at the end of the year. There is no seasonality either. It is assumed that the *lighting* consumption is the same for the winter, when the days are shorter and the natural lighting is weaker, as it is during the summer or spring. These are all big simplifications that decrease the significance of the results.

The same can be said for *electronics*, although there seems to be a separation between cooling and heating seasons. However, the consumption still does not variate a lot, there are only a few possible values for consumption, either during the day or the night.

The simulation in Energy Plus Software is based on the input of various types of data such as building characteristics, temperature, occupation or equipment data, and its performance improves with the extent of data used. This way it allows for dynamic and precise simulations when complex and thorough data analysis is performed. On the other hand, it is also possible to use simplifications and reduce the inputs in the model, in order to reduce the need of access to more complex data, at the cost of less precise results. For example, Energy Plus uses a binary variable to represent occupancy schedules. These simplifications resulted in the cyclic and unrealistic results from Energy Plus.

In this study where access to more complex data is limited, the genetic algorithm allowed for a much better understanding of the end-use consumption. The simulated load profiles show much more variability, seasonal patterns, and different regimes with different activity levels. This is a result not only of the diverse regimes created with different bounds and the constraints applied, but also the model itself, since it used real loads instead of just simulating loads based on simplified assumptions. This allows to conclude that the genetic algorithm developed has an edge over Energy Plus when there is a lack of complex, detailed data, leading to simplifications and less insightful data analysis limiting the advantages of using Energy Plus.

5.2 Model implementation on other buildings

Once validated, the model was implemented in a different building, the South Tower, to test the model performance in accurately forecasting end-uses demand. The algorithm ran with 3750 members of population and 140 iterations, for each day.

As it can be seen in Figure 28, there are a few days where the mean relative error is very high, however most of these occurrences are in the regime of reduced activity in August and in the last week of the year.

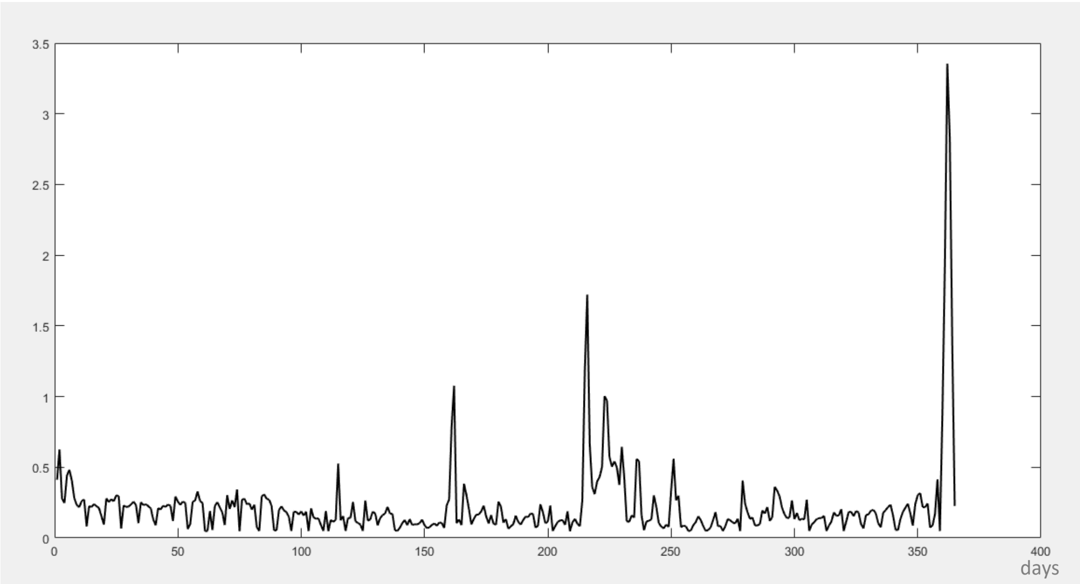

Figure 28 - Daily MRE for south tower building modelling, throughout one year

It can also be observed that the error drops cyclically on weekends, since, as previously noticed, the demand is lower leading to lower absolute errors.

In August, the real loads were significantly higher than expected, therefore the errors were high in this regime since the estimated boundaries did not allow for very high loads. In the final week of the year the loads were very low, and a lot of them were smaller than the sum of the minimum boundary for each end-use. So, once again the errors were high during this week.

Figure 29 shows the MRE for the South Tower building modelling after removing the regime of reduced activity in August (R3) and the last week of the year. Removing these two regimes from the analysis, the results improve. Despite this, there are still a few high values of MRE, and they usually occur in weekends with unexpected very low demand, leading to higher relative errors on these days, just as in the last week of the year. And, since the loads are small, the relative error increases significantly. Table 13 shows the improvement of the errors when those regimes were not taken into consideration.

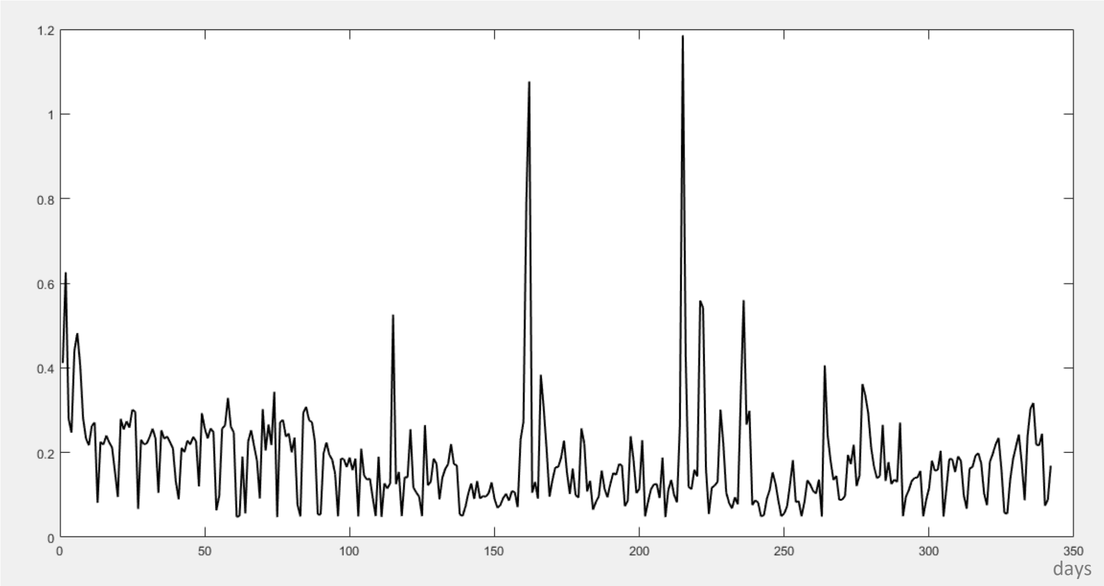

Figure 29 - Daily MRE for south tower building modelling after correction

Table 13 - Comparison of errors of simulation 3

	RMSE	*MRE*
Before correction	34.79	22.36%
After correction	32.96	17.85%

The errors improve significantly with this correction. Since the data used to determine the bounds for Civil was not available for the South Tower, this led to some gross estimations of the bounds. That and the relatively large consumption during the vacations in August in the South Tower comparing to Civil lead to the increase in errors. This shows that with better estimates for the bounds, the results improve considerably. So, despite higher errors in the South Tower simulation, there is potential to improve the model even further, as long as more concrete and precise data is available.

Table 14 shows the comparison between the end-use shares obtained from the real data and the shares estimated by the simulation. Despite some of the high errors mentioned before, the model once again reached good results, since the shares are similar for each end-use.

Table 14 - End-use shares comparison – South Tower building

End-Use Shares	*Lighting*	*Electronics*	*HVAC*	*Others*
Audit data	17%	9%	67%	7%
Simulation 3	23.3%	15.4%	53.7%	7.6%

5.3 Discussion

5.3.1 Comparison of indicators between buildings

From the analysis referred in chapter 4 it was observed that the rooms/spaces that are the largest consumers are technical rooms, by a large margin, followed by investigation labs and offices. After those, classrooms, study spaces and corridors also account for a significant share of the total consumption in a building.

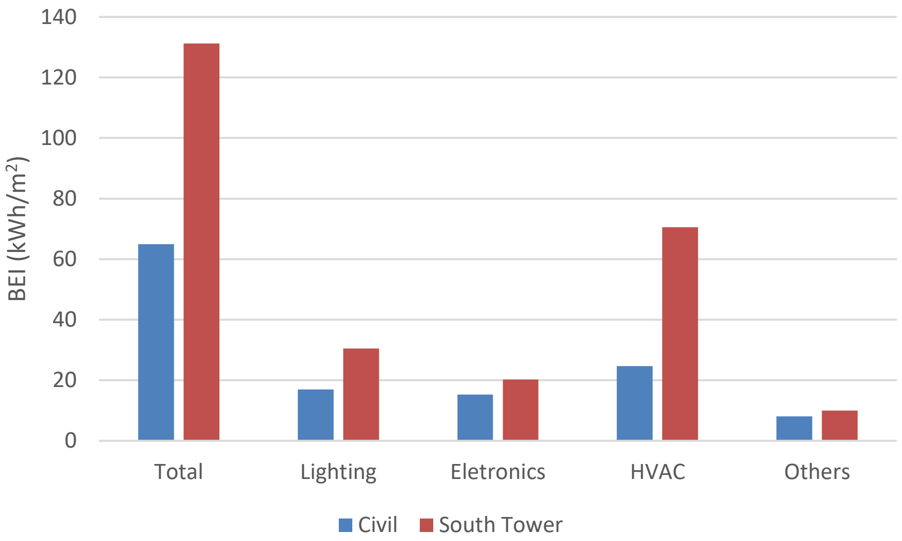

Figure 30 - Comparison of Building Energy Intensity per area unit of each end-use for both buildings

Analysing Figure 30, it can be seen that the consumption per unit of gross floor area in the South tower is higher than in the building of Civil for all end-uses. This could be explained by different reasons. The south tower is a tall building with glass facades. This results in higher solar gains, and it could lead to increased cooling demand in the summer and even in spring and autumn. These differences could also be a result of different occupancy schedules between the two buildings or could result from less energy efficient equipment in the South tower. While for *electronics* and *others* end-uses energy intensities are similar, for *lighting* there is also a larger difference, that, once again, might be explained with the architecture of the South tower, that has all the common spaces inside, leaving offices and classrooms for the exterior facades, while the Civil Building has large common spaces with natural lighting.

However, without further information regarding architecture, occupancy, equipment efficiency or the existence of any retrofitting measures taken in either building, one might try to take conclusions looking at the types of spaces in each building, and how much they represent (in area and electricity demand) for each building.

To better understand the relation between area allocation for each type of space and the Building Energy Intensity (BEI), the ratios between the South Tower BEIs and the Civil BEIs were calculated, to see how much of an increase each end-use has from one building to the other. The ratios for the area allocation were also calculated to better compare their variations. Tables 15 and 16 present these ratios.

Table 15 - Ratio of South Tower BEI over Civil BEI

End-use	*Total consumption*	*Lighting*	*Electronics*	*HVAC*	*Others*
$South\ Tower\ BEI/Civil\ BEI$	2.02	1.80	1.33	2.86	1.24

Table 16 - Ratio of area allocation between South Tower and Civil

Type of Space	*South Tower Area Allocation/Civil Area Allocation*
Classrooms	0.68
Study Space	1.76
Technical Room	1.53
Offices	0.66
Research Labs	2.20
Hallways	0.99

Looking at the technical rooms, there is a larger allocation of area to these spaces in the South tower relative to the building of Civil. Although the increase does not seem very significant, going from 4.66% to 7.12%, this represents a rise of around 50%. Considering that these rooms account for a large portion (40-60%) of the total consumption in the building this increase in area allocation should lead to a significant increase in total energy consumption, especially in *HVAC* since the consumption in these spaces is mostly *HVAC*. Furthermore, since these spaces include *HVAC* equipment like pumps or the central equipment of some systems, this means that a larger allocation to these spaces should mean there are more *HVAC* equipment in the building overall. This can easily be seen by looking at Table 15, since the *HVAC* increase is the highest.

Comparing research labs, there is a significantly larger allocation of area to these labs in the South tower relative to Civil building. These spaces consume a large quantity of electricity in *lighting*, around 40 to 70% of their total consumption is on *lighting*. These could explain why the South Tower has a higher BEI of *lighting* than the building of Civil, especially because many of the labs in the South tower are underground, so the *lighting* consumption there is constantly high.

Considering the higher allocation of area for labs, it is expected that the BEI of the *Others* end-use is higher in the South tower, which is the case. However, this increase is not significant, which means that not only the lab specific equipment is only a portion of the end-use, but also that this equipment is not a major consumer in labs.

Unlike the labs, the allocation of building area for offices is higher in the building of Civil than in the South tower. The major contributing end-use to the total consumption in offices is the *electronics* (40-60%), mostly due to the computers for every teacher/researcher. This should lead to a higher BEI of this end-use in the Civil building, however that is not the case, since the BEI is slightly higher for the South tower.

This could be a result of the bigger allocation for study spaces in detriment of classrooms in the South tower. Classrooms have low to zero *electronics* consumption, while study spaces have a much larger consumption of this end-use, due to students charging their laptops or phones in these spaces for hours. This could be the reason the BEI is more balanced than what should be expected.

From the results, it can be concluded that analysing the area allocation of a building can provide information on how much energy that building consumes and how it is distributed, as long as there is an understanding of the type of activities that occur in each space and how energetically demanding they are.

For these two buildings the technical rooms consume the most out of any space in the buildings, however this is not the case for every building in campus. In some buildings, a sizeable portion of electricity demand comes from data centres, which do not exist in the two studied buildings. So, to extrapolate to other buildings, attention to these details is especially important, and the more different the building is in comparison to the ones discussed in the paper, the more understanding of these patterns of consumption are required.

5.3.2 Comparison of indicators with other studies

Comparing the results from this work with different indicators for various educational buildings around the world [16,31], it can be seen that the BEIs obtained are closer to the values obtained for other school buildings in southern Europe compared to other regions, as seen in Table 17. These values are lower for the southern European countries. These countries have warmer climates, therefore, milder winters. This way there is less demand for heating in these countries, which can explain the lower energy indicators [37]. Since the countries in Southern Europe have similar climates, the energy demands, specifically *HVAC*, are closer than for other regions in the world, so this can explain this difference in values for BEIs. This can especially be seen for Finland, a Scandinavian country, that reaches a maximum BEI of 450 kWh/m^2. This is also observed in Taiwan, which is the country with climate closer to southern Europe climate, and has a BEI for universities in the range obtained in this work.

Table 17 - BEIs for educational buildings in different countries

Southern Europe	*BEI (kWh/m^2)*	*Other regions*	*BEI (kWh/m^2)*
Cyprus (High Schools) [16]	62.75–116.22	Slovenia [16]	192
Greece [16]	92	Taiwan (University) [16]	79
Portugal (This work)	64.9-131.7	Espoo, Finland (University) [16]	89-450
		Manhattan [31]	142

Comparing the energy indicators obtained in this work with ones obtained by R. Jing et al [38] some conclusions can be made (Table 18).

Table 18 - BEIs comparison between IST buildings and Hong Kong office buildings

End-use	*IST Buildings BEI (kWh/m^2)*	*End-use*	*Hong Kong offices BEI (kWh/m^2)*
Lighting	16.97-30.53	Lighting	33.04
Electronics	15.26-20.24	Electronics	28.32
HVAC	24.62-70.53	HVAC	160.48
Others	8.02-9.66	Others	14.16

As can be seen, *HVAC* consumption increases significantly in the Hong Kong offices. This is due to the differences in climates. Hong Kong has a subtropical climate, so in the summer, temperatures are very high, and unlike in universities, the activity levels don’t reduce in the summer, for vacations. So the *HVAC* demand is much higher. The other end-uses consumptions are similar to the ones obtained in this work, which indicates that *HVAC* is the most variable end-use due to climate differences. The higher *electronics* consumption could be due to more electronic equipment, specifically computers, in offices in comparison to university buildings. Elevators are more common and used more frequently in offices, which explains the higher *Others* BEI.

6 Conclusions

Modelling energy demand in buildings is crucial to reach the goals of reducing the energy consumption and CO_2 emissions. However, in many situations, information such on influencing parameters, as occupancy schedules or building characteristics are not available. Finding ways to develop models with more easily accessible information is therefore of crucial importance.

For this purpose, a modified genetic algorithm was developed, where the error between the hourly simulated and real loads was used, as well as the number of constraints not met, to add more information regarding the patterns of consumption for each end-use in the model. The model forecasts the electricity demand of a building and disaggregates it into 4 different end-uses: *Lighting*, *HVAC*, *electronics* and *others*. By using outdoor temperature data and equipment data to estimate the boundaries, the model achieves satisfiable results.

By using IST campus' buildings as case study, the model was validated for the Civil building, for which there was detailed data to train the model, using two approaches: one using hourly *HVAC* loads to tune the model, and other using only equipment characteristics and usage data. After validated and finetuned, the model was applied to the South Tower building, to observe the accuracy of the model on other buildings.

The simulation with the data obtained from the *HVAC* profile resulted in the lower model error, 9.77%, while, without using the *HVAC* profile, the model predicted the energy demand with an accuracy of 12.57%. The model performed better with the additional information, that allowed for a better and more detailed estimation of boundaries. The results also show that the simulated end-use shares are similar to the real ones, so again, the model disaggregates the total electricity demand into the different end-uses at a satisfiable performance. When comparing the end-uses profiles against other dynamic simulation done previously in Energy Plus Software, the model showed more variability throughout the day, week and seasons, demonstrating that it disaggregates the electricity consumption into the different end-uses in a more realistic manner that represents the patterns of consumption and how it changes throughout time. Therefore, this indicates that the model provides more insightful results than energy consumption forecasting programs such as Energy Plus, when there is limited access to real complex data.

For the South Tower, the model did not perform as well with a relative error of 22.36%. However, the bounds were obtained by a rough estimate with some simplifying assumptions, specifically that the patterns of consumption were the same between buildings throughout the different periods. By ignoring the periods where those estimates were further apart from reality, either by having less data for those periods, or because the patterns of consumption differed more, the error decreases to 17.85%. This shows the potential for better performances of the model for different buildings, when having access to data that allows better boundaries estimates. With this in mind, the model developed can be useful in situations where more complex data is not available, since it only requires total hourly consumption profiles and outdoor temperature data, as long as good boundaries and constraints can be estimated.

With the results from the simulations, end-use consumption electricity demand indexes for each building were determined. These indicators allowed for an analysis of the influence of the different types of spaces in each building, and how much total building area is allocated to them, on each end-use consumption. From the analysis, it was observed that there was a relation between the differences in area allocation and the end-use consumptions. The space that most contributed to this variation was the technical rooms. This shows that the area allocations for each type of space, can help understand the total and end-uses consumptions for that building.

As mentioned in section 5.3.2 energy indicators can point to the type of activities in a building. There were clear differences regarding end-use BEI for this work and for residences due to the difference in activities between the two types of building. Comparing to educational buildings the range of values for the total BEI in this work (64.9-131.7 kWh/m^2) were close to values obtained in other studies in countries with similar climate like Cyprus (62.75–116.22 kWh/m^2) and Greece (92 kWh/m^2).

6.1 Limitations and future work

For this work most data used is usually accessible to universities, to develop a model that can be replicated for different studies without the need to have more complex data. However, this limits the performance of the model since simpler data does not provide enough information to forecast energy demand with very low errors. Furthermore, it was only possible to validate the model for one building, as data regarding end-use profiles was not available for other buildings. And, while this data was available for Civil, it was not real data measured by smart meters or by an audit, it was obtained through a rough simulation from Energy Plus Software.

This way, future work should be done to improve the model developed in this study. The model should be validated with real data, if available, for the building of Civil. If not possible, it should be tested for another building where more data regarding end-use profiles is accessible. The relation between *lighting* consumption and solar radiation could also be analysed to potentially add more information to the model. Also, to better understand the viability of using the area allocations to analyse the consumption of each building, simulations for other buildings with more different space types should be done, specifically buildings with data centres, since they are significant consumers, with a somewhat constant consumption during the day and night. This way, a database of end-uses BEIs for university buildings with diverse activities can be created, which can serve as reference for other studies regarding energy consumption in higher education buildings.

Simulations for non-university buildings could also be done to see if the model developed can also perform well for different buildings with other types of activities.

REFERENCES

[1] International Energy Agency (IEA). https://www.iea.org/data-and-statistics/data-browser/?country=WORLD&fuel=Energy%20consumption&indicator=TotElecCons, (last accessed October 2021).

[2] L. Pires, P. Silva and J. Castro-Gomes, “A importância do consumo energético dos edifícios na Europa: soluções propostas para a sua redução,” 2005.

[3] International Energy Agency (IEA). https://www.iea.org/data-and-statistics/data-browser?country=WEOEUR&fuel=Energy%20consumption&indicator=TFCShareBySector, (last accessed October 2021).

[4] International Energy Agency (IEA). https://www.iea.org/data-and-statistics/data-browser?country=WEOEUR&fuel=CO2%20emissions&indicator=CO2BySector, (last accessed October 2021).

[5] J. Bernardo, "Eficiência Energética nos Edifícios da Administração Pública Central", *PO SEUR 03-2016-65*, 2016. https://poseur.portugal2020.pt/media/39465/apresentação-eng-joão-bernardo.pdf, (last accessed October 2021)

[6] Programa de Eficiência de Recursos na Administração Pública. https://www.ecoap.pt/ecoap-2030/, (last acessed October 2021).

[7] X. Gui, Z. Gou and Y. Lu, “Reducing university energy use beyond energy retrofitting: The academic calendar impacts,” *Energy Build.*, vol. 231, 2021

[8] T. Ueno, F. Sano, O. Saeki and K. Tsuji, "Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data", *Applied Energy*, vol. 83, pp. 166-183, 2006.

[9] G. Peschiera, J. Taylor and J. Siegel, "Response-relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data", *Energy Build.*, vol. 42, pp. 1329-1336, 2010.

[10] S. T. Tzeiranak, P. Bertoldi, D. Paci, L. Castellazzi, T. R. Serrenho, M. Economidou and P. Zangheri, "Energy Consumption and Energy Efficiency trends in the EU-28, 2000-2018", *Publications Office of the European Union*, Luxembourg, 2020. https://publications.jrc.ec.europa.eu/repository/handle/JRC120681 (last accessed October 2021)

[11] A. F. Orlando, M. P. Málaga and M. M. Huamani, “Methodology for generating electric load profiles for sizing an electric energy generation system,” *Energy Build.*, vol. 52, pp.161-167, 2012.

[12] B. J. Birt, G. R. Newsham, I. Beausoleil-Morrison, M. M. Armstrong, N. Saldanha, I. H. Rowlands, “Disaggregating categories of electrical energy end-use from whole-house hourly data,” *Energy Build.*, vol. 50, pp. 93-102, 2012.

[13] F. Amara, K. Agbossou, Y. Dubé, S. Kelouwani, A. Cardenas and S. Hosseini, “A residual load modeling approach for household short-term load forecasting application,” *Energy Build.*, vol. 187, no. March 2019, pp. 132-143

[14] R. Mena, F. Rodríguez, M. Castilla and M.R. Arahal, “A prediction model based on neural networks for the energy consumption of a bioclimatic building,” *Energy Build.*, vol. 82, pp. 142-155, 2014.

[15] B. Arregi and R. Garay, “Regression analysis of the energy consumption of tertiary buildings,” *Energy Proc.*, vol. 122, pp. 9-14, 2017.

[16] J. C. Wang, “A study on the energy performance of school buildings in Taiwan,” *Energy Build.*, vol. 133, pp. 810-822, 2016.

[17] A. Agnetis, G. de Pascale, P. Detti and A. Vicino, "Load Scheduling for Household Energy Consumption Optimization," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2364-2373, Dec. 2013.

[18] K. Song, S. Kim, M. Park and H. Lee, “Energy efficiency-based course timetabling for university buildings,” *Energy*, vol. 139, pp. 394-405, 2017.

[19] J. Chen and C. Ahn, “Assessing occupants’ energy load variation through existing wireless network infrastructure in commercial and educational buildings,” *Energy Build.*, vol. 82, pp. 540-549, 2014.

[20] M. A. M. Daut, M. Y. Hassan, H. Abdullah, H. A.Rahman, M. P. Abdullah and F. Hussin, “Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review,” *Renew. Sustain. Energy Rev.*, vol. 70, pp. 1108-1118, 2017.

[21] S. Asadi, S. S. Amiri, M. Mottahedi, “On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design,” *Energy Build.*, vol. 85, pp. 246-255, 2014.

[22] K. Li, C. Hu, G. Liu and W. Xue, “Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis,” *Energy Build.*, vol. 108, pp. 106-113, 2015.

[23] D. Hawkins, S.M. Hong, R. Raslan, D. Mumovic and S. Hanna, “Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods,” *International Journal of Sustainable Built Environment*, vol. 1, issue 1, pp. 50-63, 2012.

[24] H. Dagdougui, F. Bagheri, H. Le and L. Dessaint, “Neural network model for short-term and very-short-term load forecasting in district buildings,” *Energy Build.*, vol. 203, 2019.

[25] S. K. Satapathy, S. Dehuri, A. K. Jagadev and S. Mishra, *EEG Brain Signal Classification for Epileptic Seizure Disorder Detection,* Academic Press, pp. 1-25, 2019.

[26] E. M. de Oliveira and F. L. C. Oliveira, “Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods,” *Energy*, vol. 144, pp. 776-778, 2018.

[27] J. Massana, C. Pous, L. Burgas, J. Melendez and J. Colomer, “Short-term load forecasting in a non-residential building contrasting models and attributes,” *Energy Build.*, vol. 92, pp. 322-330, 2015.

[28] X. Yang, *Nature-Inspired Optimization Algorithms*, 2nd ed., Academic Press, pp. 91-100, 2021.

[29] H. B. Gunay, Z. Shi, I. Wilton and J. Bursill, “Disaggregation of commercial building end-uses with automation system data,” *Energy Build.*, vol. 223, 2020.

[30] F. Rosso, V. Ciancio, J. Dell'Olmo and F. Salata, “Multi-objective optimization of building retrofit in the Mediterranean climate by means of genetic algorithm application,” *Energy Build.*, vol. 216, 2020.

[31] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, V. Modi, “Spatial distribution of urban building energy consumption by end use,” *Energy Build.*, vol. 45, pp. 141-151, 2012.

[32] Rede Campus Sustentável. http://www.redecampussustentavel.pt/, (last accessed October 2021).

[33] American Society of Heating, Ventilating, and Air Conditioning Engineers (ASHRAE). Guideline 14-2002, Measurement of Energy and Demand Savings; Technical Report; American Society of Heating, Ventilating, and Air Conditioning Engineers: Atlanta, GA, USA, 2002.

[34] Energy Plus Software. https://energyplus.net/, (last accessed October 2021).

[35] AuditIST Campus Sustentável 2013. http://www.redecampussustentavel.pt/, (last accessed October 2021).

[36] Renewables Ninja. https://www.renewables.ninja/, (last accessed April 2021).

[37] S. T. Tzeiranaki, P. Bertoldi, F. Diluiso, L. Castellazzi, M. Economidou, N. Labanca, T. R. Serrenho, P. Zangheri, “Analysis of the EU Residential Energy Consumption: Trends and Determinants,” *Energies* 12, 1065, 2019.

[38] R. Jing, M. Wang, R. Zhang, N. Li and Y. Zhao, “A study on energy performance of 30 commercial office buildings in Hong Kong,” *Energy Build.*, vol. 144, pp. 117-128, 2017.