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

Nuno Miguel Martins Senra Almeida
nunommsalmeida@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal
December 2021

A B S T R A C T

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

1. Introduction

Buildings account, on average, for 40% of the energy use in Europe and one third of the CO_2 emissions [1,2]. It is believed that more than 50% of the energy consumption could be reduced by implementing measures to improve energy efficiency [3]. 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 [4,5]. 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 [6]. Therefore, it is important to implement energy saving strategies for service buildings.

Many universities have committed to reduce the energy consumption in their campi [7]. Knowing the energy demand of each building is important to analyse the current situation and find more efficient energy systems.

Modelling energy demand has, therefore, become of crucial importance. Many studies have tried to find parameters that could influence energy consumption, like climatic factors, occupancy or construction solutions (insulation, architecture).

F. Amara et al [8] found that temperature is one of the more important factors that influence consumption. R. Mena et al [9] studied the energy consumption on a solar energy research centre and found that temperature and solar radiation were both influencing factors.

J. C. Wang [10] conducted a study on energy indicators for education buildings of different education levels in Taiwan. 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.

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 occupants do not have financial incentives to reduce consumption as they do in their residences.

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-user's comfort. By convenience of accessing data and for model validation purposes, this study used IST as a case study.

This study is organized as follows. Section 2 reviews studies regarding optimization algorithms and analysis of energy indicators; Section 3 details the methodology regarding the correlation between different influencing parameters of electricity demand in buildings, followed by how the model was developed; Section 4 describes the case study and the data used; Section 5 shows the results for the model validation with the case study, as well as for the implementation for another building; Section 6 presents the conclusions.

2. Literature Review

There are many different models that try to predict energy consumption, and they are grouped into different methods. There are classic methods such as stochastic time series models, which include ARIMA, ARMA and autoregressive models, and regression models, like linear regression (LR) and multiple linear regression (MLR). 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 [11].

S. Asadi et al [12] 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.

J. Massana et al [13] 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 an 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. 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 [14].

H. B. Gunay et al [15] used a genetic algorithm to help disaggregate the total loads of an academic office building into three different end-uses, lighting and plug loads, 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 [16] 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. [10] 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 [17] estimated the energy indicators in Manhattan for different types of buildings, such as offices, residential, schools and others.

3. Methodology

A genetic algorithm (GA) was developed to forecast electricity demand in university buildings and disaggregate it into end-uses. First, data including total energy consumption, end-use shares, hourly load profiles, data equipment and outside temperature were collected and processed. Afterwards, this data was analysed to determine any relation between consumption data and possible influencing parameters. To improve the analysis, the data was separated into different regimes. The patterns of consumption vary throughout the year, so by separating the data into different regimes according to the activity should improve the analysis. Table 1 shows the different regimes. The HVAC data was also divided into heating and cooling seasons.

3.1 Correlation of electricity demand influencing parameters

The relation between HVAC consumption and outside temperature was studied. The correlation between the average HVAC consumption during the day (08:00-19:00) and the average temperature (T) was determined using the *corrcoef* function in Matlab. This was done for every regime as shown in equation 1.

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

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

Table 1 – Regimes segmentation and characterization

R1	Sundays on periods of exams/outside of school period
R2	Saturdays on periods of exams/outside of school period
R3	Reduced activity (Two weeks in august)
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.)

3.2 Prediction model

The developed genetic algorithm will produce many different possible solutions, the members of population, which will reproduce, creating offspring that are better solutions to the optimization problem. This is done throughout many generations

3.2.1 Members of population

Each member of the population (nPop) will simulate the hourly loads for the 4 end-uses for an entire day. Each member resembles a matrix with 24 by 4 dimensions, where the columns represent each end-use consumption profile, and each line represents the hourly consumption. 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).

$$\sum_{e=1}^{4} nPop_{h,e} = Total\ Hourly\ Energy_h, \ \ h = \{1,2,\dots,24\} \quad (2)$$

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

So, the model will optimize each end-use electricity consumption for each day and in the end will join the results.

3.2.2 Boundaries

It is important to guide the model to forecast the end-uses consumption as accurately as possible. To do so, it is necessary to implement information to reduce the errors and approximate the results to reality. This was done by adding bounds that limit the generated values and constraints, which will be discussed in the next section.

These bounds are estimates of the maximum and minimum possible values for each end-use consumption at any given hour. So, when generating hourly values for each end-use, the algorithm will randomly choose a value in between the bounds implemented. This way, the bounds will assure that the initial error will be smaller, and therefore, so the final solution will have a lower error, and the results will be closer to reality.

To estimate these bounds, the equipment data detailed in section 4.1 was used. The nominal power ($\bar{P}_N$) was obtained by dividing the total consumption ($Energy_{Total}$) of each end-use by the average working time of the equipment ($\overline{Work\ Time}$).

$$\frac{\sum Hours}{Number\ of\ equipment} = \overline{Work\ Time}_{\ R,End-use}\ h \quad (4)$$

$$\frac{Energy_{Total,R,End-use}}{\overline{Equipment\ Work\ Time}_{\ R,End-use}} = \bar{P}_{N,R,End-use}\ kW \quad (5)$$

The same was done for the standby consumption, and so, the upper limit (max bound) was obtained.

$$\overline{P_{N,R,End-use}} + \overline{P_{Standby,R,End-use}} = MaxBound_{R,End-use} \quad (6)$$

This was done for every end-use and regime. The minimum bounds are equal to the permanent consumption ($PermConsump$).

$$MinBound_{R,End-use} = PermConsump_{R,End-use}, \quad (7)$$

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

As HVAC data was available for the Civil building, a second method of determining HVAC bounds was used. The hourly HVAC profile was separated into the different regimes created, as well as separated by day and night, and heating and cooling season. The weekdays regimes (R6 and R7) were further divided into two groups. HVAC systems exist mainly to achieve thermal comfort of occupants, so this consumption should increase when temperatures are harsher, when compared to more moderate days. 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.

$$DailyMeanT_{d,R,Season=}\frac{\sum T_{d,R,Season}}{12\ hours} \tag{8}$$

$$MeanT_{R,Season} = \frac{\sum DailyMeanT_{d,R,Season}}{Number\ of\ Days_{R,Season}}, \tag{9}$$

$$R = \{6,7\}, \quad Season = \{Heating, Cooling\},$$

$$, d = \{1,2,\ldots, Number\ of\ Days\}$$

For each one of these groups, the average of the maximum and minimum hourly loads for each day were calculated.

$$MaxBound_{R,HVAC,Season} = \frac{\sum max\ (HVAC_{R,d,h,Season})}{Number\ of\ Days_{R,Season}}, \tag{10}$$

$$MinBound_{R,HVAC,Season} = \frac{\sum min\ (HVAC_{R,d,h,Season})}{Number\ of\ Days_{R,Season}}, \tag{11}$$

$$, h = \{8,9,\ldots,19\}, \;\; R = \{1,2,\ldots,8\},$$

$$, d = \{1,2,\ldots, Number\ of\ Days\}, \;\; Season = \{Heating, Cooling\}$$

This way, the model can be tested with two different sets of bounds to compare the performance of the two.

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 ($HVACRatio_{Real}$). 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.

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

In the simulation, the model imposes that the ratio of the generated values must be equal or higher than the previously calculated ones, as shown in equation 13.

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

$$, R = \{1,2,\ldots,8\}, \qquad Time = \{Day, Night\},$$

$$, Season = \{Cooling, Heating\}$$

Despite not knowing how the consumption changes from hour to hour, it is known that the consumption ($TotalE$) 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.

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

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

Finally, to assure the end-use shares will be close to reality, some data regarding end-use shares should be implemented. Since the algorithm runs each day at a time, it is impossible to check if the end-use shares of the entire year are correct. So, to add this information to the model, the end-use shares will be evaluated only for the weekdays, since they are the more numerous days and they are the biggest representatives of the final end-use shares.

To allow for some variation, instead of checking if the simulated end-use shares ($EUShare$) 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.

$$EUShare_{Sim} \in [EUShare_{Real} - \delta; EUS_{Real} + \delta], \tag{16}$$

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. When this happens, it is expected that the lights will be turned on and more electronic equipment or HVAC will be used. So, when one end-use is consuming more energy, the others should to. 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. This separates the periods of time when consumption is only due to standby, when occupancy is zero, and periods when there are occupants in the building, so equipment is being used. This will assure that when lights are turned on, other equipment will also be turned on.

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

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

3.2.3.3 Constraints regarding HVAC consumption and temperature

Finally, constraints that relate HVAC consumption with the ambient temperature were also added. The correlations obtained in section 3.1 were used for this purpose.

The results 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.

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 during regimes of activity ($BaseCorr$). 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. The value of $BaseCorr$ was determined by analysing the results and it will be explained in section 4.1.

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

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

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

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

$$D = Day$$

3.2.4 Reproduction and survival of the best

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)** and **the count of constraints not complied (CCount).** The error is the mean absolute error for the entire day, comparing the total simulated consumption ($SimHourlyE_h$) for each hour (summing the end-uses) with the real value for that hour ($RealHourlyE$), for the corresponding building.

$$|SimHourlyE_h - RealHourlyE_h| = Hourly\ Error_h, \quad (22)$$

$$CCost = \frac{\sum Hourly\ Error_h}{24} \quad (23)$$

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

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. If the model only sorted the members by the error, this could lead to situations where the error could be close to 0, however the model would fail to deliver end-use load disaggregation that would make sense. As such, complying with constraints is the priority of the model, followed by achieving the minimum error.

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.

3.2.4.2 Crossover and Mutation

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. 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.

3.2.4.3 Sorting and survival of the best

After reproduction parents and children are joined in the same population, so the best ones survive to the following generation. The members with lowest CCount are the best and are sorted by lowest CCost. Then the members with the second lowest CCount are the next group of members to be sorted by CCost. This proceeds until the original size of population is reached, and the worst members are eliminated.

The process of reproduction and sorting reoccurs every generation until one stopping criteria is reached, the maximum number of iterations is reached, or the mean relative error (MRE) of the best member of the population is lower than 5%. ASHRAE guideline 14 [18] states the limits of two relative errors, NMBE and CVRMSE, are 10% and

30% respectively, for hourly predictions. So, an MRE of 5% was assumed to be an acceptable error.

$$MRE_h = \frac{\sum \frac{|Total\ Hourly\ Energy_h - Real\ Hourly\ Energy_h|}{Real\ Hourly\ Energy_h}}{24} \quad (24)$$

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

4. Case Study Description

For this work, the Alameda campus of IST, in Lisbon was used, due to accessibility of data.

The total yearly consumption in 2019 for every room of every building on campus was available, separated by end-uses, namely Lighting, Electronics, HVAC and Others (pumps, elevators, lab equipment). This data was obtained by an audit provided by "Campus Sustentável" project [19].

The electricity hourly loads for 2019 were available for five buildings (Central, Civil, Mathematics, North Tower, South Tower) and were obtained by smart meters. These loads were processed in order to remove errors. Missing data was replaced by the average of the nearest hours when available, and missing data that lasted a few days were replaced by the correspondent hourly values from the week before or after.

Weather data, obtained from *renewablesninja* [20], specifically hourly profile of outside temperature in Lisbon was used.

4.1 Civil Building

The Civil building was used to develop the model since there was more available data for this building.

Equipment data was available for this building. An audit made by "Campus Sustentável" [21] registered all equipment in the building and their nominal power, total consumption, standby consumption and number of working hours. The consumption and hours were divided into 5 different regimes: AR1-Weekdays, AR2-Holidays, AR3-Saturdays, AR4-Sundays and AR5-2 weeks of vacations in August. These regimes were matched with the regimes mentioned in section 3, as show in Table 2. The permanent consumption was also available, that is, the minimum electricity consumption at any given hour, for each end-use and regime.

Table 2 – Regimes grouping

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

Real hourly HVAC profiles for 2017 and 2018 were also available, as well as hourly modelled profiles for total, lighting, electronics and HVAC consumption simulated in Energy Plus Software [22]. The profiles from 2017 were used and adjusted to 2019, so that weekdays in one year did not coincide with weekends in the other year.

Table 3 – Correlations between HVAC consumption and outside temperature per regime

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

Table 3 shows the correlations between HVAC consumption and temperature. It is observed that 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%). This is due to the low occupancy in the latter regimes. Since the occupancy is low, the need to achieve thermal comfort is very reduced, so the HVAC consumption becomes independent of the temperature.

There is also a positive relation between HVAC consumption and temperature during cooling season, and a negative one in heating season. During cooling season, when temperatures increase, HVAC consumption also increases. On the other hand, in heating season, HVAC demand increases the lower the temperature is. This relation is evident during the week (R6 and R7) but not so much on other regimes.

With this in mind, the $BaseCorr$ for R6 and R7 is 20%, while for all other regimes it is 30%. Since R8 only includes a few days of data, the correlations obtained were very different from all other regimes of lower activity. They were considered outliers, and the $BaseCorr$ for R8 is equal to the regimes of reduced activity.

4.2 South Tower

After developing and validating the model for Civil, the model was tested for another building, the South Tower. Since equipment data was not available for this building, the bounds were estimated differently.

By extrapolating the average working time of equipment and the regime share of each end-use to the data from the 2019 audit the bounds were calculated. The regime share is the percentage of total consumption attributed to each regime (equation 25). By multiplying the total yearly consumption ($TotalE$) from the 2019 consumption audit with the regime share ($RShare$), equations 4 and 5 could be used to obtain the max bounds.

$$\frac{Energy_{Total,R,End-use,Civil}}{Energy_{Total,End-use,Civil}} = RShare_{R,End-use} \quad (25)$$

$$RShare_{R,End-use} * TotalE_{End-use,South\ Tower} =$$

$$= Energy_{Total,R,End-use,South\ Tower}\ kWh/year \quad (26)$$

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

4.3 Comparison between buildings

To analyse the relation between end-use consumption and type of space it was first needed to know which rooms/spaces were the largest consumers. By looking at the total consumption data, the biggest consumers were technical rooms, research labs, hallways, offices, classrooms and study spaces. In Civil, these spaces were responsible for 92% of the total consumption in the building. In the South Tower they were responsible for 98% of the consumption. Technical rooms were by far the major consumers, 49% in Civil and 61% in the South Tower.

After that, it was important to understand how each room/space consumes energy, by finding the yearly end-use shares for each one. Table 4 presents the end-use shares ranges for Civil and South Tower.

Table 4 – Range of end-use shares per type of space

	Lighting (%)	Electronics (%)	HVAC (%)	Others (%)
Classrooms	56.19-81.14	4.78-21.18	14.08-22.63	0
Study Spaces	34.87-43.14	55.7-58.94	1.16-6.2	0
Technical Rooms	0.07	0.02-0.31	81.3-96.64	2.98-18.6
Offices	21.62-26.04	44.71-65.01	8.98-33.68	0
Research Labs	40.7-73.37	14.96-23.11	3.51-23.99	0-20.36
Hallways	70.42-94.08	1.89-3.35	2.58-22.29	0-5.41

5. Results and discussion

5.1 Validation

To validate the model, two simulations were performed for the Civil building. **Simulation 1** used bounds obtained from the HVAC profile. **Simulation 2** used bounds derived from the equipment data. The input parameters for simulation 1 were 4000 members of population and 150 iterations. To save time, and because the error stabilized before 140 iterations, the input parameters for simulation 2 were reduced to 3750 members and 140 iterations.

As expected, using the HVAC profile led to better results, as seen in Table 5. Despite this, both simulations show satisfiable performance.

Table 5 – Error comparison between simulation 1 and 2

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

Looking at the HVAC forecast of a typical month (Figure 1), there is a noticeable increase of error in the second simulation. This explains the higher error for the total consumption since the other bounds did not change. The HVAC profile allowed to obtain better estimates for the bounds. The hourly values were available, so the bounds obtained are closer to reality than estimates obtained by information from the equipment data audit.

The prediction errors are smaller during weekends. This is due to the smaller real consumption leading to lower errors.

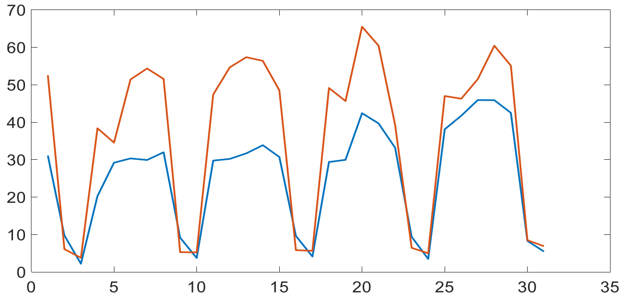

Figure 1 – Comparison of daily HVAC RMSE in the month of March (31 days): Simulation 1 (blue) vs Simulation 2 (orange)

Table 6 compares the simulated end-use shares of both simulations with the shares calculated from the energy consumption data. The simulated shares are close to the calculated values, showing the good performance of the model for both simulations.

Table 6 – End-use shares comparison – Civil building

End-Use Shares	Lighting	Eletronics	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%

The end-uses profiles of simulation 2 were compared with profiles simulated in Energy Plus Software. The developed 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. Figure 2 shows both simulated lighting profiles during a week in March.

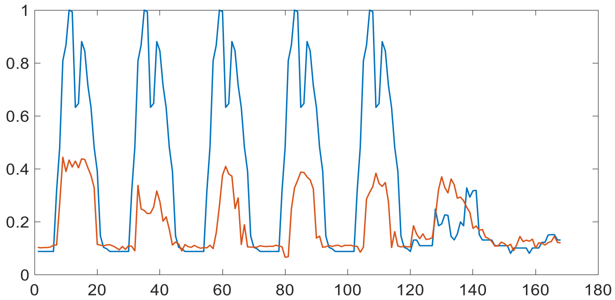

Figure 2 – Normalized lighting consumption comparison in a typical week: Energy Plus (blue) vs Simulation 2 (orange)

5.2 Implementation in South Tower

After validation, the model was implemented in another building, the South Tower, to test the model performance in accurately forecasting end-uses demand. **Simulation 3** was performed with the same input parameters as simulation 2, but for South Tower building.

The results show a few days with very high errors. However, these days occur during the vacation period in August (R3) and the last week of the year. Due to the lack of data, bounds had to be estimated assuming the patterns of consumption between buildings did not change, but this was not the case. The measured loads were considerably higher than expected in August and very low in the last week of December. This led to high errors since the bound's estimates were significantly different than they would have been with more available data.

Table 7 shows the error improvement after removing those two regimes from the results. This demonstrates how important it is to have data to obtain good estimates, as both simulations for Civil had more available data than the South Tower.

Table 7 – Error comparison of simulation 3

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

Just as in Civil, the model predicted the end-use shares for the South Tower with good results (Table 8). So, despite the higher error, these results show the potential of implementing this model to different buildings, as long as data regarding end-use consumption is available.

Table 8 – End-use shares comparison – South Tower

End-Use Shares	Lighting	Eletronics	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

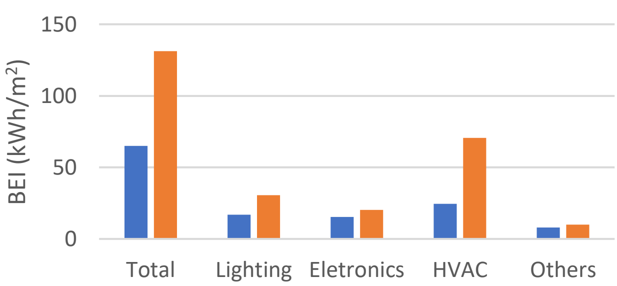

Figure 3 – Comparison of Building Energy Intensity per unit area of each end-use for both buildings

Analysing Figure 3, 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, such as building characteristics, occupancy or inefficient equipment or processes.

However, without further information regarding architecture, occupancy, equipment efficiency or the existence of any retrofitting measures taken in either buildings, 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. Table 9 and 10 present these ratios.

Table 9 – Ratio of South Tower BEI over Civil BEI

End-use	Total consumption	Lighting	Electronics	HVAC	Others
$\frac{South\ Tower\ EBI}{Civil\ EBI}$	2.02	1.80	1.33	2.86	1.24

Table 10 – Ratio of area allocation between South Tower and Civil

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

There seems to be a relation between consumption and area allocation for each type of room.

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 around 50% more. These spaces are the biggest consumers, especially regarding HVAC. This increase in area allocation should lead to a significant increase in total and HVAC consumption. This does occur as the HVAC ratio is the highest.

The South Tower allocates much more area to research labs than Civil, more than double. These spaces consume large amounts of electricity in lighting (40-70% of total consumption). Therefore, the BEI should increase for lighting, as it does since after HVAC, the biggest end-use consumption increase was lighting.

Despite the bigger allocation to labs, the "Others" end-use did not increase significantly due to the low consumption of this end-use in labs. Electronics consumption does not vary significantly, which can be explained by the area allocation of offices and study spaces. Offices and study spaces are the biggest consumers of this end-use. While Civil has a bigger area share of offices, the South Tower has a bigger share of study spaces, so they balance out.

5.3.2 Comparison of indicators with other studies

Comparing the total BEIs from this work with different indicators for various educational buildings around the world [10], the range obtained (64.9-131.7) is closer to the values obtained for other school buildings in southern Europe compared to other regions (Table 11). Since the countries in Southern Europe have warmer climates, the energy demands, specifically HVAC, are lower than for other regions in the world, so this can explain this difference in values for BEIs.

Table 11 – BEIs for educational buildings in different countries

Southern Europe	BEI (kWh/m^2)	Other regions	BEI (kWh/m^2)
Cyprus (High Schools) [10]	62.75–116.22	Slovenia [10]	192
Greece [10]	92	Taiwan (University) [10]	79
Portugal (This work)	64.9-131.7	Espoo, Finland (University) [10]	89-450
		Manhattan [17]	142

By comparing the BEIs for IST with the BEIs obtained in [23] it is noticeable the large increase in HVAC consumption in the offices (160.48 kWh/m^2), which is mainly due to the subtropical climate of Hong Kong, and the constant activity during the summer, unlike in IST, where occupancy reduces in the summer during vacations. The other end-uses are similar to the ranges obtained, electronics (28.32 kWh/m^2) and others (14.16 kWh/m^2) have higher consumptions, mainly due to more computers and more frequent use of elevators.

6. Conclusions

In this work, a modified genetic algorithm was developed, to forecast the end-uses consumption of a building. 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 use data. After validated and finetuned, the model was applied to the South Tower building, to observe the accuracy of the model on other buildings. By using outside temperature data and equipment data to estimate the boundaries, the model achieves satisfiable results.

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.

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 due to the lack of data. 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 ambient 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. It was concluded that analysing the area allocations for each type of space, can help understand the total and end-uses consumptions for that building.

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. 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=WEOEUR&fuel=Energy%20consumption&indicator=TFCShareBySector, (last accessed October 2021).

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

[3] 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).

[4] 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.

[5] 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.

[6] 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).

[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] 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

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Campus Sustentável project. https://sustentavel.tecnico.ulisboa.pt/, (last accessed October 2021).

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

[21] AuditIST Campus Sustentável 2013, https://sustentavel.tecnico.ulisboa.pt/, (last accessed October 2021).

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

[23] 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.