Politecnico di Torino
Master course in ICT for Smart Societies

ICT in Building Design

Designing Optimal Residential Building in Oslo and Forecasting of Indoor Air-Temperature for Energy-Efficient Management

Students:

Braccio Jacopo s273999

De Luca Sara s279897

Radaelli Daniele s277488

Ruberti Alessio s277489

Professors:

Chiesa Giacomo

Bottaccioli Lorenzo

Academic Year 2020/21

Contents

1 Introduction

Buildings are one of the largest consumer of primary energy and attaining their efficiency is, therefore, an important goal.

The energy consumed in buildings in developed countries comprises 20-40% of their total energy use and it is above that of industry and transport in the EU [1] and US [2]. Therefore knowing the characteristics of a building is necessary to manage and control its use. Moreover, knowing the energy use is important to promote demand analysis, energy feedback, Demand Response and Demand Side Management applications. In this project we will model a residential building, through the use of the software DisignBuilder, varying all possible configurations of the construction parameters, in order to find the scenario that optimizes the consumption. Once the optimal configuration has been found, we will perform an EnergyPlus simulation in order to verify the behaviour of the internal temperature of the building over one year and analyze the energy signature. Finally, a model useful for Predictive Model Control will be implemented using two configurations of an Multi Layer Perceptron able to forecast the internal temperatures of the buildings.

2 Modelling

2.1 Objective

Based on the objectives indicated in the course, this section will describe the energy need variations of a residential building located in Oslo, Norway. The target is to optimize the electricity need for cooling, heating and lightning, (Q).

2.2 Model of the building

The building model is a residential building prototype modelled in 3D through the use of the Design Builder (DB) software. The building is divided in three thermal zones, as shown in Figure 1, with the following activity templates:

- Zone 1: Domestic Toilet

- Zone 2: Domestic Bedroom

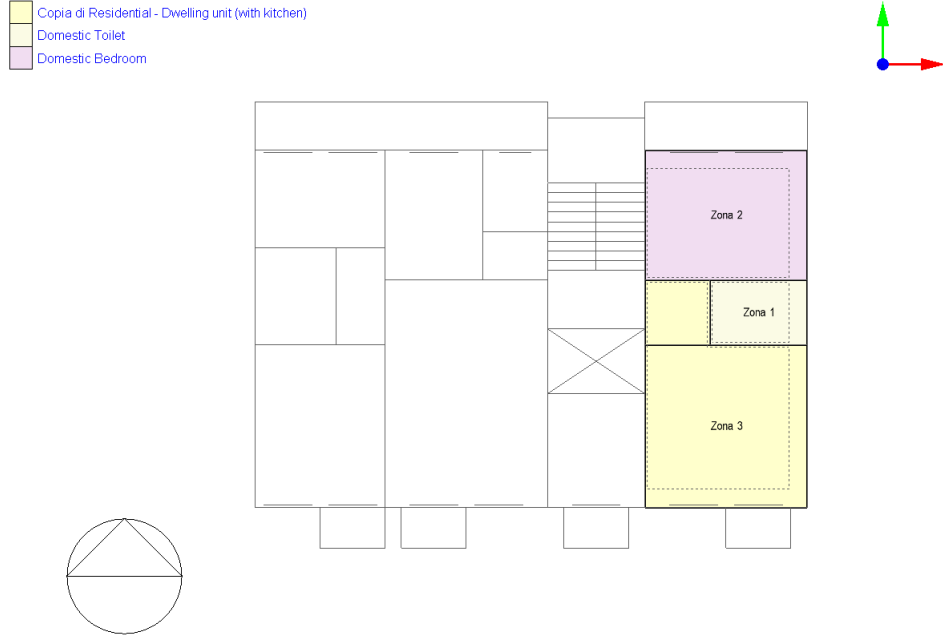- Zone 3: Dwelling unit (with kitchen)

Figure 1: Plan of the building.

People's presence is the noise added to the building characteristics: every person affects the building energy balance, adding 13 W of heat noise to the system. It is modeled using the activity schedule "Residential Occ", but since in the building there are only two people, this noise is negligible. Moreover the natural ventilation, when it is on, follows the same schedule "Residential Occ" and it activates only if the external temperature is between 16° and 24°C.

Heating and Cooling Setpoint Temperatures are initially set as:

- Heating and Heating set back: 20°C and 13°C

- Cooling and cooling set back: 26°C and 32°C

The cooling systems uses as fuel the electricity from the grid, while the heating uses gas and their annual schedule is set as follows:

- Heating: Winter schedule (on from 10 Sep until 31 May)

- Cooling: Summer schedule (on from 31 May until 10 Sep)

2.3 Optimization

2.3.1 Scenario variables

The residential building is modelled over all possible configurations of the parameters described in Table 1

Variable name	Number of variations	Variation values
Natural ventilation	2	on/off
U-value opaque surfaces	5	0 m - 0.35 m
U-value transparent surfaces	3	Single, Double, Triple glazing
Window to wall ratio (WWR)	3	5%, 15%, 50%
Orientation	4	0°, 90°, 180°, 270°
Air Change per Hour (ACH)	10	0 - 9

Table 1: Possible configurations.

In order to obtain the variations on natural ventilation and U-value of transparent surfaces, six different .idf files were created, each one for a different configuration of natural ventilation and glazing type. Instead, in order to obtain the variations on every other variable it was used Besos library on Python. For instance the variation of U-value for opaque surfaces is obtained by varying the thickness of the insulation layer of external walls. The variable ACH is not directly present in the .idf file, therefore in order to obtain this variation the value in m^3/s of flow rate present in the file was multiplied by the volume of the building and divided by 3600, thus obtaining the global flow rate.

The overall number possible combinations is 1800 for the case with natural ventilation on and 180 for the case with natural ventilation off (ACH is not a simulation variable in this case).

2.3.2 Optimization of the scenario variables

For each possible combination of variables is evaluated the total energy consumption of the building. This evaluation has been done by applying "Oslo-Fornebu" .epw file which contains the Typical Meteorological Year. This weather file is the one provided by DesignBuilder for the city of Oslo.

It turned out that the configuration with the lowest consumption had the characteristics showed in Table 2.

Variable	optimized value
Natural ventilation	on
U-value opaque surfaces	0.35 m
U-value transparent surfaces	Triple glazing
WWR	50%
Orientation	90°
ACH	9

Table 2: Best-optimized scenario.

The total energy consumption of the building with this configuration is 168.55 kWh/m^2. It is important to note that an increase in the thickness of the insulating layer leads to a proportional decrease in the floor area of the building. Although passing from a thickness of 0.01m to 0.35m the area decreases by 6m2, the lower consumption expressed in kWh/m^2 occurs in conditions of maximum insulation. However, due to the low temperatures in oslo, it is still convenient to have a highly

insulated building that reduces heat loss, it does not matter if the design expense of the structure is high.

It also appears reasonable that the best orientation of the building is with the wall with the largest windowed area oriented towards east. Moreover the result of a very high ACH is also reasonable, because the higher the air change (especially during summer) the lower the energy expended for cooling the building. Finally it appears that considering the same insulation and U-value settings, the file with ACH on is always better than the one with ACH off in terms of consumption.
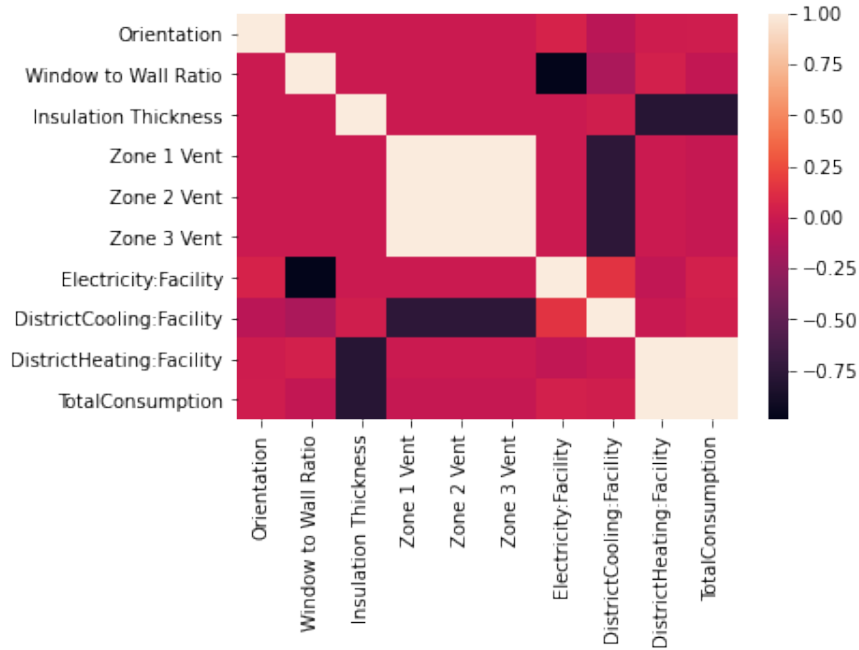
2.4 Correlation between variables

Figure 2: Heatmap of the correlations between variables

The heatmap in Figure 2 represents the correlation between the input variables used in the optimization and those obtained in output. Several considerations can be deduced:

- the orientation has almost no influence on the power consumption;

- the WWR has negative correlation the electricity consumption of the building, because more windowed surface means less use of artificial light;

- the insulation thickness has a negative correlation on the heating and total consumption;

- the natural ventilation, divided in zones, has a negative impact on the cooling consumption, this is especially clear in summer;

- the heating consumption is directly correlated with the total consumption since it is the largest component.

3 Simulation

Using the optimal parameters evaluated in the previous section, a new model of the building was created in DesignBuilder, creating a new .idf file, and a new simulation has been performed on that model. This allows us to understand the internal temperature trend over a year and check that the building configuration is correct. Then the output of the simulation, performed on the whole building, for the trend of the internal temperature was displayed creating a dashboard in *Grafana*, as shown in Figure 3.
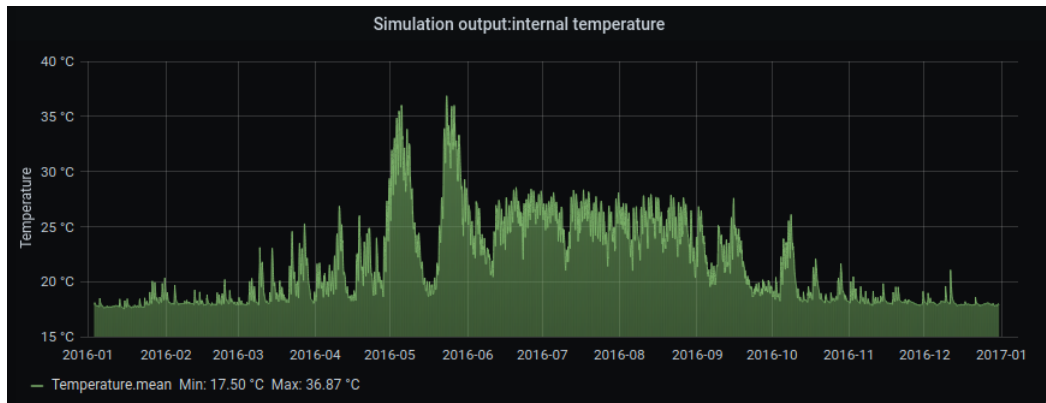

Figure 3: Simulation output: internal temperature (with overheating)

It is evident from Figure 3 that during the period between May and June internal temperatures can reach two peaks of almost 40°C. This may be due to the very high insulation of the building, because it happens in a period when the cooling system has not been activated yet. Moreover it can be noticed that, during the rest of the year, especially in summer, the trend of the internal temperature is oscillating between 20°C an 27°C. This can be attributed to the fact that, being the building highly insulated, the cooling system have difficulty in cooling the house. Moreover the fact that the value of ACH is relatively high, can bring to sudden collapses in the temperatures whenever the natural ventilation is activated.

3.1 Overheating phenomenon

The visualization of the time series of the internal temperature of the building (Figure 3) led to the realization that, with the current configurations of parameters and schedules, an overheating phenomenon occurred.
In order to solve the overheating problem, the following changes for the temperature setpoints were made:

- Heating set back: 14°C

- Cooling set back: 28°C

Moreover new schedules for both heating and cooling system were defined:

- Heating schedule:
  1 from 31 Oct until 31 Mar

0.5 from 31 Mar until 31 May
0 from 31 May until 10 Sep
0.5 from 10 Sep until 31 Oct

- Cooling schedule:
0 from 31 Oct until 31 Mar
0.5 from 31 Mar until 31 May
1 from 31 May until 10 Sep
0.5 from 10 Sep until 31 Oct

This new configuration will allow us to have more constant and linear trend of internal temperature data, better preparing the dataset for prediction. Using this configuration, it was possible to reach the range of comfort of internal temperatures, however the total energy consumption of the building increased by 30kWh. This was due to the obvious increase of consumption imposed by the new schedule of the cooling system and the heating system, that activates in set back also during mid seasons. After these modifications in the model the overheating problem was solved and the result is shown in Figure 4. It can be observed that the peaks have been shaved, however the collapse of the temperatures in between the two peaks is still present in the month of May.
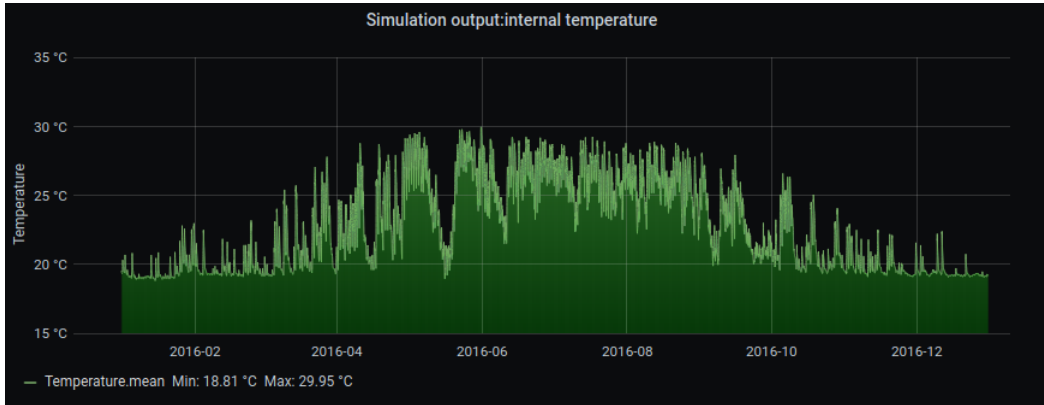

Figure 4: Simulation output: internal temperature (without overheating)

It is important to highlight the fact that the results in Figure 4 were obtained with a different weather file, with respect to the ones in Figure 3. This was done because in the forecasting (Section 4) we needed more years, in order to have sufficient data for the training. Infact the weather file has been downloaded from Weather Underground[1] using the weather station IOSLOOSL9. However the samples regarding the solar radiation were not available from this source, so information about the solar radiation has been downloaded from NREL[2]. This website allows you to select a point on the map, not a station, so the point that we selected might not be exactly the same as ISOSLOOSL9, leading to some inconsistencies in the dataset.

However, also with the current configuration, it can be observed an oscillating behaviour of the internal temperatures during the year. This can be attributed to

---

[1]http://wunderground.com

[2]http://www.nrel.gov

several factors: first, in the weather files used for the simulation, the external temperatures show a very fluctuating trend, suffering of frequent sharp collapses (for example in the month of May, going from 20°C to 4°C in about three days); then, the high value of natural ventilation in those days accentuates the phenomenon, because it can bring to a sudden cooling of the house; finally, the fact that data about solar radiation come from two different sources brings further inconsistencies.

3.2 Energy signature

As described in Annex B of the International Standard EN ISO 15603:2008, the Energy Signature is an evaluation method in which energy consumption is correlated with climatic variables aimed at representing the actual energy behaviour of the building [3]. When using the energy signature method it is common to measure the energy use in the building and plot the observations in a power versus temperature difference graph. Norway is characterized by a very varied climate: in the winters it is cold and dark for most of the day, while in the summers it is warm and predominantly sunny. This results in a large variation of energy use in building along the year. In Figure 5 the energy signature of the building with a hourly frequency is reported.
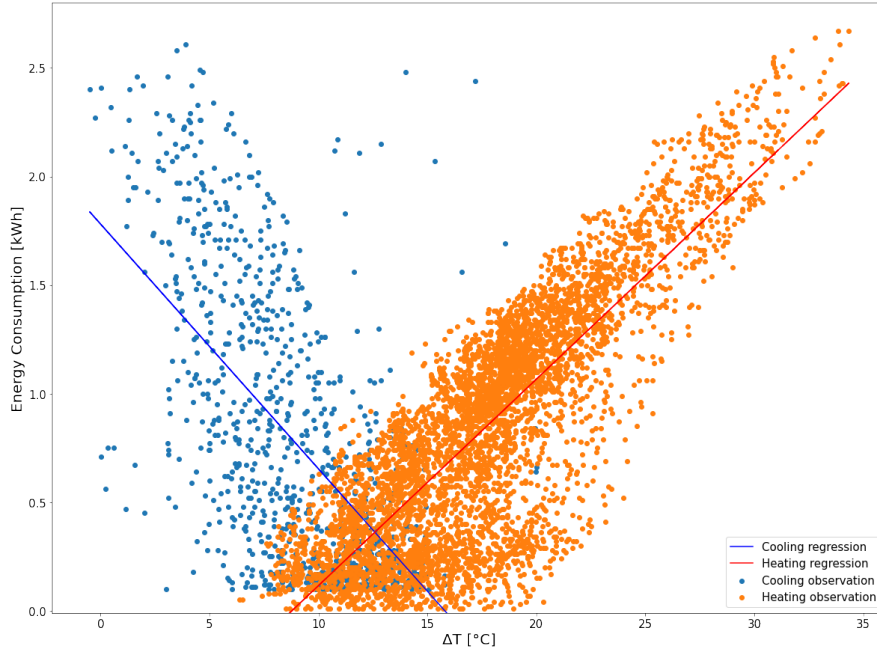

Figure 5: Energy Signature with hourly frequency

Two trends can be observed. The first one represents the energy signature of cooling that has a decreasing trend. It is easy to see that the regression line does not follow the trend of the points perfectly. This confirms that the cooling of the building is not suitable due to the high heat retained by the building. The second one is the energy signature of the heating that has an increasing trend. In this case, however, the regression line perfectly follows the trend of the points, demonstrating an excellent functioning of the heating system. This also explains the very low value of R^2 for cooling with respect to the one for heating, as show in Table 3.

In these plots, for both heating and cooling, the samples in which the value of power is equal to zero were disregarded, since in those cases the internal temperature does not depend on heating or cooling power.

	Hourly	Daily	Weekly
Heating	0.689	0.692	0.767
Cooling	0.357	0.348	0.519

Table 3: Values of R^2 for heating and cooling and for different sampling frequencies

In order to mediate any anomalous climatic condition, the energy signature has been performed also with a daily mean frequency. The result is shown in Figure 6.
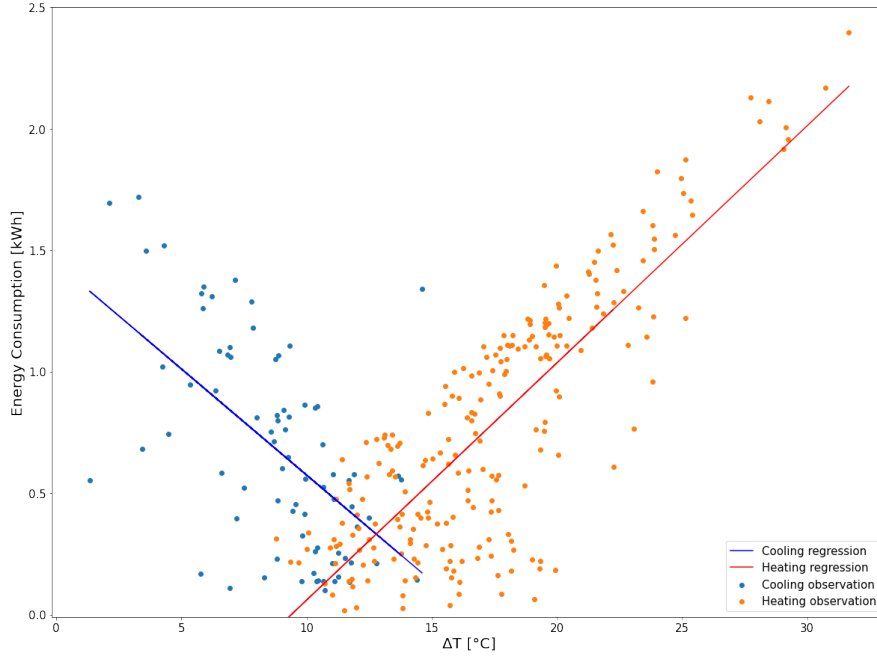

Figure 6: Energy Signature with hourly frequency

This operation wasn't enough to eliminate noise: infact the energy signature for cooling is still sparse and far from the regression line. This is further supported by the results reported in Table 3: the value of R^2 for cooling for the daily frequency is even lower than the one for the hourly frequency. Also the value of R^2 for daily frequency for heating is just slightly higher than the one for hourly frequency.

Finally the energy signature was performed with a weekly mean frequency. The result is shown in Figure 7. In the process of averaging the samples from hourly to daily and from daily to weekly, some results happened to be zero, so they were disregarded as previously. This is the reason why in the plot in Figure 7 has less data than in the dataset (46 out of 52 samples remained). The best result is obtained with a weekly frequency, as supported by the results in Table 3: with a weekly sampling frequency the value of the index R^2 increases significantly, both for heating and

8

cooling. However, even with a low frequency sampling, data are still affected by noise, due to the high variations of our dataset.
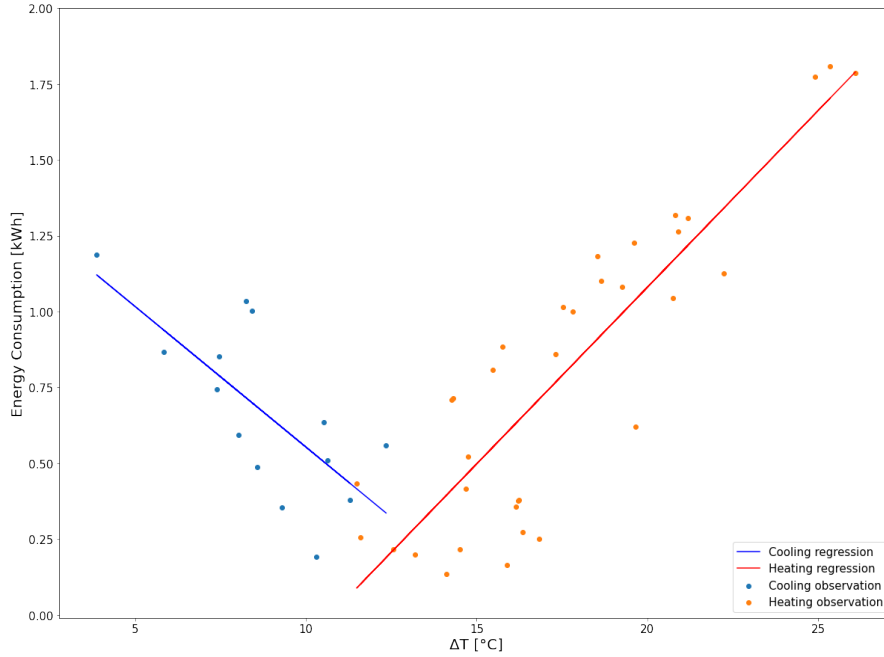

Figure 7: Energy Signature with weekly frequency.

4   Forecasting indoor temperature

The smart building concept aims to use smart technology to reduce energy consumption, as well as to improve comfort and users' satisfaction. Forecasting of the indoor temperature is necessary for the regulation of energy devices to ensure occupant comfort, as well as for energy optimization. This forecasting constitutes a complex task, because it is governed by complex physical and behavioral phenomena. It is affected by a multitude of parameters, which could be classified into three groups: outdoor conditions, building characteristics, and occupants' behavior [4].

The evaluation of internal temperature is in a state of development and often and willingly is performed with simple regression models in order to have a more complete picture of what could be the future consumption of any type of structure.

Regression models typically need to be adjusted in an "ad hoc" manner in order to capture non-linear behaviour, which arises from complex (physical) multi-variate interactions between ambient conditions, occupancy, building operating conditions and so forth [5].

Regression has always been the standard approach used to model relationship between one outcome variable and several input variables, and this can be seen both from a white-box and a black-box point of view. This means that we could use regression for analytical purposes in which a scenario is understood through physics, or for data-driven purposes in which a scenario is modeled using data alone [6]. Since the generation of data regarding our surrounding environment and ourselves is undergoing a huge growth in terms of *velocity*, *variety* and *volume*, it is clear how

9

datasets are generally too large for classical methods to have meaning. Predictive data-driven modelling now uses other means to fit models, such as machine learning, and these approaches are rapidly taking over in many disciplines.

Our proposal for data prediction is to take a machine learning approach using two different configurations of an Artificial Neural Network: a univariate configuration in which only one time series in used for forecasting, and a multivariate configuration in which more series are used to make the prediction.

4.1 Artificial Neural Networks (ANNs)

Forecasting power consumption in high energy consuming buildings requires advanced intelligent tools such as ANN. In general, ANNs are simply mathematical techniques designed to accomplish a variety of tasks. The research in the field has a history of many decades but a massive growth started in the early 1980s. Today, ANNs can be configured in various arrangements to perform a range of tasks including pattern recognition, data mining, classification, forecasting and process modelling. ANNs are composed of attributes that lead to perfect solutions in applications where we need to learn a linear or non-linear mapping [7].

Artificial neural network has found increasing consideration in forecasting theory, leading to successful applications in various forecasting domains including economic, business financial and many more. ANN can learn from examples (past data), recognize a hidden pattern in historical observations and use them to forecast future values. In addition to that, they are able to deal with incomplete information or noisy data and can be very effective especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem.

4.1.1 Multi-layer Perceptron (MLP)

ANNs consist of an inter-connection of a number of neurons. There are many varieties of connections under study, however, our treatment will analyse only one type of network, which is called the **multi-layer perceptron** (MLP). In this network, the data flows forward model used for forecasting purposes. The input nodes are the previous lagged observations, while the output provides the forecast of the future value. Hidden nodes with appropriate nonlinear transfer functions are used to process the information received by the input nodes [7]. The model can be written as:
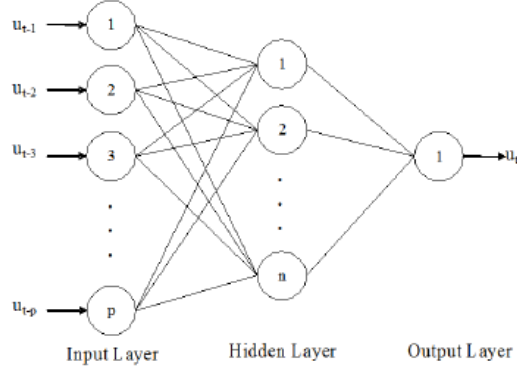
Figure 8: Topology of the MLP network.

4.2 Dataset

The dataset has been made over four years of simulation, from 2016 to 2019. The weather files of these four years have been downloaded from Weather Underground (a website that contains past weather information and that allows to download them through a python script) using the weather station IOSLOOSL9. However the samples regarding the solar radiation, as previously explained, were not available from this source, so information about the solar radiation has been downloaded from an alternative website. The simulation on E+ has been performed by applying the 4 .epw files over 4 years and the sampling of the .idf file has a frequency of 15 minutes, therefore the dataset has 4 samples per hour. This means that there is a total of 140.256 samples.

4.3 Univariate MLP

Univariate time series are a dataset comprised of a single series of observations with a temporal ordering and a model is required to learn from the series of past observations to predict the future values of the sequence.
Since the Univariate time series have a dataset based on temporal order, it is clear how, in order to create a correct predictive model, both the correct number of past values and correct number of neurons must be selected for the prediction of the future horizon.

4.3.1 Data preparation

Being the indoor temperature the goal of our forecasting, the time series that is going to be used for each room is its respective 'Zone Operative Temperature', and for the total building the 'Zone Operative Temperature' averaged over the zones, extrapolated from the dataset. Figure 9 shows the four time series of the indoor temperature used for predictions.
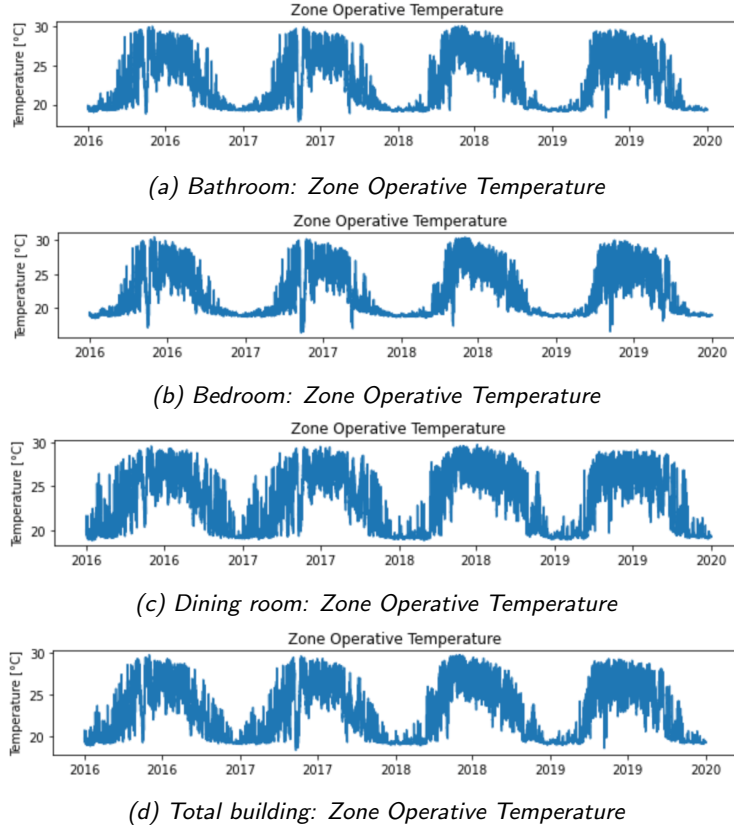
(a) Bathroom: Zone Operative Temperature

(b) Bedroom: Zone Operative Temperature

(c) Dining room: Zone Operative Temperature

(d) Total building: Zone Operative Temperature

Figure 9: Zone operative temperature for each room and total building

A time series is a sequence of numbers that are ordered by a time index. This can be thought of as a list or column of ordered values. Before a Univariate time series can be modeled, it must be prepared since the majority of practical machine learning uses supervised learning. The concept behind supervised learning is the following: $Y = f(X)$, where X are the input variables and Y the output variables. At this point an algorithm will be used to learn the mapping function from the input to the output. This concept can be applied to our MLP model, where it will learn a function that maps a sequence of past observations as input to an output observation. As such, the sequence of observations must be transformed into multiple examples from which the model can learn.

The *shift()* method is used in order to create a function that enables to transform our database to be written in a way that the future 5 hours are function of past values. Therefore, the dataset in use is in the following form: $T_{in}(t-n), ..., T_{in}(t-1)$ as input parameters and $T_{in}(t), ..., T_{in}(t+19)$ as output values, where 19 represents the 19th quarter of hour (300 minutes, hence 5 hours), where n represents the optimal lag, the maximum past horizon that gives the lowest prediction error.

4.3.2   Finding the optimal lag order

One of the most important thing in time series problems is choosing the correct number of past values that are going to be used for the forecasting. A prediction with a low number of past values will be able to predict the near future with a very

low error but with a high error instants of advanced future; whereas, a prediction that takes into account a large number of past values will tend to predict the near future with more difficulty and the distant future in a better way.

Using the Mean Absolute Error (MAE) as a performance indicator, the best past value is given by the one that will minimize the average MAE of our predictions (from the next quarter of an hour to 5 hours later).

The final optimal lag is found using the following algorithm:

Algorithm 1 Optimal lag order

1: Define the list that contains the prediction errors: MAE_list
2: Define the window of the number of past values to be considered: $window =$ range(1,20)
3: **for** lag in $window$ **do**
4:     Transform dataset in supervised learning form, with number of previous lag $= lag$
5:     Divide the dataset in training and test sets
6:     Fit MLP to the test set
7:     Prediction using the test set
8:     Append the value of the MAE to the error list MAE_list
9: **end for**
10: $optimal\_lag = \text{argmin}(MAE\_list)$

The ANN used for the prediction is MLP, with a number of input values that varies according with the lag order, one hidden layer with 30 neurons with a 'sigmoid' activation function, and an output layer. The network uses the *Adam* optimizer with a learning rate equal to 0.01. By performing the previous algorithm, we found out that the optimal number of lag for the three rooms and the total building is always 7. This could be due to the internal temperature having almost the same behaviour for all the zones in study.

4.3.3 Finding the optimal number of neurons

When designing a MLP, choosing the correct number of neurons is important: a multilayer perceptron with few neurons with respect to the number of inputs does not learn properly the mapping between inputs and outputs, whereas if the number of neurons is large, the network may overfit. That is why the MLP used in the previous section must be tuned before been used for the final prediction. The optimal number of neurons in this case is found using a grid search: for each number of neurons in the interval [1,10] a network is created and fit. The network with the lowest error is kept. It has been found that the optimal number of neurons (of the network that has 7 as optimal lag value) is 3. Figure 10 shows the final optimized design of the MLP network.
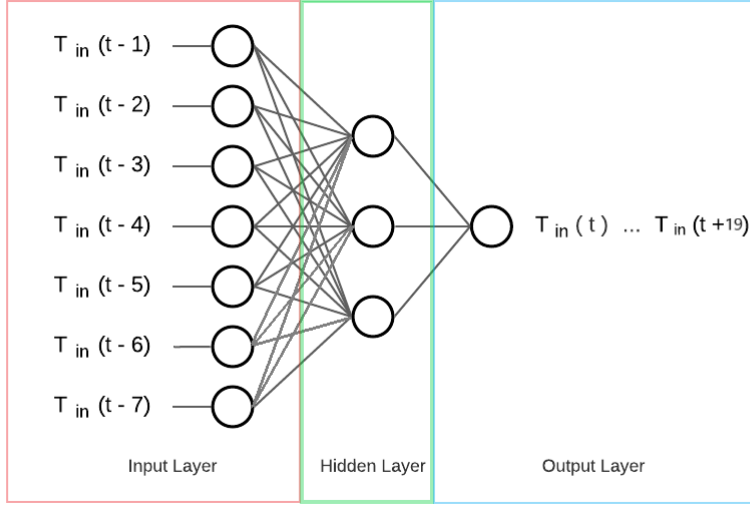
*Figure 10: Univariate MLP network topology.*

4.3.4 Results

By applying the methodology described above, predictions have been performed with a k-step forecasting, where k is the number of future time samples, that goes from 1 to 20 (i.e. 300 minutes). The indicators chosen to evaluate the performance of the network are: MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error). We identified MAE = 0.750 as the maximum acceptable threshold. This threshold represents the maximum limit within which indoor temperature changes do not impact on environmental thermal comfort perceived by the occupants [8]. Table 4 reports the results of the performance indicators for the three rooms and for the whole building. Highlighted in bold there are the values of k that represent the maximum threshold for the steps of the forecast. The results show that the maximum number of future minutes that can be predicted respecting the threshold are: 225 minutes for the Bathroom, 210 minutes for the Bedroom, 180 minutes for the Dining room, 195 minutes for the whole building. The trend of the error is increasing when the k increases, this means that a larger number of future time steps to predict implies a bigger error (both MAE and RMSE).

14

| Prediction Steps	Time [min]	Bathroom		Bedroom		Dining room		Total Building	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
k = 1	15	0.09	0.158	0.098	0.168	0.136	0.213	0.111	0.185
k = 2	30	0.114	0.199	0.124	0.212	0.2	0.302	0.133	0.238
k = 3	45	0.137	0.249	0.148	0.279	0.369	0.394	0.169	0.286
k = 4	60	0.171	0.315	0.184	0.345	0.322	0.467	0.195	0.331
k = 5	75	0.205	0.384	0.223	0.421	0.344	0.512	0.238	0.407
k = 6	90	0.249	0.462	0.261	0.49	0.372	0.568	0.280	0.477
k = 7	105	0.276	0.515	0.304	0.565	0.42	0.638	0.318	0.54
k = 8	120	0.312	0.583	0.360	0.655	0.456	0.7	0.368	0.617
k = 9	135	0.354	0.655	0.405	0.731	0.507	0.781	0.411	0.723
k = 10	150	0.390	0.721	0.472	0.833	0.574	0.874	0.511	0.818
k = 11	165	0.434	0.797	0.54	0.935	0.647	0.974	0.561	0.893
k = 12	180	0.489	0.880	0.603	1.033	**0.731**	**1.085**	0.629	0.989
k = 13	195	0.573	0.993	0.697	1.153	0.804	1.188	**0.701**	**1.09**
k = 14	210	0.645	1.093	**0.739**	**1.230**	0.909	1.321	0.827	1.232
k = 15	225	**0.701**	**1.177**	0.768	1.292	0.997	1.427	0.896	1.344
k = 16	240	0.775	1.273	0.852	1.399	1.085	1.551	0.975	1.44
k = 17	255	0.815	1.344	0.874	1.445	1.154	1.645	1.305	1.529
k = 18	270	0.862	1.417	0.92	1.533	1.206	1.725	1.108	1.628
k = 19	275	0.909	1.489	0.975	1.618	1.32	1.865	1.179	1.727
k = 20	300	1.024	1.623	1.053	1.724	1.48	2.036	1.249	1.827

Table 4: Univariate MLP performance indicators
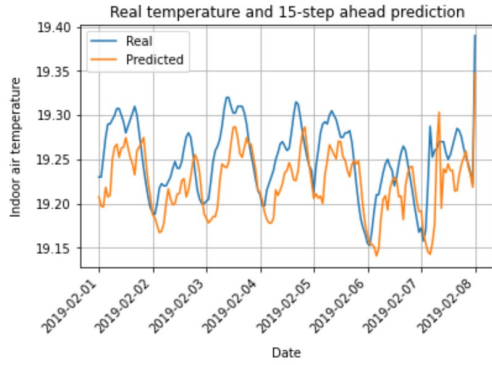
(a) Bathroom: Prediction for k = 15 (225 min)
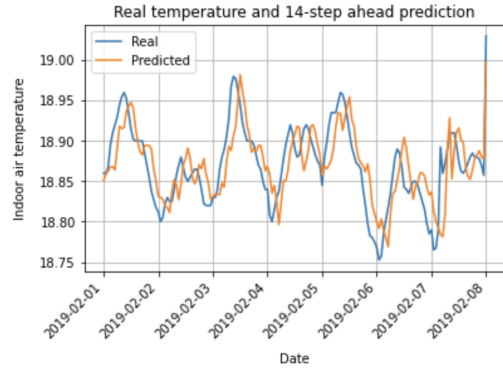
(b) Bedroom: Prediction for k = 14 (210 min)
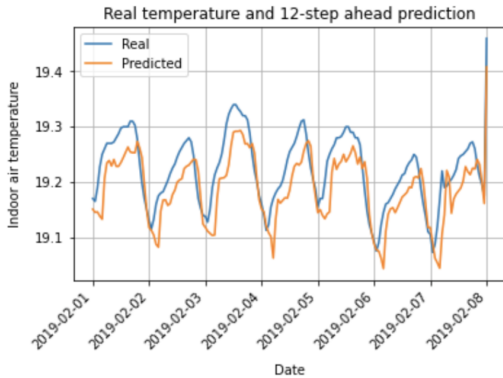
(c) Dining room: Prediction for k = 12 (180 min)
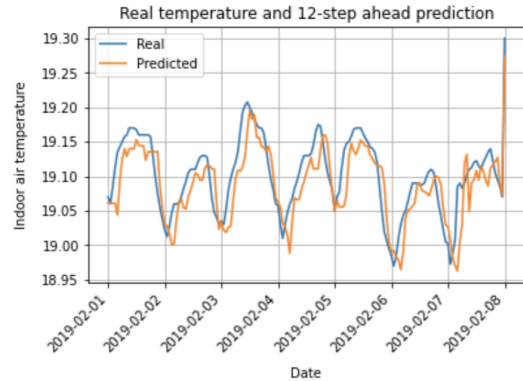
(d) Total building: Prediction for k = 13 (195 min)

Figure 11: Final prediction for the three rooms and total building.

The plots in Figure 11 show the comparison between indoor temperature predicted by the neural network, for different k, and the realistic values obtained with E+, referred to the first week of February 2019.

4.4 Multivariate MLP

The Multivariate MLP differs from the Univariate MLP since at the entrance of the neural network we do not have a single time series but more than one observation can be found at each time step. This is the only major difference that can be emphasized between the two neural networks as they have the same topology.

4.4.1 Data preparation

The time series extracted from the simulation output that will be used for the prediction are the following:

- Zone Operative Temperature [C]

- Zone Windows Total Transmitted Solar Radiation Rate [W]

- Site Outdoor Air Drybulb Temperature [C]

- DistrictHeating:Facility [J]

- DistrictCooling:Facility [J]

Data about internal temperature and solar radiation are divided by zone, while data about the power cooling and power heating are relative to the whole building. In order to perform predictions on the three separated zones, data about power were weighted over the surface of each zone, obtaining thus time series of internal temperature, solar radiation, power cooling and power heating for each of the three zones. Moreover, in order to perform predictions on the whole building, data about internal temperature and solar radiation were averaged (with respect to the surfaces of each zone), so that time series of internal temperature, solar radiation, power cooling and power heating are obtained also for the entire building. The time series of the external temperature is also kept into consideration for the predictions. Figure 12 shows the time series relative to the whole building used for prediction. The other times series sets used for the three rooms have a similar trend to the one shown. Finally, as for the Univariate case, time series must be prepared and transformed into supervised learning, using the algorithm explained in section 4.3.1.
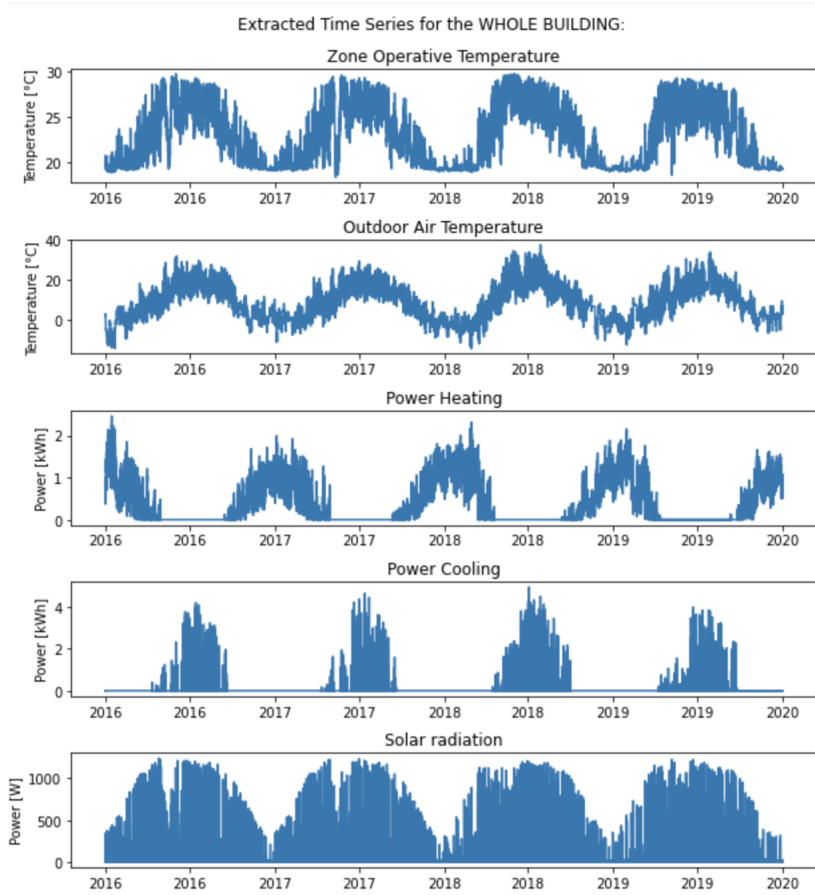
Figure 12: Time series used for predictions relative to the whole building

4.4.2 Finding the optimal lag order

Using the Mean Absolute Error (MAE) as a performance indicator, the optimal lag order will be given by the one that will minimize the average MAE of our predictions (from the next quarter of an hour to 5 hours later). The network uses the *Adam* optimizer with a learning rate equal to 0.01. By performing Algorithm 1, it was found that the optimal lag order for each time series given as input to the network (for each room and for the total building) is always 2.

4.4.3 Finding the optimal neural numbers

Also in this case the optimal number of neurons in the hidden layer has been found using a grid search. It has been found that the optimal number of neurons (of the network that has 2 as optimal lag value) is 7. Figure 13 shows the optimized design of the network used for the forecasting.

17
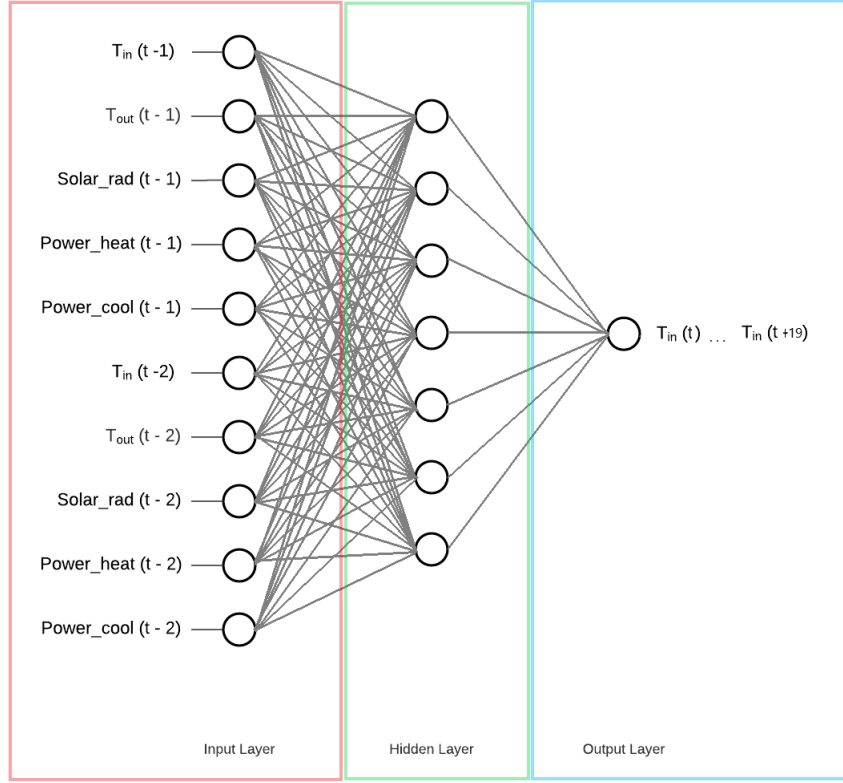
Figure 13: Multivariate MLP network topology.

4.4.4 Results

The methodology applied for the Multivariate MLP is the same used for the Univariate MLP previously explained in section 4.3, thus carrying out a prediction with a k-steps forecasting, taking both MAE and RMSE as performance indicators and setting the MAE threshold equal to 0.75 as the maximum acceptable limit.

Table 5 reports the results of the performance indicators for the three rooms and for the whole building. As done before, it was highlighted the k values representing the maximum threshold for the steps of the forecast.

Prediction Steps	Time [min]	Bathroom		Bedroom		Dining room		Total Building	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
k = 1	15	0.117	0.212	0.122	0.199	0.117	0.192	0.114	0.185
k = 2	30	0.153	0.264	0.129	0.217	0.148	0.234	0.125	0.211
k = 3	45	0.194	0.315	0.18	0.282	0.16	0.279	0.172	0.276
k = 4	60	0.250	0.387	0.189	0.305	0.187	0.316	0.177	0.296
k = 5	75	0.279	0.422	0.233	0.354	0.231	0.374	0.193	0.329
k = 6	90	0.318	0.498	0.225	0.382	0.285	0.45	0.214	0.367
k = 7	105	0.329	0.54	0.254	0.434	0.315	0.506	0.248	0.426
k = 8	120	0.391	0.625	0.311	0.515	0.366	0.582	0.298	0.503
k = 9	135	0.437	0.7	0.352	0.584	0.422	0.659	0.336	0.567
k = 10	150	0.477	0.772	0.385	0.644	0.466	0.729	0.378	0.63
k = 11	165	0.519	0.845	0.439	0.726	0.54	0.831	0.433	0.712
k = 12	180	0.573	0.929	0.488	0.8	0.612	0.931	0.487	0.794
k = 13	195	0.642	1.024	0.543	0.811	0.694	1.042	0.54	0.869
k = 14	210	0.662	1.089	0.604	0.967	**0.745**	**1.16**	0.605	0.965
k = 15	225	**0.696**	**1.153**	0.666	1.058	0.834	1.278	0.667	1.055
k = 16	240	0.77	1.236	**0.728**	**1.145**	0.971	1.401	**0.726**	**1.136**
k = 17	255	0.866	1.383	0.815	1.264	1.068	1.526	0.812	1.254
k = 18	270	0.947	1.496	0.909	1.393	1.17	1.661	0.807	1.389
k = 19	275	1.029	1.608	0.989	1.501	1.339	1.852	0.983	1.493
k = 20	300	1.1	1.706	1.066	1.612	1.374	1.917	1.062	1.606

Table 5: Multivariate MLP performance indicators.

(a) Bathroom: Prediction for k = 15 (225 min)

(b) Bedroom: Prediction for k = 13 (195 min)

(c) Dining room: Prediction for k = 13 (195 min)
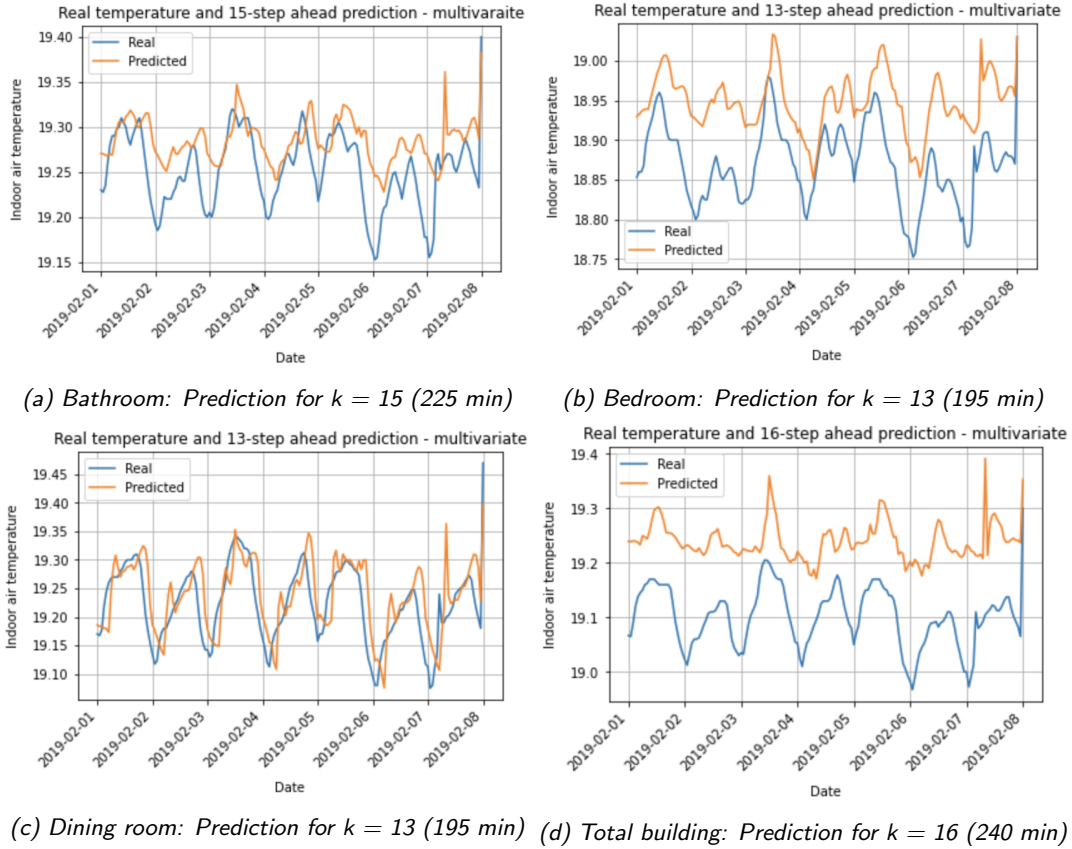

(d) Total building: Prediction for k = 16 (240 min)

Figure 14: Final predictions for the three rooms and total building.

The results show that the maximum number of future minutes that can be predicted respecting the threshold are: 225 minutes for the Bathroom, 240 minutes for the Bedroom, 210 minutes for the Dining room, 240 minutes for the whole building. Even in this case, the trend of the error is increasing when the k increases, this means that a larger number of future time steps to predict implies a bigger error (both MAE and RMSE).

Finally, plots in Figure 14 show the comparison between indoor temperature predicted by the neural network, for different k, and the realistic values obtained with E+, referred to the first week of February 2019.

5   Conclusions

The results obtained from the optimization of the features showed a major problem from the point of view of the comfort of the user: the overheating phenomenon. This problem has been solved by adjusting the schedule of heating and cooling systems and the temperatures setback. However even with these corrections the trend of the internal temperatures remained unstable, due to: the noise introduced by the solar radiation, that comes from a different station with respect to weather data; the high value of ACH that brings sudden cool-downs; the high insulation level of the entire building, that withholds heat and requires a high power cooling consumption to cool it down. Moreover, from the point of view of a possible realization of the building, there are several issues. For instance, the thickness of the insulation layer and the triple glazing are very expensive features, that would make the realization unfeasible. All of these components are due to the fact that the objective of this case study, was to optimize the total power consumption, not the comfort of the user. As a consequence we obtained a building that has a low correlation between power cooling and internal temperature. This means that a change in the power cooling, does not bring to an immediate response of the internal temperature. Which explains the very low values of R^2 in the energy signature.

In the forecasting section our goal was to predict the indoor air temperature (of the three rooms and the total building) of one year, using three years of history, with two different MLP models, hence a comparison between the two must be done.

The univariate MLP, which requires only the internal indoor air temperature as only input time series, gives satisfying prediction up to almost 4 hours for each room and the total building, with 7 past values. On the other hand, the multivariate MLP, gives the same results (or even better) with only 2 past values: this is because, having just a single knowledge, the univariate MLP needs more past trend values to correctly predict the future. Figure 11 highlights another interesting insights of the univariate MLP: the forecasted values are delayed with respect to the real ones, so the network behaves like a moving average prediction method with a window size equal to the number of optimal lag values. In order to avoid this phenomena, other kind of artificial neural networks can be used that take into account recurrency. An example are LSTM, or Long-Short-Term Memory Recurrent Neural Networks, that unlike the feedforward networks where the signals travel in the forward direction only, the data signals travel in backward directions as well as these networks have the feedback connections, enabling the network to learn deep connection between the past values.

Finally, thanks to the prediction models in the smart building field, Model Predictive Control frameworks can be used, since they outstand among other conventional methods applicable for the building control design. The basic idea of MPC is to exploit a model of the process to predict the future evolution of the system and to compute control actions by optimizing a cost function depending on these predictions (for example, reducing power consumption). However, the building dynamics are slow, and subjected to different disturbances (both internal and external), making the use of univariate MLP almost useless, since it would not be able to capture the effects of the variations of the lasts inside the building, and preferring a multivariate approach where all the possible key factors influencing the behavior of the building are considered.

References

[1] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information", *Energy and buildings*, vol. 40, no. 3, pp. 394-398, 2008.

[2] "Monthly energy review", *U.S. Energy Information Administration*, September, 2017.

[3] Gustav Nordström, *Use of energy-signature method to estimate energy performance in single-family buildings*, Luleå University of Technology, Graphic Production 2014.

[4] Huebner, G.M.; McMichael, M.; Shipworth, D.; Shipworth, M.; Durand-Daubin, M.; Summerfield, A. The reality of English living rooms, a comparison of internal temperatures against common model assumptions. Energy Build. 2013, 66, 688–696.

[5] Y. Heo and V. M. Zalava, "Gaussian process modelling for measurement and verification of building energy savings", *Energy and Buildings*, vol. 53, pp. 7-18, 2012.

[6] Aurora Gonzàlez-Vidal, A.P. Ramallo-Gonzàlez, F. Terroso-Sàenz and Antonio Skarmeta, "Data driven modeling for energy consumption prediction in smart buildings", IEEE International Conference on Big Data, 2017.

[7] A. Azadeh, S.F. Ghaderi, S. Sohrabkhani, "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors".

[8] Alessandro Aliberti, Francesca Maria Ugliotti, Lorenzo Bottaccioli, Giansalvo Cirrincione, Anna Osello, Enrico Macii, Edoardo Patti and Andrea Acquaviva, *Indoor Air-Temperature Forecast for Energy-Efficient Management in Smart Buildings*, Researchgate.