



Modelling thermal comfort in residential buildings in Portugal

Marek Utracki

Thesis to obtain Master of Science Degree in
Energy Engineering and Management

Supervisor: Prof. Carlos Augusto Santos Silva
Dr. Ricardo Manuel Anacleto Gomes

Examination Committee

Chairperson: Prof. Jorge de Saldanha Gonçalves Matos
Supervisor: Dr. Ricardo Manuel Anacleto Gomes
Member of the Committee: Dr. Diana Pereira Neves

November 2021

I declare that this document is an original work of my own authorship and that it fulfils
all the requirements of the Code of Conduct and Good Practices of the
Universidade de Lisboa.

Abstract

For several years it has been possible to observe a strong correlation between the mortality rate and Extreme Meteorological Events (EWE) in Portugal. This relationship results from low levels of thermal comfort in buildings, a result not only of the building's constructive characteristics, but also of the meteorological conditions where it is located. The problem of energy poverty is further aggravated by the aging of society and low average income, which translates into insufficient levels of air conditioning. Some attempts to predict EWE have been made, but considering either heat waves or cold waves. The RELIABLE project aims to create a model for forecasting EWE for the whole year. Thus, within the scope of this thesis, building energy simulation models are compared in terms of reliability of results and calculation times. This research work compared Energy Plus, the 5R1C model and optimization tools, in particular Random Forest and Neural Networks. The models considered the thermal characteristics of nine existing building archetypes in Montalegre, Portugal. Simulation results were compared with real temperature measurements. This work shows that the simplified 5R1C model serves the purpose of forecasting EWE effect on building indoor temperature, due to the short computation time and its suitability for large buildings.

Keywords

thermal comfort, building simulation, simplified models, extreme weather events, energy poverty

Resumo

Há vários anos que é possível constatar uma forte correlação entre a taxa de mortalidade e Eventos Meteorológicos Extremos (EME) em Portugal. Esta relação resulta de baixos níveis de conforto térmico em edifícios, fruto não apenas das características construtivas do edifício, mas também das condições meteorológicas onde este se encontra. O problema da pobreza energética é agravado ainda pelo envelhecimento da sociedade e baixo rendimento médio, que se traduz em níveis de climatização insuficientes. Algumas tentativas de prever EME foram realizadas, mas considerando ou ondas de calor ou vagas de frio. O projeto RELIABLE tem como objetivo criar um modelo para previsão de EME para todo o ano. Assim, no âmbito desta tese comparam-se os modelos de simulação energética em edifícios em termos de confiabilidade de resultados e tempos de cálculo. A pesquisa compara o Energy Plus, o modelo 5R1C e ferramentas de otimização, em particular Random Forest e Redes neuronais. Os modelos consideram as características térmicas de nove arquétipos de edifícios existentes em Montalegre, Portugal. Os resultados das simulações foram comparados com medições reais de temperatura. Este trabalho mostra que o modelo simplificado 5R1C serve o propósito de previsão do efeito de EME na temperatura interior dos edifícios, devido ao curto tempo de computação e sua adequação para grandes parques edificados.

Palavras-chave

Conforto térmico, simulação em edifícios, modelos simplificados, eventos meteorológicos extremos, pobreza energética

Table of Contents

1.	Introduction.....	15
1.1.	Context / Overview	16
1.2.	State of the art	18
1.3.	RELIABLE project.....	18
1.4.	Aim of this work.....	19
2.	Theoretical Background.....	20
2.1.	5R1C model	21
2.2.	Energy Plus	23
2.3.	Machine Learning methods	24
2.3.1.	Random Forest Regression	24
2.3.2.	Artificial Neural Networks	25
3.	Methodology	28
3.1.	Montalegre case study	29
3.2.	Energy Performance Certificates into models	29
3.3.	Monitoring Campaign	30
3.4.	The basic model.....	30
3.4.1.	5R1C.....	33
3.4.2.	Sensitivity Analysis	33
3.5.	Data input procedure in models.....	35
3.6.	Model calibration	35
3.7.	Machine Learning model: Random Forest.....	37
3.7.1.	Feature selection	37
3.8.	Artificial Neural Network.....	38
4.	Results.....	40
4.1.	Discussion about the results.....	41
4.2.	Energy Plus	42
4.3.	5R1C	43
4.4.	ML Random Forest	47

4.5.	Artificial Neural Network.....	48
4.6.	Model Calibration results	48
4.7.	Computing time	50
5.	Conclusions and recommendations	51
5.1.	Conclusions	52
5.2.	Recommendations	52
5.3.	Limitations of this work	53
6.	Appendix	58

List of Figures

Figure 1	Average consumption per dwelling (at normal climate)	17
Figure 2	Figure 2 Inability to keep home adequately warm in Europe in 2012 based on Eurostat data	17
Figure 3	5R1C model represented as electric scheme	22
Figure 4	EnergyPlus internal elements / modules	23
Figure 5	Generic scheme of Random Forest working principle	25
Figure 6	Structure of Artificial Neural Network	26
Figure 7	Construction of an artificial neuron	26
Figure 8	Activation functions: Sigmoid and ReLU	27
Figure 9	Indoor temperature and relative humidity sensor by HOBO®	30
Figure 10	One of the cubes designed in Google SketchUp.	32
Figure 11	Result from 5R1C model after first test run with comparison to HOBO measurement	33
Figure 12	Sensitivity analysis chart. One dot defines one simulation run, one line presents how the output temperature has been changing with the change of given parameter. Lines are not correlated among each other and should be considered separately.	34
Figure 13	Random Forest training result	38
Figure 14	Models comparison with HOBO measurement and outside temperature	42
Figure 15	Energy Plus results for one week in May	43
Figure 16	5R1C result graph for one week in May before fine tuning the model with infiltration values	45
Figure 17	5R1C results after introduction of infiltration value.	45
Figure 18	5R1C results for the first 7 days of February	46
Figure 19	Energy Plus results for first 7 days of February	46
Figure 20	ML Random Forest results for 7 days in May with comparison to HOBO	47
Figure 21	ML Random Forest results for 7 days in February	48

Figure 22	Sensitivity Analysis - Envelope Size	60
Figure 23	Sensitivity Analysis – Infiltration	60
Figure 24	Sensitivity Analysis - Total Internal Area and Thermal Capacity	61
Figure 25	Sensitivity Analysis - Thermal Capacity	61
Figure 26	Sensitivity Analysis - U wall	62
Figure 27	Sensitivity Analysis - U windows	62
Figure 28	Sensitivity Analysis - Ventilation Efficiency	63
Figure 29	Sensitivity Analysis - Ventilation	63

List of Tables

Table 1	Population structure in Montalegre	29
Table 2	Monitored parameters, equipment used and resolution of measurement	30
Table 3	Archetype's description derived from Energy Certificates	31
Table 4	Flow chart of methodology used in research	32
Table 5	Acceptable tolerances defined by ASHRAE Guidelines	36
Table 6	MBE and Cv(RMSE) scores of comparison of model with real measurement	49
Table 7	MBE and Cv(RMSE) scores of comparison EnergyPLUS and 5R1C models	49
Table 8	Computing times	50
Table 9	Input data for 5R1C model (archetypes description)	59

List of Acronyms

EWE	Extreme Weather Events
5R1C / RC	5 resistances 1 capacitance model
ML	Machine Learning
RF	Random Forest
ANN	Artificial Neural Network
DOE	United States Department of Energy
TFM	Transfer Method Function
GUI	Graphical User Interface
NREL	National Renewable Energy Laboratory
EPC	Energy Performance Certificate
MBE	Mean Bias Error
Cv(RMSE)	Coefficient of variation of the Root Mean Square Error
RMSD	Root Mean Square Deviation

List of Symbols

C_m	Thermal capacity (J/m ² K)
$H_{tr,w}$	Thermal transmission coefficient for walls with windows (W/K)
$H_{tr,op}$	Thermal transmission coefficient for opaque walls (W/K)
$H_{tr,em}$	Effective thermal mass (W/K)
$H_{tr,ms}$	Surface thermal mass (W/K)
H_{ve}	Thermal transmission coefficient due ventilation and infiltration losses (W/K)
$H_{tr,is}$	Thermal conductance (W/K)
θ_{air}	Air temperature (°C)
θ_e	Outdoor air temperature (°C)
θ_m	Mass temperature (°C)
θ_s	Surface temperature (°C)
θ_{sup}	Temperature of ventilation air stream (°C)
$\Phi_{HC,nd}$	Heating or cooling power (W)
Φ_{int}	Internal gains from appliances and human presence (W)
Φ_{sol}	Incident solar radiation on external surface of the envelope (W)

Chapter 1

Introduction

In this chapter context of the work is described. Next state of the art and what can be done regarding the issue described in context. Following with the description of RELIABLE project and the aim of this work.

1.1. Context / Overview

For several years now a correlation between death rate and temperature in Portugal is observed. The peak death rates can be observed in winter and summer periods even though Portugal has one of the mildest winter climates [8] [9]. Portugal is among countries with the highest mortality rates in Europe for both the summer and winter periods. The situation in which people are more likely to die in period of cold weather in countries with winter is rather mild is called “excess winter mortality paradox”. For example the ratio between Excess Winter Deaths and Heating Degree Days for Portugal is 5.7. This ratio is lower for colder European countries like Finland (4.0), Estonia (3.9) or Denmark (4.8) [10] [11]. The older buildings in Portugal (before 1990) have no thermal insulation. Furthermore, heatwaves and cold waves and the increased mortality can be explained by the fact that the buildings do not meet minimum thermal requirements or have no cooling systems [12]. This problem is being swelled by the ageing society and low average income. Elderly people have less ability to handle inadequate indoor temperatures which is the effect of fuel poverty. Although, increased mortality is connected to the age of people at this stage of the project only the dwelling is considered not the dwellers nor their age. For this is the next, future stage. There is an unofficial definition of fuel poverty: “anyone who meets in its housing particular difficulties to have necessary energy to meet its basic energy needs because of the inadequacy of its resources or of its housing conditions” [13] [14]. Such a situation falls within the scope of energy poverty. Energy poverty has no universal definition since energy poverty is displayed differently in different countries. As most European countries have no official definition for the term “energy poverty”, this state is often described as “inability to keep homes adequately warm” [15] (Figure 2). It has been proven that such situation can affect the children indirectly when spending on food are reduced to afford fuel to keep the house warm. Therefore, energy poverty can be correlated with low income and high energy costs. Portugal belongs to the group of European countries that have high inability to keep homes adequately warm according to *Alleviating fuel poverty in the EU*. [2] (Figure 2). Heating consumption in Portugal decreased over the years 2010-2015 (economic crisis and taxation followed by increase of prices).

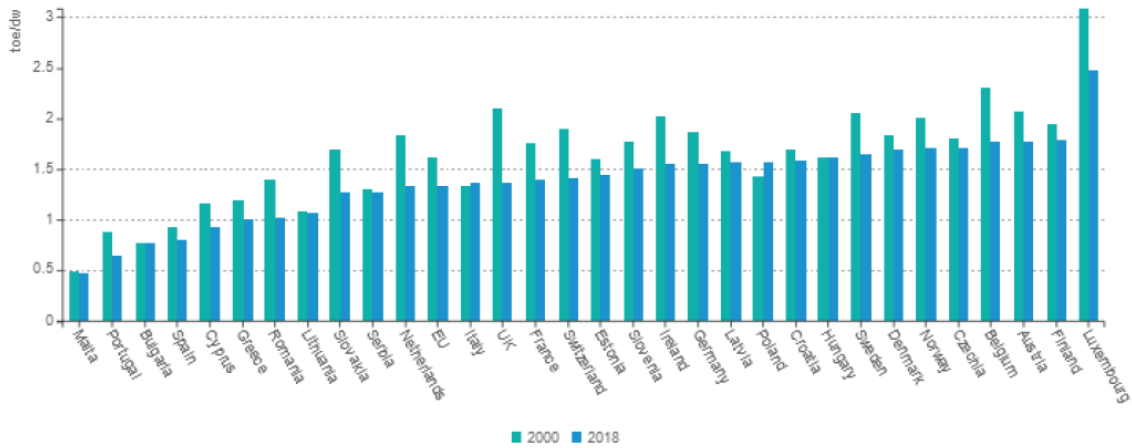


Figure 1 Average consumption per dwelling (at normal climate) [1]

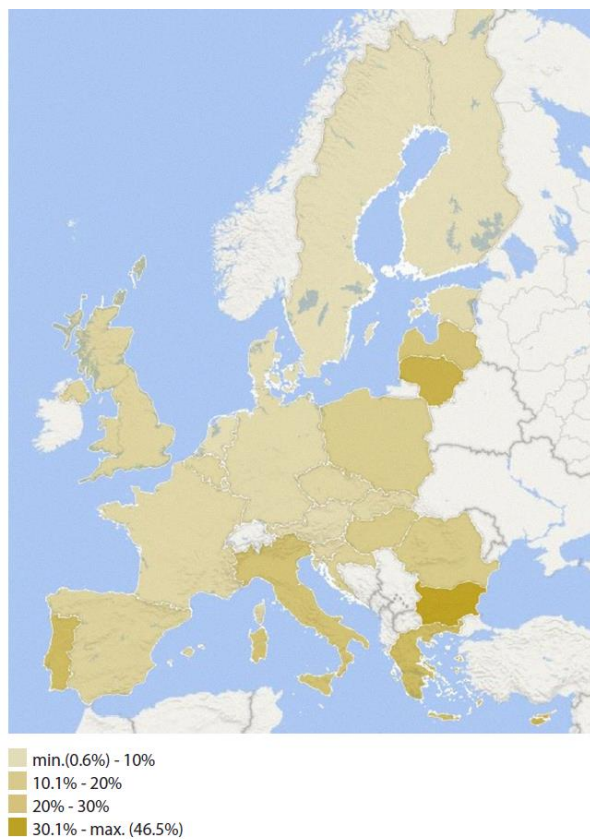


Figure 2 Inability to keep home adequately warm in Europe in 2012 based on Eurostat data [2]

Portugal when compared with neighbouring countries – for example Spain – has considerably smaller energy consumption per dwelling (Figure 1) even though both countries belong to the same climatic area. This brings a conclusion and confirms that the problem is connected with low average income. Portugal has one of the most unequal income distribution. According to Gini index, which represents the degree of inequality in wealth, Portugal with score 0.341 is 2.6 percentage points above OECD average of 0.315 [16]. Such situation results in less energy consumption and lower appetite for investments in building retrofitting.

1.2. State of the art

From the reasons presented in the previous section stems the need for a quick and nationwide solution. A solution that is cheap but does not require physical interference with the building. Over the past few years, two projects were developed in Portugal: ÍCARO and FRIESA. Project ÍCARO, developed by INSARJ, is a nationwide health warning system which detects and monitors heatwaves and determines if the forecasted number of deaths is going to exceed the expected number of deaths [17]. Initially it was created based on time series statistical model using dynamic regression and thresholds which were calibrated for Lisbon. The calibration was done concerning 1981-1991 heatwaves. Every day a report is made with the forecast for the next three days. The project is run only in the period between May and September and thus it is used only for heatwaves prediction. Its downside is the very low resolution because it divides Portugal into 5 regions. Recently there was an incentive made to update the system and upgrade it with distributed lag non-linear model (DLNM) [18]. This upgrade would allow to analyse the relation of exposure to heat and the response effect (increased mortality). Together with this upgrade the calibration for newest data would be done. FRIESA is similar project to ICARO which started in 2014. It estimates and analyses the risk of cold waves and its potential impact on mortality [19]. It is used in the period from November to March. Currently, FRIESA covers only two districts: Porto and Lisbon. Both of those projects aims to provide decision makers with appropriate tool for handling the hazardous events associated with cold and hot on the population.

1.3. RELIABLE project

The projects briefly described above have some major flaws. Namely very low resolution and the need to use two different applications for different parts of the year. Also, these applications show only outside temperature or the predicted excess of deaths. Project RELIABLE has been started to develop a tool – real-time internet application – which will provide reliable data about indoor temperature for different types of buildings across Portugal [20]. The dashboard will be a map with forecasted data of very high resolution. It will determine zones of dangerously high indoor temperatures in the summer and dangerously low indoor temperatures in the winter. The biggest advantages of this project are public availability through a websites of different public authorities (ADENE, INSA, municipal services) all year long and transparency of the data . The development of such an application will help local services more effectively focus resources in zones of concern and monitor the well-being of the citizens in danger zones.

Such an ambitious enterprise requires a lot of computational power. Simulation for the whole of Portugal requires weather data from hundreds of points. The administrative division of Portugal is 18 districts which are divided into municipalities. 308 municipalities are further divided into 3092 parishes.

Simulation for 2 weeks ahead for every hour for 9 archetypes for the 308 municipalities, assuming 1second for one iteration, would require 931392 iterations. A such number of iterations - roughly calculating - would take almost 11 days. Therefore, there is a need for a faster but reliable model

1.4. Aim of this work

This work aims to compare methods for indoor temperature predictions. Build a 5R1C model, test Machine Learning (ML) models, in particular Random Forest (RF) and Artificial Neural Networks (ANN) capabilities in this area. The models will be created based on energy certification data from buildings from different periods of construction. The results from the 5R1C, RF and ANN models will be then compared with results from EnergyPlus software and with real data. In theory, the 5R1C model should provide a faster computation of the results. If the accuracy of 5R1C, ML or ANN is reasonable and the computations times faster or comparable with Energy Plus, one of these models will be used for further proceedings in the RELIABLE project. The output from investigated models is not expected to be exact as the real measurement. The objective is to get a general model of an archetype that reflects the thermal behaviour of an envelope.

Chapter 2

Theoretical Background

This chapter describes the three models used during research (5R1C, Random Forest and Artificial Neural Network). It introduces the theory on which these models are based and gives description of the areas in which they are currently employed.

2.1. 5R1C model

The model described in international standard EN ISO 52016-1:2017 consisting of five resistances and one capacity (5R1C) (Figure 3) is proposed as a simplified way for simulation of thermal behaviour of the building zone given the external weather data. The envelope is described as an equivalent electric circuit consisting of five resistances and one capacity which solution requires employment of limited input parameters. Roberto Bruno et al. investigated the performance of 5R1C model for multi-house building and concluded that over larger periods of time the fluctuations and time shift between simplified model and real data are too high especially in houses with high thermal capacities [21]. Case described in this paper is different since data from dozens of buildings are averaged which already attenuates the possible fluctuations but also introduces some degree of inaccuracy.

The main variables employed in the model are:

- thermal transmission coefficients for walls with windows $H_{tr,w}$; for opaque walls $H_{tr,op}$. which is decoupled into $H_{tr,em}$ - the effective thermal mass which accumulates thermal energy - and $H_{tr,ms}$ - surface thermal mass;
- thermal transmission coefficient due to ventilation and infiltration losses H_{ve}
- thermal conductance $H_{tr,is}$ for coupling between surface temperature node θ_s and indoor temperature node θ_{air} ;
- thermal capacity per floor area of the envelope C_m ;

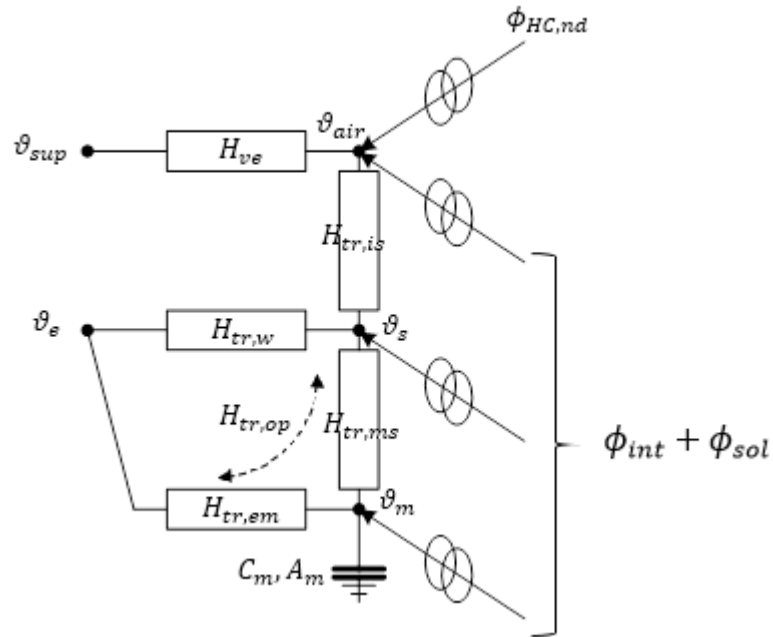


Figure 3 5R1C model represented as electric scheme [21]

Other variables in the model:

- $\Phi_{int} + \Phi_{sol}$ represents the sum of incident solar radiation on an external surface of the envelope (Φ_{sol}) and the internal gains from appliances and human presence (Φ_{int}) divided between internal surfaces (θ_s), mass node temperature (θ_m) and indoor air node temperature (θ_{air});
- $\Phi_{HC,nd}$ heating or cooling power; can be an input when maintaining desire indoor temperature or an output;
- θ_{sup} ventilation stream air temperature; if there is only natural ventilation this value is equal to the outdoor temperature, for mechanical ventilation it can be set to appropriate value;
- θ_e outdoor air temperature, provided as input from weather file;
- θ_s internal surface average temperature, obtained as a result of preliminary calculation;
- θ_m mass temperature of the opaque walls, obtained as a result of preliminary calculation;
- θ_{air} indoor air temperature; the final result of the calculation;

Since real buildings are simulated and the RC model does not have any library of materials to choose from the thermal capacity in the model is calculated manually. Materials used for equivalent parts of the building zones, namely walls, ceilings and floors, are defined by density, thickness, area and specific heat. Dimensions of the envelope, as well as the window area, are taken from Energy Plus envelope models. From the dimensions of the respective surfaces and their thermodynamic parameters, the thermal capacity was calculated. All the necessary data was carefully checked and assigned to the respective variables in the 5R1C model.

5R1C model considers direct and diffuse radiation. Algorithm calculates the angle at which sun shines on the dwelling's windows based on the latitude and longitude of the location of interest and the altitude of the sun at every hour. Latitude and Longitude have been provided manually as following: 41.828 , -7.787 .

2.2. Energy Plus

Energy Plus® is a free, open-source program developed by the United States Department of Energy (DOE). Used mainly for the simulation of energy and water consumption. Energy Plus is a modular, structure-based program build on the most popular features and capabilities of BLAST (Building Loads Analysis and System Thermodynamics) and DOE-2.1E. One of the main advantages of this program is user-defined, sub-hourly time steps. Inputs and outputs are given in text files. The programme simulates air movements and interactions between zones. EnergyPlus is using the Transfer Function Method (TFM) for the calculation of thermal loads. This function has been adopted by ASHRAE and is regarded as one of the most accurate methods for the calculation of heating and cooling loads [22]. The TFM method is to first compute heat gains from walls, windows, floors and roofs, and then by multiplying the heating loads by some heat transfer coefficients the cooling loads are calculated [22]. EnergyPlus is a simulation engine to which one provides input data in text format. But this feature gives ultimate opportunity for to build applications to interface with Energy Plus. For a better user experience and for actually seeing the building geometry that is modelled Graphical User Interface (GUI) software can be used. There are a few free GUI software, Simergy, ModelMaker but OpenStudio – developed by National Renewable Energy Laboratory (NREL) and US DOE – together with Google SketchUp are one of the most popular.

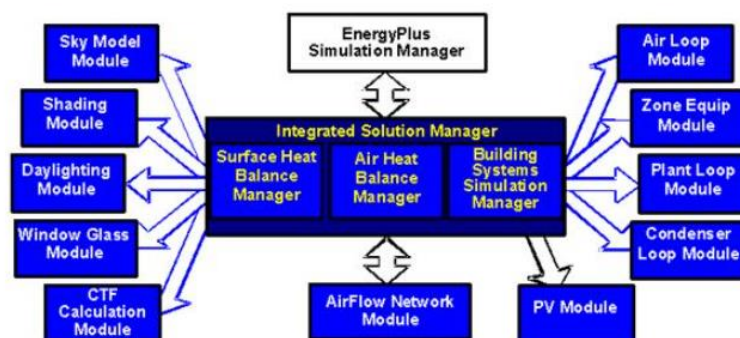


Figure 4 EnergyPlus internal elements / modules [3]

EnergyPlus based on user's description of the building, associated mechanical systems and weather data calculated heating and cooling loads necessary to maintain control setpoints through auxiliary HVAC systems and coil loads, as well as the energy consumption of primary plant

equipment and many others simulation details. All those additives are called modules (Figure 4) the structure of the model is built on the principle of management of those modules [3]. One of them is advanced fenestration calculation module which includes controllable window blinds, electrochromic glazing, heat balances through window layers to properly assign solar energy absorbed by window panes and a library of commercially available windows. Anisotropic sky model module allows for calculation of diffuse solar energy on tilted surfaces [3]. Such structure allows the model to employ only the modules that are necessary for the simulation and avoids going through the whole code. The individual modules sets global flags to inform the other models and the “Manager” about their status and whether to read input, initialize, simulate or report [3].

2.3. Machine Learning methods

In this work machine learning has been implemented in two approaches. Random Forest Regression and Artificial Neural Network has been considered. This has been done for two reasons. First, to verify if Machine Learning can be applied for simulation of EWE using general information about archetypes and second, as an attempt to make the model more autonomous and simplify the analysis. The models are described in detail in following subsections

2.3.1. Random Forest Regression

Machine Learning is a part of Artificial Intelligence branch. Random Forest is a supervised machine learning algorithm that is constructed from Decision Trees algorithms. It is applied in various industry fields such as e-commerce, health care and banking to predict outcomes or behaviour [4]. Random Forest is used to solve regression and classification problems. It utilizes ensemble learning which is a technique that employs multiple times one algorithm in multiple combinations to provide a solution to a problem.

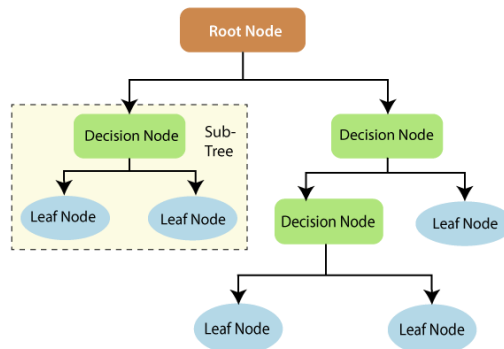


Figure 5 Generic scheme of Random Forest working principle [4]

As the name suggests Random Forest consists of multiple Decision Trees (Figure 5) generated by the Random Forest algorithm. The “forest” is trained through bagging or bootstrap aggregating, which are ensemble meta-algorithms that improve the accuracy of ML algorithms. This involves using different samples of training data rather than just one sample. The number of trees and how those trees behave can be set during the building phase of the model. However, one of the features of Random Forest is that in every generated tree a subset of features is generated randomly. Regression is the task to which Random Forest is employed in this thesis. Random Forest regression follows the simple concept of regression, but when in linear regression the function is known and steps can be traced, the function of Random Forest regression is like a blackbox – internal workings are unknown [4]. For one data point, each one of the trees predicts a value. The final output is an averaged value across all of the predicted values. This is contrary to Random Forest classification, where the output is determined by a majority-voting system [4]. In this way, not only is accuracy improved, but also the stability of the model, because it is less likely for any change in the data set to impact on the whole “forest” of trees.

2.3.2. Artificial Neural Networks

A Neural Network is made of artificial neurons – perceptrons as suggested by M. Minsky and S. Papert [23] – that receive and process input data. An Artificial Neural Network (ANN) consists of hidden layers of neurons. The most basic ANN has 3 layers: input layer, hidden layer, and output layer (Figure 6). The process starts when input data is fed into the model. Data is then processed via its layers to provide the desired output.

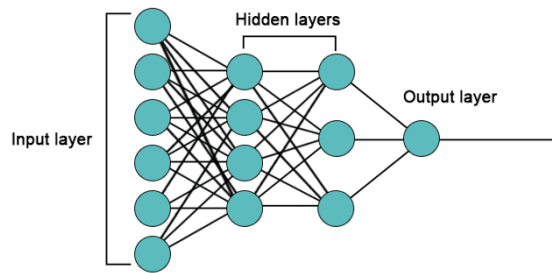


Figure 6 Structure of Artificial Neural Network [5]

Neural Network learns from the data and its structure and gives output. There are three categories of learning. When the data – the inputs and outputs – are labelled the method is called supervised learning because model is trained on how to interpret data [5]. Unsupervised learning is when ANN learns without human intervention and the data is not labelled. The output is determined only based on patterns identified within the fed data. When ANN learns depending on feedback received from the user it is called reinforcement learning [5]. For the purpose of this research supervised learning will be used. Neuron – perceptron [23] – is a node that receives multiple signals and gives one output signal (Figure 7).

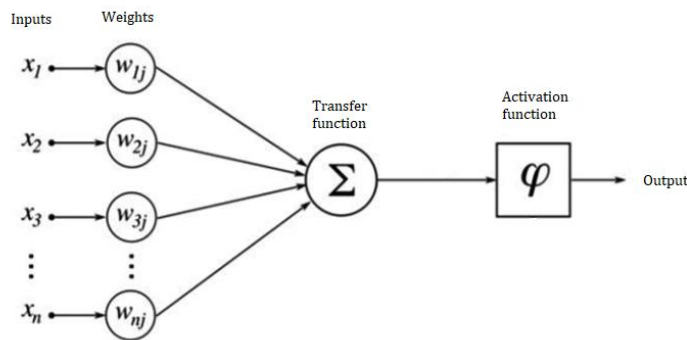


Figure 7 Construction of an artificial neuron

It is important to standardise or normalize the input values so that the neuron will be able to process them way easier. Every input signal has assigned weight. Weights are the point in which ANN learn by calculating a function called “cost function” which compares the actual output value with predicted one and adjusting the weights. The goal is to minimize the cost function. Alongside with the weights activation functions are the tools that trigger the neuron. There are a few different activation functions which can be divided into two groups: linear and non-linear [6]. Linear functions will not be confined by any range. The most used functions right now are Sigmoid and ReLU (Figure 8) where the latter is gaining even more popularity [6].

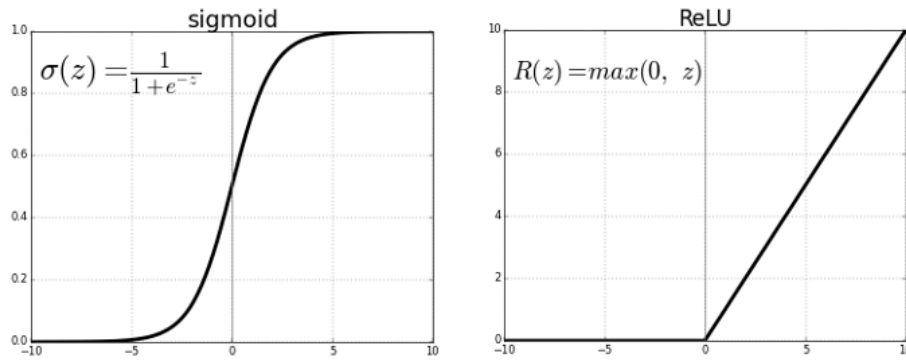


Figure 8 Activation functions: Sigmoid and ReLU [6]

Sigmoid function exists between 0 and 1. Its shape is the main reason it is used in prediction of probability as an output. ReLU on the other hand is rectified till point 0. This means that any input value which is lower than 0 will be automatically set to 0. This causes an issue of not mapping negative values correctly and decreases the ability of the model to fit or train from the data properly. The problem was solved with introduction of Leaky ReLU with additional variable which for negative values is set usually to 0.01 [6].

Chapter 3

Methodology

This chapter describes the consecutive steps that has been undertaken in this thesis. First the procedure of turning energy certificates into archetypes and models. Then the description of what is modelled – the zones. Next the procedure for calibrating and how the models will be compared, followed by the procedure of data feeding into the models. Finally the description of how the different models has been built.

3.1. Montalegre case study

The town of Montalegre has been chosen for the case study an testing the first prototype. It is a town located on the very North of Portugal in district Vila Real along the border with Spain. At the level of 1000m above sea level Montalegre has a cool Mediterranean climate. There is a short dry season in summer but overall, plenty of precipitation. Average annual temperature is around 10°C. But this town is known for extreme weather events both in winter and in summer. Local agencies have been developing very actively adaptation strategies to climate change. The weather data for the prototype and for the purpose of this thesis has been collected in nearby town of Braganca in 2019. It has municipality with large share of old population (Table 1)

Table 1 Population structure in Montalegre [24]

Age Group	Year 2001 (resident pop.)	Year 2011 (resident pop.)
0 – 14 years old	1.666	1.003
15 – 64 years old	7.609	6.053
65 and over	3.487	3.481

3.2. Energy Performance Certificates into models

Energy Performance Certificates (EPCs) for buildings in this town has been collected by Portuguese Energy Agency (ADENE). EPCs have detailed information about the dwelling construction solutions and other relevant data for its energy simulation. It contains typology, year of construction, construction solutions and information about heating and cooling systems. Following the division per construction period followed by the ADENE based on construction booms, new milestones in building regulations, nine construction periods were considered: 1919-1945; 1946-1960; 1961-1970; 1971-1980; 1981-1990; 1991-1995; 1996-2000; 2001-2005; after 2006. The specification of each archetype is shown in Table 3. This periods’ definition was established considering the similarity of construction solutions for those periods. As it is impossible to simulate each building of large urban area, there is a need to use this simplified approach in building energy simulation, usually called building archetypes [25] [26]. Such solution introduces an inaccuracy error. Thus the resulting output from models is not expected to be of spectacular accuracy when compared to real measurement, rather a behaviour that reflects in general

the temperature inside the envelope is an objective. The RELIABLE project aims to forecast extreme weather events (EWE) impact on indoor temperature and following next step in the future proceeding in the project - the impact on the dwellers.

3.3. Monitoring Campaign

The data for case study has been collected using HOBO® U-12-011 data logger (Figure 9) which was calibrated in temperature chamber by the National Laboratory for Civil Engineering. Table 2 Monitored parameters, equipment used and resolution of measurement shows the specification of the device. The location of the device was living room or bedroom since those are the rooms that are occupied for most of the time. The position of the device has been carefully chosen to not have any direct sunlight exposure on the device and to not be in the vicinity of heat sources.



Figure 9 Indoor temperature and relative humidity sensor by HOBO® [7]

Table 2 Monitored parameters, equipment used and resolution of measurement [7]

Monitored Parameters	Equipment	Resolution
Temperature (°C)	HOBO U-12-011	0.03°C at 25°C
Relative Humidity		0.03% RH

Measurements were conducted every hour for 43 days in the period from 12.05.2019 to 24.06.2019.

3.4. The basic model

When performing a building energy simulation, it is crucial to define several inputs regarding the building geometry, construction solutions, schedules of occupancy and internal gains of the building. To define

the geometry of the building model (one per archetype) EPCs provide information about the floor area and height of the dwelling as well as the number of floors. From this information it is possible to calculate the width of square box and draw the building zone. Both models – 5R1C and Energy Plus – have an option to provide a schedule of occupancy. It is hard to say how such a schedule should look like and it is rather obvious that in each dwelling it looks differently. It was therefore assumed that for the model to predict EWE which may be harmful to elderly people – those of the age 65 and older – the schedule assumes to have one occupant all the time in the dwelling, for the night the number of occupants is increased to three people. This was done on the assumption that the young occupants are either working or at school and the older occupants remain at home most of the time. Occupancy influences the indoor temperature because humans also emit heat to the environment. It is not a significant amount of heat (around 100W [27]) but when more occupants are present the overall indoor temperature is influenced by the heat produced by their bodies.

Table 3 Archetype's description derived from Energy Certificates database for Montalegre

Period of construction	floor area	average height	number of floors	U- walls	U-roof	U-floors	U- window	solar factor (g-value)	Window Wall/Ratio
1919-1945	99.7	2.6	2	1.85	2.67	1.82	4.4	0.9	8%
1946-1960	103.0	2.7	2	1.87	2.59	1.61	3.9	0.9	9%
1961-1970	93.5	2.5	2	1.59	2.45	1.90	4.3	0.9	10%
1971-1980	91.4	2.5	2	1.59	2.36	1.81	4.2	0.9	10%
1981-1990	140.7	2.7	2	1.32	2.13	1.98	4.1	0.9	12%
1991-1995	132.7	2.7	2	1.30	2.25	1.67	3.6	0.8	12%
1996-2000	117.8	2.6	3	1.18	2.21	1.60	3.5	0.8	15%
2001-2005	135.3	2.6	2	0.93	1.61	1.17	3.5	0.8	14%
after 2006	141.3	2.7	2	0.80	1.07	1.08	2.8	0.5	22%

To simulate the indoor temperature in Energy Plus one needs to provide a 3D model of the dwelling to the software. Developers of Energy Plus software recommend OpenStudio a plugin for Google SketchUp [28] for the design of the dwelling since it is free, simple to use and allows running Energy Plus simulation directly from Google SketchUp. There were nine cubes created (Figure 10). Each one has been designed according to the archetype specification.

Table 4 Flow chart of methodology used in research

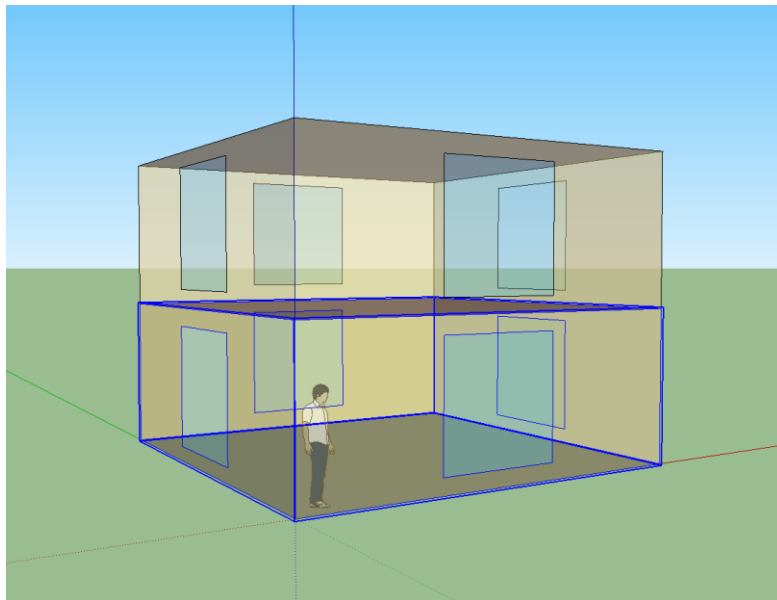
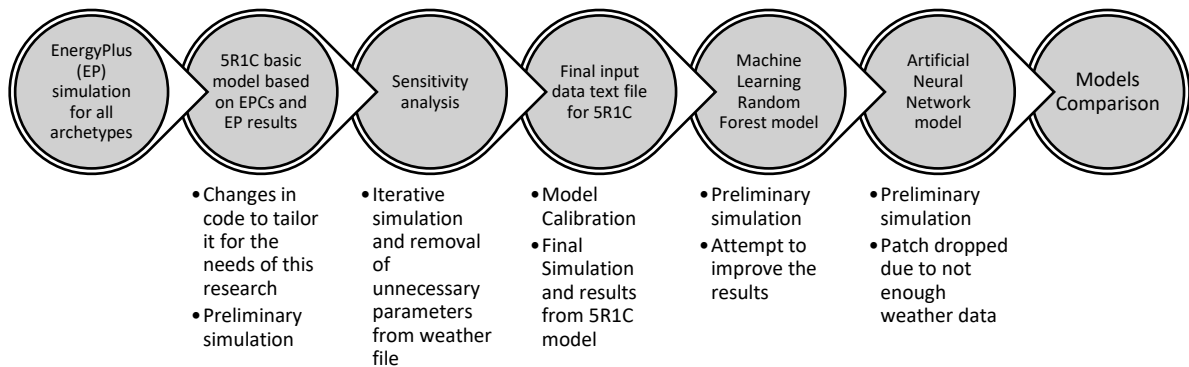


Figure 10 One of the cubes designed in Google SketchUp.

Montalegre buildings have mostly two floors therefore, the archetypes are modelled as two cubes connected in such manner that ceiling of the bottom one is the floor of the upper one. The surfaces of the cube were given unique names. Next, in the Energy Plus Launch Document, those surfaces were assigned features according to EPCs. The weather file used for simulation had information about the weather from the previous year. Data was collected in Braganca near Montalegre. The general methodology applied for the research is shown in Table 4.

3.4.1. 5R1C

The basic code (application) for the model has been accessed through *github.com*. It has been created by the Department of Architecture and Building Systems of the ETH Zürich [29]. The code has been than tailored for the needs of the research.

After creating a loop in the code to read data for the 9 archetypes a first preliminary run has been conducted. As presented on the Figure 11 below the temperatures are quite high in comparison to the data acquired by the means of HOBO device.

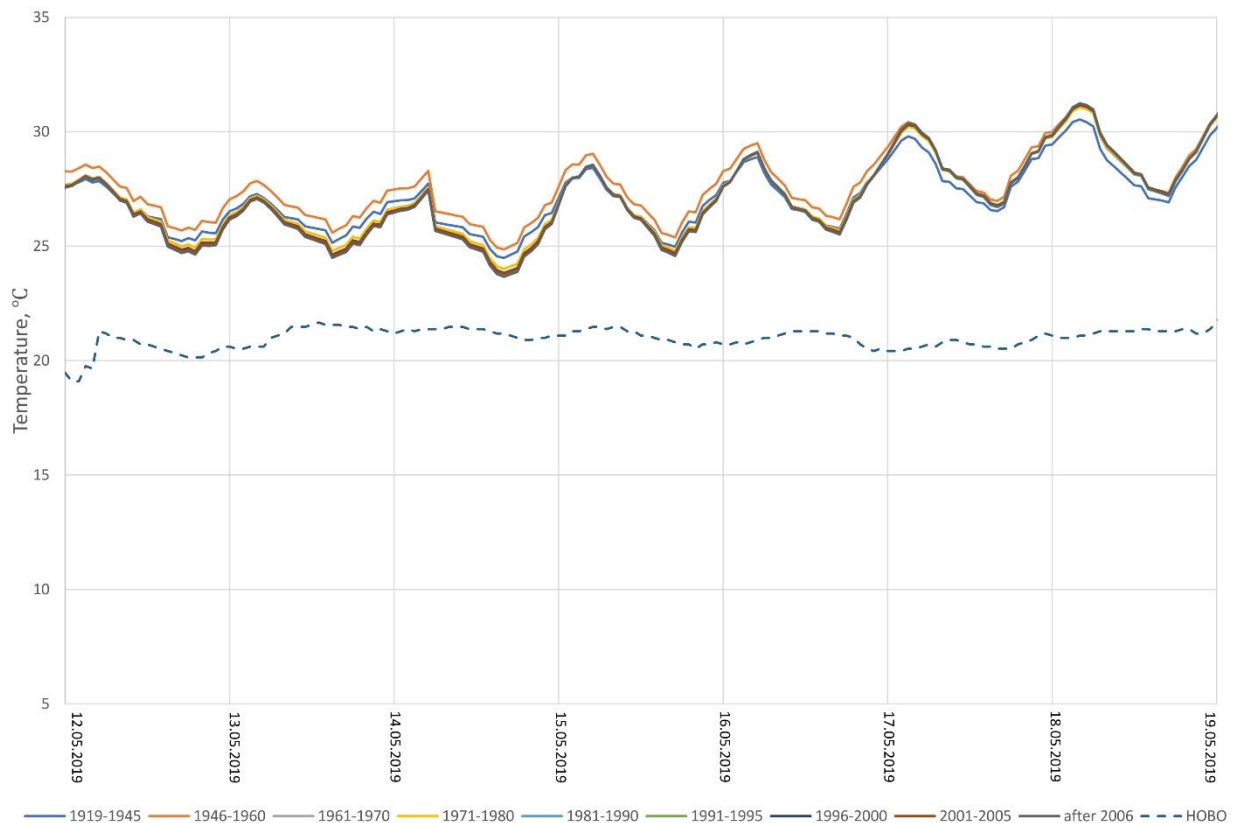


Figure 11 Result from 5R1C model after first test run with comparison to HOBO measurement

Moreover, the lines for each of the archetypes are rather close to one another which suggest there is almost no distinction between archetypes. Therefore, there was the need to conduct a sensitivity analysis to better understand the simulation parameters and outputs.

3.4.2. Sensitivity Analysis

Sensitivity analysis has been conducted to get a sense of how various input variables impact the output. It has been done by running the simulation nine times for the same weather period, with identical values

for parameters besides the one under investigation, which has been changed slightly in every simulation. The purpose of sensitivity analysis was to fine-tune the model. After preliminary simulation it turned out that the model is returning much higher temperatures than expected. It was hard to tell what impacts the results the most.

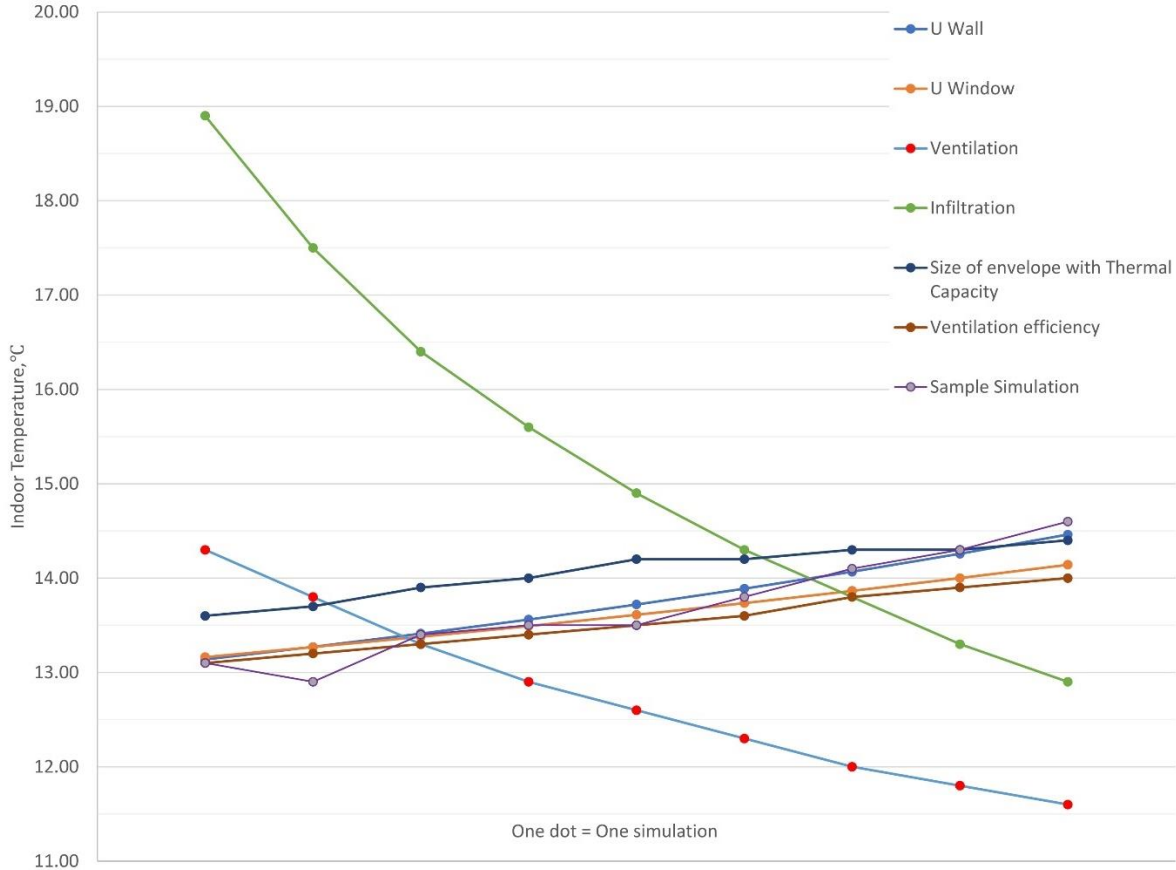


Figure 12 Sensitivity analysis chart. One dot defines one simulation run, one line presents how the output temperature has been changing with the change of given parameter. Lines are not correlated among each other and should be considered separately.

Figure 12 represents the entire sensitivity analysis. The graphs with output indoor temperature from sensitivity analysis of each consecutive parameter are in the appendix. The lines have been put together into one chart for the ease of comparison. One dot represents one simulation, one line represents how the output temperature has been changing throughout the sensitivity analysis. *U wall* and *U window* are the heat transfer capacities of walls and windows. Values of those two parameters has been decreasing with each simulation. *Ventilation* represents controlled air exchanges per hour and its value was increasing with each simulation. *Infiltration* represents the unwanted air exchanges due to imperfections/cracks of the envelope; its value has been increasing with each run. *Size of the envelope* was combined with *Thermal Capacity*, because *Thermal Capacity* is defined as “per floor area” therefore, with the increase of size of the envelope – floor area – the thermal capacity is decreasing. *Ventilation efficiency* value has been increasing and it represents the wellness of heat recovery from ventilation. It can be concluded quite easily that the biggest impact on the output temperature comes

from the infiltration and ventilation parameters. As an outcome from sensitivity analysis the infiltration parameter has been tuned to bring the output temperature into the vicinity of the real measurement.

The outcome of the sensitivity analysis is pointing to the infiltration and ventilations parameters which were not measured by any device. Manipulation of those parameters helped moving the temperatures into more reasonable range. This information forced the researcher to assume infiltration value to be 2.5 air exchanges per hour. The same was applied to every archetype.

3.5. Data input procedure in models

To provide data into EnergyPlus it is needed to specify the features of each component. The thermal properties of walls and windows. The weather file is supplied by specifying the patch to the file. The output from this software may be a CSV file with the output data that you first specify in the software.

Providing data into the RC model is quite different. Since the RC model is not software but a code written in Python, the data is read from an external CSV file. The unique parameters that must be provided to the program are the geometry of the envelope namely the areas of floors, walls and windows and the thermodynamic parameters: heat transfer coefficient [described as U in datasheet] for the walls and windows and thermal capacitance per floor area of the dwelling. Additionally, parameters like lighting utilization factor, the coefficients for ventilation and infiltration as well as ventilation efficiency can be specified if known. If the additional parameters are not provided externally the program will apply default values. The program has also an option to set the temperature for turning on heating or cooling. However, in this case it was assumed that there is no heating nor cooling in the building. Ventilation efficiency describes recuperation of heat which was also assumed to be 0. RC model considers the irradiation of the sun. Therefore, the transmittance of sunlight and radiation for the windows needs to be set. In this case, the transmittances were not known and the values from EP were taken and introduced into the RC model. Gains per person are also considered and a schedule of occupancy is needed for that. Such a schedule for the RC model was built from scratch and it does not take into account holidays and was simulated in such way that for 8 hours a day all habitants but one leave the dwelling.

3.6. Model calibration

Results from simulation were compared with real measurements conducted in a house belonging to the second archetype (1946-1960) in Montalegre. Measurements were conducted from the 12th of May to

the 26th of June. The device used to measure the indoor temperature is widely used temperature logger HOBO. It has to be assumed that during this period in the house, windows could have been opened and most probably some heating was applied which influences the measurements. Unfortunately, the measurements did not take into account abovementioned events, therefore measurements were treated like for dwelling with closed windows.

To compare the results two statistical indicators were used: Mean Bias Error (MBE) and Coefficient of variation of the Root Mean Square Error (Cv(RMSE)). MBE measures how closely the simulated data coincide with the measured data. It is an overall bias measure of the data. As shown in the equation (1) below it is the total sum of the difference between simulated and measured values in the considered period (one hour). The sum of differences is then divided by the sum of the measured values.

$$MBE(\%) = \frac{\sum_{Period} (S-M)_{interval}}{\sum_{Period} M_{interval}} \times 100\% \quad (1)$$

The final values are compensated (reduced) by negative values if present. Due to this compensation effect, MBE is often represented together with Cv(RMSE). The Coefficient of RMSE is calculated as the Root Mean Square Deviation (RMSD) normalised to the mean of observed values. To put it simpler, Cv(RMSE) measures either the variability between measured and simulated data or goodness-of-fit of the model. It specifies overall uncertainty in prediction reflecting the size of the error. It is always positive and it is insensitive to the compensation effect. The lower Cv(RMSE) the better the calibration of the model. It is calculated as follows:

$$Cv(RMSE) = \frac{RMSE_{Period}}{A_{Period}} \times 100\% \quad (2)$$

$$RMSE_{Period} = \sqrt{\frac{\sum (S-M)_{interval}^2}{N_{Interval}}} \quad (3)$$

$$A_{Period} = \frac{\sum_{Period} M_{Interval}}{N_{Interval}} \quad (4)$$

where $N_{Interval}$ is the number of time intervals considered for the monitored period.

A model is considered as 'calibrated' if the abovementioned indices meet the criteria set by the ASHRAE [30] (Table 5):

Table 5 Acceptable tolerances defined by ASHRAE Guidelines

INDEX	ASHRAE 14 (%)
-------	---------------

Cv(RMSE)	15
MBE	±5

It is important to mention that those values are not strictly set and are only guidelines to give an idea of how good is the model.

3.7. Machine Learning model: Random Forest

The next step, after having the 5R1C model tuned up, was to build a Machine Learning (ML) model based on the output from 5R1C. Machine Learning models need to be trained before being able to produce reliable results. To do that a set of carefully selected features, together with the output they produce, needs to be provided. The model finds patterns in the input data and learn what output should particular input produce. This method is called supervised machine learning. Weather file in EPW format has 35 parameters. Giving them all into the model as input would be utterly witless. The output from training on such a set of data would produce exact results as the provided input training output. Such a situation is called overfitting. In other words: the model is sensitive to noise. Overfitted model cannot correctly predict any results, because it is tailored for only one set of data. The solution to this is feature selection.

3.7.1. Feature selection

The set of 35 parameters in weather file was carefully investigated. First, each parameter, visually to have an idea of how it is behaving. Then, by the use of a built-in method for parameter selection in regression models – backward elimination. This method required running one simulation after another, analysing scores which showed how important a given parameter in a given set of parameters is or what influence on the output it has. Consecutive runs were conducted after the removal of the parameter with the highest significance score - P-value. P-value is the probability value for a given statistical model that, if the null hypothesis – general statement that there is no correlation between two measured phenomena – is true, a set of statistical observations are greater or equal in magnitude than the observed results. In other words an assumption is made – the null hypothesis - that given combination of variables do not have any effect on dependent variable. Then P-value is calculated and variable with P-value higher than the arbitrary selected threshold is removed from the dataset. Abovementioned steps are repeated until a set of features with which the overall performance of the model is in the acceptable range. Basing on

the experience of the python society the P-value has been set to 0.05 (5% significance). After conducting feature selection, a set of 6 features was selected. Result of the training is presented on Figure 13.

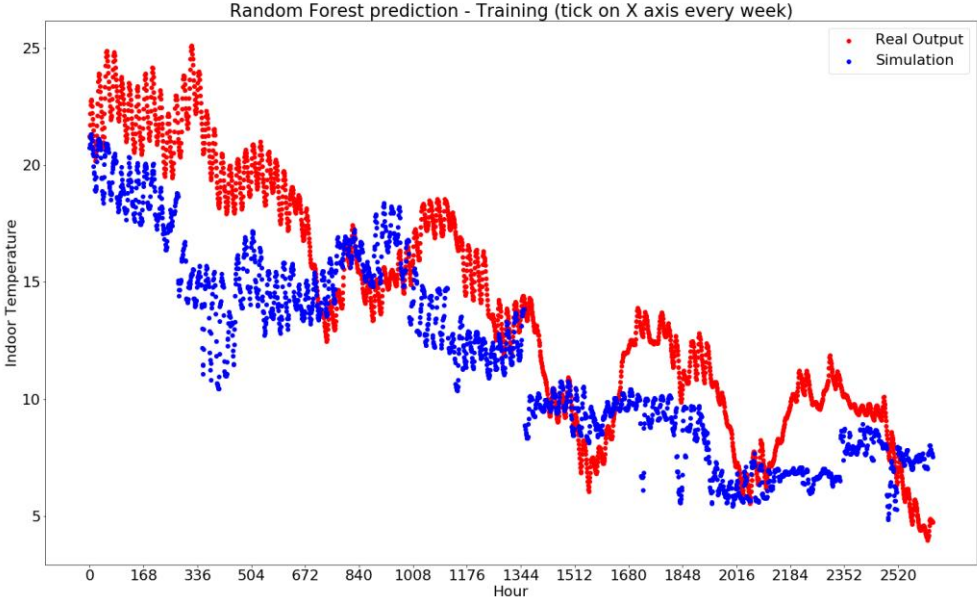


Figure 13 Random Forest training result

A common practice in regression models is using the information from the previous step. However, in this case, such practice created overfitting. The model was able to predict the output with almost 100% effectiveness basing only on dry bulb temperature. Due to this tendency model build for this work does not use the information from the previous step.

3.7.2. Artificial Neural Network

The basic structure for ANN was written in Python using TensorFlow library. Building the structure is quite simple but creating a model that is supposed to forecast behaviour of nine archetypes is much more complicated and time consuming since the approach here is separate ANN model for each archetype. Training each ANN is time consuming. Since the weather data was for one year, it was shuffled in order to give the ANN a “sense” of different output temperatures and parameters influencing the output. This resulted in shuffled sample for testing and therefore a discontinuous resulting data. Due to limited time resources and results far from acceptable level a conclusion was made that sample of

one year is not enough to properly train the ANN, the investigation of ANN usage for EWE prediction has been stopped and the results will not be presented here.

Chapter 4

Results

Discussion about the results. Comparison between models and correlation between real measurements and model results.

4.1. Discussion about the results

The results from the three models were compared for the period from 12.05.2020 13:00 to 24.06.2020 23:00 for which the real measurement was conducted. The house in which the measurement using HOBO was done is of the archetype 1946-1960. The temperatures for the rest of the archetypes will be shown also.

The chart presented below (Figure 14) shows comparison between models and real measurement for archetype 1946-1960. Lines which are semi-transparent shows results hour by hour. Each of such lines has bold line that was created using moving average method for 24h periods. The main purpose of this graph is to show the behaviour of each model for easier comparison with another and the real data. What is standing out here is the Orange RC model. This is the model before adjusting it with infiltration values. Adjusted 5R1C model, ML Random Forest and Energy Plus are following almost perfectly the real HOBO measurement. What is interesting is that Energy Plus reacts in much more “real” way – influence of thermal capacity – to the changes of outside temperature shown with thin purple line. Of course there are also other parameters influencing the model but outside temperature has the biggest share in abovementioned impact. Moreover, Energy Plus did not react as drastically to the sudden increase of temperature – at the right hand side of the chart – as RC model did. Surprising is that ML Random Forest also did not react as drastically although it was build based on RC model. The discrepancy between Energy Plus and RC models stems from the way they handle parameters. RC is simplified model thus such outliers are not concerning.

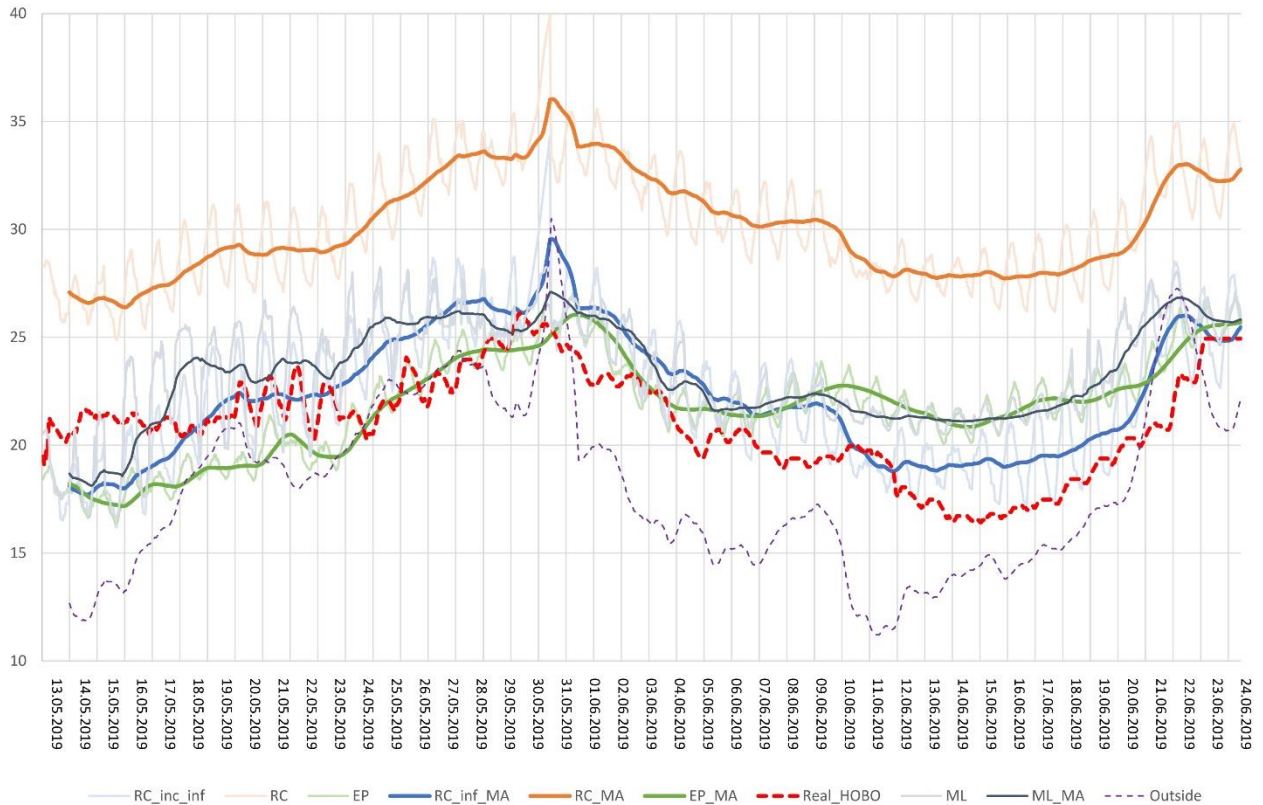


Figure 14 Models comparison with HOBO measurement and outside temperature

Comparing the results from different models with the real measurements a really good fit can be observed. Of course there are some fluctuations between models but the overall outcome is very satisfactory. But from where are those differences coming? It stems out from the methodology and theory behind models. It can be seen that EP has a tendency to rather delay the peaks. This is due to two things. First the way the model handles thermal capacity, second the amount of weather information it receives. EP takes into account all the data from weather file. Where RC as simplified model requires only few of them. The ML model is showing potential for further investigation. Although for the current amount of data the results are not as good as from 5R1C, with larger sample of data the output can be greatly improved.

4.2. Energy Plus

The Energy Plus results are one of the most uniform results (with low amplitude of oscillations). This model shows that the archetypes with the lowest variations in temperature are the archetypes 1919-1945, 1946-1960 and 1961-1970. In terms of indoor temperature those are showing the biggest similarity to one another. This tendency is also pronounced in the results from 5R1C model. This phenomenon is

connected to significantly higher thermal capacity of those archetypes. It is around 1350 kJ/m²K for 1919-1945 and 1946-1960 archetypes and 802 kJ/m²K for 1961-1970. When the thermal capacity of the rest of the archetypes is between 290 and 312 kJ/m²K. Energy Plus results are almost perfectly in phase with the real measurement.

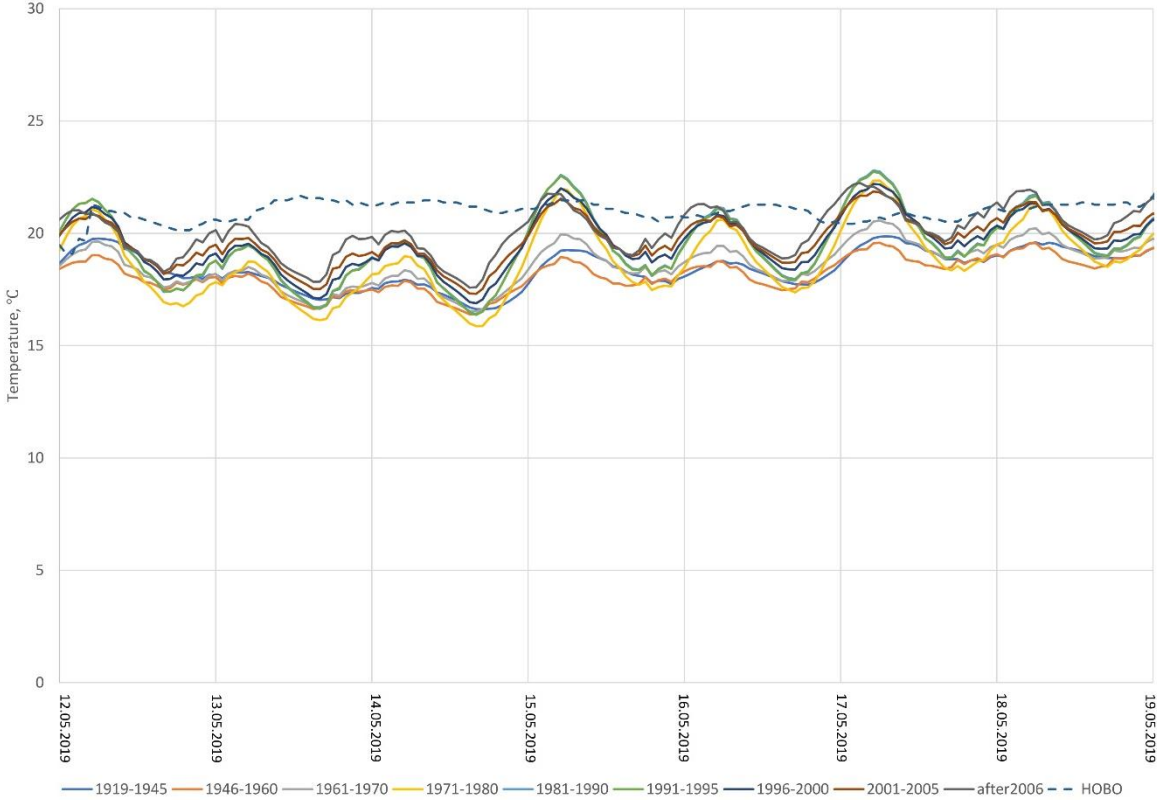


Figure 15 Energy Plus results for one week in May.

4.3. 5R1C

The most significant thing which bring 5R1C ahead Machine Learning models is that it takes into account only a few arbitrarily chosen parameters from weather file. It helps save significant amount of time on feature selection, but on the other hand it seems that there is less control over the model.

Figure 16 show the results from RC before tuning it with infiltration values. It can be seen that the values of temperature for each archetype more or less are around the same values. When infiltration value was introduced – such to bring the graph as close to the real measurement as possible which is 2.5 air exchanges per hour– the lines for each archetype are more spread and it looks more like real situation. Infiltration has been chosen due to the fact that the data collected is from various periods of time and the standards in this times were changing over time and some parameters were not considered. The infiltration parameter impacts the temperature amplitude between day and night which is visibly higher.

Moreover it introduces distinction between archetypes. This proves that although here infiltration parameter has been assumed arbitrary it should be measured and included into thermal characteristic data set of buildings or even as one of parameters measured as a part of EPCs. What is standing out from the result is the difference in temperature amplitudes between archetypes. 1919-1945 and 1946-1960 are the archetypes that present the lowest variations in temperatures. This is of course due to high thermal capacity. This two archetypes show almost the same output. Such situation stems out from the different standards in construction industry throughout the years and the fact that the calculated thermal capacity may be inaccurate in this case. Conclusion is that thermal capacity of buildings should be carefully investigated further as it is – together with infiltration and ventilation parameter – one of the most important parameters influencing the final output from the models.

When Comparing Energy Plus with 5R1C, the profile of temperature is very pronounced in the simplified model. It has a tendency to have very high oscillations for summer months when compared to the colder months. Another vivid conclusion that stems out from the results is that 5R1C model has a tendency to move out of phase quite drastically – this can be seen on the graph were comparison to HOBO is – the authors guess is that this is because of the input radiation values. The model reads them but in the real measurement those seem to have no influence on temperature. The question is “why?” Maybe there is a shadow cast on the dwelling which prevents the building to collect the heat from sunlight radiation. Moreover, the oscillations are higher in lower thermal capacity classes, which brings a conclusion that 5R1C model may be a good choice for the higher thermal capacity classes, since the results may be more reliable. This result proves other research conducted for Mediterranean weather conditions where Roberto Bruno et al. came to similar conclusions [21].

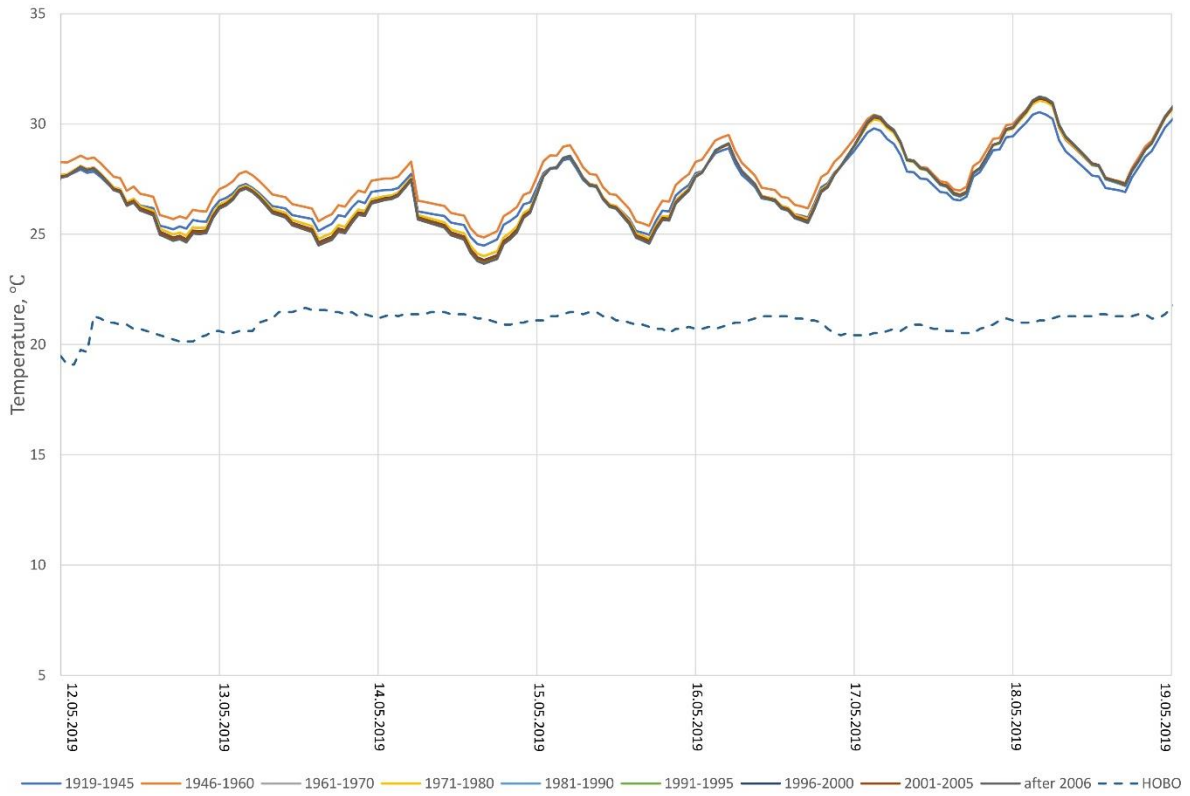


Figure 16 5R1C result graph for one week in May before fine tuning the model with infiltration values

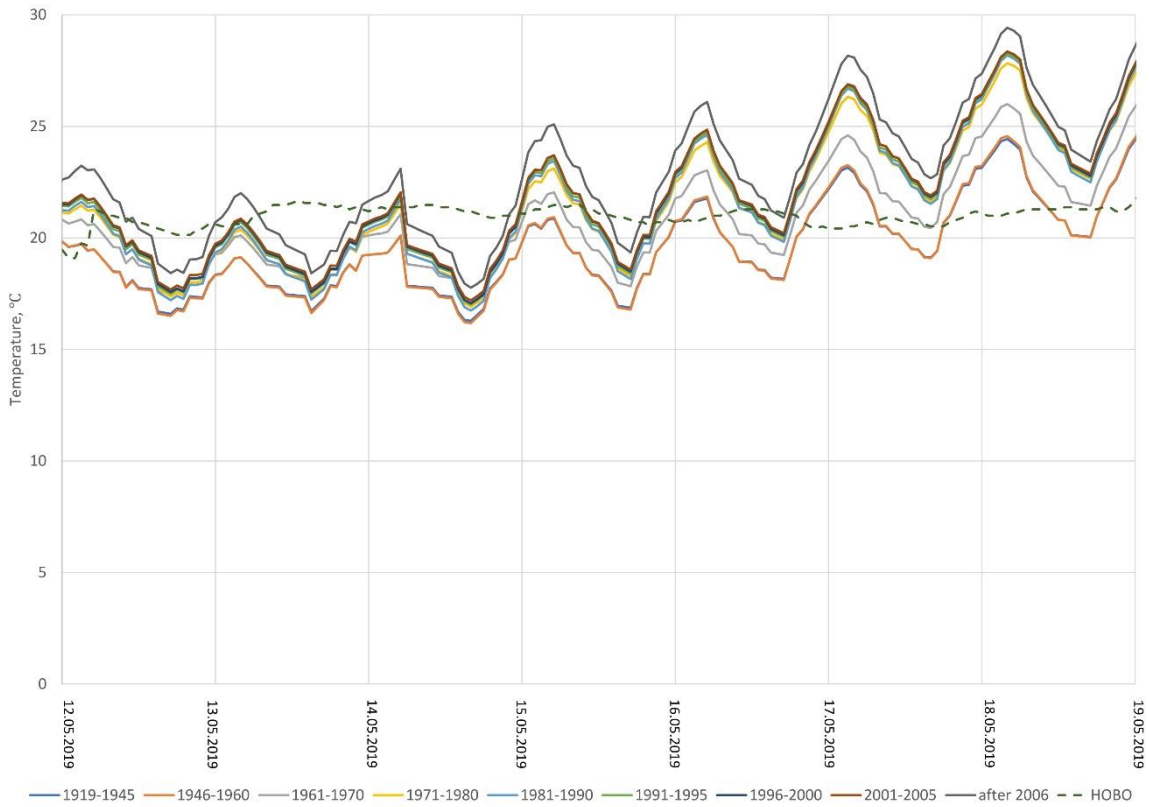


Figure 17 5R1C results after introduction of infiltration value.

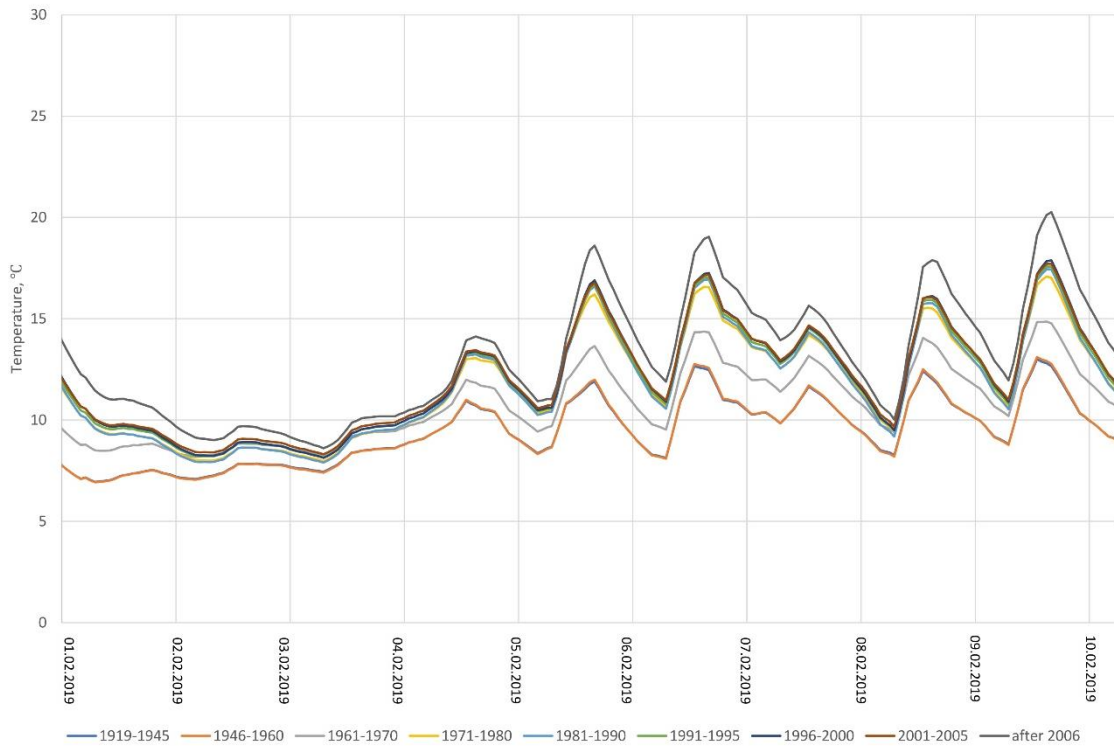


Figure 18 5R1C results for the first 7 days of February

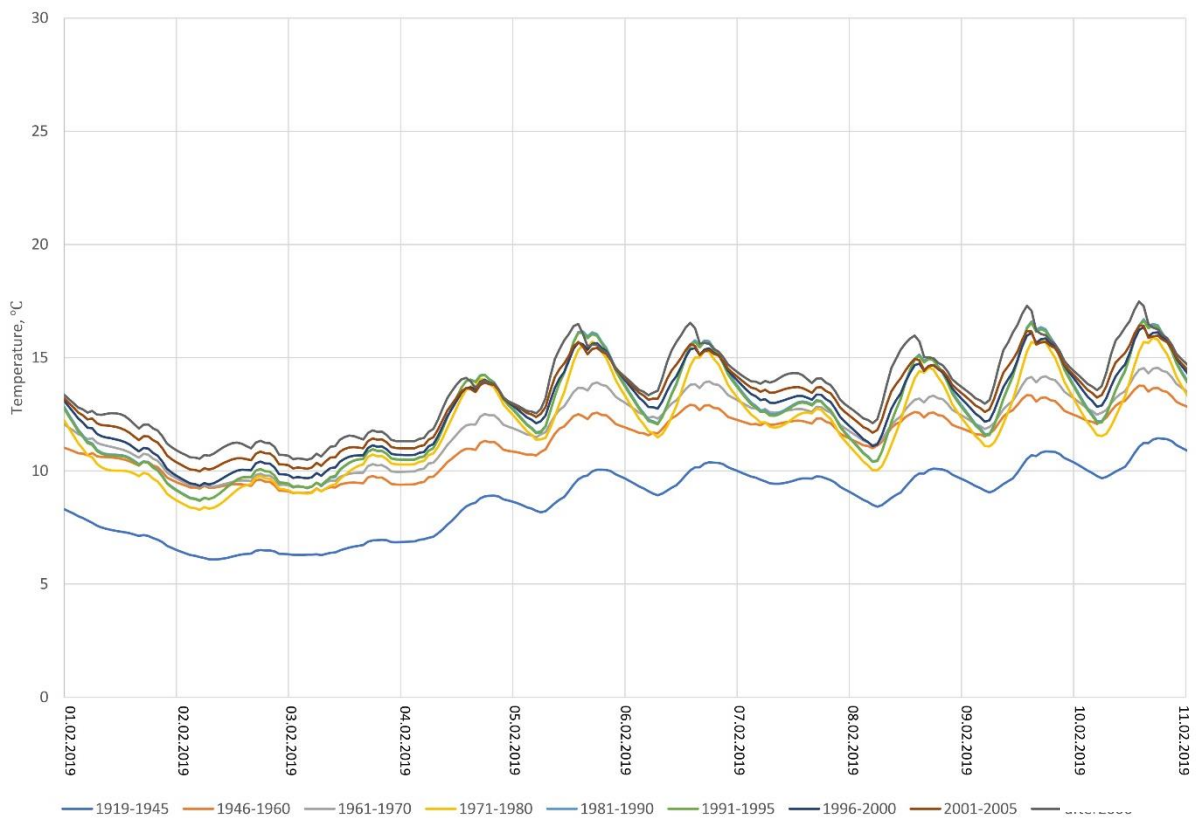


Figure 19 Energy Plus results for first 7 days of February

4.4. ML Random Forest

What can be seen almost immediately is that the resulting graph on Figure 20 is not as smooth as the 5R1C or Energy Plus models. The result gives some idea of the level at which the resulting temperatures are. High accuracy is not as demanded as distinction between archetypes which Random Forest was unable to do, at least for the data provided. In 5R1C archetypes have very pronounced distances on the graph. What is interesting is that even that the model was built on 5R1C output data, the period in February shown on Figure 21 is more similar to Energy Plus output than 5R1C. It is hard to give any reason why this happened besides the one that 5R1C was build based on Energy Plus output however, this process should already introduce some error. Overall the result is good, but lack of the ability to distinguish between archetypes disqualifies this model.

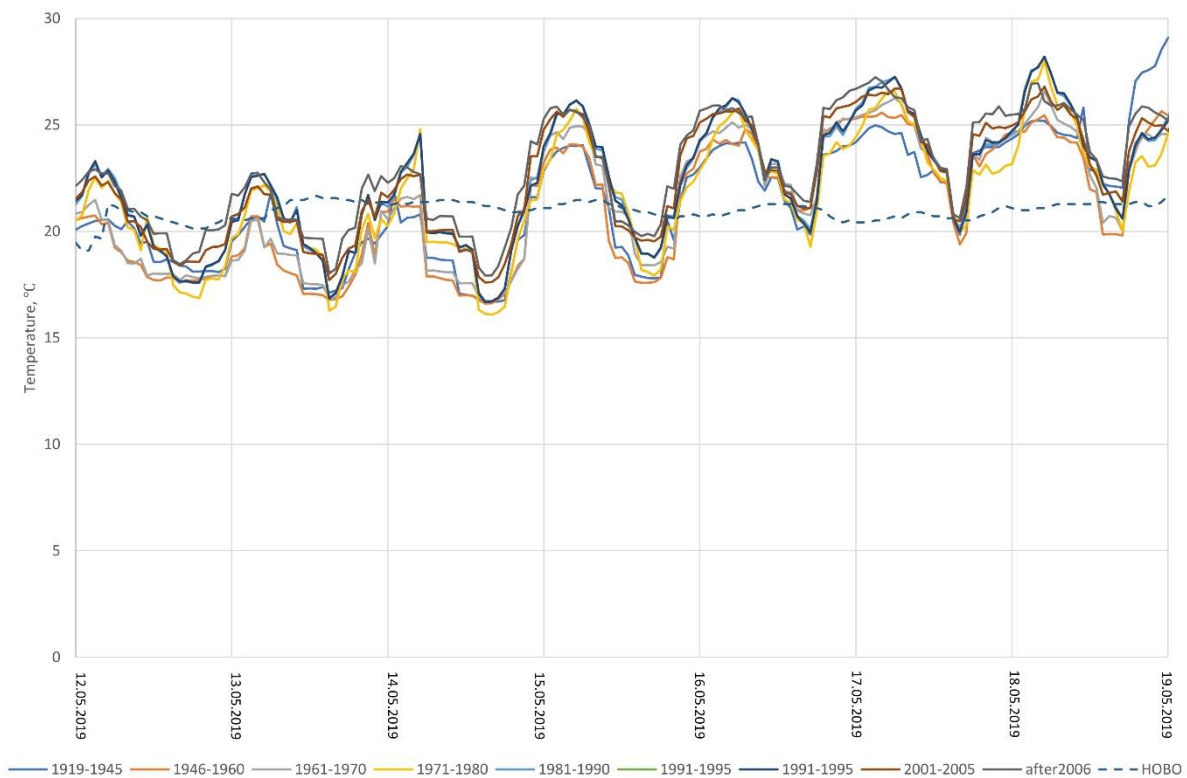


Figure 20 ML Random Forest results for 7 days in May with comparison to HOBO

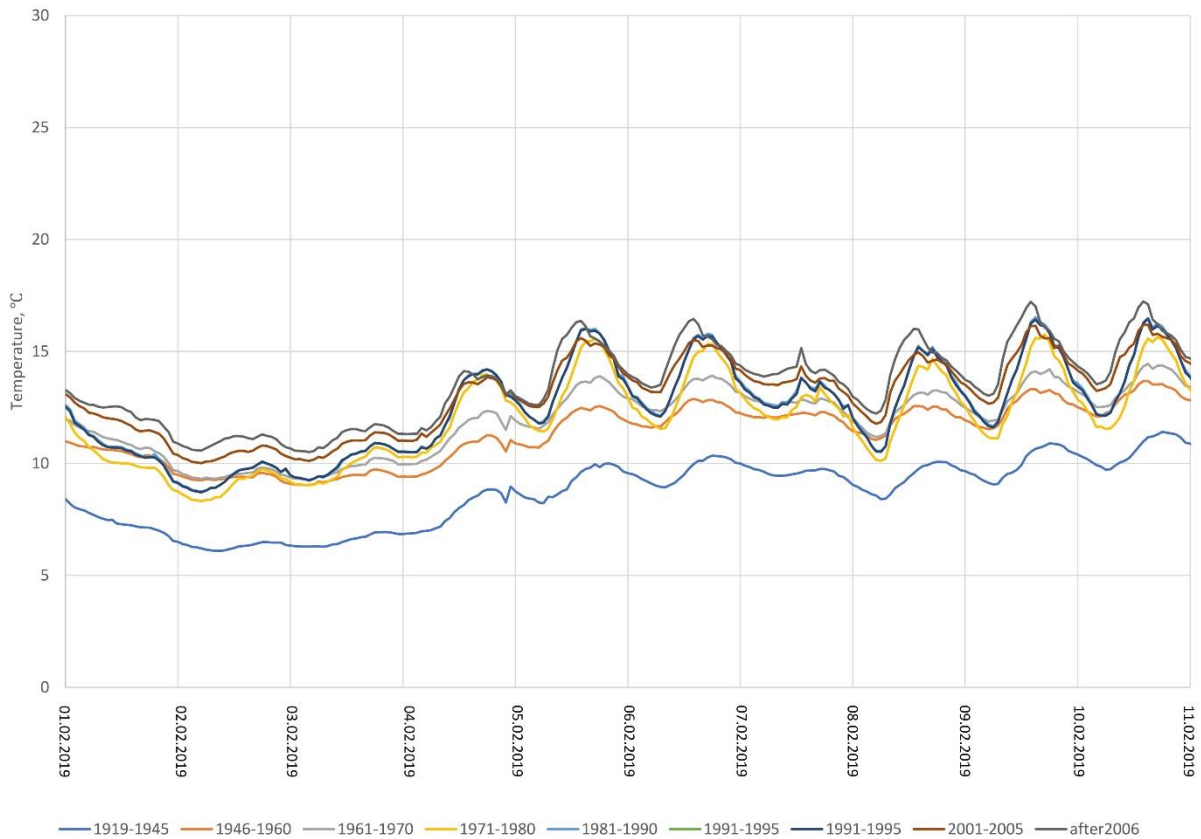


Figure 21 ML Random Forest results for 7 days in February

4.5. Artificial Neural Network

There was an attempt made to create ANN model however, due to time constraints and what the author concluded from the preliminary results, a decision was made to not pursue this topic. The historical data for one year back was not enough to simulate future temperature occurrences. Good practice in ANN field in forecasting weather events is to have at least 2-3 years of data on which the model can learn different relations and conditions.

4.6. Model Calibration results

Cv(RMSE) scores were calculated on different stages during the research. Introducing some changes into the model resulted in change of the score. Interesting is that when “improvement” was introduced and visually the graph looked much more like real values the Cv(RMSE) score showed that the correlation is weaker. The comparison is presented for 2 archetypes when comparing models to the real data. Then comparison between EnergyPlus and 5R1C is presented for the same two archetypes.

Table 6 MBE and Cv(RMSE) scores of comparison of model with real measurement

	EnergyPLUS		5R1C - 2 floors		5R1C - 1 floor	
	1919-1945	1946-1960	1919-1945	1946-1960	1919-1945	1946-1960
MBE	-11.06%	-4.02%	-12.29%	-12.35%	-18.08%	-18.07%
Cv(RMSE)	18.73%	12.90%	15.49%	15.60%	21.46%	21.43%

The scores from EnergyPLUS for 1946-1960 archetype indicate that here the performance was the best and within the ASHRAE guidelines. The real measurement has been conducted in dwelling from 1946-1960 period. Comparison has been done for two archetypes to check and confirm that the 1946-1960 archetype is the correct one due to the fact (described in section 4.3) that those two archetypes showed very similar results. Scores for 5R1C model are different and not quite good. There is distinction in 5R1C model, because it was first calculated as 1 floor building and then as two floor building as the real dwelling. In EnergyPLUS 2-floor construction is set by the 3D model configuration. In 5R1C this issue was resolved by doubling the wall and floor areas of the dwelling and increase in thermal capacity. This introduces an error because 5R1C has no option to assign the additional wall(ceiling/floor) between floors nor the thermal parameters between the two floors. Hence a conclusion can be made that 5R1C treats the geometry as one big room. However, the impact of such error cannot be measured in this case. Nevertheless resulting score for 5R1C is satisfying and is on the verge of ASHRAE guidelines when it comes to Cv(RMSE) but not when MBE value is considered. There is also almost no difference between two archetypes in both 5R1C methods which confirms what has been said earlier that there is a need to carefully investigate infiltration and ventilation as well as thermal capacity of each archetype.

Table 7 MBE and Cv(RMSE) scores of comparison EnergyPLUS and 5R1C models

	5R1C		5R1C – Increased Infiltration	
	1919-1945	1946-1960	1919-1945	1946-1960
MBE	-8.98%	-9.05%	-3.15%	-3.13%
Cv(RMSE)	12.28%	12.44%	10.39%	10.57%

There was higher correlation of those models and real measurements than between models. Correlation is higher when infiltration values are introduced. The lack of correlation may be stemming from high oscillations of 5R1C. On the other hand when looking at the Figure 14 it seems like the models are close to one another. Scores are within ASHRAE guidelines range, but the goodness of fit was expected to be

higher. The MBE value is quite low here which indicates that values are quite close between those two models.

4.7. Computing time

All the simulations were done on a personal computer Asus TravelMate-P with Intel Core i7-5500U CPU at 2.40GHz clock speed, 8GB RAM and 64bit operating system. It is obvious that the computation times will depend on the hardware used. Times are given for whole year simulations for one archetype in Table 8. 5R1C simulated all archetypes one after another, and what is interesting the last simulation took twice as much time as the first one. After consultation with software engineer there is a suspicion that the increase in computation times may be due to less memory available after each computation. This may be due to the way how saving results to a file is handled because the memory buffer is not cleared after each *write* method. This issue is not crucial at this point of research but when the future app is considered it may generate problems with the flow of the model.

Table 8 Computing times of the different simulation tool used

Model	Average Time
Energy Plus	~8.41 s
5R1C	~4.55 s
ML Random Forest	~5.58 s

Chapter 5

Conclusions and recommendations

Conclusions from research and results discussion. Recommendations for future researchers.

5.1. Conclusions

In this work different tools were used to simulate the indoor temperature of nine building archetypes. All of the models show potential, but some of them (like Machine Learning) need a bigger amount of training data and more time and effort to properly train and calibrate. Simplified model 5R1C has proven to be a good substitute to more sophisticated model like EnergyPlus. 5R1C showed outstanding performance in result-to-computing-time ratio. Energy Plus gave – as expected – the most accurate results, but the time and the effort that has to be put to configure the model is too high when considered future automatization of model. 5R1C is quite friendly and easy to use even for a newbie when it comes to supplying the data into the model and automatization process. The ease of editing text file is a very bright point of the model. The natural choice when considering automatization is Machine Learning. Random Forest gave good results but not very reliable nor consistent due to shattered sample of data resulting from a small data sample and the need for dividing the sample into train and test samples. ANN would be a great choice, but for that a much bigger sample of data is needed to train the model properly.

For the purpose of RELIABLE project 5R1C is good choice since it produced satisfactory results confirmed by calibration scores MBE and Cv(RMSE). In the future switch to ML or even ANN should be considered under the condition that enough data for training of the models is gathered. The almost twice shorter time of computation of 5R1C model shows that there is potential to conduct nationwide simulation as Portugal has more than 3 million buildings..

5.2. Recommendations

It is recommended to run the simplified model in RELIABLE project and simultaneously collecting data about the buildings over a larger period of time, weather data and training ANN. Over time – sources claim that a batch of 3 years of data is enough to get reliable results from neural network model – ANN would be ready to replace other models. A study solely focused on ANN performance using bigger batch of real data would be highly appreciated in the area on which RELIABLE project is focused. Moreover a database of real in-building measurement data (temperature, humidity, CO₂ levels) from various archetypes would dramatically improve the output of models. Especially the output of ANN model. Same approach with bigger database should be applied to other Machine Learning models to check if they will produce satisfying results. Machine Learning is a very broad area that has much more options to investigate. The tool created can be used not only by municipal services to predict the outcome of indoor temperature during EWE but also as a part of smart home systems to control HVAC equipment,

electronic blinds or systems for plant watering and gain control over energy usage at homes. It can be also used to design or redesign homes to make them more energy efficient.

5.3. Limitations of this work

The time frames in which this research has been conducted was rather short when considered the possibilities of 5R1C and Machine Learning. The results could have been even better if the parameters of the models were assigned much more carefully. For example, individual infiltration values for each archetype – this should be investigated further and measured since it impacts the output. As well as individual thermal capacities should be investigated further. Computing times of 5R1C could have been improved with optimisation of the simulation code in Python. Also, the performance of the tool created could have been checked in other locations since the data for building in Lisbon has been prepared within the RELIABLE project. This would cast some light on the versatility of the model and the ease of adaptation to new localisations.

References

- [1] ODYSSEE-MURE., "Sectoral Profile - Households 2010-2015," <http://www.odyssee-mure.eu>.
- [2] BPIE., "Alleviating fuel poverty in the EU. Investing in Home Renovation, a Sustainable and Inclusive Solution.," 2014.
- [3] EnergyPlus, "EnergyPlus Documentation," 2015. [Online]. Available: https://www.energyplus.net/sites/default/files/docs/site_v8.3.0/GettingStarted/GettingStarted/index.html. [Accessed 10 12 2020].
- [4] O. Mbaabu, "Section," 11 December 2020. [Online]. Available: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>. [Accessed 15 07 2021].
- [5] J. Nduati, "Section," 25 October 2020. [Online]. Available: <https://www.section.io/engineering-education/introduction-to-neural-networks/>. [Accessed 28 July 2021].
- [6] S. Sharma, "Towards Science," 6 September 2017. [Online]. Available: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>. [Accessed 28 July 2021].
- [7] onsetcomp.com, "HOBO Temperature/Relative Humidity Data Logger," HOBO, [Online]. Available: <https://www.onsetcomp.com/products/data-loggers/u12-011/>. [Accessed 12 December 2020].
- [8] J. Ballester, "European seasonal mortality and influenza incidence due to winter temperature variability," *Nature Climate Change*, p. 927–930, 2016.

- [9] T. F. R. J. Southgate, T. Waite, R. Harrell, S. Kovats, A. Bone, Y. Doyle and V. Murray, "Excess Winter Deaths in Europe: a multi-country descriptive analysis," *European Journal of Public Health*, p. 339–345, 2015.
- [10] C. Liddell, C. Morris and H. Thomson, "Excess winter deaths in 30 European countries 1980–2013: a critical review of methods.," *Journal of Public Health, Volume 38, Issue 4. Oxford Academic*, 2016.
- [11] S. G. Simões, V. Gregório and J. Seixas, "Mapping fuel poverty in Portugal.," in *1st Energy Economics Iberian Conference, EEIC | CIEE*, Lisbon, Portugal, 2016.
- [12] R. Oliveira, C. C. Antonio, H. Santos and E. O. Fernandes, "THE PORTUGUESE HOUSING STOCK: AND INITIAL ANALYSIS OF THE NATIONAL ENERGY CERTIFICATION DATABASE," in *Proceedings of the 7th International Conference on Mechanics and Materials in Design*, Albufeira/Portugal, 11-15 June 2017.
- [13] J. D. Healy, "Housing, Fuel Poverty and Health: A Pan-European Analysis.," 2004.
- [14] S. M. Magalhaes and V. M. Leal, "Characterization of thermal performance and nominal heating gap of the residential building stock using the EPBD-derived databases: The case of Portugal mainland.," 2013.
- [15] "An official website of the European Union," [Online]. Available: https://ec.europa.eu/energy/eu-buildings-factsheets-topics-tree/energy-poverty_en. [Accessed 27 June 2021].
- [16] J. Arnold and F. R. C., "Reducing inequality and poverty in Portugal.," *Economics Department Working Papers n° 1258.*, 2015.
- [17] S. P. d. Silva, 28 October 2021. [Online]. Available: <http://www2.insa.pt/sites/INSA/Portugues/AreasCientificas/Epidemiologia/Unidades/UnstrObser/Paginas/ICARO.aspx>.
- [18] M. N. C. S. N. Bulhosa, "The impact of heatwaves on mortality in the Lisbon district - ICARO system revisited," 2019.

- [19] S. Antunes, "Modeling and forecasting the effect of extreme cold in the population health: the basis for the development of a real-time warning system," IPMA, 2014.
- [20] IST-ID, "RELIABLE - REAL-TIME RISK WARNING SYSTEM FOR BUILDING OCCUPANTS DURING EXTREME WEATHER EVENTS," 2021. [Online]. Available: <https://reliable.tecnico.ulisboa.pt/>. [Accessed 27 07 2021].
- [21] R. Bruno, G. Pizzuti and N. Arcuri, "The prediction of thermal loads in building by means of the EN ISO13790 dynamic model: a comparison with TRNSYS," *Energy Procedia* 101, pp. 192-199, 2016.
- [22] O. M.A., Al-Rabghi and K. M. Al-Johani, "Utilizing Transfer Function Method for hourly load calculations," *Energy Convers. Vol. 38*, pp. 319-332, 1997.
- [23] M. Minsky and S. Papert, *Perceptron: An Introduction to Computational Geometry*, 1969.
- [24] Sociedade Portuguesa de Inovação, "Área de Reabilitação Urbana da Vila de Montalegre. Proposta de delimitação," Sociedade Portuguesa de Inovação, Montalegre, 2015.
- [25] C. Sousa-Monteiro, F. Causone, S. Cunha, A. Pina and S. Erba, "Addressing the challenges of public housing retrofits. 9th International Conference on Sustainability in Energy and Buildings," in *9th International conference on sustainability in Energy and Buildings*, Chania, Crete, Greece, 2017.
- [26] S. Claudia, C. Carlos, P. André, Y. Maribel and F. Paulo, "An urban building database (UBD) supporting a smart city information system.," *Energy and Buildings*, vol. 158, pp. 244-260, 2018.
- [27] Y. Cengel, *THERMODYNAMICS: AN ENGINEERING APPROACH*, 2019.
- [28] Google, "SketchUp," 2016. [Online]. Available: <https://www.sketchup.com/>. [Accessed 12 November 2020].

- [29] P. e. a. Jayathissa, "Optimising building net energy demand with dynamic BIPV shading.," *Applied Energy* 202, pp. 726-735, 2017.
- [30] G. R. Ruiz and C. F. Bandera, "Validation of Calibrated Energy Models:," *Energies*, vol. 10, 2017.
- [31] O. M.A.AI-Rabghi, "Utilizing transfer function method for hourly cooling load calculations," *Energy Conversion and Management*, pp. 319-332, 1997.

APPENDIX

Table 9 Input data for 5R1C model (archetypes description)

PERIOD	WINDOW_AREA	WALL_AREA	FLOOR_AREA	ROOM_VOL	TOT_INT_AREA	LIGHTING_LOAD	LIGHTING_control	LIGHTING_UTILISATIO N_FACTOR	LIGHTING_MAINTEN_ FACTOR	U_WALL	U_WINDOW	ACH_VENTILATION	ACH_INFILTRATION	VENT EFFIC	THERM_CAP_PER_FLO OR_AREA	SET_TEMP_HEATING	SET_TEMP_COOLING	WINDOW_SOLAR_TR ANS	WINDOW_LIGHT_TRA NS	AVG_HEIGHT	ROOF_AREA
1919-1945	9.04	120.38	99.72	255.54	228.16	2.00	300.00	0.44	0.90	1.08	4.39	0.50	2.50	0	1367195.44	1.00	90.00	0.837	0.898	2.56	72.50
1946-1960	10.34	124.82	103.02	277.49	244.23	2.00	300.00	0.44	0.90	1.09	3.89	0.50	2.50	0	1348692.54	1.00	90.00	0.837	0.898	2.69	67.90
1961-1970	11.64	114.16	93.47	237.07	301.00	2.00	300.00	0.44	0.90	0.65	4.30	0.50	2.50	0	802160.09	1.00	90.00	0.837	0.898	2.54	97.00
1971-1980	11.03	112.17	91.44	230.78	266.64	2.00	300.00	0.44	0.90	0.44	4.17	0.50	2.50	0	318471.99	1.00	90.00	0.837	0.898	2.52	75.60
1981-1990	17.76	144.26	140.67	380.75	250.64	2.00	300.00	0.44	0.90	0.34	4.08	0.50	2.50	0	292047.33	1.00	90.00	0.837	0.898	2.71	55.04
1991-1995	16.31	137.29	132.70	351.99	334.49	2.00	300.00	0.44	0.90	0.33	3.63	0.50	2.50	0	295008.74	1.00	90.00	0.837	0.898	2.65	110.44
1996-2000	15.64	116.91	117.77	303.69	189.91	2.00	300.00	0.44	0.90	0.44	3.49	0.50	2.50	0	296994.40	1.00	90.00	0.837	0.898	2.58	49.27
2001-2005	16.19	135.32	135.26	348.65	229.32	2.00	300.00	0.44	0.90	0.33	3.50	0.50	2.50	0	290881.15	1.00	90.00	0.837	0.898	2.58	78.24
after 2006	25.74	136.55	141.32	375.71	291.18	2.00	300.00	0.44	0.90	0.28	2.83	0.50	2.50	0	285339.28	1.00	90.00	0.837	0.898	2.66	81.14

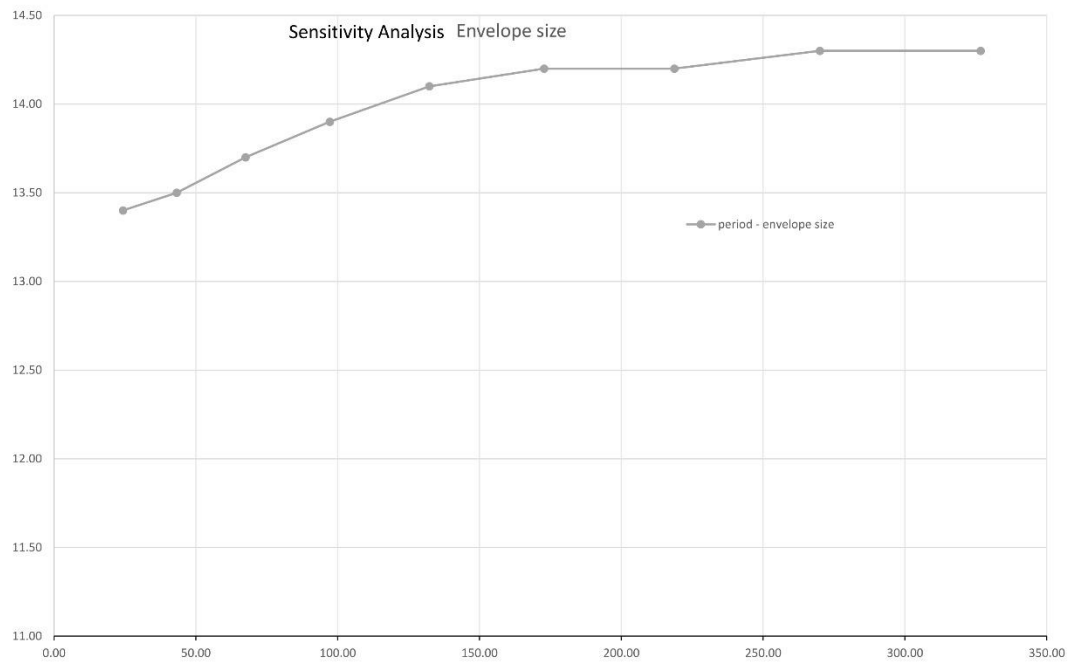


Figure 22 Sensitivity Analysis - Envelope Size

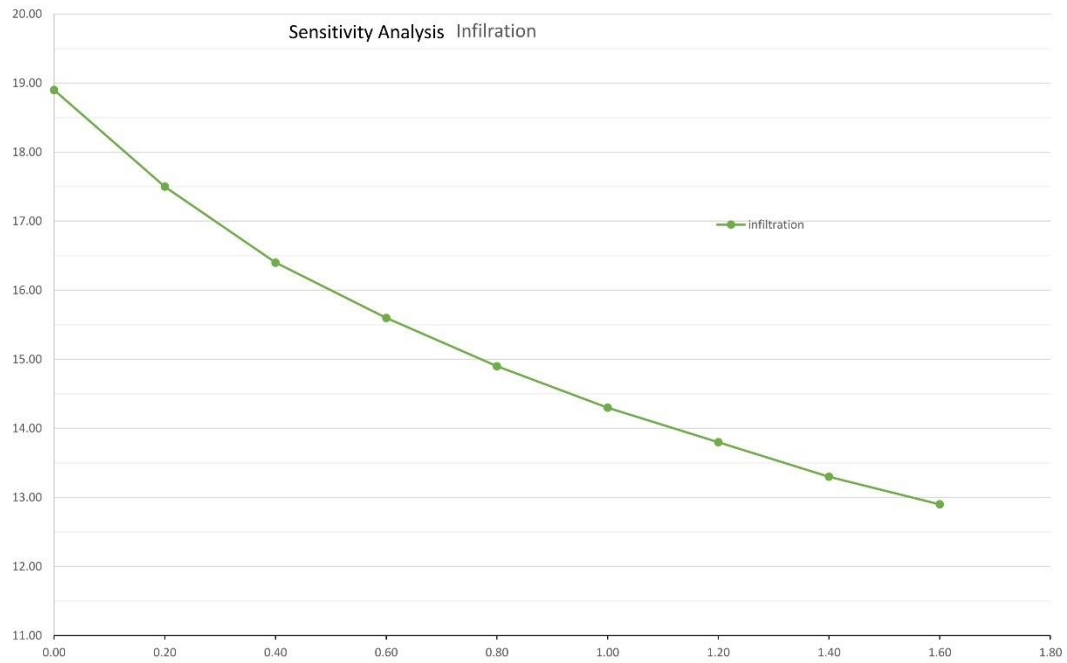


Figure 23 Sensitivity Analysis – Infiltration

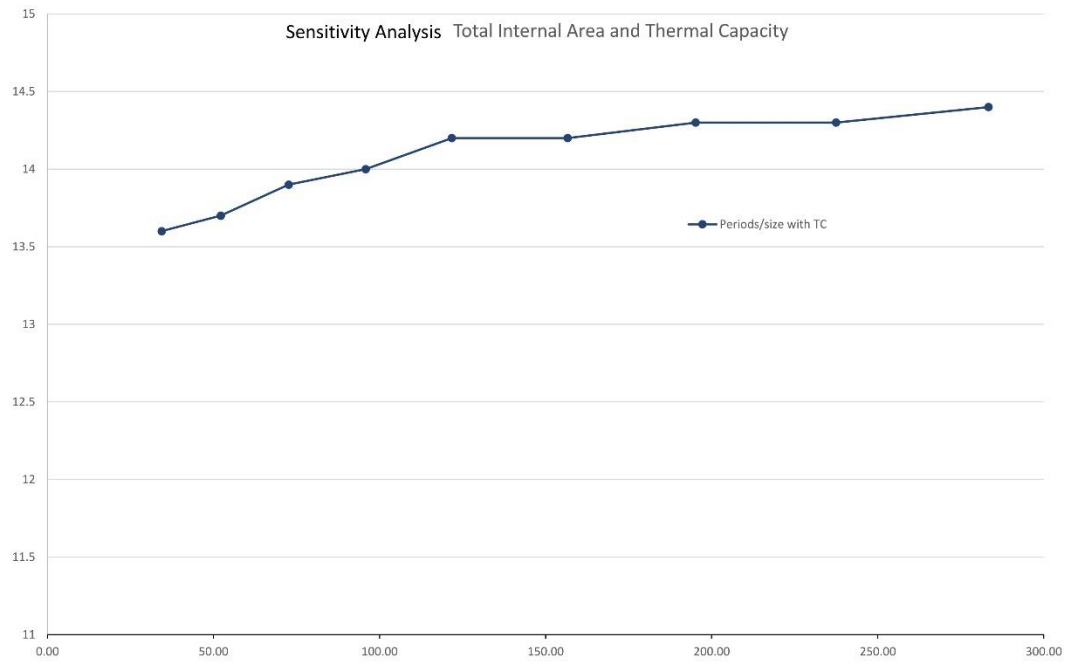


Figure 24 Sensitivity Analysis - Total Internal Area and Thermal Capacity

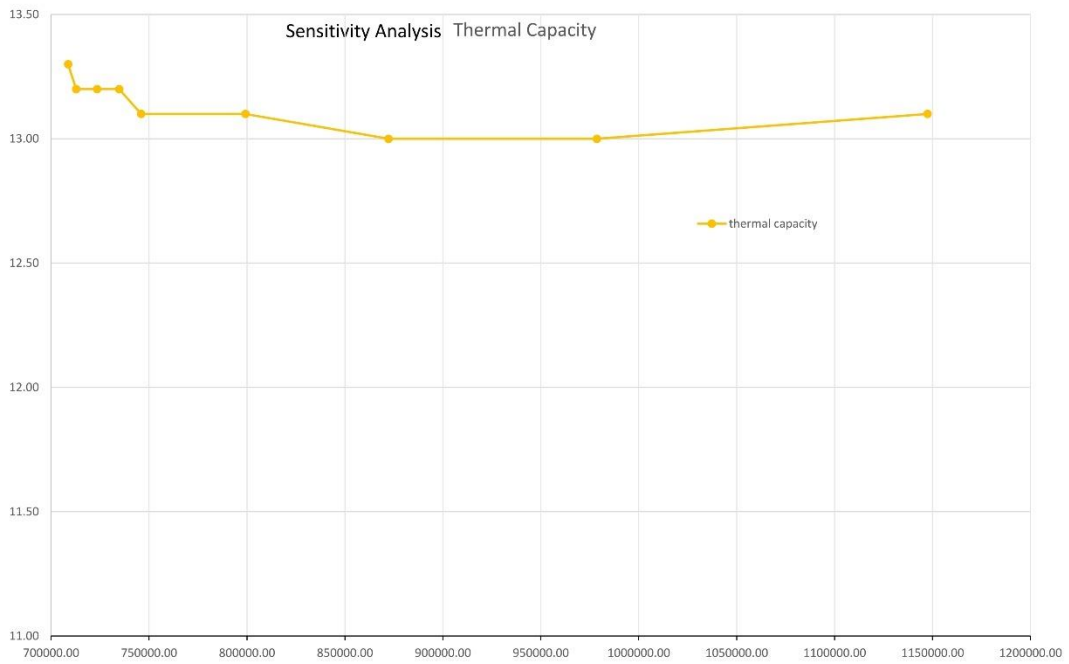


Figure 25 Sensitivity Analysis - Thermal Capacity

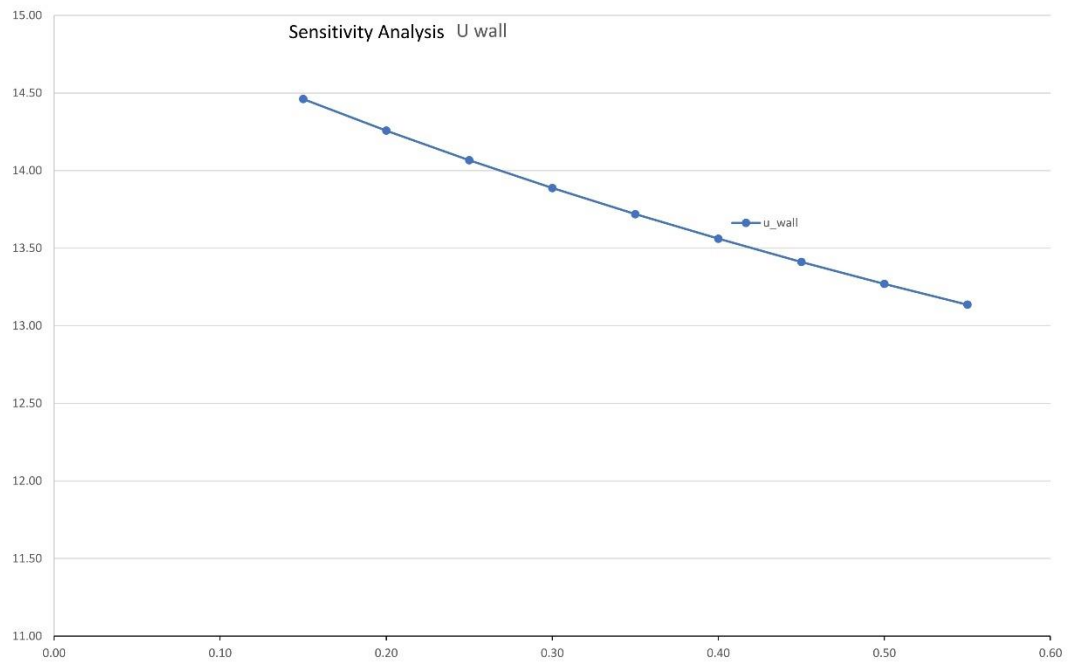


Figure 26 Sensitivity Analysis - U wall

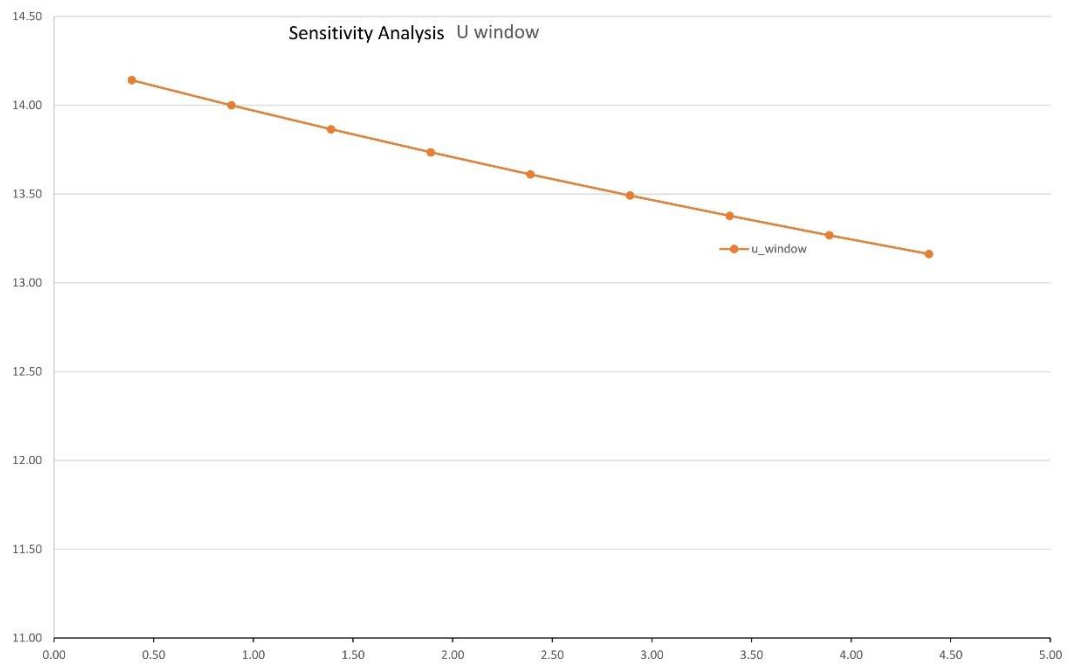


Figure 27 Sensitivity Analysis - U windows

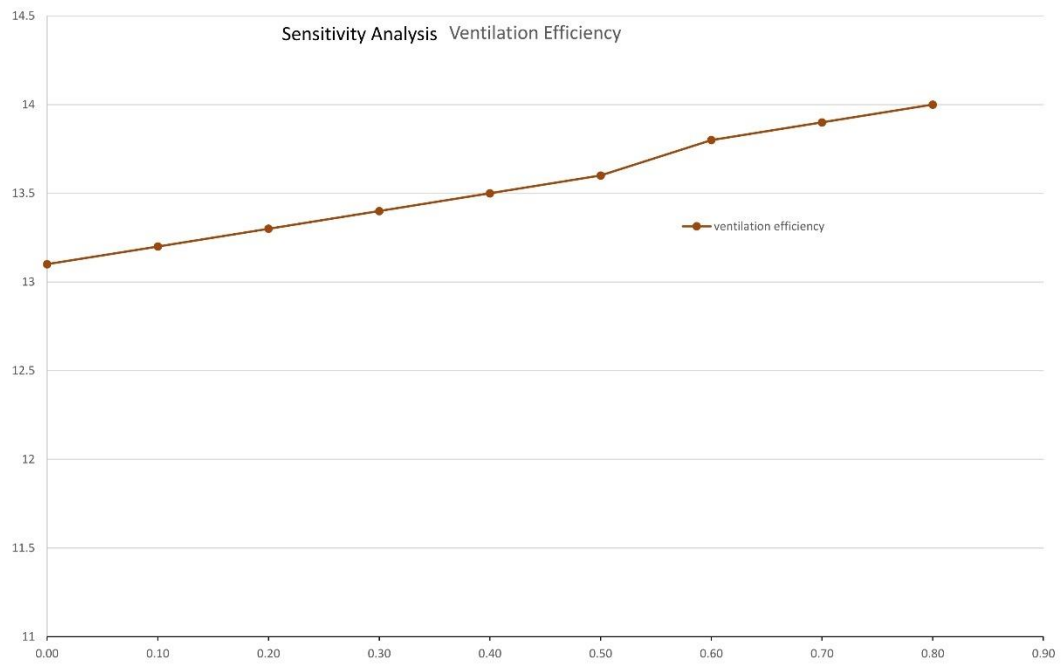


Figure 28 Sensitivity Analysis - Ventilation Efficiency

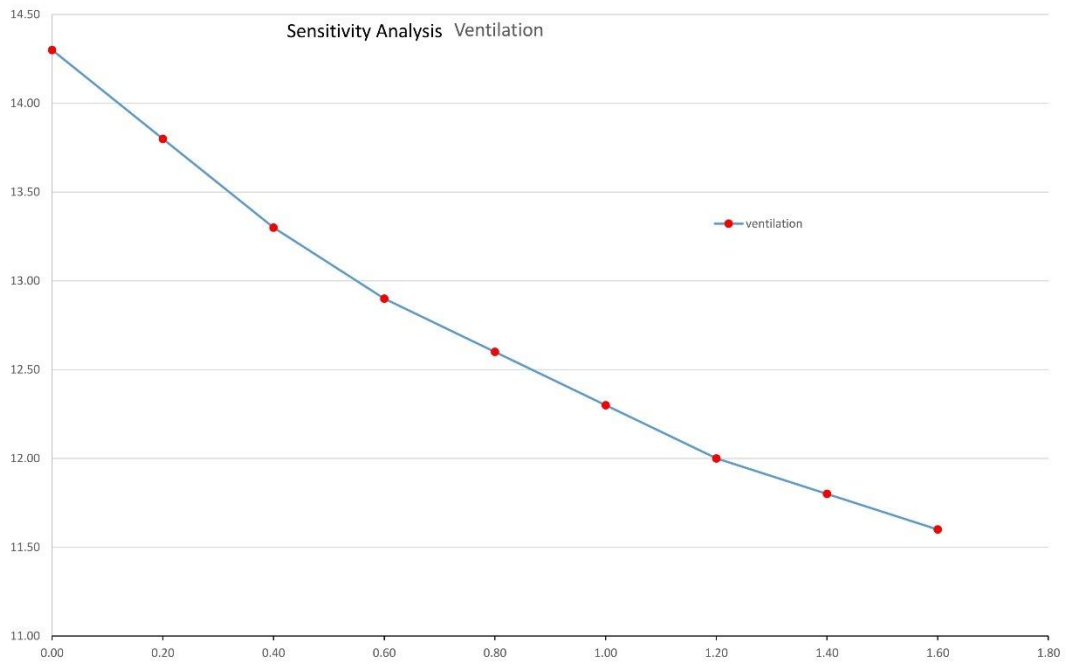


Figure 29 Sensitivity Analysis - Ventilation