Final Master Thesis

MASTER'S DEGREE IN ENERGY ENGINEERING

Machine learning for energy consumption optimization in HVAC systems

MEMORY

Author :	Joan Palat Pañella
University Supervisor :	Dr. Cecilio Angulo Bahón
Company Supervisor :	Enrico Braggion / Josep Fageda Bassagañas
Date :	July, 2018

ETSEIB

Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Abstract

This project aims to optimise energy consumption in Heating, Ventilation and Air Conditioning systems modifying their working schedules. Taking advantage of a machine learning algorithm I have developed (based on Python), schedules can be daily updated depending on multiple variables and guaranteeing comfort. To facilitate the interaction with the algorithm, I have also created a simple website to introduce the data required and to display the results.

The project has been developed in Montrol Open Solutions, SL.

ETSEIB

Contents

List of Figures

List of Tables

ETSEIB

Acronyms

HVAC Heating, Ventilation and Air-Conditioning

ASHRAE American Society Of Heating, Refrigerating And A-C Engineers

MOS Montrol Open Solutions

PLC Programmable Logic Controller

SCADA Supervisory Control And Data Acquisition

IP Internet Protocol

LR Linear Regression

SVR Support Vector Regression

RBF Radial Basis Function

ML Machine Learning

AI Artificial Intelligence

PMV Predicted Mean Vote

MSE Mean Squared Error

PC Personal Computer

I/O Input/Output

UPC Universitat Politècnica de Catalunya

Chapter 1

Introduction

1.1 Motivation

Acceptable ambient conditions are key to guarantee the wellness and productivity of both users and employees. To ensure this, Heating, Ventilation and Air-conditioning (HVAC) systems are installed all over the world. Nowadays, HVAC systems are responsible, approximately, of about a half of the energy consumption in buildings [1], what implies high environmental and economic costs. Therefore, optimising the HVAC systems' functioning should be a 'must' in every facility.

With this goal, many buildings already have automatic control systems to regulate their HVAC equipment [2], but working schedules are barely never modified due to lack of time, tools and knowledge on how to do it. Furthermore, schedules are almost always set in a conservative way, i.e. to ensure comfort conditions in a vast majority of the year. Thus, specially during non-extreme seasons, working schedules could be shortened to reduce energy consumption. The same way, during the coldest/hottests days, working schedules could be enlarged to maintain comfort.

1.2 Objectives

The main goal of this project is to develop a user-friendly tool to permit maintenance technicians to easily optimise the working schedules of their HVAC systems, based on the historical data available. Also, this tool should be capable to be installed in many different buildings without requiring a huge amount of programming hours. To do so, several sub-goals have to be accomplished:

- Understand the automatic control of the HVAC system of a pilot installation.
- Obtain historical data from the pilot installation.
- Pre-process the data to improve its quality (data cleaning).
- Choose a proper predictive model based on heat transfer knowledge, defining the required variables.
- Develop a methodology to predict future values from the current ones.
- Write an algorithm to optimise the working schedules.
- Prepare a user interface to facilitate interaction between the final user and the program.
- Integrate the developed tool with the current control system.
- Check the tool's performance.

1.3 Scope

The scope of the project is to use machine learning techniques to reduce energy consumption and/or improve thermal comfort by developing a generic and flexible tool able to be integrated with HVAC control systems and to be commercialised. So, the tool must be reliable, simple and safe.

The scope of this project includes data monitoring, processing and logging; programming in Python to develop the machine learning algorithm; comparing statistical results to choose the best-fitting predictive model; developing the website in HTML+PHP and integrating the schedules-optimising tool with a Programmable Logic Controller (PLC).

Due to the nature of the data used and to be predicted (mainly temperatures) and the capability to model it as a multi-variable linear equation as explained in Section 2.4, I have based the predictive algorithm on multi-variable regressions. Two types of regressions have been analysed: Linear Regression (LR) and Support Vector Regression (SVR) with Radial Basis Function (RBF) kernel.

All the project is based on the heating installation of a pilot building located in Salt (Girona), which is a public primary school. However, and because of several difficulties, the testing facility is the cooling system of a small room located at my house in Gavà (Barcelona). Testing process includes all the steps required to optimise the schedules, from the PLC programming and sensors' connections to the results' checking.

Assessing the energy savings derived from the implementation of the developed software would require a relatively long period, so it is out of scope of this project.

Chapter 2

Background and State of the Art

2.1 Montrol Open Solutions

Montrol Open Solutions SL (MOS) [3] is a young company founded in July 2016 located in Olot (Girona) that offers monitoring and automatic control services, specially focused on buildings' HVAC systems. The main motivation of MOS is to maximise energy efficiency.

Currently, the company is formed by a multidisciplinary team of 3 people with the following backgrounds:

- Industrial Engineer
- Telecommunications Engineer
- Energy Engineer

Figure 2.1: Montrol Open Solutions SL

2.2 Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that studies the creation of algorithms which allow machines (computers) to create models to predict future behaviours based on historical data [4].

ML is strongly related with statistics and optimisation, and it usually requires a considerable computational power (specially for applications with large amounts of historical data).

Currently, there are used quite a few ML algorithm sets, being some of them [5] [6] [7]:

- Supervised learning regression: Supervised learning regression consists in providing an equation to the machine together with historical data, and letting it just to calculate the coefficients associated to each variable.
- Unsupervised learning: Unsupervised learning consists in giving to the machine a set of variables (features) and outputs (labels) and letting it to freely (without providing any equation) treat the features to build a model to predict the label. One of the main areas of unsupervised learning are Artificial Neural Networks (ANN. ANN try to emulate the behaviour of biological neural networks in computers and machines. In neural networks, between the input data and the output, some hidden layers are found. Data is processed and weighted in each layer until the final output is computed.
- Reinforced learning: Reinforced learning lets the machine to take free/random decisions given a set of values and check the goodness of the results by using a defined cost function. Trying to minimise the cost, the machine itself learns which decisions has to take.

2.3 Thermal comfort

Thermal comfort is the condition of mind that expresses satisfaction with the thermal environment and is assessed by subjective evaluation [8]. It plays a key role in people's general wellness [9] and big amounts of energy are used to achieve it.

Even though thermal comfort is a subjective concept, some methodologies have been developed to evaluate it. The principle one is the known as Predicted Mean Vote (PMV), that has also many variations to adapt it for different conditions and environments.

PMV index predicts the mean response of a larger group of people according to the ASHRAE thermal sense scale [10] (see Table 2.1). There exists many literature about models to compute the thermal sense scale values [9].

Table 2.1: ASHRAE thermal sense scale

+3	Hot
+2	Warm
+1	Slightly warm
0	Neutral
-1	Slightly cool
-2	Cool
-3	Cold

2.4 Temperature evolution modelling

As stated in [11], the temperature heat gain a room/building heated or cooled by water radiators can be expressed as:

$$\dot{Q}_{room} = \dot{Q}_{radiator} + \dot{Q}_{wall} + \dot{Q}_{window} + \dot{Q}_{Sun} + \dot{Q}_{activity} \tag{2.1}$$

$\dot{Q}_{room}$: Room heat exchange

$$\dot{Q}_{room} = \rho_{room} V_{room} C_{v,room} \frac{dT_{room}}{dt} \tag{2.2}$$

ρ_{room}: Room air density
V_{room}: Room volume
$C_{v,room}$: Room air specific heat at constant volume
T_{room}: Room air temperature

$\dot{Q}_{radiator}$: Radiators heat exchange

$$\dot{Q}_{radiator} = U_{o,radiator} F_{radiator} \Delta T_{ml,radiator} S_{radiator} \tag{2.3}$$

$U_{o,hx}$: Radiator heat transfer coefficient
F_{hx}: Radiator correction factor
S_{hx}: Radiator equivalent surface
$\Delta T_{ml,hx}$: Radiator temperature difference logarithmic mean

$$\Delta T_{ml,hx} = \frac{(T_{water,in} - T_{air,out}) - (T_{water,out} - T_{air,in})}{\ln \frac{T_{water,in} - T_{air,out}}{T_{water,out} - T_{air,in}}} \tag{2.4}$$

$T_{water,in}$: Water temperature at the radiator inlet
$T_{air,out}$: Air temperature at the radiator outlet
$T_{water,out}$: Water temperature at the radiator outlet
$T_{air,in}$: Air temperature at the radiator inlet

$\dot{Q}_{wall}$: Walls heat exchange

$$\dot{Q}_{wall} = \frac{T_{ext} - T_{room}}{R_{air,in} + R_{wall} + R_{air,ext}} S_{wall} \tag{2.5}$$

T_{ext}: Exterior air temperature

S_{wall}: Walls equivalent surface

$R_{air,in}$: Interior air convection thermal resistance

$$R_{air,in} = \frac{1}{h_{air,in}} \tag{2.6}$$

$h_{air,in}$: Interior air convection heat transfer coefficient

R_{wall}: Wall conduction thermal resistance

$$R_{wall} = \frac{L_{wall}}{k_{wall}} \tag{2.7}$$

L_{wall}: Wall thickness k_{wall}: Wall conduction heat transfer coefficient

$R_{air,ext}$: Exterior air convection thermal resistance

$$R_{air,out} = \frac{1}{h_{air,out}} \tag{2.8}$$

$\dot{Q}_{window}$: Windows heat exchange

$$\dot{Q}_{window} = \frac{T_{ext} - T_{room}}{R_{air,in} + R_{window} + R_{air,out}} S_{window} \tag{2.9}$$

S_{window}: Window equivalent surface

R_{window}: Window conduction thermal resistance

$$R_{window} = \frac{L_{window}}{k_{window}} \tag{2.10}$$

L_{window}: Window thickness k_{window}: Window conduction heat transfer coefficient

$\dot{Q}_{Sun}$: Sun heat exchange. It is a function of many parameters, being the main ones the geographical position (latitude and longitude), solar position (altitude and azimuth) and sky state.

$\dot{Q}_{activity}$: Activity (humans and devices) heat exchange. To keep it simple, it can be expressed as a function of the room's occupation.

Substituting in Equation 2.1, grouping all the constants (and assuming thermal coefficients as constants, even though they are actually not), the resulting expression to model room temperature is:

$$a\frac{dT_{room}}{dt} = b\Delta T_{ml} + c(T_{ext} - T_{room}) + d(T_{ext} - T_{room}) + e(SolarPos) + f(Occ) \tag{2.11}$$

Dividing by coefficient 'a' (to isolate $\frac{dT_{room}}{dt}$) and accordingly defining new coefficients, the final expression is:

$$\frac{dT_{room}}{dt} = B\Delta T_{ml} + C(T_{ext} - T_{room}) + D(T_{ext} - T_{room}) + E(SolarPos) + F(Occ) \quad (2.12)$$

Chapter 3

Pilot installation

3.1 Overview

The selected building to study its temperature evolution and apply a schedule's optimisation tool is a primary school located in Salt (Girona).

The school works from 9 a.m. to 5 p.m. on workdays, and is closed on weekends. This winter, the school has had some issues to keep the temperature at comfort levels (specially on Mondays' mornings) and has been forced to close at least once due to low temperatures inside the classrooms. To solve this problem, heating system's schedule has been extended, increasing energy costs.

Figure 3.1: Pilot installation image

3.2 Heating system

For heating purposes, the school is divided in two big areas: one oriented towards north and another towards south.

The heating system is composed by a water boiler and two levels of water circuits. The first level corresponds to the primary water circuit, which goes trough the boiler and is heated up to the boiler set-point. The second level corresponds to the secondary water circuits and is split into two circuits, one for each area (north and south) that bring the hot water to the radiators that heat the rooms.

Water from the supply flow of the primary circuit is mixed with water from the return flow of each of the secondary circuits using 3-ways valves to achieve the desired flow temperature.

Lastly, two water pumps are located in parallel in each circuit to move the water. These devices rotate weekly.

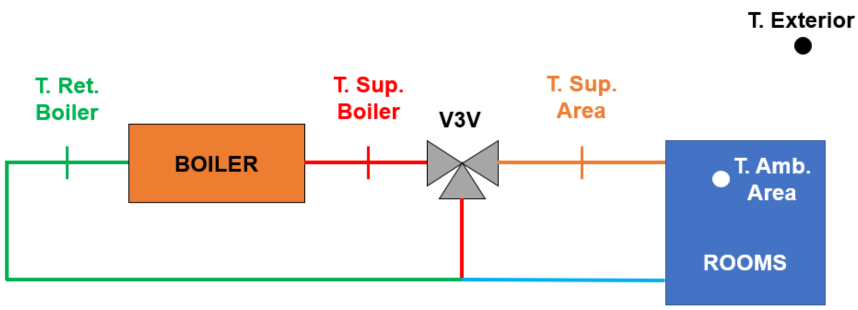


Figure 3.2: Heating system scheme

3.3 Control system

3.3.1 Architecture

The architecture of the control system is simple: the field elements (sensors and actuators) are connected to the Inputs/Outputs (I/O) module of the (PLC), which is located in a panel near the boiler's room. This I/O module is then connected via communications bus with the communications module, that is connected via Ethernet with the router, which sends and receives the information via Internet Protocol (IP) and some proprietary protocol to the Supervisory Control And Data Acquisition (SCADA) located in a remote PC.

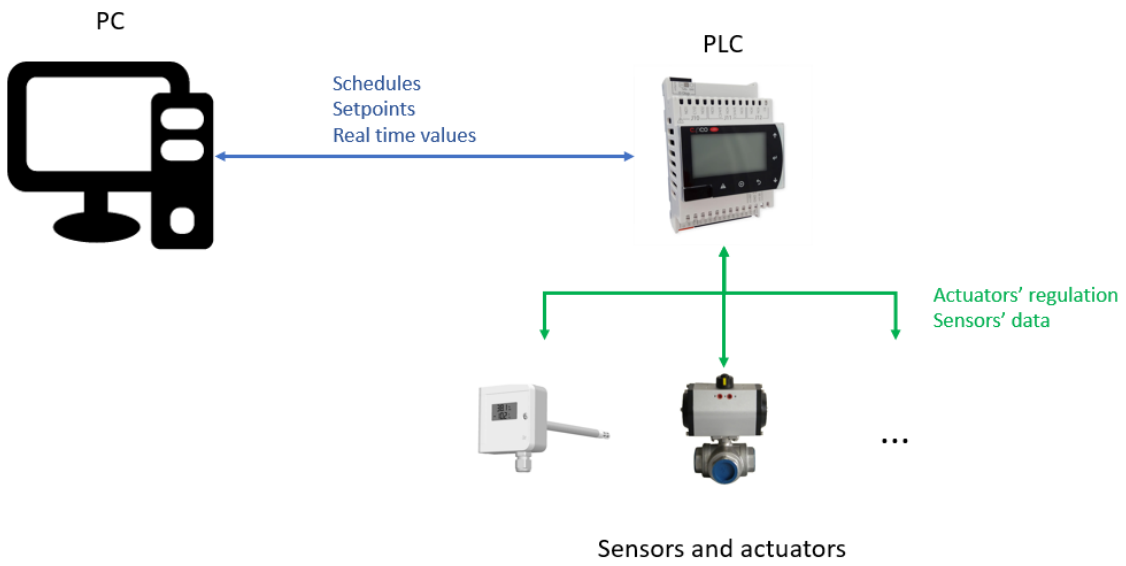


Figure 3.3: Architecture scheme

3.3.2 Field components

As introduced in subsection 3.3.1, field elements composing the control system are the controllers, the sensors and the actuators. In this case, these are the following ones:

PLC: I/O module, Net module.

Sensors: Water temperature probes and air temperature probes.

Actuators: 3-ways valve actuator, pump's motor, boiler's burner.

3.3.3 Algorithm

The goal of the control system is to maintain the rooms' temperature above a minimum temperature. To do so, when the working schedule is ON, the boiler's burner is regulated to achieve a primary circuit return temperature equal to 60°. Then, depending on the area's ambient and the exterior temperature, the secondary circuit supply temperature set-point is computed as follows:

The 3-ways valve actuators regulate to maintain the secondary circuits' supply temperature as close as possible to the set-point to heat (or not) the rooms. When the set-point is below 60°, the valves are forced to be closed.

Chapter 4

Temperature prediction model

4.1 Available data

In the controller, each variable is stored in a separate csv file that must be downloaded manually. Data is recorded every 15 min, and comprises the following variables:

- Secondary circuit supply temperature setpoint (North area) [°C]
- Secondary circuit supply temperature setpoint (South area) [°C]
- Boiler's burner regulation [%]
- 3-ways valve regulation (North area) [%]
- 3-ways valve regulation (South area) [%]
- Ambient temperature (North area) [°C]
- Ambient temperature (South area) [°C]
- Exterior temperature [°C]
- Boiler's supply temperature [°C]
- Secondary circuit supply temperature (North area) [°C]
- Secondary circuit supply temperature (South area) [°C]

4.2 Model

To predict temperature evolution, the model considered is based on multi-variable regressions and following Equation 2.12. To keep the physical meaning of the coefficients, what is predicted are ambient temperature differentials. This way, each regression coefficient is related to one type of heat transfer. Furthermore, the differential between a time t and t-1 must be predicted with only the t-1 data.

Data pre-processing is required to compute the temperature differentials of the historical data and to add some variables that are not recorded directly, such as solar position (to allow irradiance estimations) and building's occupation.

With respect to the model proposed in Section 2.4, some simplifications have been introduced:

- Due to the lack of water return temperature data, ΔT_{ml} in the radiators is assumed to be equal to the temperature difference between the secondary circuit supply temperature and the ambient temperature. $\Delta T_{ml} \approx T_{water,sup} - T_{amb}$
- Wall and windows heat transfer has been mixed together in the same coefficient, as the associated variable is the same $(T_{ext} - T_{amb})$
- Heat gain from the Sun is assumed to be independent of the sky state. This is surely not a good assumption, but there is no information on the sky state of the historical data.
- Solar position has been divided in two variables with their own coefficient: altitude and azimuth. This way, theoretically, building's orientation is better described.
- Building's occupation is assumed to be a boolean variable (1 for occupied building and 0 for unoccupied building). Building is assumed to be occupied on weekdays from 9 a.m. to 5 p.m. Therefore, in the model is not considered partial occupation.

So, Equation 2.12 can be expressed as:

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + E_1(\sin(Altitude)) + E_2(\cos(Azimuth)) + F(Occ) \quad (4.1)$$

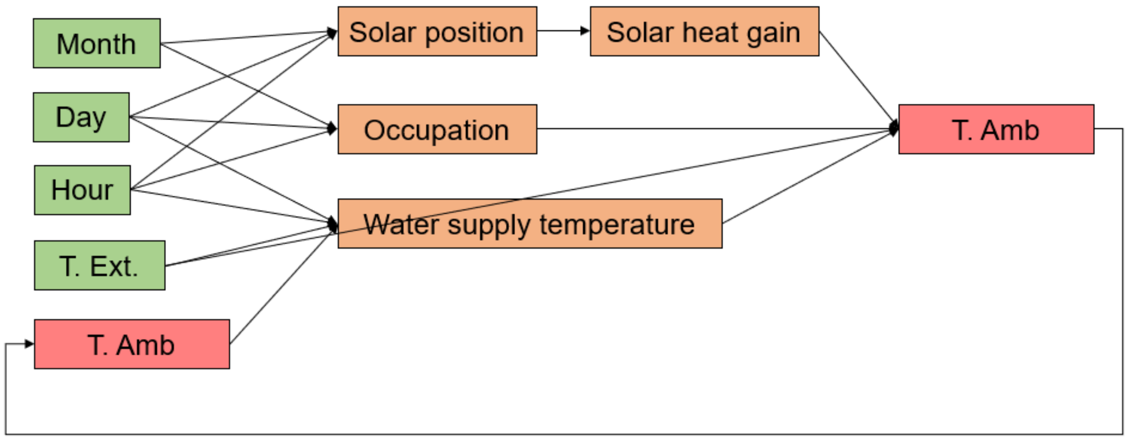


Figure 4.1: Variables scheme

Variables scheme used is summarised in Figure 4.1

4.2.1 Training and testing

Available data has been split in two big groups: training data and testing data. Training data is used to compute the coefficients and to construct the model, and testing data is used to evaluate the model's performance. In the present case, testing data has been set as the last 48h of data. This way of splitting the data is useful to test the model behaviour against what can be considered as the most similar day compared to the one it has to be predicted.

From all the available variables, not all the theoretically identified might be useful. Therefore, a study comparing various featuresâĂŹ combinations has been performed.

Also, old data is not necessarily useful to predict the temperature evolution of the building in the near future. For example, data from February might not be useful at all to predict a day in May. To prove this, the effect of setting a initial cutting date to neglect all previous data has been studied.

Lastly, two different ways of performing multi-variable regressions have been tested: linear regression and Support Vector Machines with 'rbf' kernel.

To choose the best model among all the exposed ones, mean Squared Error (MSE) when predicting the temperature evolution of the last 24-48h of the real data available (corresponding to 7th and 8th May 2018). As is proven in refer section 4.2.3, the error when predicting the ambient temperature differential presents a normal distribution around a mean value of 0. Therefore, it is easy to define the confidence interval of the prediction given the standard deviation of the error. Taking into account that is error is propagated in every prediction, the resulting expression is:

$$T_{amb,h} = T_{amb,h-1} + dT_{amb} \pm \Delta T_{amb} \tag{4.2}$$

For a 95% confidence interval:

$$\Delta T_{amb} = 2\sigma_{err,h} \tag{4.3}$$

$$\sigma_{err,h} = \sqrt{\sigma^2_{err,h-1} + \sigma^2_{err}} \tag{4.4}$$

Therefore

$$T_{amb,h} = T_{amb,h-1} + dT_{amb} \pm 2\sqrt{\sigma^2_{err,h-1} + \sigma^2_{err}} \tag{4.5}$$

4.2.2 Features

Features' combinations considered have been (and according to what exposed in Section 2.4 and Section 4.2):

- Combination 1:

  Secondary circuit supply temperature - Ambient temperature

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) \tag{4.6}$$

- Combination 2:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) \tag{4.7}$$

- Combination 3:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

  Sine of the solar altitude

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + E_1(\sin(Altitude)) \tag{4.8}$$

- Combination 4:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

  Sine of the solar altitude

  Cosine of the solar azimuth

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + E_1(\sin(Altitude)) + E_2(\cos(Azimuth)) \tag{4.9}$$

- Combination 5:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

  Sine of the solar altitude

  Cosine of the solar azimuth

  Occupation

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + E_1(\sin(Altitude)) + E_2(\cos(Azimuth)) + F(Occ) \tag{4.10}$$

- Combination 6:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

  Sine of the solar altitude

Occupation

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + E_1(\sin(Altitude)) + F(Occ) \quad (4.11)$$

- Combination 7:

  Secondary circuit supply temperature - Ambient temperature

  Exterior temperature - Ambient temperature

  Occupation

$$\frac{dT_{room}}{dt} = B(T_{water,sup} - T_{amb}) + CD(T_{ext} - T_{amb}) + F(Occ) \quad (4.12)$$

4.2.3 Results

Table 4.1: North area: MSE

	24h		48h	
	LR	SVR	LR	SVR
Comb. 1	0.43	0.43	0.82	0.86
Comb. 2	0.52	**0.25**	0.72	0.75
Comb. 3	0.64	0.26	0.82	**0.63**
Comb. 4	0.42	0.47	0.71	0.81
Comb. 5	**0.31**	0.29	0.58	0.82
Comb. 6	0.41	0.36	0.59	0.81
Comb. 7	**0.31**	0.42	**0.40**	0.67

Analysing the results, the best model to predict temperature evolution for the last 48h in the North area is, by far, LR with features' combination 7 (see Figure 4.2). For the South area, however, the best model is LR with features' combination 6 (see Figure 4.5). Complete results (and plots) are found in Annex I.

Optimal regression coefficients are summarised in Table 4.5:

Figure 4.4 and Figure 4.5 prove that the error made when predicting ambient temperature differentials presents a normal distribution around a mean value of 0 (as stated in subsection 4.2.1).

So, it is concluded that:

Table 4.2: North area: Data cutting date

	24h		48h	
	LR	SVR	LR	SVR
Comb. 1	05/03/2018 2:30	04/03/2018 17:45	05/03/2018 3:30	04/03/2018 13:45
Comb. 2	04/03/2018 16:00	05/03/2018 1:15	05/03/2018 3:15	04/03/2018 19:30
Comb. 3	05/03/2018 5:15	04/03/2018 14:45	28/02/2018 2:00	04/03/2018 14:15
Comb. 4	24/02/2018 11:30	04/03/2018 17:45	28/02/2018 2:00	04/03/2018 17:30
Comb. 5	24/02/2018 21:15	26/02/2018 23:15	28/02/2018 4:15	04/03/2018 17:15
Comb. 6	28/02/2018 3:45	28/02/2018 22:00	28/02/2018 4:15	04/03/2018 17:45
Comb. 7	01/03/2018 16:00	04/03/2018 17:30	01/03/2018 16:00	04/03/2018 19:30

Table 4.3: South area: MSE

	24h		48h	
	LR	SVR	LR	SVR
Comb. 1	0.24	0.25	0.27	0.28
Comb. 2	0.19	0.28	0.24	0.28
Comb. 3	0.15	0.12	0.21	0.22
Comb. 4	0.15	0.12	0.22	0.22
Comb. 5	0.06	**0.02**	0.11	0.15
Comb. 6	**0.04**	0.03	**0.10**	**0.13**
Comb. 7	0.09	**0.02**	0.13	0.18

- All prediction models work better for the South area rather than for the North one.
- LR and SVR present similar accuracy results, being LR slightly better for longer-term predictions.
- Adding more variables does not always increase accuracy (but does not neither decrease it neither). This can be caused by the fact that both solar altitude, azimuth and building occupation are strongly related to the hour, so taking into account all three variables may result in redundancies.
- The optimal amount of data to train the model corresponds, approximately and for all models, to 2 months before the date to be predicted. Less data make the models imprecise, but older data only disturbs the results.

Table 4.4: South table: Data cutting date

	24h		48h	
	LR	SVR	LR	SVR
Comb. 1	05/03/2018 2:30	04/03/2018 12:15	05/03/2018 3:15	01/03/2018 16:45
Comb. 2	05/03/2018 8:30	28/02/2018 14:45	05/03/2018 2:45	05/03/2018 1:00
Comb. 3	05/03/2018 4:30	04/03/2018 11:45	05/03/2018 5:15	04/03/2018 18:00
Comb. 4	05/03/2018 4:30	01/03/2018 16:15	05/03/2018 3:30	04/03/2018 18:00
Comb. 5	05/03/2018 6:15	28/02/2018 10:45	26/02/2018 22:45	01/03/2018 10:30
Comb. 6	05/03/2018 6:30	01/03/2018 10:15	26/02/2018 12:15	25/02/2018 11:00
Comb. 7	28/02/2018 15:45	04/03/2018 16:45	28/02/2018 8:30	04/03/2018 18:00

Table 4.5: Regression best-fitting coefficients

	North	South
B	0.01013563	0.00270163
CD	0.00703451	0.00398531
E_1	-	-0.02169058
E_2	-	-
F	0.0664441	0.05828716

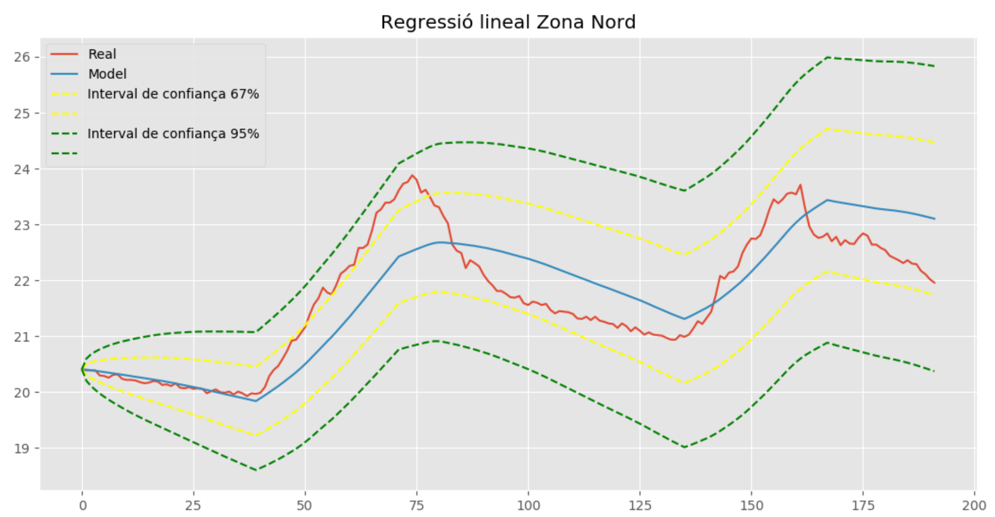


Figure 4.2: 48h temperature evolution prediction. North area.

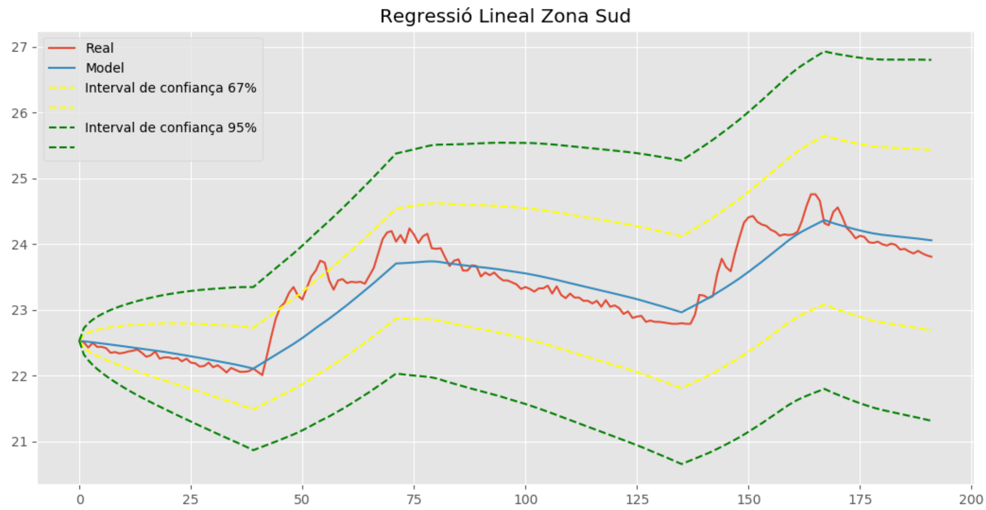


Figure 4.3: 48h temperature evolution prediction. South area.

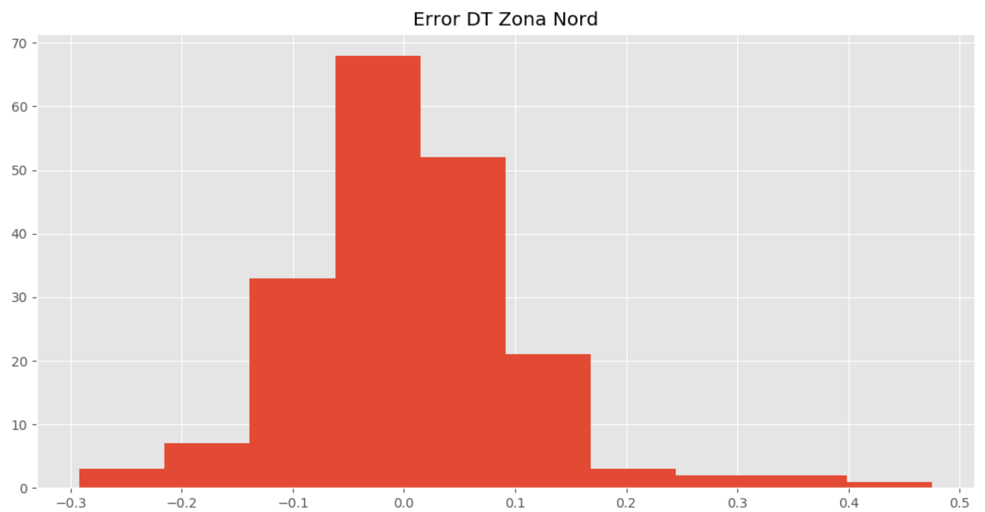


Figure 4.4: Prediction error histogram. North area.

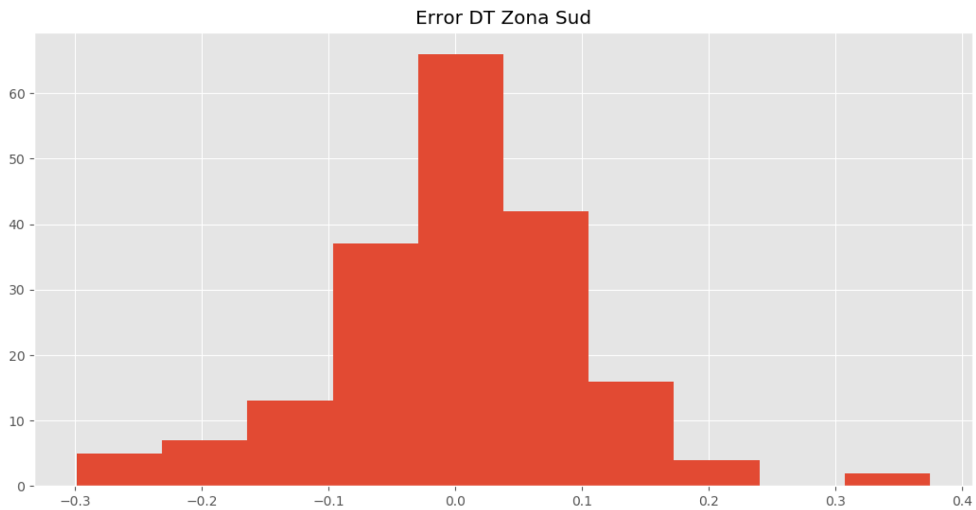


Figure 4.5: Prediction error histogram. South area.

Chapter 5

Optimisation algorithm

5.1 Algorithm description

The optimisation algorithm consists in, basically, an iterative temperature prediction process.

The required input data are:

- Current temperature
- Temperature setpoint
- Initial date and time
- Final date and time

The algorithm can be divided in two sub-processes, one computing the optimal starting date and time for the installation, and the other computing the optimal ending date and time.

To optimise starting date and time, and for energy efficiency purposes, the initial guess is that starting date/time is equal to the initial date/time introduced by the user (i.e. if the user wants thermal comfort between 9 a.m. and 5 p.m., first iteration is performed assuming the heating installation starts working exactly at 9 a.m.). Iterations' procedure is:

1. Iteration starting date/time = Initial date/time
2. Get current date/time
3. Get current ambient temperature
4. Predict temperature at initial date/time
5. (a) If initial ambient temperature > setpoint:
   Optimal starting date/time = Iteration starting date/time

   (b) Else:
   Iteration starting date/time = Iteration starting date/time - 15min
   Back to point 4

On contrary, to optimise heating shut down, the initial guess is that the ending date/time for the heating installation is equal to the initial date/time (i.e. ending date/time equal to 9 a.m., following the previous example).

1. Iteration ending date/time = Initial date/time
2. Get current date/time
3. Get current ambient temperature
4. Predict temperature at final date/time
5. (a) If ambient temperature is always > setpoint:
   Optimal ending date/time = Iteration ending date/time

   (b) Else:
   Iteration ending date/time = Iteration ending date/time + 15min
   Back to point 4

5.2 Optimisation results

Optimisation results have been evaluated by predicting the optimal working schedule of the heating installation for the 5th June 2018. Despite not being the best moment to analyse a heating system, and despite not having connection with the field elements, some valid results regarding the performance of the algorithm could be extracted.

To perform the simulation, I have assumed that, the day before (4th June 2018) at 6 p.m., the maintenance technician decides to optimise the schedule for the day after. Setting a temperature setpoint of 24.0°C and estimating a current ambient temperature of 22.0°C in both North and South areas, to be above the setpoint between 9 a.m. and 5 p.m., the optimal heating schedule would be from 6:45 a.m. to 9 a.m. in the North area and from 2:15 a.m. to 9 a.m. in the South area.

Comparing the optimal schedule results with the normal schedule, in the North area, energy savings are estimated in a 10% for this precise day. In the South area, energy consumption would be higher, but thermal comfort would be better. Main results are summarised in Table 5.1.

Table 5.1: Optimisation results

	North	South
Optimal schedule	6:45 a.m. - 9 a.m.	2:30 a.m. - 9 a.m.
Temp. at 9 a.m. (opt. sch.) [°C]	24.1	24.0
Temp. at 9 a.m. (std. sch.) [°C]	24.5	23.1
Temp. at 5 p.m. (opt. sch.) [°C]	25.5	25.2
Temp. at 5 p.m. (std. sch.) [°C]	25.8	25.2

Chapter 6

Tool integration

6.1 Communications

The optimisation tool is thought to be installed in a PC connected via TCP/IP with the PLC. Data is sent and received to/from the PLC using a communications standard protocol, being Modbus TCP the one used in this project (running on port 502).

Modbus function used has been Holding Registers (what allows reading and writing data).

Modbus adresses have been defined as follows:

- Adress 0: Enable optimal schedule
- Adress 1: Optimal initial minute
- Adress 2: Optimal initial hour
- Adress 3: Optimal initial day
- Adress 4: Optimal initial month
- Adress 5: Optimal initial year
- Adress 6: Optimal final minute
- Adress 7: Optimal final hour

- Adress 8: Optimal final day
- Adress 9: Optimal final month
- Adress 10: Optimal final year
- Adress 11-12: Current ambient temperature

6.2 Aemet API

The meteorological prediction is obtained taking advantage of the API created by the Spanish Meteorological Agency (AEMET) [12], that provides a town-based hourly prediction for the 48h after the query.

To interact with the API, a personal alphanumeric key is required. The procedure to obtain it is really easy and it only involves mail registration and validation.

Lastly, to request town-based predictions, the town's code has to be known (or consulted in a database). The two towns considered in this project (Salt and Gavà) have the codes presented in Table 6.1:

Table 6.1: AEMET API town codes

Town	Code
Salt	17155
Gavá	08089

6.3 User Interface

The user interface has been developed as a webpage programmed with HTTP and PHP, and runs over a local server (based on Apache).

The webpage is very simple, and it contains 3 pages:

- Login (see Figure 6.1): To initiate user's session. If the user and password match the defined ones, redirects to page 'Input data'.
- Input data (see Figure 6.2): To allow the introduction of data by the user. If all data is introduced properly, optimisation algorithm can be ran clicking on the button 'Launch Calculations'.
- Results (see Figure 7.1): To show the optimisation results. Once the optimisation algorithm has finished working, results are displayed. Then, the computed schedule can be sent to the PLC clicking on the button 'Send Data'.

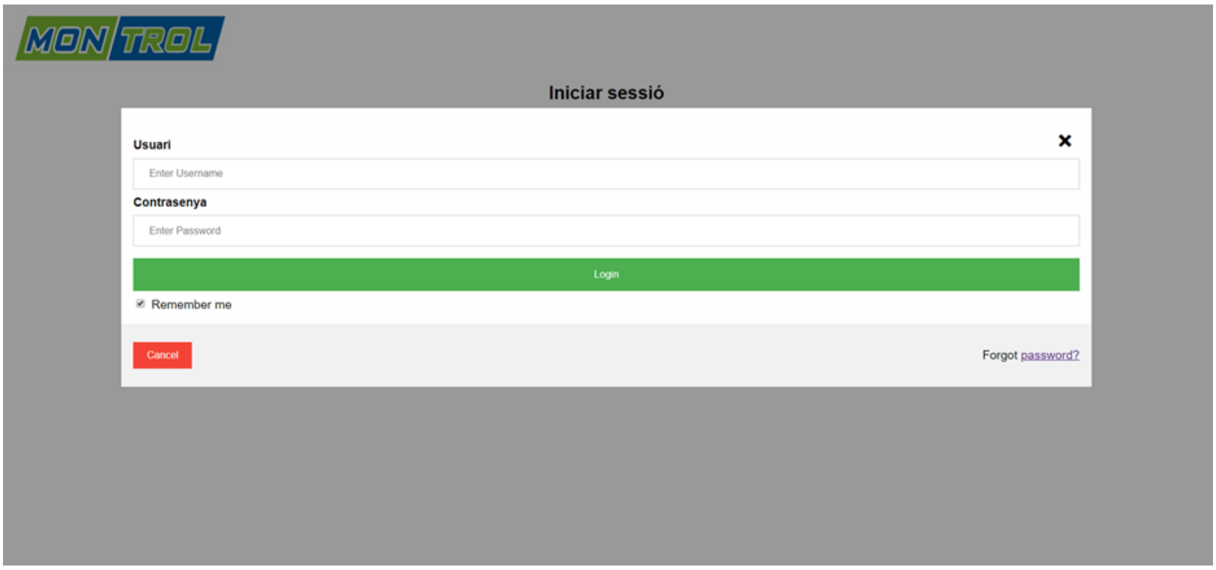


Figure 6.1: User Interface. Login.

Figure 6.2: User Interface. Data input.

Resum dades introduïdes

	Zona Nord	Zona Sud
Consigna temperatura [ºC]	24	24
Hora inici	2018-06-05T09:00	2018-06-05T09:00
Hora fi	2018-06-05T17:00	2018-06-05T17:00

Resum resultats

Enviar resultats

	Zona Nord	Zona Sud
Hora òptima encesa	2018-06-05 06:45:00	2018-06-05 02:30:00
Temperatura encesa horari òptim [ºC]	24.1	24.0
Temperatura encesa horari estàndar [ºC]	24.5	23.1
Hora òptima apagada	2018-06-05 09:00:00	2018-06-05 09:00:00
Temperatura apagada horari òptim [ºC]	25.5	25.2
Temperatura apagada horari estàndar [ºC]	25.8	25.2
Estalvi estimat [%]	10.0	-

Figure 6.3: User Interface. Results.

Chapter 7

Test installation

7.1 Overview

This test installation has been added to fully prove the tool, as testing was not feasible in the pilot installation due to several reasons. The installation consists on a small room with an air-conditioning and heating machine. The device is controlled by a thermostat, that sends an ON/OFF signal to the refrigerant compressor and to the air fan. The temperature setpoint is changed by a potentiometer, and Heat/Cool mode is chosen by a switch. Empirically, I have seen that the hysteresis band of the control is equal to 2.5°C.

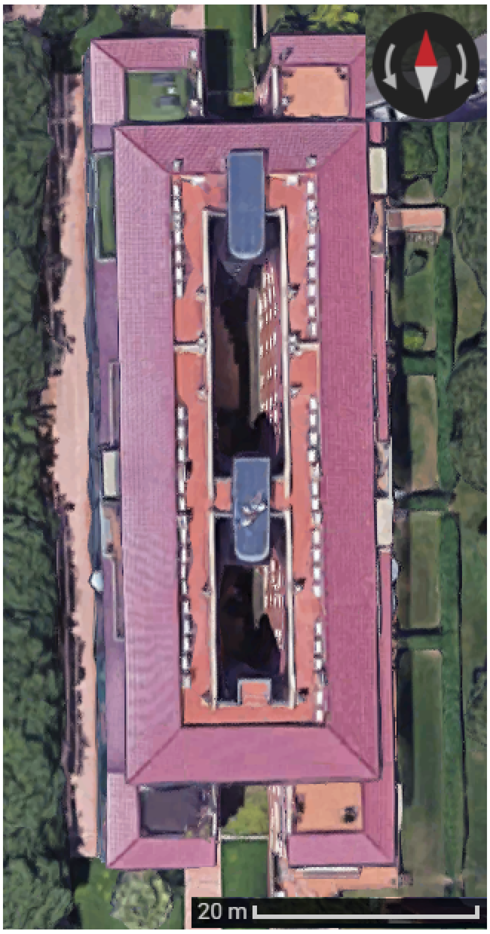


Figure 7.1: Test installation image

7.2 Prediction model

The theoretical equation to model temperature is almost identical to the one exposed in Section 2.4 and Section 4.2 but with one difference has to be mentioned: in this installation, instead of radiators, room air is cooled/heated by direct mixing with the discharge air coming from the split. However, the resulting best-fitting equation is:

$$\frac{dT_{room}}{dt} = B(T_{discharge} - T_{room}) + F(Occ) \tag{7.1}$$

Exterior temperature had also been recorded, but results when taking into account were absurd and, thus, discarded.

7.3 Results

Temperature modelling results for this test installation have not been really meaningful due to the small amount of data available. A lot of data had to be erased because an error I made when programming the PLC, that resulted in non-valid values.

However, the purpose of this test installation was to prove the integration of the schedules-optimisation algorithm with the PLC, what has been a success. Data can be read and written to the PLC and schedules can be updated according to the values computed by the software.

Chapter 8

Economic study

8.1 Costs

8.1.1 Material cost

Material costs are highly dependent on the existing elements in the facility (if any) and the number of control points required.

In the best case, meaning that all required sensors are already present in the installation and that PLCs are able to communicate via a standard communications protocol, material costs could be null.

For all other cases, cost of the hardware elements would be:

- PC: 500.00 €/unit
- PLC: 351.00 €/unit
- Exterior air temperature probe: 33.00 €/unit
- Ambient air temperature probe: 22.50 €/unit
- Duct air temperature probe: 53.50 €/unit
- Duct water temperature probe: 92.00 €/unit

- Solar irradiance sensor: 138.80 €/unit

Also, distribution panel connection and mounting (if required) and installation hours must be accounted, but they are strongly case-dependent and difficult to extrapolate from a building to another.

8.1.2 Engineering costs

Engineering costs include adapting the program to each facility, studying and choosing the best features' combinations to predict temperatures, installing all software required and integrating the schedules' optimiser tool with the HVAC control system.

All this is estimated to last for 8h, what is translated into 400 €.

8.2 Savings

Economic savings have not been studied in any real installation, as it was out of the scope of the project. However, I have performed some predictions.

For a building where only engineering costs were required (that would mean that there is already a HVAC control system working and, thus, further energy savings would come from optimising schedules), to achieve a payback of 6 months, yearly savings should be higher than the ones shown in Table 8.1 (energy prices have been obtained from [13] and [14]).

Table 8.1: Required savings

	Price [€/kWh]	Required savings [kWh/year]
Natural gas	0.05	16000
Electricity	0.13	6154

For a natural gas boiler like the one present at the pilot installation, with a power of 100 kW, assuming it is working 6 months/year during an average of 59h 45min per week (Mondays from 4:15 to 17:00 and from Tuesday to Friday from 5:15 to 17:00) and estimating an efficiency equal to 90%, saving 16000 kWh/year means a gas consumption reduction equal to a 10%.

Chapter 9

Environmental impact

9.1 Cost

To fully assess environmental costs, a Life Cycle Assessment should be performed for all the elements required to properly run the optimisation tool, but this is out of scope of the project.

Despite this, considering that the project's goal is just the developing of the software tool, and assuming that the PC where it would be installed would be running whether or not this software is implemented, environmental costs can be considered as null.

9.2 Savings

Environmental 'savings' (understanding savings as avoided emissions), are directly related to the energy savings achieved by optimising schedules.

Assuming that the software will only be installed if the payback goal of 6 months is achieved (see Section 8.2), CO_2 equivalent emissions' savings would be, at least, the figures displayed in Table 9.1. Specific emissions for electricity have been assumed to be equal to 392 gCO_2/kWh [15]. Specific emissions for natural gas have been assumed to be equal to 252 gCO_2/kWh [16].

Table 9.1: CO_2 equivalent emissions savings

	Min. En. savings [kWh/year]	Min. CO2 eq. savings [kg CO2 eq./year]
Natural gas	16000	4032
Electricity	6154	2413

Chapter 10

Conclusions

Optimising HVAC working schedules daily using ML techniques is not only technically possible (as proven in the test installation), but also feasible and sustainable.

From an environmental point of view, feasibility is evident: costs are almost null, and emissions reduction is estimated in, at least, a 10%.

From an economic point of view, feasibility depends strongly on the already existing control system and the installation requirements. Analysing just the software (that is the scope of this project), economic feasibility (i.e. payback $<$ 6 months) starts for energy savings larger than 16000 kWh for natural gas and 6154 kWh for electricity (considering prices exposed in Section 8.2). This values, despite being big, represent (for example) just a 10% consumption reduction for a 100 kW natural gas boiler.

Academically, this project has allowed me to improve (a lot) my programming skills specially with Python (which I had not touched it since my 1st engineering year) and with HTML and PHP (what I had never seen before). Also, I have reviewed a considerable part about heat transfer. Lastly, I have been able to compute and understand some statistics concepts I had studied years ago.

Personally, developing the project while working full-time has forced me to organise a lot my time (both my working time and my free time) and has made me prove my willpower.

Chapter 11

Future work

Future work will basically consist in checking the performance of the developed tool in a real installation during a long period of time, assessing energy savings and thermal comfort gains (if any).

Furthermore, the user interface should be improved making it more attractive and safe.

Lastly, some other ML techniques could be assessed. For example, reinforced learning could be interesting to let the software freely decide the schedule.

Acknowledgements

I would like to thank, first of all, my bosses and colleagues in MOS, Enrico Braggion and Josep Fageda Bassagañas for proposing me the topic of the project and for their interest and support (both material and moral).

Dr. Cecilio Angulo, from UPC, has also been a key person to develop the project, as it has supervised its development since the very beginning and has provided me very useful ideas and references.

Lastly, I would be unfair if I did not thank my family, specially for letting me turn on and off the air conditioning in a completely arbitrary way (when gathering data to test the software).

References

[1] International Energy Agency. Energy efficiency 2017. Technical report, International Energy Agency, 2017.

[2] M. Castilla, J.D. Álvarez, M. Berenguel, M. Pérez, F. Rodríguez, and J.L. Guzmán. Técnicas de control del confort en edificios. *Revista Iberoamericana de Automática e Informática Industrial*, 7, July 2010.

[3] Montrol Open Solutions SL. Montrol, June 2018. URL `http://www.montrol-e.com/es/`.

[4] Expert System. What is machine learning? a definition, 2018. URL `https://www.expertsystem.com/machine-learning-definition/`.

[5] Jay H. Lee, Joohyun Shin, and Matthew J. Realff. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Elsevier, Computers and Chemical Engineering*, 114, 2017. URL `https://www.sciencedirect.com/science/article/pii/S0098135417303538`.

[6] Chirag Deb, Siew Eang Lee, and Mattheos Santamouris. Using artificial neural networks to assess hvac related energy saving in retrofitted office buildings. *Elsevier, Solar Energy*, 163, 2018. URL `https://www.sciencedirect.com/science/article/pii/S0038092X18300963`.

[7] Daniel J. Eck. Bootstrapping for multivariate linear regression models. *Elsevier, Statistics and Probability Letters*, 134, 2018. URL `https://www.sciencedirect.com/science/article/pii/S0167715217303450`.

[8] ANSI/ASHRAE, editor. *Standard 55, Thermal Environmental Conditions for Human Occupancy*. ANSI/ASHRAE, 2013.

[9] José Domingo Álvarez, Ramon Costa-Castelló, and María del Mar Castilla. Repetitive control to improve users' thermal comfort and energy efficiency in buildings. *Energies*, 11, 2018. URL `http://www.mdpi.com/1996-1073/11/4/976/html`.

[10] Engineering Toolbox. The engineering toolbox - predicted mean vote index (pmv), June 2018. URL https://www.engineeringtoolbox.com/predicted-mean-vote-index-PMV-d_1631.html.

[11] Siyu Wu and Jian-Qiao Sun. Multi-stage regression linear parametric models of room temperature in office buildings. *Elsevier, Building and Environment*, 57, 2012. URL https://doi.org/10.1016/j.buildenv.2012.04.014.

[12] Agencia Estatal de Meteorología. Aemet open data, 2018. URL https://opendata.aemet.es/centrodedescargas/inicio.

[13] PrecioGas. Precio kwh gas natural, 2018. URL https://preciogas.com/conceptos/precio-kwh.

[14] Comisión Nacional de los Mercados y la Competencia. Comparador de ofertas de energÃŋa, 2018. URL https://www.cnmc.es/ambitos-de-actuacion/energia/comparador-ofertas-energia.

[15] Generalitat de Catalunya Oficina Catalana del Canvi Climàtic. Factor d'emissió associat a l'energia elèctrica: el mix elèctric, June 2018. URL http://canviclimatic.gencat.cat/ca/redueix_emissions/com-calcular-emissions-de-geh/factors_demissio_associats_a_lenergia/index.html.

[16] Instituto para la Diversificación y Ahorro de la Energía. Factores de emisión de co2 y coeficientes de paso a energía primaria de diferentes fuentes de energía final consumidas en el sector edificios en españa. Technical report, Gobierno de España - Ministerio de Industria, Energía y Turismo, 2014.

Annexes

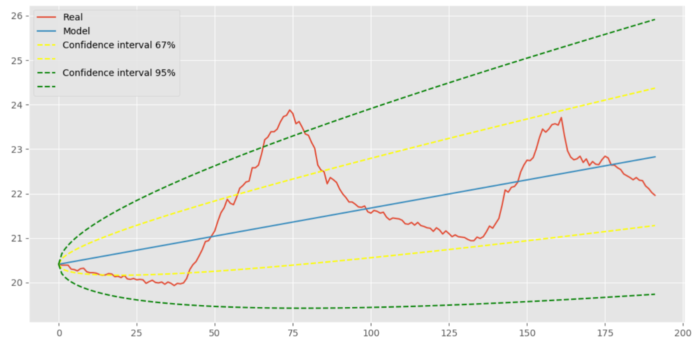
Annex I: Pilot installation prediction models' comparison

This Annex contains the main results of the various simulations performed to choose the best regression model and the best features' combinations.

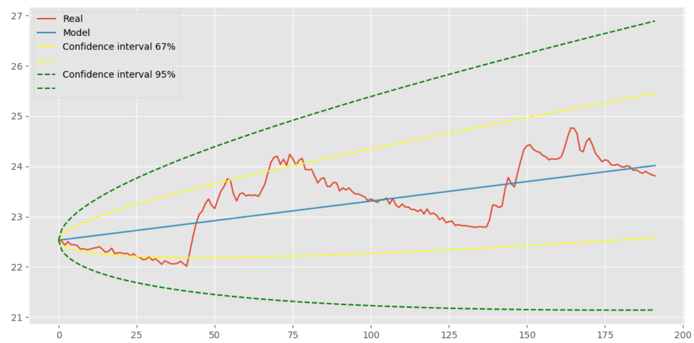
Linear Regression. Variables combination 1. Prediction time 48h

	North	South
MSE [°C]	0.82	0.27
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.55	1.44
95% conf. int. [°C]	3.09	2.88

Linear Regression. North Area. Variables combination 1. Prediction time 48h



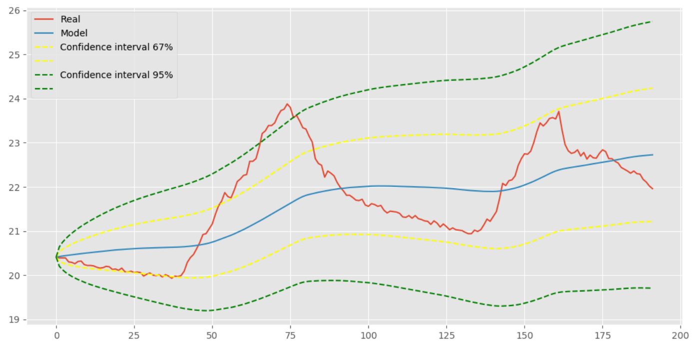
Linear Regression. South Area. Variables combination 1. Prediction time 48h



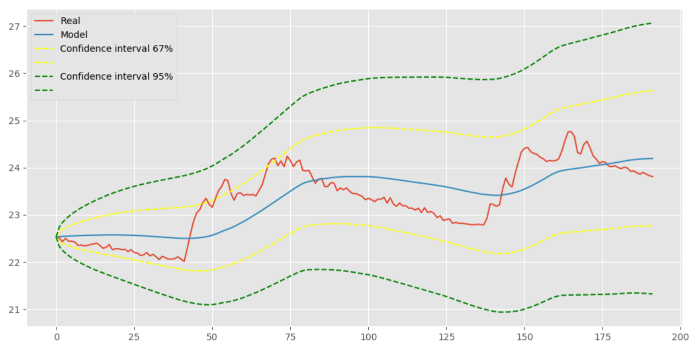
Linear Regression. Variables combination 2. Prediction time 48h

	North	South
MSE [°C]	0.72	0.24
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.51	1.44
95% conf. int. [°C]	3.02	2.87

Linear Regression. North Area. Variables combination 2. Prediction time 48h



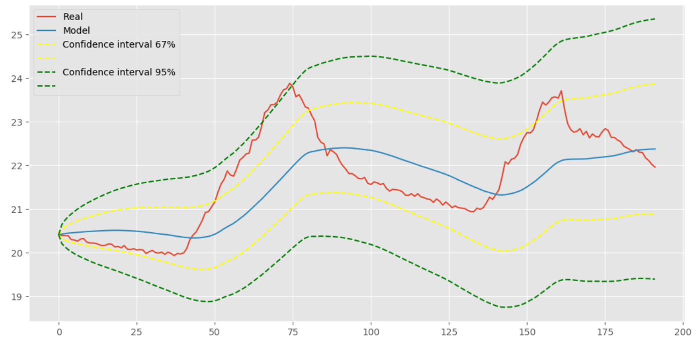
Linear Regression. South Area. Variables combination 2. Prediction time 48h



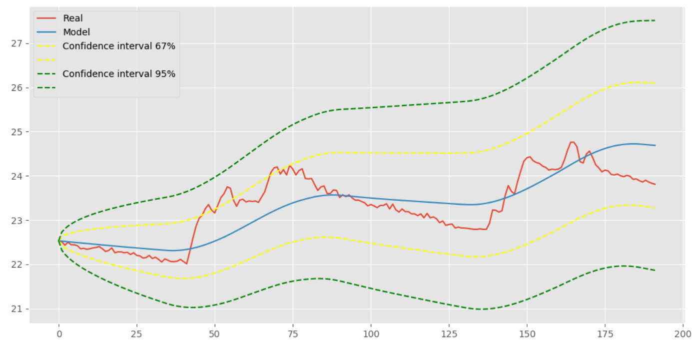
Linear Regression. Variables combination 3. Prediction time 48h

	North	South
MSE [°C]	0.72	0.21
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.49	1.41
95% conf. int. [°C]	2.98	2.83

Linear Regression. North Area. Variables combination 3. Prediction time 48h



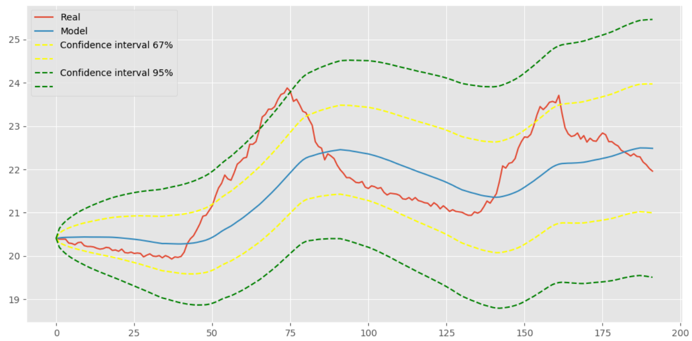
Linear Regression. South Area. Variables combination 3. Prediction time 48h



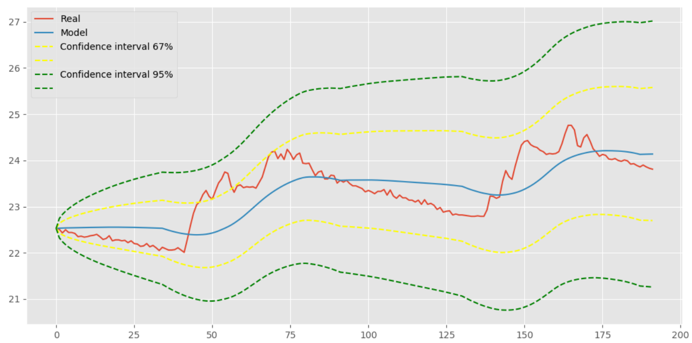
Linear Regression. Variables combination 4. Prediction time 48h

	North	South
MSE [°C]	0.71	0.22
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.49	1.44
95% conf. int. [°C]	2.98	2.88

Linear Regression. North Area. Variables combination 4. Prediction time 48h



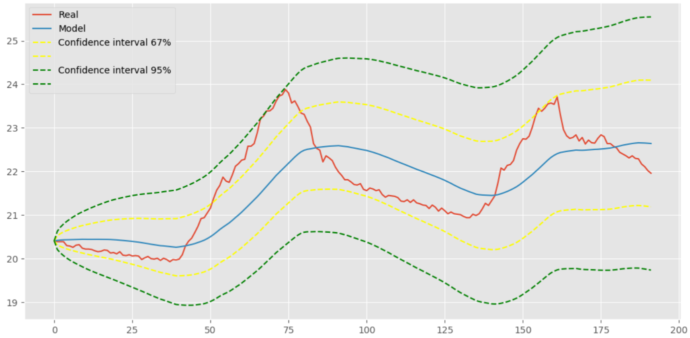
Linear Regression. South Area. Variables combination 4. Prediction time 48h



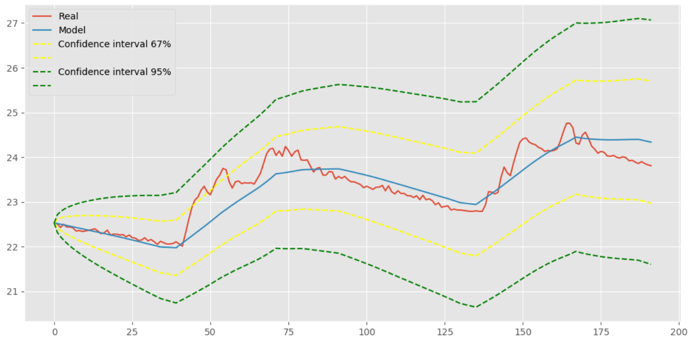
Linear Regression. Variables combination 5. Prediction time 48h

	North	South
MSE [°C]	0.58	0.11
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.45	1.37
95% conf. int. [°C]	2.90	2.73

Linear Regression. North Area. Variables combination 5. Prediction time 48h



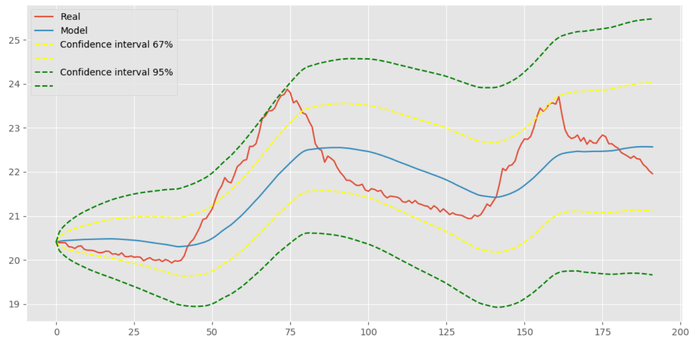
Linear Regression. South Area. Variables combination 5. Prediction time 48h



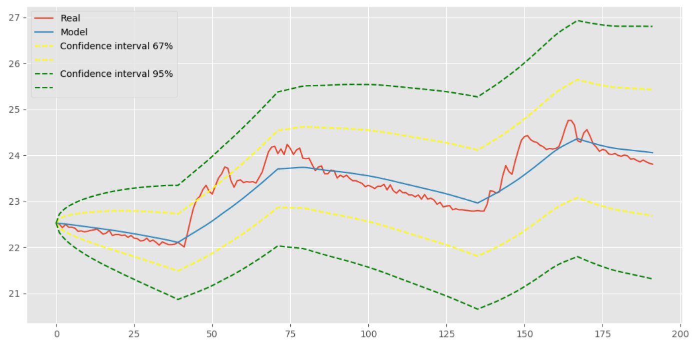
Linear Regression. Variables combination 6. Prediction time 48h

	North	South
MSE [°C]	0.59	0.10
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.45	1.37
95% conf. int. [°C]	2.91	2.74

Linear Regression. North Area. Variables combination 6. Prediction time 48h



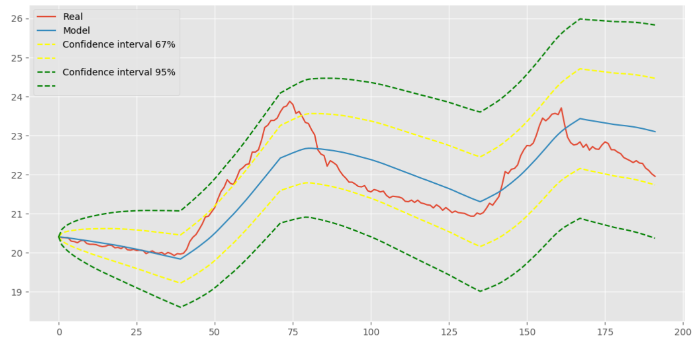
Linear Regression. South Area. Variables combination 6. Prediction time 48h



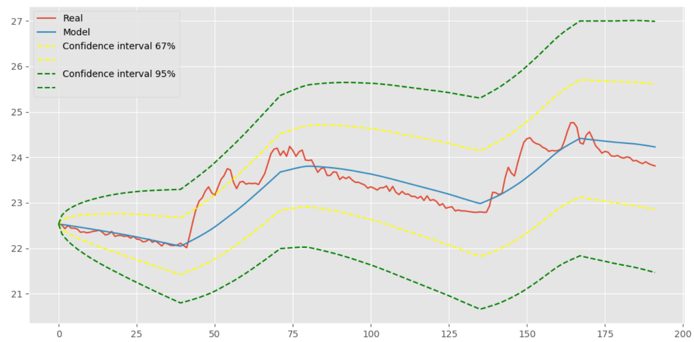
Linear Regression. Variables combination 7. Prediction time 48h

	North	South
MSE [°C]	0.40	0.13
Error st. dev. [°C]	0.10	0.10
67% conf. int. [°C]	1.36	1.38
95% conf. int. [°C]	2.73	2.76

Linear Regression. North Area. Variables combination 7. Prediction time 48h



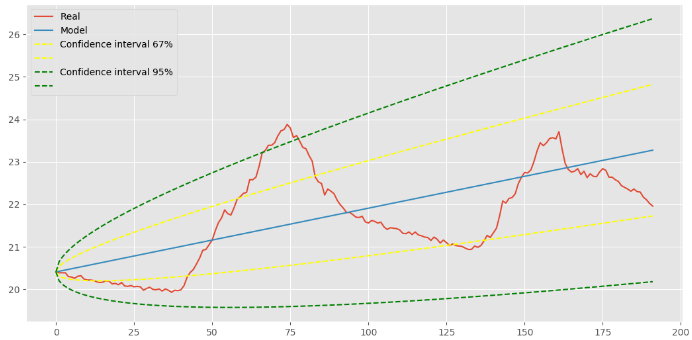
Linear Regression. South Area. Variables combination 7. Prediction time 48h



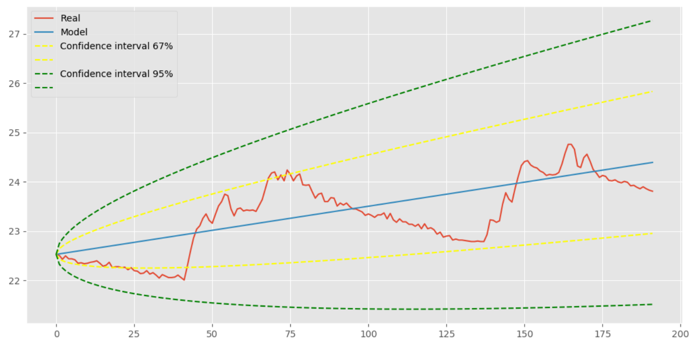
Support Vector Regression. Variables combination 1. Prediction time 48h

	North	South
MSE [°C]	0.86	0.28
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.55	1.44
95% conf. int. [°C]	3.09	2.88

Support Vector Regression. North Area. Variables combination 1. Prediction time 48h



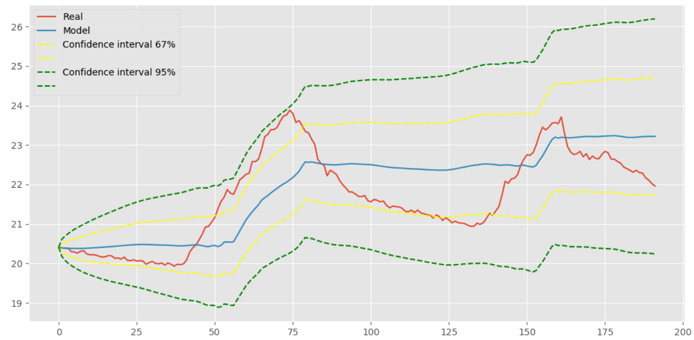
Support Vector Regression. South Area. Variables combination 1. Prediction time 48h



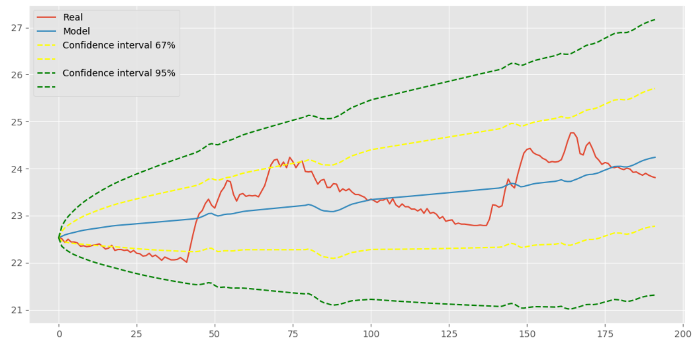
Support Vector Regression. Variables combination 2. Prediction time 48h

	North	South
MSE [°C]	0.75	0.28
Error st. dev. [°C]	0.11	0.11
67% conf. int. [°C]	1.49	1.46
95% conf. int. [°C]	2.98	2.93

Support Vector Regression. North Area. Variables combination 2. Prediction time 48h



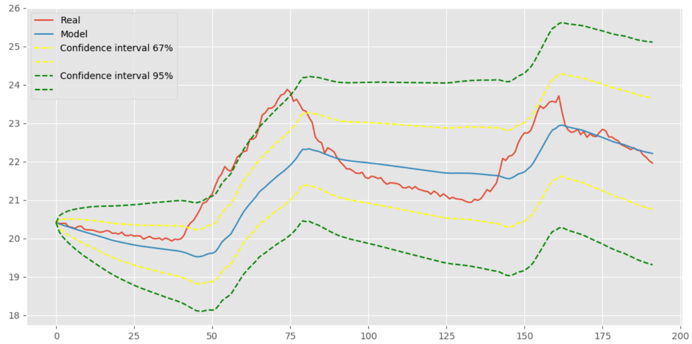
Support Vector Regression. South Area. Variables combination 2. Prediction time 48h



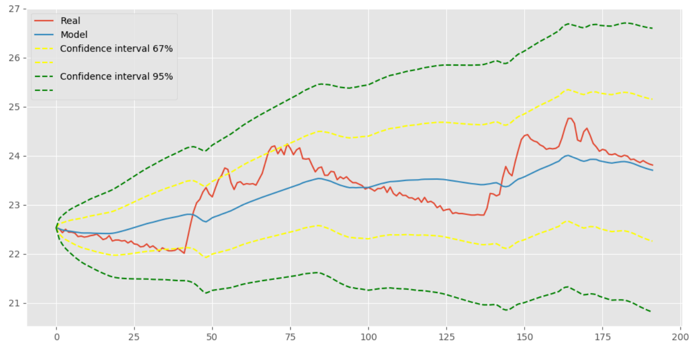
Support Vector Regression. Variables combination 3. Prediction time 48h

	North	South
MSE [°C]	0.69	0.21
Error st. dev. [°C]	0.10	0.10
67% conf. int. [°C]	1.45	1.45
95% conf. int. [°C]	2.90	2.89

Support Vector Regression. North Area. Variables combination 3. Prediction time 48h



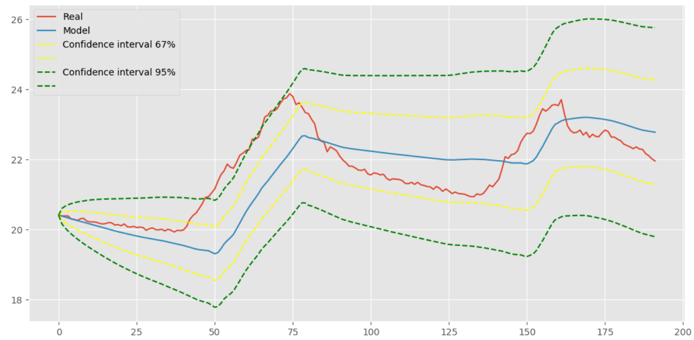
Support Vector Regression. South Area. Variables combination 3. Prediction time 48h



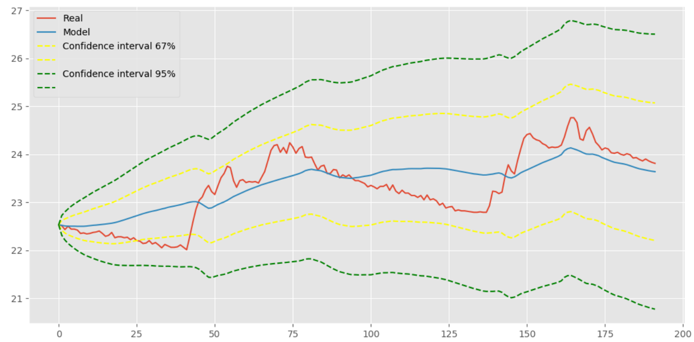
Support Vector Regression. Variables combination 4. Prediction time 48h

	North	South
MSE [°C]	0.81	0.22
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.49	1.43
95% conf. int. [°C]	2.98	2.86

Support Vector Regression. North Area. Variables combination 4. Prediction time 48h



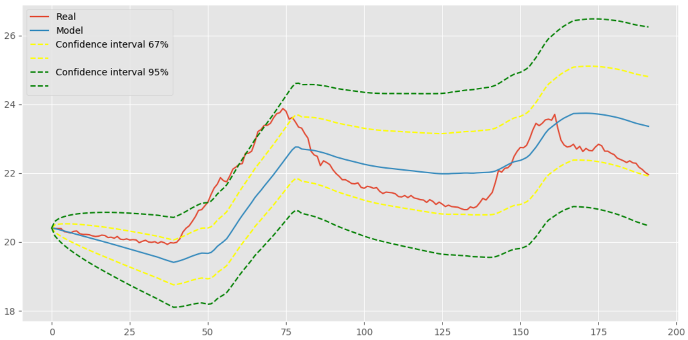
Support Vector Regression. South Area. Variables combination 4. Prediction time 48h



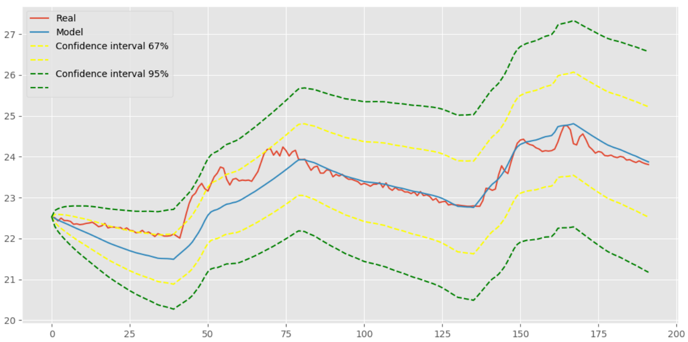
Support Vector Regression. Variables combination 5. Prediction time 48h

	North	South
MSE [°C]	0.82	0.15
Error st. dev. [°C]	0.10	0.10
67% conf. int. [°C]	1.44	1.35
95% conf. int. [°C]	2.89	2.70

Support Vector Regression. North Area. Variables combination 5. Prediction time 48h



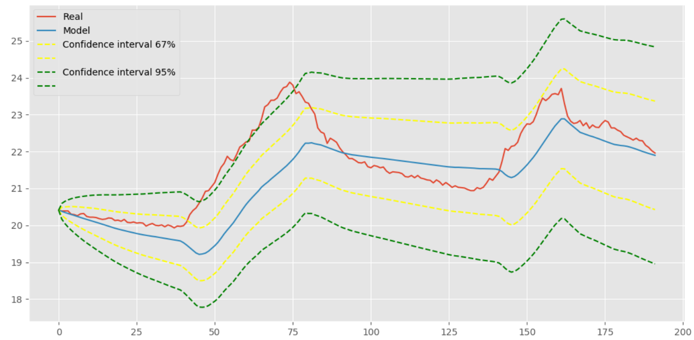
Support Vector Regression. South Area. Variables combination 5. Prediction time 48h



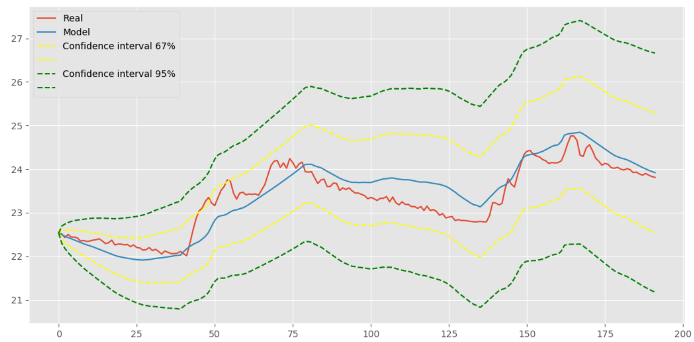
Support Vector Regression. Variables combination 6. Prediction time 48h

	North	South
MSE [°C]	0.81	0.13
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.47	1.37
95% conf. int. [°C]	2.94	2.74

Support Vector Regression. North Area. Variables combination 6. Prediction time 48h



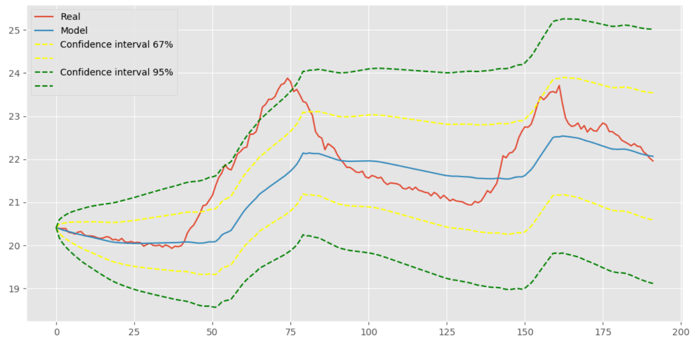
Support Vector Regression. South Area. Variables combination 6. Prediction time 48h



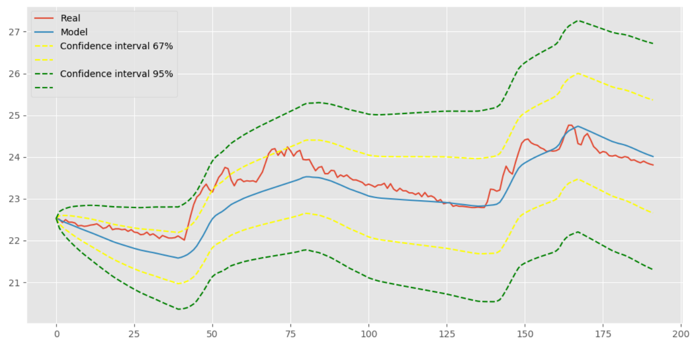
Support Vector Regression. Variables combination 7. Prediction time 48h

	North	South
MSE [°C]	0.67	0.18
Error st. dev. [°C]	0.11	0.10
67% conf. int. [°C]	1.47	1.35
95% conf. int. [°C]	2.95	2.71

Support Vector Regression. North Area. Variables combination 7. Prediction time 48h



Support Vector Regression. South Area. Variables combination 7. Prediction time 48h

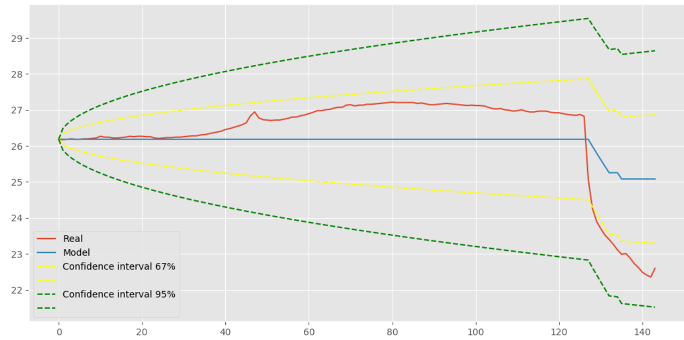


Annex II: Test installation prediction models' comparison

Linear Regression. Variables combination 1. Prediction time 12h

MSE	0.96
Error st. dev.	0.15
67% conf. int.	1.78
95% conf. int	3.56

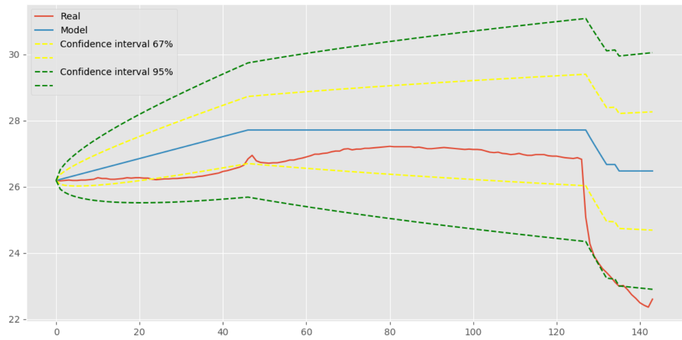
Linear Regression. Variables combination 1. Prediction time 12h



Linear Regression. Variables combination 2. Prediction time 12h

MSE	1.96
Error st. dev.	0.15
67% conf. int.	1.78
95% conf. int	3.56

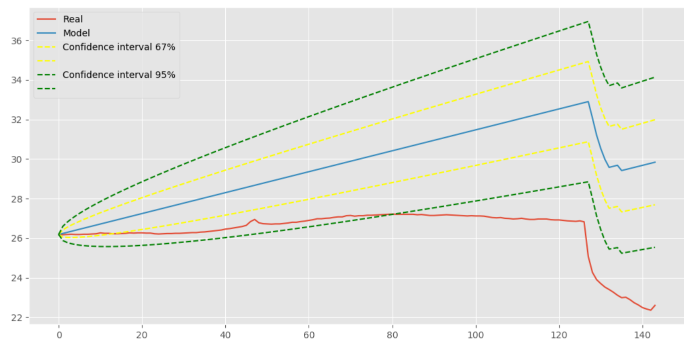
Linear Regression. Variables combination 2. Prediction time 12h



Support Vector Regression. Variables combination 1. Prediction time 12h

MSE	14.81
Error st. dev.	0.18
67% conf. int.	2.15
95% conf. int	4.30

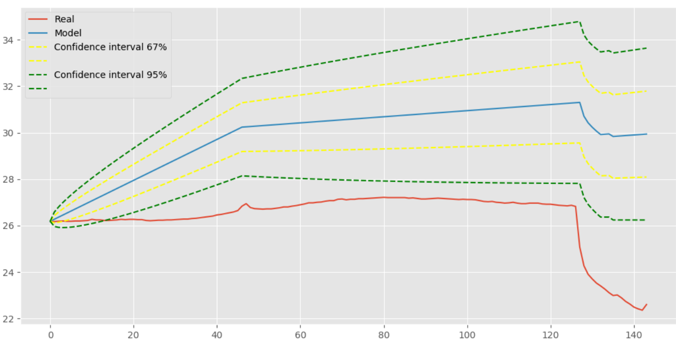
Support Vector Regression. Variables combination 1. Prediction time 12h



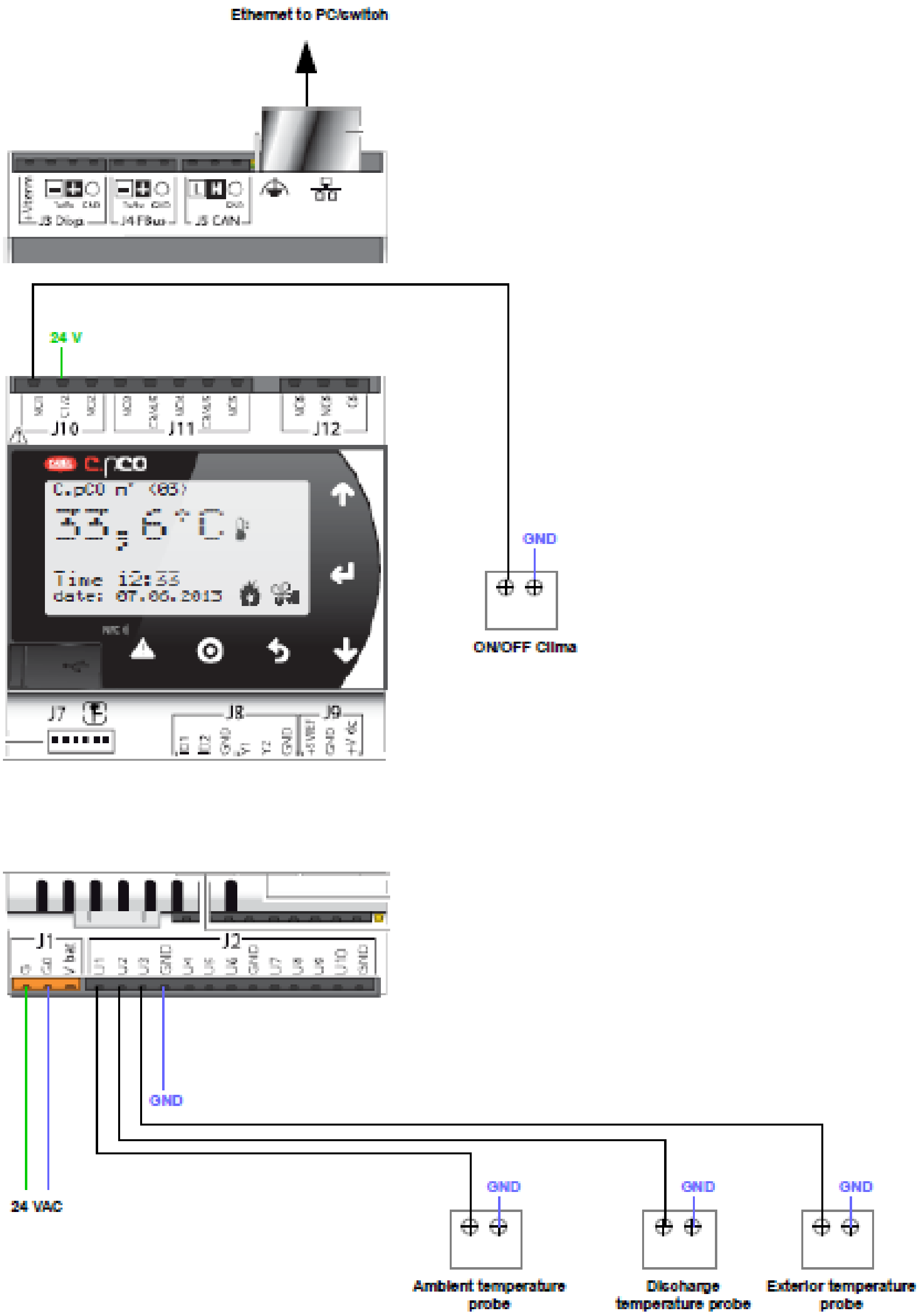
Support Vector Regression. Variables combination 2. Prediction time 12h

MSE	15.10
Error st. dev.	0.16
67% conf. int.	1.85
95% conf. int	3.70

Support Vector Regression. Variables combination 2. Prediction time 12h



Annex III: Test installation connections scheme



Annex IV: Test installation field elements user manuals

Due to the size of the user manuals, they have been located into an attached folder.