Università
degli Studi
di Padova

Dipartimento
di Ingegneria
dell'Informazione

Modeling and Identification for HVAC Systems

Laureando
Francesco Scotton

Relatore
Giorgio Picci

Dipartimento di
Ingegneria
dell'Informazione

Tesi svolta presso il
Royal Institute
of Technology
KTH

Anno 2012

Abstract

Heating, Ventilation and Air Conditioning (HVAC) systems consist of all the equipment that control the conditions and distribution of indoor air. Indoor air must be comfortable and healthy for the occupants to maximize their productivity. Moreover, HVAC energy consumption is between 20% and 40% of the total energy consumption in developed countries and accounts around 33% of the global CO_2 emissions. So the study of HVAC systems plays an important role in building science.

The aim of this project is to identify mathematical models that will be employed by intelligent control algorithms which guarantee human comfort indoors, energy saving and less CO_2 emissions at the same time. Three models, based on first-principle physical knowledge, are proposed for CO_2 concentration, temperature and humidity, respectively, for a room in the Q-building at KTH. Thermodynamic equations and an original estimation of the number of the occupiers of the room are employed.

Validation shows that models have really good performances, even with a short training dataset. Discussions on the obtained results are given and some ideas for future work are proposed.

Acknowledgments

Travel is the only thing you buy that makes you richer. This period in Stockholm has been the best experience in my life. Working here on my thesis would not have been possible without the help, guidance and support of some people who contributed to realize it.
First of all I would like to thank Prof. Bo Wahlberg for giving me the opportunity to do my thesis with the Automatic Control group.
Thanks to my supervisor Lirong Huang, my guidance for all the project period, for your patience, kindness and availability.
Thanks to my italian supervisor Prof. Giorgio Picci, who made this experience possible.
Thanks to Seyed Alireza Ahmadi, for your help for the hardware part and for involving me on the writing of your paper.
Thanks to all the people of the HVAC group: I have learnt a lot in your meetings.
Thanks to Damiano Varagnolo, for the LaTeX tips and for your help with the nonparametric part.
Thanks to Prof. Gianluigi Pillonetto, who kindly provides me the MATLAB® code for the nonparametric part.
Thanks to Antonio and Sky, who had the role of family for me during my thesis period and to all the people I had the pleasure to know in Sweden.
Last but not least, thanks to my family, for your support in my whole life, for believing in me and for giving me the possibility to do my thesis at KTH.

Contents

List of acronyms

HVAC	Heating Ventilation and Air Conditioning
IAQ	Indoor Air Quality
WSN	Wireless Network Sensor
T/H	Temperature/Humidity
SCADA	Supervisory Control and Data Acquisition
DCV	Demand Controlled Ventilation
LTI	Linear Time Invariant
MISO	Multiple Input - Single Output
ARX	Auto-Regressive with eXogenous inputs
ARMAX	Auto-Regressive Moving Average with eXogenous inputs
BJ	Box-Jenkins
OE	Output Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
LS	Least Squares
PEM	Prediction Error Methods
PLC	Programmable Logic Controller

1 Introduction

Heating, Ventilation and Air conditioning (HVAC) systems consist of components working together to introduce, distribute and condition air in a building for human comfort.

HVAC systems play a major role in the control of Indoor Air Quality (IAQ) and thermal comfort. Indeed, poor ventilation, improper temperature and humidity cause a bad "indoor environmental quality". People spend 80-90% of their time indoors: they are less productive and more often get sick with a bad IAQ, since it can cause irritation of the eyes and nose, fatigue, headache and shortness of breath. To guarantee a good IAQ, HVAC systems typically employ a control that mantain a fixed setpoint of fresh air ventilation based on the designed occupancy of the space. This is an inefficient method, since it often provides much more fresh air than necessary, especially in the spaces with frequently varying occupancy, such as laboratories and conference rooms. To cope with this problem, Demand Controlled Ventilation (DCV) is employed: DCV uses CO_2 sensors to control the supply of fresh air. In fact, people are the main source of indoor air contaminants and the variation of CO_2 concentration is a quite accurate way to monitor the occupancy, which makes it possible to set the fresh air amount according to people demand. DCV is definitely more efficient than the fixed ventilation method, since it avoids excessive ventilation and saves energy.

Over the past few years energy saving has become an important topic. The percentage of buildings contribution to the total energy consumption is between 20% and 40% in developed contries and is rapidily increasing, as the population grows and the demand for building services and comfort levels is rising [15]. HVAC systems account for the greatest amount of the energy consumption in a building. Indeed, this has markedly grown over recent years,

since comfortability is not considered luxury anymore, but it is required. For instance we can consider the U.S. situation in the year 2010 according to the Buildings Energy Data Book 2010 (U.S. Department of Energy): Figure 1.1 shows that buildings energy consumption takes 41% of the total energy consumption, while the heating, cooling and ventilation consumptions are about half of the total buildings consumption. Hence, improving the efficiency of

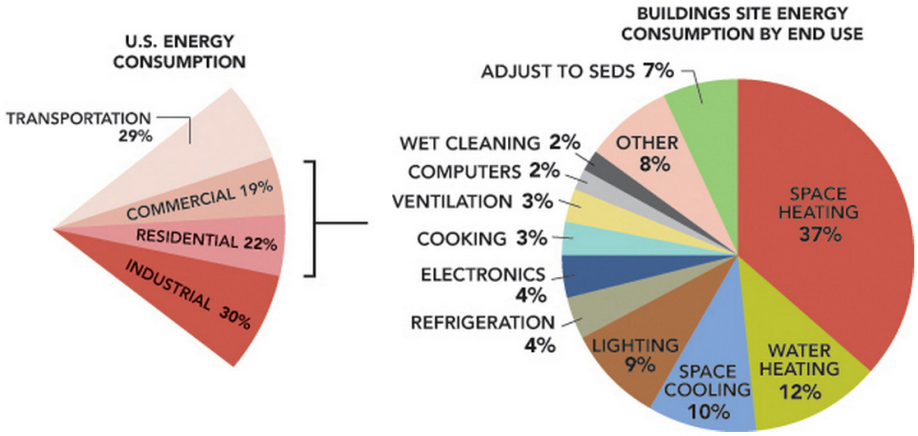


Figure 1.1: U.S. energy consumption year 2010.

HVAC systems can be very helpful for the energy saving. It has been estimated that an intelligent control could reduce HVAC systems energy consumption by 20-30% [7]. The energy saving would also have a positive effect on the CO_2 emissions as buildings account for about 33% of global CO_2 emissions. A significant part of CO_2 emissions are in fact caused by the combustion of fossil fuels to provide heating, cooling, lighting and the power for home appliances and electronic devices.

A smart way to implement an intelligent control algorithm is the employment of a Wireless Sensor Network (WSN). A WSN consists of a group of sensor nodes which are spatially distributed in a measurement area. Each sensor node is equipped with a transducer which can provide an electric signal, changes of which depend on a physical variable like temperature, pressure and illumination intensity. Sensors in the WSN can communicate among themselves and the "main" node, which receives all the data and forwards them to the processing unit.

Wired sensors are usually employed in typical HVAC systems: however they are pretty expensive and they cannot be arbitrarily located because of the cables. On the contrary, WSNs are cheaper and more flexible: a greater number of sensors can be used, offering more accurate measurements, the network can be built and turned down quite easily without changes in the rest of the environment, sensors can be reprogrammed in every moment, more complex algorithm can be implemented and sensors can be placed everywhere, as they work on stand-alone energy supply.

This project is implemented at the Royal Institute of Technology (KTH), Stockholm. Here, improving the existing HVAC system is in need. In fact, according to the analysis in [8], there are some buildings which do not use energy in a proper way. For example, half of the Q-building energy for cooling or heating is consumed during not working hours. Moreover, it is pointed out in [8] that existing temperature and CO_2 sensors at KTH are not very accurate: this is another reason why the present control is inefficient.

1.1 Previous Work

In the literature there are many work studying identifications for HVAC systems. They can be classified into two groups: *Black Box* (no a priori information) and *Grey Box* (based on physical knowledge) model approaches. Due to the difficulties in thermodynamics models, black box is the most common choice: linear parametric models, as ARX, ARMAX, BJ and OE have been successfully implemented to model HVAC systems.

Chi-Man Yiu *et al.* [4] dealt with a black box identification for an air conditioning system: a MIMO ARMAX model was estimated, parameters of which were evaluated using the Ricorsive Extended Least Squares Method (RELS), and compared it with a SISO ARMAX model.

Mustafaraj *et al.* [5] tested different temperature and humidity models for an office: BJ, ARX, ARMAX and OE were identified with the black box technique. In that environment, BJ outperformed ARX and ARMAX, but it is noticed that the difference is only in the assumption of the noise model, which could be suitable only for that system: it could not be concluded that the BJ is better than ARX and ARMAX.

Mustafaraj *et al.* developed their previous work in [6]: NoNlinear AutoRegressive models with eXternal inputs (NNARX) for temperature and humidity were estimated, and their performances were compared with linear ARX models. They also considered the effect of the carbon dioxide concentration in the models, as there was correlation between occupancy and CO_2 level, demonstrating that the accuracy of the model was improved.

Haizmann [1] used the black box method and identified an OE model for the temperature of a small conference room on the 6^{th} floor of the Q-building at the Royal Institute of Technology (KTH). This work did not obtain a robust model: it suffered of a lot of hardware problems and the author was not able to get reliable data.

Recently, Wu and Sun [7] proposed a physics-based temperature model of a room in an office building and compared it with a black box identified ARMAX model. They employed thermodynamics equations to determine the structure and the order of a linear regression model (grey box approach).

Qi and Deng [10] studied a MIMO control strategy for the Air Conditioning system (A/C) to control both indoor air temperature and humidity. The model of the A/C system was derived from energy and mass conservation principles.

Maasoumy [3] estimated a temperature model for three rooms of a building and designed an optimal control algorithm for HVAC systems. The thermal circuit approach, which uses analogies with electric circuits, was employed.

1.2 Contribution of the Thesis

This project proposes, using observed data, mathematical models that describe the dynamics of temperature, humidity and CO_2 level of a room, which can be used to implement an intelligent control for the HVAC system.

Three MISO ARMAX models are proposed for CO_2 level, temperature and humidity, respectively, following the Grey Box approach. Physical knowledge and an estimation of the number of occupiers of the room based on CO_2 level are employed. Models are identified with the measurements taken in May 2012.

Validation shows that these are really good models according to the defined metrics, even with a short dataset, and they outperform the temperature models in the previous results [1] and [7].

1.3 Overview of the Thesis

The thesis is organised as follows:

- **Chapter 2:** we introduce the background. HVAC components, WSN sensors type and topology and the plant of our test-bed are described;
- **Chapter 3:** we identify the models: structure of the models, methods of identification, data pre-processing and the identified models are described;
- **Chapter 4:** we validate the models: employed validation metrics, comparison between different model types and with past work are presented;
- **Chapter 5:** we discuss the obtained results and suggest some ideas about how this project can be continued and improved.

2
Test-bed Setting

Experiments are taken in the test-bed placed in the room A 225 (LAB3) on the 2nd floor of the Q-building at KTH. This room has an area of about 80 m^2, a volume of about 270 m^3, four windows of 0.64×4 m^2 on the external wall and one window on one internal wall of 2.5 m^2 (see Figure 2.6 for the map of the test-bed). It is equipped with an HVAC system and a Wireless Sensor Network (WSN).

2.1 HVAC System in the Test-bed

KTH campus is equipped with a SCADA system (centralized system which monitors and controls the whole plant) connected to all ventilation, heating and cooling units. In the Q-building there are three separate ventilation units for fresh air supply and several cooling and heating units to supply cooling coils and radiators. Some temperature and CO_2 sensors are located in the building and linked to the "central building system" which consists of three single board computers (complete computer built on a single circuit board), called Soft PLC. Indeed, each ventilation unit is controlled by a Soft PLC, which is connected to a respective group of sensors, cooling and heating units. Soft PLCs have the internet access and can be manually controlled, using the web interface of the SCADA system, from a remote location: it is possible to change the state of the actuators and monitor the measurements of the sensors.

Rooms with special functions, like laboratories and conference rooms, are equipped with sensors and actuators: they can be controlled by the related Soft PLC, which is programmed to keep, using respectively heating and cooling actuators and Demand Controlled Ventilation (DCV) method (see Chapter 1 for definition), a temperature of 22°C with a ± 1°C dead band and a CO_2 level

(a) Fresh air inlet **(b)** AC

(c) Radiator **(d)** Air outlet

Figure 2.1: HVAC components in the testbed.

below 850ppm.

The actuators in the test-bed are: four fresh air inlets, four AC units, two outlets and four radiators, whose pictures are depicted in Figure 2.1. We will explain more in details how these components work in the following sections.

2.1.1 Fresh Air Inlet

In Figure 2.2 it is depicted a schematic drawing of the fresh air inlet.

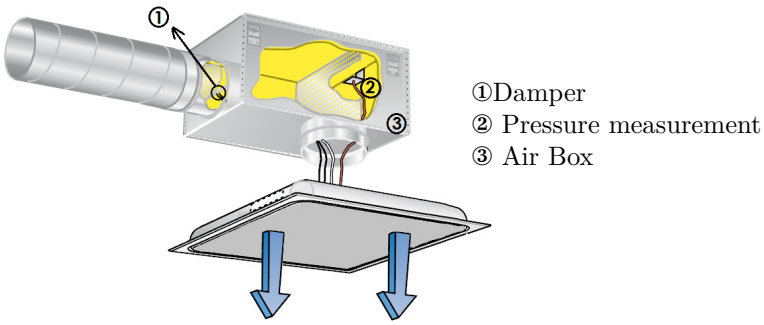


Figure 2.2: Inlet working.

Fresh air is supplied in the duct by one of the ventilation units, while the damper regulates the airflow which comes into the room. The opening percentage of the damper can be set in the SCADA web interface. The pressure measurement in the air box is used during the installation process: the position of the damper

can be fixed depending on the desired airflow. The airflow q ($\frac{l}{s}$) is calculated as

$$q = k\sqrt{p}, \tag{2.1}$$

where p is the pressure (Pa) and k is a factor given by the company which produces the inlet. The position of the damper can then be adjusted to obtain the value of measured pressure that gives, using the equation (2.1), the desired airflow . If the damper position is regulated with a motor (not fixed), it is still possible, using the pressure measurement, to check the value of the airflow at given damper positions.

2.1.2 Air Conditioning

In Figure 2.3 a schematic drawing of the air conditioning unit in the test-bed is illustrated. Its working is based on the *induction principle*.

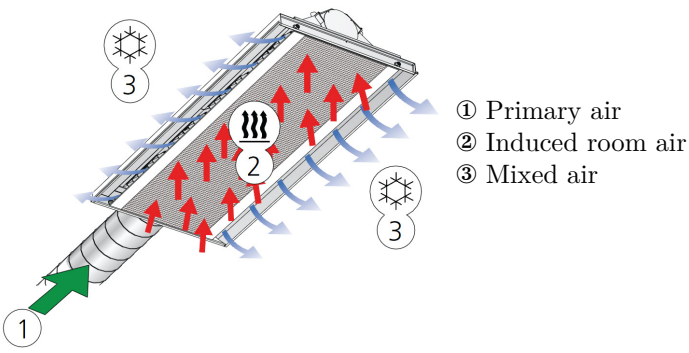


Figure 2.3: AC working.

The primary air, supplied by one of the ventilation units, is injected in a plenum, a housing where it is created and stored air with a greater pressure than the atmospheric one. The plenum is equipped with nozzles of various sizes, which are little pipes from which the air can be discharged. Due to the high pressure in the plenum, the air comes out through the nozzles at high velocity and creates a zone with lower pressure, as an increase in the velocity produces a decrease on the pressure. This depression induces the room air to be sucked up through the heat exchanger, which consists of a coil where chilled water flows. The sucked room air is then cooled with the heat exchanger, mixed with the primary air and discharged into the room from the sides of the device, as we can see in Figure 2.3. The AC unit can be used also to heat the room: the only difference with the cooling process is that the water that is flowing in the coil is hot. The maximum airflow coming from the AC in the testbed is about $20\frac{l}{s}$.

It should be noticed that the effect of the air coming out from the AC on the CO_2 level is negligible: the mixed air has a CO_2 level close to the room one and the airflow is relatively low if it is compared with the inlet. Because of

this fact, which has also been verified in practice, the effect of the AC is not considered in the CO_2 model.

As for the inlet, there is a damper in the primary air duct whose opening percentage can be set from the SCADA web interface. The heat exchanger is activated according to the state of the damper (open or closed).

2.1.3 Radiator

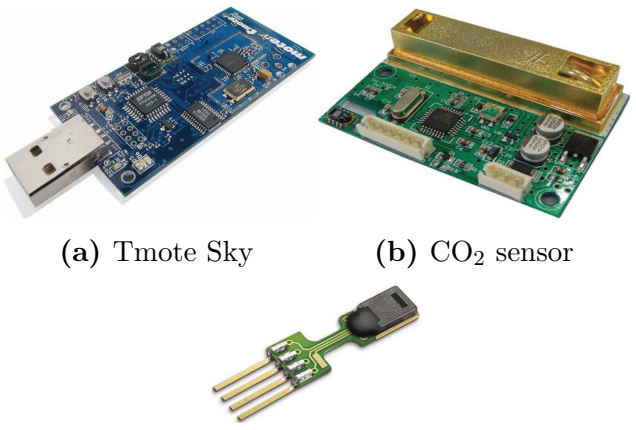
Similar to the other components, the radiator is equipped with a valve, whose opening percentage can be set from the SCADA interface. In this project the radiator was always turned off, as in the period when measurements were taken the weather was warm enough. Therefore the functioning of the radiator will not be treated in detail.

2.1.4 Air Outlet

As it can be seen in Figure 2.1d, the air outlet is just an hole in the wall, where a tube which allows the air to flow through is located. In the tube there is a damper whose opening percentage can be regulated from the SCADA web interface. As the tube communicates with the room outside, when the damper is completely open the room air is then discharged.

2.2 WSN in the Test-bed

Our WSN uses wireless sensors TMote Sky whose example is depicted in Figure 2.4a. TMote Sky follows industry standards, like USB and IEEE 802.15.4: it



(a) Tmote Sky **(b)** CO_2 sensor

(c) T/H sensor

Figure 2.4: Sensors in the testbed.

is then easy to interface it with other devices. TMote Sky is equipped with onboard temperature, humidity and light sensors. However, it was noticed in

[1] that the integrated temperature and humidity sensor measurements were influenced by the heat of the TMote Sky surrounding components. To overcome this problem, each TMote Sky was located in a box and another temperature and humidity sensor (Sensirion® SHT71, see Figure 2.4c) was used, which was fixed outside the box (isolated from the TMote board). TMote Sky is able to be attached to other external sensors: each CO_2 sensor (Soha Tech® SH-300-DTH, see Figure 2.4b) was then connected to a TMote Sky.

Motes (TMote Sky equipped with either temperature and humidity or CO_2 sensors) are organised in a *star* topology: as it can be noticed in Figure 2.5, the central node, called *coordinator*, receives the data from the other sensors and forwards them to the main mote called *base*, which is connected to a laptop. The data processing is accomplished using LabVIEW, a National Instruments® software. LabVIEW was chosen because it allows to implement real-time data acquisition and complex control algorithms: it is then a suitable environment for the future works. Furthermore, it offers the possibility to create a graphical interface and monitor the data capture, giving the user an easy way to notice errors in the system.

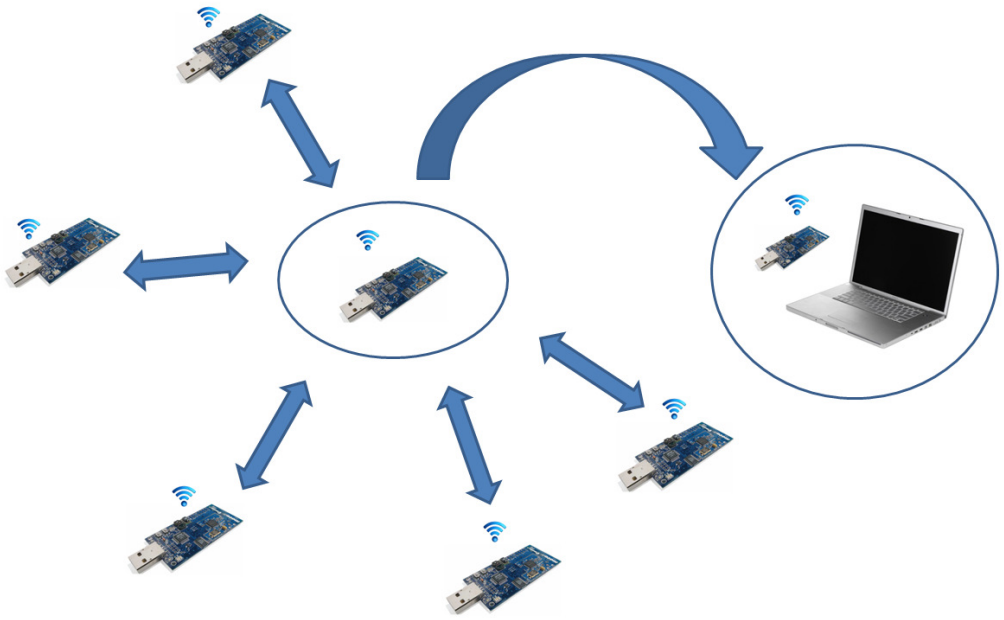


Figure 2.5: Star topology.

2.3 Test-bed Map

Figure 2.6 illustrates how sensors are located inside and outside the test-bed room: orange circles stand for temperature and humidity sensors and green one for CO_2 level. Numbers in the circles are the identifiers of the sensors. Locations of the ducts of air inlet, air outlet and AC, the radiator and the

windows can also be seen. Moreover, we use different colours to distinguish the "outside" walls (brown) from the "inside" one (black).

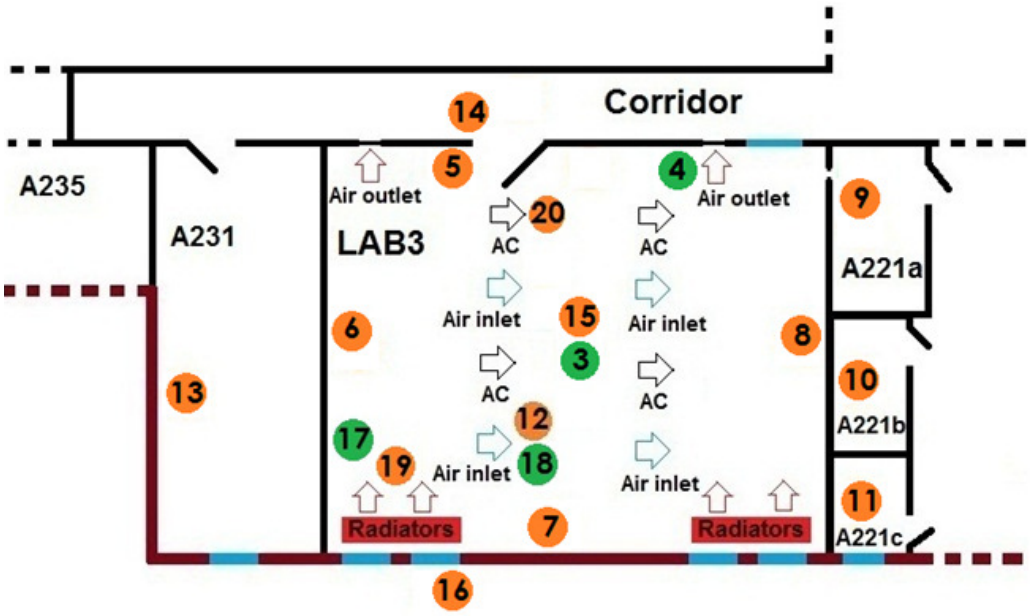


Figure 2.6: Test-bed LAB3: temperature and humidity sensors (orange circle), CO_2 level sensors (green circle), air outlet ducts (red arrow), fresh air inlet ducts (blue arrow), Air Conditioning ducts (black arrow), Radiators (red arrow), Outside wall (Brown line), windows (blue line).

3
Modeling and Identification

In this chapter we are going to describe the identification procedure applied to the HVAC system in the test-bed. The main problem of the identification is the limited set of data avalaible. Indeed, we have just collected measurements less than one week. The reason is that, to take informative measurements, a sufficient number of occupiers in the test-bed is needed. Unfortunately, this happens only when students have been in the test-bed to do some experiments for one of their courses. All models are based on the presence of people in the test-bed and we can obtain informative data only when the number of occupiers varies.

We have divided the "good" measurements into two groups:

- identification data: measurements from 16/05/2012 10:56:00 to 18/05/2012 08:00:00 are used to identify the model;
- validation data: measurements from 18/05/2012 09:05:00 to 19/05/2012 17:30:00 are used to validate the models.

3.1 Physics-based Models

Three MISO models, for CO_2 level, temperature and humidity, are proposed to model the HVAC system in the test-bed.

3.1.1 CO_2 Concentration

The CO_2 concentration relies on the number of the occupiers of the room and on the effect of the air inlet and air outlet ducts. When inlets and outlets are

turned on, room air is discharged through the outlets, while air with a lower CO_2 concentration is flowing in the room at the same time.

As we see in Figure 2.6, in our test-bed we have four CO_2 sensors: 3,4 and 17 which measure the room concentration and 18 which gives the concentration of the fresh air of the inlet. The mean of measurements of sensors 3 and 4 is used as room CO_2 concentration.

So the equation that represents the dynamics of CO_2 level in the room is

$$\frac{d}{dt}C(t) = kN(t) + \beta_{IN}\Big[I(t) - C(t)\Big] + d(t), \tag{3.1}$$

where $C(t)$ is the CO_2 concentration in the room, $N(t)$ is the number of occupiers of the room, $I(t)$ is the CO_2 concentration of the air inlet duct, $k > 0$ and $\beta_{IN} > 0$ are the unknown parameters and $d(t)$ is the disturbance.

Applying the *forward difference approximation* we obtain

$$\frac{d}{dt}C(t-1) \approx \frac{C_n - C_{n-1}}{\Delta t}, \tag{3.2}$$

where Δt is the sampling interval, C_n and C_{n-1} are the sampled value $C(n\Delta t)$ and $C((n-1)\Delta t)$. Substituting (3.2) in (3.1) we obtain

$$C_n - (1 - \beta_{IN}\Delta t)C_{n-1} = k\Delta t N_{n-1} + \beta_{IN}\Delta t I_{n-1} + \Delta t d_{n-1},$$

which can be rewritten

$$(1 + a_1 q^{-1})C_n = K_N N_{n-1} + K_I I_{n-1} + K_d d_{n-1}, \tag{3.3}$$

where $K_N = kK_d$, $K_I = \beta_{IN}K_d$, $K_d = \Delta t$ and $a_1 = K_I - 1$.

The relation above can be seen as a difference equation of a MISO **ARMAX** model (where θ is the vector of the parameters to be estimated and n_i the number of inputs)

$$A(q,\theta)y(t) = \sum_{i=1}^{n_i} B_i(q,\theta)u_i(t - n_{ki}) + C(q,\theta)e(t). \tag{3.4}$$

Estimation of the Number of Occupiers

One of the original ideas of this project is to use the room CO_2 concentration to estimate the number of the occupiers of the room. Our idea is to use the model found for the CO_2 level in the previous section to get an estimation of the number of the occupiers. Neglecting the noise, from (3.3) it follows that

$$N_{n-1} \approx \frac{C_n + a_1 C_{n-1} - K_I I_{n-1}}{K_N}. \tag{3.5}$$

With this approximation it is possible to use measurements of CO_2 concentration instead of those of the number of people in the temperature and humidity models. For our application it is not required to know the exact number of people but how much they affect the level of CO_2, temperature and humidity.

3.1.2 Temperature

We propose two different models for the temperature, depending on how the measurements of sensors 5,6,7,8,15 of the test-bed (see Figure 2.6) will be treated.

In both models we apply the concepts of *thermal resistance* of [3], which associate the conduction of heat with the electrical resistance. In particular, under the *steady-state* condition, the heat flow for *conduction* is given by

$$q_{cond} = \frac{kA}{L}(T_1 - T_2), \tag{3.6}$$

where k ($\frac{\mathrm{W}}{\mathrm{mK}}$) is the thermal conductivity (which depends on the wall material), A is the surface of the wall, L is the thickness of the wall, T_1 and T_2 are the temperatures of two sides of the same wall. If we define the *thermal resistance for conduction* R_{cond} ($\frac{\mathrm{K}}{\mathrm{W}}$) as

$$R_{cond} = \frac{L}{kA},$$

Equation (3.6) becomes

$$q_{cond} = \frac{T_1 - T_2}{R_{cond}}. \tag{3.7}$$

In the same way we can define the *thermal resistance for convection* starting from the formula of heat for *convection* under steady state

$$q_{conv} = hA(T_s - T_{\mathrm{inf}}),$$

where h ($\frac{\mathrm{W}}{\mathrm{m^2K}}$) is the convection heat transfer coefficient, A the surface of the wall, T_s is the temperature of the wall surface, T_{inf} is the temperature of the air inside (INT) or outside (EXT) the room (see Figure 3.1). Defining $R_{conv} = \frac{1}{hA}$ ($\frac{\mathrm{K}}{\mathrm{W}}$) we obtain a relation similar to (3.7).

The equivalent resistance R_{eq} ($\frac{\mathrm{K}}{\mathrm{W}}$) that describes the heat transfer for a wall is given by

$$R_{eq} = \frac{1}{h_1 A} + \frac{L}{kA} + \frac{1}{h_1 A},$$

as it is depicted in Figure 3.1 (Figure 2 in [3]), where T_{INT}, T_{EXT}, $T_{S,1}$, $T_{S,2}$ are respectively the inside, outside, internal surface and external surface temperatures and q_x is the overall heat given by $q_x = \frac{T_{INT} - T_{EXT}}{R_{eq}}$.

Mean Temperature Model

In this model we consider the temperature of the room T_i as the mean of the measurements of sensors 5,6,7,8,15,

$$T_i = \frac{T_5 + T_6 + T_7 + T_8 + T_{15}}{5}.$$

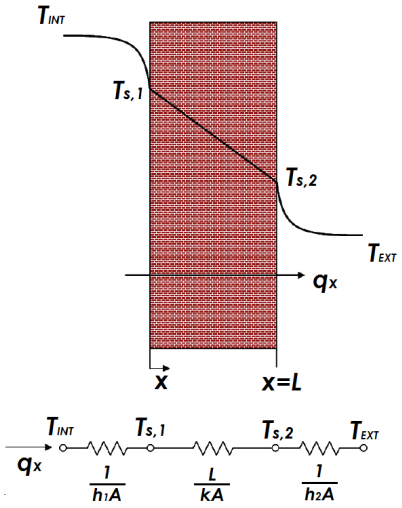


Figure 3.1: Heat transfer through a plane wall: T_{INT}, T_{EXT}, $T_{S,1}$, $T_{S,2}$ are respectively the inside, outside, internal surface and external surface temperatures and q_x is the overall heat (Figure 2 in [3]).

For a room in *steady-state*, the energy balance can be expressed as (following [7] and [3])

$$\rho V C \frac{d}{dt} T_i = \dot{m}_{AC} C_{AC} (T_{AC} - T_i) + \dot{m}_{air} C_{air} (T_{air} - T_i) + \sum_{j \in N_i} R_{ji}^{-1} (T_j - T_i) + d\Big(t, N(t - \Delta t), \phi(t)\Big), \quad (3.8)$$

where T_i is the temperature of the air in the test-bed, T_{AC} is the temperature of the air of the AC duct, T_{air} is the temperature of the air inlet duct, T_j, $j \in N_i$ are the temperatures of the neighbor rooms and outside (see Figure 2.6, e.g. room A231, room A221b), N_i is the set of the neighbor rooms, N is the number of persons, ϕ represents the effect of solar flux, d is the overall disturbance, V is the room volume, ρ is the room air density, C ($\frac{\text{J}}{\text{kg}\cdot\text{K}}$) is the specific heat capacity of the air in the room, $\dot{m}_{AC}$ ($\frac{\text{kg}}{\text{s}}$) is the mass flow rate of the AC duct , $\dot{m}_{air}$ ($\frac{\text{kg}}{\text{s}}$) is the mass flow rate of the air inlet duct, R_{ji} ($\frac{\text{K}}{\text{W}}$) the thermal resistance of the wall that divedes two adjacent rooms.
We can approximate d in (3.8) as a linear function of N and consider (3.5) as the estimate of the number of the occupiers obtaining

$$\rho V C \frac{d}{dt} T_i \approx \dot{m}_{AC} C_{AC} (T_{AC} - T_i) + \dot{m}_{air} C_{air} (T_{air} - T_i) + \sum_{j \in N_i} R_{ji}^{-1} (T_j - T_i) + \Big[c_0 C(t) + c_1 C(t - \Delta t) - f_1 I(t - \Delta t)\Big] + v(t), \quad (3.9)$$

where $v(t)$ is the disturbance and the other symbols are defined in previous sections.

To obtain the discrete model, we choose a "weighted" version of the *central difference approximation*

$$\frac{d}{dt}T_i \approx k_1 \frac{T_{i,n+1} - T_{i,n}}{\Delta t} + k_2 \frac{T_{i,n} - T_{i,n-1}}{\Delta t}, \tag{3.10}$$

where $k_1 > 0$, $k_2 > 0$ and $k_1 + k_2 = 1$. Applying the approximation (3.10) to (3.9) we obtain

$$\begin{aligned} T_{i,n+1} + \beta_{i1} T_{i,n} + \beta_{i2} T_{i,n-1} = \alpha_{AC} T_{AC,n} + \alpha_{air} T_{air,n} \\ + \sum_{j \in N_i} \alpha_{ji} T_{j,n} + \alpha_0 C_n + \alpha_1 C_{n-1} + \alpha_I I_{n-1} + \alpha v_n, \end{aligned}$$

where $\alpha_{AC} = \alpha \dot{m}_{AC} C_{AC}$, $\alpha_{air} = \alpha \dot{m}_{air} C_{air}$, $\alpha_{ji} = \frac{\alpha}{R_{ji}}$, $\alpha_0 = \alpha c_0$, $\alpha_1 = -\alpha c_1$, $\alpha_I = -\alpha f_1$, $\alpha = \frac{\Delta t}{\rho V C k_1}$, $\beta_{i1} = \frac{1-2k_1}{k1} + \alpha_{AC} + \alpha_{air} + \sum_{j \in N_i} \alpha_{ji}$, $\beta_{i2} = \frac{k_1 - 1}{k1}$. This model can be seen as an ARMAX model and rewritten as

$$\begin{aligned} (1 + \beta_{i1} q^{-1} + \beta_{i2} q^{-2}) T_n = \alpha_{AC} T_{AC,n-1} + \alpha_{air} T_{air,n-1} + \sum_{j \in N_i} \alpha_{ji} T_{j,n-1} \\ + (\alpha_0 + \alpha_1 q^{-1}) C_{n-1} + \alpha_I I_{n-2} + \alpha v_{n-1}. \end{aligned} \tag{3.11}$$

Central Temperature Model

In this model we assume that the temperature of the room T_i is the measure provided by sensor 15 which is located in the middle of the room. We consider the measurements of sensors 5,6,7,8 as the temperatures of the wall surfaces: we insert in the formula also the effect of the heat transfer of the wall for convection.

In this case the energy balance becomes

$$\begin{aligned} \rho V C \frac{d}{dt} T_i = \dot{m}_{AC} C_{AC} (T_{AC} - T_i) + \dot{m}_{air} C_{air} (T_{air} - T_i) \\ + \sum_{j \in N_i} R_{ji}^{-1} (T_j - T_i) + \sum_{k \in N_{wall}} h_{wk} S_{wk} (T_{wk} - T_i) \\ + d\Big(t, N(t - \Delta t), \phi(t)\Big), \end{aligned}$$

where N_{wall} is the set of sensors that are close to the wall, h_{wk} $(\frac{\mathrm{W}}{\mathrm{m^2K}})$ is the convection heat transfer coefficient of the wall k, S_{wk} is the surface of the wall k. Following the same procedure of the mean temperature model, we have

$$\begin{aligned}(1+\beta_{i1}q^{-1}+\beta_{i2}q^{-2})T_n &= \alpha_{AC}T_{AC,n-1}+\alpha_{air}T_{air,n-1}\\ &+\sum_{j\in N_i}\alpha_{ji}T_{j,n-1}+(\alpha_0+\alpha_1 q^{-1})C_{n-1}+\alpha_I I_{n-2}\\ &+\sum_{k\in N_{wall}}\alpha_{wk}T_{wk,n-1}+\alpha v_{n-1},\end{aligned} \tag{3.12}$$

where $\alpha_{wk}=\alpha h_{wk}S_{wk}$, $\beta_{i1}=\frac{(1-2k_1)}{k1}+\alpha_{AC}+\alpha_{air}+\sum_{j\in N_i}\alpha_{ji}+\sum_{k\in N_{wall}}\alpha_{wk}$.

3.1.3 Humidity

The model for the humidity is obtained in a similar way to the model of the temperature. The humidity of the room is considered to be equal to the mean of the humidities provided by sensors 5,6,7,8 and 15.

We assume that the humidity of the room depends on the humidity of the air coming out from the inlet and the AC ducts, the temperature of the room, the number of occupiers of the room and other unknown factors. So the equation is

$$\frac{d}{dt}H=\lambda_{AC}(H_{AC}-H)+\lambda_{air}(H_{air}-H)+h(t,T_i,N),$$

where H_{AC} and H_{air} are respectively the humidities of the air coming from the AC and the inlet and $h(t,T_i,N)$ is the disturbance that depends on time, room temperature and number of persons.

We can approximate $h(t,T_i,N)$ with a function that is linear in T_i and N and obtain, using for N the approximation (3.5),

$$\begin{aligned}\frac{d}{dt}H(t)\approx{}&\lambda_{AC}(H_{AC}-H)+\lambda_{air}(H_{air}-H)+\lambda_T T_i\\ &+\Big[c_0C(t)+c_1C(t-\Delta t)-f_1I(t-\Delta t)\Big]+z(t),\end{aligned}$$

where $z(t)$ is the disturbance, λ_{AC} and λ_{air} $[\frac{1}{\mathrm{s}}]$ are unknown gains, λ_T $(\frac{\%\mathrm{RH}}{\mathrm{sK}})$ is the temperature unknown gain and the other terms has the same meaning as in Section 3.1.2.

Applying the *"weighted" central difference approximation* as in the temperature model (see Section 3.1.2) we obtain

$$\begin{aligned}H_{n+1}+\delta_{i1}H_n+\delta_{i2}H_{n-1}={}&\gamma_{AC}H_{AC,n}+\gamma_{air}H_{air,n}+\gamma_T T_{i,n}\\ &+\gamma z_n+\gamma_0 C_n+\gamma_1 C_{n-1}+\gamma_I I_{n-1},\end{aligned}$$

where $\gamma_{AC}=\lambda_{AC}\gamma$, $\gamma_{air}=\lambda_{air}\gamma$, $\gamma_0=c_0\gamma$, $\gamma_1=-c_1\gamma$, $\gamma_I=-f_1\gamma$, $\gamma=\frac{\Delta t}{k_{H1}}$, $\delta_i=\frac{1-2k_{H1}}{k_{H1}}+\gamma_{AC}+\gamma_{air}$, $\delta_{i2}=\frac{k_{H1}-1}{k_{H1}}$, k_{H1} the weighting constant of the difference approximation. The model can be written as

$$(1+\delta_{i1}q^{-1}+\delta_{i2}q^{-2})H_n = \gamma_{AC}H_{AC,n-1}+\gamma_{air}H_{air,n-1}+\gamma_T T_{i,n-1} \\ +\gamma z_{n-1}+(\gamma_0+\gamma_1 q^{-1})C_{n-1}+\gamma_I I_{n-2}, \tag{3.13}$$

which is an ARMAX model.

3.2 Identification Methods

The models proposed above are ARMAX systems of some orders.

For all the models, the Prediction Error Method (**PEM**) is used to identify the unknown parameters. A SISO model can be considered as example

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t),$$

where θ is the vector of the parameters to be estimated, $u(t)$ is the input, $y(t)$ is the output and $e(t)$ is a zero mean white noise with finite variance. The PEM method minimizes, with respect to θ, the function of the prediction error $\varepsilon_F(t,\theta)$

$$V_N(\theta, Z^N) = \frac{1}{2N}\sum_{t=1}^{N}\varepsilon_F^2(t,\theta), \tag{3.14}$$

where Z^N is a vector which contains the collected input-output data ($Z^N = [y(1)\ u(1)\ y(2)\ u(2)\ \ldots\ y(N)\ u(N)]$). The prediction error is given by the difference between the output and the predictor values, $\varepsilon_F(t,\theta) = y(t,\theta) - \hat{y}(t,\theta)$, where

$$\hat{y}(t,\theta) = H^{-1}(q,\theta)G(q,\theta)u(t) + [1 - H^{-1}(q,\theta)]y(t).$$

For the central temperature model, parameters are also estimated with the Least Squares (LS) method, as one of our objectives is to compare our work with [7], where this method was used. After the temperature model is rewritten in the form $y = X\beta$, where y and X are the observed data and β is the parameter to be estimated, it is used the backslash operator in MATLAB® to apply the LS method ($\beta = (X^TX)^{-1}X^Ty$).

For the PEM method, the System Identification Toolbox in MATLAB® (see [12]) with the command `armax` is employed, which allows to specify the orders of the model. Actually we do not know the order of the polynomial $C(q)$ in the ARMAX model (see (3.4)), because in all models, CO_2 (3.3), mean temperature (3.11), central temperature (3.12) and humidity (3.13), a generic noise is considered. Therefore different orders for $C(q)$ are tried for each model. Worth noticing, one of the differences of the two types of identification for the temperature is the treatement of the noise: with LS it is considered as an unknown disturbance without any assumption on the structure, while with PEM its structure is determined choosing the order of $C(q)$.

The three models are also identified using the **structured state space** model estimation. Given a system in the state space form (θ is the parameter to be estimated)

$$\begin{cases} x(n+1) = \overline{A}(\theta)x(n) + \overline{B}(\theta)u(n) + v(n) \\ y(n) = \overline{C}(\theta)x(n) + e(n) \end{cases},$$

where $v(\cdot)$ and $e(\cdot)$ are uncorrelated zero mean white noises with covariance matrices $R_1(\theta)$ and $R_2(\theta)$ respectively, it is possible to determine the equation of the one-step predictor and then apply again the PEM method, minimizing the cost (3.14) with respect to θ. The optimal one-step predictor of $y(t)$ is given by the Kalman filter

$$\begin{aligned} &\hat{x}(n+1|n) = \overline{A}(\theta)\hat{x}(n|n-1) + \overline{B}(\theta)u(n) + K(\theta)\Big[y(t) - \overline{C}(\theta)\hat{x}(t|t-1)\Big], \\ &\hat{y}(n|n-1) = \overline{C}(\theta)\hat{x}(n|n-1), \end{aligned}$$

where

$$K(\theta) = \overline{A}(\theta)P(\theta)\overline{C}^T(\theta)\Big[\overline{C}(\theta)P(\theta)\overline{C}^T(\theta) + R_2(\theta)\Big]^{-1},$$

and $P(\theta)$ is the solution of the Riccati equation

$$P(\theta) = \overline{A}(\theta)P(\theta)\overline{A}^T(\theta) + R_1(\theta) - K(\theta)\overline{C}(\theta)P(\theta)\overline{A}^T(\theta).$$

We can consider again the CO_2 model to give an example of this approach. A state space realization of the CO_2 model (3.3) is

$$\begin{cases} x(n+1) = -a_1 x(n) + \begin{bmatrix} K_N & K_I \end{bmatrix} \begin{bmatrix} N(n) \\ I(n) \end{bmatrix} \\ C(n) = x(n) + e(n) \\ x(0) = x_0 \end{cases},$$

where $x(\cdot)$ is the state, x_0 is the initial condition, $e(\cdot)$ is a white noise and the other symbols have the same meaning of the equation (3.3). This fixed structure model is created with the MATLAB® function `idgrey` and unknown parameters a_1, K_N and K_I are identified with the PEM method (`pem` function).

Actually, this is a realization of a particular ARMAX model ($C(q,\theta) = A(q,\theta)$, see (3.4))

$$C_n = \frac{K_N}{1 + a_1 q^{-1}} N_{n-1} + \frac{K_I}{1 + a_1 q^{-1}} I_{n-1} + e(n). \tag{3.15}$$

Equation (3.15) could look like an Output Error model (OE) equation: however, the MISO OE in general has, for each input, transfer functions with different denominators. This model is used to try a different structure of the noise, as it

is nothing known about it. A similar approach is used for both temperature and humidity models.

Recently, a new identification approach based on **nonparametric** estimation of impulse responses has been proposed [16]. In the last part of the project, we will use this method to see what happens with a completely different approach. However, because of the limited time left, we do not analyze deeply the theory of this approach and we will give the reader only an idea about it.

We briefly recall that parametric approaches usually fix the maximal complexity of the model a priori, before seeing the data (e.g., orders of numerators and denominators for ARX, ARMAX, BJ). Nonparametric techniques instead do not pose this limitation, in the sense that, before seeing the data, the maximal complexity of the model is virtually unbounded. Here we focus on a particular and novel nonparametric technique, that looks directly for the *impulse response* of the system within an infinite-dimensional space of functions, without any guess or physical formulation about the most suitable structure, but considering that the to-be identified system is LTI. Indeed, the space of the possible estimation outcomes is restricted to the continuous with some derivatives bounded energy functions, which have also to decay esponentially to zero, as BIBO stable LTI systems impulse responses are requested.

To be more specific, if we consider the general MISO system (m inputs)

$$y_t = \sum_{i=1}^{\infty} f_i u_{t-i} + \sum_{i=0}^{\infty} g_i e_i, \tag{3.16}$$

where, for each time instant $t \in \mathbb{Z}$, $y_t \in \mathbb{R}$, $e_t \in \mathbb{R}$ and $u_t \in \mathbb{R}^{m\times 1}$, $f_t \in \mathbb{R}^{1\times m}$ and $g_t \in \mathbb{R}$, then the 1-step ahead linear predictor is given by

$$\hat{y}_{t|t-1} = \sum_{k=1}^{m} \left[\sum_{i=1}^{\infty} h_i^k u_{t-i}^k \right] + \sum_{i=1}^{\infty} h_i^{m+1} y_{t-i},$$

where $h_i^k = \{h_{i,t}^k\}_{t\geq 0}$, $k = 1 \dots m+1$ are the predictor impulse responses. The nonparametric model that we consider directly estimate these impulse responses. Therefore, for the MISO system (3.16), while the parametric identification method computes

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{N} \Big(y_t - \hat{y}_{t|t-1}(h_\theta) \Big)^2,$$

where we highlighted the fact that the impulse response relies on the parameter θ, the nonparametric method

$$\hat{h} = \arg\min_{h\in\mathcal{H}} \sum_{t=1}^{N} (y_t - \hat{y}_{t|t-1}(h))^2 + \eta ||h||_{\mathcal{H}}^2,$$

where $\mathcal{H}$ is the set of possible estimation outcomes[1] and $\eta||h||_{\mathcal{H}}^2$ is the term that weights the regularity of the function, namely, it takes into account exactly

[1]It can be shown that this is a dense subset of the continuous functions [16].

our a priori believes: the system has to be a stable LTI and so the impulse response must decay exponentially. This nonparametric method then trades off the fitting of the data and the regularity of the estimation outcome, and belongs to the family of the so-called Tikhonov regularization methods [17].

The MATLAB® function `SSpline`, which implements this method, is employed to identify the CO2 and the temperature models. The function returns an `idpoly` object for the identified model.

The code of `SSpline` has kindly been provided by Gianluigi Pillonetto, Assistant Professor of Control and Dynamic Systems at the Department of Information Engineering of the University of Padova, and we are very thankful to him.

3.3 Data Pre-processing

Before being used for identification, the data collected in the test-bed will be preprocessed in some way.
For the raw data we have the following problems:

- each sensor may lose data;
- measurements of different sensors are not aligned, namely, not all the measurements are available at the same time instant;
- we have non-accurate measurements of the temperature, humidity and CO_2 level of the air coming out from the inlet.

Interpolation is used to solve the first two problems. First, a new time vector is defined, whose components differed by the *chosen* sampling time t_{Int} (180s). The *actual* sampling time in the test-bed is 30 seconds: this means that every 30 seconds each sensor is supposed to send a measurement of either temperature and humidity or CO_2 level to the coordinator (see Section 2.2). However, data are resampled, because of the slow dynamics which we are dealing with, for temperature, humidity and CO_2 level.

The MATLAB® function `interp1` is employed, which returns, using the known measurements, the interpolated points corresponding to t_{Int}. The function allows to choose different methods of interpolation: linear and cubic spline interpolation are tried.

Let us now consider the third problem. As seen in Section 2.1, it is possible to manually open or close the damper between the inlet duct and the central ventilation system. The problem is that even with the damper completely closed, there is a low air flow coming out from the duct. As the sensor used to monitor the CO_2 concentration of the air inlet is placed just outside the duct (see Figure 3.2), the measurements provided by this sensor are always equal to the lower concentration air coming from the central ventilation system. We use this measurement when the inlet is turned on; when it is off the inlet

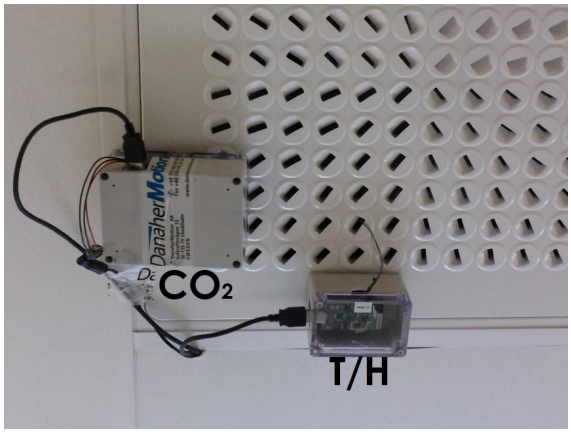


Figure 3.2: T/H and CO_2 sensors of the fresh air inlet.

concentration is *assumed* to be equal to the **sensor 17** one. Indeed sensor 17, as it can be seen in Figure 2.6, is pretty near to the inlet. This solution is reasonable: if no air is flowing in the room, with the damper closed, the CO_2 concentration of the inlet should gradually come close to the room CO_2 concentration near the inlet, and then the closest CO_2 sensor is considered. There is a similar problem for the temperature of the inlet: following the same approach, temperature measurements of **sensor 19** (see Figure 2.6) are taken when the inlet is off. Actually, the problem of the weak airflow coming with the closed damper exists only during "work time", from 6:00 to 17:00. In fact, out of this range of time, the central ventilation system (not under our control) is *automatically* turned off and no air is coming in the ducts of both AC and inlet (and with damper closed either). This is considered as a disturbance: indeed, when the weak flow is not coming into the room, the CO_2 and temperature increase, but the models do not take it into account with any input terms. However, we will see in Chapter 4 that this effect is really relevant and in future work it should not considered just as disturbance.

Last but not least, we consider removing the physical equilibrium offsets from the data. In fact, the identified model is a linearized version of the true system around the operating point. A more accurate model can be obtained using data without offset, since linear models are not able to explain arbitrary differences between the input and output. However, our models will not have all zero inputs and hence it is not such a big problem. To see this fact, let us consider the CO_2 model without noise

$$C_n = -a_1 C_{n-1} + K_N N_{n-1} + K_I I_{n-1}. \tag{3.17}$$

Supposed the system stable ($|a_1| < 1$), if data with offsets are used, in free response it will close zero instead of the equilibrium point, with bad consequences on its performances. Nevertheless, the concentration of the inlet I on the right-hand side of (3.17) can never be zero (see Section 2.1) and the system will never be in free response. Therefore, the main problem that one could

have with data with offsets (system in free response which goes to zero) never happens with our formulation, and good models can be identified using data with offsets. By the way, we have tried to determine an equilibrium point to shift the data around for CO_2, temperature and humidity, but better results can only be obtained with CO_2 model. Indeed, CO_2 level is not too sensitive to weather changes as temperature and humidity: the minimum value of CO_2 is used as equilibrium point.

A common way to deal with raw data is to remove the mean to all the dataset both identification and validation data. However, our models intend to be used for the online control and hence we should not use the mean of "future" data (validation data). A tradeoff could be to subtract the mean of the identification data to the validation data, but it is not possible due to the too much different conditions of the two datasets.

3.4 Identified Models

In this section we are going to identify the models proposed in Section 3.1.

3.4.1 CO_2 Models

As we have seen in Section 3.1, in order to identify the CO_2 model (3.3), we need to determine the parameters a_1, K_N and K_I. These parameters will be estimated with two methods which differ on the assumption on the noise structure, polynomial model estimation (**poly**) and structured state space model estimation (**ss**).

Parameters Estimation by Polynomial Approach

With this method we assume a general ARMAX structure for the model and the parameters, which are estimated using the PEM method, are depicted in Table 3.1.

a_1	K_N	K_I
-0.8964	2.108	0.0579

Table 3.1: Parameters of CO_2 model (3.3) identified by polynomial approach.

Parameters Estimation by State Space Approach

In Table 3.2 we can see the parameters estimated with the state space method, which sets the $C(q)$ polynomial equal to the $A(q)$ one (see Section 3.2). As we can notice in Table 3.2, the parameters of this method are close to those of the polynomial approach, and we guess that also the performances will be similar.

a_1	K_N	K_I
-0.8569	2.686	0.0812

Table 3.2: Parameters of CO_2 model (3.3) identified by state space approach.

3.4.2 Temperature Models

In Section 3.1.2 we have described the two types of temperature model that we consider: the first is based on the average temperature of the room and the second on the temperature of the sensor located in the middle of the room. We want to see the difference between the two models because single sensor measurements are noisy and this may cause more problems in the identification procedure. As we have done in Section 3.1.2, we will call them respectively mean temperature and central temperature models.

It should be pointed out that in the temperature models we do not consider the measurements coming from the other rooms and outside. The reasons are:

- the sensor placed outside is too sensitive to the sun and the wheather conditions, giving unlikely measurements of temperature and humidity. Probably, the sensor is not suitable to stay outside;

- the conditions of the other rooms surrounding do not affect the test-bed too much and neglecting them does not change anything.

Mean Temperature Models

The model that we consider is (removing the terms of the surrounding of the room $\sum_{j\in N_i} \alpha_{ji} T_{j,n}$):

$$\begin{aligned}(1+\beta_{i1}q^{-1}+\beta_{i2}q^{-2})T_n = \alpha_{AC}T_{AC,n-1} + \alpha_{air}T_{air,n-1} \\ +(\alpha_0+\alpha_1 q^{-1})C_{n-1} + \alpha_I I_{n-2} + \alpha v_{n-1}.\end{aligned} \tag{3.18}$$

As for the CO_2, the model (3.18) model is identified with different noise structure with polynomial and state space methods. The obtained parameters are listed in Table 3.3.

	poly	ss
β_{i1}	-1.179	-0.8655
β_{i2}	0.2567	0.1381
α_{AC}	0.0043	0.0106
α_{air}	0.0762	0.2719
α_0	0.0013	0.0027
α_1	$-9.96\cdot10^{-4}$	$-17.2\cdot10^{-4}$
α_I	$-3.06\cdot10^{-4}$	$-9.8\cdot10^{-4}$

Table 3.3: Parameters of temperature model (3.18) identified by polynomial (poly) and state space (ss) approaches respectively.

Central Temperature Models

For this model, temperature measurements from sensor 15 (see Figure 2.6) are considered as temperature of the room. The model, removing the contribute $\sum_{j\in N_i} \alpha_{ji} T_{j,n}$ becomes:

$$\begin{aligned}(1+\beta_{i1}q^{-1}+\beta_{i2}q^{-2})T_n = \alpha_{AC}T_{AC,n-1} + \alpha_{air}T_{air,n-1} \\ +(\alpha_0+\alpha_1 q^{-1})C_{n-1} + \alpha_I I_{n-2} \\ + \sum_{k\in N_{wall}} \alpha_{wk}T_{wk,n-1} + \alpha v_{n-1}.\end{aligned} \tag{3.19}$$

The obtained parameters can be seen in Table 3.4.

	poly	ss
β_{i1}	-1.037	-0.0951
β_{i2}	0.1036	0.2211
α_{AC}	0.0086	0.0033
α_{air}	0.1331	0.3221
α_0	0.0022	0.0027
α_1	-0.0018	-0.0027
α_I	$-4.66\cdot10^{-4}$	$-5.58\cdot10^{-5}$
α_{w1}	-0.0502	-0.0871
α_{w2}	0.03	1.195
α_{w3}	-0.0451	-0.2799
α_{w4}	-0.004	0.0166

Table 3.4: Parameters of temperature model (3.19) identified by polynomial (poly) and state space (ss) approaches respectively.

Symbols $w1$, $w2$, $w3$ and $w4$ stand for the walls near sensor 5, sensor 6, sensor 7 and sensor 8 (see Figure 2.6), respectively.

3.4.3 Humidity Models

The identified parameters of the humidity model (3.13) for the two identification approaches (polynomial and state space) are listed in Table 3.5.

	poly	**ss**
δ_{i1}	-1.6696	-0.1487
δ_{i2}	0.6844	-0.6206
γ_{AC}	$-2.83 \cdot 10^{-4}$	0.0085
γ_{air}	0.0147	0.2693
γ_T	$2.16 \cdot 10^{-4}$	-0.0814
γ_0	0.0016	0.0068
γ_1	-0.0018	-0.0073
γ_I	$4.98 \cdot 10^{-5}$	$4.9 \cdot 10^{-4}$

Table 3.5: Parameters of humidity model (3.13) identified by polynomial (poly) and state space (ss) approaches respectively.

3.4.4 Nonparametric Estimated Models

The nonparametric approach is tried to identify a different model for CO_2 and temperature. As seen in Section 3.2, the MATLAB® function `SSpline` is just employed to get the `idpoly` object of the identified models. We will not treat any details about the identified models, since we do not analyze deeply the theory of the method. It is anyway possible to study their performances, as the System Identification Toolbox can be employed with `idpoly` objects.

This attempt is to see what happens with a novel Black Box method and give some ideas for future work.

4
Validation of Models

In this chapter each identified model in Section 3.4 will be analyzed, showing its performances under the metrics which we are going to define in the following section.

4.1 Validation Metrics

When the models of CO_2, temperature and humidity are identified, it is necessary to define some metrics to evaluate their goodness. Our validation method consists of three metrics: Fit, Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Fit is defined as

$$fit := 100 \cdot \left(1 - \frac{|| \hat{\mathbf{y}} - \mathbf{y} ||}{|| \mathbf{y} - \frac{1}{N}\sum_{i=1}^{N} y(i) ||}\right), \tag{4.1}$$

where N is the number of components of the validation dataset, $\mathbf{y}$ is the $N \times 1$ vector of the output measurements and $\hat{\mathbf{y}}$ is the $N \times 1$ vector of the **simulated** model output. More precisely, each component of $\hat{\mathbf{y}}$ is the model response calculated using initial conditions and current and past values of input measurements, but not output measurements as prediction does. MAE is defined as

$$MAE := \frac{1}{N}\sum_{i=1}^{N} | \, y(i) - \hat{y}(i) \, |,$$

and MSE as

$$MSE := \frac{1}{N}\sum_{i=1}^{N} (y(i) - \hat{y}(i))^2,$$

where the symbols have the same meaning of equation (4.1). As we can guess from their names, MAE gives an average of the absolute errors, while MSE an average of the squared errors. The main difference is that with MAE the errors are weighted equally, while with MSE larger errors are weighted more heavily due to the square.

4.2 CO_2 Models

As seen in Section 3.4.1 and 3.4.4, three CO_2 models have been obtained: poly, ss and nonparametric.

4.2.1 CO_2 Poly and SS Models

Given the parameters a_1, K_N and K_I (see Section 3.4.1) it is possible to predict with iteration and therefore validate the model. The comparison between the CO_2 measurements and the predicted output for the two models is depicted in Figure 4.1: the fit results 46.5% for the polynomial approach and 47.5% for the state space one. The two models have similar performances, as it could be guessed looking at the parameters listed in Table 3.1 and 3.2. This result is also confirmed by MAE and MSE coefficients, as it can be noticed in Table 4.1.

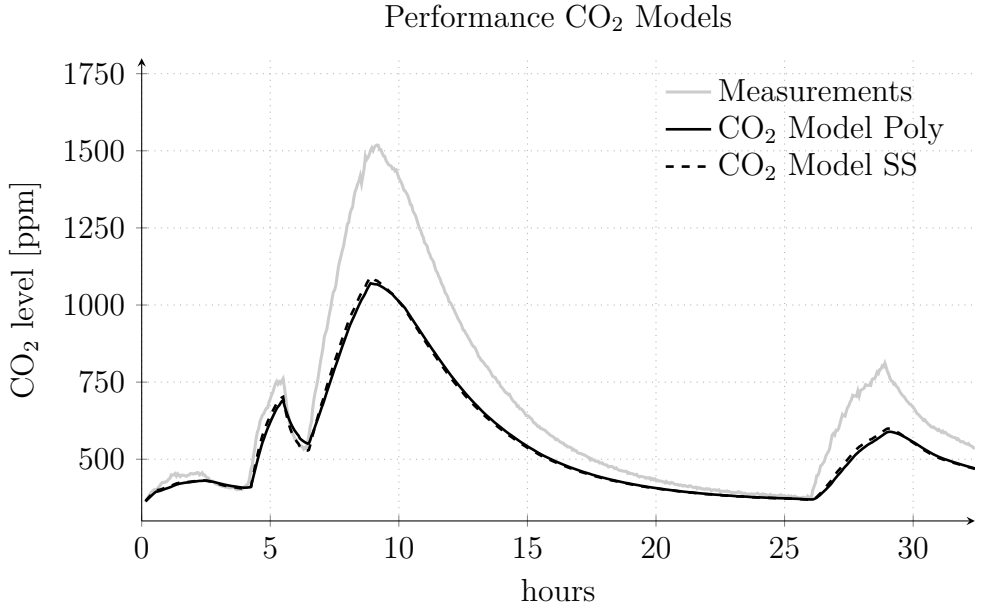


Figure 4.1: Validation of CO_2 models: real measurements and predictions of model (3.3) with parameters identified by polynomial (**fit 46.5%**) and state space (**fit 47.5%**) approaches.

	MAE (ppm)	MSE (ppm)
CO_2 poly	111.4	$2.64 \cdot 10^4$
CO_2 ss	109.3	$2.54 \cdot 10^4$

Table 4.1: Comparison beween MAE and MSE of CO_2 model (3.3) identified by polynomial (poly) and state space (ss) approaches.

4.2.2 Nonparametric CO_2 Model

The CO_2 model has also been identified with the Non Parametric method (**NP**), which uses the same inputs but different model structure. Performances are shown in Figure 4.2 and Table 4.2. Figure 4.2 and Table 4.2 show that

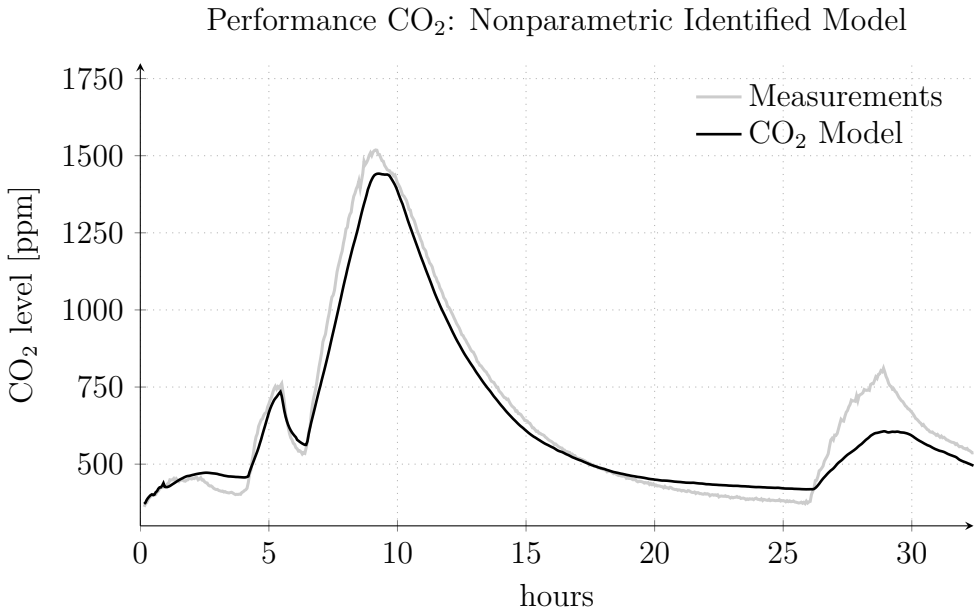


Figure 4.2: Validation of the CO_2 model identified with NonParametric (NP) method: real measurements and predictions of CO_2 NP model, **fit 77.8%**.

	MAE (ppm)	MSE (ppm)
CO_2 poly	111.4	$2.64 \cdot 10^4$
CO_2 NP	49.1	$4.25 \cdot 10^3$

Table 4.2: Comparison beween MAE and MSE of the CO_2 model identified by polynomial (poly) and NonParametric (NP) approaches.

the model identified with the NonParametric method predicts well the changes in the test-bed.

4.2.3 Estimation for the Number of Occupiers

Figure 4.3 shows the comparison between the estimation and the record of the number of occupiers. The estimation appears acceptable. Moreover, for our purpose, the high accuracy of this estimation is not required. Indeed, it is mainly used to get a theoretical formulation for the temperature and humidity models, which replaces the number of occupiers with CO_2 measurements, as inputs of the models.

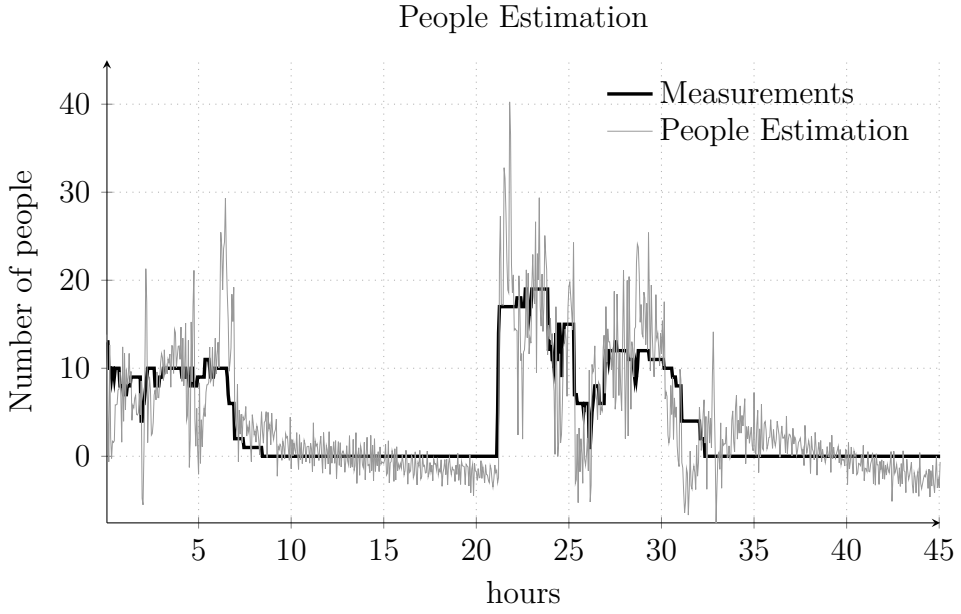


Figure 4.3: Estimation of the number of the people in the test-bed: real measurements and estimation (3.5) using **identification** data.

4.3 Temperature Models

Temperature models identified in Section 3.4.2 are here validated. A temperature model is also identified with the nonparametric method, using the same data of the mean temperature model.

4.3.1 Mean Temperature Models

The performances of the models are shown in Figure 4.4 and Table 4.3. It should be noticed that the ss model is better than the poly one: as we have already said for the CO_2, it should not be considered a general result. In fact, the type of the noise is never known and it is possible that for another test-bed

the structure assumed by the state space estimation could be unsuitable. It may be better not to assume a specific structure of the noise, as we have done in the polynomial estimation.

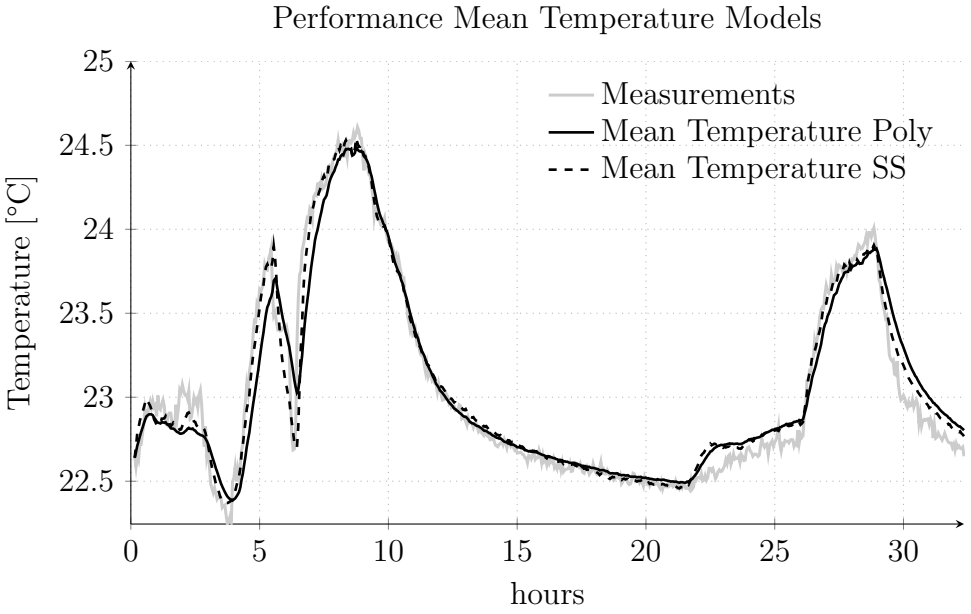


Figure 4.4: Validation of mean temperature models: real measurements and predictions of model (3.18) with parameters identified by polynomial (**fit 75.3%**) and state space (**fit 82.6%**) approaches.

	MAE (°C)	MSE (°C)
Temperature poly	0.12	0.0311
Temperature ss	0.0875	0.0178

Table 4.3: Comparison beween MAE and MSE of mean temperature model (3.18) identified by polynomial (poly) and state space (ss) approaches.

4.3.2 NonParametric Temperature Model

The nonparametric method is also employed to identify the mean temperature model: Figure 4.5 and Table 4.4 show the performances of this model. To identify this model, data without offsets are employed, since the same data used for the mean temperature model gives bad results. However, this is a tricky issue, since temperature equilibrium point is sensitive to the weather changes (see Section 3.3). The best result is obtained removing the minimum value of *identification* data to the overall dataset. Figure 4.5 and Table 4.4 show that the mean temperature model outperforms the nonparametric one.

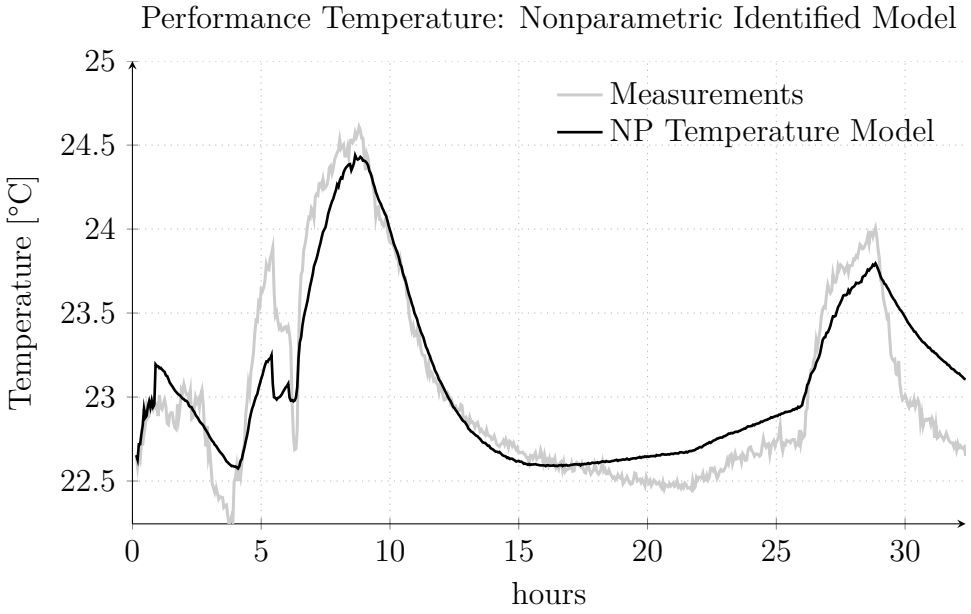


Figure 4.5: Validation of temperature model identified with NonParametric (NP) method: real measurements and predictions of NP temperature model, (**fit 59.6%**).

	MAE (°C)	MSE (°C)
Temperature poly	0.12	0.0311
Temperature NP	0.1739	0.05

Table 4.4: Comparison beween MAE and MSE of temperature model identified by polynomial (poly) and NonParametric (NP) approaches.

4.3.3 Central Temperature Models

It should be kept in mind that in this model the effect of the heat transfer across the walls is considered using measurements of sensors 5,6,7 and 8 (see Figure 2.6). The identified parameters, the behaviour with the validation data and MAE and MSE coefficient can be seen respectively in Table 3.4, Figure 4.6 and Table 4.5. In this case the ss approach is better than the poly one.

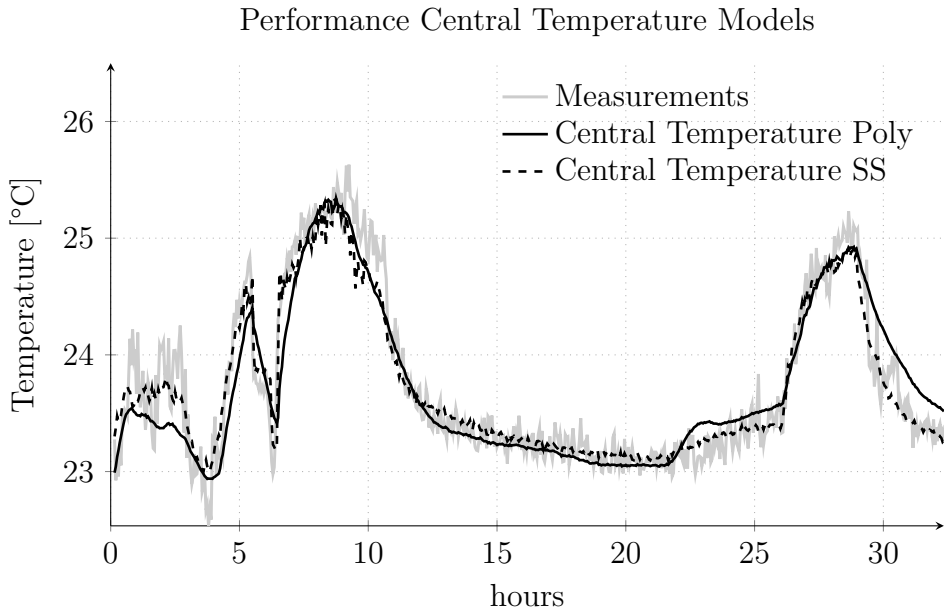


Figure 4.6: Validation of central temperature models: real measurements and predictions of model (3.19) with parameters identified by the polynomial (**fit 64.5%**) and state space (**fit 72.7%**) approaches.

	MAE (°C)	MSE (°C)
Temperature poly	0.2036	0.0769
Temperature ss	0.1522	0.0438

Table 4.5: Comparison beween MAE and MSE of mean temperature model (3.19) identified by polynomial (poly) and state space (ss) approaches.

4.4 Humidity Models

The parameters of humidity model (3.13) have been estimated in Section 3.4.3. The performances of the models are shown in Figure 4.7 and Table 4.6. In this case, the better humidity model is obtained with the general structure ARMAX model (poly).

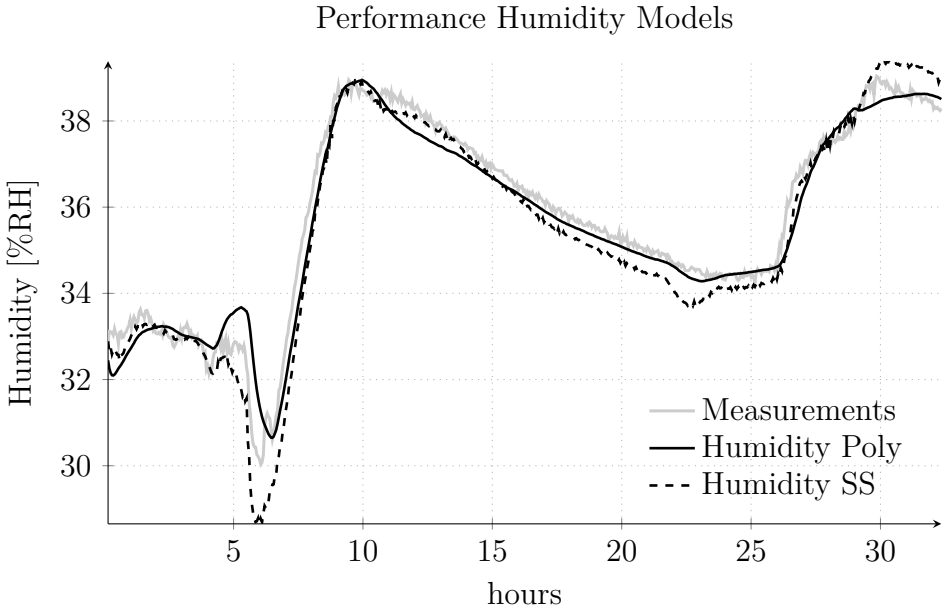


Figure 4.7: Validation of humidity models: real measurements and predictions of model (3.13) with parameters identified by polynomial (**fit 80.5%**) and state space (**fit 73.8%**) approaches.

	MAE (%RH)	MSE (%RH)
Humidity poly	0.3327	0.2044
Humidity ss	0.4461	0.339

Table 4.6: Comparison beween MAE and MSE of the humidity model (3.13) identified by polynomial (poly) and state space (ss) approaches.

4.5 Validate the Models in a Different Period

The identified models (identification data 16/05/2012 10:56:00 to 18/05/2012 08:00:00) are tested in the duration of about 1 day of the following month, from 12/06/2012 08:55:00 to 13/06/2012 14:00:00. In this period we weirdly had people in the test-bed and we collected data. In Figure 4.8, 4.9, 4.10, 4.11 and 4.12 we can see the behaviour of the poly CO_2, the NP CO_2, the poly mean temperature, the NP mean temperature and the humidity models. The ss models are slightly better (for CO_2 and temperature) in this situation, but we decide to report only the more general poly models results. We observe that the performances of the CO_2 and temperature models are good. The humidity one is not so good due to the very limited set of data that we dealt with.

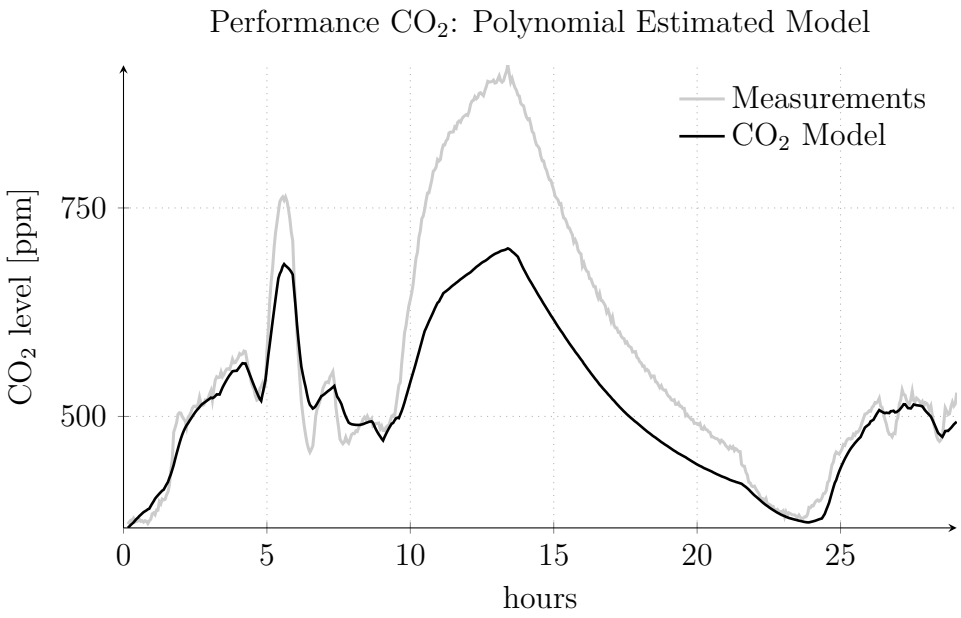


Figure 4.8: Validation of CO_2 model: real measurements and predictions of model (3.3) with parameters identified by polynomial approach (**fit 40.7%**) using validation data of a *different* month.

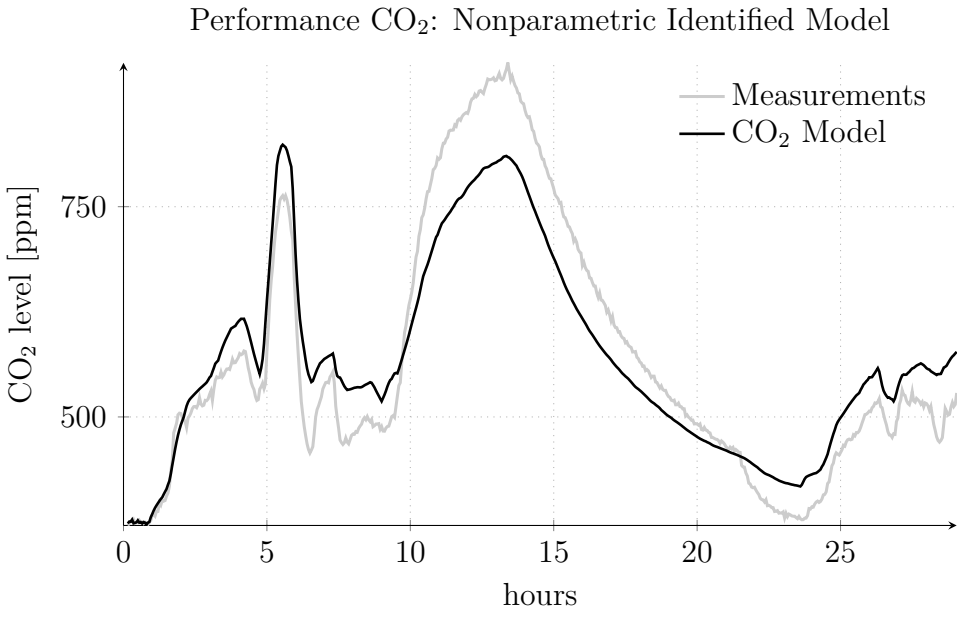


Figure 4.9: Validation of CO_2 NP model: real measurements and predictions of CO_2 NP model (**fit 64%**) using validation data of a *different* month.

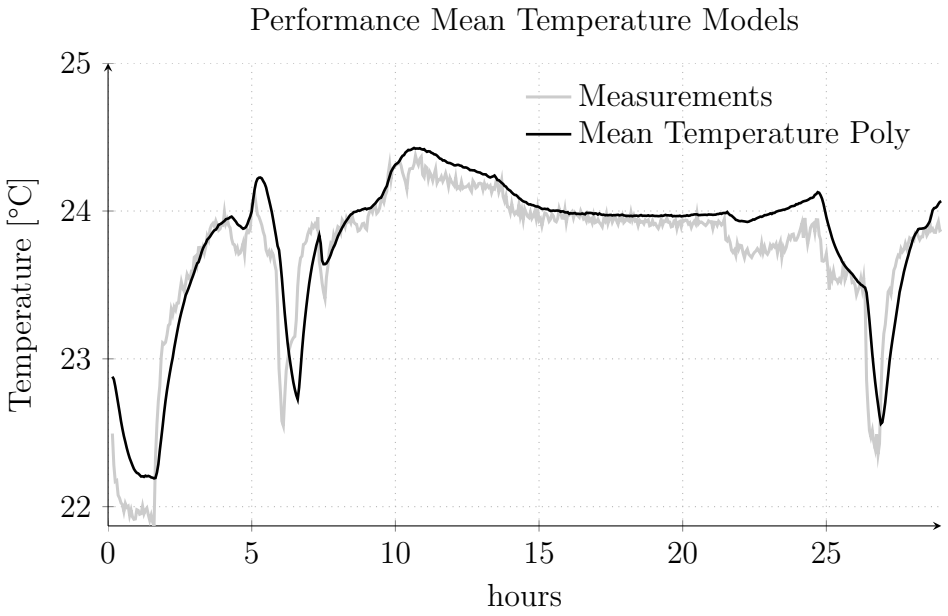


Figure 4.10: Validation of temperature mean model: real measurements and predictions of model (3.18) with parameters identified by polynomial approach (**fit 57.1%**) using validation data of a *different* month.

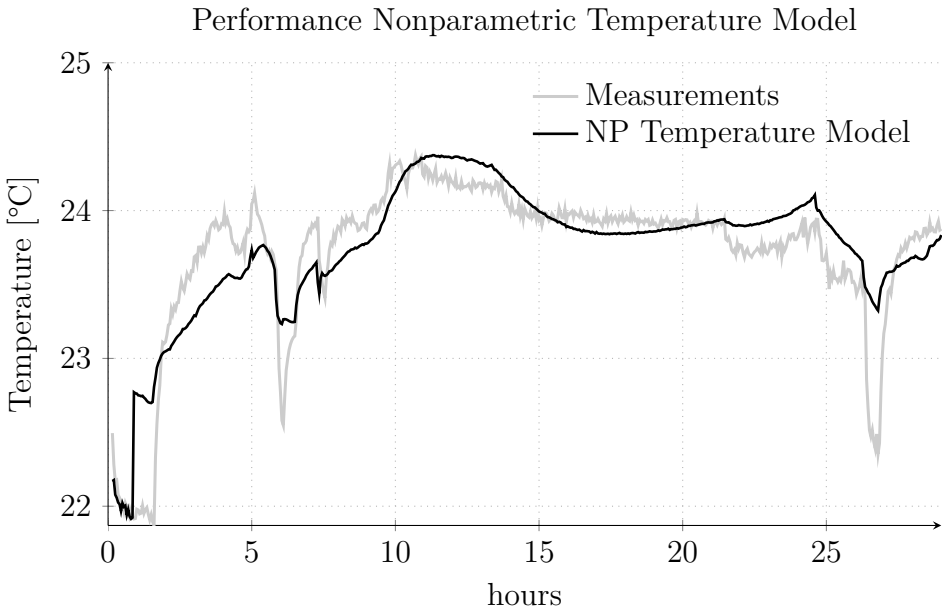


Figure 4.11: Validation of the NP temperature model: real measurements and predictions of the NP temperature model (**fit 51.1%**) using validation data of a *different* month.

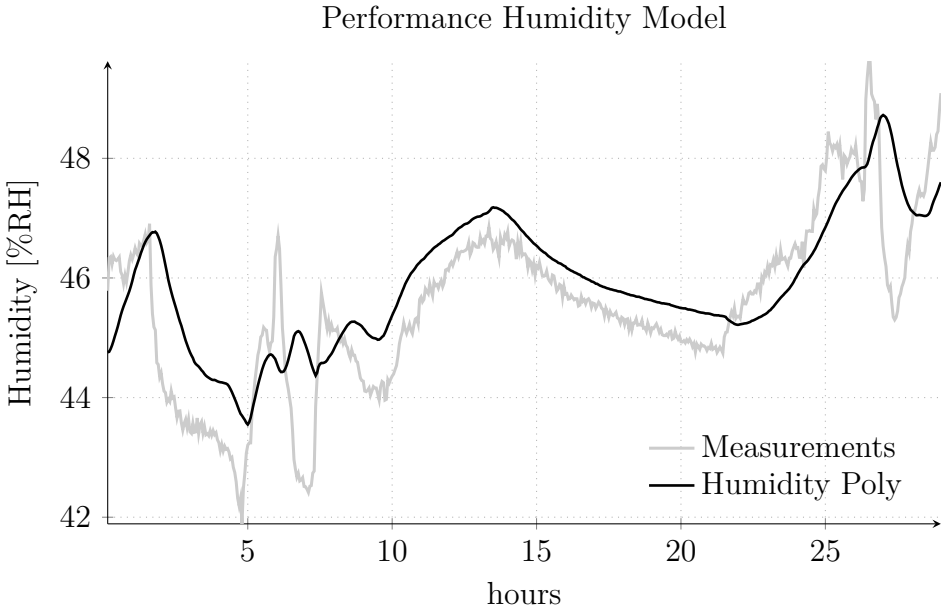


Figure 4.12: Validation of the humidity mean model: real measurements and predictions of model (3.13) with parameters identified by polynomial (**fit 34.3%**) using validation data of a *different* month.

Different from the CO_2 level, the temperature and the humidity models are sensitive to the weather changes. Probably, the set of the identification data for the humidity is so different from the day considered in June and the model does not work as well as it did for predictions of the days in May. This indicates that the models should be trained in a longer set of data under different weather conditions "to be prepared" for all types of validation dataset.

It is noticed that the nonparametric approach has good performances also in this analysis: our suggestion is then to try again to use this method in future work and test its performances in a longer set of data to be sure to have a general result. However, it should be pointed out that this method is a Black Box approach, which neglects all physical characteristics of the system.

4.6 Comparisons with Previous Work

So far we have shown the performances of our models implemented in two ways. Since temperature models are our main results, it is useful to compare them with some previous work in this field.

4.6.1 Comparison with the Reference

Our idea is based on [7]: the innovation is to consider the effect of the people in the temperature dynamics, introducing some terms based on the CO_2 level.

The model presented in the reference is given by

$$T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t\Big[P_{rm}T_{rm}(n) + P_{disch}T_{disch}(n) + P_{oa}T_{oa}(n) + \Delta E(n)\Big], \tag{4.2}$$

where T_{rm} is the temperature of the room, T_{disch} is the discharge air temperature, T_{oa} is the outside air temperature, Δt is the sample interval, ΔE is the disturbance and P_{rm}, P_{disch} and P_{oa} are the parameters to be identified. To make a fair comparison, a new model should be identified for our test-bed. Since our test-bed does not provide us with enough accurate outside measurements, the model (4.2) is not applicable in our case. The effect of the outside temperature may be included through the wall and the window temperatures, which are considered a linear function of T_{rm} and T_{oa} (see [7] for details). We consider the following model

$$T_{rm}(n+1) = T_{rm}(n-1) + 2\Delta t\Big[P_{rm}T_{rm}(n) + P_{disch}T_{disch}(n) + P_{wa_i}T_{wa_i}(n) + P_{wd_i}T_{wd_i}(n) + \Delta E(n)\Big], \tag{4.3}$$

where T_{wa_i} is the temperature of the internal surface of the external wall, T_{wd_i} is the temperature of the internal surface of the window and P_{wa_i} and P_{wd_i} are other two parameters to be identified. The reason to consider only the external wall with its window temperatures is due to the fact that the temperature difference between the room and the outside air is 2200% greater than the difference between two rooms in average, and the heat convenction is more involved there. There are still three problems: there is no sensor to give the measurement of the temperature of the window surface, we have two air discharge devices (AC and inlet) and the model (4.3) is unstable. In fact, an equivalent equation is

$$(1 - 2\Delta t P_{rm}q^{-1} - q^{-2})T_{rm}(n) = 2\Delta t\ \Big[P_{disch}T_{disch}(n-1) + P_{wa_i}T_{wa_i}(n-1) + P_{wd_i}T_{wd_i}(n-1) + \Delta E(n-1)\Big],$$

and we can see that the polynomial on the left-hand side is unstable for every P_{rm}. Therefore, our solutions are:

- consider only one term $P_{out}T_{out}$ for both wall and window, using the measurements of the sensor 7 (see Figure 2.6), which is pretty close to the window and the wall;
- introduce a new parameter to get a stable model, as it was done in Section 3.1.2, employing the weighted central difference approximation, see equation (3.10);
- split the discharge air term into two terms $P_{AC}T_{AC}$ and $P_{inl}T_{inl}$, since in our test-bed we have both the AC and the inlet which discharge air, see Chapter 2.1.

Finally, the model that we will use is

$$(1 + P_{rm1}q^{-1} + P_{rm2}q^{-2})T_{rm}(n) = P_{AC}T_{AC}(n-1) + P_{inl}T_{inl}(n-1) + P_{out}T_{out}(n-1) + P_E\Delta E(n-1), \tag{4.4}$$

where it is applied the same approach of Section 3.1.2. Unknown parameters P_{rm1}, P_{rm2}, P_{AC}, P_{inl} and P_{out} are then estimated with the Least Squares method.

The performances of model (4.4) are compared with the central temperature model, which in this case is identified with the Least Squares Method, as in [7]. We can see the identified parameters for the two models in Table 4.7 and the performances in Figure 4.13 and Table 4.8.

β_{i1}	-0.696
β_{i2}	-0.2087
α_{AC}	0.0111
α_{air}	0.1424
α_0	0.0025
α_1	-0.0021
α_I	$-4.32\cdot10^{-4}$
α_{w1}	-0.0668
α_{w2}	0.0567
α_{w3}	-0.07
α_{w4}	0.0283

(a)

P_{rm1}	-0.7556
P_{rm2}	-0.1406
P_{AC}	0.0046
P_{inl}	0.0389
P_{out}	0.0665

(b)

Table 4.7: Parameters of central temperature model (3.19) (a) and reference model (4.4) (b) identified by LS method.

	MAE (°C)	MSE (°C)
Central temperature model LS	0.2106	0.0785
Reference model	0.2623	0.1197

Table 4.8: Comparison beween MAE and MSE of the reference model (4.4) and the central temperature model (3.19) identified by LS method.

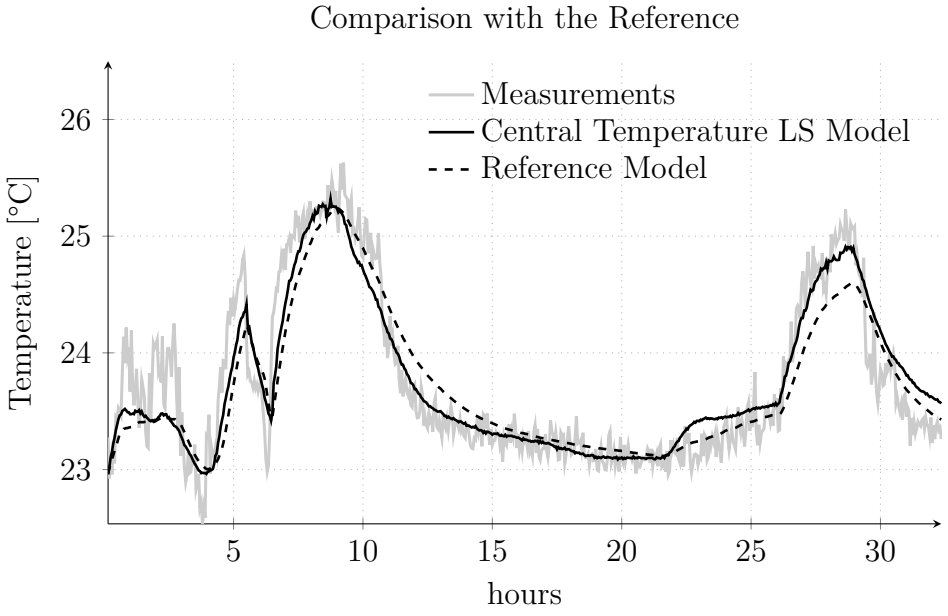


Figure 4.13: Comparison with the reference: real measurements, predictions of the reference model (**fit 56%**) and predictions of the central temperature model with parameters identified by **LS** method (**fit 64.4%**).

4.6.2 Previous Thesis Work

Another similar work is Florian Christian David Haizmann's master thesis project [1], which was done last year at KTH. Florian identified an Output Error (OE) model for a small conference room on the 6$^{\text{th}}$ floor of the Q-building at KTH, using the black box method. The model only considered the *temperature* and had the following structure

$$y(t) = \sum_{i=1}^{2} \left(\frac{B_i(q)}{F_i(q)} u_i(t-1) \right) + e(t),$$

where $y(t)$ is the temperature, u_1 is the binary input related to the AC (on/off), u_2 is the number of occupiers of the room, $B_1(q) = b_{11}$, $B_2(q) = b_{21} + b_{22}q^{-1}$, $F_1(q) = 1 + f_{11}q^{-1}$, $F_2(q) = 1 + f_{21}q^{-1} + f_{22}q^{-2}$ are the polynomials whose coefficients have to be identified and e is a white noise.

Since our test-bed is completely different from the one in Florian's, we need to identify a new model, employing his technique and introducing some adjustments to make a fair comparison. In our test-bed there is also the air inlet (see Section 2.1) as input, whose effect has to be considered in the model. When the black box method is used, the orders of the polynomials to be identified are just a guess, looking for the best result with some tries. Finally, the model is

$$y(t) = \sum_{i=1}^{3} \left(\frac{B_i(q)}{F_i(q)} u_i(t-1) \right) + e(t), \tag{4.5}$$

where $y(t)$ is the temperature, u_1 is the number of occupiers of the room, u_2 is the *binary* input related to the AC (on/off), u_3 is the *binary* input related to the inlet (on/off), $B_1(q) = b_{11} + b_{12}q^{-1}$, $B_2(q) = b_{21} + b_{22}q^{-1}$, $B_3(q) = b_{31} + b_{32}q^{-1}$, $F_1(q) = 1 + f_{11}q^{-1} + f_{12}q^{-2}$, $F_2(q) = 1 + f_{21}q^{-1} + f_{22}q^{-2}$, $F_3(q) = 1 + f_{31}q^{-1} + f_{32}q^{-2}$ are the polynomials, coefficients of which are the parameters to be identified and e is a white noise. In Table 4.9 we can see the obtained coefficients of the model (4.5).

b_{11}	0.03535
b_{12}	-0.03462
b_{21}	-0.00707
b_{22}	0.00781
b_{31}	-0.5589
b_{32}	0.5427

(a)

f_{11}	-1.122
f_{12}	0.1255
f_{21}	-1.986
f_{22}	0.9863
f_{31}	-0.9635
f_{32}	-0.03428

(b)

Table 4.9: Parameters of temperature model (4.5): B_i coefficients (a) and F_i coefficients (b) for $i = 1, 2, 3$.

Figure 4.14 and Table 4.10 show the perfomances of this model. Clearly, this

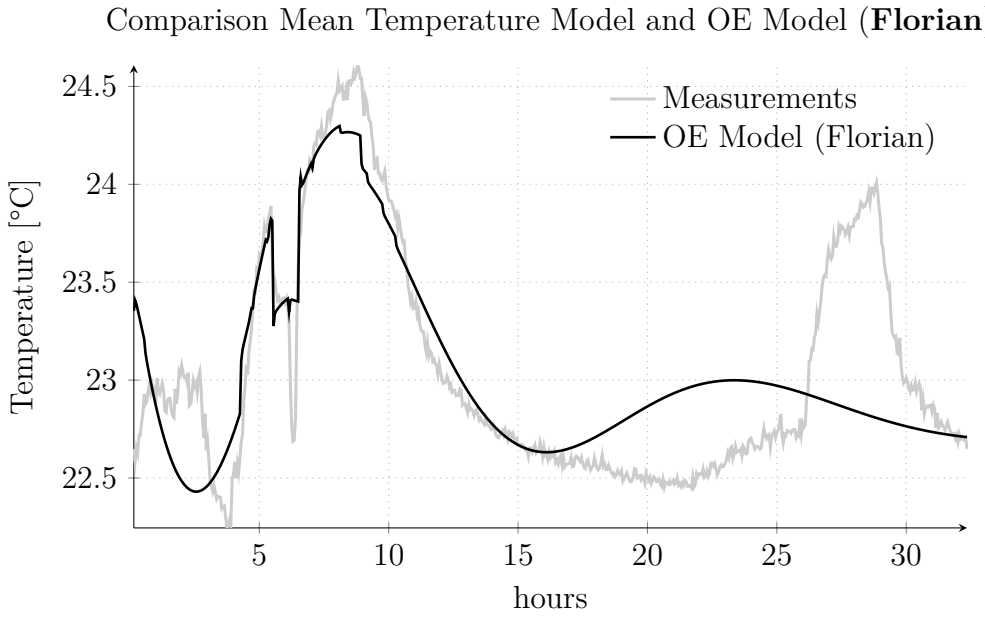


Figure 4.14: Validation of temperature model (4.5): real measurements and predictions of OE model (4.5) identified with the approach of the previous thesis work (**fit 36.8%**).

model is not suitable for our test-bed: in most of the data the output of the model is far from the real measurements, especially when all inputs are zero (from about sample 200) and the model is in free response. The last peak of the temperature, where this model has more difficulties, is due to the automatic close of the inlet (see Section 3.3).

	MAE (°C)	MSE (°C)
Temperature poly	0.12	0.0311
OE model (Florian)	0.2704	0.1386

Table 4.10: Comparison beween MAE and MSE of the temperature model (4.5) and the mean temperature model (3.18) identified by polynomial approach.

It should be underlined that:

- Florian chose a mean equilibrium temperature value to subtract to the data (see Section 3.3), according to the data that he had collected: in our test-bed we obtain the better result removing the minimum temperature value of the *identification* data. The same value is employed to remove trends in the validation data.
- in this model both AC and inlet are binary inputs (0 off, 1 on) while in our model we use the temperature of the air coming out from the two HVAC components.

5 Conclusions and Future Work

In this chapter we summarize our results and give some ideas for future work.

5.1 Discussion of Our Method and Improvements

Three physics based models for CO_2 level, temperature and humidity, respectively, have been proposed and identified. The performances of these models are satisfying: models are able to predict the changes in the test-bed, even with a short training dataset. We have derived the physics-based temperature and humidity models in a novel way by employing an estimation of the number of the occupiers of the room based on the CO_2 level. Experimental results show that it has been a good idea and our models outperform those in the previous work ([1] and [7]), as seen in Chapter 4.

Identified models have also been validated in a day with data collected one month later the identification period (see Section 4.5): the aim was to analyse their behaviour under different conditions. CO_2 and temperature models still work well, while the humidity results are not satisfying, which implies that the models should be trained with more data.

As final part of the project, we have also tried a novel method of non-parametric identification, which is described in [16]. We directly employed its MATLAB® implementation: models show good performances, but it should be kept in mind that all physical characteristics are neglected and a larger dataset should be uset to get a general result.

Before moving on to the next stage, one should solve some problems. The sensor which gave the external temperature should be replaced with one less

sensitive to the sun and the weather conditions. It could be interesting to consider the effect of the outside conditions in the model and compare the performances with ours. Moreover, we think that the contribute of neighbor rooms can definitely be neglected: only the heat convection of the inside walls surfaces should be taken into account, as we have done in our actual central temperature model (see equation (3.19)). However, sensors should be located to be in contact with the wall surface, and not just near, as in our test-bed.

Another important issue to be analyzed is to determine the equilibrium point for temperature and humidity according to the weather changes: this problem was also noticed in the previous thesis work [1], but it is still not solved. An accurate value of the equilibrium point could be helpful to get better results, as the identified model is a linearized version of the real system around that value (see Section 3.3).

We should consider the problem of the inlet measurements which we discussed in Section 3.3. When the damper of the inlet was closed, a weak airflow continued to come into the room, introducing a relevant disturbance. Measurements of respectively sensor 17 and sensor 19 were therefore employed for CO_2 and temperature when the inlet was off, because they were the nearest sensors. In future work these sensors should be placed closer, in order to get more accurate measurements when the inlet is off. A solution should be found also for the problem of the automatic closing of the central ventilation system (see Section 3.3). To understand the importance of this problem Figure 4.1 and 4.4 can be considered: the peak of CO_2 and temperature around sample 550, where inputs were all zero, was only due to the closing of the system. This effect should be considered in the models.

The main problem of this work is that we have a short dataset. It is indispensable to be sure to get at least one-two months of data in the next stage of this work, in order to have enough time to solve possible new problems and identify robust models. In fact, in the literature, months of measurements are normally used for the identification. In this way, it could be possible to make our models more robust and suitable for all seasons.

5.2 A New Approach: Hybrid Models

In Chapter 4 we have seen that the identified models work pretty well, even with the limited set of data that we have employed. Moreover, models are able to predict changes in the test-bed with different working conditions of AC and inlet. This is a really good point for our models, since the dynamics when AC and inlet are on or off is completely different.

We therefore think that a good idea may be to identify multiple models according to the state of the inlet or the AC: AC off inlet off, AC off inlet on, AC on inlet off and AC on inlet on. To limit the number of models we suggest to start to get different models for the two states of just the inlet: in fact, the

airflow of the inlet is definitely stronger than the AC one, following a more marked changes in the dynamics.

Bibliography

[1] Florian Christian David Haizmann, *Identification of an Indoor Air Quality Model for a Conference Room using a Wireless Sensor Network*, Master thesis at KTH (2012)

[2] Lennart Ljung, *System Identification: Theory for the user*, Prentice Hall PTR, Upper Saddle River and NJ, 2 edition, 1999

[3] Mehdi Maasoumy, *Modeling and Optimal Control Algorithm Design for HVAC Systems in Energy Efficient Buildings*, Master thesis at the Berkeley University (2011)

[4] Jacob Chi-Man Yiu, Shengwei Wang, *Multiple ARMAX modeling scheme for forecasting air conditioning system performance*, Energy Conversion and Management 48 (2007) 2276-2285

[5] G. Mustafaraj ,J. Chen, G. Lowry, *Development of room temperature and relative humidity linear parametric models for an open office using BMS data*, Energy and Buildings 42 (2010) 348-356

[6] G. Mustafaraj ,J. Chen, G. Lowry, *Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office*, Energy and Buildings 43 (2011) 1452-1460

[7] Siyu Wu, Jian-Qiao Sun, *A physics-based linear parametric model of room temperature in office buildings*, Building and Environment 50 (2012) 1-9

[8] S.A. Ahmadi, I. Shames, F. Scotton, L. Huang, H. Sandberg, K.H. Johansson, B. Wahlberg, *Towards More Efficient Building Energy Management Systems*, Knowledge, Information and Creativity Support Systems (KICSS) conference 2012

[9] Catalin Teodosiu, Raluca Hohota, Gilles Rusaou,en, Monika Woloszyn, *Numerical prediction of indoor air humidity and its effect on indoor environment*, Building and Environment 38 (2003) 655-664

[10] Qi Qi, Shiming Deng, *Multivariable control of indoor air temperature and humidity in a direct expansion (DX) air conditioning (A/C) system*, Building and Environment 44 (2009) 1659-1667

[11] Moteiv Corporation, *Tmote Sky: Ultra low power IEEE 802.15.4 compliant wireless sensor module*, http://www.snm.ethz.ch/pub/uploads/Projects/tmote_sky_datasheet.pdf

[12] *System Identification Toolbox*, http://www.mathworks.it/products/sysid/

[13] T. Barker, *Climate Change 2007: Synthesis Report*, http://www.ipcc.ch

[14] Ed Bas, *Indoor Air Quality A Guide for Facility Managers*, Fairmont Press and Marcel Dekker, Lilburn and Ga and New York, 2 edition, 2004

[15] Luis Perez-Lombard, Jose Ortiz, Christine Pout, *A review on buildings energy consumption information*, Energy and Buildings 40 (2008) 394-398

[16] G. Pillonetto, G. De Nicolao (2010, January), *A new kernel-based approach for linear system identification*, Automatica 46(1), 81-93

[17] G. Wahba (1990), *Spline models for observational data*, SIAM

[18] Swegon, *Air Diffusers*, www.swegon.com

[19] Swegon PACIFIC, *Integrated climate beam*, www.swegon.com