



**HAL**  
open science

# A data-driven methodology to predict thermal behavior of residential buildings using piecewise linear models

M.H. Benzaama, L.H. Rajaoarisoa, B. Ajib, S. Lecoeuche

## ► To cite this version:

M.H. Benzaama, L.H. Rajaoarisoa, B. Ajib, S. Lecoeuche. A data-driven methodology to predict thermal behavior of residential buildings using piecewise linear models. *Journal of Building Engineering*, 2020, 32, 10.1016/j.jobe.2020.101523 . hal-03224418

**HAL Id: hal-03224418**

**<https://hal.science/hal-03224418v1>**

Submitted on 17 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# A data-driven methodology to predict thermal behaviour of residential buildings using piecewise linear models

M. H. Benzaama<sup>a,\*</sup>, L. H. Rajaoarisoa<sup>b</sup>, B. Ajib<sup>b</sup>, S. Lecoeuche<sup>b</sup>

<sup>a</sup>COMUE Normandie University - ESITC-Caen Research Lab, 14610 Epron, France

<sup>b</sup>IMT Lille Douai, Univ. Lille, Informatics and Control Systems, F-59000 Lille, France

## Abstract

Nowadays, data-driven approaches are a good way to estimate very efficient black box models for different engineering systems. This class of model is well recognized by its outstanding performances to describe the overall behavior of one system based on its input-output relationships without any physical knowledge. In the context of building modeling, this approach is particularly well suited to predict future temperatures or energy consumption in a building. This paper presents an innovative method that uses input-output data to establish reliable and suitable thermal behavior models for residential buildings, especially for existing buildings where only measurements are available and no numerical models are at the disposal of the facility managers. The main paper contributions consist in the design of a new methodology based on the adaptation of a switched model estimation technique and in its validation to model accurately building thermal behaviors. The paper describes different stages needed to reproduce faithfully complex behaviors : data collection, PieceWise affine Auto-Regressive eXogenous (PWARX) identification technique, sensitivity analysis... It also explains how the procedure and the data-driven estimation algorithm are efficient in extracting sub-model parameters and sequence that give an outstanding ability to reproduce thermal dynamics of buildings, requiring the only collection of available data. The effectiveness of our methodology is discussed through experiments on different buildings located in the North of France. Indeed, through a comparative study between the piecewise ARX model and other existing models such as nonlinear ARX, indexed ARX and ARX models, the PWARX model gives good results in terms of indoor temperature estimation with 78.48% accuracy.

**Keywords:** Thermal behavior, Smart-metering, Data-driven models, PWARX, experiments on real buildings

## 1. Introduction

Control of energy consumption continues to be the primary concern in all areas of research [9; 11; 16]. Indeed, the optimization of energy consumption allows for improving the energy performance of the building. Furthermore, to design a set of optimal control, a thermal dynamics modeling step is necessary [20]. However, this task is complicated due to various factors influencing the thermal behavior, particularly by [19]: (i) climate, (ii) building envelope, (iii) building services and energy systems, (iv) building operation and maintenance, (v) occupant activities and behavior and (vi) indoor environmental quality provided.

Having an accurate thermal model plays a vital role in improving prediction and evaluating energy performance. In the literature, three main categories of modeling approaches have been considered [32]:

- white box models which are based on physical knowledge of the system and thermal balance equations. These

are often obtained through energy simulation software like EnergyPlus [4], TRNSYS [10], etc;

- black box models which use only measured input/output data and statistical estimation methods (e.g. [24; 26]);
- grey box models, a mix of the first two categories above. They use input/output data as well as some a priori knowledge on the system. A popular grey-box model is the equivalent RC networks [24; 34; 36].

A comparative study between these different models was done in [1]. The main conclusions of this study are that: (i) the use of white box model often requires important set-up and computation time; (ii) it also involves a large number of inputs to define the model, such as the composition of the building envelope for example [13]. In some studies, it is difficult, if not impossible, to recover this input [30]. To overcome this problem, data-driven methods have emerged in building framework. We find from the literature that the most commonly used data-driven techniques for building thermal modeling and energy performances prediction are based on the ARX (AutoRegressive eXogeneous) model [21; 31]. However, despite its performances, such as quick implementation, a good accuracy..., this model has important limits, mainly due to the estimation of a unique thermal model. A unique model

\*I am corresponding author

Email addresses: mohammed-hichem.benzaama@esitc-caen.fr (M. H. Benzaama), lala.rajaoarisoa@imt-lille-douai.fr (L. H. Rajaoarisoa), balsam.ajib@imt-lille-douai.fr (B. Ajib), stephane.lecoeuhe@imt-lille-douai.fr (S. Lecoeuche)

may not consider dynamic changes due to usages, equipment configurations or external factors, such as wind and solar radiation, that influence the thermal dynamics of the building.

Today, we can find several data-driven models derived from the ARX model. One can read the following references for further information [2; 15; 17; 22; 27; 29; 37]. Each of them is mainly differentiated by the parameter estimation techniques and the structure of the model. For instance, in [22], the authors tested the Fractional order Auto-Regression with eXogenous variable (FARX) model on building integrated energy systems. The model has been validated using the input-output data retrieve of a residential building simulated with software IES<VE>. As a result, it has been shown that the FARX model gives better accuracy than ARX one. Also, the results in [27] and [35] show that the Non-linear ARX (NARX) model performance was significantly greater than the one of ARX model. However, each technique has its own advantages and inconveniences, and one of our motivation is to discuss on how to recover the best data-driven model structure for simulating the thermal behavior of the building. Issues that one may remark into the following papers [33; 37] and [1; 22]. The first response to this discussion has been presented in [17] where the best prediction method includes a combination of two separate time-indexed ARX models to improve the prediction accuracy of the cooling load over different forecasting periods. In the following, we will also contribute to this discussion by making a comparison of existing ARX-derived techniques.

### 1.1. Paper contribution

In this paper, a black-box identification approach by means of a PieceWise affine auto-Regressive eXogenous (PWARX) model will be developed in the framework of building thermal modeling. This new methodology takes profit from recent advances in the hybrid system identification community. Hybrid and PWARX systems are heterogeneous dynamic systems that combine simultaneously continuous and discrete dynamics. These systems are helpful to introduce expert knowledge in the data-driven models, especially when various behaviors or uses have to be explained. They can be represented by switching models, i.e. by a set of continuous-submodels indexed by a discrete mode or a specific building management system setting. However, the idea of using partially observed regime switching models for building thermal modeling has not received satisfactory solutions apart from a few attempts which do not combine real world learning applications such as: missing data, dependent time series and noisy observations, see [14; 28]. Also, these works mainly highlight the necessity of using different models to represent different sorts of dynamics in a building. It was argued too in [2] and has been proved in the context of the prediction of buildings energy consumption into [3]. Then, the purpose of this paper will be to highlight the interest of using PWARX systems to model building thermal dynamics, in its 3 aspects: presentation of its scientific foundations, with regard to conventional ARX techniques; discussion around its interest to give explanations on different thermal dynamics; experimental comparison of the results obtained with different state-of-the-art methods.

Besides, this paper also discusses the selection of suitable inputs to define the best model structure. This will be done by performing sensitivity analysis and testing several configurations. This step aims to state on the influence of each input on the accuracy of the model and on the quality of identified parameters. Thus, the PWARX model should be able to explain the true thermal and energy behaviors of the building by identifying the use scenarios. Finally, the effectiveness of our methodology will be shown by presenting thermal modeling results for different building architecture located in the north of France. In particular, we assess the performance of our model through a comparative study between the piecewise ARX model and other existing models such as nonlinear ARX, indexed ARX and ARX models.

### 1.2. Paper outline

The rest of the paper is organized as follows. The formal definitions of the ARX model and its derivatives, as well as PWARX model are introduced in Section 2. The system identification technique used to identify and validate the PWARX model is detailed in Section 3. To illustrate the effectiveness of the proposed method, experimental results obtained from a student residential are provided in Section 4. Finally, conclusions and further works are given in Section 5.

## 2. Background

### 2.1. ARX model

Let us first consider an ARX model, using an input/output ( $y, u$ ) representation, writing as follows:

$$y(t) = -a_1y(t-1) - \dots - a_{n_a}y(t-n_a) + b_1u(t-1) + \dots + b_{n_b}u(t-n_b) + e(t) \quad (1)$$

where  $n_a$  and  $n_b$  are the model order,  $a_i$  and  $b_i$  are the model coefficients, and  $e(t) \in \mathbb{R}^{n_e}$  is a white noise process. In other terms, the ARX model can be defined by the following relations:

$$A(z)y(t) = B(z)u(t) + e(t) \quad (2)$$

with

$$A(z) = 1 + a_1z^{-1} + \dots + a_{n_a}z^{-n_a} \quad (3)$$

and

$$B(z) = b_1z^{-1} + \dots + b_{n_b}z^{-n_b} \quad (4)$$

where  $z$  is a backward shift operator.

So, for available input measurements, we can estimate the output ( $\hat{y}$ ) at each time  $t$  by the ARX model as follows:

$$\hat{y}(t) = B(z)u(t) + e(t) + (1 - A(z))y(t). \quad (5)$$

Finally, in compact form the estimate of the output can be writing as:

$$\hat{y}(t) = \varphi^T(t)\theta \quad (6)$$

with

$$\varphi(t) = [y(t-1) \dots y(t-n_a) \quad u(t-1) \dots u(t-n_b)]^T \quad (7)$$

and

$$\theta = [a_1 \cdots a_{n_a} \ b_1 \cdots b_{n_b}]^T. \quad (8)$$

In these equations,  $\varphi$  represents the regression vector and  $\theta$  the parameter vector. In the context of the design of a building thermal model,  $y$  generally represents the indoor temperature, and  $u$  represents the measured factors that influence the temperature evolution.

Taking the thermal ARX model example in [5], the indoor thermal behavior ( $T_i$ ) depends on the following inputs: the outdoor air temperature ( $T_o$ ), solar gains ( $Ra$ ), internal gains ( $Q_u$ ) and thermal gains ( $Q_h$ ). Thus, the relationship between the input-output is expressed by:

$$T_i(t) = - \sum_{j=1}^{n_a} a_j T_i(t-j) + \sum_{r=1}^{n_b} (b_{1,r} T_o(t-r) + b_{2,r} Ra(t-r) + b_{3,r} Q_h(t-r) + b_{4,r} Q_u(t-r) + v(t)). \quad (9)$$

## 2.2. NARX model

NARX is a nonlinear auto-regressive network with exogenous inputs. It has a recurrent dynamic nature and is commonly used in time-series modeling [6]. The formal definition of the NARX model is given by the following expression [8]:

$$y(t) = f([u(t-1) \cdots u(t-n_b) \ y(t-1) \cdots y(t-n_a)]) + e(t) \quad (10)$$

where  $y$  is the output variable,  $u$  is the deterministic externally variable, and  $f$  represents the nonlinear function.  $n_a$  and  $n_b$  represent the model order. Thus, the next value of the output signal depends on its previous values and those of the input signal (exogenous input) [6].

Basically, a NARX model can be implemented by using a feed-forward neural network to approximate the function  $f$ . This network can be used for nonlinear filtering of noisy signals or for predicting future behavior [23]. Moreover, it converges faster, and is less susceptible to variations and dependencies. It has typically better generalization abilities than other networks [15].

As an example, let us consider the NARX structure described in [15]. It uses as output the indoor temperature ( $T_i$ ) and as input variables the outdoor air temperature ( $T_o$ ), solar gains ( $Ra$ ), internal gains ( $Q_u$ ) and thermal gains ( $Q_h$ ). They are necessary to predict the internal temperature as follows:

$$T_i(t) = f([T_i(t-1), \cdots, T_i(t-n_a); T_o(t-n_k), \cdots, T_o(t-n_k-n_b); Ra(t-n_k), \cdots, Ra(t-n_k-n_b); Q_h(t-n_k), \cdots, Q_h(t-n_k-n_b); Q_u(t-n_k), \cdots, Q_u(t-n_k-n_b)]) + v(t), \quad (11)$$

## 2.3. Indexed ARX model

Several researchers introduce a more efficient ARX model by including physical understanding of the building in the model structure. This kind of model is called indexed ARX model. This model is more accurate than typical regression models and more efficient, in terms of computation time for instance. So, the ARX model is indexed with respect to different time intervals, different modes or configurations and

different inputs to improve accuracy and efficiency. The  $(n-q-m)$ -indexed model is then defined as a ARX model while considering for example  $n$  past times for the output ( $\mathbf{y} : \{y(t), y(t-1), \dots, y(t-n)\}$ ), a  $q$  modes or configurations and  $m$  input numbers ( $\mathbf{u} : \{u_1, u_2, \dots, u_m\}$ ). The input-output relationship can be written by the following expression:

$$y(t) = -a_{1,q}y(t-1) - \cdots - a_{n,q}y(t-n) + b_{1,q}u_1(t) + \cdots + b_{m,q}u_m(t) + e(t). \quad (12)$$

Practically, a 4-3-5 ARX indexed model has been defined in [27] by the following configurations: the load prediction *Load* as output and temperature  $T_o$ , humidity  $H_o$ , solar radiation  $Ra$ , wind speed  $W_d$  and people occupancy  $O_c$  as inputs, can be written as follows:

$$Load(t) = b_{1,q}T_o(t) + b_{2,q}H_o(t) + b_{3,q}W_d(t) + b_{4,q}Ra(t) + b_{5,q}O_c(t) + a_{1,q}Load(t-1) + a_{2,q}Load(t-2) + a_{3,q}Load(t-3) + a_{4,q}Load(t-4) + a_{5,q}. \quad (13)$$

**Remark 1.** Here three modes describe respectively : the daytime (8 AM to 9 PM), the transition time (6 AM to 8 AM), and the nighttime (9 PM to 6 AM).

## 2.4. PWARX model

PieceWise affine ARX model is a special class of non-linear models, obtained by splitting the state-input domain into a finite number of polyhedral regions. It can model a large number of physical processes such as residential building systems with several inherent operating modes or configurations. The general structure of a PWARX model is:

$$y(t) = f(\varphi(t)) + e(t) \quad (14)$$

with  $f$  as a piecewise affine map of the following form:

$$f(\varphi(t)) = \begin{cases} \theta_1^T \bar{\varphi}(t) & \text{if } \sigma(t) = 1 \\ \vdots & \\ \theta_q^T \bar{\varphi}(t) & \text{if } \sigma(t) = q, \end{cases} \quad (15)$$

where  $\bar{\varphi} = [\varphi^T \ 1]^T$  is the extended regression vector.  $\sigma(t)$  is the switching rule defined by:

$$\sigma(t) = i \text{ iff } \varphi(t) \in \mathfrak{R}_i, \text{ for } i = 1, \dots, q; \quad (16)$$

and  $\{\theta_i\}_{i=1}^q$  are the parameter vectors that define the sub models.  $\{\mathfrak{R}_i\}_{i=1}^q$  represent a complete partition of the region  $\mathfrak{R} \subset \mathbb{R}^n$ , with  $n = n_e n_a + (n_b + 1)$ , and each region is a convex polyhedron with:

$$\mathfrak{R}_i = \{\varphi \in \mathbb{R}^n : H_i \bar{\varphi} \leq \mathbf{0}\} \quad (17)$$

where  $H_i$  and  $\mathbf{0}$  are respectively a matrix of appropriate dimensions defining the limit of the region partitioning the set of regression vector and the null vector. To complete and give more details about the definition given above, figure 1 shows the partitioning into  $\{\mathfrak{R}_i\}_{i=1}^3$  region of the regression space according to scenario described by each parameter vector  $\{\theta_i\}_{i=1}^3$  (see [12] for more details).



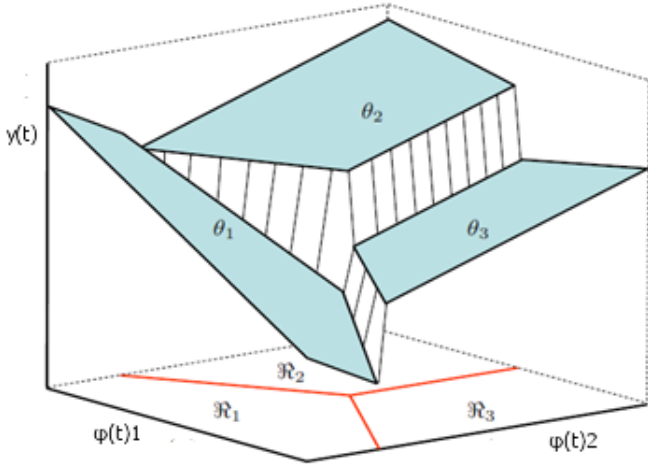


Figure 1: Representation of a PWARX system with 3 sub-models [12].

So PWARX model is defined by a set of ARX sub-models that help to analyze and predict the various thermal behaviors of the building. Each sub-ARX model is assumed to describe one mode or configuration [12].

Basically, the physical presentation of thermal behavior of indoor air by considering the  $q^{th}$  mode and with  $m$  input variables is defined by:

$$T_i(t) = \sum_{j=1}^{n_a} a_{j,q} T_i(t-j) + \sum_{v=1}^m \sum_{j=1}^{n_b} b_{j,q} u_v(t-j), \quad (18)$$

and in terms of outdoor air temperature ( $T_o$ ), outdoor humidity ( $H_o$ ), solar radiation ( $Ra$ ) and heating power ( $Pw$ ), it can be presented as:

$$T_i(t) = \sum_{j=1}^{n_a} a_{j,q} T_j(t-j) + \sum_{r=1}^{n_b} (b_{r,q}^1 T_0(t-r) + b_{r,q}^2 H_0(t-r) + b_{r,q}^3 Pw(t-r) + b_{r,q}^4 Ra(t-r)). \quad (19)$$

So, based on the input/output pair generated by the system described by relation (18), the identification of the PWARX model consists in (i) determining the number of sub-models in order to have the best fitting between the measured and estimated output, (ii) associating the data of each affine sub-model respectively, (iii) and finally, estimating each parameter vector  $\theta_i$  associated with these sub-models. Details about this identification procedure is given in the following section.

### 3. PWARX thermal model identification methodology

In this paper, we present an original data-based method to identify and estimate a set of models that are able to reproduce the variety of building thermal behaviors (Figure 2). So, to estimate all parameters of the PWARX model, we adopt an identification procedure defined by the following steps. Firstly, we design the experiment and the system necessary to collect the data. Then, we acquire some measurements from different

room parts, without requesting specific usages or functioning conditions. Our case study consists of student residential buildings located in Douai (France). Different scenarios will be considered during the measurement campaigns, such as the orientation (rooms situated at East/West), a different heating sequence and various periods of acquisition in the objective to provide sufficient data for discussing on the validation of our new methodology.

Secondly, we proceed to the determination of the model structure by a sensitivity study according to several parameters. Indeed, the sensitivity analysis makes it possible to identify the parameters of influence of the model for the prediction and to choose well (or validate) the model structure for a maximum precision and a minimum computation time in the context of building thermal modeling. The main objective of this study is then to give a good approach to estimate switched thermal dynamics and then to reproduce complex building thermal behaviors based on a data-driven modeling. Details are given in the next paragraphs.

#### 3.1. Smart metering

Our objective is to compute a thermal model for a building without any particular prior knowledge on the system. To this end, more extensive and uniform data is required, to provide more general information about the thermal behavior of each building studied. Thus, a smart supervisory, control and data acquisition (2SCADA) system is designed. Figure 3 illustrates its overall architecture. The system is based on wireless sensors (developed by CLEODE<sup>TM</sup> company) that allow measuring the indoor and outdoor temperature ( $T_i, T_o$ ), relative humidity ( $H_i, H_o$ ), solar radiation and diffuse solar radiation in the building ( $R_a, R_{ai}$ ).

Figure 4 shows examples of different trends for the indoor and outdoor temperature (on the top), and the indoor and outdoor humidity (on the bottom). Each sensor is defined with a sampling time equals to  $T_s = 2.4min$ .

Complementary to the smart-metering system, a control system is used to control (ON/OFF) the heating operation for a few scenarios. Table 1 below defines each sensor characteristics used for collecting data. These components are based on wireless system network (WSN) technology which is today the most popular technology used in smart-metering systems. Moreover, many WSN platforms have been developed [7], ZigBee/IEEE 802.15.4 protocols are often part of them. These protocols are a global hardware and software standard designed for WSN requiring high reliability, low cost, low power, scalability, and low data rate [25].

The data are collected, using the smart-metering and the control systems, in order to analyse and predict the thermal behaviors of the building for several non-defined operating modes.

#### 3.2. Estimation of the PWARX model

Here, we justify the choice to model the building thermal behavior as a switching system by its ability to detect changes in thermal dynamics resulting from a given operating mode,



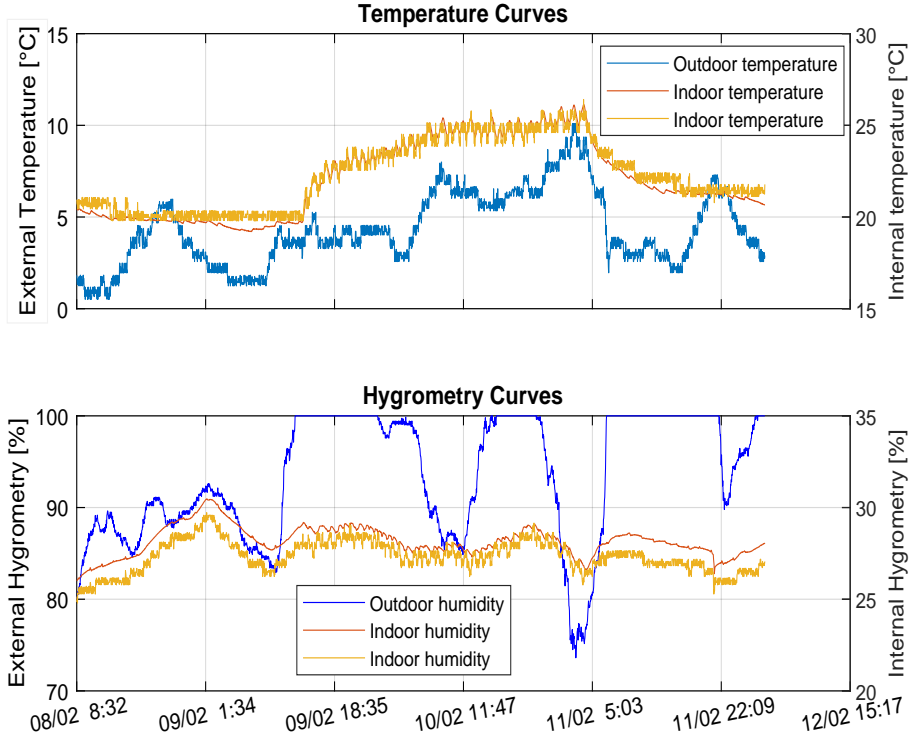


Figure 4: Example of collected data during one week: Temperature and hygrometry curves

Table 1: Wireless sensor characteristics

Measured parameter	Measuring range	Accuracy	Type of sensor	Units
Temperature	$-40$ to $125^{\circ}C$	$\pm 0, 4^{\circ}C$ (max), at $-10$ to $85^{\circ}C$	Thermistor	$[^{\circ}C]$
Heat flux	$0 - 2000W/m^2$	$40\mu V/(W/m^2)$	HFP	$[W/m^2]$
Relative humidity	$0$ to $100\%$	$\pm 3\%$ (max), at $0 - 80\%$	Capacitive polymer	$[\%]$
Solar radiation	$1 - 65536Lux$	$\pm 100Lux$	Light sensor	$[Lux]$ to $[W/m^2]$
Electrical power	up to $3500W$	up to $16A$ on $220V$	Power meter	$[W]$

least square technique while minimizing:

$$V(\theta_i) = \frac{1}{N_i} \sum_{t=1}^{N_i} (y(t) - \theta_i^T \bar{\varphi}(t))^2 \quad (20)$$

where  $N_i$  is the sample number.

*B- Data re-affectation and model estimation.* The aim is to reduce the number of clusters by reclassifying each data point and to estimate the parameters of each sub-model. To do that, we try to minimize the Euclidean distance (Figure 6) between

each pair of data  $(x(i), x(j))$  by:

$$d_i^j = \|x(i) - x(j)\| \quad (21)$$

as well as the error between the measured output and the output of each sub-model is minimal. Moreover, data  $x(i)$  is ranked according to the regression vector and the measured output as the following expression:

$$x(i) = [\varphi(i)^T, y(i)]^T. \quad (22)$$

Thus, the re-affectation of data  $x(i)$  is done relative to their nearest neighbors (c-cnn). Formally, data will migrate towards

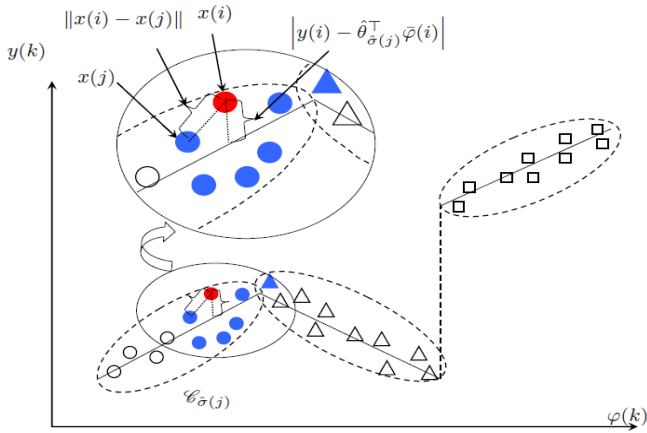


Figure 5: Data re-affectation into three affine sub-classes in the regression space [12].

the most representative clusters according to a specific decision rule:

$$\psi_j^i = \exp(-\alpha_{\hat{\sigma}(j)} \|x(i) - x(j)\|^2 - \beta_{\hat{\sigma}(j)} (y(i) - \hat{\theta}_{\hat{\sigma}(j)}^T \varphi(i)))^2 \quad (23)$$

with  $\hat{\theta}_{\hat{\sigma}(j)}^T$  is the parameter vector corresponding to the class  $C_{\hat{\sigma}(j)}$ .  $\alpha_{\hat{\sigma}(j)}$  and  $\beta_{\hat{\sigma}(j)}$  are positive parameters. Thereafter, we need to update the class number  $C = \{C_1, \dots, C_{\bar{q}}\}$  where  $\bar{q} \leq \Lambda$ . And, to decide on the re-affectation of the data  $x(i)$  to the remaining class  $\bar{q}$ , we use the membership criterion of  $x(i)$  to the class  $C_p$ , defined by:

$$P(x(i) \in C_p) = \frac{\sum_{j|x(j) \in \{\Gamma_c(x(i) \cap C_{\bar{q}})\}} \psi_j^i}{\sum_{j=1}^c \psi_j^i}, \quad p \in \{1, \dots, \bar{q}\} \quad (24)$$

where  $\bar{q}$  determine the persistent number of class in order to have the best fitting between the measured and estimated output. Let notice that  $P(x(i) \in C_p) = 1$  if the elements of  $\Gamma(x(i)) \in C_p$  and  $P(x(i) \in C_p) = 0$  if the element of  $\Gamma(x(i)) \notin C_p$  [12].

*C-Convergence criterion.* The convergence is done by comparing the earlier parameter vectors  $\theta^{(r)} = [\hat{\theta}_1^{(r)}, \dots, \hat{\theta}_{\bar{q}}^{(r)}]$  and the posterior parameter vectors  $\theta^{(r+1)}$  following:

$$\|\theta^{(r+1)} - \theta^{(r)}\| \leq \nu \quad (25)$$

where  $r$  is the index of the iteration and  $\nu$  is an arbitrary threshold defined by the user. Generally, one use  $\nu = 10^{-5}$ .

*D- Sensitivity analysis.* A study of the influence of each input onto the model structure is presented in this article by comparing the FIT criterion value for different configurations. The FIT represents the similarity between the measured output  $y$  and the output  $\hat{y}$  predicted by the model. It is given by the following equation:

$$FIT = (1 - \frac{\|\hat{y} - y\|}{\|y - \bar{y}\|}) \times 100\% \quad (26)$$

where  $\bar{y}$  and  $\hat{y}$  are respectively the mean and the estimate of the measures  $y(t)$ . Once a model with a satisfactory FIT has been estimated, it is possible to evaluate the influence of each parameter on the model and on the prediction accuracy. To this end, we analyze the sensitivity of each parameter by identifying and characterizing the: (i) input uncertainties, (ii) model orders, and (iii) class number.

Brief, this analysis allows us to appreciate the contribution of each input variable onto the FIT. This contribution is computed by using the following influence index defined by:

$$W_{ui} = \frac{\hat{\rho}_i}{\sum_{i=1}^m \hat{\rho}_i} \times 100\%, \quad (27)$$

where  $\hat{\rho}_i$  is the contribution estimated of each input obtained from the computation of  $F\hat{T}_k$ . This last is the estimation of the FIT subject to the influence of inputs considered for the  $k^{th}$  model structure.

It can be expressed by:

$$F\hat{T}_k = \sum_{i=1}^{N_i} \sum_{t=1}^m \xi_i(k) \hat{\rho}_i u_i(t), \quad \text{for each model structure } k \quad (28)$$

with

$$\xi_i(k) = \begin{cases} 1 & \text{if the } i^{th} \text{ input is considered for the model} \\ & \text{structure } k, \\ 0 & \text{otherwise.} \end{cases} \quad (29)$$

Concerning the other terms,  $u_i(t)$  represents each input variable used for the identification of PWARX model. Model structure  $k$  is then the configuration considered, in terms of input number, to estimate thermal model parameters. For instance scenario where only two inputs between  $m$  are considered.

Given the influence index  $W_{ui}$  for each input signal, we can realize an immediate analysis of how much the uncertainty on the measurements can affect the indoor air temperature prediction. Thus, this procedure should allow to properly structure the model in terms of input data in order to increase the accuracy of the model and reduce the calculation time.

## 4. Experimental results and discussion

### 4.1. Case studies

To show the effectiveness of our methodology, experiments on the Lavoisier student residential building, located in Douai, in the north of France (Figure 6) have been made. The total area of the building is approximately  $3500m^2$  and it is subdivided into 2 sub-building parts. The first part is composed by 4 floors and the second one by 5 floors. Each floor has respectively for each sub-building 10 and 32 rooms of  $11m^2$  with a double glazed window size ( $135cm \times 110cm$ ). Notice that for this building, the envelop is composed by brick walls with a thickness  $22cm$  and a glass wool insulation with a thickness  $14cm$ .

So, we launched experiments on the second sub-building during this study. Herein, half of the rooms are located in



Figure 6: (a) Lavoisier student Residence in Douai (b) Lavoisier student Residence 3D model

the east side and the other half are located in the west side. To recover a persistent database for identifying the underlying causal thermal model of the building, several experiment scenarios have been taken into account: two different orientations "East-West", heating sequence and different floor levels (1<sup>st</sup> and 3<sup>rd</sup> floors). In addition, rooms operate on different thermal excitations driven by a heating system supply and are influenced by natural solar radiation. Precisely, the heating supply will be controlled on two different sequences such as a controlled sequence and a random heating sequence.

Also, the same experimental protocol was renewed for Condorcet student residence located in Douai (Figure 7). The difference of two buildings is mainly the envelop. Indeed, for Condorcet we have a block wall with a thickness 20cm combined with polystyrene insulation with thickness 10cm. These two buildings are considered with the aim of validating the model for its ability to reproduce the building thermal behaviors faithfully only using a data collection.

So, for each scenario we evaluate the thermal environment thanks to different sensors and actuators placed in each room as it is illustrated onto Figure 8 above (example for west oriented room). Once again, readers can refer to Table 1 for sensor and actuator characteristics. Moreover, the location of the temperature sensors makes it possible to measure the indoor air temperature of the irradiated zone ( $T_{i1}$ ) and the shaded zone ( $T_{i2}$ ) for the two orientations (i.e. East-West), that we add the solar radiation sensor (noted  $Lux_{ext}$  and  $Lux_{int}$  on the figure), the heating power consumption measure (noted  $Ch_1$  and  $Ch_2$  on the figure), the outdoor humidity and temperature ( $H_e$  and  $T_e$ ).

Besides, during the measurement campaigns each collected data associated with each scenario is saved in a database in CSV format. This is possible thanks to an in-house developed application based on LABVIEW<sup>TM</sup> and NetBeans<sup>TM</sup> software's. The historical data is used after that to identify the PWARX-thermal model according to the experimental planning detailed after.

#### 4.2. Experimental design

During this experience we would firstly like to compare two rooms of the Lavoisier residence according to its orientation

(i.e. east and west). This stage aims to find a model that validates the indoor thermal behavior, for these two orientations under the same conditions.

##### 4.2.1. East side

The experimental study was made from February 07<sup>th</sup> - 12<sup>th</sup> February, 2018. The evolution of the outdoor temperature varies between 0 and 10°C for an outdoor humidity ranging between 70 and 100% as indicated respectively in Figure 9a and 9b. Moreover, Figure 10a and 10b show the evolution of indoor air temperature (black solid line) for two different sensor positions ( $T_{i1}$  and  $T_{i2}$ ) and the heating power demand (black dot line), during the period defined before. The temperature varies between 19 and 25°C and the electrical power consumed between 0 and 1400W.

In the other hand, the solar radiation received on the east wall and the radiation incident on the room is shown in figure 11a and 11b. We can see that solar radiation received by the wall reaches a maximum value of 11.5W/m<sup>2</sup>, while the incident solar radiation on the floor is approximately 0.45W/m<sup>2</sup>. And finally, two scenarios are considered in this study: the first one is the free evolution scenario (i.e. without heating supply), and the second one consists to apply a random activation of the heating source. This is renewed also for the west side.

##### 4.2.2. West side

For the same period, i.e. from 07-02-2018 to 12-02-2018, the solar radiation received on the west wall reaches a maximum value of 95W/m<sup>2</sup> while the solar radiation received on the floor takes a maximum value of 42.5W/m<sup>2</sup> (see Figure 12). Moreover, Figure 13 below shows that the indoor air temperature of the shaded area varies between 19 and 26°C, however for the irradiated area, the indoor temperature varies between 19 and 31°C, which generates a difference of 7°C between the two positions.

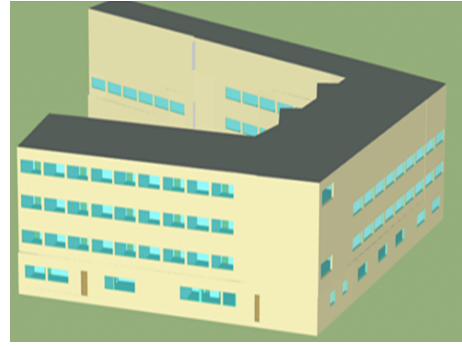
##### 4.2.3. Analysis for different heating sequences

In this part, experiments are made from 13/02 to 17/02, with the following conditions: (i) two west oriented rooms respectively at the 1<sup>st</sup> and 3<sup>rd</sup> floor are considered, (ii) weather conditions are the same in terms of outside temperature, outdoor





(a)



(b)

Figure 7: (a) Condorcet student Residence in Douai. (b) Condorcet student Residence 3D model

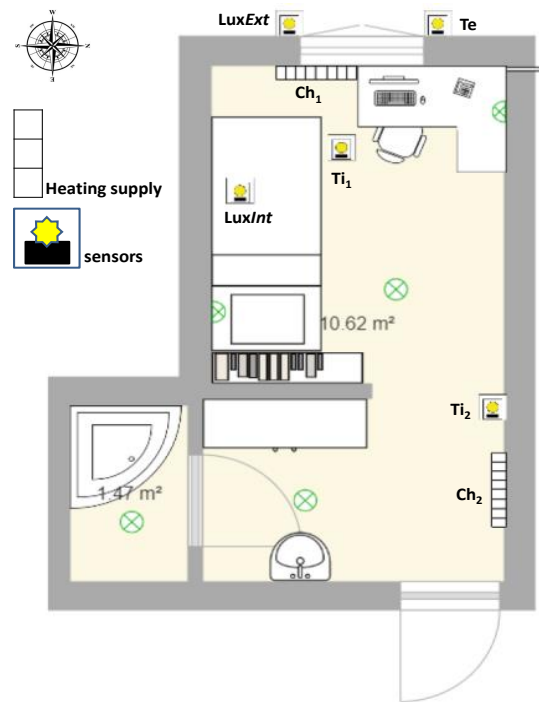
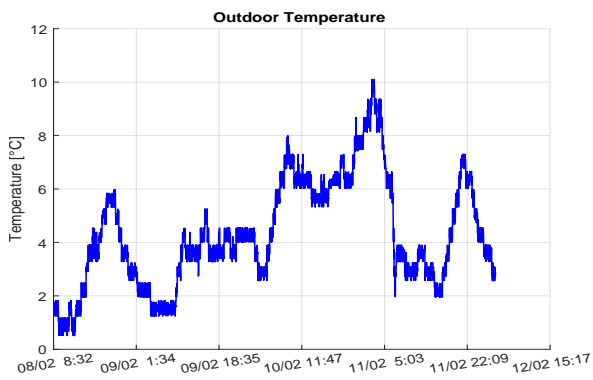
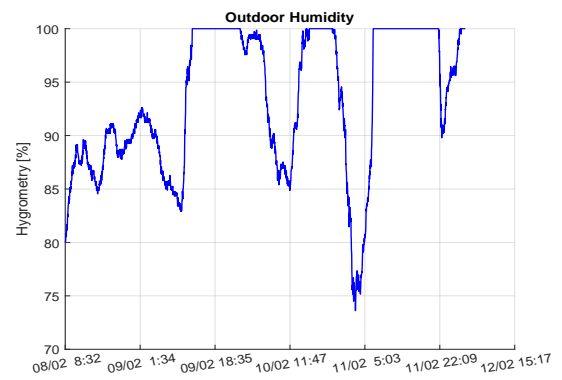


Figure 8: Location of sensors and actuators in rooms



(a)



(b)

Figure 9: (a) Outdoor temperature and (b) Outdoor humidity curves

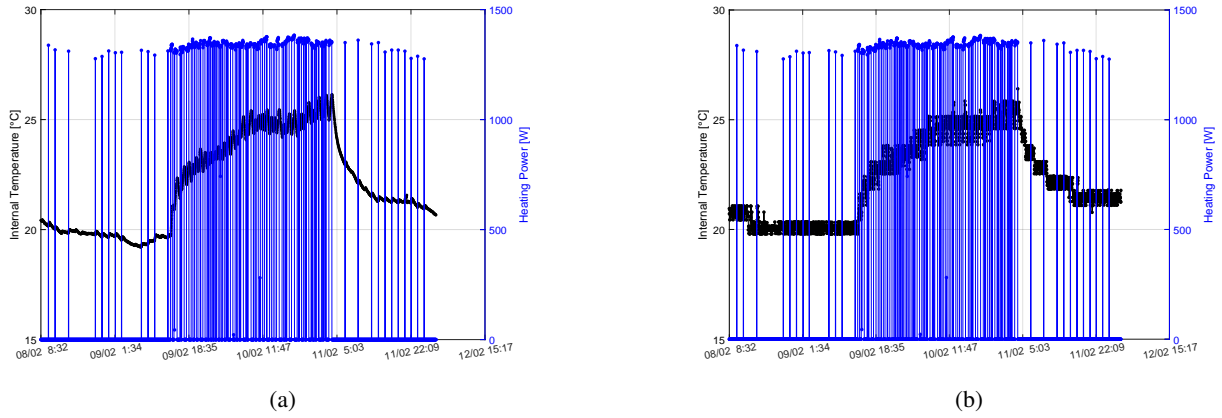


Figure 10: Heating power and indoor air temperature (a): position 1 and (b): position 2

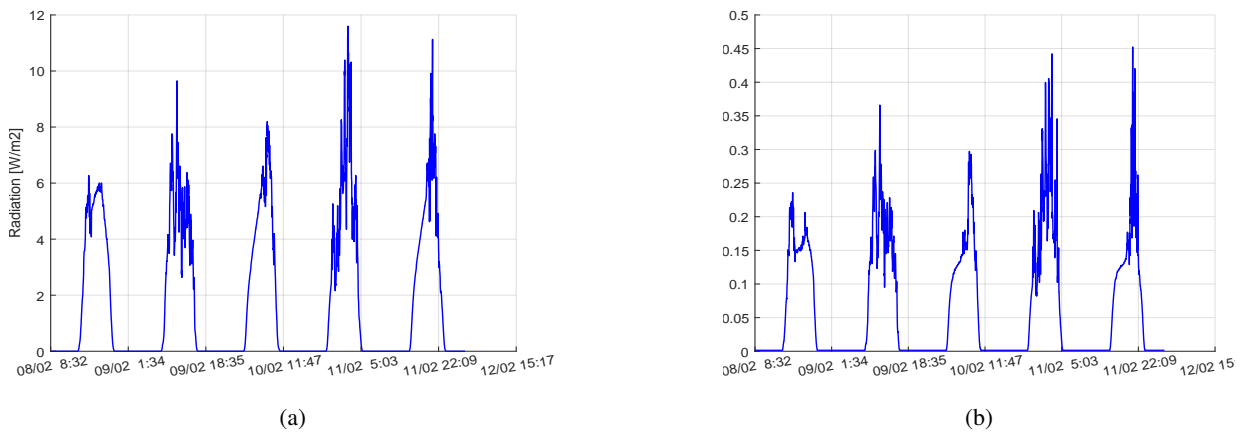


Figure 11: (a) Outdoor solar radiation. (b) Solar radiation incident on the floor

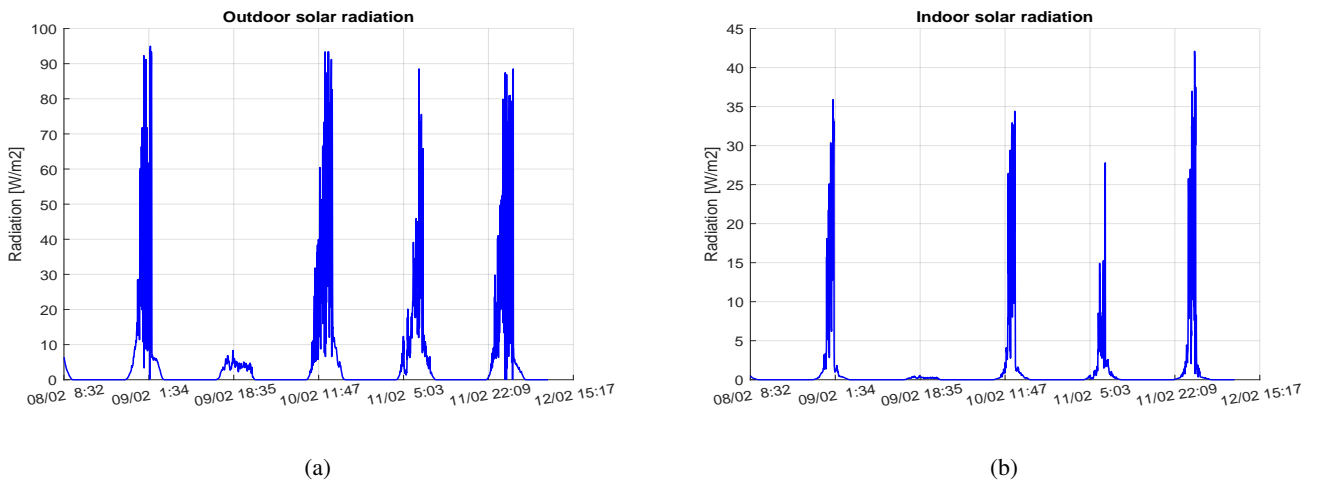
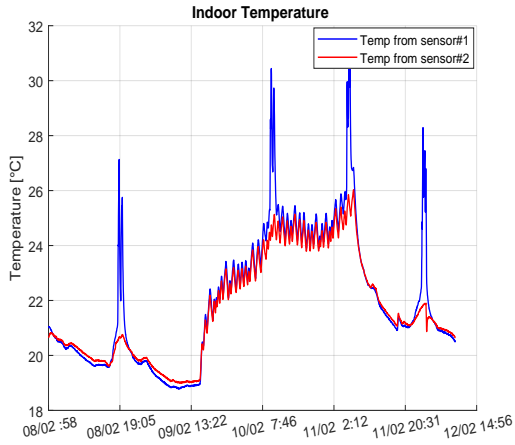


Figure 12: (a) Solar radiation. (b) Solar radiation on the floor

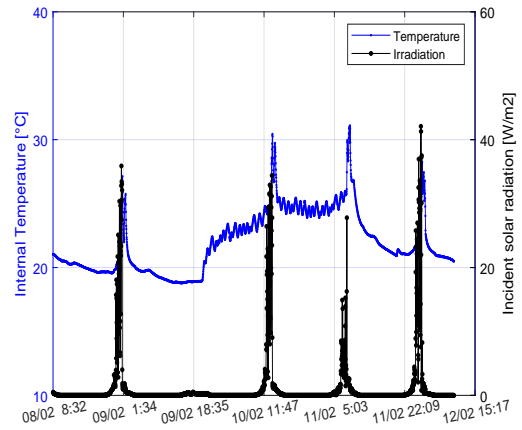
humidity and solar radiation, (iii) two distinct modes of heating system functioning are launched.

As example, we can observe in Figure 14, the west side receives a solar radiation power which rises to  $90\text{W}/\text{m}^2$ . On the other hand, to meet the third condition cited before, the 3<sup>rd</sup>

floor heating system is controlled in a random manner, which causes a random variation of the indoor thermal behavior (Figure 16b). While, the room on the 1<sup>st</sup> floor evolves with two different scenarios: free evolution (i.e. heating system OFF) and heating system activated (Figure 15b).



(a)



(b)

Figure 13: (a) Indoor air temperature ( $T_{i1}$  and  $T_{i2}$ ). (b) Solar radiation effect on indoor air temperature.

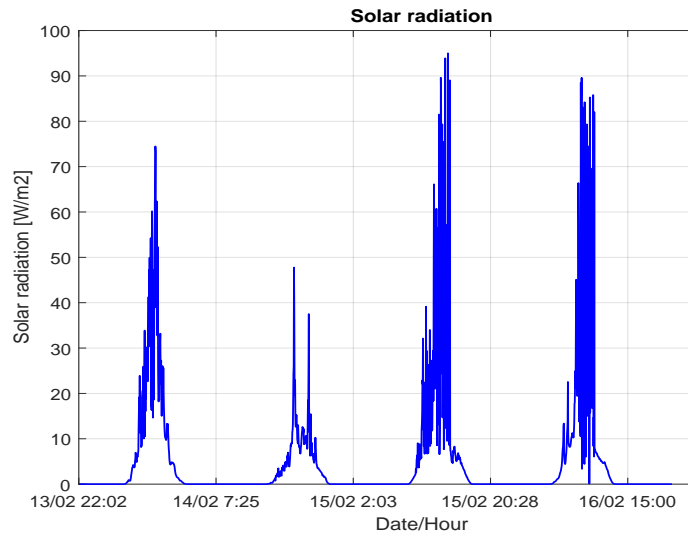
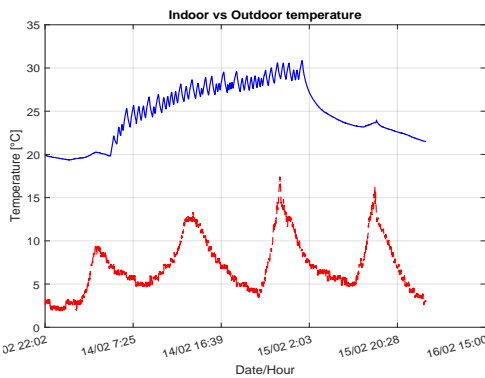
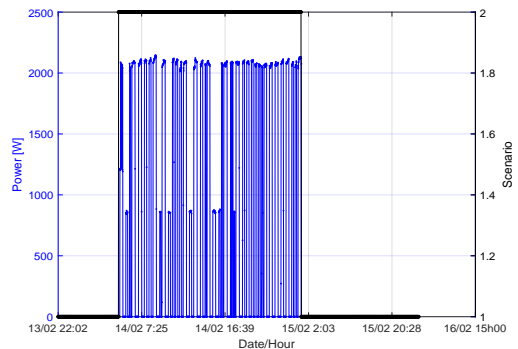


Figure 14: Solar radiation.



(a)



(b)

Figure 15: (a) Indoor air temperature for the 1<sup>st</sup> floor. (b) Heating scenario for the 1<sup>st</sup> floor.

Figure 15a shows that the indoor temperature for the first stage varies between 20 and 31°C for an outside temperature

ranging between 2 and 10°C. On the other hand, the interior temperature for the third stage varies between 19 and 26°C as

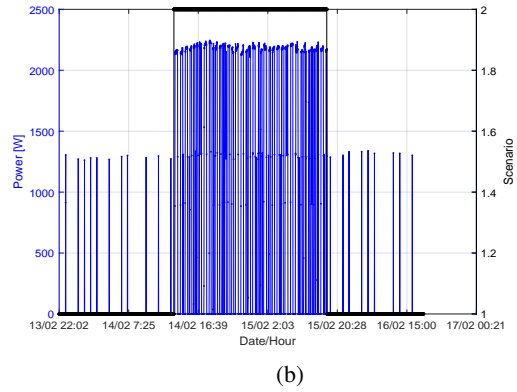
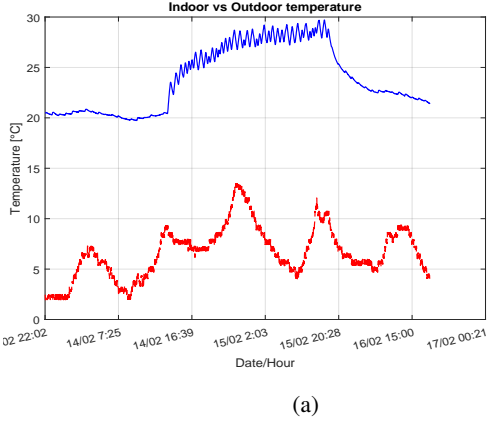


Figure 16: (a) Indoor air temperature for the 3<sup>rd</sup> floor. (b) Heating scenario for the 3<sup>rd</sup> floor.

shown in Figure 16a.

#### 4.3. PWARX-thermal model validation

The objective of this part is to propose a model validation process in order to present the precision of estimated parameters and to evaluate how the choice of hyper-parameters (that define the model structure) could influence the results. For example, in [27], authors have taken only the outside temperature as inputs using an order number equals to  $n_a = n_b = 4$ . In contrast, authors of [33] used the outdoor temperature, outdoor humidity, solar radiation and wind speed as inputs for the same order number  $n_a = n_b = 4$ . However, the choice of initial model structure hyper-parameters was not justified in most of related works. Approach for analyzing and validating the structure of the model will be presented in the next paragraphs.

##### 4.3.1. Sensitivity analysis

Sensitivity analysis is used to determine which is the best structure for the model to faithfully reproduce the thermal behavior on a variety of measurement campaigns. Thus, the structural analysis serves to regulate the algorithm by making an analysis on the hyper-parameters such as  $n_a, n_b$ , class number ( $C$ ) and the input choices.

The determination of the inputs and their influence on the structure of the model are first presented in this section. To this end, we estimate a model derived by following configurations:

- $n_a = 2, n_b = 2, n_k = 0, C = 200, \beta_0 = 0.5, \gamma_0 = 0.5$ .
- Inputs measured: Outdoor temperature ( $T_o$ ), outdoor humidity ( $H_o$ ), solar radiation ( $Ra$ ), heating power ( $Pw$ ).
- Output measured: Indoor Temperature  $T_i$ .

By varying the system order  $n_a$  and  $n_b$  in Table 2, we can see several model structures (1 to 4) and the FIT values obtained with each of them. We observe that the best estimation is reached when  $n_a$  and  $n_b$  correspond to an order of 5 with a FIT equals to 78.48%. This is achieved with  $T_o, H_o, Ra$  and  $Pw$  as inputs. The influence of each input on the accuracy of

Model	Model structure			Inputs selected	FIT
	$n_a$	$n_b$	$C$		
1	2	2	200	$T_o, H_o, Ra, Pw$	71.41%
2	3	3	200	$T_o, H_o, Ra, Pw$	64.70%
3	4	4	200	$T_o, H_o, Ra, Pw$	51.60%
4	5	5	200	$T_o, H_o, Ra, Pw$	78.48%
4 <sub>A</sub>	5	5	200	$T_o, H_o,$	15.27%
4 <sub>B</sub>	5	5	200	$Ra, Pw,$	55.68%
4 <sub>C</sub>	5	5	200	$T_o, H_o, Ra$	24.61%
4 <sub>D</sub>	5	5	200	$Ra, H_o, Pw$	70.38%

Table 2: Best fitting per model structure: (Top) according to system order, (Bottom) according to selected inputs for the 4<sup>th</sup> model

this model (resp. accuracy of the parameters) has been also investigated. Results in the bottom part of Table 2 show the impact of inputs on the FIT's value. We can note that the FIT decreases or increases according to the inputs selected. The worst model corresponds to a structure composed only of  $T_o$  and  $H_o$  (FIT=15.27%). By adding the solar radiation to this structure, the FIT increases and becomes 24.61%. The model structure with  $Ra$  and  $Pw$  as inputs gives a better FIT compared to the model structure with  $T_o$  and  $H_o$  with a difference of 40.41%. By adding the outdoor humidity to this structure, the FIT increases and becomes 70.38%.

To summarize, Figure 17 represents the effect of the system orders  $n_a$  and  $n_b$  (varying from 2 to 5) and c-nearest-neighbors  $c$  (varying from 100 to 300). The results show that, with all inputs, the best model corresponds to  $n_a = n_b = 5$  and  $c = 200$  neighbors.

On the other hand, Figure 18 and 19 show a comparison between the model structures "4<sub>D</sub>, 4<sub>A</sub> and 4<sub>C</sub>" as defined in table 2 (indoor estimated temperature by red solid line, indoor measured temperature by black solid line). Results indicate that solar radiation and heating power are essential in order to estimate models with accuracy and precision. The absence of the heating power as input (Model 4<sub>C</sub>) reduces the FIT by 53.87% compared to model 4 (Figure 18b), while the absence

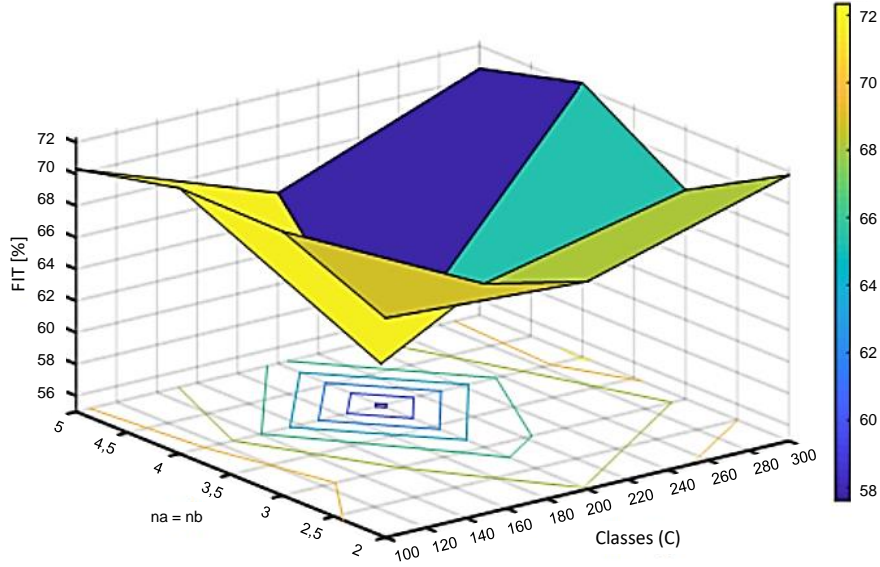


Figure 17: Effects of c-nearest-neighbors ( $c$ ) and the system order ( $n_a, n_b$ ) on the FIT.

of the solar radiation and heating power (Model 4<sub>A</sub>) reduces the FIT by 63.21% compared to model 4 (Figure 18a). We have also noted that the absence of the outdoor temperature (Model 4<sub>D</sub>) reduces the FIT by 8.1% compared to model 4 as shown in Figure 18b.

Figure 20 shows the influence index of each input variable on the model structure. It is mainly based on the sensitivity analysis conducted before and computed in order to classify the set of inputs (see equation (27)). The results show that the heating power ( $P_w$ ) is the most influential input with a relative influence index of 70%. This is due particularly to the impact of the heating power on the indoor thermal behavior. For the next, we remark that solar radiation ( $R_a$ ) affects with an influence index of 20%. Let note that the thermal response time of solar radiation is greater than of heating power. Then, the indoor thermal behavior is influenced by the diffuse solar radiation which depends on the building orientation, glazing type and weather conditions. At the end, outside humidity ( $H_o$ ) and temperature ( $T_o$ ) are less influential, with computed indexes of 8% and 4% respectively.

To conclude this analysis, modeling in efficient way the thermal behavior of the building requires at least the heating power and the solar radiation as inputs of the model. The other terms can be used simply to add further information to the model.

#### 4.3.2. PWARX and ARX derived models

This section presents the validation of the PWARX-thermal model by comparing it with ARX and its derived models. We choose the PWARX structure model equivalent to the model type number 4 (see Table 2). Moreover, for the building investigated, the algorithm estimates 3 operating modes describing the thermal behavior of the building. Figure 21 shows the estimated switching evolution of these operating modes

(black dashed line) based on data collected from the east-facing room.

Precisely, three operating modes (i.e. discrete state) were detected with respect solar radiation ( $R_a$ ) and the use of the heating system ( $P_w$ ); which is simply the logical consequence of results obtained during the sensitivity analysis above. For instance, the first sub-model  $\langle SM1 \rangle$  (resp. first operating mode) corresponds to the activation of the heating system (ON). The second sub-model  $\langle SM2 \rangle$  (resp. second operating mode) defines the situation where heating system is (OFF) and solar radiation is greater than  $2.19 W/m^2$ , while the third sub-model  $\langle SM3 \rangle$  (resp. third operating mode) corresponds to the situation where heating system is (OFF) and solar radiation less than  $2.19 W/m^2$ . This is summarized in table 3 below.

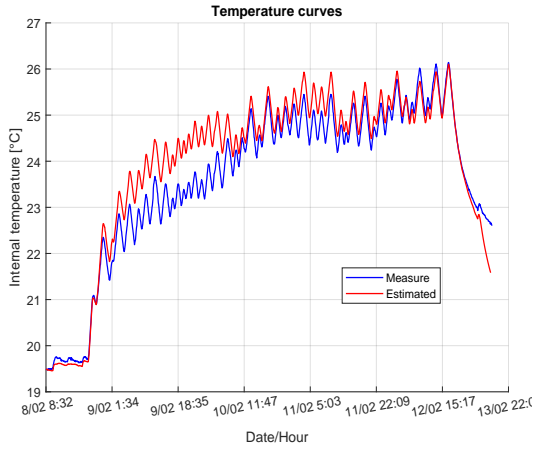
Modes	$P_w$	$R_a$
Mode 1	ON	- $W/m^2$
Mode 2	OFF	$\geq 2.9 W/m^2$
Mode 3	OFF	$\leq 2.9 W/m^2$

Table 3: Switching conditions

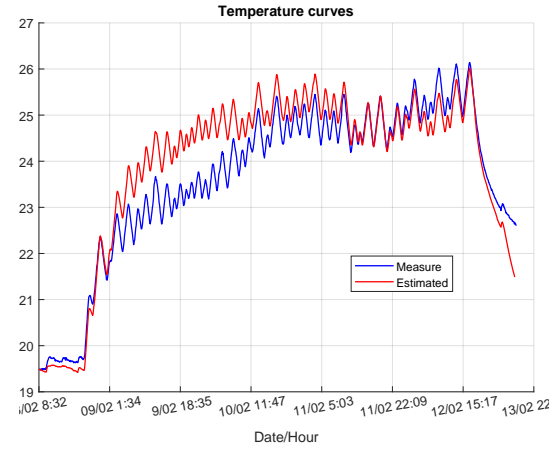
Furthermore, we can see on Figure 21, the validation of these switching conditions by comparing the estimated operating modes (black dashed line) with the predicted (red dot line) by the rules defined in table 3. And finally, each vector of parameters associated with each operating mode is given in Table 4, as numerical application example for the PWARX-thermal model identified. Basically, for the  $q^{th}$  mode and with  $m$  input variables, we have:

$$T_i(t) = \sum_{j=1}^{n_a} a_{j,q} T_i(t-j) + \sum_{v=1}^m \sum_{j=1}^{n_b} b_{j,q} u_v(t-j). \quad (30)$$

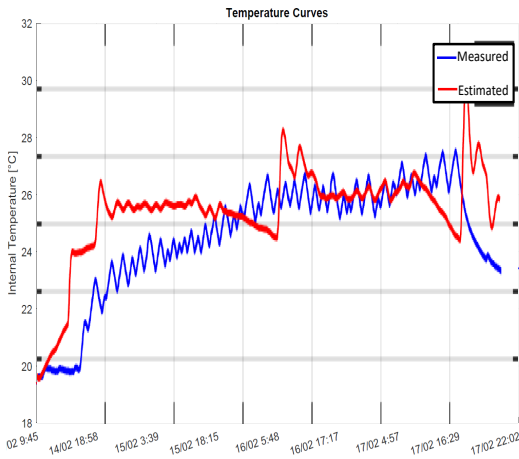




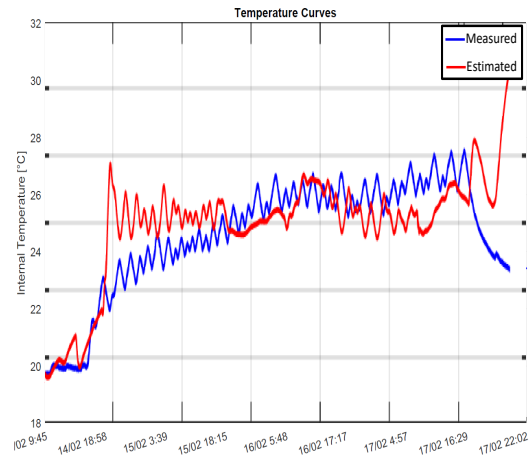
(a)



(b)

Figure 18: Training data: (a) Model Structure "4". (b) Model structure "4<sub>D</sub>" without outside temperature

(a)



(b)

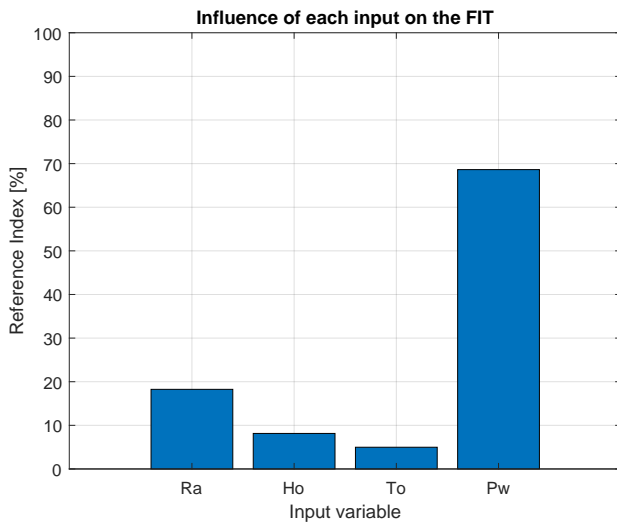
Figure 19: Training data: (a) Model Structure "4<sub>A</sub>" without solar radiation and heating power. (b) Model structure "4<sub>C</sub>" without heating power

Figure 20: Influence index

Parameters	<i>S</i> M1 sub-model 1	<i>S</i> M2 sub-model 2	<i>S</i> M3 sub-model 3
$a_1$	1,1562	1,2750	0,8826
$a_2$	-0,1606	-0,2917	0,1128
$b_{11}$	-0,0031	0,0001	0,0040
$b_{12}$	0,0031	0,0014	-3,9678e-05
$b_{21}$	-0,0017	0,0016	-0,0261
$b_{22}$	0,0019	-0,0016	0,0189
$b_{31}$	2,5085e-06	5,8055e-05	0,0326
$b_{32}$	-8,9926e-07	-5,4697e-05	-0,0201
$b_{41}$	3,3243e-05	3,4460e-05	3,3891e-05
$b_{42}$	4,1579e-05	3,2496e-05	4,3921e-05
$e$	0,0364	0,3707	0,7884

Table 4: Each behavioral model parameters.

As an example, let us considerate the following numerical rep-

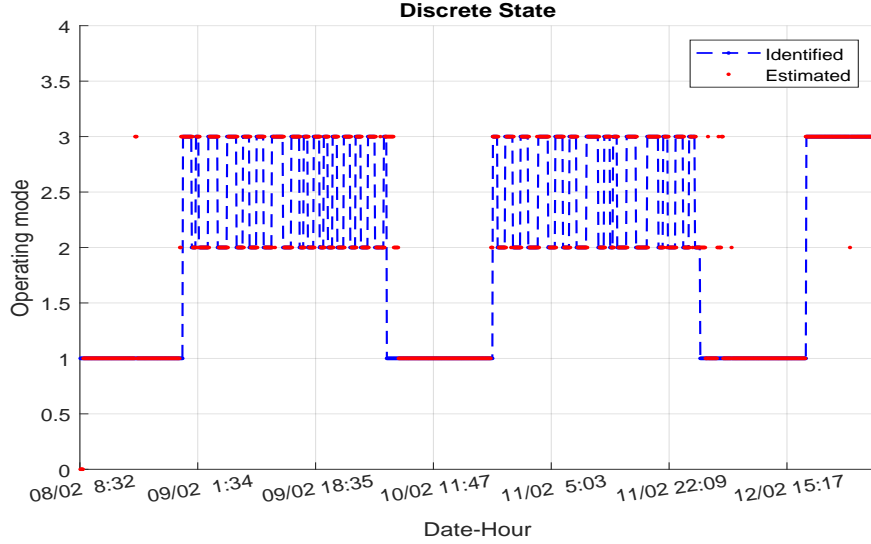


Figure 21: Operating mode validation.

resentation for mode  $q = 1$  (resp.  $SM1$ ):

$$\begin{aligned}
 T_i(t) = & -1.1562T_i(t-1) + 0.1606T_i(t-2) - 0.0031T_o(t-1) \\
 & + 0.0031T_o(t-2) - 0.0017H_o(t-1) + 0.0019H_o(t-2) \\
 & + 2.508e-06Ra(t-1) - 8.99e-07Ra(t-2) \\
 & + 3.32e-05Pw(t-1) + 4.15e-05Pw(t-2) + 0.03
 \end{aligned} \tag{31}$$

Moreover, regarding Figure 22, one can see the comparison of performance of the PWARX-thermal model with other ARX-derived models. The two bottom sub-figures in Figure 22 illustrate the estimation of the indoor temperature with ARX model (green solid line) and indexed-ARX (black solid line) one, compared with the measurements (black solid line). We remark here, on the one hand, that the building thermal behavior is not well recovered by the ARX model, and on the other hand, the operating points defined for the indexed-ARX model described in [27] are not really suitable to explain building configurations.

So, by confronting with PWARX model, the NARX model can give good results (Figure 22 — top and left part) in terms of indoor temperature estimation (pink solid line), compared with the measurements (black solid line). However it remains a challenging task to interpret without proper knowledge of the internal structure of the model. In other terms, it is difficult for building experts for instance to understand physical phenomena, especially when various behaviors or uses have to be explained. As we can see, PWARX-thermal model makes possible the understanding of the system behavior in a much more intuitive and easy way than other models.

Finally, Table 5 gives a numerical quantification of the performance of each model based on the FIT criteria. The results confirm that the PWARX model structure is the best for prediction purposes, being able to estimate accurately the real temperature according to different scenarios or configurations, influencing the thermal behavior of the building.

Model	ARX	Indexed ARX	NARX	PWARX
FIT	21.41%	34.93%	73.55%	78.48%

Table 5: Performance comparison between PWARX model and ARX-derived models

#### 4.3.3. Impact of the orientations and floor levels

The parameter's estimation of the PWARX model was made with the measurement campaign from February 07<sup>th</sup> - 12<sup>th</sup> February, 2018, for the east orientation and using training data as presented in section 4.1. In this section, we want to validate our model with other scenarios, firstly by considering data collected from west oriented room for the period February 14<sup>th</sup> - 17<sup>th</sup> February, 2018. Figure 23 shows the indoor temperature estimation results (red solid line) with a  $FIT = 60.02\%$ . Thus, we observe with satisfaction that the structure of the estimated model is adapted to east and west orientations.

In the next, the same PWARX-thermal model was tested with data collected from the room on the 3<sup>rd</sup> floor and located in the west-facing, from February 14<sup>th</sup> to 17<sup>th</sup>, 2018. The thermal behavior of this room is mostly different compared to the reference room one due to the orientation, floor level, heating scenarios and weather conditions. Once again the result in Figure 24 shows there is good agreement between the experimental measure (black solid line) and the prediction (red solid line) with a  $FIT = 77.66\%$ .

#### 4.3.4. PWARX-thermal building methodology tested on another building

Authors of [18] mentioned in their study, the same sets of ARX model inputs may not apply directly to a different building type and location due to measurement properties. From this context, hereafter we verified if the reference PWARX model structure estimated for Lavoisier building can reproduced the indoor thermal behavior of two east-west oriented rooms, parts of Condorcet building. The purpose of this step

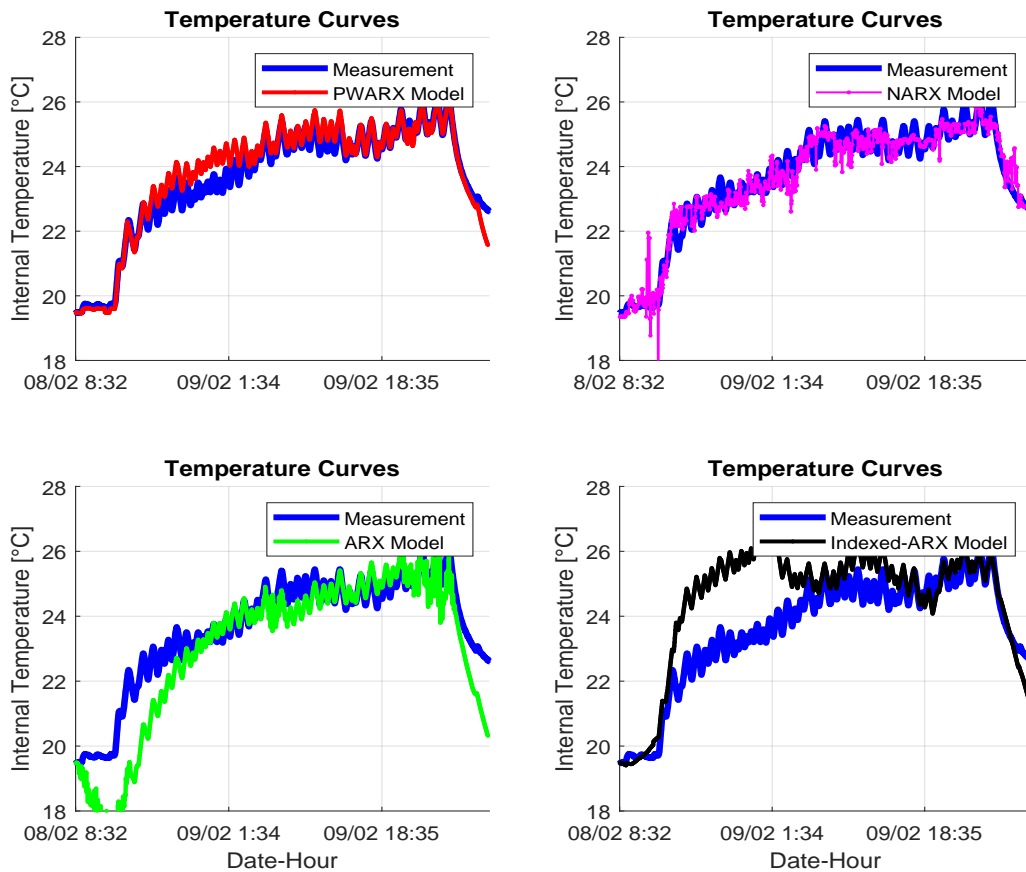


Figure 22: Indoor air temperature validation: Comparison between PWARX, NARX, ARX and Indexed-ARX models.

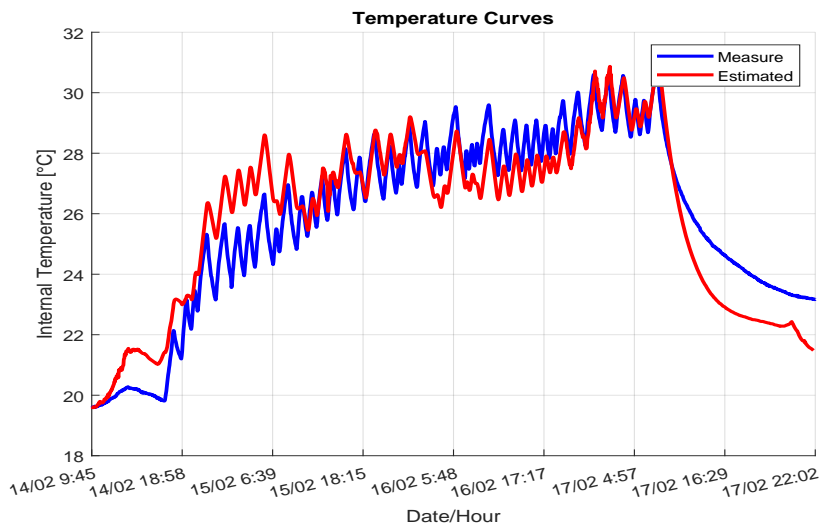


Figure 23: Validation of Temperature (West Side),  $FIT = 60.02\%$  (14-02 / 17-02).

is to show that the reference model structure may also be suitable for any building's archetype. Besides, it is important to note that the thermal behavior conditions are not also similar due to different heating scenarios and weather conditions.

So, Figure 25 below gives us any ideas on indoor temperature dynamics. Indeed, herein the indoor thermal behavior,

respectively for rooms in east and west side, can be estimated by the PWARX model with a FIT of 57.40% and 77.66% for both rooms. These results are acceptable and in line with configurations given before.

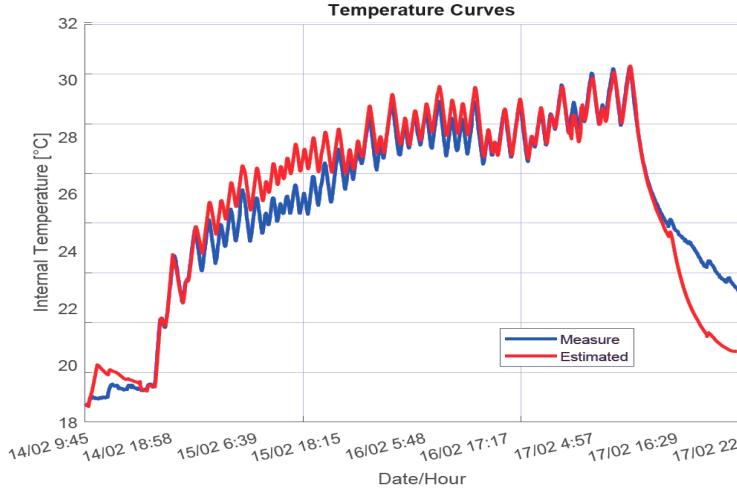
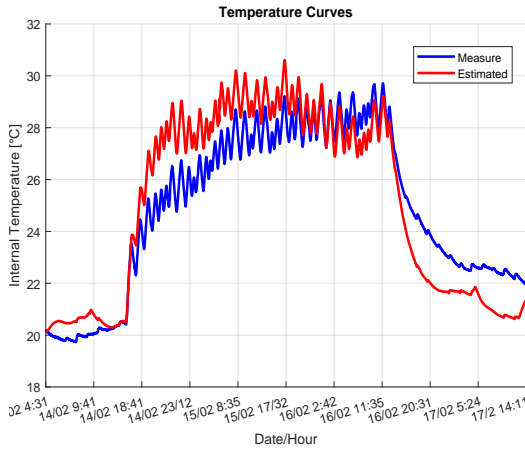
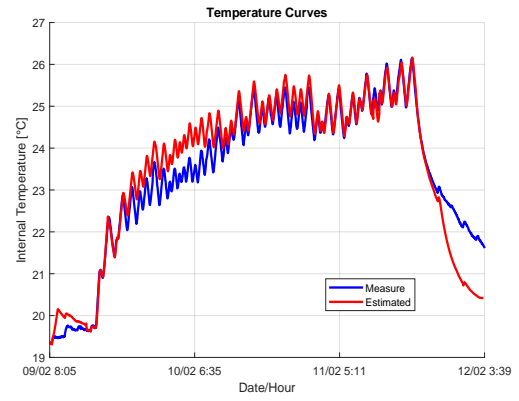


Figure 24: Validation of Temperature (West Side, 3<sup>rd</sup> Floor),  $FIT = 77.66\%$  (14-02 / 17-02).



(a)



(b)

Figure 25: (a) Validation of Temperature (East Side),  $FIT = 57.40\%$  (14-02/17-02). (b) Validation of Temperature (West Side),  $FIT = 77.66\%$  (09-02/12-02)

#### 4.3.5. Comparison between models and evaluation of thermal behavior similarities

Figure 26 below shows the comparison procedure that could be implemented in order to compare the thermal behaviors of different rooms or buildings, using our PWARX methodology, in general manner. In fact, this procedure consists in testing a reference PWARX-thermal model on other measurements derived for other rooms to be compared. In this example, five rooms belonging to both the Condorcet and Lavoisier residences are used, one as a referenced model (Room 1) and 4 others to be compared. The objective is to analyse the ability of the reference model to recover another thermal behavior identified, based on a particular case study (examples: heating mode, orientation, weather conditions), but also to detect possible true different thermal dynamics. Briefly, it will allow us to verify the similarity of the thermal behavior of each room with respect to the reference room from Lavoisier building (noted here Room 1). The FIT's difference between the Room 1 and the other rooms will be considered as a similarity

rate ( $SR$ ) and makes it possible to quantify the thermal behavior discrepancy. This last can be computed by the following expression:

$$SR = 100 - (FIT_{room_1} - FIT_{room_i}), \text{ with } i = 2, \dots, 5. \quad (32)$$

where  $FIT_{room_i}$  is computed using the measurements collected on room  $i$  and the PWARX model estimated on room 1. The FIT results in Table 6 show that Rooms 1 and 5 probably have the same behavior (similarity rate equal to 99.18% as shown in Figure 27). Of course, both rooms are located in two different residences (Condorcet, Lavoisier) but this can be explained by the fact they are in the same orientation and level floor, and have also the same heating equipment. Furthermore, the comparison between Room 1 and Room 2 shows the floor level difference impacts on the thermal behavior of the rooms. For this case, the similarity of the thermal behavior of both is around 81.54%, as illustrated in Figure 27.

Now, if we consider the thermal behavior of the Rooms 1 and 4, it is not really similar related to the orientation. Indeed,

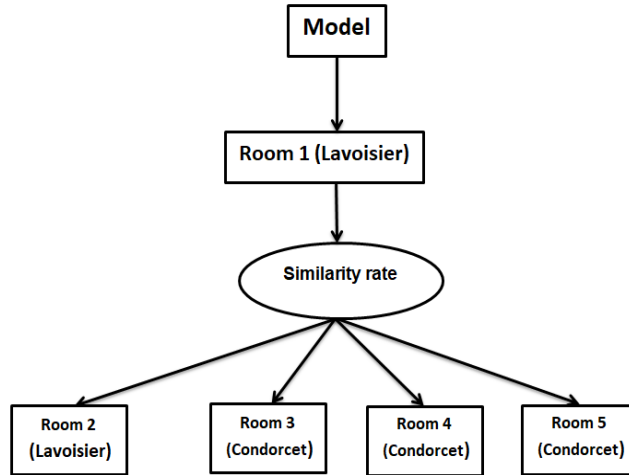


Figure 26: Thermal behavior comparison procedure.

Test on data collected from:	Scenarios			FIT
	Orientation	Heating mode	Floor	
Room 1 (Lavoisier)	East	Random	3 <sup>rd</sup>	78.48%
Room 2 (Lavoisier)	West	Random	1 <sup>st</sup>	60.02%
Room 3 (Condorcet)	East	Controlled	1 <sup>st</sup>	49.15%
Room 4 (Condorcet)	West	Random	3 <sup>rd</sup>	57.40%
Room 5 (Condorcet)	East	Random	3 <sup>rd</sup>	77.66%

Table 6: Cross-validation results

the data collected from Room 4 have lower solar energy values than those from Room 1. As a result, the similarity rate decreases significantly and reaches 78.92% (see Figure 27). Finally, Room 3 can be considered as the worst case for this validation. In fact, as the heating operating equipment and floor level are different from the initial scenario, the similarity rate obtained is around 70.67% compared to the reference Room 1. Overall proportionality of the similarity rate is illustrated in Figure 27.

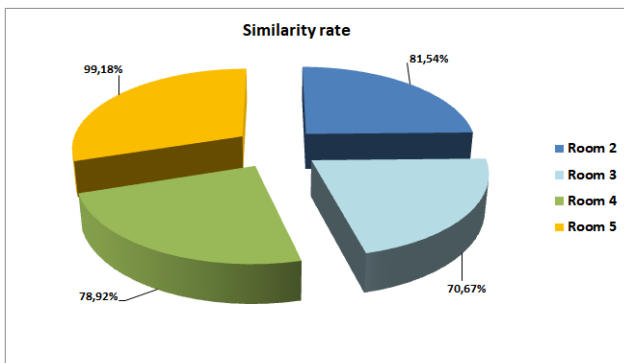


Figure 27: Similarity rate.

In summary, experimental data allows to interpret easily the thermal behavior of different rooms (resp. buildings). In other terms, it is possible to compare each room (resp. building) to

another reference one by deducing in particular the orientation and/or the floor level are/is different or similar of the reference for instance. On the other hand, we can use the similarity rate value to state if a reference model is suitable to also explain the thermal behavior of any building's archetype, even if the configurations are not identically exact due to different heating scenarios and/or weather conditions.

## 5. Conclusions

In this paper, we present a novel methodology for modeling the thermal behavior of residential buildings by using the PWARX model. The identification procedure is detailed to obtain the model and its parameters. Also, several analyses have been conducted to have a suitable structure for the model, as well as to validate it. The first of them is the sensitivity analysis which is necessary to determine what inputs will be required in order to guarantee the accuracy of the model and the quality of identified parameters. We can conclude that solar radiation and heating power are the main influential inputs for modeling thermal behavior with an influence index of 20% and 70% respectively. On the other hand, we also remark that model parameters are impacted by the location and the orientation of the building. Once again, it has been shown that the PWARX model is more suitable for any configuration than the ARX and its derived models. Indeed, through a comparative study between the piecewise ARX model and other existing



models such as nonlinear ARX, indexed ARX and ARX models, the PWARX model gives good results in terms of indoor temperature estimation with 78.48% accuracy. Furthermore, a comparison procedure details how the PWARX model could be used to estimate and to compare the thermal behavior of different archetype buildings.

Hence, the future works will be directed to extend this approach for predicting energy performance certificates for existing buildings, for helping to derive strategies for better energy efficiency of the Heating, Ventilation and Air-Conditioning system (HVAC) while taking into consideration the occupant's activities and comforts, and usage modes.

**Acknowledgements** This work was supported by the European project "SHINE: Sustainable Houses in Inclusive Neighborhoods". A project granted by Interreg 2 Seas and the European Regional Development Fund.

## References

- [1] A. Abdul, F. Alan S., J.-S. Farrokh, and R. Kaamran. Development and performance comparison of low-order black-box models for a residential hvac system. *Journal of Building Engineering*, 15:137–155, 2018.
- [2] B. Ajib. Data-driven building thermal modeling using system identification for hybrid systems. Technical report, Ph.D. report, Lille University and IMT Lille Douai, 2018.
- [3] B. Ajib, S. Lefteriu, S. Lecoeuche, A. Caucheteux, and J. Gauvrit. Prediction of standardized energy consumption of existing buildings based on hybrid systems modeling and control. Technical report, IEEE Conference on Decision and Control (CDC), Miami Beach, FL, USA, 2018.
- [4] A. Al-janabi, M. Kavgic, A. M. Zadeh, and A. Azzouz. Comparison of energyplus and ies to model a complex university building using three scenarios: Free-floating, ideal air load system, and detailed. *Journal of Building Engineering*, 22:262–280, 2019.
- [5] Amasyali and El-Gohary. A review of data driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205, 2018.
- [6] H. Asgar, X. Chen, M. Morini, M. Pinelli, R. Sainudiin, P. R. Spina, and M. Venturini. Narx models for simulation of the start-up operation of a single shaft gas turbine. *Applied Thermal Engineering*, 93:368–376, 2016.
- [7] R. M. Badreddine. Optimized energy management for a smart building multi-sources multi-charges: different principles of validations. Technical report, Ph.D. report, Université de Grenoble, 2012.
- [8] M. H. Beale, M. Hagan, and H. Demuth. Neural network tool box user's guide. Technical report, The MathWorks, Inc., USA, 2011.
- [9] D. Benjamin. Improving the energy efficiency of an innovative low temperature heating and cooling solution. Technical report, Architecture, Urban space development, Université de Grenoble, 2011.
- [10] M. Benzaama, S. Menhoudj, K. Kontoleon, A. Mokhtari, and M. Lekhal. Investigation of the thermal behavior of a combined geothermal system for cooling with regards to algeria's climate. *Sustainable Cities and Society*, 43:121–133, 2018.
- [11] D. Bond, W. Clark, and M. Kimber. Configuring wall layers for improved insulation performance. *Appl Energy*, 112:235–245, 2013.
- [12] K. Boukharouba, L. Bako, and S. Lecoeuche. Identification of piecewise affine systems based on dempster-shafer theory. Technical report, IFAC Symposium on System Identification, SYSID, Saint Malo, France, 2009.
- [13] D. Crawley, J. Hand, K. M., and G. B.T. Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, 43(4):661–673, 2008.
- [14] P. Fazenda, P. Lima, and P. Carreira. Context based thermodynamic modeling of buildings spaces. *Energy and Buildings*, 124:164–177, 2016.
- [15] F. Ferracuti, A. Fonti, L. Ciabattini, S. Pizzuti, A. Artoni, L. Helsen, and G. Comodi. Data-driven models for short-term thermal behaviour prediction in real buildings. *Applied Energy*, 204:1375–1387, 2017.
- [16] P. Gori, C. Guattari, L. Evangelisti, and F. Asdrubali. Design criteria for improving insulation effectiveness of multilayer walls. *Int J Heat Mass Transf*, 103:349–359, 2016.
- [17] Y. Guo, E. Nazarian, J. Ko, and K. Rajurkar. Hourly cooling load forecasting using time-indexed arx models with two-stage weighted least squares regression. *Energy Conversion and Management*, 80:46–53, 2014.
- [18] P. Herie, M. Nadia, R. Marie, B. Rachid, and M. Eric. Modeling of a building system and its parameter identification. *Journal of Electrical Engineering Technology*, 8(5):975–983, 2013.
- [19] Y. Hiroshi, H. Tianzhen, and N. Natasa. Total energy use in buildings - analysis and evaluation methods. *Energy and Buildings*, 152(IEA EBC annex 53):124–136, 2017.
- [20] B. Hubert. Rules for modeling energy systems in low consumption buildings. Technical report, ENSM Paris, 2014.
- [21] Y. Junjing, S. Mattheos, L. Siew Eang, and D. Chirag. Energy performance model development and occupancy number identification of institutional buildings. *Energy and Buildings*, 123:192–204, 2016.
- [22] C. Lin, B. Biswajit, and M. C. David. Fractional order models for system identification of thermal dynamics of buildings. *Energy and Buildings*, 133:381–388, 2016.
- [23] T. Lin, B. Horne, P. Tino, and C. Giles. Learning long-term dependencies in narx recurrent neural networks. *IEEE Trans Neural Network*, 7:1329–1338, 1996.
- [24] K. Liu, T.-Z. Liu, P. Jian, and Y. Lin. The re-optimization strategy of multi-layer hybrid building's cooling and heating load soft sensing technology research based on temperature interval and hierarchical modeling techniques. *Sustainable Cities and Society*, 38:42–54, 2018.
- [25] T. Markus and E. Morris. Buildings, climate and energy. Technical report, Pitman Publishing Limited, London, 1980.
- [26] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer. Identifying services for short-term load forecasting using data driven models in a smart city platform. *Sustainable Cities and Society*, 28:108–117, 2017.
- [27] M. Mechaqrane and M. Zouak. A comparison of linear and neural network arx models applied to a prediction of the indoor temperature of a building. *Appl Neural Comput*, 13:32–37, 2004.
- [28] M. Paulus, C. T., E. David, and C. Culp. Algorithm for automating the selection of a temperature dependent change point model. *Energy and Buildings*, 87:95–104, 2015.
- [29] T. Reddy and D. Claridge. Using synthetic data to evaluate multiple regression and principal component analysis for statistical modeling of daily building energy consumption. *Energy and Buildings*, 21:35–44, 1994.
- [30] K. Rick, v. S. Jos, and S. Henk. Inverse modeling of simplified hygrothermal building models to predict and characterize indoor climates. *Building and Environment*, 68:87–99, 2013.
- [31] Z. Roberto, F. Gustavo, H. Oliveira, and N. Mendes. Development of regression equations for predicting energy and hygrothermal performance of buildings. *Energy and Buildings*, 40:810–820, 2008.
- [32] S. Royer, S. Thil, T. Talbert, and M. Polit. A procedure for modeling buildings and their thermal zones using co-simulation and system identification. *Energy and Buildings*, 78:231–237, 2014.
- [33] R. Sarwar, H. Cho, S. J. Cox, P. J. Mago, and R. Luck. Field validation study of a time and temperature indexed autoregressive with exogenous (arx) model for building thermal load prediction. *Energy and Buildings*, 119:483–496, 2017.
- [34] J. Siroky, F. Oldewurtel, J. Cigler, and S. Prvara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88:3079–3087, 2011.
- [35] M. Soleimani-Mohseni, B. Thomas, and P. Fahlen. Estimation of operative temperature in buildings using artificial neural networks. *Energy and Buildings*, 38(6):635–640, 2006.
- [36] S. Wang and X. Xu. Simplified building model for transient thermal performance estimation using ga-based parameter identification. *International Journal of Thermal Sciences*, 45:419–432, 2006.
- [37] K. Yun, R. Luck, P. J. Mago, and H. Cho. Building hourly thermal load prediction using an indexed arx model. *Energy and Buildings*, 54:225–233, 2012.