



HAL
open science

Data-driven building thermal modeling using system identification for hybrid systems

Balsam Ajib

► **To cite this version:**

Balsam Ajib. Data-driven building thermal modeling using system identification for hybrid systems. Automatic. Ecole nationale supérieure Mines-Télécom Lille Douai, 2018. English. NNT : 2018MTLD0006 . tel-02899639

HAL Id: tel-02899639

<https://theses.hal.science/tel-02899639>

Submitted on 15 Jul 2020

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.

IMT Lille Douai



Université de Lille



THESE

présentée en vue
d'obtenir le grade de

DOCTEUR

en
Automatique
par

Balsam AJIB

DOCTORAT DE L'UNIVERSITE DE LILLE
DELIVRE PAR IMT LILLE DOUAI

Data-driven building thermal modeling using system identification for hybrid systems

Soutenue le 16 Novembre 2018 devant le jury d'examen:

Rapporteur	Thierry POINOT	Professeur, Université de Poitiers
Rapporteur	Monika WOLOSZYN	Professeur, Université Savoie Mont Blanc
Examineur	Didier DEFER	Professeur, Université d'Artois
Examineur	Gilles GUERASSIMOFF	Maître-assistant, Mines ParisTech
Directeur de thèse	Stéphane LECOEUCE	Professeur, IMT Lille Douai
Encadrant	Antoine CAUCHETEUX	Ingénieur/chercheur, CEREMA
Encadrant	Sanda LEFTERIU	Maître-assistant, IMT Lille Douai

Unité de Recherche Informatique et Automatique URIA
IMT Lille Douai

Ecole Doctorale Sciences Pour l'Ingénieur EDSPI

To my parents ...

Acknowledgements

I express my sincere appreciation to those who have contributed to this thesis and supported me in different ways during the last three years.

First, I would like to express my sincere gratitude to my advisor Prof. Stéphane Lecoeuche for the opportunity that he gave me, for his valuable contributions, suggestions, and comments. I would like to thank him for his patience, motivation, and immense knowledge. His guidance sure helped me at various stages of my research.

I am grateful to my supervisors, Dr. Antoine Caucheteux and Dr. Sanda Lefteriu for their insightful comments and suggestions. I really appreciate their willingness to help me whenever I need any clarifications. I thank them for their generosity, availability and support.

My sincere gratitude to my reading committee, Prof. Monika Woloszyn and Prof. Thierry Poinot for their remarks and suggestions. I also acknowledge the rest of my jury members, Prof. Didier Defer and Dr. Gilles Guerassimoff for their presence, remarks and the interesting discussion we had.

I would like to thank my fellow labmates for the endless discussions and for all the fun we have had in the last three years. A very special thanks goes to Enjie and Ziad, for being the best officemates one can ever have. Their presence in the office made work much more bearable, they even made it fun with all the crazy moments and endless discussions that we had. I also would like to thank the rest of the group: Monika, Pablo, Johann, Tony and Pau for all the stimulating discussions that we had. A special thought goes also to Raghda and Loulou. These people made my stay in Douai a lot easier and gave me the best memories of these three years.

I would also like to thank my "lebanese group": Joelle, Noura, Joseph, Reine and Eliane for all the crazy and happy moments we had in the past years. I particularly remember my first night in Douai, I met Joelle, and found in her my relief. I can't imagine how my first six months would have passed in Condorcet without the fun I had with her, Noura, Mohammed, Mark, Evi and Pau.

I would like to thank all my fellow labmates in DIA. I had the pleasure to meet and work with them. In particular I thank Hajer, Sido, Hichem, Aziz, Lucien, Lee, Koceila, Imen, Djidula, Baya, Sang, Hang, Abir, Eric, Jacques, Guillaume, Anthony, Lala, Muriel and many others. Their presence made everything easier.

I would like to thank my dearest friends Mira and Firas, not just for the last three years, but for the last eight years. One cannot imagine more understanding, listening and caring friends. They are the kind of people that will always be there to cheer you up, to give advice and to go into the deepest discussions with you.

Words cannot express the feelings I have for my parents and my sisters Loulou, Farah, Bosh, Hiba and Rana for their constant unconditional love and support. I would not be here if it was not for them. I dedicate this thesis to my parents and thank them for all the hard work and sacrifices they had to make to get me where I am now. A special thought goes also to May for her love and moral support during these past years.

Last, but not least, I thank Dany for his continued and unfailing love, support and understanding during my pursuit of Ph.D degree. He was involved in every single detail, and he was happy to do it. He was always around, celebrating my happiest moments and helping me to keep things in perspective during hard times. I greatly value his contribution and deeply appreciate his belief in me. Words would never say how grateful I am to him. I consider myself the luckiest in the world to have such a loving and caring partner and family, standing beside me with their love and unconditional support.

Contents

Acknowledgements	v
List of Figures	xii
List of Tables	xiii
List of Abbreviations	xv
List of Symbols	xviii
General Introduction	1
Background	1
Motivation	4
Contribution	6
1 State of the art in building thermal modeling and energy performance prediction	9
1.1 Introduction	9
1.2 "White box" models	11
1.2.1 CFD approach	11
1.2.2 Zonal approach	12
1.2.3 Nodal or multi-zonal approach	14
1.2.4 Summary and discussion on the physical models	16
1.3 "Black box" models	17
1.3.1 System Identification via parametric models	18
1.3.2 Artificial Neural Networks (ANNs)	21
1.3.3 Support Vector Machines (SVMs)	23
1.3.4 Genetic Algorithms (GAs)	24
1.3.5 Decision trees	25
1.3.6 Summary and discussion on the statistical models	26
1.4 "Grey box" models	27
1.4.1 Thermal networks based "grey box" models	27
1.4.2 Other "grey box" examples	29
1.4.3 Summary and discussion on the "grey box" approaches	30
1.5 Conclusion	31
2 Hybrid system approach: motivation, background and applied methodology	33
2.1 Introduction	33
2.2 Theoretical Foundations	35
2.2.1 Lumped RC thermal model for a building	35
2.2.2 Physical interpretation of model switches	38
2.2.3 Equivalence between PWARX models and local transfer functions derived from the thermal network	39

2.3	Background on hybrid systems identification	41
2.4	PWARX algorithm	44
2.5	Conclusion	47
3	Evaluation of the proposed modeling approach	49
3.1	Introduction	49
3.2	Important points about data collection	50
3.3	Model performance evaluation criteria	50
3.4	Test case 1: Eco-Confort room in Douai	51
3.4.1	Description of the test case	51
3.4.2	Scenario 1: heating ON/OFF	52
3.4.2.1	Results for the estimation data set	53
3.4.2.2	Results for the validation data set	55
3.5	Test case 2: Virtual room created with Dymola software	56
3.5.1	Description of case study	57
3.5.2	Scenario 2: heating ON/OFF, ventilation switch, shades open/closed and change in occupancy	57
3.5.2.1	Results for the estimation data set	58
3.5.2.2	Results for the validation data set	61
3.6	Occupied building	62
3.6.1	Description of the case study	63
3.6.2	Combining PWARX model with SVM classifiers	65
3.6.2.1	Need for an automated classification technique	65
3.6.2.2	SVM classifiers	67
3.6.2.3	Estimating regions in the regression space using SVM	68
3.6.2.4	Predicting discrete states using SVM	70
3.6.3	Estimation results	71
3.6.4	Validation results	73
3.6.4.1	SVM validation	73
3.6.4.2	PWARX validation results	73
3.7	Discussion of results and conclusion	75
4	Predicting the energy consumption of buildings using a data driven PWARX model	79
4.1	Need for estimating the energy consumption of buildings	79
4.2	Proposed control methodology	80
4.2.1	Simulate the identified models using standardized scenarios in closed loops with PID controllers	81
4.2.2	Evaluation of the energy consumption	85
4.3	Case study-HUMIBATEX Angers	85
4.4	Results	87
4.4.1	Estimation and validation results	87
4.4.1.1	Before insulation	87
4.4.1.2	After insulation	89
4.4.2	Evaluating the energy consumption before and after insulation	89
4.5	Conclusion	93
	Conclusion and future work	95
	References	99

Abstract

List of Figures

1	(a) Share of each sector in the final energy consumption in the EU in 2015. (b) Share of buildings sector constituted by household and services sectors. <i>Source: Eurostat, EU-28, 2015</i>	2
1.1	Building thermal modeling approaches	10
1.2	Graphical representation of a Neural Network	22
1.3	Graphical representation of the activation function in an ANN	22
1.4	Graphical representation of a decision tree	26
1.5	Example of an RC thermal network for buildings [Bacher and Madsen, 2011]	28
2.1	RC model of the building	35
2.2	Example of repartition of a PWARX system with 3 sub-models	45
2.3	Three sub-models in the regression space with schematic explanation of the distances taken into account in Eq. (2.27) [Boukharouba, 2011].	46
3.1	Eco-Confort room used to collect measurements for scenario 1	52
3.2	Model inputs for scenario 1: heating ON/OFF	53
3.3	Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for scenario 1	54
3.4	Estimation results for scenario 1 (heating ON/OFF). From the top to the bottom: measured and estimated temperature; detected discrete states (number of configurations); heating power input.	54
3.5	Prediction results over two days for scenario 1 (heating ON/OFF). Figure on top: measured and predicted temperature using PWARX. Figure in the bottom: heating power input.	55
3.6	Prediction results for scenario 1 (heating ON/OFF) using PWARX and classical ARX with two model structures: ARX 551 that had the best Fit and ARX 541 to test the same structure as the one found for the PWARX model	56
3.7	Virtual room created with Dymola software used to generate data for scenario 2	57
3.8	Weather inputs for the estimation and validation periods for scenario 2	58
3.9	Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for scenario 2	59
3.10	Estimation results for scenario 2 (heating, ventilation, shades and occupants switch). From the top to the bottom: temperatures estimated by PWARX model and obtained for the virtual room; detected discrete states; air change rate and occupancy scenarios; shades and heating scenarios.	60

3.11	Prediction results for scenario 2 (heating, ventilation, shades and occupants switch). From the top to the bottom: temperatures predicted by PWARX model and obtained for the virtual room, detected discrete states; air change rate and occupancy scenarios; shades and heating scenarios.	62
3.12	Prediction results for scenario 2 using a classical ARX model with two structures: ARX 321 to test the same structure as the one found for the PWARX model, ARX 431 that had the best Fit=34%.	62
3.13	Weather data recorded for January 2014	65
3.14	House data recorded for January 2014	66
3.15	Ventilation system data recorded for January 2014	66
3.16	Bi-dimensional linear classification using SVM	68
3.17	non-linear SVM classification by feature space transformation	69
3.18	SVM learning	69
3.19	SVM classifier functioning	71
3.20	Predicting the temperature using SVM and PWARX algorithm	71
3.21	Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for the case of the occupied building	72
3.22	PWARX estimation results for the third test case of the occupied building	72
3.23	PWARX simulation results for the whole prediction period	74
3.24	Comparison between PWARX and ARX simulation results for the whole prediction period	74
3.25	One hour ahead prediction using PWARX	76
4.1	Temperature regulation by closed loop control	81
4.2	Control response characteristics	84
4.3	Control loop diagram	85
4.4	Building under study-HUMIBATEX Angers	86
4.5	PWARX estimation results before insulation	88
4.6	PWARX validation results before insulation	88
4.7	PWARX estimation results after insulation	89
4.8	Validation after insulation	90
4.9	Ambient air temperature (used as input) and ventilation mode for the simulated scenario	91
4.10	Control results before insulation	91
4.11	Control results after insulation	92

List of Tables

3.1	Comparison of validation criteria computed for prediction results of scenario 1 using PWARX and ARX models	56
3.2	Scenario 2 configurations	58
3.3	Configurations related to each sub-model estimated in scenario 2	61
3.4	Comparison of validation criteria computed for prediction results of scenario 2 using PWARX and ARX models	62
3.5	Validation criteria for prediction results of PWARX and ARX models	75
3.6	Validation criteria for prediction results of PWARX and ARX models	75
4.1	Scenarios applied for this case study	86
4.2	PID controllers parameters for each sub-model	90
4.3	Energy performance gain obtained by our approach and classical physics approaches	92

List of Abbreviations

ABV	Air Blowing Ventilation
ANN	Artificial Neural Network
ARMAX	AutoRegressive Moving Average models with eXogeneous inputs
ARX	AutoRegressive models with eXogeneous inputs
BES	Building Energy Simulation
BJ	Box-Jenkins
CFD	Computational Fluid Dynamics
DFV	Double Flux Ventilation
DHW	Domestic Hot Water
EBC	Energy in Buildings and Communities program
ELC	Extended Linear Complementarity
EU	European Union
EUI	Energy Use Intensity
GA	Genetic Algorithm
GDP	Gross Domestic Product
HTC	Heat Transfer Coefficient
HVAC	Heating Ventilation and Air Conditioning
IEA	International Energy Agency
LC	Linear Complementarity
LTI	Linear Time Invariant
MAE	Mean Absolute Error
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MLD	Mixed Logical Dynamical
MMPS	Max-Min-Plus-Scaling
MPC	Model Predictive Control
MVHR	Mechanical Ventilation system with Heat Recovery
NAC	Normalized Annual Consumption
NARX	Nonlinear AutoRegressive models with eXogeneous inputs
NNARX	Neural Network AutoRegressive models with eXogeneous inputs
OE	Output Error
PCM	Phase Change Materials
PID	Proportional Integral Derivative
POMA	Pressurized zOnal Model with Air-diffuser
PRISM	PRInceton Scorekeeping Method
PV	Photo Voltaic
PWA	PieceWise Affine
PWARX	PieceWise AutoRegressive models with eXonegeneous input
RBF	Radial Basis Function
RC	Resistance Capacitance

RMSE	Root Mean Squared Error
RT	Réglementation Thermique-French (Thermal Regulations for buildings)
SARX	Switched AutoRegressive models with eXogeneous input
SI	System Identification
SISO	Single Input Single Output
SVM	Support Vector Machine
UK	United Kingdom
VGS	Vertical Greenery Systems

List of Symbols

$\alpha\phi_s$	solar gains on the outer surface of the exterior wall (W)
$\bar{\varphi}$	extended regression vector
$\beta\phi_s$	solar gains transmitted through the windows (W)
ϕ_h	heat flow from the heaters (W)
ϕ_{occ}	heat flow from the occupants (W)
θ	vector of parameters defining the model
φ	regression vector
C_e	thermal capacitance of exterior wall (JK ⁻¹)
C_{film}	thermal capacitance of the air film next to the heater (JK ⁻¹)
C_h	thermal capacitance of the heating system (JK ⁻¹)
C_{in}	thermal capacitance of the air inside the room and furniture (JK ⁻¹)
C_{occ}	thermal capacitance of the human body (JK ⁻¹)
n_a	order of the output variable
n_b	order of the input variable
n_k	pure delay between the input and the output variables
n_u	number of input variables
n_y	number of output variables
R_{e1}	heat transfer coefficient between the exterior environment and the envelope (exterior convective and radiative resistance + half of the conduction resistance for the envelope, we consider the mean temperature in the middle of the walls) (KW ⁻¹)
R_{e2}	heat transfer coefficient between the room environment and the envelope (interior convective and radiative resistance + half of the conduction resistance for the envelope) (KW ⁻¹)
R_{film}	heat transfer coefficient between the exterior environment and the air film close to the heater and exterior wall (KW ⁻¹)
R_h	convective and radiative resistance of the heating system (KW ⁻¹)
R_{occ}	conductive+convective+radiative heat transfer resistance between the human body surface and the indoor air temperature) (KW ⁻¹)

R_{sh}	thermal resistance of the shades (KW^{-1})
R_{vent}	resistance for natural ventilation (KW^{-1})
R_{wd}	thermal resistance of windows (KW^{-1})
T_{ex}	outdoor ambient temperature (K)
T_e	middle temperature of the envelope (K)
T_{film}	temperature of air film close to the heater and exterior wall (K)
T_h	surface temperature of the heating system(K)
T_{in}	room air temperature (K)
T_{occ}	temperature of the human body surface (K)
T_{surr}	mean air temperature of surrounding offices (K)
cNN	c nearest neighbors
R^2	Coefficient of determination
R_{adj}^2	Adjusted coefficient of determination

General Introduction

As its title indicates, this thesis treats the subject of building thermal modeling using data-driven techniques, where data collected from the system is used to find a suitable mathematical model of the building. This thesis proposes a special class of models called hybrid models for representing the thermal dynamics in a building. Hence, the introductory part of this work provides a general background on the importance of building thermal modeling together with the motivation behind this thesis and its main contributions.

Why is having accurate models for the thermal behavior of buildings so important?

As the quality of human life has been improving for the past decades, a discernible increase in the energy consumption has been observed and this has risen concerns about climate change and energy shortage. Hence, the European Commission has adopted an energy efficient plan in March 2011 which was followed, in October 2012, by a directive (2012/27/EU) on the energy efficiency. The aim was to implement actions in a common framework to promote energy efficiency and also to foresee the establishment of indicative national energy efficiency targets for 2020. European Union (EU) countries then agreed on a new 2030 framework for climate and energy, including EU-wide targets and policy objectives for the period between 2020 and 2030. The EU wishes to achieve a secure and sustainable energy system to meet its long-term 2050 target of cutting greenhouse gas emissions to 80% below 1990 levels [[EUROPA, 2012](#); [European Commission, 2011](#)].

Among all economic sectors, the building sector is a major energy consumer. Therefore, a directive specifically dedicated to the energy performance of buildings was established in 2010 (DIRECTIVE 2010/31/EU) [[EPBD, 2010](#)]. This directive stated that the building sector is responsible for the biggest part of the final energy consumption ¹ with a share that goes up to 40% and also 36% of CO₂ emissions in Europe. Fig. 1 shows the distribution of the final energy consumption in the EU per sectors in 2015. The residential sector, which includes the households sector and the services sector, accounts for 39% of the final energy consumption in the EU. The sector being in full expansion, its energy consumption is bound to increase. Therefore, measures such as the reduction of energy consumption and the use of energy from renewable sources in the building sector constitute important measures needed to reduce the EU energy dependency and greenhouse gas emissions.

¹Final energy consumption is the total energy consumed by end users, such as households, industry and agriculture. It excludes energy used by the energy sector for deliveries and transformation. Primary energy consumption measures the total energy demand of a country. It covers consumption of the energy sector itself, losses during transformation (for example, from oil or gas into electricity) and distribution of energy, and the final consumption by end users.

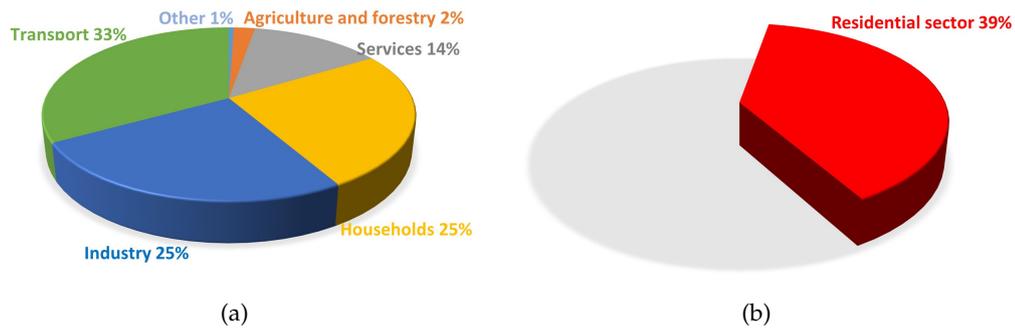


FIGURE 1: (a) Share of each sector in the final energy consumption in the EU in 2015. (b) Share of buildings sector constituted by household and services sectors. *Source: Eurostat, EU-28, 2015*

Energy consumption of buildings is related to various factors including: architecture, thermo-physical properties of the building elements and construction materials, climate zone, quality of the installed systems for Heating, Ventilation and Air Conditioning (HVAC) and occupants' behavior, just to name a few. Therefore, to meet the goals set by the EU nations, one has to act from several directions. The most effective way to improve the energy efficiency without reducing the indoor environment quality and the tenants well-being is to design the building in a way to take maximum profit from the environment and the natural energy sources (like day-light, natural ventilation, passive solar heating during the cold season, passive cooling in the warm season, etc.). When designing a building in cold climates, the heat losses through the envelope and ventilation should be reduced (by using high thickness insulating materials, high performance double or triple glazed windows and highly efficient heat recovery units coupled with mechanical ventilation systems) and the solar gains should be maximized. In hot climates, insulation materials help reduce the heat fluxes entering the building from the outdoor environment and solar shading devices protect sun facing windows. Early constructed buildings rarely have these specifications, hence, retrofitting and renovation should be considered for these buildings to improve their energy efficiency, along with the installation of energy efficient systems. Still, in all these cases, occupants play a major role in managing the energy consumption of the building as they interact with control systems and building elements to reach their own personal desired level of comfort. This is done in different ways like turning the air-conditioning ON or OFF, adjusting the thermostat temperature, opening and closing windows, adjusting blinds, using hot water, etc. Some solutions were proposed with the aim of encouraging tenants to optimize their energy use [Guerassimoff and Thomas, 2015]. Despite the variety of research work dedicated to modeling the occupants' behavior [Vorger, 2014; Amayri et al., 2016], this remains one of the most complex tasks in building thermal modeling.

At the building level, the potential actions for improving energy efficiency are facilitated by the availability of accurate models providing a description of the thermal behavior of buildings and a good estimate of their energy performance. A model is a description of the real physical system. A thermal model of a building is a model that describes the heat transfer phenomena occurring in the system. The purpose of the model is to observe and predict the thermal response of the building under normal use. The need for a model describing the thermal behavior of a building

arises in many stages of its life cycle and for many purposes among which we cite the following:

- At the design stage, the thermal model helps predict the general thermal behavior of the building in terms of its location, architecture and materials used. At this stage, the thermal model could give insight into the best architecture and materials to be used to improve the energy efficiency of the building. The surface, orientation and composition of windows could, for example, change in order to maximize or minimize the heat gain from the sun depending on the position and location of the building.
- The appropriate design of Heating Ventilation and Air Conditioning (HVAC) systems also depends on the estimate of maximum loads needed for heating and/or cooling the building which could be calculated using the thermal models. The maximum load defines the best size for the HVAC systems.
- Another important information that the thermal models deliver at the design stage is the compliance of the energy efficiency of the building with the regulations of each country. According to the legislation issued by the EU, the member states must set minimum energy performance requirements for new buildings, for the major renovation of buildings and for the replacement or retrofit of building elements (heating and cooling systems, roofs, walls and so on). In France, the new thermal regulation, RT2012, sets the consumption of primary energy in new constructions to a maximum of 50kWh/m²/year [RT 2012]. Owners are therefore required to undertake an energy study of their proposed new construction to prove that it will have a primary energy consumption lower than 50 kWh/m²/year. On completion of the works, owners will also be required to attest that the energy performance has been achieved, for which purpose a post-completion study will also need to be undertaken. In order to assess the compliance of new buildings to this regulation, tools were introduced to provide the calculation of regulatory indicators on the energy efficiency of the building and the degree of comfort of occupants. The calculations can be made by specialist consultants using computer software programs specifically designed for this purpose. The studies are done by creating the building model based on its architectural and physical properties and then simulating the model under normal usage scenarios of occupancy, set point temperature, lighting and ventilation for each type of buildings.
- In the early lifetime of a building, there might be a disagreement between the energy use estimated in the design phase and the actual energy use. To examine this discrepancy, a model is needed to check the operation of the building. The simulation should have the occupancy schedules that reflect the actual occupancy of the building. These simulations may be used at this stage to refine and optimize control strategies and to bring adjustments to the building (like replacing the HVAC equipment) to achieve the requirements of the project as initially set by the owner and as designed by architects and engineers. This is known as building commissioning.

- The gain in the energy efficiency of a building is highly related to the efficiency of its energy systems. The biggest share of the energy consumption in a building comes from its HVAC systems. Therefore, it is important to simulate the response of the building under different HVAC control strategies and choose the most efficient solution. In this context, the comfort of the occupants is an important factor to keep in mind, which is why modeling the indoor air temperature of a building or a building zone and the control of HVAC systems has become an important research topic in this area. This could be done at the design stage or later on during operation.
- A big share of buildings, in France and worldwide, were built before the existence of any thermal regulation. The thermal renovation of these buildings plays an important role in reducing the energy consumption and the environmental impact of the building sector. A thermal model is used in this case to investigate the impact of different renovation strategies applied to a building and hence deciding on the most effective ones.
- Recently, the advances that were made in the smart grid² revealed the necessity for building models providing real time observations and short time predictions of their energy need in order to support the decision making process. Furthermore, the concept of smart buildings and positive energy buildings that use renewable energies to fulfill their energy needs, offers high flexibility in energy demand and this also represents a challenge to the smart grid and should be predicted using accurate thermal models.
- A thermal model could also be used for detecting faults in the energy systems in a building. On big sites, a complex heating or cooling system exists with multiple components and subsystems including cooling towers, chillers, pumps, heat exchangers, etc. A fault in one of these components often leads to a failure in the whole system. A thermal model could be used in this case for detecting and locating the faults.

Hence, a model describing the thermal behavior of a building system is needed in the design phase, as well as during operation, for the implementation of actions that contribute to enhancing its energy efficiency. During the last decades, many modeling techniques were developed for this purpose, some of them being particularly adapted to the design phase and others being more suitable for operation.

Advantages and limitations of existing techniques in modeling the thermal behavior of buildings

A model can be formulated either as a set of equations based on the underlying physics or as a set of purely mathematical equations, with no physical interpretation, relating the inputs to the outputs of a system. Existing thermal models for buildings are divided into three main categories: "white box" models, "black

²The smart grids are modern electric power grid infrastructures for enhanced efficiency and reliability through automated control, high-power converters, modern communications infrastructure, sensing and metering technologies, and modern energy management techniques based on the optimization of demand, energy and network availability (definition from [Kolokotsa, 2016]).

box" models and "grey box" models. "White box" models are based on solving the equations that describe the physical heat transfer phenomena in a building coupled to aerodynamic and hygroscopic transfers. They are also known as physical models or forward modeling approaches. "Black box" models, on the other hand, do not require any physical information. They are based on mathematical equations developed based on samples of data representing the system from which they were collected. "Grey box" models use data and a model based on physical interpretation to determine the unknown parameters of the model, hence combining these former approaches.

"White box" models are generally seen as robust models for modeling the thermal behavior of buildings because they are based on equations that represent the actual physical behavior of the system. They have been used for more than 60 years now in the building industry and they are very effective in the design phase. However, because of their dependency on the physical equations, these models need a detailed description of the physical system. An extensive knowledge about the building geometry, the type of materials used for construction, as well as details about the HVAC systems in the building, the occupants' behavior and the outdoor environment (nearby buildings, vegetation, etc.) are required. Therefore, "white box" models are mostly used in the preliminary design phase where a detailed description of the building is available. For existing buildings, a detailed description is not always available, and even if it was, errors related to the aging of materials, or to the fact that the builder did not follow the specifications exactly might lead to bias between the actual behavior and the simulated one. Researchers tried to find alternative models to deal with the lack of information in this case. They saw that "black box" or "grey box" models, being constructed from measured data, provide a feasible alternative to the characterization of existing buildings. The main limitation of these latter models is that they need rich sets of data collected during the operation of the building, therefore they cannot be used in the design phase. They would, however, make very good candidates for providing simple models that could be employed for real time observations and short time predictions in order to implement the best control strategies, or for helping the facility manager with the building maintenance and optimization. "White box" models, being more complex and more time consuming, show limitations from this point of view.

In most EU countries, half of the residential stock was built before 1970, when the first thermal regulations were introduced [EUROPA]. Thus, it is important to predict the thermal behavior and energy performance of these buildings, optimize their operation and control and integrate them in the growing smart grid. Having easily automated models that give quick results using actual data from these buildings during normal usage would offer tremendous benefits in terms of time and costs. "Grey box" and "black box" models represent both relatively easy to implement models and are built using data collected from the real system. "Grey box" models start from a preliminary description of the building (hence they rely heavily on expert knowledge) and the values of the parameters are found by solving an optimization problem to minimize the error between the data and the model. This is a growing research topic, but issues related to the optimization problem like the correctness of the model structure, the initial guess for the parameters and the large number of parameters make the application of these approaches rather limited. On the other hand, "black box" models offer advantages like accuracy, simplicity, ease of implementation, low computational cost and little need for expertise and time for building the model. Hence, in this thesis, we propose to

study the feasibility of reproducing the thermal behavior of buildings using "black box" models constructed entirely from input output data of the system.

Contribution of the thesis

In this thesis, we present a novel data driven technique for modeling the thermal behavior of buildings based on a hybrid system approach. A hybrid system is a system that consists of both continuous and discrete dynamics. A building is a complex system characterized by a nonlinear behavior and the occurrence of different events like opening/closing windows and shades, turning the heating ON/OFF, etc. The appearance of these events leads to a different dynamic thermal behavior in the building which cannot be described by linear models. The main limitation of most "black box" techniques used in the literature is the assumption of a linear structure. Therefore, this work proposes to describe the thermal behavior of buildings via a nonlinear model by linearizing it around different operating points. This leads to a hybrid structure combining the continuous nature of the thermal behavior in the building and the discrete dynamics that result from the occurrence of different configurations or events. We use a System Identification (SI) technique called Piece Wise AutoRegressive eXogeneous inputs (PWARX) for clustering the data into several groups representing different configurations, each configuration being modeled by a mathematical equation.

The contributions of the thesis can be summarized in the following points:

- A lumped Resistance Capacitance (RC) model of a building which employs switches to account for sudden changes due to turning the heating on/off and opening/closing the windows and the shades.
- A new data-driven approach for modeling the thermal behavior of building zones which uses hybrid system identification together with support vector machines for data classification.
- A methodology for predicting the energy consumption of buildings by integrating the hybrid model in a hybrid control loop.

The manuscript is organized as follows:

- Chapter 1 summarizes the state-of-the-art related to the thermal modeling of buildings. Research related to all three modeling approaches "white box", "black box" and "grey box" along with their advantages and drawbacks is presented. The common theme of the work included in this state of the art is that its objective is to predict the thermal behavior of buildings and their energy performance.
- Chapter 2 presents the motivation for using a hybrid system approach for the thermal modeling of buildings. The equations derived from a building described as an RC circuit are used to prove that the occurrence of some sudden events like turning the heating ON/OFF, opening/closing the windows or shades and the presence or absence of occupants leads to different configurations characterized by a change in the value of some parameters appearing in the transfer function of the system thus yielding a new dynamical model. The proposed PWARX approach is also presented in detail. It is based on a

clustering technique to group similar data points (representing the same configuration) and an estimation technique based on least squares to estimate the model behind each cluster.

- Chapter 3 details the various test cases employed for the numerical validation of the proposed PWARX technique in modeling the indoor air temperature of buildings. These range from a small unoccupied room to an entire occupied building. We showcase in this chapter the ability of this technique to describe and predict the thermal behavior of the buildings when different configurations occur.
- Chapter 4 describes a full methodology aiming at predicting the energy consumption of buildings under normal use. The previously proposed approach is completed by integrating the identified hybrid model in a closed loop with a hybrid control for simulating standardized scenarios and predicting the amount of energy consumed for each of them. A test case of a building located in France, for which renovation work was done, is presented. We show that the improvement in the energy performance due to renovation work can be accurately and quickly predicted using our developed approach based on data collected from the system without any knowledge on the physical details of the building.
- The last part of the thesis is dedicated to summarizing these research findings, drawing a conclusion on the previous chapters and outlining some suggestions for future work.

Chapter 1

State of the art in building thermal modeling and energy performance prediction

This first chapter gives an overview of the state of the art on building thermal and energy modeling. Such models can generally be divided into three main categories: "white box", "black box" and "grey box" models. This chapter presents the most important work done within each one of these areas.

1.1 Introduction

The thermal modeling of buildings is generally characterized by two thermal behavior: the static thermal behavior and the dynamic thermal behavior. In the static thermal behavior, energy assessment is performed in steady state conditions while considering a limited number of factors. They are usually applied to simplify the thermal model and to overcome the limitations of computing resources. They are mostly used for regulatory calculations and energy labeling in order to compare different performances in standard use conditions. They can also be used to calculate relative parameters like the heat loss coefficient. A static model is not usually seen as a robust model to give accurate and precise information about the real energy consumption of buildings because important factors, like the inertia of the building and the occupants' behavior, are not taken into consideration. On the other hand, the dynamic thermal simulation seeks to understand the thermal exchange between the building and its environment. It studies the transient response of the building under different weather conditions, as well as under the influence of the behavior of the occupants, the use of electrical devices, the characteristics of the heating/cooling system, etc. They are capable of capturing the building and systems dynamics and are seen as robust models for predicting the thermal behavior of buildings. Therefore, this study will focus only on dynamic models.

Before detailing the methods employed for the characterization of the thermal behavior of buildings, some terms have to be defined. The model is the set of equations that compute an output variable in terms of input variables. For instance, the output of the thermal model of a building could be the indoor air temperature and the inputs could be, among others, all weather related data. We call parameters the elements of the equations that define the model. They represent the link between the output/s of the model and the inputs. For example, in a linear regression, the parameters are the coefficients that are multiplied by

the input variables, while in a conduction heat transfer equation, the thermal conductivity is a parameter. There exist different types of equations, i.e., models, to represent a system and different techniques for identifying their parameters. In what follows, the main techniques employed for building simulations are presented.

Building thermal modeling approaches are divided into three main categories: the physics based approaches known as "white box", the data based approaches known as "black box" and a third approach joining the two previous ones called "grey box". Physics based techniques consist in solving the equations that describe the physical phenomena occurring in a building. Three main physics based thermal building models are currently used: the Computational Fluid Dynamics (CFD), the zonal and the multi-zonal also known as nodal methods. The "black box" models are based on equations that are determined based on measured data of the system. Many techniques are employed for this, among which we cite the System Identification (SI) techniques like regression models, machine learning techniques like SVM, Artificial Neural Networks (ANN), decision trees and optimization techniques like Genetic Algorithms (GAs). The third and last approach is the "grey box". It consists in joining physics based equations and measurements from the system in order to construct the model. This could be done in different ways as will be shown later on, by using different physics based and data-based techniques. The classification of these modeling approaches is presented in Fig. 1.1. A state of the art on building thermal modeling and energy performance prediction is presented in many papers in which the authors explain the difference between these approaches [Koulamasa et al., 2018; Wang and Zhai, 2016; Soto and Jentsch, 2016; Harish and Kumar, 2016; Fouquier et al., 2013; Wang et al., 2012]. Details on each modeling approach together with examples of research papers exemplifying each of them will be presented hereafter.

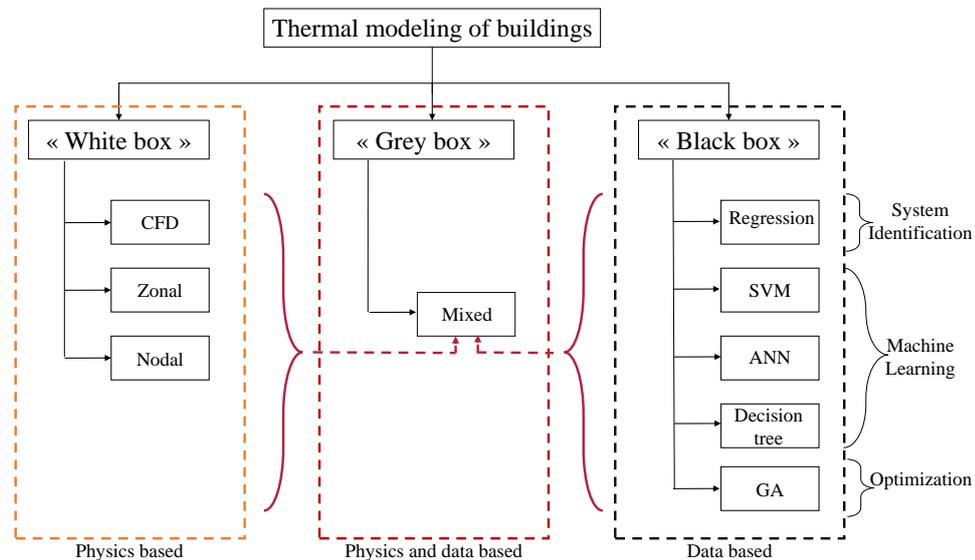


FIGURE 1.1: Building thermal modeling approaches

1.2 "White box" models

"White box" models are based on physical equations that describe the thermal, aerodynamic and hygric transfer between the building and its environment. The choice of equations to implement and the associated phenomena differs from one model to another. They can also be distinguished by the level of details in the employed numeric resolution. The most common physics based approaches used for building thermal modeling are the CFD approach, the multi-zonal or nodal approach and the zonal approach. They are usually incorporated in software tools with a graphical interface to enter the parameters relative to the building architecture, materials used in construction, occupants and energy systems. These are known as Building energy Simulation (BES) tools performing dynamic thermal simulations and they have been accepted as powerful tools for modeling the thermal dynamics of a building and its energy performance. This modeling approach begins with a detailed description of the building, its systems, components and environment like weather data, neighboring buildings, etc., then a thermodynamic building model is built using fundamental engineering principles. It is particularly adapted to the design of buildings and has been historically used for this purpose. Nowadays it is commonly used for verifying the compliance of buildings with the thermal regulations and for evaluating their energy performance.

1.2.1 CFD approach

Computational Fluid Dynamics (CFD) was first introduced to the building industry in the 1970s [Nielsen, 1976]. It is considered the most complete but also the most complicated approach in building thermal simulations. In fact, each building zone is decomposed into a large number of small volumes generated by a mesh, which is why this method is considered to be three-dimensional. It is known for offering a detailed fluid flow description and is mainly used in the building community for predicting indoor and outdoor airflow, pressure, temperature, humidity, pollutants and chemical species distribution. The approach is based on a numerical solution of the governing equations for fluid flow, namely, Navier-Stokes equations.

Many CFD software tools are available, to name a few: Fluent [Fluent, 2012], Massachusetts Institute of Technology MIT-CFD, PHOENICS-CFD [CHAM, 2005], etc. For buildings applications, the CFD approach is mostly dedicated to investigating the effect of airflow from different aspects. For example, the impact of surrounding buildings on the wind characteristics was investigated by [Lee, 2017]. The aim was to study the effect of the number of surrounding buildings considered during a simulation on the resultant wind pressure on the facades and roof of a target building which could be helpful to estimate the infiltration rate. The study was done in Copenhagen city center where a building was under observation and the main variable was the total number of layers of buildings around it. Four different number of layers were tested by taking four different radial distances from the target building. The center was digitally constructed and the simulations were performed using ANSYS Fluent [Fluent, 2012]. CFD was also successfully employed for studying the natural ventilation in buildings. [Prakash and Ravikumar, 2015] studied the effect of the size and orientation of window openings on the thermal comfort of occupants and indoor air flow characteristics in naturally ventilated buildings. The paper focused on studying the effects of window openings on

adjacent walls. The CFD technique was employed to identify a new set of strategies to find the best location of window openings for a three-dimensional room model. In the same context, [Gao and Lee, 2011] evaluated the influence of different opening configurations on natural ventilation performance of a residential unit in Honk Kong. It was found that better natural ventilation performance could be achieved when the two openings groups were positioned in opposite directions or perpendicular to each other. [Song and Meng, 2015] used Fluent to study the improvement of ventilation design in school buildings. The study focused on one classroom where a field study was carried to collect the actual data of boundary conditions for CFD simulation. Four different ventilation systems were simulated, compared and analyzed. The four cases correspond to the number of windows open and the operation of exhaust fans and ceiling fans. In another work, [Guo et al., 2015] proposed a methodology for optimizing the natural ventilation in a green building through CFD wind environment simulation from three aspects, i.e., site planning, building shape and building envelope. The goal is to help architects in designing green buildings that are energy efficient and environmentally friendly by a complete qualitative analysis and evaluation of building natural ventilation. In the same context, [Yang and Jiang, 2017] used CFD to simulate the natural ventilation in a college students dormitory in Jinan. Based on the analyses, the paper puts forward a design strategy and optimization method of natural ventilation in the dormitory. In other works, researchers employed CFD for studying the comfort of occupants regarding the indoor air temperature and velocity. [Zhang et al., 2017] studied the thermal comfort in ancient architecture towers in Beijing using a floor-type fan coil air-conditioner for controlling the physical environment. CFD was used for calculating the indoor temperature and velocity fields in the summer period for one ancient tower building in Beijing.

In these papers, CFD was mostly employed for airflow characterization. For the evaluation of the energy consumption of a building, loads prediction and indoor air temperature prediction, the CFD approach is frequently coupled with BES tools and that will be presented later on in Sect. 1.2.3. In fact, the integration of building energy simulation and computational fluid dynamics programs can provide more accurate predictions about building energy use and indoor environment due to the complementary information provided by the two programs.

1.2.2 Zonal approach

The zonal approach is a simplification of the CFD approach. It consists in dividing each building zone into several cells (small parts of a room) in which the air is assumed to be perfectly mixed with a homogeneous temperature and humidity, hence it is considered as a two-dimensional approach. Zonal models are a promising way to predict air movement in a room with respect to comfort conditions and gradient of temperature because they require smaller computation times compared to the CFD approach. The airflow and temperature distributions are determined by solving the pressure field and the mass and energy balance equations in each cell. It is possible to evaluate the spatial distribution of the temperature, pressure, concentration or air velocity using a reasonable computational time, it is also convenient for investigating thermal comfort, indoor air quality and energy analysis. SimSPARK [Mora, 2003], developed from the object-oriented simulation environment SPARK [SPARK, 2003; Sowell and Haves, 2001], is a frequently employed

building modeling software based on the zonal approach. Another such software is POMA (Pressurized zOnal Model with Air-diffuser) [Haghighat et al., 2001], developed to predict airflow patterns and temperature distribution in a ventilated room.

The zonal approach was initially developed to study the interaction between the terminal unit and the rest of the room [Lebrun, 1971; Laret, 1980; Sandberg, 1987; Bouia and Dalicieux, 1991; Wurtz, 1995]. Later on, it was used for different purposes like modeling temperature stratification [Allard and Inard, 1992], prediction of contaminant distribution [Huang et al., 2005], design [Bozonnet et al., 2005], thermal comfort evaluation [Musy et al., 2002], modeling solar domestic hot water systems [Kenjo et al., 2007], etc. [Megri and Haghighat, 2007] presented a review on the development and applications of the zonal modeling approach for simulating the indoor environment of buildings. They separated the experimentally based zonal models from the generalized zonal models and others that have seen additional developments. The first category has been driven by experimental studies like the one presented by [Ngendakumana, 1991] where the author studied the temperature distribution in a five-zone scheme. He suggested seven airflow paths that were obtained from experimental observation. Then he developed an analytical model, based on the energy conservation equation and some complementary equations like the convective heat flow equation and the power heat flow equation, to predict the temperature in the zones. Similar experiments were conducted and different analytical solutions were proposed [Laret, 1980]. Over the years, the developed models were extended by integrating branches for commercial heating systems and the displacement ventilation. On the other hand, the generalized zonal models are based on the mass and energy conservation equations in different cells, and no assumptions for airflow direction are required. In this context, the object-oriented environment SPARK [Sowell and Haves, 2001] was created.

[Tittlein et al., 2008] compared the effects of two types of ventilation techniques (counter-flow ventilation and a single-flow ventilation) on the energy efficiency of a passive house located in France. They described the thermo-aeraulic behavior in buildings using the SimSpark platform to investigate the effect of counter-flow ventilation on the heating demand and the effect of adding overhangs above windows. In recent works, the zonal approach was mainly employed for predicting airflows and temperature distribution as it is less time consuming than CFD models and more accurate than nodal models which will be detailed in the next section. [Fang et al., 2017] implemented a new algorithm in a developed zonal model computer program based on the POMA software [Haghighat et al., 2001]. The purpose of this algorithm is to predict the indoor thermal environment of a lab room (temperature distribution and airflows) under a natural convection condition and mechanical ventilation. An underfloor air distribution strategy for which the diffusers are set in the floor and they deliver fresh air bottom up rather than top down is used. To assess the accuracy of the implemented algorithm, the simulation results were compared to experimental measurements. [Megri and Yu, 2015] developed a new zonal model (POMA⁺) to predict temperature and airflow distribution within a room in a building. The model was evaluated by comparison to other existing models and calibrated using experimental results. [Kuznik et al., 2011] presented a numerical modeling of a Double Skin Facade (DSF) using a zonal model approach for the mass transfer based on the pressure difference in the DSF. The radiative and convective heat transfers were

also taken into account. The numerical modeling has been validated using the experimental data collected from a full-scale DSF during the summer season. Once the numerical model has been validated, it was used to study the influence of different airflow rates through the air channels of the facade and different angles of the solar shading devices on the heat transfer in the DSF.

1.2.3 Nodal or multi-zonal approach

This method is the simplest between the three "white box" approaches. The building is divided into several zones where a zone generally represents one or multiple rooms, a wall, the exterior environment, but it can also represent a specific load like the loads from occupants or electrical devices. Each zone is a node described by the same value of the physical parameters: temperature, pressure, concentration, etc. Therefore, the state variables (temperature, humidity, pollutant concentration, pressure, etc.) are assumed to be homogeneous within the zone, hence, this approach is considered to be uni-dimensional. For each node, the thermal transfer equations are solved. The nodal approach is mostly used to quantify the energy consumption of a building. It is also suitable for design and system sizing since it provides a rapid solution.

A considerable number of software tools employing the nodal approach is at the users' disposal like EnergyPlus [Crawley et al., 2001], TRNSYS [Trnsys, 2000], IDA-ICE [IDA-ICE, 2014], Pleiades+Comfie [Peuportier and Sommereux, 1990], ESP-r [Clarke and McLean, 1988], etc. These tools are particularly dedicated to the modeling of the energy consumption (heating and cooling), estimating the thermal comfort of occupants but also for verifying the regulatory requirements in terms of the building energy performance.

EnergyPlus was used in [Shabunko et al., 2016] to benchmark the energy performance of 400 residential buildings. The EnergyPlus models calculated the energy use per year for each of these buildings and the results were verified using the power consumption data. [Recht et al., 2014] presented a methodology to study the reliability of Comfie when a passive house is considered. The software was coupled to the statistical software R. The process included a sensibility study and an uncertainty propagation in order to quantify the simulation errors related to the uncertainty of the input parameters. In other types of work, BES tools were used to model the thermal behavior and energy consumption of buildings containing Phase Change Materials (PCM) in their walls. For example, [Sang et al., 2017] investigated the effect of integrating PCM in the construction of a test room on the thermal comfort, the temperature fluctuation and the space heating. They used EnergyPlus and the results were compared to experimental measurements. On the other hand, [Al-Saadi and Zhai, 2015] used TRNSYS for simulating PCM-enhanced walls in order to help design engineers and architects to select effective PCM thermal properties to achieve maximum savings in heating and cooling loads. The model was then validated using experimental data from literature. Another application field for BES tools is to study the impact of different kinds of facades on the energy performance of buildings. For example, [Dahanayake and Chow, 2017] studied the impact of VGS (Vertical Greenery systems) on the energy performance of buildings. VGS represent the concept of integrating vegetation into vertical facades of buildings that in addition to aesthetic benefits, provide diversified benefits including air quality improvement, minimizing the heat island effect, improvement of the

building thermal performance [Kenai et al., 2018], noise reduction by absorption and additional supply of oxygen. The study proposed a mathematical model for the VGS that was integrated into EnergyPlus. In [Andelkovic et al., 2016], the authors studied the reliability of EnergyPlus for modeling double skin facades. The study was carried out on a multi-story building in Belgrade with naturally ventilated double skin facade. The simulation results for the cavity air temperature and velocity were compared with the data of a long term experimental analysis of the building. [Hilliaho et al., 2015] examined the suitability of IDA-ICE software tool for the glazed space energy simulation in theory and practice. They analyzed how the program performs in predicting the temperature and energy consumption of flats with highly glazed space and compared the results to the actual field measurement case. In another context, a mixed-mode predictive control system was proposed by [Zhao et al., 2016] by coupling Matlab [MATLAB, 1998] and EnergyPlus in order to optimize HVAC energy consumption while meeting the individual thermal comfort preference. In a recent work, [Nageler et al., 2018] proposed a detailed comparison of 1D building energy simulation tools with 3D CFD simulation and measurement data from a test bed (Test-Box), exposed to natural environmental conditions. They selected Dymola [Dymola and Catia, 2015], EnergyPlus, IDA ICE and TRNSYS for comparison against highly accurate real measurement data and dynamic CFD results. They used the same boundary conditions for all tools and compared the prediction accuracy of the air temperature against each other and against measured data.

As mentioned in Sect. 1.2.1, BES tools are frequently coupled with CFD for the thermal modeling of indoor environments. In a recent work, [Tian et al., 2018] presented a review of the motivations, methods and applications linking the CFD and BES tools. This coupling takes the advantages of both techniques: the BES performs rapid calculations related to thermal load predictions that can provide boundary conditions to CFD. CFD, on the other hand, provides information to BES related to local airflow in order to study the HVAC control and improve load calculation. For example, [Zhai et al., 2002] coupled a BES tool to a CFD software to predict the cooling or heating demand in an office. The authors used EnergyPlus to calculate the heating or cooling demand, and they used MIT-CFD to simulate the air flow and find the temperature distribution inside an office in Boston. They implemented a quasi-dynamic coupling strategy for which, the BES part first produces a set of surface temperatures and a heating load at the first hour and passes them to the CFD part. Based on these boundary conditions, the CFD part calculates the flow and temperature distributions for the next time step. Then, the BES part uses the room air temperatures and the convective heat transfer coefficients from the CFD results for the next hour running, and so on. In a similar work, [Pan et al., 2010] coupled EnergyPlus and Fluent to estimate the cooling load of an atrium in hot and humid regions. In the same way as before, EnergyPlus provides Fluent with the boundary conditions, and Fluent provides EnergyPlus with the air temperature distribution in order to calculate the right cooling load.

Another nodal approach consists in using a thermodynamic equivalent circuit to model the thermal behavior of buildings. This type of modeling is used in some software tools libraries like the ones developed for Dymola. Dymola [Dymola and Catia, 2015] is a modeling and simulation environment based on the open-source equation-based and object-oriented modeling language Modelica. Some libraries containing dynamic simulation models for building energy and control

systems were developed for it [Burhenne et al., 2013; Wetter et al., 2015; Videla and Lie, 2006]. Modeling the building using a thermal network is very popular because of its simplicity and low computational cost [Crabb et al., 1987; Boyer et al., 1996]. Thermal networks provide an alternative physics-based modeling technique which uses the analogy with electrical circuits by writing the model as a Resistance Capacitance (RC) network [Crabb et al., 1987]. The electric resistances and capacitances are equivalent to thermal ones, the electric potential is equivalent to the temperature gradient and the electric current is equivalent to the heat flux. This basic physical approach is simpler when compared to complex detailed thermodynamic equations, thus having much lower computational cost but enough accuracy and precision. First, the model structure as an RC network is determined using physical insight and afterwards, the parameter values are obtained from the thermal properties of construction materials. However, for existing buildings, an accurate and detailed description is not always available. Errors due to the aging of materials or poor knowledge of the construction characteristics might lead to an inappropriate estimation of the parameters, the values of R and C being dependent on material properties. In this case, these parameters are most commonly identified from input-output data of the building which leads us to the "grey box" modeling approach as will be explained in Sect. 1.4. We recall here some of the works done using the electric analogy as a "white box" model, i.e., without using data for estimating the parameters. [Goyal and Barooah, 2012] proposed to predict the temperature and humidity in a multi-zone building by modeling the building using the electric analogy. The goal was to propose a low order model that would be suitable for HVAC control and monitoring. [Fraisie et al., 2002] studied how to transform a multi-layer wall into a three resistances and four capacities model (3R4C). They also presented a global analogical model of a building by connecting a water loop model (1R2C) with two wall models (3R4C), which they integrated into the simulation software TRNSYS.

1.2.4 Summary and discussion on the physical models

This section presented the three main physics based approaches used in the literature of building simulations. The CFD approach is the most complete (3D) and therefore the most complex approach because it uses less assumptions than the other two approaches. In building modeling, it is especially dedicated to the description of airflow and the distribution of temperature. Its main disadvantage is its significant computation time and its complexity due to the fact that a detailed description of the building and a fine mesh are required. It also requires a high level of expertise and knowledge of fluid dynamics equations and its results are sensitive to boundary conditions. Another limitation is that it is not possible to use CFD simulations for energy and loads prediction. The zonal approach is a simplification of the CFD approach (2D), it is also dedicated in particular to the description of airflow and distribution of temperature and it provides detailed investigation of the thermal comfort, indoor air quality and energy analysis. It provides an improvement over the well-mixed assumption used for nodal models and it gives faster results than CFD models. Another advantage is that it can be easily coupled with other BES tools. The limitation of the zonal model lies in its lack of capability in modeling turbulent flows therefore one has to go back to CFD simulations. It also keeps a certain level of complexity and therefore, the user must have a minimum knowledge about fluid mechanics. The last approach is the nodal

or multi-zonal approach. It is based on the assumption that indoor air variables are uniform in each zone (1D). The advantage of the nodal approach is its ability to describe the behavior of a multiple zone building on a large time scale with a small computation time. It is particularly adapted to the evaluation of the energy consumption of buildings and the description of the evolution of an averaged temperature in a zone. However, this approach is unable to study the air quality inside a zone, it is inaccurate for large volume rooms where thermal stratification takes place and cannot consider the local effects of heat load or pollutant. To solve these problems, the nodal approach is commonly coupled with the CFD approach. Thermal networks are a specific way to model the thermal behavior of buildings using the nodal approach. The advantages of this approach are its low computation time and the limited number of parameters that should be defined by the user.

Each one of these approaches has its own advantages and limitations. The choice of the type to be used highly depends on the goal behind the application and the expected results. It also depends on the complexity of the phenomena observed, the parameters investigated, and the degree of required accuracy. "White box" models respect the physical aspect of the phenomena occurring in the building, and thus, lead to physically significant parameters. They outperform "grey box" and "black box" models at the design stage as they do not rely on measured data. However, in order to apply these methods, a deep level of detail about building geometry and description of material properties and existing HVAC systems is required. This leads to a large number of parameters to be set, whose values are not always available and precise. Moreover, an expert is needed to build the model, calibrate it and interpret the results. For existing buildings, finding a detailed and accurate description about its construction materials is not always possible. The properties of the materials actually installed could be different than the ones reported in the literature and the builder may not have followed all the original specifications. Moreover, most of the time, discrepancies between the expected performance of the building at the design stage and the one observed during operation exist due to the fact that the occupants' behavior plays a major role in determining the energy performance of the building. This is one of the most important difficulties that exist when dealing with "white box" models. This could be solved by calibrating the model using measurements from the building, however, this requires a big number of measurements that are rather complicated and uncertain [Amayri et al., 2016; Caucheteux et al., 2013, 2016]. Therefore, to overcome the limitations related to the availability of the physical description of the building, and to offer simple models that need little expertise but maintain an acceptable accuracy in representing the thermal behavior of the building during operation, researchers have tried to find alternative approaches namely "black box" and "grey box" models.

1.3 "Black box" models

"Black box" models are based on mathematical equations derived from input-output data acquired from the system. This faculty of generating a model only from measurements implies that "black box" models do not need any physical details about the building geometry or physical properties about the construction components. They do need however a rich amount of data to accurately represent the behavior of the system [Scanu, 2017]. They are easier to implement and less time consuming

than "white box" models, that is why, the last decade has witnessed an increased interest in this type of modeling. Many data-driven techniques have been employed for modeling the thermal behavior of buildings for different purposes including, but not limited to, prediction of their energy consumption [Amiri et al., 2015], estimation of the physical parameters [Rouchier et al., 2015], fault detection [Turner et al., 2017] and development of simple but accurate temperature models that can be employed for implementing control strategies [Mustafaraj et al., 2010]. This latter is a very important aspect of these models because the HVAC systems are responsible for the biggest share of the energy consumption in a building [Pérez-Lombard et al., 2008], hence, a lot of efforts have been and are still dedicated to finding the best control strategies that guarantee the thermal comfort of occupants and at the same time, optimize the energy consumption of the building. Model Predictive Control (MPC) is a promising strategy for achieving both objectives, however it is based on a model representing the temperature behavior, and the control would be as effective as the accuracy of the model. In this context, "black box" models represent an advantage over "white box" models by offering simple, accurate and low computational time consuming models that could be used for real-time command and control.

The most common data-driven techniques for building thermal modeling and energy performance evaluation are mainly classified between System Identification (SI) and Machine Learning (ML) techniques, however, one can find other statistical and optimization methods implemented for the same purpose. In the following, some of the main techniques employed and related works are summarized.

1.3.1 System Identification via parametric models

System identification (SI) is the science of building mathematical models of dynamical systems from measured data [Ljung, 1998]. The idea is to predict a certain output variable as a combination of the input variables and an error term. We focus on parametric models, i.e., models employing a finite-dimensional parameter vector. A detailed description of different types of models used in SI, including parametric models, is given in the book [Ljung, 1998].

One of the most implemented SI techniques in building thermal modeling is linear regression. The widely used energy signature method [Fels et al., 1986] employs the simplest regression model that estimates the energy consumption of a building in terms of the ambient air temperature. Another widely used model equation representing a dynamic system in a simple form is the AutoRegressive with exogeneous inputs (ARX) equation. An ARX model aims at finding a mathematical relationship between the output of the system and its past instances (the autoregressive part that represents the dynamic aspect of the system) and the inputs (the exogeneous variables). The relationship in its simplest form for one output $y(t)$ and one input $u(t)$ is given as:

$$y(t) + a_1y(t-1) + \dots + a_{n_a}y(t-n_a) = b_{n_k}u(t-n_k) + \dots + b_{n_k+n_b-1}u(t-n_b-n_k+1) + e(t), \quad (1.1)$$

where $y(t)$ is the output at instant $t \in \mathbb{Z}$, $u(t)$ is the input at instant t and $e(t)$ is the effect of disturbances (measurement noise, etc.) assumed to have a normal distribution $\mathcal{N}(0, \sigma^2)$. n_a and n_b are fixed beforehand and they are the model orders which represent the memory of the system or the maximum relevant past terms and n_k is the pure delay between them [Ljung, 1998]. The coefficients b_i and a_i are

the parameters to be determined. This equation represents the output in terms of a single input, but it can be extended to multiple inputs and outputs. Eq. 1.1 can be rewritten as

$$y(t) = -a_1y(t-1) - \dots - a_{n_a}y(t-n_a) + b_{n_k}u(t-n_k) + \dots + b_{n_b+n_k-1}u(t-n_b-n_k+1) + e(t) \quad (1.2)$$

where the current output $y(t)$ is expressed in terms of its n_a past values and n_b past input values. The estimated output is:

$$\hat{y}(t) = -a_1y(t-1) - \dots - a_{n_a}y(t-n_a) + b_{n_k}u(t-n_k) + \dots + b_{n_b+n_k-1}u(t-n_b-n_k+1). \quad (1.3)$$

The parameters to identify can be collected into a vector

$$\theta = [-a_1 \quad \dots \quad -a_{n_a} \quad b_{n_k} \quad \dots \quad b_{n_b+n_k-1}]^\top \quad (1.4)$$

and the past values of the output and input into the regression vector

$$\varphi(t) = [y(t-1) \quad \dots \quad y(t-n_a) \quad u(t-n_k) \quad \dots \quad u(t-n_b-n_k+1)]^\top. \quad (1.5)$$

Hence, Eq. (1.2) can be rewritten as

$$y(t) = \theta^\top \varphi(t) + e(t) \quad (1.6)$$

and Eq. (1.3) as

$$\hat{y}(t) = \theta^\top \varphi(t). \quad (1.7)$$

The solution to these equations, i.e., the value of the parameter vector θ , is found by applying the least squares estimation that consists in minimizing the cost function defined as the 2-norm of the difference between the estimated and the measured output vectors, as a function of the unknown parameters θ :

$$J(\theta) = \sum_{t=1}^N (y(t) - \hat{y}(t))^2 = \sum_{t=1}^N (y(t) - \theta^\top \varphi(t))^2, \quad (1.8)$$

where N is the number of samples. The minimum for this cost function occurs when the gradient with respect to the parameter θ is equal to zero. The solution to this optimization problem is provided by the normal equation:

$$\hat{\theta} = (\Phi^\top \Phi)^{-1} \Phi^\top Y \quad (1.9)$$

where $Y = [y(1) \quad \dots \quad y(N)]^\top$ and $\Phi = [\varphi(1) \quad \dots \quad \varphi(N)]^\top$.

The state-of-the-art of building thermal simulations is rich with publications employing this kind of models. It is used for modeling and predicting the indoor air temperature and humidity, the energy consumption, physical parameter identification and fault detection [Amasyali and El-Gohary, 2018; Fouquier et al., 2013; xiang Zhao and Magoulès, 2012; Jiménez et al., 2008a; Lee et al., 1996]. The authors of [Parti and Parti, 1980] were the first to propose a new method for the prediction of the end-use appliance specific energy consumption in buildings using linear regression in 1980. The idea was to deduce monthly and yearly residential end-use consumption from the sum of several end-use consumption from household invoices added to a noise term.

[Mustafaraj et al., 2010] investigated the potential of using linear parametric models to predict room temperature and relative humidity for different time scales (30 min or 2 h ahead) for an office in London throughout the summer, autumn and winter seasons. They compared the performance of Box–Jenkins (BJ), AutoRegressive with eXternal inputs (ARX), AutoRegressive Moving Average with eXternal inputs (ARMAX) and Output Error (OE) models in identifying the thermal behavior of the office. The goal is to use these models for improving the performance of the thermal environment control system. External and internal climate data, recorded over the summer, autumn and winter seasons, were used to build and validate the models. Similarly, [Ríos-Moreno et al., 2007] investigated to what extent linear ARX and ARMAX could be used in order to predict the interior air temperature of a building. The results are shown for classrooms of a university in México. Outside air temperature, global solar radiation flux, outside air relative humidity and air velocity were used as the input variables. The measures were done by sampling every 5 min during a period of 36 days; the first two thirds of the data record were used in order to determine the model coefficients and the remaining data for validation. Many other works can be classified in this context, i.e., predicting indoor air temperature for control purposes. [Freire et al., 2008] presented a methodology for obtaining linear regression equations that represent the indoor air temperature and relative humidity for two building models. The collected data was obtained from simulations performed with a building simulation tool. [Royer et al., 2014] developed a procedure for modeling a building and its thermal zones using system identification techniques. The goal is to build a scalable and reusable model that can model the indoor air temperature on the zone as well as the building level and that can be applied on different buildings. The authors chose a state space model structure because it is more suitable for modeling MIMO (Multiple Inputs Multiple Outputs) systems with the outputs being the air temperature in the considered thermal zone and the inputs being the outdoor temperature, direct normal solar radiation and HVAC power in the considered thermal zone.

In other works, researchers used system identification to assess the energy consumption in buildings. One of the earliest well known methods that used linear regression for estimating energy consumption is the Princeton Scorekeeping Method (PRISM) [Fels et al., 1986] that has been used broadly to analyze conservation and refurbishment measures in buildings in the US. It is a two variable linear regression model that uses monthly billing data for one year from a dwelling to find a weather-adjusted Normalized Annual Consumption (NAC). Later on, [Rabl, 1988] presented the dynamic inverse models based on system identification. The author provides an overview of different methods for implementing the identification process in building thermal modeling with a particular attention focused on their physical interpretation. All the methods are tested with data from an office building. [Bauer and Scartezzini, 1998] proposed a regression method to handle both heating and cooling calculations simultaneously by dealing with internal and solar gains. In more recent papers, [Arregi and Garay, 2017; Geng et al., 2018; Mottahedi et al., 2015; Amiri et al., 2015] used regression techniques for estimating the energy performance of buildings. A review of data-driven models applied for the prediction of the energy consumption in buildings is presented by [Amasyali and El-Gohary, 2018].

Regression models are also used to determine the values of some parameters. The equivalence between the structure of physics based models, like the differential equations derived from an RC network of a building, and "black box" models,

like linear regression, makes it possible to find the physical parameters of the building from the identified parameters of the regression model. [Jiménez et al., 2008b] presented an application of the system identification toolbox IDENT in Matlab [MATLAB, 1998] to estimate the thermal properties of a wall component by comparing parametric models to the physical ones derived from the RC network equivalent of the component. In another work, [Naveros et al., 2015] used in-situ measured data and the equivalence between an ARX model and an RC network model for identifying the parameters (U-value, dynamic solar energy transmittance and effective heat capacity) of a simple homogeneous wall.

Regression models are also applied for detecting faults in the operation of a building, the faults being most commonly located on the HVAC systems level. For example, [Turner et al., 2017] introduced a data-driven automated building HVAC fault detection method that uses a recursive least-squares model approach. The authors used only the indoor and outdoor temperatures to produce the model. The parameters of the model are then observed and estimated in real time and the faults are detected when the model parameters deviate from their values computed during normal system operation. In another work, [Xiao et al., 2011] presented a fault diagnosis strategy based on a regression model and a set of centrifugal rules for centrifugal chillers. Faults are detected by the residuals which are the differences between the fault indexes calculated by the model and the actual measurements from the system, the fault indexes being characteristic quantities of the chiller.

1.3.2 Artificial Neural Networks (ANNs)

The Artificial Neural Network (ANN) is a nonlinear statistical learning technique inspired by biological neural networks. It was introduced in 1959 [Lettvin et al., 1959]. The network usually consists of an input layer, some hidden layers and an output layer. Each layer is made up of some interconnected neurons which have an activation function. Each neuron is connected to neurons of a previous layer through adaptable weights (see Fig. 1.2). The network uses a learning mode, in which an input is presented to the network along with the desired output and the weights are adjusted so that the network attempts to produce the desired output. The weights after training contain meaningful information whereas before training they are random and have no meaning.

Three types of parameters are therefore used to define ANNs: the interconnection pattern between neurons of different layers; the learning process of updating the weights of the interconnection; and the activation function that converts a neuron's weighted input to its output activation as in Fig. 1.3. A first step would be to properly choose the inputs according to the desired outputs. An initial weight w_i is chosen randomly for each input neuron. Then, the learning algorithm adjusts the weights of each neuron. Fig. 1.3 illustrates the calculation done at the neuron level. The node receives weighted activation of other nodes through its incoming connections. First, these are added up (summation). An activation function is then applied to the summation result, the outcome is the activation of the node. For each of the outgoing connections, this activation value is multiplied with the specific weight and transferred to the next node.

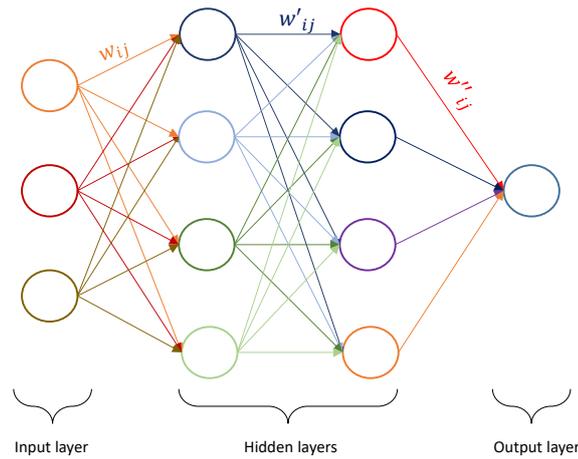


FIGURE 1.2: Graphical representation of a Neural Network

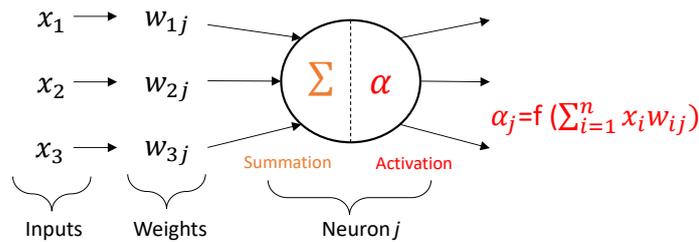


FIGURE 1.3: Graphical representation of the activation function in an ANN

During the past two decades, ANNs have been applied to predict various types of building energy use, such as the overall building energy consumption, cooling and heating loads, electricity consumption, sub-level components operation and optimization and estimation of usage parameters. A number of applications using ANNs in energy systems are presented by [Kalogirou, 2000]. In 2006, the same author [Kalogirou, 2006] did a brief review of the use of ANNs in energy applications particularly in buildings, including solar water heating systems, air flow distribution inside a room, prediction of energy consumption, indoor air temperature, HVAC system analysis, etc. In early publications, [Kalogirou and Bojic, 2000] presented an artificial neural network for the prediction of the energy consumption of a passive solar building. In this study, the training data was generated by a simulation program. The input parameters were related to: insulation (characterizing whether thermal insulation was used on all walls or not), masonry thickness, a function characterizing whether the heat transfer coefficient was constant or variable, and time of day. The output is a single value of the simulated energy consumption of the building in kWh. Various network architectures, such

as three, four and five-layers, have been investigated aiming at finding the one that could result in the best overall performance. [Olofsson et al., 1998] predicted the annual heating demand of a number of small single family buildings in the north of Sweden. As model inputs they used the difference between indoor and outdoor temperatures, a building-climate perspective obtained from a dynamic energy simulation software and a measure of inhabitant influence. Later, [Olofsson and Andersson, 2001] developed a neural network which makes long-term energy demand (the annual heating demand) predictions based on short-term (typically 2–5 weeks) measured data with a high prediction rate for single family buildings. [Mechaqrane and Zouak, 2004] used a Neural Network AutoRegressive with eXogenous input (NNARX) model to predict the indoor temperature of a residential building. They also compared the performance of this model with a linear ARX model. [Kumar et al., 2013] gave a review on energy analysis of a building using artificial neural networks. In recent works, [Khayatian et al., 2016] used an ANN for predicting heat demand indicators to validate a data set of energy certificates. The number of inputs and hidden neurons was optimized in order to achieve better accuracy. [Magalhães et al., 2017] aimed at producing tools to assess the relationship between heating energy use and indoor temperatures at different levels of occupants' behavior (in terms of where, when and at what temperature households heat their dwellings). [Deb et al., 2018] aimed to develop prediction models for HVAC related energy saving in office buildings. The modeling made use of data gathered from reports that entail building and energy consumption data for 56 office buildings in Singapore. ANN was also employed for fault detection and diagnosis in building energy systems. For example, [Lee et al., 2004] used a general regression neural network to detect and diagnose faults in a building air-handling unit and [Hou et al., 2006] developed a strategy combining a rough set approach and an ANN method to detect and diagnose sensor faults based on the past running performance data in HVAC systems.

1.3.3 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are supervised learning models used for classification and regression analysis. They are considered as robust and accurate methods in the field of data mining. The original algorithm for SVM was introduced by Vapnik in 1963 [Vapnik, 1963] while the current incarnation was published in 1995 [Cortes and Vapnik, 1995] and has been increasingly used in research and industry for various analyses such as regression, classification and nonlinear function approximation. The ability of SVMs in solving nonlinear problems made them useful for forecasting the energy consumption with high accuracy. Therefore, the past years have witnessed increased number of studies conducted on building energy analysis with a primary goal of predicting the energy consumption or temperature.

For the buildings heating/cooling load predictions, most of the SVM used is for regression. Therefore, the basic theory behind SVM for regression is presented in this section. Given a set of data points $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ($x_i, y_i \in \mathbb{R}$), SVM approximates the function of the following form

$$f(x) = \langle \omega, \phi(x) \rangle + b \quad (1.10)$$

with $\langle \cdot, \cdot \rangle$ a scalar product, $\phi(x)$ representing a high-dimensional feature space and ω and b are estimated by an optimization problem that corresponds to a minimization of the regularized risk function norm

$$\min_{\omega, b, \zeta_i, \zeta_i^*} \frac{1}{2} \|\omega\|_2 + C \frac{1}{n} \sum_{i=1}^N (\zeta_i + \zeta_i^*) \quad (1.11)$$

subject to

$$\begin{cases} y_i - \langle \omega, \phi(x_i) \rangle - b \leq \epsilon + \zeta_i \\ \langle \omega, \phi(x_i) \rangle + b - y_i \leq \epsilon + \zeta_i^*, i = 1, \dots, N \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \quad (1.12)$$

The first term ω is the regularized term and it controls the smoothness of the model. The second term is the empirical error measured by the ϵ -insensitive loss function. C is a regularization parameter that represents a tradeoff between the flatness of f and the maximal tolerated deviation larger than ϵ imposed by the user, and last, but not least, ζ_i and ζ_i^* are two slack variables allowing a flexibility in the constraints.

The basic idea of SVM for regression is to transform a nonlinear relation between x and y into a linear map. That is done by sending the nonlinear problem to a high-dimensional space called feature space and to perform a linear regression in this feature space. Therefore, a kernel function, defined as the dot product in the feature space $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ is introduced. With the introduction of kernels, all necessary computations can be performed directly in the input space, without having to compute the map $\phi(x)$. The most popular kernel functions are: linear $k(x_i, x_j) = x_i \cdot x_j$; polynomial $k(x_i, x_j) = (x_i \cdot x_j + c)^d$; Radial Basis Function (RBF) $k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$. Using different kernel functions, one can construct different learning machines with arbitrary types of decision surfaces.

[Dong et al., 2005] presented the first study using SVM for energy consumption prediction in buildings. The authors examined the feasibility and applicability of SVM in building load forecasting. They predicted the monthly electricity consumption of four randomly selected commercial buildings in Singapore. Three input features were considered: monthly mean outdoor dry-bulb temperature, relative humidity and global solar radiation. They used three years data to train the model and the derived model was applied to predict the landlord utility for one year. [Lai et al., 2008] employed SVM to forecast the energy consumption of a residential building. The data included one year and three months of daily recordings of electrical consumption and climate data. The predictions were done for the last month. In another work, [Li et al., 2009a] and [Li et al., 2009b] used SVM to predict the hourly cooling load in an office building in China and compared it with other statistical methods. [Fu et al., 2015] used SVM to predict next day electricity load of public buildings with sub-metering devices.

1.3.4 Genetic Algorithms (GAs)

The Genetic Algorithm (GA) is a stochastic optimization technique deduced from an analogy with the process of natural selection and evolution theory of Darwin. It was introduced in 1975 [Fogel et al., 1975], further developed in 1989 [Goldberg,

[2014] and has been used in building simulation since 1990.

In building simulations, GA is used to find the proper equation of a prediction model able to fit the problem. Three types of equations are frequently adopted:

- linear: $y = w_1x_1 + \dots + w_nx_n$,
- quadratic: $y = w_1x_1 + \dots + w_nx_n + w_lx_1x_2 + \dots + w_mx_1x_n + w_px_2x_3 + \dots + w_qx_{n-1}x_n + w_rx_1^2 + \dots + w_sx_n^2 + w_t$,
- exponential: $y = w_0 + w_1x_1^{w'_1} + \dots + w_nx_n^{w'_n}$,

where y is the output, (x_1, \dots, x_n) are the input variables and w_i and w'_i are real valued weights. The task of a GA is to find the best weights. It is a powerful optimization method that is able to deal with complex problems with a large number of inputs.

In building modeling, GA is used to find a prediction model of the energy consumption, for the optimization of the equipment/load demand and for finding an optimal estimate of the parameters in an energy model. [Sadeghi et al., 2011] aimed at estimating the consumption of electricity per capita in the residential sector based on economic indicators in Iran using the genetic algorithm approach. Three forms of equations (linear, quadratic and exponential equations) were developed to estimate the electricity demand, the developed models were validated with actual data and the best estimated model was selected. In another work, [Ozturk et al., 2005] studied the annual electricity consumption estimation in Turkey. They implemented two prediction models of the annual electricity consumption for both the industrial and the total Turkish demand. [Canyurt et al., 2005] developed three forms of equations to better analyze energy use and make future predictions based on GA approach. The models developed in the quadratic form are applied to Turkey. The estimation is based on the Gross Domestic Product (GDP), population, import, export, house production, cement production and basic house appliances consumption figures. [Ooka and Komamura, 2009] used GA for design purposes. They applied two genetic algorithms to propose an optimal design method for building energy systems. The goal of the method was to ensure the energy efficiency of the system and the best operational planning for heating and cooling taking into consideration important criteria like energy consumption and CO₂ emissions. The first GA was for the equipment capacity, and the second one for the operation planning. Another application of GA is parameter identification. For example, [Wang and Xu, 2006] used GA to identify the parameters of their energy model. The building energy model consisted of the simplified models of the building envelopes and the simplified model of the building internal mass. The parameters of the building envelopes are identified using available physical properties, however, the parameters of the building internal mass model are identified using monitored operation data and optimized using GA.

1.3.5 Decision trees

A decision tree is a technique to partition data into groups using a tree-like chart. It is therefore a graph consisting of a root node and a couple of branch nodes. The root nodes correspond to the input data. These root nodes are split into different

groups based on the predictor variables. The split data is then disseminated to sub-nodes as branches. The data on the sub-nodes can undergo further split or not. In a decision tree model, each non-leaf node represents one feature, each branch of the tree represents a different value for a feature, and each leaf node represents a class of prediction as shown in Fig. 1.4.

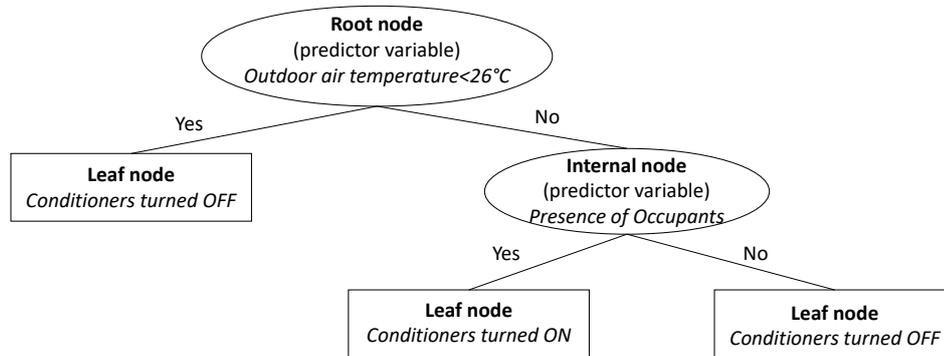


FIGURE 1.4: Graphical representation of a decision tree

Decision tree is a relatively new approach for building thermal modeling and energy consumption prediction, therefore, limited number of publications using this type of modeling is found in the state of the art. [Yu et al., 2010] developed a building energy demand predictive model based on the decision tree method. The method is applied to estimate residential building energy performance indexes by modeling building Energy Use Intensity (EUI) levels. The method provided the combination of significant factors as well as the threshold values that will lead to high building energy performance. [Hong et al., 2012] developed a decision support model for selecting the facility expected to be effective in generating energy savings and making the facility improvement program more effective. Energy consumption data were collected from 6282 elementary schools. The decision tree was used to establish groups of schools based on their energy consumption. Then, the accuracy of prediction was improved using a combination of genetic algorithms, artificial neural networks, and multiple regression analysis. The application of a prediction model after forming a cluster based on the energy consumption using decision trees improved the prediction accuracy.

1.3.6 Summary and discussion on the statistical models

This section described five data-driven methods employed in the literature for modeling the thermal behavior of buildings and predicting their energy consumption. Linear regression is frequently employed in this domain for its simplicity, prediction accuracy and low computational cost. Another advantage of linear models is that they are easier to use in control schemes than nonlinear models and need less amount of data. However the main limitation remains their linear structure which makes them unable to describe nonlinear phenomena. ANNs are also widely used

in the building community. The ability of ANNs in solving nonlinear problems made them effective for forecasting the energy consumption of buildings because of the complexity of the latter system. The ANN models used in the literature show relatively good performance but a huge limitation of this method is its lack of interpretability and the large amount of learning and relevant data required. The third technique presented was the SVM. SVM algorithms present an advantage by their ability to handle nonlinear problems with a less amount of data than ANN, however, their main limitation remains in the choice of the appropriate kernel function and their high computational time. GAs are also employed for building energy consumption prediction and for optimizing the functioning of some equipment or the value of some parameters in the model. The advantage of this method is its ability to deal with complex problems with a large number of inputs. It is also a powerful optimization method. However, a main drawback of this method is that one can never be sure that the GA has generated the optimal solution. It is also a time consuming method. The last data-driven technique presented was the decision trees. Their main advantage is their technical simplicity, they are interpretable flowchart-like tree structures that enable the user to extract useful information. In comparison to other data-driven approaches, decision trees are easy to understand, less complex and are not expensive in terms of computation time. However, their prediction accuracy is not very high and they still have limited performance when it comes to non linear problems.

For the "white box" models we saw that the choice of the software tool or technique depends on the desired output of the application itself. However, in the case of statistical techniques, this does not hold as it is hard to say which technique is more suitable for which type of application. Choosing one statistical technique over another is mainly related to the desired simplicity of the model and the richness of available data. The level of interpretability could also be a factor to consider for choosing one method over another. For instance, linear regression provides more interpretable results and a better representation of the dynamicity of the system than the other "black box" techniques. In order to overcome the physical interpretability of the model, some researchers proposed a mix between "black box" and "white box" techniques to introduce the so-called "grey box" models.

1.4 "Grey box" models

"Grey box" models were introduced as a complementary approach between physics based and data-driven models. The goal behind joining these two is to remove the drawbacks of each one of them when applied alone. These would be the lack of physical interpretation and the need of a large amount of data for "black box" models and the need of a detailed physical description of the system and the complexity found in "white box" models. A "grey box" model is therefore a model that respects the physical aspect of the system without the need of a detailed physical knowledge about it or a big amount of data.

1.4.1 Thermal networks based "grey box" models

One of the most common "grey box" approaches employed for building thermal modeling is the approach based on a representation of the building as an RC thermal

network as shown in Fig. 1.5. It uses data collected from the system to estimate the values of the parameters (resistances and capacitances) and the differential equations derived from the network to model the thermal behavior of the building¹. The interior environment is modeled by a temperature node T_i and a capacitance C_i , the heater is modeled by an input heating power ϕ_h , the solar radiation on the envelope is also represented by an input heating power $A_w\phi_s$, the building envelope is represented by a temperature node inside the envelope T_e , two resistances on each side of the node R_{ie} and R_{ea} and a capacitance C_e , the solar radiation inside the building is represented by the heat input $A_e\phi_s$ while the ambient temperature is represented by the node T_a . The differential equations describing the thermal behavior of the building are derived from the network and the parameters of the network, i.e., the resistances and capacitances, are identified using statistical techniques.

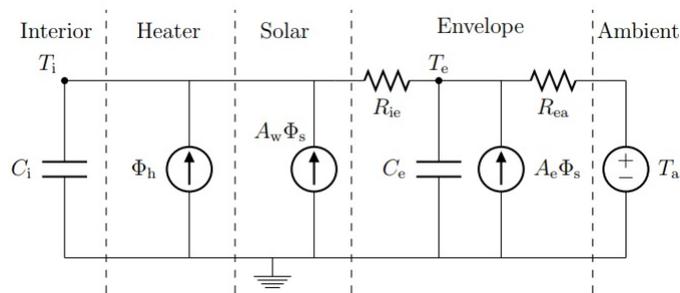


FIGURE 1.5: Example of an RC thermal network for buildings [Bacher and Madsen, 2011]

One of the earliest research work applying the analogy between the electric circuit and a thermal network for a whole building is presented by [Sonderregger, 1978]. The house is represented by a single electrical analog, composed of one capacitor and a few resistors and the parameters are fitted to recorded time histories of weather and measurements of heating load and indoor temperature in the house. In the following years, many researchers have employed this type of modeling for describing the building thermal behavior. We cite in particular the work of [Madsen and Holst, 1995; Andersen et al., 2000] in which the authors describe the continuous-time modeling of the heat dynamics of a building based on discrete-time data and statistical methods. The statistical methods are used in parameter estimation and model validation, while physical knowledge is used in forming the model structure. The suggested lumped parameter model is thus based on thermodynamics and formulated as a system of stochastic differential equations. The parameters in the continuous-time model are estimated by the maximum likelihood method where a Kalman filter is used in calculating the likelihood function. This work was extended after and has led to one of the most cited papers in this domain [Bacher and Madsen, 2011] in which the authors suggested a procedure based on "grey-box" models for the identification of suitable models for the heat dynamics of a building. They formulated a hierarchy of models of increasing complexity based on prior physical knowledge and they suggested a forward selection strategy enabling the modeler to iteratively select suitable models of increasing complexity.

¹When the R and C parameters are determined using physical properties obtained from the technical sheets of a building, the thermal network model is considered as "white box" since it relies completely on the physical knowledge of the building.

They applied their procedure on a single storey building and the performance of the models was compared using likelihood ratio tests. More works can be found from the 90's until the present day. For example, the work of [Boyer et al., 1996] shows the adaptation of the nodal analysis linked to the case of a building thermal behavior where the authors considered the case of conduction into a wall using the thermo-electric analogy. [Wang and Xu, 2006] proposed a method to simplify the building thermal model and to identify the parameters of the simplified model. For building envelopes, the model parameters can be determined using the physical details. However, since the building internal mass involves various components, it is very difficult to obtain the detailed physical properties. Therefore, the building internal mass is represented by a thermal network of lumped thermal mass and the parameters are identified using GA estimators. [Park et al., 2013] modeled a building system by a second order RC thermal network and the thermal parameters of the model were identified using different linear parametric models. [Reynders et al., 2014] proposed the identification of reduced-order grey-box models of increased complexity. They analyzed the robustness of the identified models for day-ahead predictions as well as the physical interpretation of the identified parameters. Five model types were investigated ranging from 1st to 5th order models. [Harb et al., 2016] also presented a model identification approach for forecasting the building thermal response based on grey-box models. They compared four models in their ability to forecast the building indoor temperature behavior. They found that a two-capacitor (second order) model structure with an additional consideration of the indoor air as a mass-less node (4R2C-model) enables the most accurate qualitative prediction of the indoor temperature. In a very recent work, [Brastein et al., 2018] investigated the dispersion of parameter estimates of a grey box model based on a lumped parameter thermal networks by use of randomization. They showed that a significant dispersion in the parameter estimates exists when using randomized initial conditions for a numerical optimization algorithm.

[Ferracuti et al., 2017] provided a comparison of three data-driven models for short-term thermal behavior prediction in a real building. The considered building models are: three lumped element grey-box models of first, second and third order, an ARX and a Nonlinear ARX (NARX). The models identification is performed by means of real measured data. Nevertheless the quantity and quality of the available input data, all the data-driven models show good accuracy in predicting short-term behavior of the real building both in winter and summer. Among the grey-box models, the third order one shows the best performance.

1.4.2 Other "grey box" examples

Although the thermal network model is one of the most popular "grey box" models used, different structures and approaches exist in which the authors couple BES tools with statistical techniques in order to obtain an accurate thermal model. For example, [Lü et al., 2015] presented a methodology based on a physical–statistical approach designed to account for building heterogeneity that arises from weather conditions, energy flows, occupancy, etc. Stochastic parameters are introduced into the physical model and the statistical time series model is formulated to reflect model uncertainties and individual heterogeneity in buildings. [Zhao et al., 2014] focused on learning the behavior of occupants in an office building. They developed a data mining approach using office appliance power consumption data to learn the occupants' behavior. The impact of the occupants' behavior on the energy

consumption of the HVAC system was investigated by simulating EnergyPlus models. [Qin et al., 2012] coupled CFD with a machine learning technique for the prediction of the thermal dynamic behavior in an atrium. CFD enables the prediction of indoor thermal data for buildings while an energy simulation model provides a whole building energy analysis. An ANN is used as an integrating tool to couple the energy simulation model and the CFD model. In an original work, [Macarulla et al., 2017] studied the potential of creating predictive models using a stochastic grey-box modeling approach to predict CO₂ concentrations in a room that are most commonly estimated by the differential equation of the tracer-gas mass balance. A set of stochastic differential equations are first defined and then, the model parameters are estimated using a maximum likelihood method. The approach therefore combines physical knowledge and information embedded in the monitored data to identify a suitable parametrization for the model. [Hasan et al., 2014] presented in their paper a new control approach that depends on a Monte Carlo method to generate an optimal heating plan. The approach is a minimization problem with the following constraints: the building inertia, the energy dynamic pricing and the thermal comfort of occupants. For this purpose, they used a second order state space model, for which the parameters were estimated using input-output data. The output of the developed model is the interior temperature while the inputs are the outside temperature, the energy supplied by the heating system and the solar and interior gains. The data sets used for estimating the parameters and validating the model concern historical recorded data of interior, exterior temperatures and the associated supply of energy.

1.4.3 Summary and discussion on the "grey box" approaches

As we can see from the previous examples, the advantages of "grey box" approaches are clear as they are a compromise between "white box" and "black box" models. They present the advantages of both techniques and thus remove some important drawbacks.

The lumped thermal network method is frequently employed because, from the physical point of view, it is easy to implement when compared to other physics based methods. The main limitation was in determining the right value of the R and C parameters. This was solved by combining this representation with statistical techniques using measurements from the system in order to find the most accurate values of the parameters. However, the limitation of this method remains in the fact that the structure of the network cannot be easily determined. Therefore, one has to test many orders and construct multiple thermal networks in order to find the best one, which is time consuming and also needs a certain level of expertise in the domain. The accuracy of the model will depend on the available data and the optimization technique used. Moreover, the model is represented by first order differential equations so it remains linear and probably unable to describe the existing nonlinearities, a problem attempted to be solved by having stochastic differential equations. Other types of "grey box" models exist and it is hard to have one single axe of comparison because of the big variety of methods that could be coupled for different purposes.

1.5 Conclusion

In this first chapter, a short selection of papers in the field of building thermal modeling and energy performance prediction was presented. The focus was on dynamic models to the detriment of static ones. Some of the presented articles were related to the air flow in buildings and occupants' comfort. As one can see, the field has been at the heart of the research community for almost 50 years now. A rich state of the art exists with a big variety of techniques classified into three groups: the physics based, the data based and the last one joining both aspects.

First, the physics based category was presented for which three sub-categories were defined, each category being adapted for a certain type of applications. Then, the "black-box" approach was presented where the models are developed by measuring the input and output data of the system and by fitting a mathematical function to the data. The main limitation of "black box" models is the need of a rich data set of measurements collected from the system to identify an accurate model [Scanu, 2017]. This implies that these models cannot be used at the design stage, or for deciding on the most efficient renovation strategies. The fact that they do not rely on the physical aspect of the system can be seen as an advantage in terms of the non-necessity for any information about the physical properties of the system, but it can also be seen as a limitation because of the lack of physical interpretability of their parameters. The use of this kind of models is intended to complement the use of classical "white box" models by trying to overcome some of their limitations, namely the need for a large number of precise physical parameters and their inability to accurately model the occupants' behavior which leads to discrepancies between the expected and real output. "Black box" models, being based on data collected from the system during operation phase, could be the solution for providing an accurate description of the thermal behavior of buildings taking into account the occupants' behavior. Another advantage of some data-driven techniques is the simplicity of the model compared to "white box" models which makes them more attractive for control purposes, an aspect that is very important, given that control strategies are increasingly used to ensure occupants' comfort and optimize the energy consumption of HVAC systems. A compromise between the "white box" models and the "black box" models can be found in the last modeling approach presented, the "grey box" techniques. They consist in coupling the two previous models and remove some of their drawbacks. However, a certain level of expertise will still be required in order to develop such models. All other aspects such as computation time, accuracy, physical knowledge and required data will depend on the types of techniques employed and a lot of combinations exist. The choice of this modeling technique depends on the expertise of the modeler, the desired outcome and the available data. This type of techniques seems to be a promising field for the near future, their development being highly dependent on the cooperation between the "white box" and "black box" communities.

This work focuses on "black box" models, in particular, system identification applied to hybrid systems, i.e., systems that exhibit both continuous and discrete dynamics. This is motivated by the desire of offering simple, easy to implement but also accurate models representing the dynamic aspect of the thermal behavior of existing buildings for which the physical knowledge may be lacking. The difficulties presented by classical ARX models in terms of their inability to model nonlinear phenomena are overcome by the idea of using a particular class of hybrid system

identification techniques called Piece Wise ARX (PWARX), dedicated to the modeling of nonlinear systems. The next chapter presents in more details the motivation behind using this type of modeling techniques for describing the thermal behavior of buildings.

Chapter 2

Hybrid system approach: motivation, background and applied methodology

This chapter introduces the hybrid dynamical systems approach proposed in this thesis for modeling the thermal behavior of buildings. After a general background on hybrid models, the motivation behind using this type of model for describing the thermal dynamics of a building system is given in Sect. 2.2. Sect. 2.3 presents a concise literature review on hybrid system identification while Sect. 2.4 details the PWARX algorithm applied in this thesis in its original form. Finally, Sect. 2.5 concludes this chapter.

2.1 Introduction

The first chapter recalled some of the state-of-the-art in the area of the thermal simulations for buildings and the evaluation of their energy performance. Despite the variety of options available, finding a reliable model describing the thermal behavior of a building is not an easy task considering the complex and nonlinear processes governing the building environment. Moreover, a building is characterized by the occurrence of different events like opening/closing windows or changing the heating mode. These events could greatly affect the thermal dynamics in the building and therefore, the model should take them into account. "White box" models being derived from physics laws present advantages over the other types of modeling techniques with their ability of modeling such events. However, a number of limitations that were detailed in the previous chapter, typically related to the need of a high number of precise parameters, the complexity of the model and the need of measurements to calibrate it, has pushed researchers to try to find complementary modeling approaches. On the other hand, the main limitation of other types of state-of-the-art modeling techniques is that changes in the ventilation rate or the occupancy level are most commonly disregarded or treated as disturbances using stochastic approaches [Madsen and Holst, 1995]. Moreover, the limitation of system identification techniques commonly used for building thermal modeling [Ljung, 1998] is the hypothesis of a linear model that makes them yield an incomplete description.

This thesis presents a methodology based on a nonlinear data-driven technique for modeling the thermal dynamics of buildings. This methodology uses input-output measurements: the inputs are the typical ones related to solar radiation,

heating and temperature in the surrounding environment and the output is the temperature in the building zone. No physical knowledge is needed, therefore, the proposed approach is particularly suitable for existing buildings. It is based on a switched system representation, with an easy to implement, reliable and automated model generation, capable of characterizing the continuous behavior of the temperature when discrete events affect the system. The method detects different configurations by joining data points representing the same behavior in one class and then identifying an appropriate model for it. Instead of a single model representing the dynamics of the temperature inside a building, this thesis proposes to model the building system via a set of continuous sub-models, each one of them representing certain dynamics.

This methodology takes profit of recent advances in the hybrid system identification community. Hybrid systems are heterogeneous dynamical systems that combine simultaneously continuous and discrete dynamics. For example, the air temperature evolution in a building is continuous, but could be disrupted by discrete events, after which, it goes back to being continuous. This type of systems can be represented (in a state space form or input-output form) by a set of continuous sub-models where each one represents a discrete mode [Vidal et al., 2002; Doucet et al., 2001; Bemporad and Morari, 1999a]. The idea of using hybrid models for building thermal modeling has been virtually unexplored in the literature of building dynamic thermal modeling up to now. [Luo and Ariyur, 2010] and [Lin et al., 2012] emphasized the significant effect of the convective heat transfer through an open door on the thermal dynamics and the energy savings inside a zone which motivates the use of different models for the case of an open or closed door and that is what [Lin et al., 2012] did in their paper. [Yun et al., 2012] presented a time and temperature indexed ARX model for thermal load prediction in buildings. This means that they used different sets of coefficients in the ARX predictive equation depending on the time of the day and the ambient temperature. [Paulus et al., 2015] employed change point models or piecewise linear regression models to predict the energy consumption in buildings. They proposed an algorithm to automatically select a correct model shape based solely on outdoor temperature. Five different model shapes were proposed and the correct change point model shape was determined through a series of tests. In these works, the necessity of using different models for representing different kinds of dynamics in a building is highlighted. Moreover, [Fazenda et al., 2016] investigated context-based thermodynamic modeling of buildings using an RC thermal network where the model changes according to the configuration. They used a context-dependent reduced thermal network model and many configurations were tested like changing the heating mode, the ventilation rate, etc. This work falls in the same category as ours, the general idea being the same, i.e. changing the model according to the configuration/context. However, in this previous work, the authors use a physical approach, and consider that the number of contexts (configurations) and the boundary conditions that control the transitions are observable and previously known. In our work, these transitions are detected automatically by the algorithm based solely on input-output data of the system, without any prior knowledge on the occurrence time of the transitions, or the number of discrete states. This means that anytime a new dynamic behavior of the temperature in the building is detected for one reason or another, the algorithm estimates a new model to describe the new behavior.

For this purpose, a particular class of hybrid switching models called PWARX (PieceWise Autoregressive eXogenous input) introduced in [Boukharouba, 2011] is employed in this thesis because of its suitability for this type of application, namely the fact that no previous knowledge about the number of sub-models/discrete states or events that may exist in a building is available. In the next section, this type of modeling is motivated through an equivalence with the thermal conditions in a building system in terms of the occurrence of discrete events changing the underlying physical processes.

2.2 Theoretical Foundations

This section presents the motivation behind using switching models for building thermal modeling from a physical and mathematical point of view. For this, the example of a simple building modeled using the thermal network approach is discussed. On the basis of the physical equations derived from the RC thermal model, it is shown how a configuration change could lead to a change in the dynamic behavior of the indoor temperature in terms of its model structure and parameters [Ajib et al., 2017].

2.2.1 Lumped RC thermal model for a building

We consider the case of a typical small occupied building subject to normal weather conditions. The RC model is shown in Fig. 2.1.

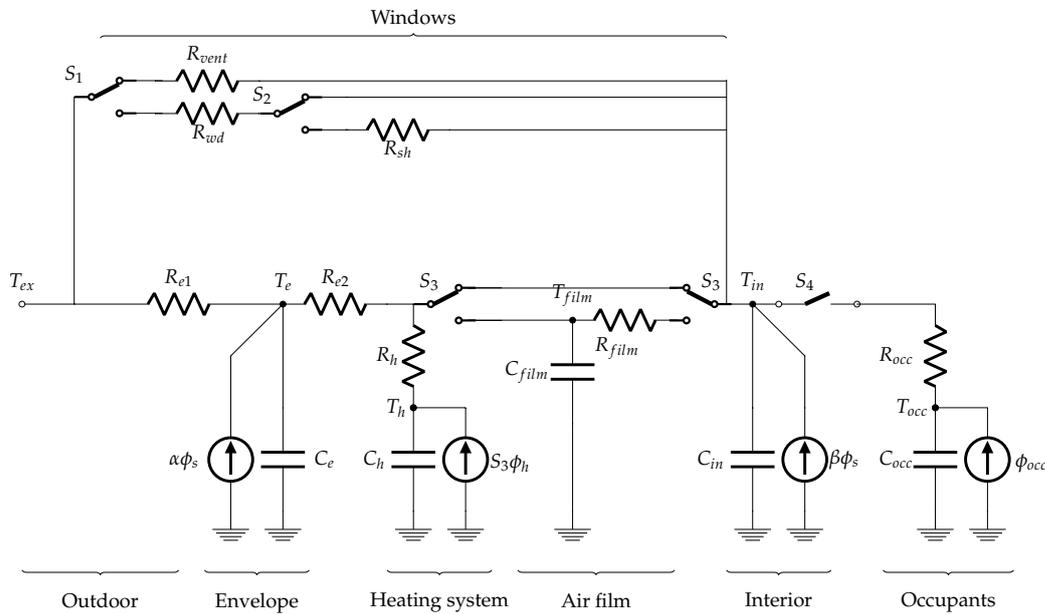


FIGURE 2.1: RC model of the building

In this RC model, one capacitance and two resistances (C_e , R_{e1} , R_{e2}) are used to model the envelope. The heating system is modeled by one capacitance (C_h), one resistance (R_h) and a heat flow input (ϕ_h). The occupants are modeled in the same

way (C_{occ} , R_{occ} , ϕ_{occ})¹. One capacitance is considered for modeling the indoor environment (C_{in}). An extra one is considered for modeling the indoor air temperature in case the heating is ON (C_{film}), as will be explained later on. Windows and shutters are considered of a low thermal capacity and are therefore represented solely by resistances (R_{wd} , R_{sh}). One resistance (R_{vent}) is used to model the thermal resistance for ventilation when the windows are open. Besides the heat flow from the heating system and the occupants, the outdoor temperature (T_{ex}), solar radiation (on the exterior walls ($\alpha\phi_s$) and transmitted to the building through the windows ($\beta\phi_s$)) are the remaining model inputs. The model output is the indoor air temperature. The configuration changes are represented via switches S_1 , S_2 , S_3 and S_4 as follows:

- Windows $\begin{cases} S_1 = 0 (\bar{S}_1 = 1), S_1 \text{ up, windows open} \\ S_1 = 1 (\bar{S}_1 = 0), S_1 \text{ down, windows closed.} \end{cases}$
- Shades $\begin{cases} S_2 = 0 (\bar{S}_2 = 1), S_2 \text{ up, shades open,} \\ S_2 = 1 (\bar{S}_2 = 0), S_2 \text{ down, shades closed.} \end{cases}$
- Heating $\begin{cases} S_3 = 0 (\bar{S}_3 = 1), S_3 \text{ up, heating OFF,} \\ S_3 = 1 (\bar{S}_3 = 0), S_3 \text{ down, heating ON.} \end{cases}$
- Occupants $\begin{cases} S_4 = 0 (\bar{S}_4 = 1), S_4 \text{ up, absence of occupants,} \\ S_4 = 1 (\bar{S}_4 = 0), S_4 \text{ down, presence of occupants.} \end{cases}$

A third, fourth or fifth order thermal network model is hence obtained depending on the position of the switches S_3 and S_4 , as shown in Fig. 2.1. This figure shows the building with its main components: the interior environment of the room (air volume, equipment and furniture), the heating system, the windows (controlling the ventilation), the envelope and the occupants. This lumped model, based on simplified physical equations, explains why different configurations occurring in a building lead to a hybrid model.

The parameters are detailed in the list of symbols. Eq. (2.1) presents the coupled differential equations derived from this model:

$$\begin{aligned}
C_e \dot{T}_e &= \frac{T_{ex} - T_e}{R_{e1}} + \bar{S}_3 \frac{T_{in} - T_e}{R_{e2}} + S_3 \frac{T_{film} - T_e}{R_{e2}} + \alpha\phi_s \\
S_3 C_{film} \dot{T}_{film} &= S_3 \frac{T_e - T_{film}}{R_{e2}} + S_3 \frac{T_h - T_{film}}{R_h} + S_3 \frac{T_{in} - T_{film}}{R_{film}} \\
C_h \dot{T}_h &= S_3 \frac{T_{film} - T_h}{R_h} + \bar{S}_3 \frac{T_{in} - T_h}{R_h} + S_3 \phi_h \\
C_{in} \dot{T}_{in} &= \frac{T_{ex} - T_{in}}{R_{eq}} + \bar{S}_3 \frac{T_e - T_{in}}{R_{e2}} + \bar{S}_3 \frac{T_h - T_{in}}{R_h} + S_3 \frac{T_{film} - T_{in}}{R_{film}} + \\
&\quad S_4 \frac{T_{occ} - T_{in}}{R_{occ}} + \beta\phi_s \\
S_4 C_{occ} \dot{T}_{occ} &= S_4 \left(\frac{T_{in} - T_{occ}}{R_{occ}} + \phi_{occ} \right)
\end{aligned} \tag{2.1}$$

with $R_{eq} = \bar{S}_1 R_{vent} + S_1 (R_{wd} + S_2 R_{sh})$.

¹In the literature, for simplification purposes, the heat gain from the occupants and the heating system are most of the time considered as input flows [Bacher and Madsen, 2011; Reynders et al., 2014]. In this work, the goal behind the proposed thermal network is to investigate the effect of different configurations on the structure and the parameters of the thermal model. Therefore, these components are not simplified by an input heat source, they are presented by their capacitance and resistance to properly analyze their effect on the model for different configurations.

These equations are rewritten in state space form as shown in Eq. (2.2):

$$\underbrace{\begin{bmatrix} \dot{T}_e \\ S_3 \dot{T}_{film} \\ \dot{T}_h \\ \dot{T}_{in} \\ S_4 \dot{T}_{occ} \end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix} -\frac{1}{C_e} \left(\frac{1}{R_{e1}} + \frac{S_3}{R_{e2}} + \frac{S_3}{R_{e2}} \right) & \frac{1}{C_e} \frac{1}{R_{e2}} & 0 & \frac{1}{C_e} \frac{S_3}{R_{e2}} & 0 \\ \frac{1}{C_{film}} \frac{S_3}{R_{e2}} & -\frac{1}{C_{film}} \left(\frac{1}{R_{e2}} + \frac{1}{R_h} + \frac{1}{R_{film}} \right) & \frac{1}{C_{film}} \frac{S_3}{R_h} & \frac{1}{C_{film}} \frac{S_3}{R_{film}} & 0 \\ 0 & \frac{1}{C_h} \frac{1}{R_h} & -\frac{1}{C_h} \left(\frac{S_3}{R_h} + \frac{S_3}{R_h} \right) & \frac{1}{C_h} \frac{S_3}{R_h} & 0 \\ \frac{1}{C_{in}} \frac{S_3}{R_{e2}} & \frac{1}{C_{in}} \frac{1}{R_{film}} & \frac{1}{C_{in}} \frac{S_3}{R_h} & -\frac{1}{C_{in}} \left(\frac{1}{R_{eq}} + \frac{S_3}{R_{e2}} + \frac{S_3}{R_h} + \frac{S_3}{R_{film}} + \frac{S_4}{R_{occ}} \right) & \frac{1}{C_{in}} \frac{1}{R_{occ}} \\ 0 & 0 & 0 & \frac{S_4}{C_{occ}} \frac{1}{R_{occ}} & -\frac{1}{C_{occ}} \frac{1}{R_{occ}} \end{bmatrix}}_A \underbrace{\begin{bmatrix} T_e \\ S_3 T_{film} \\ T_h \\ T_{in} \\ S_4 T_{occ} \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} \frac{\alpha}{C_e} & 0 & \frac{1}{C_e} \frac{1}{R_{e1}} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{S_3}{C_h} & 0 & 0 \\ \frac{\beta}{C_{in}} & 0 & \frac{1}{C_{in}} \frac{1}{R_{eq}} & 0 \\ 0 & 0 & 0 & \frac{S_4}{C_{occ}} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \phi_s \\ \phi_h \\ T_{ex} \\ \phi_{occ} \end{bmatrix}}_{u(t)}. \quad (2.2)$$

Eq. (2.2) is a state-space equation with the state vector $x(t)$ containing the temperatures of the exterior wall, the air film next to the radiator (controlled by S_3), the heater, the indoor air and the occupants' body (controlled by the presence of occupants via S_4) while the input vector $u(t)$ contains the solar radiation ϕ_s , the heat flux from the heater ϕ_h , the ambient temperature T_{ex} and the heat flux from the occupants ϕ_{occ} . State vector components are internal variables describing the dynamics of the system and evolving with time (first derivative) under the influence of the inputs. Matrix A is the state matrix and B is the input matrix. The output equation

$$y(t) = Cx(t) + Du(t) \quad (2.3)$$

is an algebraic equation expressing the measured quantity $y(t)$ as a linear combination between state and input vectors with coefficients given by the entries of the matrices C and D , respectively. In our case, the indoor temperature is the model output, hence $C=[0 \ 0 \ 0 \ 1 \ 0]$ and $D=0$, since there is no direct influence of the inputs on the outputs. The transfer function is obtained by applying the Laplace transform to both the state Eq. (2.2) and the output Eq. (2.3) and by finding the ratio output/input ($H(s)=Y(s)/U(s)$). The transfer function can be directly determined from the state-space representation as follows:

$$H(s) = C(sI - A)^{-1}B + D = \frac{P(s)}{q(s)} = \left[\frac{p_1(s)}{q(s)} \quad \dots \quad \frac{p_{n_u}(s)}{q(s)} \right], \quad (2.4)$$

where s is the Laplace variable, a complex number, n_u is the number of inputs (4 in this case). For Linear Time Invariant (LTI) systems, $H(s)$ is a rational function, hence it can be expressed as a ratio of two polynomials as in Eq. (2.4), with $p_j(s)$ and $q(s)$ being polynomials of order n_a and n_b respectively:

$$p_j(s) = \sum_{i=0}^{n_b} b_{ij} s^i, \quad j = \overline{1, n_u} \quad (2.5)$$

$$q(s) = \sum_{i=0}^{n_a} a_i s^i. \quad (2.6)$$

These polynomials define the transfer function, but the coefficients a_i and b_{ij} lack any physical interpretation. The system zeros and poles are found as the roots

of the numerator and denominator polynomials, respectively, and they offer some physical insight into the behavior of the system. The poles of the system are also the eigenvalues of the system matrix A and represent the time constants related to the system dynamics [Rabl, 1988]. In building systems, the time constant is a function of the product between the thermal resistances and capacitances ($\tau = RC$). Any change in the values of the elements of matrices A or B will change the coefficients of the transfer function (see Eq. (2.4)). Changes in A affect the calculated eigenvalues, and thereby, the values of the time constants.

2.2.2 Physical interpretation of model switches

Let us consider the two configurations induced by the switch S_3 , i.e. heating ON/OFF, and the resulting equations for each configuration, taking into account the presence of occupants ($S_4 = 1$).

Heating OFF ($S_3 = 0, \bar{S}_3 = 1$):

$$\underbrace{\begin{bmatrix} \dot{T}_e \\ \dot{T}_h \\ \dot{T}_{in} \\ \dot{T}_{occ} \end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix} -\frac{1}{C_e}(\frac{1}{R_{e1}} + \frac{1}{R_{e2}}) & 0 & \frac{1}{C_e} \frac{1}{R_{e2}} & 0 \\ 0 & -\frac{1}{C_h} \frac{1}{R_h} & \frac{1}{C_h} \frac{1}{R_h} & 0 \\ \frac{1}{C_{in}} \frac{1}{R_{e2}} & \frac{1}{C_{in}} \frac{1}{R_h} & -\frac{1}{C_{in}}(\frac{1}{R_{eq}} + \frac{1}{R_{e2}} + \frac{1}{R_h} + \frac{1}{R_{occ}}) & \frac{1}{C_{in}} \frac{1}{R_{occ}} \\ 0 & 0 & \frac{1}{C_{occ}} \frac{1}{R_{occ}} & -\frac{1}{C_{occ}} \frac{1}{R_{occ}} \end{bmatrix}}_A \underbrace{\begin{bmatrix} T_e \\ T_h \\ T_{in} \\ T_{occ} \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} \frac{\alpha}{C_e} & 0 & \frac{1}{C_e} \frac{1}{R_{e1}} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\beta}{C_{in}} & 0 & \frac{1}{C_{in}} \frac{1}{R_{eq}} & 0 \\ 0 & 0 & 0 & \frac{1}{C_{occ}} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \phi_s \\ \phi_h \\ T_{ex} \\ \phi_{occ} \end{bmatrix}}_{u(t)} \quad (2.7)$$

In this case, the heat flow is off, the temperature of the room T_{in} is considered to be homogeneous, the air film next to the radiator does not have a distinct temperature T_{film} , and therefore, the system is represented by a fourth order model as in Eq. (2.7) (the model order is given by the dimension of the system matrix A).

Heating ON ($S_3 = 1, \bar{S}_3 = 0$):

$$\underbrace{\begin{bmatrix} \dot{T}_e \\ \dot{T}_{film} \\ \dot{T}_h \\ \dot{T}_{in} \\ \dot{T}_{occ} \end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix} -\frac{1}{C_e}(\frac{1}{R_{e1}} + \frac{1}{R_{e2}}) & \frac{1}{C_e} \frac{1}{R_{e2}} & 0 & 0 & 0 \\ \frac{1}{C_{film}} \frac{1}{R_{e2}} & -\frac{1}{C_{film}}(\frac{1}{R_{e2}} + \frac{1}{R_h} + \frac{1}{R_{film}}) & \frac{1}{C_{film}} \frac{1}{R_h} & \frac{1}{C_{film}} \frac{1}{R_{film}} & 0 \\ 0 & \frac{1}{C_h} \frac{1}{R_h} & -\frac{1}{C_h} \frac{1}{R_h} & 0 & 0 \\ 0 & \frac{1}{C_{in}} \frac{1}{R_{film}} & 0 & -\frac{1}{C_{in}}(\frac{1}{R_{eq}} + \frac{1}{R_{film}} + \frac{1}{R_{occ}}) & \frac{1}{C_{in}} \frac{1}{R_{occ}} \\ 0 & 0 & 0 & \frac{1}{C_{occ}} \frac{1}{R_{occ}} & -\frac{1}{C_{occ}} \frac{1}{R_{occ}} \end{bmatrix}}_A \underbrace{\begin{bmatrix} T_e \\ T_{film} \\ T_h \\ T_{in} \\ T_{occ} \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} \frac{\alpha}{C_e} & 0 & \frac{1}{C_e} \frac{1}{R_{e1}} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{C_h} & 0 & 0 \\ \frac{\beta}{C_{in}} & 0 & \frac{1}{C_{in}} \frac{1}{R_{eq}} & 0 \\ 0 & 0 & 0 & \frac{1}{C_{occ}} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \phi_s \\ \phi_h \\ T_{ex} \\ \phi_{occ} \end{bmatrix}}_{u(t)} \quad (2.8)$$

When the heating is turned ON, the temperature of the air close to the radiator increases significantly compared to the temperature of the air on the other side of the room. This high temperature air film increases the heat exchange with the

surrounding environment (in our case, the outside and the air on the other side of the room). We therefore no longer consider one homogeneous temperature T_{in} representing the room, instead, we add the air film temperature T_{film} and the new heat exchange coefficient with the environment R_{film} . The system in this case is represented by a fifth order model as in Eq. (2.8). The transition from one configuration (heating ON) to another (heating OFF) changes the structure and the values of the coefficients in matrices A and B . The change in the values of the elements in matrix A leads to new eigenvalues and new time constants. Changes in matrix B will change the numerators of the transfer function (2.5) and, consequently, the system zeros. Therefore, switching the heating ON/OFF changes the dynamical behavior of the indoor air temperature and implies a new representative dynamic hybrid model.

In the same way we demonstrate that the presence/absence of occupants, induced by toggling the switch S_4 , leads to a change in the model structure and the parameters in the matrices. Opening/closing windows or shades leads also to a switched system. In fact, the equivalent resistance R_{eq} changes if we open/close the windows or the shades. In case of open windows ($S_1 = 0$), we have $R_{eq} = R_{vent}$ ². However, if the windows are closed and the shades are open ($S_1 = 1, S_2 = 0$), then $R_{eq} = R_{wd}$ (windows resistance), and in the case of closed shades ($S_1 = 1, S_2 = 1$), $R_{eq} = R_{wd} + R_{sh}$ (windows and shades resistances). In this case, the value of R_{eq} in matrices A and B changes with respect to the configuration, implying also a change in the dynamics of the temperature.

This section corresponds to the first contribution of this thesis as it showed why hybrid or switching models are appropriate to represent the dynamics of the temperature in the building. The above physical and mathematical analysis showed that the events that occur in a building like opening/closing the windows and shades, turning the heating ON/OFF or the presence of occupants have a direct effect on the model structure and parameters. These sudden events, represented by switches, lead to different state space models and therefore local transfer functions that characterize each mode or configuration. This hybrid model will join the discrete aspect of the system related to the transition between different configurations, and the continuous aspect of the temperature evolution that lies behind each configuration. We can thus assume that the output follows at any time instant one of these sub-models which are more accurate to reproduce specific dynamics associated to a variety of discrete modes. A discrete mode corresponds to a unique local configuration and a discrete event leads to a jump between two modes. Therefore, the second contribution of this thesis will be to show how one can build such models from input-output data.

2.2.3 Equivalence between PWARX models and local transfer functions derived from the thermal network

In this section, we establish the equivalence between the interpretation given in Sect. 2.2.1 and 2.2.2 using transfer functions derived from the state space representation of the building system and PWARX models. In those sections, it was shown how the change in some of the parameters in the state and input matrices, caused by the change in the configuration of the building, lead to local transfer functions

²The ventilation resistance through the windows is much lower than the conduction heat transfer resistance.

representing different dynamics. In this thesis, we use a PWARX model, which is a collection of ARX models in input-output form and not in transfer function form, however, these two representations (input-output and transfer function) of the same system are equivalent.

First of all, the transfer function in Eq. (2.4) is represented in the continuous-time domain, the discrete-time equivalent of this function is obtained by applying the z -transform. The transfer function $H(s)$ corresponds to $H_d(z)$, where the index d denotes the discrete form and z is a complex number, through the bilinear transformation, also known as Tustin's method [Matsuno, 1984]:

$$H_d(z) = H\left(\frac{2}{T} \frac{z-1}{z+1}\right). \quad (2.9)$$

It maps points on the $j\omega$ axis ($\text{Re}[s] = 0$) in the s -plane to the unit circle ($|z| = 1$) in the z -plane [Matsuno, 1984]. The discrete-time transfer function is therefore written similarly as in Eq. (2.4):

$$H_d(z) = \left[\frac{p_1(z)}{q(z)} \quad \dots \quad \frac{p_{n_u}(z)}{q(z)} \right] \quad (2.10)$$

where

$$p_j(z) = \sum_{i=0}^{n_b} b_{ij} z^{-i}, \quad (2.11)$$

$$q(z) = \sum_{i=0}^{n_a} a_i z^{-i} \quad (2.12)$$

are different than $p_j(s)$ and $q(s)$ in (2.5) and (2.6) but, by abuse of notation, we use the same notation.

We recall now the ARX equations presented in Sect. 1.3.1, i.e., Eq. (1.2) and Eq. (1.3) and for simplicity we consider $n_k = 1$:

$$y(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b) + e(t) \quad (2.13)$$

and

$$\hat{y}(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b). \quad (2.14)$$

If we introduce

$$q(z) = 1 + a_1 z^{-1} + \dots + a_{n_a} z^{-n_a}, \quad (2.15)$$

and

$$p(z) = b_1 z^{-1} + \dots + b_{n_b} z^{-n_b}, \quad (2.16)$$

where z^{-1} is the backward shift operator such that $z^{-i} y(t) = y(t-i)$, we obtain

$$q(z)y(t) = p(z)u(t) + e(t), \quad (2.17)$$

thus, Eq. (2.13) becomes

$$y(t) = \frac{p(z)}{q(z)} u(t) + \frac{1}{q(z)} e(t). \quad (2.18)$$

Therefore, a transfer function having the same structure of polynomials constituting the discrete-time transfer function derived from a state space model (Eq. (2.10), (2.11), (2.12)) can be derived from an ARX model. The PWARX model consists of a group of different ARX models where each one corresponds to a certain configuration of the building. The parameters of the ARX models change from one mode to the other. This is equivalent to the change in the numerator and denominator parameters of the transfer function when switching from one configuration to another as explained in Sect. 2.2.2. Each ARX model is therefore equivalent to a local transfer function describing the dynamics of the system for a certain mode. Therefore, the use of PWARX is consistent with the interpretation given in Sect. 2.2.1 and 2.2.2 by finding local transfer functions that correspond to different discrete states in a building system.

2.3 Background on hybrid systems identification

Hybrid systems can be used to describe real phenomena with discontinuous or nonlinear behavior. Different methods for the identification of hybrid systems exist and they have attracted increasing attention due to their potential applications. Most efforts in this domain have been devoted to the identification of PieceWise Affine (PWA) models [Sontag, 1981]. However, other model structures can be found in the literature for representing dynamic hybrid systems like Mixed Logical dynamical (MLD) models [Bemporad and Morari, 1999b], Linear Complementarity (LC) models [Schumacher et al., 2000], Extended Linear Complementarity (ELC) models [De Schutter, 2000], Max-Min-Plus-Scaling (MMPS) models [De Schutter and Van den Boom, 2001] and Markov jump linear models [Costa et al., 2006; Do Valle Costa et al., 2012]. Given the equivalence between PWA models and some of these classes of hybrid models that was first established in 2000 by [Bemporad et al., 2000], PWA system identification techniques can be regarded as general hybrid system identification techniques that are suitable for modeling many types of switched and hybrid systems. Moreover, the universal properties of PWA maps make them attractive for nonlinear system identification via multiple linearizations at different functioning points.

This class of hybrid models has actually proven to be effective in problems involving complex nonlinear systems with large data sets. It was successfully applied in various application domains, namely, computer vision [Vidal and Ma, 2006], electromechanical and automotive systems like DC motors [Canty and O'Mahony, 2009], suspension systems [Xu et al., 2016] and wind turbines [Vařak et al., 2011], systems biology [Imura et al., 2010; Vries et al., 2009], environmental systems like open channel systems [Boukharouba et al., 2010], water tanks [Mayer et al., 2015; Joseph et al., 2015, 2017], food industry [Xin et al., 2018] etc. This type of modeling has also given relatively satisfactory results for the modeling of a greenhouse [Rajaoarisoa et al., 2012], a system that could be considered close to a building system. A more exhaustive background on PWA models, their advantages and applications can be found in [Xu and Xie, 2014].

A PWA system is a collection of affine systems sharing the same continuous state and the switching between them is indexed by a discrete-valued signal called the discrete state. PWA models are obtained by partitioning the state-input domain into a finite number of non-overlapping convex polyhedral regions and by considering

linear/affine sub-systems in each region [Paoletti et al., 2007]. The identification of PWA models from measurements is a challenging problem involving the estimation of various parameters as will be explained later on and a clustering problem to associate each data point to the corresponding sub-model [Ohlsson and Ljung, 2011].

PWA systems can be represented in state-space form and in input-output form. In state-space form, a discrete-time switched affine system is represented by the following equations:

$$\begin{aligned} x(t+1) &= A_{\sigma(t)}x(t) + B_{\sigma(t)}u(t) + f_{\sigma(t)} + \omega(t) \\ y(t) &= C_{\sigma(t)}x(t) + D_{\sigma(t)}u(t) + g_{\sigma(t)} + v(t), \end{aligned} \quad (2.19)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_u}$ and $y(t) \in \mathbb{R}^{n_y}$ are respectively the continuous state, the input and the output of the system at time $t \in \mathbb{Z}$; n , n_u and n_y being respectively the number of states, number of inputs and number of outputs. $\omega(t) \in \mathbb{R}^n$ and $v(t) \in \mathbb{R}^{n_y}$ are noise terms. $\sigma(t) \in \{1, \dots, s\}$ is the discrete state. It selects the active mode (affine model) at time t , s being the number of modes. A_i , B_i , C_i , D_i , f_i and g_i , $i = \{1, \dots, s\}$ are matrices and vectors of real values and appropriate dimensions describing each affine dynamics. Hence, model (2.19) can be seen as a collection of affine models with continuous state $x(t)$, connected by switches indexed by the discrete state $\sigma(t)$. In PWA systems [Sontag, 1981], $\sigma(t)$ is determined by a polyhedral partition of the state and input space and is given by the rule

$$\sigma(t) = i \iff \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \in \Omega_i, \quad (2.20)$$

where $\{\Omega_i\}_{i=1}^s$ are convex polyhedra defining a complete partition of the state-input domain $\Omega \subseteq \mathbb{R}^{n+n_u}$.

PieceWise affine ARX (PWARX) models represent the input-output version of PWA models. They are defined by introducing the regression vector (for a Multiple Input Multiple Output (MIMO) system)

$$\varphi(t) = [y(t-1)^\top \dots y(t-n_a)^\top u(t-n_k)^\top \dots u(t-n_b-n_k+1)^\top]^\top \quad (2.21)$$

where $u(t) \in \mathbb{R}^{n_u}$ is the input vector and $y(t) \in \mathbb{R}^{n_y}$ is the measured output of the system at time $t \in \mathbb{Z}$ (cf. Eq. (1.2-1.6)). The output $y(t)$ is expressed as

$$y(t) = \theta_{\sigma(t)}^\top \bar{\varphi}(t) + e(t) \quad (2.22)$$

with $\bar{\varphi}(t) = \begin{bmatrix} \varphi(t) \\ 1 \end{bmatrix}$ the extended regression vector to take into account the constant element of the affine model; n_a , n_b , n_k are respectively the output order, the input order (relevant past terms) and the pure delay between them. $e(t) \in \mathbb{R}^{n_y}$ is the noise term assumed to have a normal distribution $\mathcal{N}(0, \sigma^2)$. $\sigma(t) \in \{1, \dots, s\}$ is the discrete state, s is the number of modes, and $\{\theta_i\}_{i=1}^s$ is the collection of parameter vectors defining the ARX model for each mode i .

For PWARX systems, the switching is determined by a polyhedral partition of the regression domain where the sets $\{\mathfrak{R}_i\}_{i=1}^s$ form a complete partition of the regression space $\mathfrak{R} \subseteq \mathbb{R}^d$, with $d = n_y \cdot n_a + n_u \cdot (n_b + n_k - 1)$. Each region \mathfrak{R}_i is a

convex polyhedron with

$$\mathfrak{R}_i = \{\bar{\varphi} \in \mathbb{R}^{d+1} : H_i \bar{\varphi} \leq \mathbf{0}\}, \quad (2.23)$$

with H_i , a matrix of appropriate dimensions and $\mathbf{0}$, the null vector. This means that for these models the discrete state $\sigma(t)$ is given by

$$\sigma(t) = i \iff \varphi(t) \in \mathfrak{R}_i. \quad (2.24)$$

By introducing the PWA map f of the following form:

$$f(\varphi) = \begin{cases} \theta_1^\top \bar{\varphi}(t), & \text{if } \varphi(t) \in \mathfrak{R}_1 \\ \vdots \\ \theta_s^\top \bar{\varphi}(t), & \text{if } \varphi(t) \in \mathfrak{R}_s \end{cases} \quad (2.25)$$

the PWARX model ((2.22), (2.24)) takes the form of a nonlinear ARX model:

$$y(t) = f(\varphi(t)) + e(t). \quad (2.26)$$

PWARX models are a sub-class of SARX (Switched ARX) models. SARX [Vidal et al., 2002] models are generalized switching models because the transition from one sub-model to another is arbitrary while in case of PWARX the governing laws for switching from one sub-model to another depends on the regression vector (commutation law defined by a polyhedral partition of the regression space).

In general, the identification problem for PWARX models is defined as follows:

Problem 1 Given a collection of N input-output pairs $(y(t), u(t))$, $t = 1, \dots, N$, estimate the model orders n_a and n_b , the pure delay n_k , the number of modes s , the parameter vector for each sub-model $\theta_{i=1}^s$ and the regions $\mathfrak{R}_{i=1}^s$.

Remark. In the case of PWARX models, estimating the regions \mathfrak{R}_i implies the estimation of the discrete state $\sigma(t)$ according to Eq. (2.24).

The difficulty in solving this problem depends on which quantities are assumed to be known or fixed a priori. The difficulty also lies in the fact that the identification problem includes a clustering problem where each data point has to be associated to the most suitable sub-model. Many identification techniques exist in the literature in order to solve this problem [Paoletti et al., 2007]. Most of these identification techniques consider that the orders n_a and n_b and the pure delay n_k of the system are fixed while for the rest of the parameters, different assumptions are made.

In this thesis, we use the PWARX identification technique introduced by [Boukharouba et al., 2009]. This technique considers n_a , n_b and n_k as fixed inputs to the algorithm, while the rest of the parameters are unknown. Therefore, Problem 1 becomes:

Problem 2 Given a collection of N input-output pairs $(y(t), u(t))$, $t = 1, \dots, N$, the model orders n_a and n_b and the pure delay n_k , estimate the number of modes s , the parameter vector for each sub-model $\theta_{i=1}^s$ and the regions $\mathfrak{R}_{i=1}^s$.

Details on the methodology are presented in the next section.

2.4 PWARX algorithm

As different dynamic behavior occur in a building system, the novelty of the proposed approach in this thesis is to model these different configurations using a data-driven approach based on hybrid system identification [Boukharouba, 2011; Bako, 2008]. The challenges behind this modeling technique are, on the one hand, finding the discrete states based solely on input-output measurements and, on the other hand, estimating the continuous model associated to each discrete state. In our application, the proposed method detects different configurations in the building by grouping data representing the same behavior in one sub-class, and finds the proper mathematical equation in the form of an affine autoregressive model with exogeneous inputs for each sub-mode [Ajib et al., 2016]. In this section, we review the proposed PWARX method for modeling the thermal behavior of buildings from collected measurements.

For our application of modeling the indoor air temperature of a building having a variety of local configurations and nonlinear dynamics, a PWARX model is a set of sub-models where each sub-model is an ARX equation (Eq. (1.2)) representing a certain configuration or thermal behavior of the building zone. The method consists of a clustering procedure to group data points in sub-classes associated to each sub-model and an estimation problem to estimate the parameter vector for every state/configuration found. The commutation between one discrete mode and another that could correspond to a building functioning change, is based on a decision rule acquired from the regression data. For example, in Fig. 2.2, the regression data is distributed into three regions $\mathfrak{R}_{i=1,2,3}$, each region corresponding to a configuration, and a parameter vector θ_i is associated to each one of them. The borders of the regions are estimated in terms of the components of the regression vector (past inputs and outputs). Data belonging to region \mathfrak{R}_1 follows the model defined by the parameter vector θ_1 that represents a certain dynamic behavior of the temperature and so on. Afterwards, these sub-models can be used to predict the temperature depending on the configuration. For simplicity, the sub-models in this algorithm are considered to have the same structure, which could be of the highest order, in terms of the orders n_b and n_a in Eq. (1.2) and the pure delay between them n_k [Ljung, 1998]. The values of n_a , n_b and n_k are inputs to the algorithm, which should be supplied by the user based on insight.

The general structure of a PWARX model was introduced in Sect. 2.3 (cf. Eq. (2.25), (2.26)). The PWARX identification algorithm used in this work for modeling the thermal behavior was introduced by Boukharouba [Boukharouba, 2011; Boukharouba et al., 2009]. It simultaneously treats clustering of regression data and parameter estimation. The clustering of data uses a novel similarity measure based on a mixture of the Euclidean distance between data in the regression vector on one hand, and on the other hand, the error between the prediction output of the neighbouring sub-models and the true output of the system. The parameter estimation is performed using the least squares technique and the algorithm is implemented in Matlab [MATLAB, 1998].

Given N data vectors $X(i) = [\varphi(i)^\top, y(i)]^\top$, where i is the index of a data point $i = 1, \dots, N$, from measurements of the system, the approach consists of three steps:

1. Initialization,

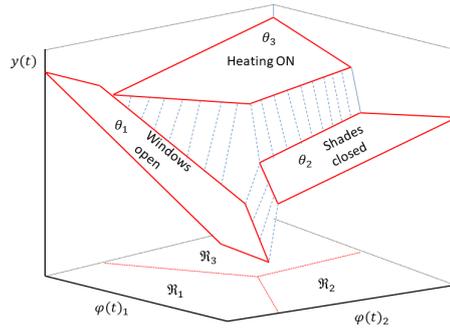


FIGURE 2.2: Example of repartition of a PWARX system with 3 sub-models

2. Data re-affectation and model estimation,
3. Convergence test.

For the initialization part, the number of sub-models is set to $s = N$, i.e. each data point is a sub-model. N clusters are obtained, $C = \{C_1, \dots, C_N\}$, where $C_i = \{X(i)\}$. To these N clusters, the initial parameter vector $\Theta^{(0)} = [\theta_1^{(0)}, \dots, \theta_N^{(0)}]$ is assigned, where $\theta_i^{(0)}$ is calculated by considering the data $X(i)$ of the sub-model C_i and its c nearest neighbors, denoted as cNN , and by using the least squares technique on these $c + 1$ data. The parameter c is set by the user.

For data re-affectation, the aim is to reduce the number of clusters by reclassifying each data point and to estimate the parameters of each sub-model. Data will migrate towards the most representative clusters according to a specific decision rule. The less representative clusters become empty and thus are eliminated. Let $\Gamma_c(X(i))$ be the set of the cNN of $X(i)$ and let $X(j) \in \Gamma_c(X(i)), j = 1, \dots, c$, be one of its neighbors that belongs to $C_p, p \in \{1, \dots, \bar{s}\}$, \bar{s} being the number of detected sub-models changing in each iteration. We introduce the following similarity measure used for the membership decision rule:

$$\phi_j^i = \exp(-\alpha_p d_j^i - \beta_p e_i^2) \quad (2.27)$$

where \exp is the exponential function,

$$d_j^i = \|X(i) - X(j)\|_2, \quad (2.28)$$

is the Euclidean distance between $X(i)$ and the neighbor $X(j)$, and

$$e_i = y(i) - \theta_p^\top \bar{\varphi}(i) \quad (2.29)$$

is the error between the output of the sub-model p and the measured output $y(i)$, as shown in Fig. 2.3. θ_p is the parameter vector associated to cluster C_p . The similarity measure has a maximum value equal to 1 if $X(i) = X(j)$ and the error between the real output $y(i)$ and the constructed output $\theta_p^\top \bar{\varphi}(i)$ is equal to zero. Lastly, α_p and β_p

The less representative classes become empty and are therefore eliminated. Parameters α_p and β_p , $p \in \{1, \dots, s\}$, as well as the number of classes s are recalculated at each iteration.

The procedure ends when the stopping criterion is achieved:

$$\| \Theta^{(r+1)} - \Theta^{(r)} \| \leq \eta, \quad (2.34)$$

r being the iteration index and η a threshold set by the user, or when the number of iterations reaches a maximum $r > r_{max}$.

Algorithm 1 PWARX

Step 1: Initialization

Set c , the number of nearest neighbors

Number of iteration $r=0$, number of classes $s = N$

Compute $\Theta^{(0)} = [\theta_1^{(0)}, \dots, \theta_N^{(0)}]$

Repeat
Step 2: Data re-affectation and model estimation

$r=r+1$

for $i=1, \dots, N$ **do**

for $j=1, \dots, c / X(j) \in \Gamma_c(X(i))$ **do**

$$\phi_j^i = \exp(-\alpha_p d_j^2 - \beta_p e_i^2)$$

$$P(X(i) \in C_p) = \frac{j/X(j) \in \{\Gamma_c(X(i) \cap C_p)\} \phi_j^i}{\sum_{j=1}^c \phi_j^i}, \quad p \in \{1, \dots, s\}$$

end for

$X(i)$ migrates towards the class that achieves the highest P

end for

Eliminate empty classes

s =number of non-empty classes

for $i=1, \dots, s$ **do**

 Compute $\theta_i^{(r)}$ using least squares technique

 Update α and β

end for

Until
Step 3: Convergence test

$$\| \Theta^{(r+1)} - \Theta^{(r)} \| \leq \eta \text{ or } r > r_{max}$$

2.5 Conclusion

In this chapter we introduced the hybrid system identification technique proposed for modeling the thermal behavior of buildings based on the idea that a building is a complex system exhibiting different dynamics induced by sudden events. The first section was dedicated to the validation of this proposal by means of physics-based equations. The analysis was based on an RC thermal network model of a typical building and showed that the events or configurations that exist in a building (like opening/closing doors or windows and changing the

heating state) change the parameters and the structure of the model representing the system dynamics. Hence, one can conclude that the thermal behavior of the building could be represented by means of different local models representing specific dynamics. The hybrid PWARX technique is perfectly adapted to this purpose as it consists in switching between a set of linear sub-models sharing the same continuous state. The method clusters data available from the system into different regions and estimates the dynamic model that best fits the data in each cluster.

The following chapter presents some application examples to showcase the performance of the PWARX technique in modeling the thermal behavior of buildings and buildings zones. Different cases are tested ranging from a small unoccupied room in an office building to a whole occupied residential building.

Chapter 3

Evaluation of the proposed modeling approach

This chapter details the results obtained for modeling the thermal behavior of buildings using the proposed PWARX technique. Three complementary test cases are presented: the first one uses real measurements from an unoccupied room subject to simple scenarios, the second one uses synthetic data generated by a software for more complex scenarios while the third one represents the most complete case as it uses real measurements collected from an occupied building. The goal behind these experiments is to evaluate the performance of a hybrid PWARX model in representing the thermal behavior of different types of buildings, under different conditions. The performance of the model is evaluated via a set of commonly used validation criteria, namely, the *Fit*, the Mean Absolute Error (*MAE*), the Root Mean Squared Error (*RMSE*) and the adjusted coefficient of determination R_{adj}^2 .

3.1 Introduction

To study the feasibility of predicting the thermal behavior of buildings using a PWARX model, several test cases were tested. For data-driven models, one has to have at their disposal two separate data sets: the training data set and the validation data set. The training data set is used to train the model, to find the parameters such that a good fit is obtained. In the case of the PWARX model employed here, the training data set is used to find the number of sub-models (number of modes or configurations) and the parameter vector for each one of them as in Eq. (2.25). The validation data set is used to investigate the accuracy of the estimated model [Aajib et al., 2018a]. It should contain the same configurations that were modeled using the estimation data set, otherwise the prediction will not be accurate enough. All data provided for each test case is hence divided into a training data set and a validation data set. A special attention should be given here to the sampling time that should be small enough to capture the dynamics of the temperature. Considering the model in Eq. (2.2), the sampling rate has to be chosen by considering the lowest RC value representing the fastest dynamics ($\tau = RC$), however, in our case, it is unknown.

The first case is for an unoccupied room located in an office building in the North of France. Data recorded over several days during winter time is used to train a PWARX model and afterwards the model is used to predict the temperature profile in the room for the next 48 hours. The configurations tested for this first case are related to the heating system (heating ON/OFF). The second test case is for a numerical virtual room created on Dymola software. In a similar way, synthetic

data over several days is used to train the model and then predict the temperature for future days, however in this case, several configurations related to the heating system, ventilation rate, shades position and occupants' presence are tested. The last test case is for a whole occupied building located in England. For this case, measurements over several years are provided. The data for 10 days of each season is used to train the model and then the predictions are done for the three next days of each season. For the prediction phase, we couple the PWARX technique to an SVM classification technique. The SVM model is used to classify new data (the validation data) in the corresponding clusters so that the right sub-model could be used to predict the temperature as will be explained in Sect. 3.6.2.

To determine the accuracy of the estimated model, its performance is evaluated via a number of validation criteria typically used in the literature. These are introduced in the third section of this chapter, but first, important points about data collection are presented. Afterwards, three sections are presented where each provides a description of a separate case study together with the applied scenarios and the estimation and validation results. At last, Sect. 3.7 is dedicated to the results analysis.

3.2 Important points about data collection

Before beginning the data collection phase, it is important to identify the output and the potential inputs of the model. The output is usually defined based on the purpose of the modeling work in a way that respects the causality of the system. In our case, the output of the thermal model is the indoor air temperature. The inputs are usually the ones related to meteorological data including the ambient air temperature, the ambient humidity, the solar radiation, and also the ones related to the heating and ventilation systems like the heat input and the ventilation air flow rate when available. A building is usually subject to a lot of disturbances like air infiltration and occupants' behavior, therefore it is a complex evolving system exhibiting various kinds of dynamics. Hence, the collected data should be rich enough to capture the various thermal behaviors that a building could be subject to. It can have both rapid and slow dynamics, the rapid ones being mainly caused by the ventilation process while slow ones are represented by the heat transfer through the envelope. Therefore, a special attention should be given to the sampling time that should be small enough to capture the fast dynamics (in the order of minutes).

3.3 Model performance evaluation criteria

The model accuracy is computed by means of the goodness of *Fit*, the Mean Absolute Error (*MAE*), the Root Mean Squared Error (*RMSE*) and the adjusted coefficient of determination R_{adj}^2 of respective equations:

$$Fit = 100. \left(1 - \frac{\sqrt{\sum_{t=1}^N (y_{est}(t) - y(t))^2}}{\sqrt{\sum_{t=1}^N (y(t) - \bar{y})^2}} \right), \quad (3.1)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_{est}(t) - y(t)|, \quad (3.2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_{est}(t) - y(t))^2}, \quad (3.3)$$

$$R_{adj}^2 = 1 - (1 - R^2) \frac{N - 1}{N - n_u - 1}, \quad (3.4)$$

with the coefficient of determination R^2 being defined as:

$$R^2 = 1 - \frac{\sum_{t=1}^N (y(t) - y_{est}(t))^2}{\sum_{t=1}^N (y(t) - \bar{y})^2}. \quad (3.5)$$

In these equations, y is the measured output, $\bar{y} = \frac{1}{N} \sum_{t=1}^N y(t)$ its mean, y_{est} is the output of the estimated model, N is the number of samples and n_u is the number of inputs. R_{adj}^2 is function of R^2 and it is computed to take into account the number of variables in the model by adding a penalty for having a large number of independent variables (n_u). For an accurate model, the Fit must be close to 100 %, the MAE and the RMSE must be close to 0 and the R_{adj}^2 must be close to 1.

3.4 Test case 1: Eco-Confort room in Douai

The first test case discussed in this thesis considers real measurements from an unoccupied room in an office building where the possible configurations are induced by turning the heating ON/OFF.

3.4.1 Description of the test case

The first case study is a 27 m^2 room located on the ground floor of an office building in Douai in the North of France. It has one wall in contact with a hallway inside the building. Two walls and the roof are in contact with other offices, while one wall with two windows with shades is in contact with the outside as shown in Fig. 3.1. An incomplete partition wall divides the room into two. Heating is provided by three electric heaters with 1000 W maximum heating power.

The room is equipped with wireless sensors measuring the value of various parameters:

- three sensors measuring the air temperature and humidity at different locations in the room,
- one sensor per heater measuring its power and energy consumption,
- one sensor on each window to indicate whether the windows are open or closed in addition to a similar sensor on the door,
- two motion sensors,
- one luminosity sensor,

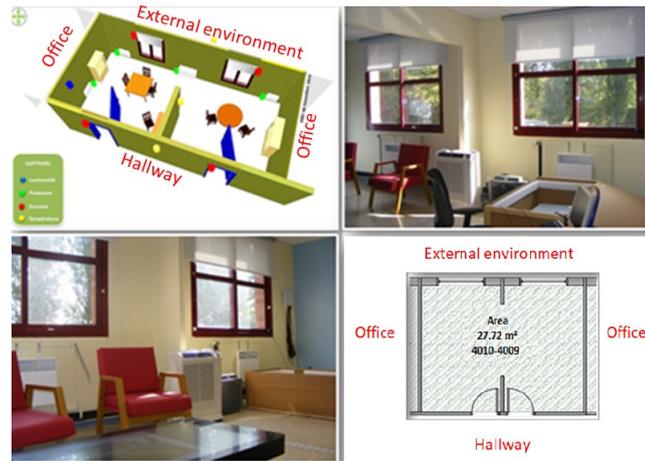


FIGURE 3.1: Eco-Confort room used to collect measurements for scenario 1

- one sensor for measuring the temperature and humidity of the ambient air outside,
- four temperature and humidity sensors are placed in adjacent offices and the hallway.

The data acquisition and command system installed in the room allows the control of the equipment (radiators, shades) and the storage of all measured data. The PWARX algorithm is applied to model and predict the temperature in the room. The output of the model is therefore the air temperature. The temperature of the room is considered to be equal to the mean of the temperatures measured by the three sensors. The inputs taken into account are: heating power, meteorological data (ambient temperature and humidity, solar radiation power) and surrounding humidity and temperature. The heating power is the sum of the power generated by the three radiators, while the surrounding temperature and humidity are considered equal to the average values of those measured in the adjacent offices and the hallway. For the solar radiation, the value measured by a small weather station located in Douai nearly 2 km away from the room is used because no such measurements are available on site.

3.4.2 Scenario 1: heating ON/OFF

We consider a simple scenario investigating the possible configurations due to turning the heating ON/OFF. The temperature in the room is set to 22°C between 06 AM and 06 PM, and 18°C during the night, for the period ranging from 14/01/2017 at 12 PM to 20/01/2017 at 09 AM, except for 19/01/2017 where the temperature was set to 26°C between 06 AM and 06 PM. The sampling time for all data is taken as 5 minutes. Using data measured in the room, adjacent rooms and meteorological data, we apply the PWARX algorithm detailed in Sect. 2.4 for modeling the temperature behavior in the room. The 5 minutes sampled data was then divided into two disjoint sets: the estimation data, considered as the first two thirds of the whole data set (from 14 to 18/01/2017), is used for the identification of PWARX models; the validation data, considered as the last third (from 18 to 20/01/2017),

is used to determine the ability of the model to predict the temperature behavior using the model found in the estimation step.

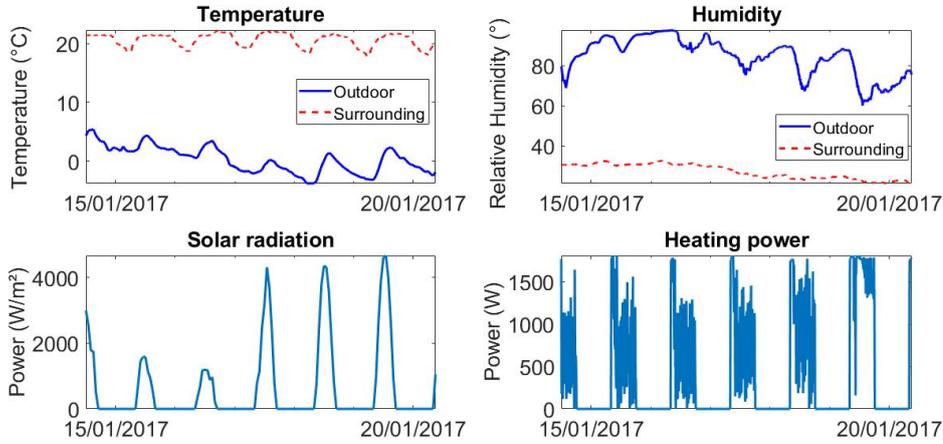


FIGURE 3.2: Model inputs for scenario 1: heating ON/OFF

Fig. 3.2 shows the inputs for the whole period. The top left figure shows the ambient and the surrounding temperature. The ambient temperature outside is negative for the last days of the scenario while the surrounding temperature is maintained at 22°C during the day and drops to 18°C during the night. The ambient relative humidity changes between 60 and 96% for the outdoors and between 24 and 32% inside the building. A low amount of solar radiation was received during the first three days, while for the last three days it goes up to nearly $300\text{ W}/\text{m}^2$. The heating power curve shows that, to attain the set point temperature of 22°C during the day, the heating is switched ON between 06 AM and 06 PM, with some short cut offs to maintain the temperature at 22°C , while for the rest of the period, to maintain the set point temperature of 18°C , the heating remains OFF.

3.4.2.1 Results for the estimation data set

When applying the PWARX algorithm to the estimation data set, the following parameters need to be specified prior to the simulation: the orders of the model: n_a , n_b , and n_k (Eq. (1.2)), and c , the number of nearest neighbors. Since these parameters are unknown, we loop over orders n_a and n_b between 1 and 8, n_k between 0 and 1 and c between 10 and 200, and retain the values that give the most accurate results. The number of nearest neighbors is the parameter that has the largest influence on the number of sub-models. This parameter should be chosen such that we attain a good compromise between the accuracy of the estimated hybrid model and the number of sub-models to keep the model as simple as possible.

The most accurate model is found for $n_a = 5$, $n_b = 4$ and $n_k = 1$. Fig. 3.3 shows the number of detected sub-models and the accuracy of the estimated model in terms of the number of nearest neighbors c . This same figure shows that the number of detected sub-models decreases in a non-linear way as the number of nearest neighbors increases and it stabilizes at 2 for $c \geq 100$. For these values of c , the *Fit* of the estimated model is quasi stable, ranging between 93 and 94%. The highest value of the *Fit* was 94.04% and it was recorded for $c = 100$. For low numbers of nearest

neighbors, the estimation was not accurate. In fact, having a low number of nearest neighbors leads to a high number of sub-models. This increases the risk of classifying some data points in the wrong sub-model which leads to a bad estimation of the temperature for these data points and decreases the Fit significantly.

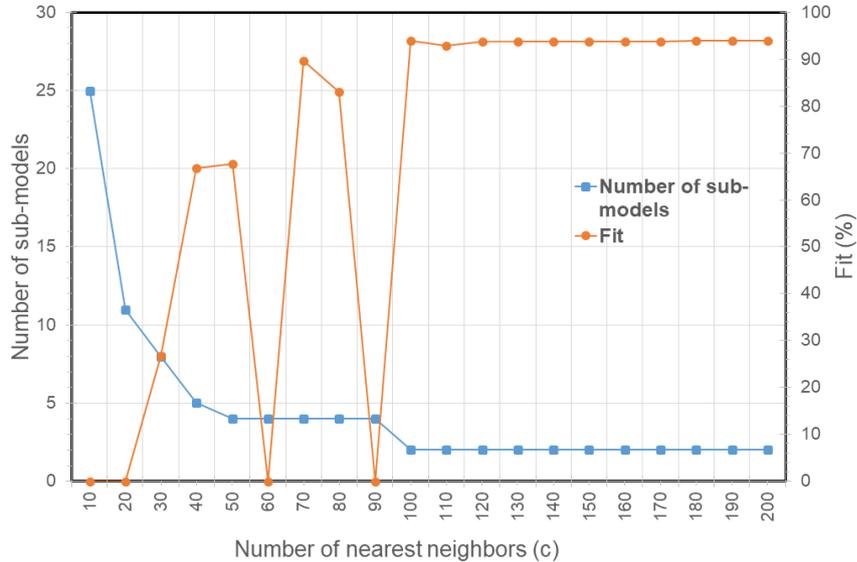


FIGURE 3.3: Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for scenario 1

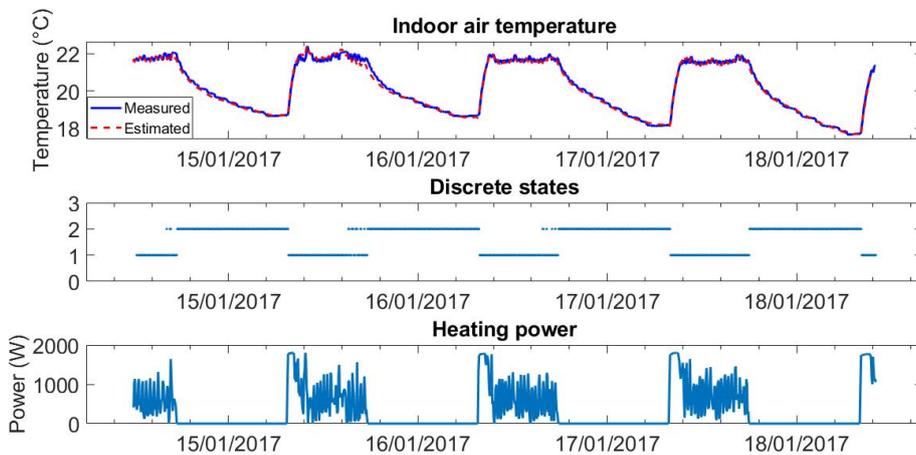


FIGURE 3.4: Estimation results for scenario 1 (heating ON/OFF). From the top to the bottom: measured and estimated temperature; detected discrete states (number of configurations); heating power input.

The best estimation result is presented in Fig. 3.4. This figure shows that, during this period, two configurations/discrete states were detected correctly in terms of the heating ON/OFF: the first one (sub-model 1) corresponds to the configuration heating ON to maintain 22°C during the day, while the second one (sub-model 2) represents the configuration heating OFF. These two configurations are induced by switch S_3 of the model presented in Sect. 2.2. In fact, the dynamics of the system

when the heating is turned ON are different than the dynamics when the heating is turned OFF for the reasons explained in the same section. The parameters of the thermal model are no other than the coefficients of the transfer function (2.18). Therefore, it is suitable to consider that they change when passing from one of these two configurations to the other, as proved in Sect. 2.2, thus generating two sub-models. For this estimation part we obtain: $Fit_{est} = 94.22\%$, $MAE_{est} = 0.06^\circ\text{C}$, $RMSE_{est} = 0.07^\circ\text{C}$ and $R_{adj,est}^2 = 0.99$ for $n_a = 5$, $n_b = 4$, $n_k = 1$ and $c = 100$.

3.4.2.2 Results for the validation data set

The models found in the estimation part are used to predict the behavior of the temperature for the last two days of the period. To do that, the model computed for discrete state 1 is used for predicting the temperature for the last two days when the heating is ON, while the model computed for discrete state 2 is used for predicting the temperature when the heating is OFF. The prediction of the indoor air temperature is hence based on the following rule:

$$\begin{aligned} \sigma(t) = 1 &\Leftrightarrow \hat{T}(t) = \theta_1^\top \bar{\varphi}(t) \quad \text{if } HP(t) > 0 \\ \sigma(t) = 2 &\Leftrightarrow \hat{T}(t) = \theta_2^\top \bar{\varphi}(t) \quad \text{if } HP(t) = 0. \end{aligned} \quad (3.6)$$

where $\sigma(t)$ is the discrete state, θ_1 and θ_2 are the parameter vectors of sub-models 1 and 2 respectively. $\hat{T}(t)$, $HP(t)$ and $\bar{\varphi}(t)$ are respectively the predicted indoor air temperature, the heat power input and the extended regression vector at time t . In this case the switching rule depends solely on the heat power input because this was clear from the estimation results. However, in more complex cases, the switching rule could be related to more than one variable of the regression vector. Validation results are shown in Fig. 3.5: a good match can be seen between the predicted and measured temperatures with a mean error of $MAE_{pred} = 0.17^\circ\text{C}$, a $Fit_{pred} = 91.16\%$, a $RMSE_{pred} = 0.2^\circ\text{C}$ and a $R_{adj,pred}^2 = 0.99$.

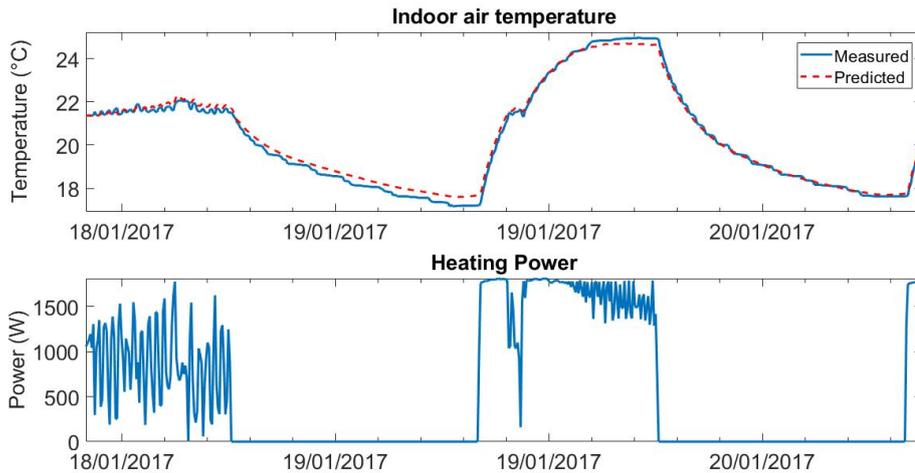


FIGURE 3.5: Prediction results over two days for scenario 1 (heating ON/OFF). Figure on top: measured and predicted temperature using PWARX. Figure in the bottom: heating power input.

Remark. The PWARX model was able to accurately predict the temperature during the last day even though the set-point temperature is 26°C instead of 22°C .

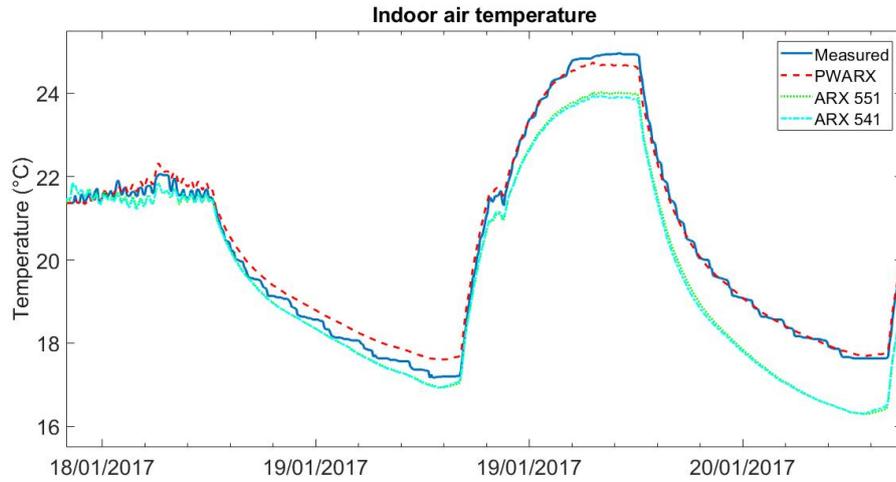


FIGURE 3.6: Prediction results for scenario 1 (heating ON/OFF) using PWARX and classical ARX with two model structures: ARX 551 that had the best Fit and ARX 541 to test the same structure as the one found for the PWARX model

Model structure (n_a, n_b, n_k)	Fit (%)	MAE (°C)	RMSE (°C)	R^2_{adj}
PWARX 541	91.16	0.17	0.21	0.99
ARX 541	65.19	0.64	0.82	0.87
ARX 551	66.5	0.62	0.78	0.88

TABLE 3.1: Comparison of validation criteria computed for prediction results of scenario 1 using PWARX and ARX models

The use of ARX models for predicting the thermal behavior of buildings is frequent in the state-of-the-art (cf. Sect. 1.3.1). Therefore, to evaluate the performance of PWARX, the prediction results are compared to those of an ARX model estimated using the system identification toolbox of Matlab [Ljung, 1998]. Different model structures are tested and the orders yielding the best prediction result are retained. For the sake of comparison, the same orders as the ones found before for PWARX ($n_a = 5$, $n_b = 4$ and $n_k = 1$) are also tested. The comparison is shown in Fig. 3.6 (ARX 551 stands for the ARX model computed for $n_a = 5$, $n_b = 5$, $n_k = 1$). The ARX models fail to properly predict the temperature in the room with a highest $Fit_{pred} = 67\%$ for $n_a = 5$, $n_b = 5$, $n_k = 1$, while the prediction Fit using PWARX was up to 91%. For an ARX model having the same orders as the ones found for the PWARX model $n_a = 5$, $n_b = 4$, $n_k = 1$, a $Fit_{pred} = 64\%$ is found. Table 3.1 summarizes these results.

3.5 Test case 2: Virtual room created with Dymola software

The second test case is for a numerical model of a room created with Dymola software using Modelica building library [Wetter et al., 2015]. The motivation behind using a software to generate a synthetic scenario is to create complementary and more complex scenarios to validate our proposed approach.

then applied during the first 5 days of February for validation. The first part of the first day for both estimation and validation data sets was not considered in order to disregard the influence of the initial conditions in Dymola.

The ambient temperature and humidity, the solar radiation and the heating power are the inputs (recall that for this virtual room, all room surfaces are in contact with the outside). Fig. 3.8 presents the weather inputs for both the estimation and validation parts. The number of occupants, the ventilation rate and the state of shades are not given as inputs to the PWARX model. The change in the temperature behavior due to these events is detected automatically using algorithm 1. The sampling time is 5 minutes. The applied configurations are presented in Table 3.2. They are also presented in the bottom plots of Fig. 3.8.

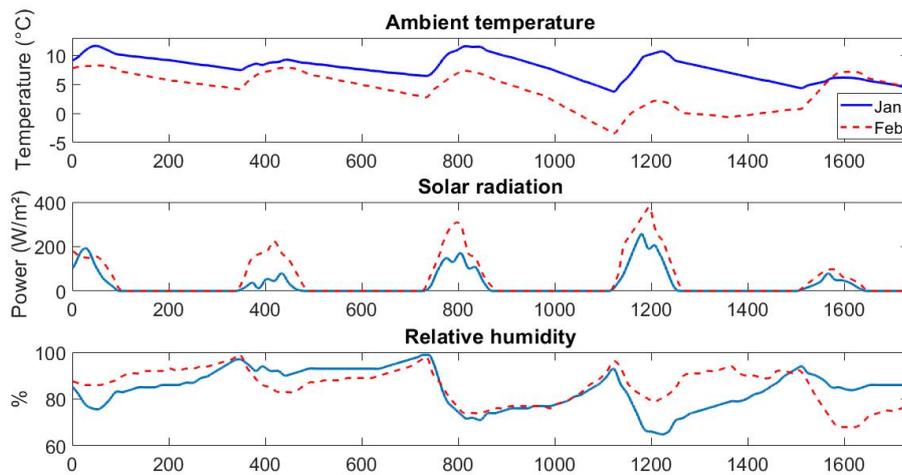


FIGURE 3.8: Weather inputs for the estimation and validation periods for scenario 2

Day	Heating (W)	Shades	Ventilation (vol/h)	Occupants
1	0	Closed	0.2	0
2	1000	Closed	0.2	0
3	0	Open (for 12 hours)	0.2	0
4	1000	Closed	1 (every 4 hours)	0
5	0	Closed	0.2	4(for 12 hours)

TABLE 3.2: Scenario 2 configurations

3.5.2.1 Results for the estimation data set

To find the best model structure, the orders of the models and the number of nearest neighbors were integrated in a loop. n_a changed between 1 and 8, n_b between 1 and n_a (because the causality of the model implies that $n_b \leq n_a$, n_k between 0 and 1 and c between 10 and 300. The most accurate model was found for $n_a = 3$, $n_b = 2$, $n_k = 1$ and $c = 100$. The number of nearest neighbors, required for the clustering phase in algorithm 1, has the biggest influence on the number of detected sub-models. Fig. 3.9 shows the number of detected sub-models and the Fit of the estimated model with respect to the number of nearest neighbors c for $n_a = 3$, $n_b = 2$ and $n_k = 1$.

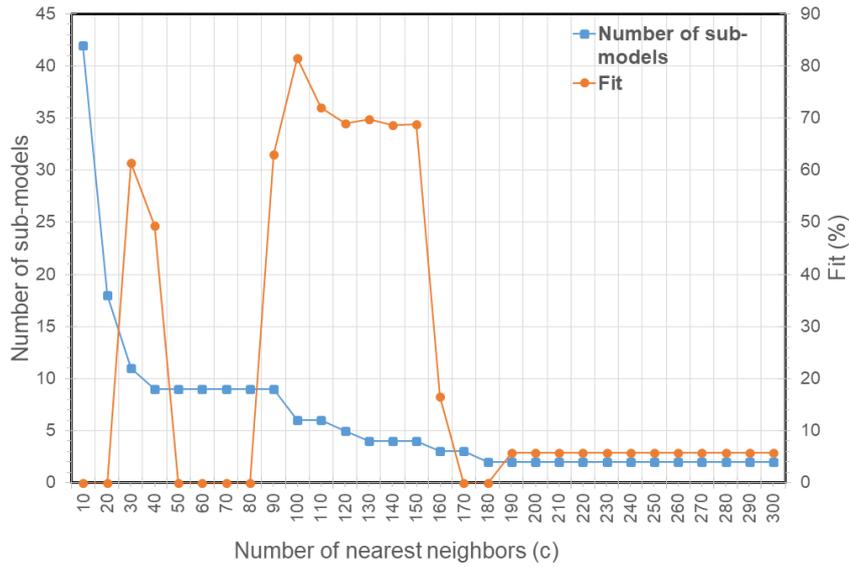


FIGURE 3.9: Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for scenario 2

As in the previous case, this figure shows that none of the two curves changes linearly with respect to the number of nearest neighbors. The number of detected sub-models decreases with respect to the number of nearest neighbors to attain a constant equal to 2 after a value of nearest neighbors equal to 180. For this scenario, an acceptable model is found for c in the range of 30-40 and 90-150, while for the rest of the cases the model does not correctly represent the behavior of the temperature. The Fit stabilizes when the number of detected sub-models stabilizes. The best Fit was found for $c = 100$. The low accuracy for a small number of nearest neighbors, i.e., high number of sub-models is due to misclassification which might lead to big errors that decrease the Fit significantly. The low accuracy that was found for high values of nearest neighbors show that a low number of sub-models is not enough to represent the temperature behavior in this scenario.

Fig. 3.10 shows the estimation results obtained for this scenario for $n_a = 3$, $n_b = 2$, $n_k=1$ and $c = 100$ with $Fit_{est} = 81.55\%$, $MAE_{est} = 0.44^\circ\text{C}$, $RMSE_{est} = 0.64^\circ\text{C}$ and $R_{adj,est}^2 = 0.96$. This same figure shows that, during the identification phase, 6 sub-models are estimated:

- For the first day, the scenario corresponds to a free evolution of the temperature (no heating, no occupants, low ventilation rate, shades closed). We can observe a switching between sub-models 2 and 3. The switch between sub-models 2 and 3 is induced by the solar radiation effect (Fig. 3.8). Therefore, when the solar radiation is at its peak ($>150 \text{ W/m}^2$), sub-model 3 is applied and, in the other case, the temperature behavior is represented by sub-model 2.
- Sub-model 5 corresponds to the second day configuration presented in Table 3.2.
- For the third day where the shades are open, the temperature behavior is once again modeled by sub-models 2 and 3 and the switching between them is due

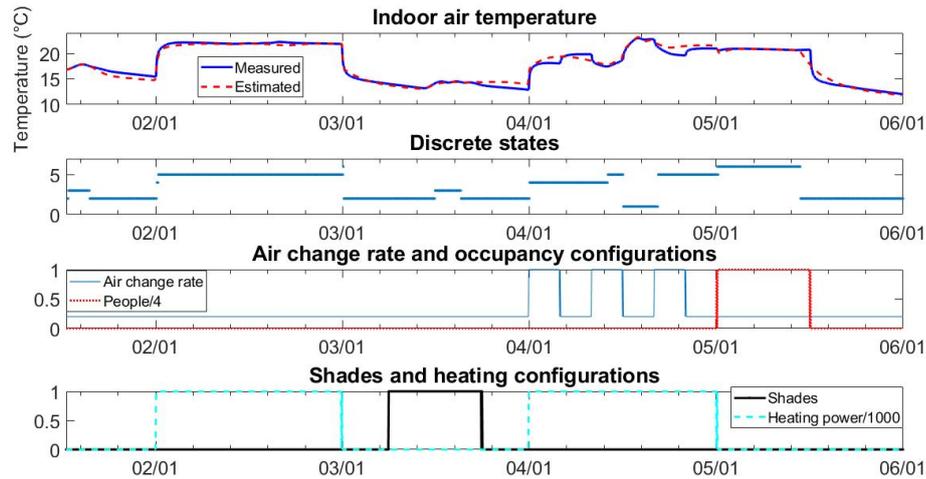


FIGURE 3.10: Estimation results for scenario 2 (heating, ventilation, shades and occupants switch). From the top to the bottom: temperatures estimated by PWARX model and obtained for the virtual room; detected discrete states; air change rate and occupancy scenarios; shades and heating scenarios.

to the solar radiation power level. As seen in the temperature plot, opening the shades did not have a big effect on the indoor air temperature and the reason behind that is the fact that the window is facing the North. Therefore, the models describing the temperature behavior are the same as the ones estimated in Day 1.

- For the fourth day the heating is turned ON again and the air change rate switches between a minimum of 0.2 vol/h and a maximum of 1 vol/h every 4 hours. For this day, sub-models 1, 4 and 5 are estimated. Sub-model 4 represents the behavior of the temperature at the beginning of the day when the ventilation rate is equal to 1 vol/h. It continues when the ventilation rate drops to 0.2 vol/h but with a less accurate estimation. Sub-model 5 represents the behavior of the temperature for rest of the day with one switch to sub-model 1 when the solar radiation is at its peak and a degraded accuracy when the ventilation rate is high as the temperature curves in Fig. 3.10 show. This result and the result for day 2 helped us conclude that sub-model 5 occurs when the heating is ON and the ventilation rate is low and sub-model 4 occurs when the heating is ON and the ventilation rate is high.
- For the fifth day, the first half is characterized by the presence of four occupants without any heating. The new behavior of the temperature is represented by sub-model 6. The rest of the day is similar to Day 1 so once again the behavior is represented by sub-model 2 with a possible switching to sub-model 3 if the solar radiation is high enough.

The correspondence between each of the identified sub-models and the tested configurations is summarized in Table 3.3. It can be noticed that an analysis should be done after the estimation part in order to understand the physical reasons behind the switching. To make the switching clear, we chose to test long configurations for this scenario.

Remark. In the case of a real occupied building a lot of configurations can appear because of the randomness of the tenants' behavior, hence, the most dominant

sub-models should be retained for the prediction phase in order to keep the hybrid model as simple as possible.

Sub-model	Solar radiation (W/m ²)	Heating (W)	Ventilation rate (vol/h)	Occupants
1	> 150	1000	0.2	0
2	< 150	0	0.2	0
3	> 150	0	0.2	0
4	< 150	1000	1	0
5	< 150	1000	0.2	0
6	< 150	0	0.2	4

TABLE 3.3: Configurations related to each sub-model estimated in scenario 2

3.5.2.2 Results for the validation data set

For validating the estimated model, the same configurations of the estimation part are applied again but this time for the first five days of February. Hence, the sub-models estimated for a certain configuration are used to predict the temperature for the same configuration. For example, sub-model 5 is applied when the heating is turned ON, sub-models 2 and 3 are applied for a free evolution of the temperature without heating, ventilation or occupants, with a switching between 2 and 3 that corresponds to the solar radiation level, etc. The switching rule is hence based on testing the values at instant t of the influencing variables as presented in Table. 3.3, these variables being known from the prediction phase. For instance,

$$\begin{aligned}
 \sigma(t) = 2 &\Leftrightarrow \hat{T}(t) = \theta_2^\top \bar{\varphi}(t) \quad \text{if} \\
 &HP(t) = 0, VR(t) = 0.2, Occ(t) = 0, SR(t) < 150W/m^2 \\
 \sigma(t) = 3 &\Leftrightarrow \hat{T}(t) = \theta_3^\top \bar{\varphi}(t) \quad \text{if} \\
 &HP(t) = 0, VR(t) = 0.2, Occ(t) = 0, SR(t) > 150W/m^2,
 \end{aligned} \tag{3.7}$$

where $\sigma(t)$ is the discrete state, θ_2 and θ_3 are the parameters vectors of sub-models 2 and 3 respectively. $\hat{T}(t)$, $HP(t)$, $VR(t)$, $Occ(t)$, $SR(t)$ and $\bar{\varphi}(t)$ are respectively the predicted indoor air temperature, the heat power input, the ventilation rate, the number of occupants, the solar radiation input and the extended regression vector at time t . The same applies for the other sub-models following the values in Table 3.3. The simulation for the whole validation period is shown in Fig. 3.11. A good Fit_{pred} of 80.39% is found with a MAE_{pred} of 0.54°C, a $RMSE_{pred}$ of 0.66°C, and a $R_{adj,pred}^2 = 0.96$.

A comparison with the prediction results of an ARX model is also provided for this test case. Fig. 3.12 shows the results obtained for both the PWARX model and classical ARX models. The best Fit_{pred} is found for an ARX model of structure $n_a = 4$, $n_b = 3$ and $n_k = 1$ with $Fit = 34\%$. For an ARX model having the same orders as the ones found for PWARX model $n_a = 3$, $n_b = 2$, $n_k = 1$ a $Fit_{pred} = 27\%$ is found. Once again, PWARX models show a better performance than classical ARX models. One can see that when more configurations came into play, the classical models became less accurate. Table 3.4 summarizes the obtained results.

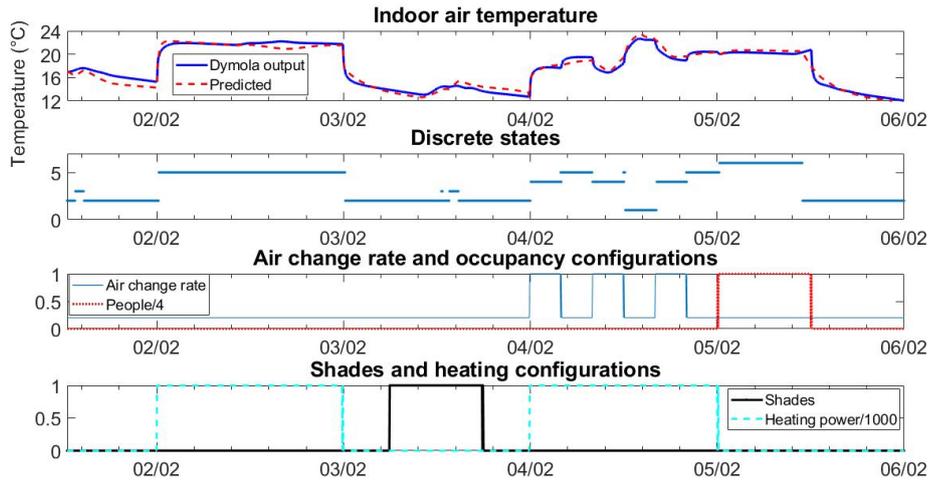


FIGURE 3.11: Prediction results for scenario 2 (heating, ventilation, shades and occupants switch). From the top to the bottom: temperatures predicted by PWARX model and obtained for the virtual room, detected discrete states; air change rate and occupancy scenarios; shades and heating scenarios.

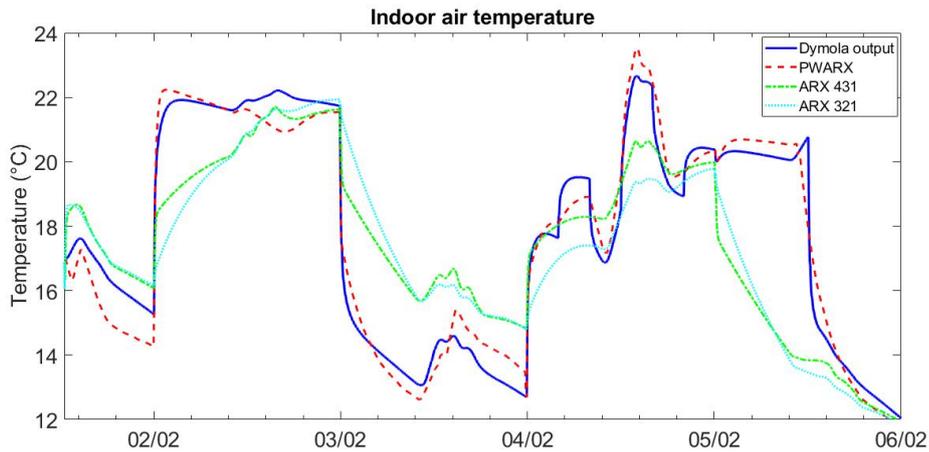


FIGURE 3.12: Prediction results for scenario 2 using a classical ARX model with two structures: ARX 321 to test the same structure as the one found for the PWARX model, ARX 431 that had the best Fit=34%.

Model structure	Fit (%)	MAE (°C)	RMSE (°C)	R_{adj}^2
PWARX 321	80.39	0.54	0.66	0.96
ARX 321	25.95	1.91	2.5	0.45
ARX 431	32.96	1.71	2.26	0.55

TABLE 3.4: Comparison of validation criteria computed for prediction results of scenario 2 using PWARX and ARX models

3.6 Occupied building

In this section, data collected from an occupied building located in the United kingdom (UK) is used for evaluating the performance of the PWARX approach in modeling the indoor air temperature of an occupied building. The data was

provided in the context of the Annex 71 meetings ¹.

3.6.1 Description of the case study

The case study is a south-facing end-terrace house built in Gainsborough, UK. It is a two-story dwelling with a total floor area of 67 m^2 [Sodagar and Starkey, 2016]. The living room, kitchen, toilet and entrance hall are located on the ground floor. Two bedrooms and a bathroom are situated on the first floor. Also, a small technical room is located on the landing of the first floor, housing the metering equipment for the PhotoVoltaic (PV) system as well as the metering for the ventilation system. The house has been monitored for 3 years, starting October 2012 until November 2015. The building was used by two adults and one child up to January 2013. In March 2013, new tenants (1 adult and 2 children) have moved in. Due to tenancy change, the house was vacant and unheated in January and February 2013.

Space heating and Domestic Hot Water (DHW) production are provided by a gas boiler. Space heating is controlled using two room thermostats, in the hall and one of the bedrooms. The gas consumption of the heating system is monitored with volumetric gas meters. Sub-meters to differentiate between space heating and DHW production were not installed. Therefore, the energy uses for domestic hot water and space heating need to be estimated based on the gas consumption data. A water meter is available that measures the total mains (cold and warm) consumption of the house. Ventilation is provided using a Mechanical Ventilation system with Heat Recovery (MVHR). The system is equipped with a monitoring system that registers temperature and relative humidity of supply and return air, as well as the electricity consumption of the ventilation unit. In addition to the electricity consumption of the MVHR, the total electricity use of the house is monitored together with the output of the PV system. Apart from the measurements of the MVHR no sub-metering has been installed for the electricity consumption. Hence, the measurements cover everything from lighting to appliances, pumps etc.

Time series data is provided for this house over the period from October 2012 until November 2015 with a 5 min interval. The dataset covers:

- CO_2 concentration (ppm) in the lounge (living room),
- On-site external air temperature ($^{\circ}\text{C}$) and relative humidity (%),
- Gas consumption meter (m^3),
- Total electricity consumption (kWh) ,
- Supply and return temperature of MVHR ($^{\circ}\text{C}$),
- Relative humidity supply and return air of MVHR (%),
- Electricity consumption MVHR (kWh),
- PV production (kWh),

¹Annex 71 is a European project on building energy performance assessment based on optimized in-situ measurements organized by the International Energy Agency (IEA) as a part of the Energy in Buildings and Communities Program (EBC).

- Temperature ($^{\circ}\text{C}$) and relative humidity (%) in the lounge,
- Temperature ($^{\circ}\text{C}$) and relative humidity (%) in bedroom 1,
- Mains water consumption (m^3).

As only the outdoor temperature and relative humidity were measured on site, weather data from a nearby weather station is included as an additional data set. Hourly averaged outdoor temperature, wind speed, wind direction and global horizontal solar radiation is collected from a weather station nearby.

One of the problems encountered in this test case was the absence of sub-metered data for the energy use of the space heating and for the production of domestic hot water. Consequently, a splitting of the gas consumption has to be performed. A split of this gas consumption was provided by the Annex organizers. The separation of the data is obtained by a deterministic method that assumes that domestic hot water and space heating cannot occur simultaneously. The assumption is based on common practice technologies for gas boilers, but could not be confirmed by the technical data available for the heating system. Given that there is no domestic hot water storage tank and that production of space heating does not coincide with domestic hot water, the method assumes that when gas use occurs on moments of water consumption, the gas is used for domestic hot water production. Clearly, this deterministic approach has some limitations:

- 5 minute sample time: for short hot water withdrawals (less than 5 min), the method will allocate the gas consumption of the entire 5 min sample to domestic hot water.
- If in winter, water consumption coincides with space heating, the gas use is always allocated to domestic hot water, even when in practice it could be space heating and cold water withdrawal.

For this test case, since data is available throughout full years, it would be interesting to model the indoor air temperature for all seasons. However, taking the 5 min sampled data over a full year leads to a very large number of samples that needs a large computation power and time. Therefore, the models estimation and validation are done using 5 min sampled data over the four seasons of the year by considering the first 10 days for each one. The estimation data set covers the periods going from 01/01/2014 to 07/01/2014 for winter; 01/04/2014 to 07/04/2014 for spring; 01/07/2014 to 07/07/2014 for summer and 01/10/2014 to 07/10/2014 for autumn. The validation is done for the three following days (08-10) of each season. This gives a total of 8064 samples for the estimation part and 3456 samples for validation.

The output of the model is the air temperature in the lounge of the house and the following inputs are considered:

- Weather data: Ambient Temperature and relative humidity, solar radiation, wind direction and speed,
- Consumption data in the house: gas consumption (for hot water and heating separated), water consumption and mains electricity consumption,
- CO_2 concentration in the lounge of the house,

- Ventilation system data: supply air temperature and humidity and the electricity consumption of the mechanical ventilation system.

Figures 3.13, 3.14 and 3.15 show the variables for the first 10 days of January 2014.

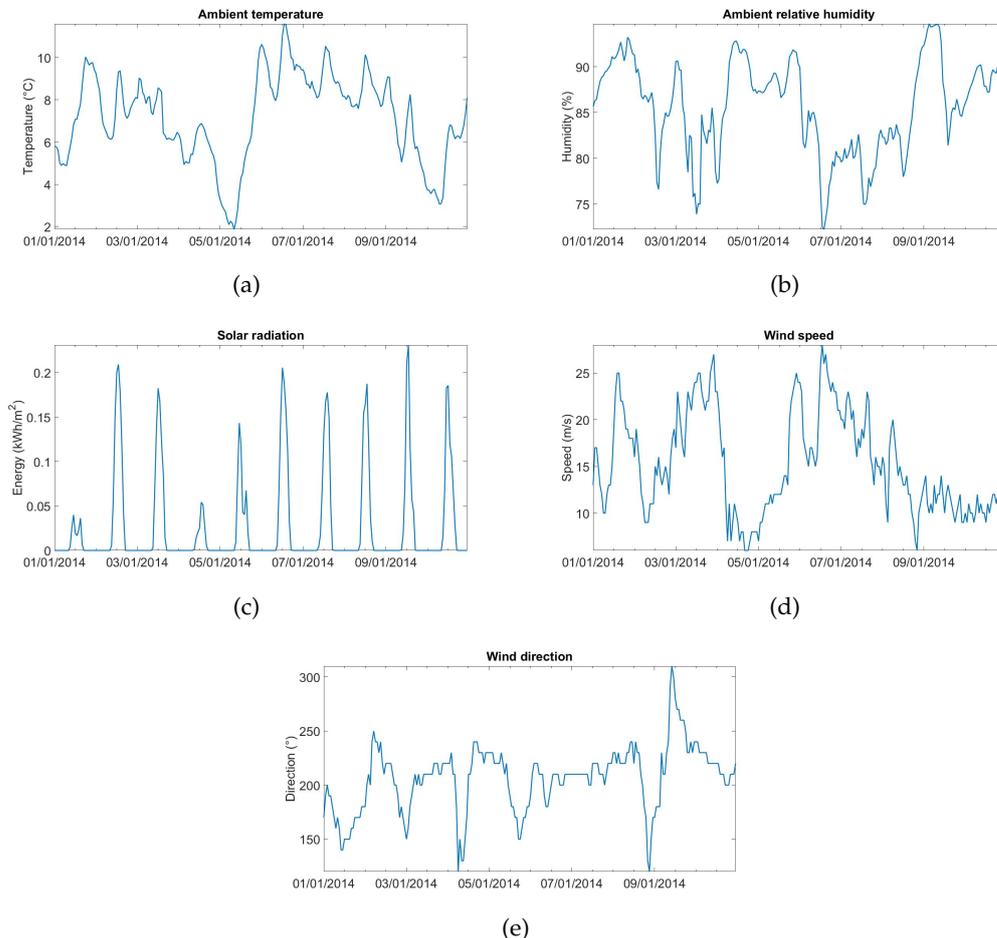


FIGURE 3.13: Weather data recorded for January 2014

3.6.2 Combining PWARX model with SVM classifiers

The key for validating the PWARX model is to identify the decision rule for switching between one discrete mode and another in order to determine the corresponding sub-model that each new data point belongs to. This section shows how to use the SVM classification technique for identifying this rule.

3.6.2.1 Need for an automated classification technique

For a PWARX model, the commutation rule is determined by a polyhedral partitioning of the regression domain where each sub-model is valid on a well-defined region \mathcal{R}_i ((2.23), (2.24)) as explained in Sect. 2.3. Estimating these regions would be the final step for the identification problem, cf. Problem 1. Recall that for the first two test cases, since simple configurations lasting a few hours each were tested, it was possible to do the link between the estimated sub-models and the variables responsible for the switching. Therefore, the regions of validation of

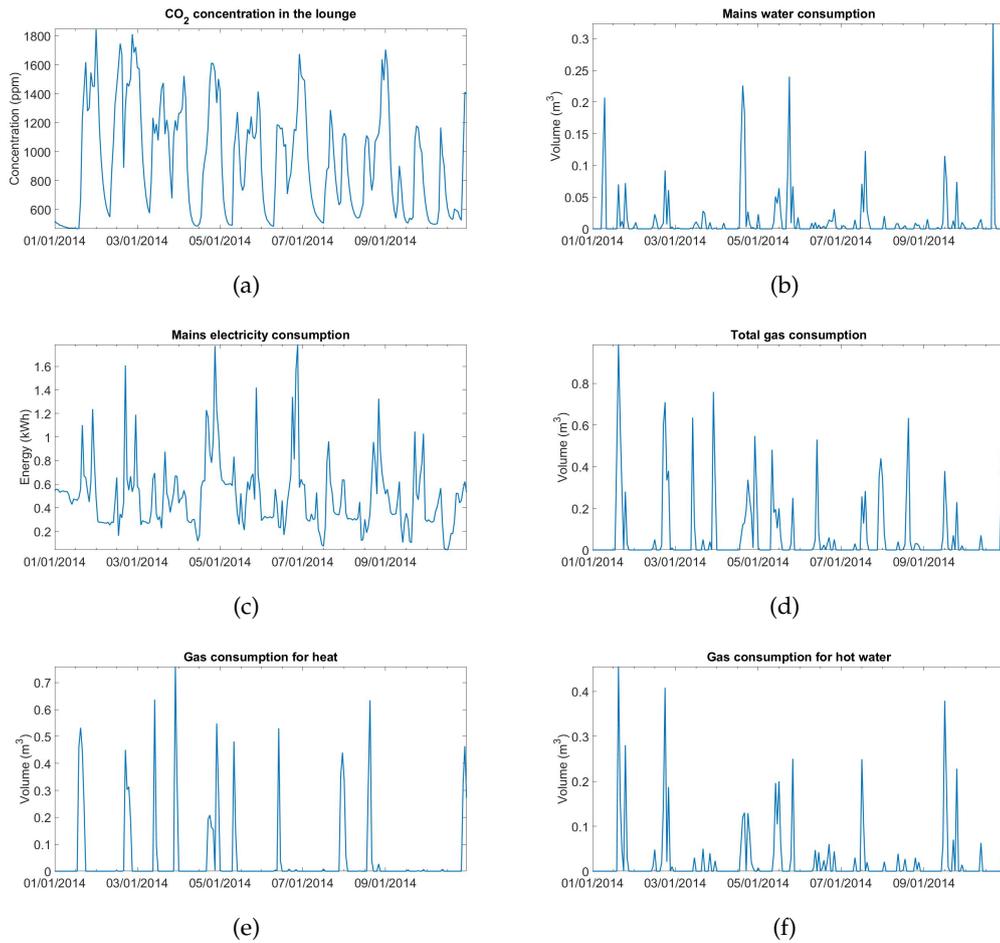


FIGURE 3.14: House data recorded for January 2014

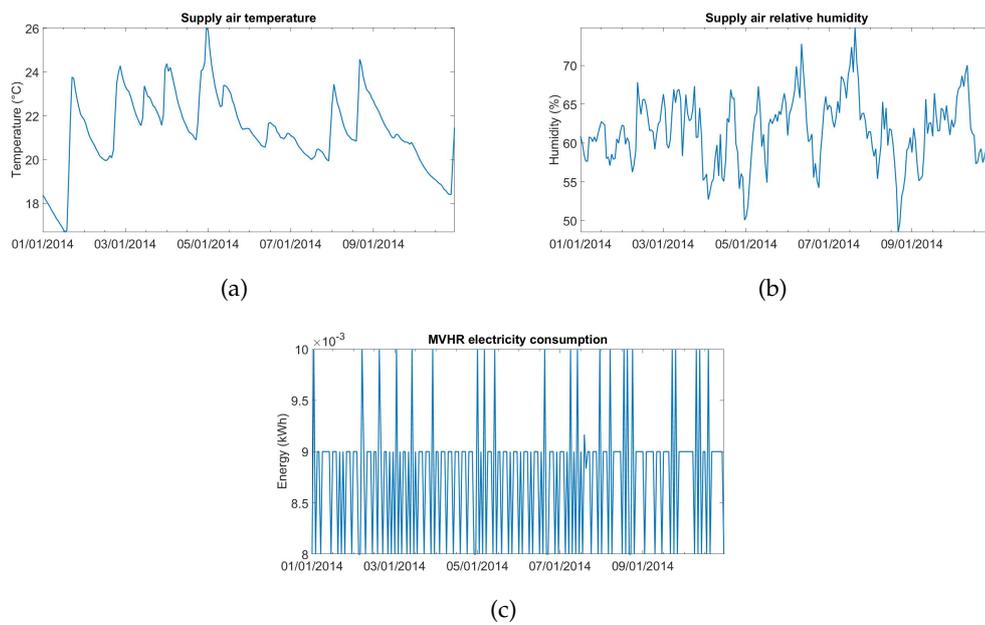


FIGURE 3.15: Ventilation system data recorded for January 2014

each sub-model were clear and the rule was extracted directly from the values of influencing variables. Moreover, the influencing variables were considered to be known for the prediction phase. For instance, in scenario 1 (Sect. 3.4.2), it was clear from the estimation results that the commutation between sub-model 1 and 2 is due to the heating mode, therefore, for the prediction, the decision rule for switching was based on the heat power input (as in Eq. (3.6)), this latter considered to be known for the whole prediction period. In scenario 2 (Sect. 3.5.2), the decision rule for switching between different sub-models became more complex because of the occurrence of a higher number of configurations and influencing variables (heat input, solar radiation, number of occupants, ventilation rate). The decision rule for this second case was also based on the values of these influencing variables at each time instant t as in (3.7), the values of these variables also supposed known for the whole prediction phase.

For realistic test cases, the PWARX model estimated from data could generate modes for which it is difficult to understand the physical configuration which generated them. This could be due to the fact that one might not be able to fully observe the operation of the building or because of having several switching signals generating a large number of configurations. For this reason, for the validation phase of this third test case, the PWARX technique (algorithm 1) is coupled with a classification technique, namely, SVM, in order to identify the decision rule for commutation between the discrete modes. Using this classification technique, the equations of the borders of the convex polyhedrals separating the regression space can be determined (see Eq. (2.23)), and therefore, new data (validation data) could be classified in the right region based on the values in the regression vector.

3.6.2.2 SVM classifiers

SVMs are supervised learning methods for classifying new data points based on the acquired labeled data used to train the classifier [Vapnik, 2013]. In other words, given labeled training data, the algorithm outputs an optimal hyperplane which separates the regions defined by the labeled data and is used for categorizing new data. In a two-dimensional space, this hyperplane is simply a line dividing the plane in two parts, each class laying on one side. For example, given the linearly separable data points shown in Fig. 3.16(a), the idea is to find the equation

$$f(X) = w^\top X + b, \quad (3.8)$$

of the line separating the blue circular points from the red triangular ones as shown in Fig. 3.16(b), with w being the weight vector, X the input vector and b the bias. The values of w and b are optimized by the SVM algorithm in order to maximize the margin around the separating hyperplane.

In some cases, the data could not be linearly separable in the original input space. Therefore, the data is mapped to a higher dimensional space called feature space, by applying what is known as the "kernel trick", after which, the separation could be linear in this higher dimensional space. This transformation is defined as

$$\Phi : X \rightarrow \phi(X). \quad (3.9)$$

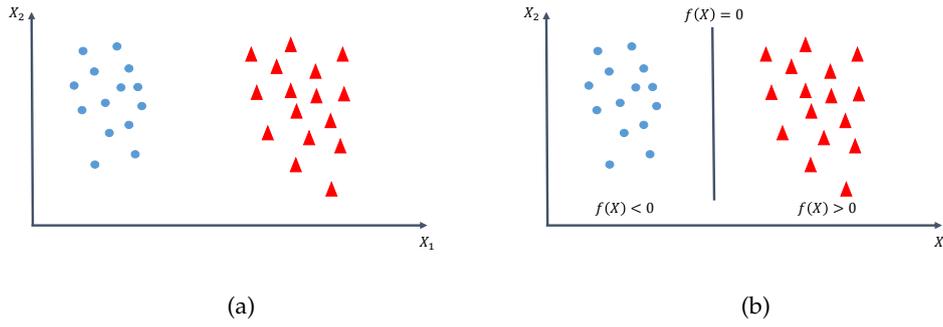


FIGURE 3.16: Bi-dimensional linear classification using SVM

Hence, the separating plane in the new feature space becomes

$$f(\phi(X)) = w^\top \phi(X) + b. \quad (3.10)$$

The kernel trick avoids the explicit mapping, i.e., the identification of $\phi(X)$. In fact, the SVM linear classifier relies on a dot product between data point vectors. In the new feature space, this becomes an inner product between vectors in the new space. The Kernel function satisfies this inner product:

$$k(\mathbf{X}, \mathbf{X}') = \langle \phi(\mathbf{X}), \phi(\mathbf{X}') \rangle, \quad (3.11)$$

which allows us to estimate the linear classifier in the new space without explicitly knowing $\phi(X)$, but by applying a kernel function to the data in the original space.

Different types of Kernel functions exist, among which we cite the following:

- Linear Kernel

$$K(X, X') = X \cdot X'. \quad (3.12)$$

- Polynomial Kernel

$$K(X, X') = (X \cdot X' + 1)^d \quad (3.13)$$

where d is the degree of the polynomial.

- Radial basis Kernel

$$K(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2\sigma^2}\right) \quad (3.14)$$

For example, the data in Fig. 3.17(a) cannot be separable by a line in the original space. That is why a new feature is added ($X_3 = X_1^2 + X_2^2$) and the data is projected to a higher dimension space containing the additional feature. In this new space, a hyperplane could separate the transformed data as in Fig. 3.17(b).

3.6.2.3 Estimating regions in the regression space using SVM

The PWARX algorithm used in this work is based on a clustering and an estimation procedure. The clustering aims at dividing regression data in groups characterized by their own dynamics and associating each regression vector to the corresponding group label. Hence, each group becomes defined on a certain region of the

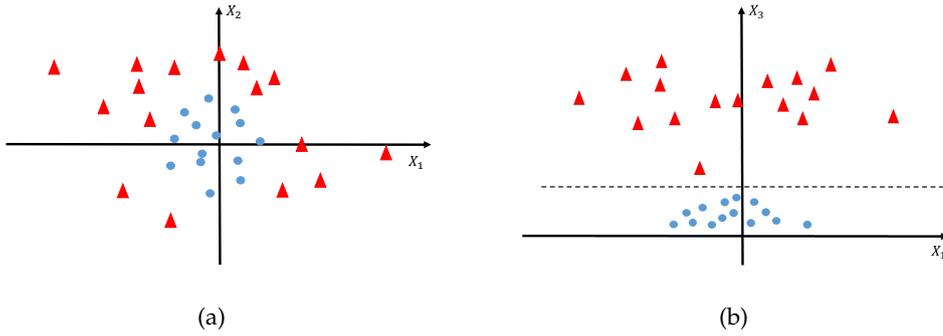


FIGURE 3.17: non-linear SVM classification by feature space transformation

regression space. Then, for the estimation part, a parameter vector is computed for each cluster and it represents the dynamic behavior of the building for that specific region of the regression space.

SVMs use the set of clustered data formed by the regression vectors and their associated label to identify the optimum hyperplanes that separate the data in the regression space.

The regression vector

$$\varphi(t) = [y(t-1) \quad \dots \quad y(t-n_a) \quad u(t-n_k) \quad \dots \quad u(t-n_b-n_k+1)]^T \quad (3.15)$$

contains n_a past values of the output and $n_b + n_k - 1$ past values of the input u . Given N regression vectors, the PWARX associates a label $s_i, i = 1, \dots, s$ to each regression vector. The SVM uses the labeled regression vectors to compute the vectors ω and the parameter b defining the equations of the hyperplanes dividing the regression space into s regions:

$$\omega^T \varphi(t) + b = 0. \quad (3.16)$$

Fig. 3.18 shows this learning step of the SVM algorithm.

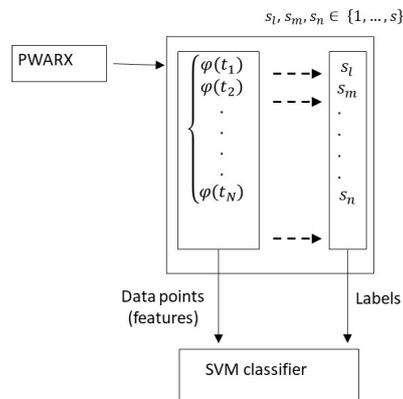


FIGURE 3.18: SVM learning

The SVM algorithm involves an optimization technique to calculate the values

of the parameters ω and b defining the hyperplanes such that the distance between the hyperplane and the closest data point is maximal. For more details about the calculation steps, the reader could refer to [Yu and Kim, 2012].

3.6.2.4 Predicting discrete states using SVM

Once the parameters of the hyperplane equations are computed, regression data for the prediction phase, φ' can be assigned to the corresponding region by computing the value of $\omega^\top \varphi'(t) + b$.

For example, let us consider that the regression vector contains only two parameters: the previous output and one input at the previous time

$$\varphi(t) = [y(t-1) \ u(t-1)]^\top, \quad (3.17)$$

and that two regions were estimated by the PWARX algorithm. Since this is a two-dimensional space, the SVM computes the equation of the line that separates these two regions. Let the equation of this line be

$$\omega_1 y(t-1) + \omega_2 u(t-1) + b = 0. \quad (3.18)$$

For a new regression vector

$$\varphi'(t) = [y'(t-1) \ u'(t-1)]^\top, \quad (3.19)$$

if

$$\omega_1 y'(t-1) + \omega_2 u'(t-1) + b > 0, \quad \varphi'(t) \in \mathfrak{R}_1 \Leftrightarrow \sigma(t) = 1,$$

else

$$\varphi'(t) \in \mathfrak{R}_2 \Leftrightarrow \sigma(t) = 2.$$

This is the simple case of a two-dimensional regression space as the one shown in Fig. 3.19. In practice, the dimension of the regression space is $d = n_a + n_u(n_b + n_k - 1)$ for one output of order n_a , n_u inputs of order n_b and a pure delay n_k . Each region \mathfrak{R}_i is therefore a convex polyhedron delineated by one or several planes of equation

$$\omega_1 \varphi(t, 1) + \omega_2 \varphi(t, 1) + \dots + \omega_d \varphi(t, d) + b = 0. \quad (3.20)$$

The discrete state $\sigma(t)$ for the prediction phase is found by evaluating the product in Eq. (3.20) of each identified plane and the regression vector for prediction $\varphi'(t)$ knowing that each region \mathfrak{R}_i is delineated by a number of planes such that

$$\sigma(t) = i \Leftrightarrow \varphi'(t) \in \mathfrak{R}_i \Leftrightarrow H_i \bar{\varphi}'(t) \leq \mathbf{0} \quad (3.21)$$

where H_i is a matrix containing the parameters $w_{j=1\dots d}$ and b of each plane that borders the region \mathfrak{R}_i . Fig. 3.20 shows the procedure of predicting the indoor air temperature using both SVM and PWARX algorithms. After the learning step (Fig. 3.18), the identified SVM classifier uses the new prediction data to find the region to which each data point belongs by following the rules presented above. The discrete state being identified, the PWARX algorithm uses it to estimate the indoor air temperature by multiplying the extended regression vector and the parameter vector corresponding to the identified region. For SVM classification, we use the SVM library LIBSVM, which handles multi-class classification problems [Chang and Lin, 2011].

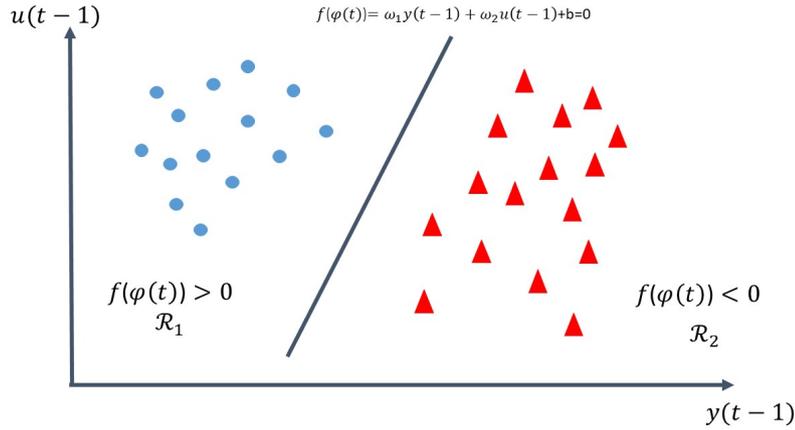


FIGURE 3.19: SVM classifier functioning

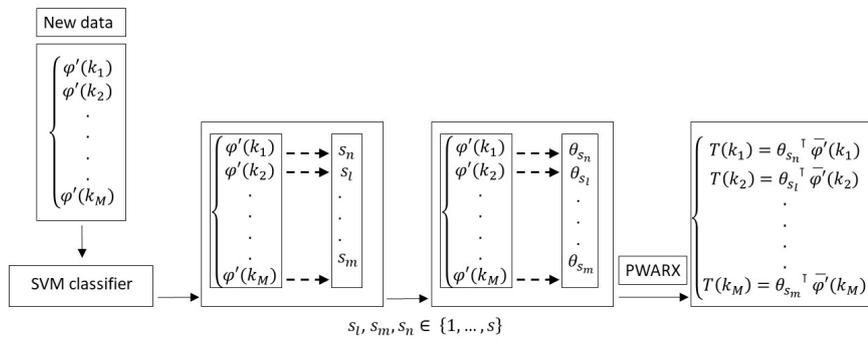


FIGURE 3.20: Predicting the temperature using SVM and PWARX algorithm

3.6.3 Estimation results

The orders of the model are determined by testing all orders less than 8 and the best result is found for $n_a = 5$, $n_b = 4$ and $n_k = 1$. Fig. 3.21 shows the number of detected sub-models and the accuracy of the estimated model in terms of the number of nearest neighbors c . The number of detected sub-models decreases as the number of nearest neighbors increases and it stabilizes at 3 for $c \geq 600$. For these values of c , the *Fit* of the estimated model is quasi stable, ranging between 65 and 66%. The highest value of the *Fit* was 70.38% and it was found for $c = 200$, but this led to 15 detected sub-models. Therefore, we choose the next best *Fit* found for a higher c , hence, lower number of sub-models, to keep the model simple. The choice was $c = 400$, with a *Fit* = 70% and 6 detected sub-models.

Fig. 3.22 shows the PWARX estimation results for $n_a = 5$, $n_b = 4$, $n_k = 1$ and $c = 400$. The number of detected sub-models is 6. The figure shows that some of the sub-models are common to all four seasons (sub-models 3 and 4) while others are specific to one season, like sub-model 5 only appearing in the summer period. Some of the other sub-models can be dominant for some seasons more than the

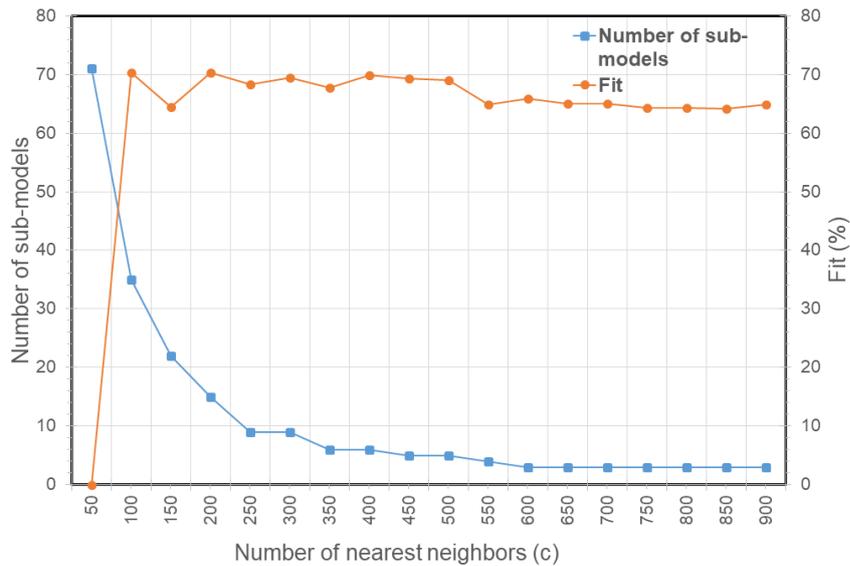


FIGURE 3.21: Number of detected sub-models and accuracy of the estimated model in terms of the number of nearest neighbors c for the case of the occupied building

others, like sub-model 3, that is more dominant in the spring and summer periods than the winter and autumn periods. Having a large number of inputs renders the relationship between the switching instances and the inputs to the model difficult to interpret, hence, the SVM classification technique presented in the previous section is used for the automated classification. The following values of the error criteria are found for this estimation $Fit_{est} = 70\%$, $MAE_{est} = 0.39^{\circ}C$, $RMSE_{est} = 0.5^{\circ}C$ and $R_{adj}^2 = 0.9$.

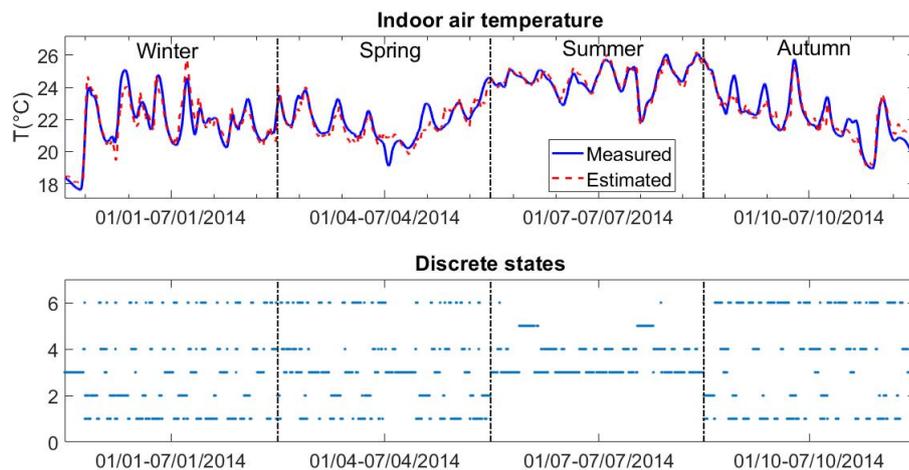


FIGURE 3.22: PWARX estimation results for the third test case of the occupied building

3.6.4 Validation results

As explained in Sect. 3.6.2, an SVM classifier is used in this case study for validation purposes as this classifier allows to determine the discrete state to be assigned to the regression data in the prediction phase.

3.6.4.1 SVM validation

The classifier is trained using labeled data obtained from the PWARX algorithm. First of all, to assess the accuracy of the classification, the labeled samples of regression data obtained by the PWARX algorithm are divided into training and testing samples. For the current example, labeled data from the first three seasons is used to train the classifier and the data for the last season (autumn) is used for determining the accuracy of the classification. The best classification is found for a polynomial Kernel function of third degree with an accuracy of 97.8%.

3.6.4.2 PWARX validation results

In the previous examples, the prediction was done for the whole prediction period. For this example, the prediction will also be done for different steps ahead. In a residential building, predicting the indoor air temperature for several time scales ahead could be needed for implementing control strategies. Since this test case is for a real occupied building, it represents a more complete study, therefore, it would be interesting to investigate predictions for a near horizon.

The one step ahead prediction considers that $\hat{y}(t)$, which is the estimated output, is computed from measured outputs up to instant $t - 1$ as follows:

$$\hat{y}^{(1)}(t) = f[\hat{y}^{(0)}(t-1) \dots \hat{y}^{(0)}(t-n_a), u(t-1) \dots u(t-n_b)], \quad (3.22)$$

with $\hat{y}^{(0)}(t-1)$ is the actual measured value of the output at instant $t-1$, i.e. $y(t-1)$.

Based on this one step ahead formula, the multiple steps ahead predictions are derived by recursively applying the same prediction formula [Zhang and Ljung, 2004]. The k^{th} steps ahead prediction of the output is hence a function of the output at the $k-1$ step and the inputs as follows (we consider a pure delay of $n_k = 1$ for simplicity purposes),

$$\hat{y}^{(k)}(t) = f[\hat{y}^{(k-1)}(t-1) \dots \hat{y}^{(k-1)}(t-n_a), u(t-1) \dots u(t-n_b)]. \quad (3.23)$$

The two steps ahead prediction of the output is therefore calculated by

$$\hat{y}^{(2)}(t) = f[\hat{y}^{(1)}(t-1) \dots \hat{y}^{(1)}(t-n_a), u(t-1) \dots u(t-n_b)], \quad (3.24)$$

and so on.

We call simulation, the results obtained for the prediction of the output for the entire prediction period (which can be viewed as infinite steps ahead predictions), namely the 8th to the 10th of the first month of each season. This is given as

$$\hat{y}(t) = f[\hat{y}(t-1) \dots \hat{y}(t-n_a), u(t-1) \dots u(t-n_b)]. \quad (3.25)$$

In this case, the output at instant t is computed in terms of previous outputs simulated by the model without using any measurements. Recall that for the first two test cases, only simulation results were given.

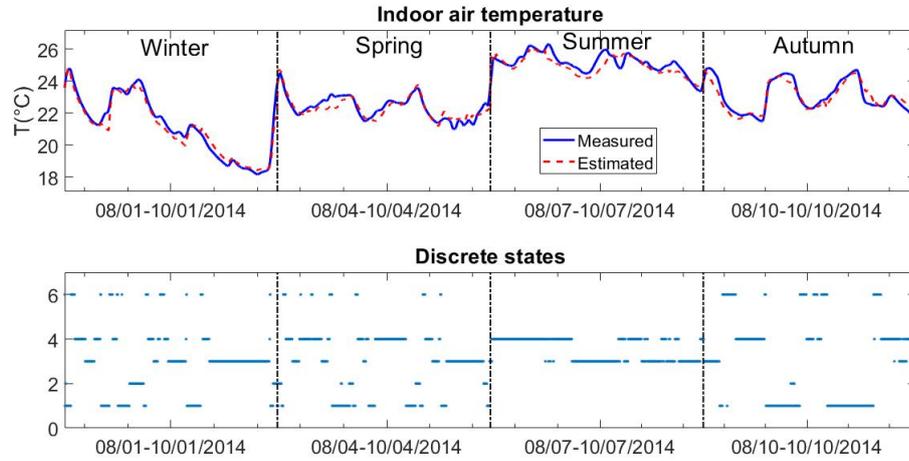


FIGURE 3.23: PWARX simulation results for the whole prediction period

The simulation results are given in Fig. 3.23. By observing the measured and simulated temperature curves, one can see that the prediction is mostly accurate during winter time while for spring and summer periods, the prediction seems less accurate. Sub-models 3 and 4 are once again found most dominant for the summer period while sub-model 5 is missing. The following results are found for this simulation: $Fit_{pred} = 80.2\%$, $MAE_{pred} = 0.28^{\circ}\text{C}$, $RMSE_{pred} = 0.36^{\circ}\text{C}$ and $R^2_{adjpred} = 0.96$.

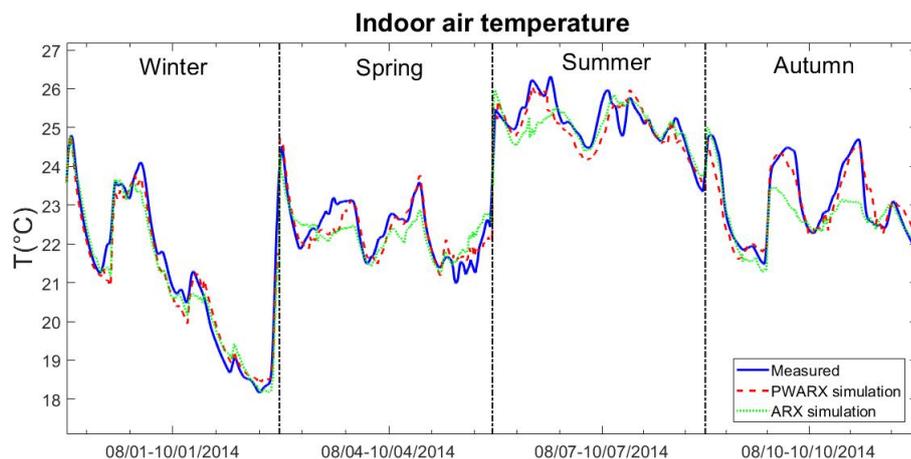


FIGURE 3.24: Comparison between PWARX and ARX simulation results for the whole prediction period

As for the previous test cases, a comparison between the simulation results obtained with the PWARX model and those obtained with an ARX model is provided. For the ARX model the same orders seem to be the most accurate. The results are given in Fig. 3.24 and they show that the prediction for winter time is

close, however, the PWARX model seems to have more accurate predictions for the remaining seasons. Table 3.5 shows the prediction results obtained for each season separately for both ARX and PWARX models. The estimation data is the same, but the prediction is done for each season separately. The following values are found for the ARX model prediction for the whole period: $Fit_{pred} = 72.9\%$, $MAE_{pred} = 0.36^\circ\text{C}$, $RMSE_{pred} = 0.5^\circ\text{C}$ and $R^2_{adjpred} = 0.92$.

Season	Fit (%)	MAE ($^\circ\text{C}$)	RMSE ($^\circ\text{C}$)	R^2_{adj}
PWARX				
Winter	80.1	0.29	0.37	0.96
Spring	54.16	0.25	0.32	0.78
Summer	34.71	0.31	0.4	0.56
Autumn	63.99	0.27	0.35	0.86
ARX				
Winter	78.26	0.3	0.41	0.95
Spring	33.66	0.4	0.47	0.55
Summer	26.57	0.32	0.45	0.45
Autumn	35.17	0.43	0.63	0.57

TABLE 3.5: Validation criteria for prediction results of PWARX and ARX models

Table 3.6 shows the performance of the PWARX and ARX models for 1 step ahead prediction (5min), 6 steps ahead prediction (30 min) and 12 steps ahead prediction (one hour), as well as the simulation for the whole period. The results in this table show a better performance of the PWARX model. For small time horizon (5 min and 30 min ahead), the performance of these two models is very close, however, when the prediction is for a longer time scale, the PWARX model outperforms the ARX model. Fig. 3.25 shows the prediction for one hour ahead (12 steps) by the PWARX model.

Prediction Step	Fit (%)	MAE ($^\circ\text{C}$)	RMSE ($^\circ\text{C}$)	R^2_{adj}
PWARX				
1 (5 min)	98.86	0.013	0.02	0.999
6 (30 min)	94.45	0.063	0.102	0.996
12 (1 hour)	89.66	0.116	0.1913	0.989
Simulation (12 days)	80.22	0.285	0.366	0.96
ARX				
1 (5 min)	98.85	0.012	0.02	0.999
6 (30 min)	93.56	0.078	0.119	0.995
12 (1 hour)	88.32	0.147	0.216	0.986
Simulation (12 days)	72.97	0.36	0.5	0.926

TABLE 3.6: Validation criteria for prediction results of PWARX and ARX models

3.7 Discussion of results and conclusion

This chapter aimed at investigating the performance of the PWARX approach in modeling the indoor air temperature of buildings. For this purpose, three test cases

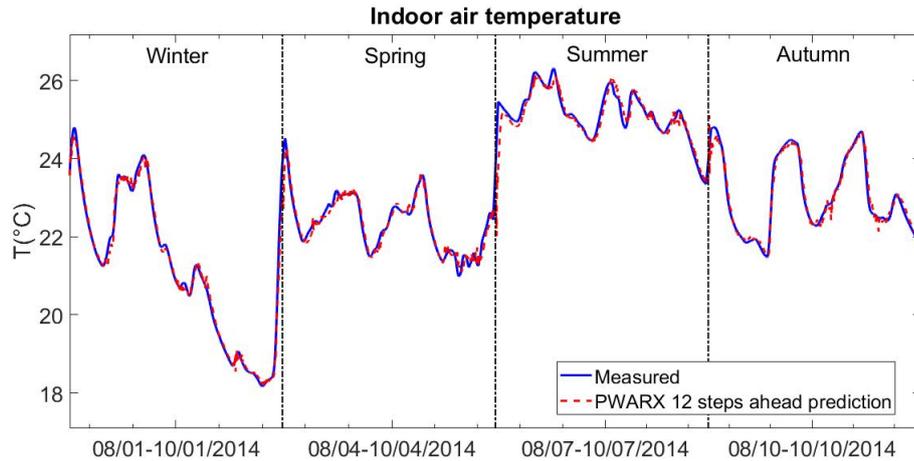


FIGURE 3.25: One hour ahead prediction using PWARX

were presented ranging from the easiest to the most complex. The first two test cases are rather academic. In the first one, the algorithm used data measured in the Eco-confort room where configurations related to the heating mode (ON/OFF) were identified. For the second one, the algorithm used data simulated by a numerical model of a room where the configurations were related to the heating mode (ON/OFF), the ventilation mode, the presence of occupants and the position of shades. The third test case was for an occupied building monitored throughout several years. The results showed that the PWARX model was able to detect different configurations in each test case and to predict the indoor air temperature with a higher accuracy than classical ARX models.

It was observed that the number of detected models and therefore the accuracy of the model depends on the chosen number of nearest neighbors c . A compromise between the number of sub-models and the accuracy of results has to be found. An analysis done after the estimation of the sub-models could relate the switching to the physical reasons. However, sometimes, due to the interference of multiple configurations, the complexity of a building system and the lack of observability of the occupants' behavior, it is hard to determine the physical reasons behind the switching.

The discrete states found for the first two test cases were related to the applied configurations, however, for the third test case, no clear interpretation of the switching could be derived. Hence, for the prediction phase, the discrete state for the prediction data of the first two test cases was determined based on the values of the influencing variables (heating, solar radiation, ventilation, etc.). However, for the third test case, the number of influencing variables was high, hence, the discrete state was determined in an automated way by using SVM classification.

In the three test cases, the PWARX models showed better performance than classical ARX models. The difference in their performance becomes important when the behavior of the temperature is highly influenced by sudden events leading to a different dynamic behavior. This was observed for the second case, where the PWARX clearly outperformed the ARX approach. In the third test case, the performance of both ARX and PWARX was similar for short time ahead prediction.

However, when the prediction horizon becomes longer, the PWARX outperforms the classical ARX model.

In this chapter, the use of PWARX for modeling the indoor air temperature of a building was validated through several test cases. In the next chapter, the use of PWARX models for predicting the energy consumption of buildings is presented. For that purpose, the models are incorporated in a hybrid control loop for simulating standard scenarios.

Chapter 4

Predicting the energy consumption of buildings using a data driven PWARX model

This chapter presents a contribution to the evaluation of the energy consumption of buildings using data-driven approaches. To do that, the PWARX modeling technique, coupled with the SVM classification technique, are integrated in a closed loop with controllers designed for each one of the continuous sub-models identified in the modeling stage. The proposed methodology is applied on a test cell located in Angers, France, for which measurements provided before and after applying insulation are used to predict the energy consumption in each case and assess the gain in energy efficiency.

4.1 Need for estimating the energy consumption of buildings

Previous chapters showed how to use PWARX algorithm for modeling the indoor air temperature of a building. This chapter discusses how to integrate this approach in a control loop for the purpose of predicting the energy consumption of a building previously modeled via a data driven approach. Because of the energy and environmental impact of the building sector, many countries have adopted legislations addressing the issue of building energy consumption and efficiency. Hence, the need to verify the compliance of the building to the regulations has risen. Moreover, the assessment of the energy consumption of buildings is the key element before taking any energy efficient actions, whether they pertain to the construction materials and envelope or to the HVAC systems. In France, the energy performance of the building is estimated by simulating the building thermal and energy response to so-called regulatory scenarios. These are predefined standard scenarios related to the occupancy of the building, the set point temperature, the lighting schedule, the ventilation, etc. Standard behavior of the occupants is considered and the calculation of the energy consumption is based on this usage. Different scenarios are possible according to the type of building (residential, hotels, schools, etc.). For example, the set point temperature in winter in a residential building is 19°C when the house is occupied and 16°C when the occupants are away.

This chapter presents a methodology showing how to use the PWARX model developed from measurements as a tool for a fast and accurate estimation of the energy consumption of a building subject to well defined scenarios. This methodology is dedicated for existing buildings because of its dependency on data measured

during the operation of the building. It can also use simulated data if measured data is not available. This could be useful in case the original thermal model of the building is too complicated to be used for simulating standard scenarios and evaluating its energy performance. In this case, the original complex model is used to generate simulated data that is used by the PWARX algorithm to estimate a simpler model (a PWARX model) that can be used later on for estimating the energy performance of the building.

The performance of the proposed methodology for predicting the energy consumption of buildings is tested on an experimental cell located in Angers, France. This cell has undergone renovation work, hence, two measurement campaigns were conducted, one prior and the other after renovation. During each measurement campaign, different scenarios were applied inside the cell. The proposed methodology is used to predict the energy consumption of the cell before and after insulation for the same scenario and assess the improvement in the energy efficiency due to the insulation work. The result is then compared to the reduction percentage of the overall Heat Transfer Coefficient (HTC) of the cell due to insulation, the HTC being calculated using the physical and thermal properties of the envelope before and after insulation.

The chapter is organized as follows. In Sect. 4.2, the concept of the proposed control methodology is presented. In Sect. 4.3 a description of the case study of a test cell in Angers, France is provided, and the obtained results are shown in Sect. 4.4 while Sect. 4.5 is dedicated to the conclusion.

4.2 Proposed control methodology

The general idea behind this methodology is to predict the amount of energy needed to maintain the temperature in the building close to a reference temperature set by the user, the building being subject to normal weather conditions. This is done by integrating the model in a control loop. Designing an effective control loop for regulating the indoor air temperature starts with creating a simple but accurate model of the thermal behavior of the building at hand. In this case, the PWARX model built from measurements is used. The control loop is presented in Fig. 4.1.

We call process variable the system variable to be controlled, in our case, the air temperature of the building zone. This variable is predicted by the PWARX model and fed into the control loop. The difference between the measured value of the process variable and the desired set point is given to the controller that uses it to determine the desired heat input to drive the system. This is the main principle behind the closed control loop.

In our case, we use the PWARX technique presented in Sect. 2.4 for modeling the indoor air temperature using inputs like the heating power and meteorological data. The model is then integrated in a closed control loop to calculate the energy needed to keep the temperature in the building within a certain range following a scenario that we impose.

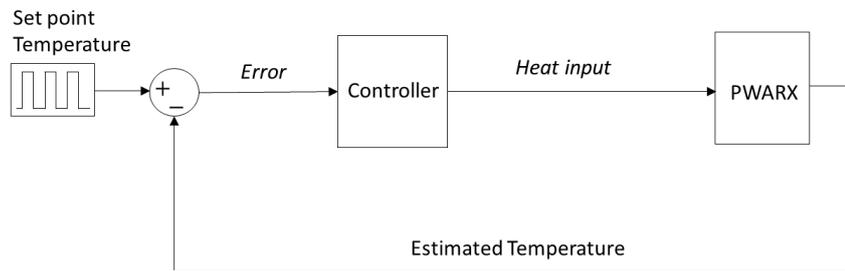


FIGURE 4.1: Temperature regulation by closed loop control

The PWARX model is a hybrid model yielding a set of sub-models representing different configurations or functioning modes of the building (heating ON/OFF, ventilation ON/OFF, etc.). An important challenge at this point would be to design a hybrid controller that corresponds to the hybrid aspect of the PWARX model of the building and to select the right controller according to the temperature dynamics. The decision rule for switching between one sub-model and another is acquired from the regression data. In the previous chapter, it was shown how the SVM classification technique is used to estimate the equations of the borders defining the regions in the regression space, in order to estimate the discrete state of new data. The same rule will be applied here for switching between controllers.

The first steps of the methodology have already been detailed in the previous chapters. First of all the collection of data should respect the points mentioned in Sect. 3.2. Once the inputs and outputs are selected and the data is collected, the next step would be to identify dynamic thermal models using the PWARX algorithm presented in Sect. 2.4. The reader should refer to Sections 2.3, 2.4 and 3.6.2 for more details about the PWARX model structure and the main steps of the estimation algorithm and the validation procedure using the SVM classification technique.

4.2.1 Simulate the identified models using standardized scenarios in closed loops with PID controllers

In order to predict the energy consumption of the building, a closed loop simulation is performed using the identified PWARX model. A first step would be to choose and design an appropriate type of controllers. The most common controllers used in the buildings industry are PID (Proportional Integral Derivative) controllers [Åström and Hägglund, 1995]. The choice of this type of controllers is motivated by the fact that they are known for the robustness of their performance, their simplicity and the availability of many algorithms to tune their parameters: proportional, integral and derivative coefficients, which are varied to obtain a response satisfying the specifications. These specifications refer to the rise time, the settling time, the overshoot or the steady-state error of the closed-loop response with respect to the

set point temperature.

The controller is designed based on the specifications and the model available for the plant, which are assumed known. The standard continuous-time PID controller is defined by

$$c(t) = K(e(t) + \frac{1}{T_i} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt}) \quad (4.1)$$

where $c(t)$ is the control signal and $e(t)$ is the error between the measured process variable and the reference variable (set point). The control signal is thus a sum of three terms: the P-term (which is proportional to the error), the I-term (which is proportional to the integral of the error), and the D-term (which is proportional to the derivative of the error). The controller parameters are the proportional gain K , the integral time T_i , and the derivative time T_d . The integral, proportional and derivative part can be interpreted as control actions based on the past, the present and the future.

Eq. (4.1) can be rewritten in an expanded form to give another universal expression of the control function

$$c(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt}, \quad (4.2)$$

where $K_p = K$, $K_i = \frac{K}{T_i}$ and $K_d = KT_d$ are respectively the proportional, integral, and derivative terms coefficients to be tuned. The continuous-time transfer function of a PID controller is found by taking the Laplace transform of Eq. (4.2):

$$C(s) = K_p + \frac{K_i}{s} + K_d s. \quad (4.3)$$

This equation can be transformed from continuous-time to discrete-time using one of the available methods: forward rectangular, backward rectangular or bilinear transformation. We consider the bilinear method for which

$$s = \frac{2}{T_s} \frac{z-1}{z+1}, \quad (4.4)$$

where T_s is the sampling time and z is the backward shift operator. The discrete-time transfer function is therefore:

$$C_d(z) = K_p + K_i \frac{T_s z + 1}{2 z - 1} + K_d \frac{2 z - 1}{T_s z + 1}. \quad (4.5)$$

Sometimes, a derivative filter is also considered in order to achieve a better performance for the controller [Åström and Hägglund, 1995]. In this case the transfer function becomes:

$$C_d(z) = K_p + K_i \frac{T_s z + 1}{2 z - 1} + K_d \frac{1}{T_f + \frac{T_s z + 1}{2 z - 1}}, \quad (4.6)$$

where T_f is the derivative filter time constant.

The hybrid PWARX model being formed by a set of discrete-time sub-models,

a set of discrete-time parallel PID controllers will be integrated in the closed loop. The design of the PID controllers (tuning the three parameters) is done using the PID Tuner application of Matlab based on the transfer function of the model. In Sect. 2.2.3, we established the equivalence between discrete transfer functions and models in input-output form. For estimating the controllers, the tuning application is based on the transfer function of the model, hence, PWARX models are transformed into discrete-time transfer function models. Since the PWARX model is a set of Multiple Inputs Single Output (MISO) sub-models, for each sub-model j , a transfer function can be derived from the ARX sub-models for each input i as follows:

$$\begin{aligned} H_j^i(z) &= \frac{T(t)}{u_j^i(t)} = \frac{B_j^i(z)}{A_j(z)} \\ &= \frac{b_{j,n_k}^i z^{-n_k} + \dots + b_{j,n_k+n_b-1}^i z^{-n_k-n_b+1}}{1 + a_{j,1} z^{-1} + \dots + a_{j,n_a} z^{-n_a}} \end{aligned} \quad (4.7)$$

where $j = 1, \dots, s$ is the sub-model index, $i = 1, \dots, n_u$ is the input index, z^{-m} is the backward shift operator such that $z^{-m}T(t) = T(t - m)$ and finally $b_{j,l}^i$, $l = n_k, \dots, n_b + n_k - 1$ and $a_{j,k}$, $k = 1, \dots, n_a$ are the corresponding coefficients (components of the parameter vector θ) multiplied by the input i and the temperature output of sub-model j respectively. Hence, the control parameters are computed by the tuning application for each local transfer function, thus generating a set of controllers corresponding each to certain temperature dynamics (modes, configurations) of the building.

The performance of the control system is done by investigating the response to a step function used as the set point command variable. The response is usually quantified by the waveform characteristics: rise time, settling time and overshoot presented in Fig. 4.2. Rise time is the amount of time the system takes to go from 10% to 90% of the final value (steady-state) and settling time is the time required by the process variable to settle within a certain percentage (2% for example) of the steady-state value. Overshoot is the percentage of the final value by which the process variable overshoots the final value. Steady-state error is the final difference between the process variable and set point. The stability of the closed-loop system, i.e., having a bounded output for every bounded input, is also an important aspect to investigate. Changing the parameters of the PID controller changes the value of these characteristics. For instance, increasing the proportional gain will increase the speed of the control system response. However, if the proportional gain is too large, the system might become unstable. The effect of the integral response is to drive the steady-state error to zero. The derivative component attempts to reduce overshoot and increase the speed of the overall control system response. However, if the derivative time is too large the control system can become unstable.

The PID tuner graphical interface allows to see the effect of changing the controller parameters on the response of the system. The initial parameters are computed by the PID tuner based on the transfer function of the system, then the parameters are altered in order to get the desired response. For this application, the parameters are altered so that the response can be as quick as possible (reducing rise and settling time) while maintaining the stability of the loop and an acceptable overshoot (less than 10%).

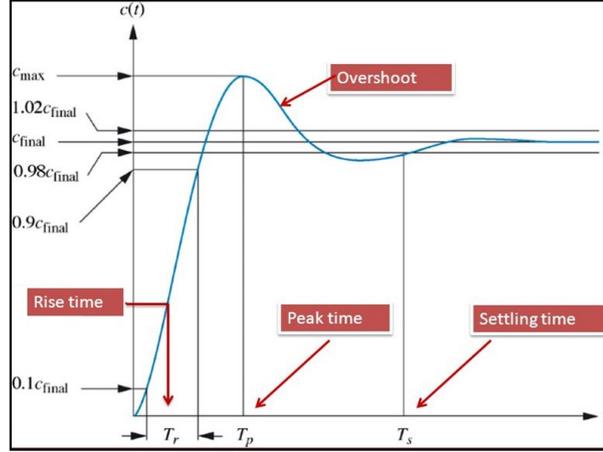


FIGURE 4.2: Control response characteristics

The design of the controller will be based solely on the SISO transfer function $H_j^{heat}(z)$ for the heat input for every sub-model j ($u_j^{heat}(t)$), because the controller will act on the heat power input to drive the temperature to its reference.

$$\begin{aligned}
 H_j^{heat}(z) &= \frac{T(t)}{u_j^{heat}(t)} = \frac{B_j^{heat}(z)}{A_j(z)} \\
 &= \frac{b_{j,n_k}^{heat}z^{-n_k} + \dots + b_{j,n_k+n_b-1}^{heat}z^{-n_k-n_b+1}}{1 + a_{j,1}z^{-1} + \dots + a_{j,n_a}z^{-n_a}}.
 \end{aligned}
 \tag{4.8}$$

We then design a parallel PID controller for each $H_j^{heat}(z)$, $j = 1, \dots, s$ as in the following equation:

$$C_j(z) = K_p^j + K_i^j \frac{T_s z + 1}{2 z - 1} + K_d^j \frac{1}{T_f^j + \frac{T_s z + 1}{2 z - 1}}.
 \tag{4.9}$$

where K_p^j is the proportional gain, K_i^j is the integral gain and K_d^j is the derivative gain and T_f^j is the derivative filter time constant for controller C_j corresponding to the j^{th} sub-model.

As mentioned earlier, an important challenge is to select the right controller corresponding to each discrete mode of the building. The switching between the different controllers will follow the same decision rule acquired for the switching between the discrete sub-models. This will result in a hybrid PID control, with each controller being designed for a certain operating range. In this methodology, the operating range for the controller is the same as the one found for the sub-model to which it is related. In building thermal simulation, multiple research papers have already introduced hybrid control to HVAC systems [Witrant et al., 2009; Shein et al., 2012].

Lastly, the designed PID controller is integrated in the loop represented in Fig. 4.3 and its operation is detailed in the following. The algorithm starts by initializing the temperature and the heat input. Using these values and the values of the other

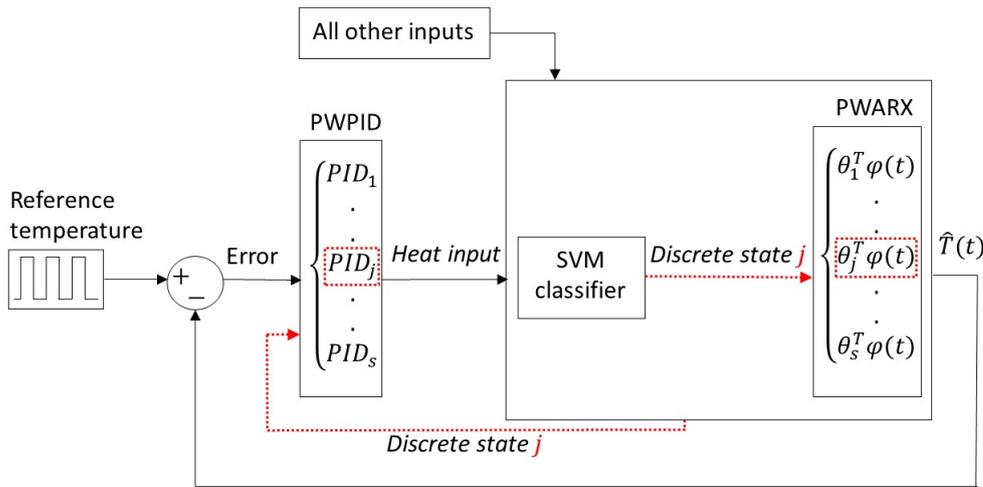


FIGURE 4.3: Control loop diagram

inputs like the ambient temperature and the solar radiation (which are no other than the variables composing the regression vector), the SVM classifier finds the current discrete state j . The discrete state allows to choose the j^{th} sub-model of the PWARX model which also uses the same inputs for predicting the temperature at the current time t . The same discrete state j is used to choose the corresponding PID controller. The error between the reference temperature and the one predicted by the PWARX model is therefore fed to the j^{th} PID controller that tries to minimize the error by delivering the convenient heat input. The heat input is then used with the other inputs to find the discrete state (using SVM) and the temperature at the next time step ($t + 1$) and so on [Ajib et al., 2018b].

4.2.2 Evaluation of the energy consumption

For the cell considered in this test case, the energy consumption is the power delivered by the radiator to maintain the set point temperature for the whole period of the scenario. The amount of energy consumed during a scenario of length L with a sampling time S and a heat power input at each time sample t ($u_{heat}(t)$) is computed as:

$$E(kWh) = \sum_{t=1}^L (u_{heat}(t)(kW) \cdot S(hour)). \quad (4.10)$$

4.3 Case study-HUMIBATEX Angers

The accuracy of the proposed methodology in predicting the energy consumption due to heating is evaluated on the following case study. It is a small building with a surface area of approximately $16m^2$ located in Angers, France and shown in Fig. 4.4. The building is separated into a humid zone ($6 m^2$) and a dry zone ($10 m^2$). During the measurement campaign, the dry zone was subject to insulation work. Hence, measurements collected before and after insulation in this zone will be

used to investigate the ability of the proposed methodology to evaluate the energy performance of the cell. The accuracy of this estimation is tested by calculating the gain in the energy performance brought by renovation and then comparing it to the



FIGURE 4.4: Building under study-HUMIBATEX Angers

reduction in the overall Heat Transfer Coefficient (HTC) of the dry zone before and after insulation, calculated based on the thermal properties of the construction and the insulation material. The HTC calculates the overall heat lost by the building envelope. The reduction in the HTC after insulation is highly linked to the gain in the energy performance.

TABLE 4.1: Scenarios applied for this case study

<i>Scenario</i>	<i>Start date</i>	<i>End date</i>	<i>Insulation</i>	<i>Heat</i>	<i>DFV</i>	<i>ABV</i>
1	18/09/2015	08/10/2015	-	-	-	-
2	14/10/2015	05/11/2015	-	✓	-	-
3	05/11/2015	04/12/2015	-	✓	✓	-
4	04/12/2015	07/01/2016	-	✓	-	✓
5	05/02/2016	04/03/2016	✓	✓	-	✓
6	04/03/2016	01/04/2016	✓	✓	✓	-
7	01/04/2016	29/04/2016	✓	✓	-	-
8	29/04/2016	27/05/2016	✓	-	-	-
9	27/05/2016	17/06/2016	✓	-	✓	-
10	17/06/2016	18/07/2016	✓	-	-	✓

The building is equipped with various sensors measuring weather data (ambient temperature and humidity, solar radiation and rain level) on each facade. Sensors are installed inside the cell to measure the temperature and humidity inside each zone at different height levels. Heating is provided in the dry zone via two electric heaters of 750W maximum power each. Two types of ventilation mechanisms are implemented, a Double Flux mechanical Ventilation (DFV) and a direct Air Blowing Ventilation (ABV). The DFV uses two networks of ducts for supply and return air, each equipped with a ventilator, the air flow being predetermined. The heat of the exhaust air conducted outside the building is recovered in order to heat the new incoming air. The ABV sucks the air from outside, heats it via an electric resistance to avoid condensation, and then blows it directly into the building which creates

an over-pressure. Then the air is naturally pushed outside via ventilation grilles. Sensors are used in the air ducts to measure the temperature and humidity of air conducted to and from the cell. Sensors are also used to measure the energy consumption of the mechanical systems (radiators, ventilation systems, humidifier and dehumidifier). The building is subject to different scenarios related to the heating and ventilation before and after insulation. These are summarized in Table 4.1. Sampling time is 10 minutes.

4.4 Results

In this section, the results for modeling the air temperature of the dry zone of the Humibatex building using the PWARX algorithm 1 before and after insulation are presented, along with the results obtained for the control loop implemented to estimate the energy consumption of the zone before and after insulation for predefined scenarios.

4.4.1 Estimation and validation results

Two data sets corresponding to measurements before and after insulation are provided, hence, two PWARX models are estimated. Each data set is divided into two: the first one contains the first half of each scenario and is dedicated to the model identification process while the second one contains the other half of each scenario and is used for validation. To reduce the errors caused by the discontinuity of data at the beginning of each new scenario, the first n_a data points of each new scenario are set equal to the measured ones for both estimation and validation. The output of the PWARX model is the indoor air temperature. The inputs to the model are: the ambient dry bulb temperature, the ambient relative humidity, the global solar radiation, the heating power from the radiators, the temperature of the air blown into the zone from both ventilation mechanisms and the air flow at the diffuser level.

To assess the precision of the model estimation and validation, the *Fit* (Eq. (3.1)), *MAE* (Eq. (3.2)), *RMSE* (Eq. (3.3)) and R_{adj}^2 (Eq. (3.4)) are used.

4.4.1.1 Before insulation

Fig. 4.5 shows the results obtained for modeling the indoor air temperature of the dry zone before insulation (first four scenarios in Table 4.1). For $n_a = 4$, $n_b = 3$, $n_k = 1$ and $c = 500$, four sub-models are found. The first sub-model corresponds to the free evolution of the temperature without any heating or ventilation (scenario 1). The others correspond to different behaviors of the temperature caused by the heating and ventilation. For this estimation we find $Fit_{est} = 75\%$, $MAE_{est} = 0.2^\circ\text{C}$, $RMSE_{est} = 0.28^\circ\text{C}$ and $R_{adj,est}^2 = 0.94$.

SVM classifiers are trained using labeled regression data from the estimation period. They are then used to classify validation data into one of the four identified regions and the corresponding sub-model is used to simulate the temperature in each region (cf. Sect. 3.6.2). The validation results are shown in Fig. 4.6. A good accuracy is found with $Fit_{pred} = 68\%$, $MAE_{pred} = 0.27^\circ\text{C}$, $RMSE_{pred} = 0.4^\circ\text{C}$ and

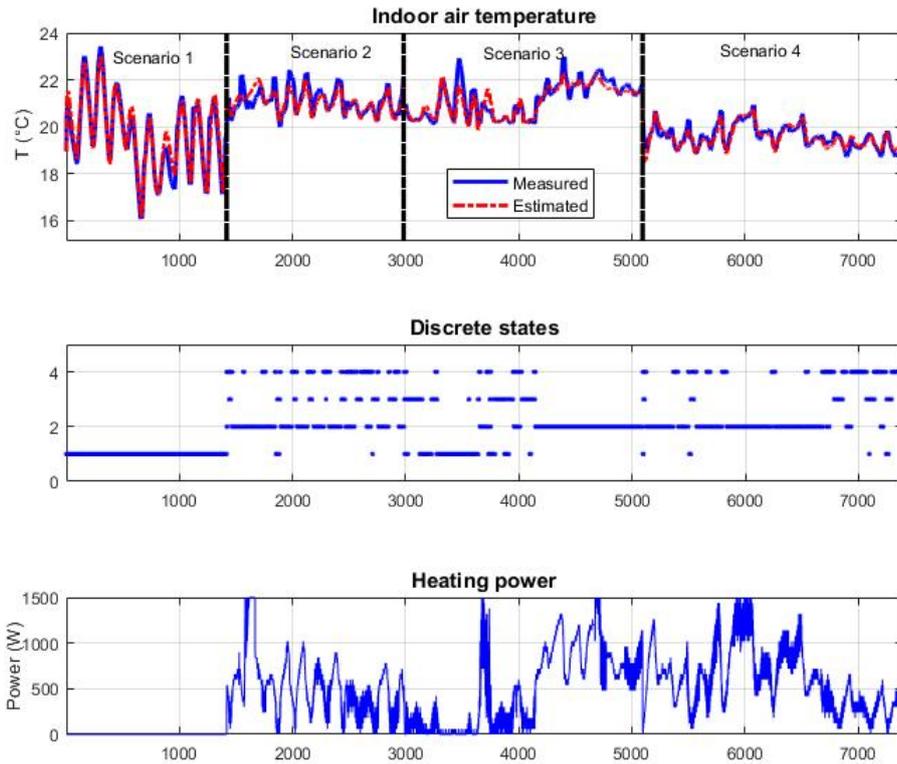


FIGURE 4.5: PWARX estimation results before insulation

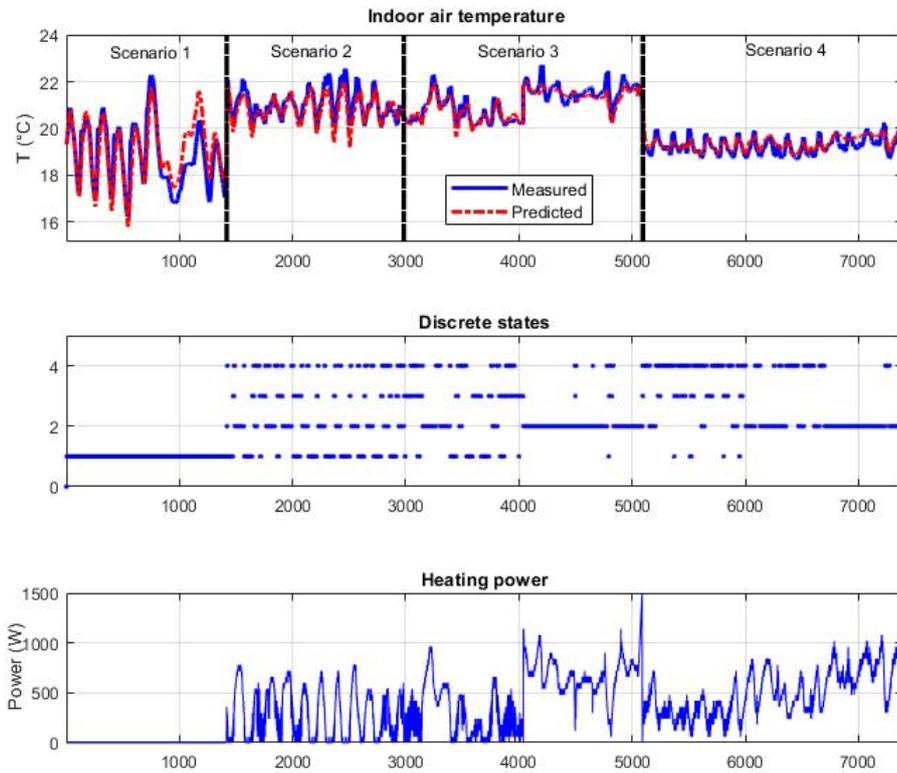


FIGURE 4.6: PWARX validation results before insulation

$$R_{adj,pred}^2 = 0.9.$$

4.4.1.2 After insulation

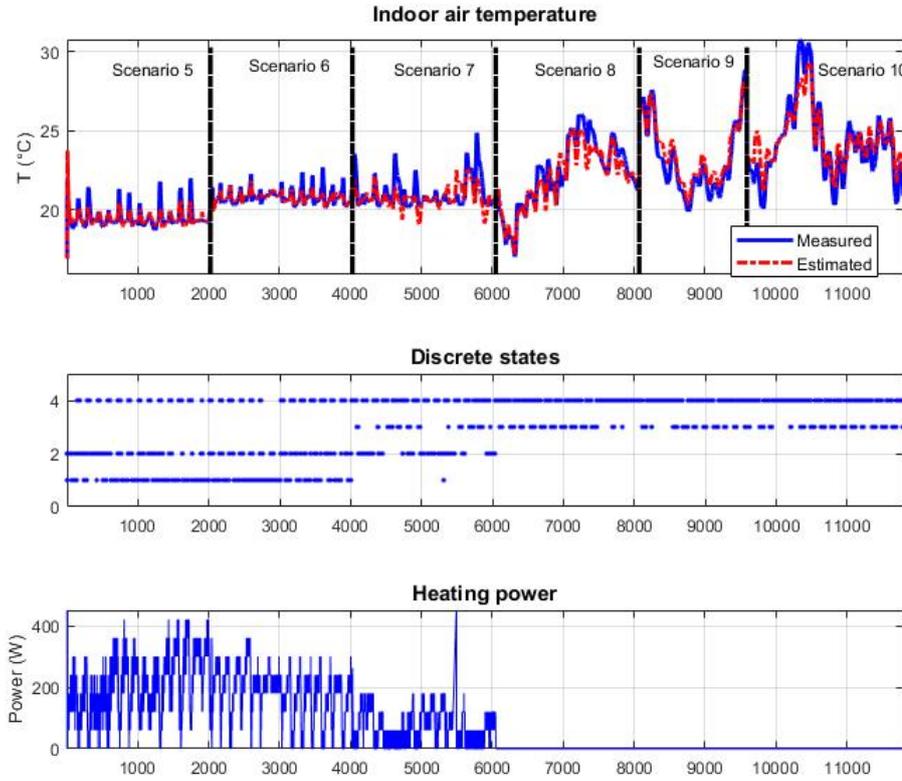


FIGURE 4.7: PWARX estimation results after insulation

The same procedure is applied on data collected in the building after insulation (last 6 scenarios). The estimation results are shown in Fig. 4.7. Four sub-models were also obtained but this time for $n_a = n_b = 5$, $n_k = 1$ and $c = 500$. Two sub-models are estimated for the free evolution of the temperature without heating or ventilation (scenario 8), the same sub-models are also found for an evolution with ventilation (scenarios 9 and 10), the reason behind this being that during the last three scenarios, the ambient temperature was quite high and in the same range of the indoor air temperature, so the ventilation did not considerably change the behavior of the temperature inside the building. The other two sub-models appear when the heating is ON. This estimation gives $Fit_{est} = 70.75\%$, $MAE_{est} = 0.46^\circ\text{C}$, $RMSE_{est} = 0.68^\circ\text{C}$ and $R_{adj,est}^2 = 0.91$. In the same way, an SVM classifier is trained using labeled regression data from the estimation part and then used to classify validation data in order to determine the discrete state at each instant t . The validation results are presented in Fig. 4.8 with $Fit_{pred} = 69\%$, $MAE_{pred} = 0.56^\circ\text{C}$, $RMSE_{pred} = 0.76^\circ\text{C}$ and $R_{adj,pred}^2 = 0.9$.

4.4.2 Evaluating the energy consumption before and after insulation

To evaluate the energy consumption of the building, the PWARX models identified in Sect. 4.4.1 are integrated in a closed loop with PID controllers and simulated for

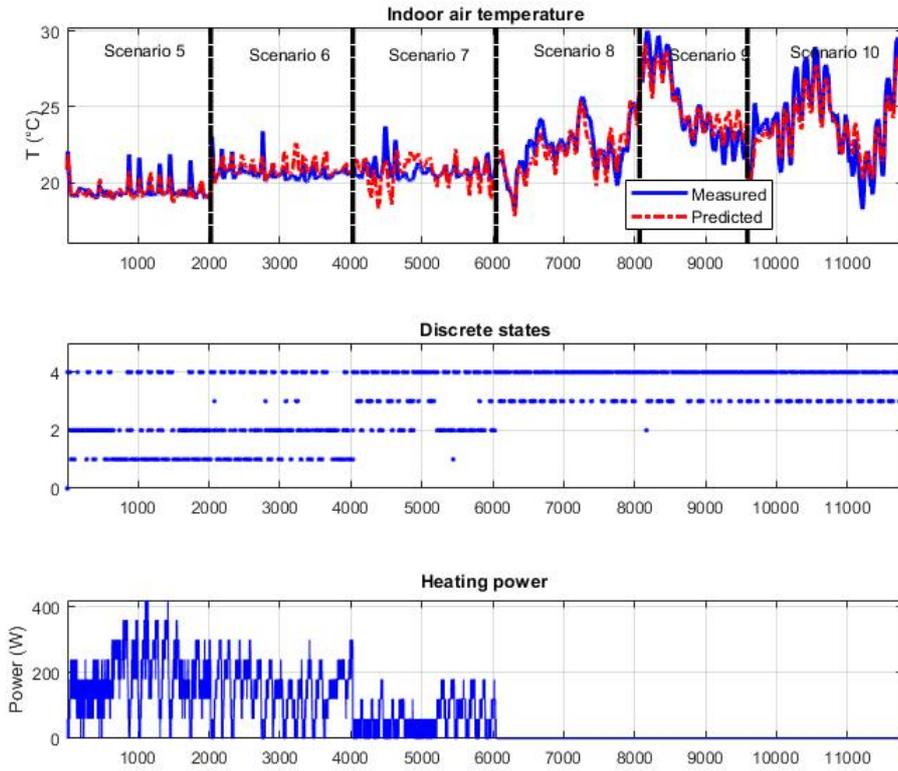


FIGURE 4.8: Validation after insulation

a predefined scenario. The switching between different controllers is based on the discrete state determined by the SVM classification.

TABLE 4.2: PID controllers parameters for each sub-model

<i>Sub-model</i>	K_p	K_i	K_d	T_f
1 before ins.	1720	0.0948	$4.2 * 10^6$	685
2 before ins.	$9.43 * 10^4$	4.93	$-5.87 * 10^7$	954
3 before ins.	-3580	-0.497	$-4.05 * 10^6$	1090
4 before ins.	1860	0.259	$2.27 * 10^6$	1230
1 after ins.	1270	0.107	$2.84 * 10^6$	985
2 after ins.	870	0.128	$-2.78 * 10^4$	333
3 after ins.	369	0.012	$2.66 * 10^6$	858
4 after ins.	502	0.0258	$1.25 * 10^6$	760

The details on the parameters of the controllers found for all the sub-models before and after insulation are given in Table 4.2 (before ins.: before insulation; after ins.: after insulation). The controllers are then integrated in the closed loop (Fig. 4.3) in order to estimate the heating power required for maintaining the temperature in the desired rang defined by the scenario. This range is defined by the user, it could be for example the scenario set by the regulations. In this case, the set point temperature in the scenario is fixed to 21°C between 6 AM and 8 PM and 18°C for the rest of the time.

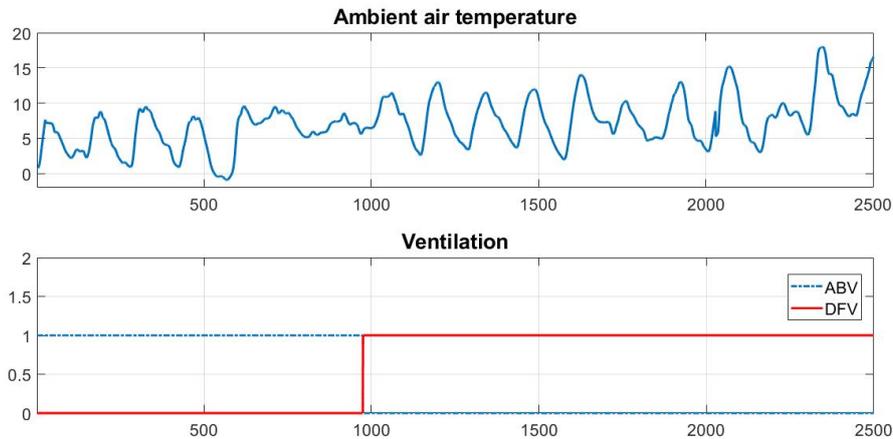


FIGURE 4.9: Ambient air temperature (used as input) and ventilation mode for the simulated scenario

The heat input provided by the command is limited to a minimum of 0W and a maximum of 1500W that corresponds to the maximum heating power that could be delivered by the radiators. The ambient air temperature and the ventilation mode for the simulated scenario are shown in Fig. 4.9. In order to assess the improvement in the energy consumption brought by the insulation work, the same scenario is applied for the models found before and after insulation and the same inputs are given to both loops. The scenario is chosen to cover the two ventilation modes (as shown in Fig. 4.9). The control results obtained before and after insulation are respectively shown in Fig. 4.10 and Fig. 4.11.

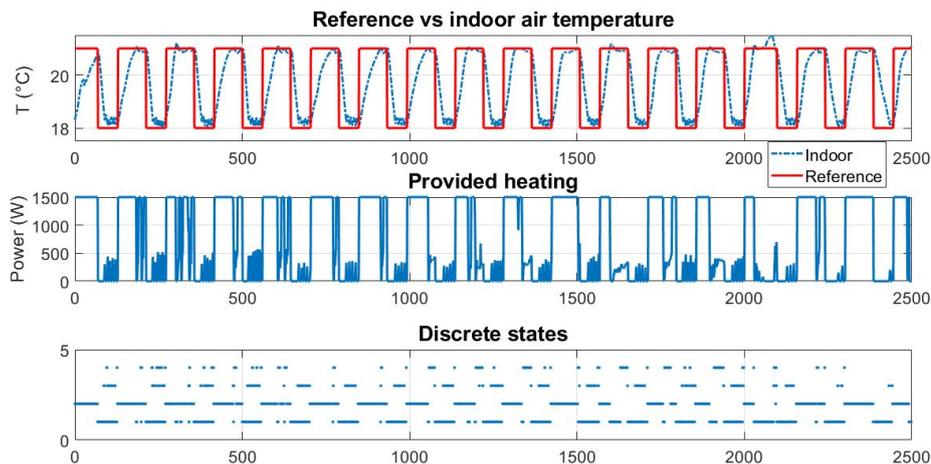


FIGURE 4.10: Control results before insulation

Before insulation, the heating provided inside the building is quickly dissipated through the construction. Therefore, the temperature takes a longer time to attain its set point compared to the case after insulation, where the heating is preserved inside the zone for a longer time and thus the temperature attains its set point faster. In the same way, before insulation, the temperature quickly drops after the heating is turned off, which is not the case after insulation. One can also notice that the

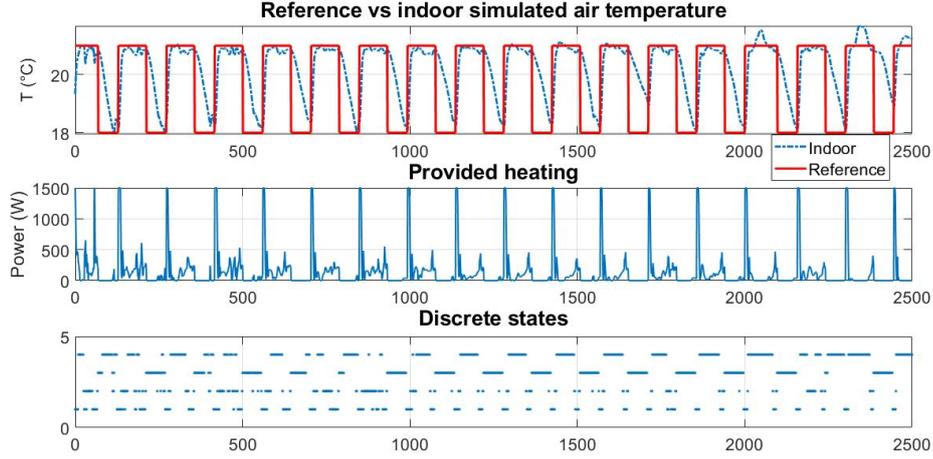


FIGURE 4.11: Control results after insulation

control for the installed heaters is performing better (close to reference temperature) after insulation while before insulation, the maximum heating power installed is not sufficient to quickly attain the 21°C maximum set temperature.

TABLE 4.3: Energy performance gain obtained by our approach and classical physics approaches

	<i>Before insulation</i>	<i>After insulation</i>	<i>Gain</i>
Our approach (kWh)	295	60	79%
HTC (W/K)	38.3	9.6	75%

The total energy consumed during this 17 days scenario (Fig. 4.9) is calculated before and after insulation.

In order to evaluate the accuracy of this methodology, the reduction in the overall HTC of the dry zone of the building was estimated using physical calculation based on the properties of the installed insulation materials. The HTC is calculated ¹ before and after insulation using the following equation:

$$HTC = \sum(S_i * U_i * b_i) + \sum(\psi_i * L_i * b_i). \quad (4.11)$$

The first part of the equation represents the heat losses through the envelope where $S_i(m^2)$ is the surface area of wall i , $U_i(W/m^2.K)$ is its heat transfer coefficient that depends on the thermal properties of the wall materials as indicated by the manufacturers. The second part represents the heat losses through thermal bridges ², where $L_i(m)$ is the length of the thermal bridge, $\psi_i(W/m.K)$ represents the heat loss coefficient of the linear bridge, it is estimated based on the calculation rules

¹These calculations were performed by the CEREMA (Centre d'Etudes et d'expertise sur les Eisques, l'Environnement, la Mobilité et l'Aménagement) center located in Angers, France, where the test cell is built.

²A thermal bridge is an area of a building construction which has a significantly higher heat transfer than the surrounding materials due to a break in the insulation, assembly junctions (roof to wall, wall to floor, etc.), corners, etc. Thermal bridges result in additional heat losses.

indicated in the French thermal regulation. Lastly, b_i is a coefficient related to the adjacent space, for instance, if the wall is exposed to the outdoor environment, $b_i = 1$ while if it is exposed to a heated area, b_i is close to zero. This is the conventional method defined by the French thermal regulations for calculating the overall HTC of an existing building [Th-Bât]. It does not take into consideration losses due to ventilation or air infiltration. This will lead to a certain difference between the value estimated by this method and the one estimated by the proposed methodology. This latter being based on data measured in the test cell, the losses through ventilation or infiltration are embedded in the model parameters.

The obtained values are shown in Table 4.3. The gain estimated by physics based methods is 75% which is close to the value of 79% found by the proposed methodology.

The described methodology is therefore suitable for estimating the gain in the energy performance of a certain building provided the availability of measurements. The estimation of this gain using the HTC method requires detailed information about the construction of the building. The advantage that the proposed data-based methodology presents is that it provides the amount of energy consumed during a certain period of time for a defined scenario. It therefore can be used for assessing the energy performance of a building for standard scenarios.

4.5 Conclusion

In this chapter, a methodology for evaluating the energy performance of buildings based on data driven techniques is presented. It relies on the PWARX hybrid system identification technique presented in Sect. 2.4 for modeling the indoor air temperature of the building combined with the SVM classification technique and integrating a hybrid PID controller. The methodology was applied on a test cell located in Angers, France, using data from experiments before and after insulation work. The results showed that the reduction in the energy consumption of the test cell due to renovation was accurately and quickly estimated using an automated technique based entirely on real measurements.

This methodology can therefore be used to assess the energy performance of a building by simulating it in a closed loop for standard scenarios. The methodology relies on measurements collected from the building without any information about its physics. Therefore, it is a quick way of estimating the energy performance of existing buildings provided the availability of real measurements. It can also be used to test different control strategies and choose the most efficient ones in terms of occupants' comfort and energy savings. The control in this test case was only applied on the heating power, however, it can also be applied for other inputs like ventilation rate.

Conclusion and future work

Having an accurate model to describe the thermal dynamics of a building is needed for implementing actions aiming at reducing the energy consumption of the building sector and its environmental impact. For this purpose, building thermal modeling has seen important developments for the past 50 years. Various modeling approaches have been proposed, some based on physical equations describing the heat, air and humidity transfer between the building and its environment, others based on mathematical equations constructed entirely from measurements collected from the building during operation. This thesis presents an original application of a data-driven technique for modeling the thermal behavior of buildings. The novelty of the proposed approach lies in using a hybrid system representation, a system that exhibits both continuous and discrete dynamics, and in treating sudden events like opening the windows or turning the heating on as discrete events affecting the temperature dynamics which generate a new continuous model.

Chapter 1 presented the three main modeling categories, namely, "white box", "black box" and "grey box". This chapter detailed the techniques used within each category along with their advantages and limitations, as well as a number of publications employing them. It was concluded that even though "white box" models have been historically adopted for representing the thermal behavior of a building system, discrepancies between the predicted behavior and the actual one has been noticed throughout the years. This is due to three main factors:

- The parameters related to the building envelope and systems given to the model are not accurate due to multiple reasons like the fact that the builder did not respect the specifications or the materials properties have changed over time, etc.
- The hypotheses and simplifications adopted by the model disregard important factors such as air infiltration or the latent heat transfer.
- The occupants' behavior considered in the model highly differs from the actual one.

To overcome these limitations, researchers have proposed to use complementary approaches such as the ones based on data measured from the building as they eliminate some of the error sources. The main limitation of data-driven models is their need of a rich amount of data that is representative of the behavior of the building for different conditions. Another limitation of most of the data-driven models is the usual assumption of a linear model structure that is not representative of the complex phenomena occurring in the building. This thesis addresses this last issue by proposing a nonlinear model with a hybrid structure, combining linear continuous dynamics with discrete events corresponding to the linearization of a nonlinear model around operating points.

The thesis considers a special class of data-driven techniques used for the identification of hybrid systems called PWARX [Boukharouba et al., 2009]. Such systems are a collection of affine sub-models where each sub-model is an ARX equation representing a certain configuration of the building characterized by its particular dynamics. The use of a switching model is mathematically interpreted and motivated in Chapter 2 using equations derived from the lumped RC thermal circuit of a typical building. This first contribution shows how the events that occur in a building like opening/closing windows or shutters, changing the heating mode or the occupancy level can lead to a change in the structure and parameters of the transfer function representing the system. This leads to a set of local transfer functions changing according to the configuration, thus supporting the proposal made in this thesis that consists in having a different model for each configuration. In the second part of this same chapter, the PWARX algorithm is presented in details. It is based on a clustering and an identification procedure that occur simultaneously. The goal of the clustering step is to group similar data points in one group and then for the identification part, the least squares method is applied to the data in each cluster to find the representative model coefficients. As a result, several sub-models are obtained where each one represents particular dynamics in the building. An important aspect in this methodology is the fact that the number of sub-models is not known a priori, which corresponds to the type of application considered in this thesis as we do not have previous knowledge about the configurations that might occur in a building. The number of sub-models is identified by the technique based on the dynamics represented by the data.

The PWARX technique is used in this thesis to model the indoor air temperature of the building. The validation of this methodology represents the second contribution of the thesis and is presented in Chapter 3 for several test cases. Data collected from different types of buildings, ranging from a simple case of a small unoccupied room to an occupied residential building, was used to test the performance of the PWARX models in predicting the indoor air temperature. The results showed the capability of the PWARX technique to detect different thermal behavior in a building. These are due to events leading to different dynamics as the ones identified in Chapter 2 like turning the heating ON/OFF, increasing the ventilation rate (opening windows), changing the occupancy, or to a certain range of important factors like the solar radiation. The results also showed that the prediction of the indoor air temperature using PWARX models is more accurate than the prediction using classical ARX models. A clear out-performance of the PWARX models was noticed when the previously mentioned events become more frequent. Due to the large number of inputs and influencing factors for the case of the occupied building, it was hard to determine the cause behind each switching instance, thus an automated classification technique was proposed to determine the switching rule. Sect. 3.6.2 showed how to use the SVM classification technique for assigning new data to the right region allowing to use the right sub-model for the prediction phase. Some sub-models are detected for a specific season, while others were dominant throughout a whole. The prediction for this test case was performed for several time horizons (k-steps ahead). The results show that when the prediction horizon is very small (5 min), both ARX and PWARX give an accurate prediction, however, as the prediction horizon increases, the PWARX outperforms the ARX.

The last contribution was detailed in Chapter 4, where it was shown how to extend the methodology to predict the energy consumption of a building subject

to standard or well defined scenarios. This is useful for having a quick evaluation of the energy performance of an existing building. The implementation of this methodology was done by integrating the PWARX model and the SVM classification technique in a hybrid control loop that uses PID controllers. It was validated using data collected before and after the renovation of a building located in France, the goal being to predict the gain in its energy performance. The results showed that this methodology was able to accurately estimate the gain in the energy performance due to renovation work. In fact, a good agreement was found between the gain estimated by the proposed methodology and the reduction in the overall HTC coefficient calculated using physical equations. This has led us to conclude that the developed methodology can be used to assess the energy performance of a building subject to standard scenarios.

This thesis was a feasibility study of the potential of using PWARX for the thermal modeling and the prediction of the energy consumption of buildings. Therefore, the most straightforward solutions available for the classification and control were used, but other more advanced ones could yield better results. The results presented in this thesis can be pursued further in several directions.

- The results obtained with the PWARX technique can be benchmarked against a "white box" model for the same test case. For this purpose, data measured for the building as well as a model of the same building using a BES tool are needed.
- The methodology proposed in the last chapter can be used for the evaluation of the energy performance of existing buildings for standard scenarios, provided the availability of measurements. This could be useful for benchmarking the energy performance of existing buildings with respect to the existing regulations. This type of information could be useful for deciding on strategies for the renovation of the park.
- The strength of this methodology lies in its ability to detect any change in the behavior of the indoor air temperature in a building. The offline phase in which the sub-models are identified is done already through the estimation phase. Therefore, a potential application would be to use this methodology for the online prediction where data measured in the building is continuously fed to the algorithm that detects any change in the behavior, assigns data to the corresponding sub-model and predicts with a high accuracy the indoor air temperature for a near horizon. Since the switching rule relies on the regression data that contains previous values of the output, then having measured values of this output will increase the accuracy of the classification and hence the prediction. The sub-models could also be updated after a certain period of time using the collected measurements. This helps adapting the model to the changes that might be occurring in the building either due to the tenants' behavior or weather conditions.
- The PID hybrid control loop was used offline for estimating the energy performance of buildings in the last chapter. The control loop could be used for the online prediction and control by using data measured from the system. This contributes to the implementation of better control strategies which enhances the performance and energy efficiency of the system. In this context, the use of model predictive control (MPC) is interesting to investigate. MPC has already

shown good performance in terms of the energy efficiency of the building and the occupants' comfort [Afram and Janabi-Sharifi, 2014]. Since it relies on the system model, it would be as effective as the model representing the behavior of the building, and the PWARX has shown good performance for predicting future behavior of the temperature. MPC uses measurements from the system in order to decide on future control actions. Combining the PWARX online prediction with MPC could lead to the implementation of better control strategies and contribute to ensuring the comfort of the occupants and the energy efficiency of the building.

- For the prediction phase, the classification is of high importance. The PWARX technique seems to be very effective in detecting different behavior and modeling them but new data should be assigned to the right region to have an accurate prediction. Therefore, further work could be done for the classification part. SVM is one of the most popular and performing classification techniques used in the literature. A compromise could be defined between the use of SVM automated classification rules and some experts rules based on the physical knowledge of the building and the phenomena occurring in it and with its environment.
- The algorithm used in this thesis supposes that the sub-models have the same structure in terms of the orders n_a , n_b and the delay n_k . In the future, models with different structures for each sub-model could be investigated. In fact, the interpretation in Sect. 2.2.1 and 2.2.2 shows that the structure of the model changes from one configuration to the other. An appropriate model structure could be defined using physics rules, the model could then be reduced using appropriate techniques, and its order could be given to the PWARX algorithm.

The PWARX approach presented in this thesis could also be used in favor of other modeling techniques. Detecting different thermal behavior could be, for example, used for adapting the parameters of a building RC thermal network for each configuration, thus increasing the accuracy of the prediction.

Bibliography

- [Afram and Janabi-Sharifi 2014] AFRAM, Abdul ; JANABI-SHARIFI, Farrokh: Theory and applications of HVAC control systems – A review of model predictive control (MPC). In: *Building and Environment* 72 (2014), S. 343 – 355. – URL <http://www.sciencedirect.com/science/article/pii/S0360132313003363>. – ISSN 0360-1323
- [Ajib et al. 2017] AJIB, Balsam ; LEFTERIU, Sanda ; CAUCHETEUX, Antoine ; LECOEUICHE, Stéphane: Building thermal modeling using a hybrid system approach. In: *Preprints of the 20th World Congress The International Federation of Automatic Control IFAC*, URL <https://hal.archives-ouvertes.fr/hal-01720374>, 2017
- [Ajib et al. 2018a] AJIB, Balsam ; LEFTERIU, Sanda ; CAUCHETEUX, Antoine ; LECOEUICHE, Stéphane: Predicting the air temperature of a building zone by detecting different configurations using a switched system identification technique. Submitted to *Building and Environment*. (2018)
- [Ajib et al. 2018b] AJIB, Balsam ; LEFTERIU, Sanda ; CAUCHETEUX, Antoine ; LECOEUICHE, Stéphane ; GAUVRIT, Jordan: Prediction of standardized energy consumption of existing buildings based on hybrid systems modeling and control. In: *57th IEEE Conference on Decision and Control (CDC), December, 2018*
- [Ajib et al. 2016] AJIB, Balsam ; LEFTERIU, Sanda ; CAUCHETEUX, Antoine ; TITTELEIN, Pierre ; LECOEUICHE, Stéphane: Building thermal modeling using switching systems. In: *International Building Performance Simulation Association IBPSA, Champs-sur-Marne, France, 2016*
- [Al-Saadi and Zhai 2015] AL-SAAD, Saleh N. ; ZHAI, Zhiqiang: A new validated TRNSYS module for simulating latent heat storage walls. In: *Energy and Buildings* 109 (2015), S. 274 – 290. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815303273>. – ISSN 0378-7788
- [Allard and Inard 1992] ALLARD, F ; INARD, C: Natural and mixed convection in rooms: prediction of thermal stratification and heat transfer by zonal models. In: *Proceedings of ISRACVE, 1992*, S. 335–342
- [Amasyali and El-Gohary 2018] AMASYALI, Kadir ; EL-GOHARY, Nora M.: A review of data-driven building energy consumption prediction studies. In: *Renewable and Sustainable Energy Reviews* 81 (2018), S. 1192 – 1205. – URL <http://www.sciencedirect.com/science/article/pii/S1364032117306093>. – ISSN 1364-0321
- [Amayri et al. 2016] AMAYRI, Manar ; ARORA, Abhay ; PLOIX, Stephane ; BANDHYOPADYAY, Sanghamitra ; NGO, Quoc-Dung ; BADARLA, Venkata R.: Estimating occupancy in heterogeneous sensor environment. In: *Energy and Buildings* 129 (2016), S. 46 – 58. – ISSN 0378-7788

- [Amiri et al. 2015] AMIRI, Shideh S. ; MOTTAHEDI, Mohammad ; ASADI, So-mayeh: Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S. In: *Energy and Buildings* 109 (2015), S. 209 – 216. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815303133>. – ISSN 0378-7788
- [Andelkovic et al. 2016] ANDELKOVIC, Aleksandar S. ; MUJAN, Igor ; DAKIC, Stojanka: Experimental validation of a EnergyPlus model: Application of a multi-storey naturally ventilated double skin façade. In: *Energy and Buildings* 118 (2016), S. 27 – 36. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816301037>. – ISSN 0378-7788
- [Andersen et al. 2000] ANDERSEN, Klaus K. ; MADSEN, Henrik ; HANSEN, Lars H.: Modelling the heat dynamics of a building using stochastic differential equations. In: *Energy and Buildings* 31 (2000), Nr. 1, S. 13 – 24. – URL <http://www.sciencedirect.com/science/article/pii/S0378778898000693>. – ISSN 0378-7788
- [Arregi and Garay 2017] ARREGI, Beñat ; GARAY, Roberto: Regression analysis of the energy consumption of tertiary buildings. In: *Energy Procedia* 122 (2017), S. 9 – 14. – URL <http://www.sciencedirect.com/science/article/pii/S1876610217328886>. – CISBAT 2017 International Conference Future Buildings Districts – Energy Efficiency from Nano to Urban Scale. – ISSN 1876-6102
- [Åström and Hägglund 1995] ÅSTRÖM, Karl J. ; HÄGGLUND, Tore: *PID controllers: theory, design, and tuning*. Bd. 2. Instrument society of America Research Triangle Park, NC, 1995
- [Bacher and Madsen 2011] BACHER, Peder ; MADSEN, Henrik: Identifying suitable models for the heat dynamics of buildings. In: *Energy and Buildings* 43 (2011), Nr. 7, S. 1511 – 1522. – URL <http://www.sciencedirect.com/science/article/pii/S0378778811000491>. – ISSN 0378-7788
- [Bako 2008] BAKO, Laurent: *Contribution à l'identification de systèmes dynamiques hybrides*, Université des Sciences et Technologie de Lille-Lille I, Dissertation, 2008
- [Bauer and Scartezzini 1998] BAUER, M. ; SCARTEZZINI, J.-L.: A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings. In: *Energy and Buildings* 27 (1998), Nr. 2, S. 147 – 154. – URL <http://www.sciencedirect.com/science/article/pii/S0378778897000352>. – ISSN 0378-7788
- [Bemporad et al. 2000] BEMPORAD, Alberto ; FERRARI-TRECCATE, Giancarlo ; MORARI, Manfred: Observability and controllability of piecewise affine and hybrid systems. In: *IEEE transactions on automatic control* 45 (2000), Nr. 10, S. 1864–1876
- [Bemporad and Morari 1999a] BEMPORAD, Alberto ; MORARI, Manfred: Control of systems integrating logic, dynamics, and constraints. In: *Automatica* 35 (1999), Nr. 3, S. 407–427
- [Bemporad and Morari 1999b] BEMPORAD, Alberto ; MORARI, Manfred: Control of systems integrating logic, dynamics, and constraints. In: *Automatica* 35 (1999), Nr. 3, S. 407 – 427. – ISSN 0005-1098

- [Bouia and Dalicieux 1991] BOUIA, H ; DALICIEUX, P: Simplified modeling of air movements inside dwelling room. In: *Proceedings of Building Simulation'91 Conference*, 1991, S. 106–110
- [Boukharouba et al. 2010] BOUKHAROUBA, K ; DUVELLA, E ; BAKO, L ; LECOEUICHE, S: Multimodeling vs piecewise affine modeling for the identification of open channel systems. In: *Large Scale Complex Systems Theory and Applications* Bd. 9, 2010, S. 474–479
- [Boukharouba 2011] BOUKHAROUBA, Khaled: *Modélisation et classification de comportements dynamiques des systemes hybrides*, Université Lille 1, Dissertation, 2011
- [Boukharouba et al. 2009] BOUKHAROUBA, Khaled ; BAKO, Laurent ; LECOEUICHE, StÅ et al.: Identification of piecewise affine systems based on dempster-shafer theory. In: *System Identification* Bd. 15, 2009, S. 1662–1667
- [Boyer et al. 1996] BOYER, H. ; CHABRIAT, J.P. ; GRONDIN-PEREZ, B. ; TOURRAND, C. ; BRAU, J.: Thermal building simulation and computer generation of nodal models. In: *Building and Environment* 31 (1996), Nr. 3, S. 207 – 214. – URL <http://www.sciencedirect.com/science/article/pii/0360132396000017>. – ISSN 0360-1323
- [Bozonnet et al. 2005] BOZONNET, Emmanuel ; BELARBI, Rafik ; ALLARD, Francis: Modelling solar effects on the heat and mass transfer in a street canyon, a simplified approach. In: *Solar Energy* 79 (2005), Nr. 1, S. 10–24
- [Brastein et al. 2018] BRASTEIN, O.M. ; PERERA, D.W.U. ; PFEIFER, C. ; SKEIE, N.-O.: Parameter estimation for grey-box models of building thermal behaviour. In: *Energy and Buildings* 169 (2018), S. 58 – 68. – URL <http://www.sciencedirect.com/science/article/pii/S0378778817331791>. – ISSN 0378-7788
- [Burhenne et al. 2013] BURHENNE, Sebastian ; WYSTRCIL, Dominik ; ELCI, Mehmet ; NARMSARA, Sattaya ; HERKEL, Sebastian: Building performance simulation using Modelica: analysis of the current state and application areas. In: *13th International Conference of the International Building Performance Simulation Association*, 2013
- [Canty and O'Mahony 2009] CANTY, Niel ; O'MAHONY, Tom: Design considerations for piecewise affine system identification of nonlinear systems. In: *Control and Automation, 2009. MED'09. 17th Mediterranean Conference on IEEE* (Veranst.), 2009, S. 157–162
- [Canyurt et al. 2005] CANYURT, Olcay E. ; OZTURK, Harun K. ; HEPBASLI, Arif ; UTLU, Zafer: Estimating the Turkish residential–commercial energy output based on genetic algorithm (GA) approaches. In: *Energy Policy* 33 (2005), Nr. 8, S. 1011 – 1019. – URL <http://www.sciencedirect.com/science/article/pii/S030142150300332X>. – ISSN 0301-4215
- [Caucheteux et al. 2013] CAUCHETEUX, A ; SABAR, A E. ; BOUCHER, V: Occupancy measurement in building: A litterature review, application on an energy efficiency research demonstrated building. In: *International Journal of Metrology and Quality Engineering* 4 (2013), Nr. 2, S. 135–144

- [Caucheteux et al. 2016] CAUCHETEUX, Antoine ; AMAYRI, Manar ; PLOIX, Stéphane: Mesure de l'occupation pour l'évaluation de la performance énergétique des bâtiments : plan de mesures et incertitudes. In: *International Building Performance Simulation Association IBPSA, Champs-sur-Marne, France, 2016*
- [CHAM 2005] CHAM: 2005. – <http://www.cham.co.uk/ChmSupport/tr000.php>
- [Chang and Lin 2011] CHANG, Chih-Chung ; LIN, Chih-Jen: LIBSVM: a library for support vector machines. In: *ACM transactions on intelligent systems and technology (TIST)* 2 (2011), Nr. 3, S. 27
- [Clarke and McLean 1988] CLARKE, JA ; MCLEAN, D: ESP-A building and plant energy simulation system. In: *Strathclyde: Energy Simulation Research Unit, University of Strathclyde* (1988)
- [Cortes and Vapnik 1995] CORTES, Corinna ; VAPNIK, Vladimir: Support-vector networks. In: *Machine learning* 20 (1995), Nr. 3, S. 273–297
- [Costa et al. 2006] COSTA, Oswaldo Luiz V. ; FRAGOSO, Marcelo D. ; MARQUES, Ricardo P.: *Discrete-time Markov jump linear systems*. Springer Science & Business Media, 2006
- [Crabb et al. 1987] CRABB, JA ; MURDOCH, N ; PENMAN, JM: A simplified thermal response model. In: *Building Services Engineering Research and Technology* 8 (1987), Nr. 1, S. 13–19
- [Crawley et al. 2001] CRAWLEY, Drury B. ; LAWRIE, Linda K. ; WINKELMANN, Frederick C. ; BUHL, Walter F. ; HUANG, Y J. ; PEDERSEN, Curtis O. ; STRAND, Richard K. ; LIESEN, Richard J. ; FISHER, Daniel E. ; WITTE, Michael J. et al.: EnergyPlus: creating a new-generation building energy simulation program. In: *Energy and buildings* 33 (2001), Nr. 4, S. 319–331
- [Dahanayake and Chow 2017] DAHANAYAKE, K.W.D. Kalani C. ; CHOW, Cheuk L.: Studying the potential of energy saving through vertical greenery systems: Using EnergyPlus simulation program. In: *Energy and Buildings* 138 (2017), S. 47 – 59. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816317480>. – ISSN 0378-7788
- [De Schutter and Van den Boom 2001] DE SCHUTTER, B ; BOOM, TJJ Van den: Model predictive control for max-min-plus-scaling systems. In: *American Control Conference, 2001. Proceedings of the 2001* Bd. 1 IEEE (Veranst.), 2001, S. 319–324
- [De Schutter 2000] DE SCHUTTER, Bart: Optimal control of a class of linear hybrid systems with saturation. In: *SIAM Journal on Control and Optimization* 39 (2000), Nr. 3, S. 835–851
- [Deb et al. 2018] DEB, Chirag ; LEE, Siew E. ; SANTAMOURIS, Mattheos: Using artificial neural networks to assess HVAC related energy saving in retrofitted office buildings. In: *Solar Energy* 163 (2018), S. 32 – 44. – URL <http://www.sciencedirect.com/science/article/pii/S0038092X18300963>. – ISSN 0038-092X
- [Do Valle Costa et al. 2012] DO VALLE COSTA, Oswaldo L. ; FRAGOSO, Marcelo D. ; TODOROV, Marcos G.: *Continuous-time Markov jump linear systems*. Springer Science & Business Media, 2012

- [Dong et al. 2005] DONG, Bing ; CAO, Cheng ; LEE, Siew E.: Applying support vector machines to predict building energy consumption in tropical region. In: *Energy and Buildings* 37 (2005), Nr. 5, S. 545 – 553. – URL <http://www.sciencedirect.com/science/article/pii/S0378778804002981>. – ISSN 0378-7788
- [Doucet et al. 2001] DOUCET, Arnaud ; GORDON, Neil J. ; KRISHNAMURTHY, Vikram: Particle filters for state estimation of jump Markov linear systems. In: *IEEE Transactions on signal processing* 49 (2001), Nr. 3, S. 613–624
- [Dymola and Catia 2015] DYMOLA ; CATIA: *version 2015*. Dassault systemes, 2015
- [EPBD 2010] EPBD: On the energy performance of buildings. In: *Official journal of the European Union, Directive 2010/31/EU of the European Parliament and of the Council* (2010)
- [EUROPA] EUROPA: . – <https://ec.europa.eu/energy/en/eu-buildings-factsheets>
- [EUROPA 2012] EUROPA: Directive 2012/27/EU on Energy Efficiency, Amending Directives 2009/125/EC and 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC. OJ L315, 14 November 2012. In: *Official Journal of the European Union* (2012), Nr. 06. – <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=EN>
- [European Commission 2011] EUROPEAN COMMISSION: *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Energy Roadmap 2050*. 2011. – <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0885&from=EN>
- [Fang et al. 2017] FANG, Hong ; ZHAO, Dongliang ; TAN, Gang ; DENZER, Anthony: Study of underfloor air distribution using zonal model-based simulation and experimental measurements. In: *Energy and Buildings* 152 (2017), S. 96 – 107. – URL <http://www.sciencedirect.com/science/article/pii/S0378778817307272>. – ISSN 0378-7788
- [Fazenda et al. 2016] FAZENDA, Pedro ; LIMA, Pedro ; CARREIRA, Paulo: Context-based thermodynamic modeling of buildings spaces. In: *Energy and Buildings* 124 (2016), S. 164 – 177. – ISSN 0378-7788
- [Fels et al. 1986] FELS, Margaret F. et al.: PRISM: an introduction. In: *Energy and Buildings* 9 (1986), Nr. 1-2, S. 5–18
- [Ferracuti et al. 2017] FERRACUTI, Francesco ; FONTI, Alessandro ; CIABATTONI, Lucio ; PIZZUTI, Stefano ; ARTECONI, Alessia ; HELSEN, Lieve ; COMODI, Gabriele: Data-driven models for short-term thermal behaviour prediction in real buildings. In: *Applied Energy* 204 (2017), S. 1375 – 1387. – URL <http://www.sciencedirect.com/science/article/pii/S0306261917305032>. – ISSN 0306-2619
- [Fluent 2012] FLUENT: 2012. – <http://www.ansys.com/Products/Fluids/ANSYS-Fluent>
- [Fogel et al. 1975] FOGEL, LJ ; OWENS, AJ ; WALSH, MJ: *Adaptation in natural and artificial systems*. 1975

- [Foucquier et al. 2013] FOUCQUIER, Aurélie ; ROBERT, Sylvain ; SUARD, Frédéric ; STÉPHAN, Louis ; JAY, Arnaud: State of the art in building modelling and energy performances prediction: A review. In: *Renewable and Sustainable Energy Reviews* 23 (2013), S. 272 – 288. – URL <http://www.sciencedirect.com/science/article/pii/S1364032113001536>. – ISSN 1364-0321
- [Fraisie et al. 2002] FRAISSE, Gilles ; VIARDOT, Christelle ; LAFABRIE, Olivier ; ACHARD, Gilbert: Development of a simplified and accurate building model based on electrical analogy. In: *Energy and Buildings* 34 (2002), Nr. 10, S. 1017 – 1031. – ISSN 0378-7788
- [Freire et al. 2008] FREIRE, Roberto Z. ; OLIVEIRA, Gustavo H. ; MENDES, Nathan: Development of regression equations for predicting energy and hygrothermal performance of buildings. In: *Energy and Buildings* 40 (2008), Nr. 5, S. 810 – 820. – URL <http://www.sciencedirect.com/science/article/pii/S0378778807001739>. – ISSN 0378-7788
- [Fu et al. 2015] FU, Yangyang ; LI, Zhengwei ; ZHANG, Hao ; XU, Peng: Using Support Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices. In: *Procedia Engineering* 121 (2015), S. 1016 – 1022. – URL <http://www.sciencedirect.com/science/article/pii/S1877705815029252>. – The 9th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC) joint with the 3rd International Conference on Building Energy and Environment (COBEE), 12-15 July 2015, Tianjin, China. – ISSN 1877-7058
- [Gao and Lee 2011] GAO, C.F. ; LEE, W.L.: Evaluating the influence of openings configuration on natural ventilation performance of residential units in Hong Kong. In: *Building and Environment* 46 (2011), Nr. 4, S. 961 – 969. – URL <http://www.sciencedirect.com/science/article/pii/S0360132310003215>. – ISSN 0360-1323
- [Geng et al. 2018] GENG, Yang ; JI, Wenjie ; LIN, Borong ; HONG, Jiajie ; ZHU, Yingxin: Building energy performance diagnosis using energy bills and weather data. In: *Energy and Buildings* (2018), S. –. – URL <https://www.sciencedirect.com/science/article/pii/S0378778817340690>. – ISSN 0378-7788
- [Goldberg 2014] GOLDBERG, DE: Genetic algorithms in search, optimization, and machine learning, addison-wesley, reading, ma, 1989. In: *Google Scholar* (2014)
- [Goyal and Barooah 2012] GOYAL, Siddharth ; BAROOAH, Prabir: A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. In: *Energy and Buildings* 47 (2012), S. 332 – 340. – URL <http://www.sciencedirect.com/science/article/pii/S0378778811006001>. – ISSN 0378-7788
- [Guerassimoff and Thomas 2015] GUERASSIMOFF, Gilles ; THOMAS, Johann: Enhancing energy efficiency and technical and marketing tools to change people's habits in the long-term. In: *Energy and Buildings* 104 (2015), S. 14 – 24. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815300931>. – ISSN 0378-7788
- [Guo et al. 2015] GUO, Weihong ; LIU, Xiao ; YUAN, Xu: Study on Natural Ventilation Design Optimization Based on CFD Simulation for Green Buildings. In: *Procedia Engineering* 121 (2015), S. 573 – 581. – URL <http://www.sciencedirect.com/science/article/pii/S1877705815029252>.

- [com/science/article/pii/S1877705815027642](http://www.sciencedirect.com/science/article/pii/S1877705815027642). – The 9th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC) joint with the 3rd International Conference on Building Energy and Environment (COBEE), 12-15 July 2015, Tianjin, China. – ISSN 1877-7058
- [Haghighat et al. 2001] HAGHIGHAT, Fariborz ; LI, Yin ; MEGRI, Ahmed C.: Development and validation of a zonal model—POMA. In: *Building and environment* 36 (2001), Nr. 9, S. 1039–1047
- [Harb et al. 2016] HARB, Hassan ; BOYANOV, Neven ; HERNANDEZ, Luis ; STREBLOW, Rita ; MÜLLER, Dirk: Development and validation of grey-box models for forecasting the thermal response of occupied buildings. In: *Energy and Buildings* 117 (2016), S. 199 – 207. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816300688>. – ISSN 0378-7788
- [Harish and Kumar 2016] HARISH, V.S.K.V. ; KUMAR, Arun: A review on modeling and simulation of building energy systems. In: *Renewable and Sustainable Energy Reviews* 56 (2016), S. 1272 – 1292. – URL <http://www.sciencedirect.com/science/article/pii/S1364032115014239>. – ISSN 1364-0321
- [Hasan et al. 2014] HASAN, Ola A. ; DEFER, Didier ; SHAHROUR, Isam: A simplified building thermal model for the optimization of energy consumption: Use of a random number generator. In: *Energy and Buildings* 82 (2014), S. 322 – 329. – URL <http://www.sciencedirect.com/science/article/pii/S037877881400560X>. – ISSN 0378-7788
- [Hilliaho et al. 2015] HILLIAHO, Kimmo ; LAHDENSIVU, Jukka ; VINHA, Juha: Glazed space thermal simulation with IDA-ICE 4.61 software—Suitability analysis with case study. In: *Energy and Buildings* 89 (2015), S. 132 – 141. – URL <http://www.sciencedirect.com/science/article/pii/S0378778814011141>. – ISSN 0378-7788
- [Hong et al. 2012] HONG, Taehoon ; KOO, Choongwan ; JEONG, Kwangbok: A decision support model for reducing electric energy consumption in elementary school facilities. In: *Applied Energy* 95 (2012), S. 253 – 266. – URL <http://www.sciencedirect.com/science/article/pii/S0306261912001511>. – ISSN 0306-2619
- [Hou et al. 2006] HOU, Zhijian ; LIAN, Zhiwei ; YAO, Ye ; YUAN, Xinjian: Data mining based sensor fault diagnosis and validation for building air conditioning system. In: *Energy Conversion and Management* 47 (2006), Nr. 15, S. 2479 – 2490. – URL <http://www.sciencedirect.com/science/article/pii/S0196890405003031>. – ISSN 0196-8904
- [Huang et al. 2005] HUANG, Hongyu ; HAGHIGHAT, Fariborz ; LEE, Chang-Seo: An Integrated Zonal Model for Predicting Indoor Airflow, Temperature, and VOC Distributions. In: *ASHRAE Transactions* 111 (2005), Nr. 1
- [IDA-ICE 2014] IDA-ICE, IDA: *indoor climate and energy 4.1*. 2014
- [Imura et al. 2010] IMURA, Jun-Ichi ; KASHIMA, Kenji ; KUSANO, Masami ; IKEDA, Tsukasa ; MOROHOSHI, Tomohiro: Piecewise affine systems approach to control of biological networks. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 368 (2010), Nr. 1930, S. 4977–4993

- [Jiménez et al. 2008a] JIMÉNEZ, M.J. ; MADSEN, H. ; ANDERSEN, K.K.: Identification of the main thermal characteristics of building components using MATLAB. In: *Building and Environment* 43 (2008), Nr. 2, S. 170 – 180. – URL <http://www.sciencedirect.com/science/article/pii/S0360132306002976>. – Outdoor Testing, Analysis and Modelling of Building Components. – ISSN 0360-1323
- [Jiménez et al. 2008b] JIMÉNEZ, M.J. ; MADSEN, H. ; ANDERSEN, K.K.: Identification of the main thermal characteristics of building components using MATLAB. In: *Building and Environment* 43 (2008), Nr. 2, S. 170 – 180. – Outdoor Testing, Analysis and Modelling of Building Components. – ISSN 0360-1323
- [Joseph et al. 2015] JOSEPH, Cyril ; GEORGE, VI ; NAYAK, Poonam: Piecewise Affine Modeling of a Hybrid 3-Tank System. (2015)
- [Joseph et al. 2017] JOSEPH, Cyril ; KRISHNA, Bipin ; GEORGE, VI: Generation of Piecewise-Affine Model from A Mixed-Logic Dynamical Model of a 3-Tank System. In: *Journal of Advanced Research in Dynamical and Control Systems* 9 (2017), Nr. 2, S. 19–23
- [Kalogirou 2000] KALOGIROU, Soteris A.: Applications of artificial neural-networks for energy systems. In: *Applied Energy* 67 (2000), Nr. 1–2, S. 17 – 35. – URL <http://www.sciencedirect.com/science/article/pii/S0306261900000052>. – ISSN 0306-2619
- [Kalogirou 2006] KALOGIROU, Soteris A.: Artificial neural networks in energy applications in buildings. In: *International Journal of Low-Carbon Technologies* 1 (2006), Nr. 3, S. 201–216
- [Kalogirou and Bojic 2000] KALOGIROU, Soteris A. ; BOJIC, Milorad: Artificial neural networks for the prediction of the energy consumption of a passive solar building. In: *Energy* 25 (2000), Nr. 5, S. 479 – 491. – URL <http://www.sciencedirect.com/science/article/pii/S0360544299000869>. – ISSN 0360-5442
- [Kenai et al. 2018] KENAI, Mohamed-Amine ; LIBESSART, Laurent ; LASSUE, Stéphane ; DEFER, Didier: Impact of plants occultation on energy balance: Experimental study. In: *Energy and Buildings* 162 (2018), S. 208 – 218. – URL <http://www.sciencedirect.com/science/article/pii/S0378778817325495>. – ISSN 0378-7788
- [Kenjo et al. 2007] KENJO, Lana ; INARD, Christian ; CACCAVELLI, Dominique: Experimental and numerical study of thermal stratification in a mantle tank of a solar domestic hot water system. In: *Applied Thermal Engineering* 27 (2007), Nr. 11, S. 1986 – 1995. – URL <http://www.sciencedirect.com/science/article/pii/S1359431106004248>. – ISSN 1359-4311
- [Khayatian et al. 2016] KHAYATIAN, Fazel ; SARTO, Luca ; DALL’O’, Giuliano: Application of neural networks for evaluating energy performance certificates of residential buildings. In: *Energy and Buildings* 125 (2016), S. 45 – 54. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816303322>. – ISSN 0378-7788
- [Kolokotsa 2016] KOLOKOTSA, Dionysia: The role of smart grids in the building sector. In: *Energy and Buildings* 116 (2016), S. 703 – 708. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815005447>. – ISSN 0378-7788

- [Koulamasa et al. 2018] KOULAMASA, C. ; KALOGERAS, A.P. ; PACHECO-TORRES, R. ; CASILLAS, J. ; FERRARINI, L.: Suitability analysis of modeling and assessment approaches in energy efficiency in buildings. In: *Energy and Buildings* 158 (2018), S. 1662 – 1682. – URL <http://www.sciencedirect.com/science/article/pii/S0378778817305649>. – ISSN 0378-7788
- [Kumar et al. 2013] KUMAR, Rajesh ; AGGARWAL, R.K. ; SHARMA, J.D.: Energy analysis of a building using artificial neural network: A review. In: *Energy and Buildings* 65 (2013), S. 352 – 358. – URL <http://www.sciencedirect.com/science/article/pii/S0378778813003459>. – ISSN 0378-7788
- [Kuznik et al. 2011] KUZNIK, Frédéric ; CATALINA, Tiberiu ; GAUZERE, Lucie ; WOLOSZYN, Monika ; ROUX, Jean-Jacques: Numerical modelling of combined heat transfers in a double skin façade – Full-scale laboratory experiment validation. In: *Applied Thermal Engineering* 31 (2011), Nr. 14, S. 3043 – 3054. – URL <http://www.sciencedirect.com/science/article/pii/S1359431111002985>. – ISSN 1359-4311
- [Lai et al. 2008] LAI, Florence ; MAGOULES, Frederic ; LHERMINIER, Fred: Vapnik’s learning theory applied to energy consumption forecasts in residential buildings. In: *International Journal of Computer Mathematics* 85 (2008), Nr. 10, S. 1563–1588
- [Laret 1980] LARET, Louis: *Contribution au développement de modèles mathématiques du comportement thermique transitoire de structures d’habitation*. Université de liege faculté des sciences appliquees laboratoire de physique du batiment, 1980
- [Lebrun 1971] LEBRUN, Jean: *Exigences physiologiques et modalités physiques de la climatisation par source statique concentrée*, Rédaction et administration: Prof. L. Leloup (Institut de mécanique), Dissertation, 1971
- [Lee 2017] LEE, Daniel Sang-Hoon: Impacts of surrounding building layers in CFD wind simulations. In: *Energy Procedia* 122 (2017), S. 50 – 55. – URL <http://www.sciencedirect.com/science/article/pii/S1876610217329119>. – CISBAT 2017 International Conference Future Buildings Districts – Energy Efficiency from Nano to Urban Scale. – ISSN 1876-6102
- [Lee et al. 2004] LEE, Won-Yong ; HOUSE, John M. ; KYONG, Nam-Ho: Subsystem level fault diagnosis of a building’s air-handling unit using general regression neural networks. In: *Applied Energy* 77 (2004), Nr. 2, S. 153 – 170. – URL <http://www.sciencedirect.com/science/article/pii/S0306261903001077>. – ISSN 0306-2619
- [Lee et al. 1996] LEE, Won-Yong ; PARK, Cheol ; KELLY, George E.: Fault detection in an air-handling unit using residual and recursive parameter identification methods. In: *Transactions-American Society Of Heating Refrigerating And Air Conditioning Engineers* 102 (1996), S. 528–539
- [Lettvin et al. 1959] LETTVIN, Jerome Y. ; MATURANA, Humberto R. ; MCCULLOCH, Warren S. ; PITTS, Walter H.: What the frog’s eye tells the frog’s brain. In: *Proceedings of the IRE* 47 (1959), Nr. 11, S. 1940–1951
- [Li et al. 2009a] LI, Qiong ; MENG, Qinglin ; CAI, Jiejun ; YOSHINO, Hiroshi ; MOCHIDA, Akashi: Applying support vector machine to predict hourly cooling load in the building. In: *Applied Energy* 86 (2009), Nr. 10,

- S. 2249 – 2256. – URL <http://www.sciencedirect.com/science/article/pii/S0306261908003176>. – ISSN 0306-2619
- [Li et al. 2009b] LI, Qiong ; MENG, Qinglin ; CAI, Jiejun ; YOSHINO, Hiroshi ; MOCHIDA, Akashi: Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. In: *Energy Conversion and Management* 50 (2009), Nr. 1, S. 90 – 96. – URL <http://www.sciencedirect.com/science/article/pii/S0196890408003154>. – ISSN 0196-8904
- [Lin et al. 2012] LIN, Yashen ; MIDDELKOOP, Timothy ; BAROOAH, Prabir: Issues in identification of control-oriented thermal models of zones in multi-zone buildings. In: *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on IEEE (Veranst.)*, 2012, S. 6932–6937
- [Ljung 1998] LJUNG, Lennart: *System identification*. Springer, 1998
- [Luo and Ariyur 2010] LUO, Qi ; ARIYUR, Kartik B.: Building thermal network model and application to temperature regulation. In: *2010 IEEE International Conference on Control Applications*, 2010, S. 2190–2195
- [Lü et al. 2015] LÜ, Xiaoshu ; LU, Tao ; KIBERT, Charles J. ; VILJANEN, Martti: Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. In: *Applied Energy* 144 (2015), S. 261 – 275. – URL <http://www.sciencedirect.com/science/article/pii/S0306261914012689>. – ISSN 0306-2619
- [Macarulla et al. 2017] MACARULLA, Marcel ; CASALS, Miquel ; CARNEVALI, Matteo ; FORCADA, Núria ; GANGOLELLS, Marta: Modelling indoor air carbon dioxide concentration using grey-box models. In: *Building and Environment* 117 (2017), S. 146 – 153. – URL <http://www.sciencedirect.com/science/article/pii/S0360132317300823>. – ISSN 0360-1323
- [Madsen and Holst 1995] MADSEN, H. ; HOLST, J.: Estimation of continuous-time models for the heat dynamics of a building. In: *Energy and Buildings* 22 (1995), Nr. 1, S. 67 – 79. – URL <http://www.sciencedirect.com/science/article/pii/S037877889400904X>. – ISSN 0378-7788
- [Magalhães et al. 2017] MAGALHÃES, Sara M. ; LEAL, Vítor M. ; HORTA, Isabel M.: Modelling the relationship between heating energy use and indoor temperatures in residential buildings through Artificial Neural Networks considering occupant behavior. In: *Energy and Buildings* 151 (2017), S. 332 – 343. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816311264>. – ISSN 0378-7788
- [MATLAB 1998] MATLAB: The mathworks MATLAB User’s guide. In: *Inc., Natick, MA* 5 (1998), S. 333
- [Matsuno 1984] MATSUNO, Yoshimasa: *Bilinear transformation method*. Elsevier, 1984
- [Mayer et al. 2015] MAYER, Barbara ; KILLIAN, Michaela ; KOZEK, Martin: Management of hybrid energy supply systems in buildings using mixed-integer model predictive control. In: *Energy Conversion and Management* 98 (2015), S. 470 – 483. – URL <http://www.sciencedirect.com/science/article/pii/S0196890415001971>. – ISSN 0196-8904

- [Mechaqrane and Zouak 2004] MECHAQRANE, A ; ZOUAK, M: A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. In: *Neural Computing & Applications* 13 (2004), Nr. 1, S. 32–37
- [Megri and Haghghat 2007] MEGRI, Ahmed C. ; HAGHIGHAT, Fariborz: Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications. In: *Hvac&R Research* 13 (2007), Nr. 6, S. 887–905
- [Megri and Yu 2015] MEGRI, Ahmed C. ; YU, Yao: New calibrated zonal model (POMA+) for temperature and airflow predictions. In: *Building and Environment* 94 (2015), S. 109 – 121. – URL <http://www.sciencedirect.com/science/article/pii/S0360132315300627>. – ISSN 0360-1323
- [Mora 2003] MORA, Laurent: *Prédiction des performances thermo-aérauliques des bâtiments par association de modèles de différents niveaux de finesse au sein d'un environnement orienté objet*, Université de la Rochelle, Dissertation, 2003
- [Mottahedi et al. 2015] MOTTAHEDI, Mohammad ; MOHAMMADPOUR, Atefeh ; AMIRI, Shideh S. ; RILEY, David ; ASADI, Somayeh: Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes. In: *Procedia Engineering* 118 (2015), S. 622 – 629. – URL <http://www.sciencedirect.com/science/article/pii/S1877705815021505>. – Defining the future of sustainability and resilience in design, engineering and construction. – ISSN 1877-7058
- [Mustafaraj et al. 2010] MUSTAFARAJ, G. ; CHEN, J. ; LOWRY, G.: Development of room temperature and relative humidity linear parametric models for an open office using BMS data. In: *Energy and Buildings* 42 (2010), Nr. 3, S. 348 – 356. – URL <http://www.sciencedirect.com/science/article/pii/S0378778809002394>. – ISSN 0378-7788
- [Musy et al. 2002] MUSY, Marjorie ; WINKELMANN, Frederick ; WURTZ, Etienne ; SERGENT, Anne: Automatically generated zonal models for building air flow simulation: principles and applications. In: *Building and Environment* 37 (2002), Nr. 8, S. 873 – 881. – URL <http://www.sciencedirect.com/science/article/pii/S0360132302000501>. – ISSN 0360-1323
- [Nageler et al. 2018] NAGELER, P. ; SCHWEIGER, G. ; PICHLER, M. ; BRANDL, D. ; MACH, T. ; HEIMRATH, R. ; SCHRANZHOFER, H. ; HOCHENAUER, C.: Validation of dynamic building energy simulation tools based on a real test-box with thermally activated building systems (TABS). In: *Energy and Buildings* 168 (2018), S. 42 – 55. – URL <http://www.sciencedirect.com/science/article/pii/S0378778818307424>. – ISSN 0378-7788
- [Naveros et al. 2015] NAVEROS, I. ; GHIAUS, C. ; RUÍZ, D.P. ; CASTAÑO, S.: Physical parameters identification of walls using ARX models obtained by deduction. In: *Energy and Buildings* 108 (2015), S. 317 – 329. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815302620>. – ISSN 0378-7788
- [Ngendakumana 1991] NGENDAKUMANA, Philippe: *Modélisation simplifiée du comportement thermique d'un bâtiment et vérification expérimentale*. (1991)
- [Nielsen 1976] NIELSEN, Peter V.: Flow in air conditioned rooms. In: *Ph. D. thesis from the Technical University of Denmark* (1976)

- [Ohlsson and Ljung 2011] OHLSSON, Henrik ; LJUNG, Lennart: Identification of Piecewise Affine Systems Using Sum-of-Norms Regularization. In: *IFAC Proceedings Volumes* 44 (2011), Nr. 1, S. 6640 – 6645. – ISSN 1474-6670
- [Olofsson and Andersson 2001] OLOFSSON, T. ; ANDERSSON, S.: Long-term energy demand predictions based on short-term measured data. In: *Energy and Buildings* 33 (2001), Nr. 2, S. 85 – 91. – URL <http://www.sciencedirect.com/science/article/pii/S037877880000682>. – ISSN 0378-7788
- [Olofsson et al. 1998] OLOFSSON, T. ; ANDERSSON, S. ; ÖSTIN, R.: Energy load predictions for buildings based on a total demand perspective. In: *Energy and Buildings* 28 (1998), Nr. 1, S. 109 – 116. – URL <http://www.sciencedirect.com/science/article/pii/S0378778898000097>. – ISSN 0378-7788
- [Ooka and Komamura 2009] OOKA, Ryoza ; KOMAMURA, Kazuhiko: Optimal design method for building energy systems using genetic algorithms. In: *Building and Environment* 44 (2009), Nr. 7, S. 1538–1544
- [Ozturk et al. 2005] OZTURK, Harun K. ; CEYLAN, Halim ; CANYURT, Olcay E. ; HEPBASLI, Arif: Electricity estimation using genetic algorithm approach: a case study of Turkey. In: *Energy* 30 (2005), Nr. 7, S. 1003 – 1012. – URL <http://www.sciencedirect.com/science/article/pii/S0360544204003391>. – ISSN 0360-5442
- [Pan et al. 2010] PAN, Yiqun ; LI, Yuming ; HUANG, Zhizhong ; WU, Gang: Study on simulation methods of atrium building cooling load in hot and humid regions. In: *Energy and Buildings* 42 (2010), Nr. 10, S. 1654 – 1660. – URL <http://www.sciencedirect.com/science/article/pii/S0378778810001374>. – ISSN 0378-7788
- [Paoletti et al. 2007] PAOLETTI, Simone ; JULOSKI, Aleksandar L. ; FERRARI-TRECCATE, Giancarlo ; VIDAL, René: Identification of Hybrid Systems A Tutorial. In: *European Journal of Control* 13 (2007), Nr. 2, S. 242 – 260. – ISSN 0947-3580
- [Park et al. 2013] PARK, Herie ; MARTAJ, Nadia ; RUELLAN, Marie ; BENNACER, Rachid ; MONMASSON, Eric: Modeling of a building system and its parameter identification. In: *J. Electr. Eng. Technol* 8 (2013), Nr. 5, S. 975–983
- [Parti and Parti 1980] PARTI, Michael ; PARTI, Cynthia: The total and appliance-specific conditional demand for electricity in the household sector. In: *The Bell Journal of Economics* (1980), S. 309–321
- [Paulus et al. 2015] PAULUS, Mitchell T. ; CLARIDGE, David E. ; CULP, Charles: Algorithm for automating the selection of a temperature dependent change point model. In: *Energy and Buildings* 87 (2015), S. 95 – 104. – ISSN 0378-7788
- [Peuportier and Sommereux 1990] PEUPORTIER, Bruno ; SOMMEREUX, Isabelle B.: Simulation tool with its expert interface for the thermal design of multizone buildings. In: *International Journal of Solar Energy* 8 (1990), Nr. 2, S. 109–120
- [Prakash and Ravikumar 2015] PRAKASH, D. ; RAVIKUMAR, P.: Analysis of thermal comfort and indoor air flow characteristics for a residential building room under generalized window opening position at the adjacent walls. In: *International Journal of Sustainable Built Environment* 4 (2015), Nr. 1, S. 42 – 57. – URL <http://www.sciencedirect.com/science/article/pii/S2212609015000047>. – ISSN 2212-6090

- [Pérez-Lombard et al. 2008] PÉREZ-LOMBARD, Luis ; ORTIZ, José ; POUT, Christine: A review on buildings energy consumption information. In: *Energy and Buildings* 40 (2008), Nr. 3, S. 394 – 398. – URL <http://www.sciencedirect.com/science/article/pii/S0378778807001016>. – ISSN 0378-7788
- [Qin et al. 2012] QIN, Rong ; YAN, Da ; ZHOU, Xin ; JIANG, Yi: Research on a dynamic simulation method of atrium thermal environment based on neural network. In: *Building and Environment* 50 (2012), S. 214 – 220. – URL <http://www.sciencedirect.com/science/article/pii/S0360132311003805>. – ISSN 0360-1323
- [Rabl 1988] RABL, A: Parameter estimation in buildings: methods for dynamic analysis of measured energy use. In: *Journal of Solar Energy Engineering* 110 (1988), Nr. 1, S. 52–66
- [Rajaoarisoa et al. 2012] RAJAOARISOA, L. H. ; M’SIRDI, N. K. ; BALMAT, J. F.: Micro-climate optimal control for an experimental greenhouse automation. In: *Communications, Computing and Control Applications (CCCA), 2012 2nd International Conference on*, 2012, S. 1–6
- [Recht et al. 2014] RECHT, Thomas ; MUNARETTO, Fabio ; SCHALBART, Patrick ; PEUPORTIER, Bruno: *Analyse de la fiabilité de COMFIE par comparaison à des mesures. Application à un bâtiment passif*. 2014
- [Reynders et al. 2014] REYNDERS, G. ; DIRIKEN, J. ; SAELENS, D.: Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. In: *Energy and Buildings* 82 (2014), S. 263 – 274. – URL <http://www.sciencedirect.com/science/article/pii/S0378778814005623>. – ISSN 0378-7788
- [Rouchier et al. 2015] ROUCHIER, Simon ; WOLOSZYN, Monika ; BEJAT, Timea: Identification of Envelope Hygrothermal Properties Based on In-situ Sensor Measurements and Stochastic Inverse Methods. In: *Energy Procedia* 78 (2015), S. 943 – 948. – URL <http://www.sciencedirect.com/science/article/pii/S1876610215017579>. – 6th International Building Physics Conference, IBPC 2015. – ISSN 1876-6102
- [Royer et al. 2014] ROYER, Sullivan ; THIL, Stéphane ; TALBERT, Thierry ; POLIT, Monique: Black-box modeling of buildings thermal behavior using system identification. In: *IFAC Proceedings Volumes* 47 (2014), Nr. 3, S. 10850 – 10855. – URL <http://www.sciencedirect.com/science/article/pii/S1474667016433392>. – 19th IFAC World Congress. – ISSN 1474-6670
- [RT 2012] RT 2012: . – <http://www.rt-batiment.fr/batiments-neufs/reglementation-thermique-2012/presentation.html>
- [Ríos-Moreno et al. 2007] RÍOS-MORENO, G.J. ; TREJO-PEREA, M. ; CASTAÑEDA-MIRANDA, R. ; HERNÁNDEZ-GUZMÁN, V.M. ; HERRERA-RUIZ, G.: Modelling temperature in intelligent buildings by means of autoregressive models. In: *Automation in Construction* 16 (2007), Nr. 5, S. 713 – 722. – URL <http://www.sciencedirect.com/science/article/pii/S0926580506001208>. – ISSN 0926-5805

- [Sadeghi et al. 2011] SADEGHI, Hossein ; ZOLFAGHARI, Mahdi ; HEYDARIZADE, Mohamad: Estimation of electricity demand in residential sector using genetic algorithm approach. (2011)
- [Sandberg 1987] SANDBERG, M: A model for ventilation by displacement. In: *Room Vent-87* (1987)
- [Sang et al. 2017] SANG, Yizhou ; ZHAO, Joe R. ; SUN, Jiaojiao ; CHEN, Bin ; LIU, Sen: Experimental investigation and EnergyPlus-based model prediction of thermal behavior of building containing phase change material. In: *Journal of Building Engineering* 12 (2017), S. 259 – 266. – URL <http://www.sciencedirect.com/science/article/pii/S2352710217301341>. – ISSN 2352-7102
- [Scanu 2017] SCANU, Lisa: *Towards archetypes of self-tuned models for connected buildings*, Université Grenoble Alpes, Theses, November 2017. – URL <https://tel.archives-ouvertes.fr/tel-01736999>
- [Schumacher et al. 2000] SCHUMACHER, JM ; WEILAND, Siep ; HEEMELS, WPMH: Linear complementarity systems. In: *SIAM journal on applied mathematics* 60 (2000), Nr. 4, S. 1234–1269
- [Shabunko et al. 2016] SHABUNKO, V. ; LIM, C.M. ; MATHEW, S.: EnergyPlus models for the benchmarking of residential buildings in Brunei Darussalam. In: *Energy and Buildings* (2016). – URL <http://www.sciencedirect.com/science/article/pii/S0378778816301931>. – ISSN 0378-7788
- [Shein et al. 2012] SHEIN, Wai W. ; TAN, Yasuo ; LIM, Azman O.: PID controller for temperature control with multiple actuators in cyber-physical home system. In: *Network-based information systems (NBIS), 2012 15th international conference on IEEE* (Veranst.), 2012, S. 423–428
- [Sodagar and Starkey 2016] SODAGAR, Behzad ; STARKEY, Diane: The monitored performance of four social houses certified to the Code for Sustainable Homes Level 5. In: *Energy and Buildings* 110 (2016), S. 245–256
- [Sonderegger 1978] SONDEREGGER, Robert C.: Dynamic models of house heating based on equivalent thermal parameters. (1978)
- [Song and Meng 2015] SONG, Jiafang ; MENG, Xiangquan: The Improvement of Ventilation Design in School Buildings Using CFD Simulation. In: *Procedia Engineering* 121 (2015), S. 1475 – 1481. – URL <http://www.sciencedirect.com/science/article/pii/S187770581502901X>. – The 9th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC) joint with the 3rd International Conference on Building Energy and Environment (COBEE), 12-15 July 2015, Tianjin, China. – ISSN 1877-7058
- [Sontag 1981] SONTAG, Eduardo: Nonlinear regulation: The piecewise linear approach. In: *IEEE Transactions on automatic control* 26 (1981), Nr. 2, S. 346–358
- [Soto and Jentsch 2016] SOTO, Aner M. ; JENTSCH, Mark F.: Comparison of prediction models for determining energy demand in the residential sector of a country. In: *Energy and Buildings* 128 (2016), S. 38 – 55. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816305527>. – ISSN 0378-7788

- [Sowell and Haves 2001] SOWELL, Edward F. ; HAVES, Philip: Efficient solution strategies for building energy system simulation. In: *Energy and buildings* 33 (2001), Nr. 4, S. 309–317
- [SPARK 2003] SPARK, LBNL: 2.0 reference manual. In: *Lawrence Berkeley National Laboratory and Ayres Sowell Associates Inc* (2003)
- [Th-Bât] TH-BÂT: Règles TH-U pour les bâtiments existants. .
– <https://www.rt-batiment.fr/batiments-existants/rt-existant-globale/documents-dapplication.html>
- [Tian et al. 2018] TIAN, Wei ; HAN, Xu ; ZUO, Wangda ; SOHN, Michael D.: Building energy simulation coupled with CFD for indoor environment: A critical review and recent applications. In: *Energy and Buildings* 165 (2018), S. 184 – 199. – URL <http://www.sciencedirect.com/science/article/pii/S0378778817339002>. – ISSN 0378-7788
- [Tittelein et al. 2008] TITTELEIN, P ; WURTZ, E ; ACHARD, G: Simspark platform evolution for low-energy building simulation. In: (2008), Nr. 6, S. 25–29
- [Trnsys 2000] TRNSYS, A: Transient System Simulation Program. In: *University of Wisconsin* (2000)
- [Turner et al. 2017] TURNER, W.J.N. ; STAINO, A. ; BASU, B.: Residential HVAC fault detection using a system identification approach. In: *Energy and Buildings* 151 (2017), S. 1 – 17. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816316590>. – ISSN 0378-7788
- [Vapnik 1963] VAPNIK, Vladimir: Pattern recognition using generalized portrait method. In: *Automation and remote control* 24 (1963), S. 774–780
- [Vapnik 2013] VAPNIK, Vladimir: *The nature of statistical learning theory*. Springer science & business media, 2013
- [Vašak et al. 2011] VAŠAK, Mario ; HURE, Nikola ; PERIĆ, Nedjeljko: Identification of a discrete-time piecewise affine model of a pitch-controlled wind turbine. In: *MIPRO, 2011 Proceedings of the 34th International Convention IEEE* (Veranst.), 2011, S. 744–749
- [Vidal et al. 2002] VIDAL, René ; CHIUSO, Alessandro ; SOATTO, Stefano: Observability and identifiability of jump linear systems. In: *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on* Bd. 4 IEEE (Veranst.), 2002, S. 3614–3619
- [Vidal and Ma 2006] VIDAL, René ; MA, Yi: A unified algebraic approach to 2-D and 3-D motion segmentation and estimation. In: *Journal of Mathematical Imaging and Vision* 25 (2006), Nr. 3, S. 403–421
- [Videla and Lie 2006] VIDELA, Juan I. ; LIE, Bernt: Library for modeling and simulating the thermal dynamics of buildings. In: MARQUARDT, W. (Hrsg.) ; PANTELIDES, C. (Hrsg.): *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering* Bd. 21. Elsevier, 2006, S. 1777 – 1782. – URL <http://www.sciencedirect.com/science/article/pii/S1570794606803056>. – ISSN 1570-7946

- [Vorger 2014] VORGER, Éric: *Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment*, Ecole Nationale Supérieure des Mines de Paris, Dissertation, 2014
- [Vries et al. 2009] VRIES, Dirk ; VERHEIJEN, Peter J. ; ARNOLD, J: Identification. Hybrid system modeling and identification of cell biology systems: perspectives and challenges. In: *IFAC Proceedings Volumes* 42 (2009), Nr. 10, S. 227–232
- [Wang and Zhai 2016] WANG, Haidong ; ZHAI, Zhiqiang (: Advances in building simulation and computational techniques: A review between 1987 and 2014. In: *Energy and Buildings* 128 (2016), S. 319 – 335. – URL <http://www.sciencedirect.com/science/article/pii/S0378778816305692>. – ISSN 0378-7788
- [Wang and Xu 2006] WANG, Shengwei ; XU, Xinhua: Simplified building model for transient thermal performance estimation using GA-based parameter identification. In: *International Journal of Thermal Sciences* 45 (2006), Nr. 4, S. 419 – 432. – ISSN 1290-0729
- [Wang et al. 2012] WANG, Shengwei ; YAN, Chengchu ; XIAO, Fu: Quantitative energy performance assessment methods for existing buildings. In: *Energy and Buildings* 55 (2012), S. 873 – 888. – URL <http://www.sciencedirect.com/science/article/pii/S0378778812004410>. – Cool Roofs, Cool Pavements, Cool Cities, and Cool World. – ISSN 0378-7788
- [Wetter et al. 2015] WETTER, Michael ; MARCO, Bonvini ; THIERRY-STÉPHANE, Nouidi ; WEI, Tian ; WANGDA, Zuo: Modelica Buildings library 2.0. In: *Building Simulation 2015 Hyderabad, India*, 2015
- [Witrant et al. 2009] WITRANT, Emmanuel ; MOCANU, Stéphane ; SENAME, Olivier: A hybrid model and MIMO control for intelligent buildings temperature regulation over WSN. In: *8th IFAC workshop on time delay systems, 2009*
- [Wurtz 1995] WURTZ, E: *Three-dimensional modeling of thermal and airflow transfers in building using an object-oriented simulation environment*, PhD Thesis, Paris, Ecole Nationale des Ponts et Chaussées, Dissertation, 1995
- [Xiao et al. 2011] XIAO, Fu ; ZHENG, Chaoyu ; WANG, Shengwei: A fault detection and diagnosis strategy with enhanced sensitivity for centrifugal chillers. In: *Applied Thermal Engineering* 31 (2011), Nr. 17, S. 3963 – 3970. – URL <http://www.sciencedirect.com/science/article/pii/S1359431111004078>. – SET 2010 Special Issue. – ISSN 1359-4311
- [Xin et al. 2018] XIN, Jianbin ; NEGENBORN, Rudy R. ; LIN, Xiao: Piecewise affine approximations for quality modeling and control of perishable foods. In: *Optimal Control Applications and Methods* 39 (2018), Nr. 2, S. 860–872
- [Xu and Xie 2014] XU, Jun ; XIE, Lihua: *Control and estimation of piecewise affine systems*. Elsevier, 2014
- [Xu et al. 2016] XU, Li ; TSENG, H E. ; HROVAT, Davor: Hybrid model predictive control of active suspension with travel limits and nonlinear tire contact force. In: *American Control Conference (ACC), 2016 IEEE (Veranst.)*, 2016, S. 2415–2420
- [Yang and Jiang 2017] YANG, Dong ; JIANG, Lihui: Simulation study on the Natural Ventilation of College Student' Dormitory. In: *Procedia Engineering* 205 (2017),

- S. 1279 – 1285. – URL <http://www.sciencedirect.com/science/article/pii/S1877705817350531>. – 10th International Symposium on Heating, Ventilation and Air Conditioning, ISHVAC2017, 19-22 October 2017, Jinan, China. – ISSN 1877-7058
- [Yu and Kim 2012] YU, Hwanjo ; KIM, Sungchul: SVM tutorial—classification, regression and ranking. In: *Handbook of Natural computing*. Springer, 2012, S. 479–506
- [Yu et al. 2010] YU, Zhun ; HAGHIGHAT, Fariborz ; FUNG, Benjamin C. ; YOSHINO, Hiroshi: A decision tree method for building energy demand modeling. In: *Energy and Buildings* 42 (2010), Nr. 10, S. 1637 – 1646. – URL <http://www.sciencedirect.com/science/article/pii/S0378778810001350>. – ISSN 0378-7788
- [Yun et al. 2012] YUN, Kyungtae ; LUCK, Rogelio ; MAGO, Pedro J. ; CHO, Heejin: Building hourly thermal load prediction using an indexed ARX model. In: *Energy and Buildings* 54 (2012), S. 225 – 233. – URL <http://www.sciencedirect.com/science/article/pii/S0378778812003933>. – ISSN 0378-7788
- [Zhai et al. 2002] ZHAI, Zhiqiang ; CHEN, Qingyan ; HAVES, Philip ; KLEMS, Joseph H.: On approaches to couple energy simulation and computational fluid dynamics programs. In: *Building and Environment* 37 (2002), Nr. 8, S. 857 – 864. – URL <http://www.sciencedirect.com/science/article/pii/S0360132302000549>. – ISSN 0360-1323
- [Zhang and Ljung 2004] ZHANG, Qinghua ; LJUNG, Lennart: Multiple steps prediction with nonlinear ARX models. In: *IFAC Proceedings Volumes* 37 (2004), Nr. 13, S. 309 – 314. – URL <http://www.sciencedirect.com/science/article/pii/S1474667017312417>. – 6th IFAC Symposium on Nonlinear Control Systems 2004 (NOLCOS 2004), Stuttgart, Germany, 1-3 September, 2004. – ISSN 1474-6670
- [Zhang et al. 2017] ZHANG, Qunli ; JIAO, Yuqing ; CAO, Mingkai ; JIN, Liwen: Simulation analysis on summer conditions of ancient architecture of tower buildings based on CFD. In: *Energy Procedia* 143 (2017), S. 313 – 319. – URL <http://www.sciencedirect.com/science/article/pii/S1876610217364548>. – ISSN 1876-6102
- [xiang Zhao and Magoulès 2012] ZHAO, Hai xiang ; MAGOULÈS, Frédéric: A review on the prediction of building energy consumption. In: *Renewable and Sustainable Energy Reviews* 16 (2012), Nr. 6, S. 3586 – 3592. – URL <http://www.sciencedirect.com/science/article/pii/S1364032112001438>. – ISSN 1364-0321
- [Zhao et al. 2016] ZHAO, Jie ; LAM, Khee P. ; YDSTIE, B. E. ; LOFTNESS, Vivian: Occupant-oriented mixed-mode EnergyPlus predictive control simulation. In: *Energy and Buildings* 117 (2016), S. 362 – 371. – URL <http://www.sciencedirect.com/science/article/pii/S0378778815302681>. – ISSN 0378-7788
- [Zhao et al. 2014] ZHAO, Jie ; LASTERNAS, Bertrand ; LAM, Khee P. ; YUN, Ray ; LOFTNESS, Vivian: Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. In: *Energy and Buildings* 82 (2014), S. 341 – 355. – URL <http://www.sciencedirect.com/science/article/pii/S0378778814005714>. – ISSN 0378-7788

Modélisation thermique des bâtiments à partir des mesures en utilisant l'identification de systèmes hybrides

Le secteur du bâtiment est un consommateur énergétique majeur, par conséquent, un cadre d'actions a été décidé au niveau international dans le but de limiter son impact. Afin de mettre en œuvre ces mesures, il est nécessaire d'avoir à disposition des modèles offrant une description fiable du comportement thermique des bâtiments. A cet effet, cette thèse propose l'application d'une nouvelle technique guidée par les données pour la modélisation thermique des bâtiments en se basant sur l'approche des systèmes hybrides, caractérisés par des dynamiques continues et événementielles. Ce choix est motivé par le fait qu'un bâtiment est un système complexe caractérisé par des phénomènes non-linéaires et l'apparition de différents événements. On utilise les modèles affines par morceaux ou PWARX pour l'identification de systèmes hybrides. C'est une collection de sous-modèles affines représentant chacun une configuration caractérisée par une dynamique particulière. Le manuscrit commence par un état de l'art sur les principales techniques de modélisation thermique des bâtiments. Ensuite, le choix d'une approche hybride est motivé par une interprétation mathématique basée sur les équations d'un circuit RC thermique. Ceci est suivi par une brève présentation des modèles hybrides et une description détaillée de la méthodologie utilisée. On montre ensuite comment utiliser la technique SVM pour classifier les nouvelles données. Enfin, l'intégration des modèles PWARX dans une boucle de contrôle hybride afin d'estimer le gain en performance énergétique d'un bâtiment après rénovation est présentée. La méthodologie est validée en utilisant des données issues de cas d'études variés.

Mots-clés: Modélisation thermique des bâtiments, Identification de systèmes, systèmes hybrides, modèles à commutation, PWARX.

Data-driven building thermal modeling using system identification for hybrid systems

The building sector is a major energy consumer, therefore, a framework of actions has been decided on by countries worldwide to limit its impact. For implementing such actions, the availability of models providing an accurate description of the thermal behavior of buildings is essential. For this purpose, this thesis proposes the application of a new data-driven technique for modeling the thermal behavior of buildings based on a hybrid system approach. Hybrid systems exhibit both continuous and discrete dynamics. This choice is motivated by the fact that a building is a complex system characterized by nonlinear phenomena and the occurrence of different events. We use a PieceWise AutoRegressive eXogeneous inputs (PWARX) model for the identification of hybrid systems. It is a collection of sub-models where each sub-model is an ARX equation representing a certain configuration in the building characterized by its own dynamics. This thesis starts with a state-of-the-art on building thermal modeling. Then, the choice of a hybrid system approach is motivated by a mathematical interpretation based on the equations derived from an RC thermal circuit of a building zone. This is followed by a brief background about hybrid system identification and a detailed description of the PWARX methodology. For the prediction phase, it is shown how to use the Support Vector Machine (SVM) technique to classify new data to the right sub-model. Then, it is shown how to integrate these models in a hybrid control loop to estimate the gain in the energy performance for a building after insulation work. The performance of the proposed technique is validated using data collected from various test cases.

Keywords: Building thermal modeling, system identification, hybrid systems, switching models, PWARX.