



**HAL**  
open science

# Towards archetypes of self-tuned models for connected buildings

Lisa Scanu

► **To cite this version:**

Lisa Scanu. Towards archetypes of self-tuned models for connected buildings. Automatic Control Engineering. Université Grenoble Alpes, 2017. English. NNT : 2017GREAT071 . tel-01736999

**HAL Id: tel-01736999**

**<https://theses.hal.science/tel-01736999>**

Submitted on 19 Mar 2018

**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.

## THÈSE

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES**

Spécialité : **Indiquer la spécialité qui figure sur votre carte  
d'étudiant**

Arrêté ministériel : 25 mai 2016

Présentée par

**Lisa SCANU**

Thèse dirigée par **Stéphane PLOIX, Professeur, Université  
Grenoble-Alpes**, et  
codirigée par **Etienne WURTZ, Directeur de recherche, CEA**

préparée au sein du **Laboratoire de Sciences pour la  
Conception, l'Optimisation et la Production de Grenoble**  
dans **l'École Doctorale Electronique, Electrotechnique,  
Automatique, Traitement du Signal (EEATS)**

## **Vers des archétypes de modèles auto-configurables pour le bâtiment connecté**

Thèse soutenue publiquement le **10 novembre 2017**,  
devant le jury composé de :

**Madame Elena Palomo del Barrio**

Professeur, CIC Energigune, Présidente

**Monsieur Christian Ghiaus**

Professeur, INSA de Lyon, Rapporteur

**Monsieur Christian Inard**

Professeur, Université de La Rochelle, Rapporteur

**Monsieur Dirk Saelens**

Professeur, Faculty of Engineering Science, KU Leuven, Examineur

**Monsieur Stéphane Ploix**

Professeur, Université Grenoble-Alpes, Directeur de thèse

**Monsieur Etienne Wurtz**

Directeur de recherche, CEA, Co-Directeur de thèse







A Couga



## Remerciements



3 ans, ou plus exactement 2 ans 11 mois et 2 jours, c'est le temps qu'aura duré cette aventure. Aventure oui, parce qu'une thèse ce n'est ni un travail, ni des études ; c'est un entre-deux. On garde des études l'apprentissage permanent, l'encadrement et le statut d'étudiant. Mais encore faut-il y ajouter la régularité, l'autonomie, la gestion des délais et du projet de thèse dans sa globalité plus propre au domaine professionnel "classique". Le tout assaisonné de beaucoup de remise en question tant professionnelle que personnelle. Une thèse rien qu'à expliquer ce n'est déjà pas simple. Et pourtant c'est essentiel, parce qu'une thèse on le l'achève pas seul, cela est avant tout un travail d'équipe, une collaboration étroite, impliquant collègues, amis et famille. Et c'est donc cette grande équipe que je tiens ici à remercier pour m'avoir encouragée, défiée, boostée, enrichie tout au long de ces trois ans!

Tout d'abord un immense merci à Fred sans qui cette aventure n'aurait jamais commencée. Merci pour ton enthousiasme sans borne parfois teintée d'utopisme, merci pour la confiance que tu m'as accordée il y a maintenant plusieurs années et pour la foi indéfectible que tu as en tes étudiants! Ce fut un plaisir de travailler avec toi et j'espère qu'on aura de nouvelles occasions d'imaginer le monde de demain!

Je tiens également à adresser mes remerciements à mes deux directeurs de thèse : Stéphane Ploix et Etienne Wurtz. Stéphane, merci pour ta présence et ta créativité bouillonnante. Etienne, merci pour ta confiance et pour avoir su doser avec subtilité exigence et tolérance. Mais aussi un grand merci à Pierre qui m'a encadré au jour le jour pendant ces trois ans. Pierre, merci pour ton écoute et ton soutien.

Merci à mon jury de thèse de m'avoir fait l'honneur d'évaluer mes travaux. Merci à Monsieur Christian GHIAUS, Professeur à l'INSA Lyon et à Monsieur Christian INARD, Professeur à l'université de la Rochelle d'avoir rapporté mes travaux. Merci pour vos retours et vos questions constructives qui ont contribué à la version finale de ce manuscrit. Merci à eux, à Dirk Saelens, pour m'avoir de surcroît accueillie

## iv | Remerciements

6 mois à Leuven, et à Elena Palomo, ma présidente de jury, pour les discussions pertinentes durant la soutenance et pour l'intérêt qu'ils ont porté à mon travail.

Un grand merci maintenant aux membres de mes deux laboratoires d'accueil qui ont été présents au jour le jour et qui ont su, tour à tour, m'épauler, me faire grandir, me faire rire. Côté GSCOP, merci Gricha pour le soutien inconditionnel que tu m'as témoigné et pour ta "théorie des graphes pour les nuls" qui m'a permis d'entrer en contact avec le reste du labo, merci Lucie pour ton écoute et ta patience mais aussi pour nos mythiques soirées karaoké, merci Lucas pour ton humour inégalable et tes conseils avisés, merci Matthieu pour ta bonne humeur contagieuse et pour les soirées passées à refaire le monde, merci Amr pour les nombreux échanges et ton optimisme, merci Manar pour ta gentillesse. Plus globalement, un grand merci à Clément, Tom, Franck, Alexandre, Florence, Hugo, Nicolas, Aurélie, Mahendra pour les nombreux fous rires, tournois de coinche, apéros, échanges qui ont largement contribué à ma motivation et ma bonne humeur au quotidien! Merci également à Marie-Jo et Fadila pour votre aide sur le plan administratif et vos encouragements répétés. Côté CEA, un merci tout particulier à mes co-bureaux du début et acolytes de thèse : Rozenn et Clément. Rozenn, merci pour les nombreuses sessions papotage, squash défouloirs et soirées bar déjantées; tu as été une vraie valve de décompression sans laquelle j'aurai très certainement explosée en vol! Clément, merci pour les heures passées à refaire le monde, tes dons de chocolat aux pauses café, ton écoute et ton empathie! Merci Ana, Patricia, Thomas, Léa pour avoir participé à la très bonne ambiance du labo et avoir su instaurer un climat de confiance et d'entraide. Merci Sarah pour ces nombreuses heures penchées sur le concept d'identifiabilité et pour m'avoir sauvée du naufrage au milieu de cet océan d'équations, merci Jeanne pour ton aide sur les analyses de sensibilité et ta réactivité. Merci Arnaud pour ton soutien et ton regard extérieur, merci Aurélie pour ta disponibilité, ton aide précieuse sur tout un tas de sujets techniques et humains, merci Nico pour ta confiance inébranlable, merci Franck pour m'avoir supportée pendant ces derniers mois en tant que co-bureau, merci Timea pour être toujours venue aux nouvelles et m'avoir encouragée.

Je souhaite également remercier le programme européen "PhD School" d'InnoEnergy qui m'a offert l'opportunité de m'ouvrir au monde de l'entrepreneuriat et de l'innovation via tout un panel de formations à Cracovie, Barcelone, Grenoble, Leuven ou encore Amsterdam. Un immense merci à Isabelle Schuster, Marion Decquick, Christine Dominjon, Fabien Gauthier et Richard Biaggioni pour leur investissement et leur énergie à développer ce programme. Un grand merci à tous ceux qui, au détour d'une

formation, ont attisé ma curiosité scientifique, m'ont enrichie de leurs expériences et m'ont apporté une vraie ouverture culturelle : Ivan, Rasha, Arthur, Trang Thu, Lucio, Arpit, et bien d'autres que j'oublie sûrement!

D'autres petites mains ont également contribué au bon déroulement de la rédaction, soutenance et pot de thèse. Merci donc à Claude pour son aide précieuse sur les subtilités de la langue anglaise. Merci à Céd pour ses innombrables relectures, skype et commentaires lors de la préparation de la soutenance! Merci à Rémy, Léa et Claire pour l'aide pour le pot de thèse.

Merci également aux copains grenoblois qui ont su me sortir de ma tour d'ivoire pendant les moments de gros rush et qui ont été là pour partager succès et déconfitures, bières et sushis, sorties ski et après-midi jeux, travaux de peinture et cinés... Merci Elo pour ta bonne humeur permanente, tes projets de refaire le monde, ta présence et ton empathie. Merci Ju pour nos sessions squash aussi vite contrebalancées par nos soirées sushis et/ou bar. Merci Guegue, Carole, Damsou, Minouche, Pauline, Léa, Sophie, Djoj, Maëlle, Tim, Cécile, Dago pour les innombrables soirées bar et/ou repas partagés qui ont ponctué ces trois ans et souvent compliqués les réveils matinaux !

Un immense merci aux copains de toujours (ou presque) dispersés de par le monde qui ont su être là malgré les kilomètres : Coco, Doudou, Céd, Fanny, Pompon, Chalague, Claire, Momo, Ben, Vital, Tim, Clemmy, Laurie, Boubou, Florent, Auré, Florian Julie, Francis, Rémy! Merci pour vos petits mots, vos encouragements, votre connerie, votre sens de la fête, nos discussions interminables. Merci de m'avoir permis de prendre le large lors des nombreux weekends plus ou moins intelligents (mais souvent moins quand même) mais aussi vacances loin de la thèse. Merci de m'avoir fourni le carburant nécessaire à l'accomplissement de cette thèse, merci de m'avoir supportée (et je reconnais que ça n'a pas du toujours être très simple)! Juste merci!

Enfin, un grand merci à ma famille, qui a toujours cru en moi, quelque soit mon projet. Merci pour votre soutien matériel et immatériel.



# Contents



List of symbols . . . . .	xiii
Introduction . . . . .	1
<b>I Sociology of energy and models for new energy services . . . . .</b>	<b>3</b>
<b>1 Context: Sociology of energy . . . . .</b>	<b>4</b>
<b>2 Energy management services . . . . .</b>	<b>7</b>
2.1 Objective. . . . .	7
2.2 Description of the services. . . . .	8
2.3 Inherent constraints . . . . .	10
<b>3 Literature review on models . . . . .</b>	<b>11</b>
3.1 Black-box models . . . . .	11
3.2 Physical models . . . . .	13
3.3 Parameter estimation. . . . .	18
<b>4 Identifiability . . . . .</b>	<b>20</b>
4.1 Literature review. . . . .	21
4.2 Tools . . . . .	22
<b>5 Problem statement . . . . .</b>	<b>23</b>
<b>II Study cases and candidate self-tuning model structures . . . . .</b>	<b>25</b>
<b>1 Presentation of the case study . . . . .</b>	<b>26</b>
1.1 Architecture and instrumentation . . . . .	26
1.2 Solar gains . . . . .	27

1.3	Estimators . . . . .	28
<b>2</b>	<b>Description of model structures . . . . .</b>	<b>29</b>
2.1	Structural description . . . . .	29
2.2	Building blocks . . . . .	30
2.3	Reference model . . . . .	34
2.4	Model with no capacitor . . . . .	36
2.5	Model with two capacitors . . . . .	37
2.6	Models with 3 capacitors . . . . .	41
2.7	Model with 4 capacitors . . . . .	44
2.8	Summary . . . . .	44
<b>3</b>	<b>Acceptability and selection criterion . . . . .</b>	<b>45</b>
3.1	Tools . . . . .	46
3.2	Test of acceptability . . . . .	46
3.3	Selection method . . . . .	47
3.4	Sensitivity analysis of the parameters . . . . .	47
<b>4</b>	<b>Conclusion . . . . .</b>	<b>49</b>
<b>III</b>	<b>Learning behavioural models for one day time horizon . . . . .</b>	<b>51</b>
<b>1</b>	<b>Introduction . . . . .</b>	<b>52</b>
<b>2</b>	<b>End-user energy services requirements . . . . .</b>	<b>53</b>
<b>3</b>	<b>Multi-regression models for energy services . . . . .</b>	<b>54</b>
3.1	Implementation . . . . .	54
3.2	The case of June. . . . .	60
3.3	Multi-regression inspired from physics . . . . .	62
3.4	Limits . . . . .	66
3.5	Conclusion . . . . .	67

<b>4</b>	<b>Meta-optimization</b>	67
4.1	Principle of the meta-optimization	67
4.2	Application to the different structures	69
4.3	Limits	75
4.4	Conclusion	78
<b>5</b>	<b>Meta-heuristic</b>	79
5.1	Principle	80
5.2	Configuration	81
5.3	Test of sensibility to the initial point	83
5.4	Application to the different model structures	84
5.5	Sensitivity analysis	87
5.6	Observability of the model	96
5.7	Reliability of the model along the seasons	99
5.8	Physical interpretation	101
<b>6</b>	<b>Conclusion</b>	105
<b>IV</b>	<b>From building description to end-user energy services</b>	109
<b>1</b>	<b>Objectives</b>	110
<b>2</b>	<b>Study case</b>	110
2.1	Architecture and instrumentation	111
2.2	Estimators	112
2.3	Removal of contact sensors	112
<b>3</b>	<b>Mono-zone model applied to a multi-zone study case</b>	116
3.1	Application to the "garden" bedroom	116
3.2	Two-neighbouring zones model applied to multi-zone study case	118
3.3	Impact of a false temperature	119

<b>4</b>	<b>Generation of the model</b>	120
4.1	Objectives and methodology	121
4.2	Configuration: Gathering information from users	122
4.3	Storage of the information	122
4.4	Model configuration: Generating equations	123
<b>5</b>	<b>Implementation</b>	129
5.1	Configuration of the genetic algorithm	130
5.2	Choice of the right configuration	130
5.3	Study of the convergence	132
<b>6</b>	<b>Application to an energy service</b>	134
<b>7</b>	<b>Conclusions</b>	137
	<b>Conclusions and future works</b>	139
	<b>Résumé en français</b>	143
<b>1</b>	<b>Introduction</b>	143
<b>2</b>	<b>Sociologie de l'énergie et modèles pour les services énergétiques</b>	144
2.1	Sociologie de l'énergie	144
2.2	Services énergétiques	145
2.3	Etat de l'art sur les modèles	146
2.4	Problématique	147
<b>3</b>	<b>Cas d'études et structures de modèles auto-configurables</b>	148
3.1	Cas d'études mono-zone	148
3.2	Blocs élémentaires et structures de modèles	149
3.3	Méthodes de sélection et d'acceptabilité des modèles	152
<b>4</b>	<b>Apprentissage de modèles à l'horizon 24 heures</b>	153
4.1	Pré-requis pour les services énergétiques	153
4.2	Modèles ARX	154

4.3	Méta-optimisation . . . . .	156
4.4	Optimisation génétique . . . . .	157
4.5	Conclusion . . . . .	160
<b>5</b>	<b>De la description du bâtiment aux services énergétiques . . . . .</b>	<b>160</b>
5.1	Objectifs . . . . .	160
5.2	Cas d'études multi-zones . . . . .	161
5.3	Modélisation des ouvertures de fenêtres . . . . .	161
5.4	Génération automatique du modèle . . . . .	162
5.5	Implémentation . . . . .	163
5.6	Application aux services énergétiques . . . . .	163
<b>6</b>	<b>Conclusions et perspectives . . . . .</b>	<b>164</b>
	<b>Bibliography . . . . .</b>	<b>179</b>
	<b>Table of figures . . . . .</b>	<b>194</b>
	<b>Table of tables . . . . .</b>	<b>196</b>



# List of symbols



## Greek symbols

Symbol	Units	Description
$c_{p,air}$	$J.m^{-3}.K^{-1}$	Specific heat capacity of the air
$\Gamma_{in}$	ppm	CO <sub>2</sub> concentration in the studied room
$\Gamma_{cor}$	ppm	CO <sub>2</sub> concentration in the room next door
$\phi_{in}$	W	Internal gains
$\rho_{air}$	$kg.m^{-3}$	Air density
$\sigma_{air}$	-	Dissatisfaction criteria due to air quality
$\sigma_T$	-	Dissatisfaction criteria due to thermal comfort
$\zeta_D$	-	Position of the door
$\zeta_W$	-	Position of the window

## Roman symbols

Symbol	Units	Description
$B_m$	$W.pers^{-1}$	Heat produces by person in average
$C_a$	$J.K^{-1}$	Heat capacity of the air
$C_i$	$J.K^{-1}$	Heat capacity of the fictive wall
$C_{w,n}$	$J.K^{-1}$	Heat capacity of the interior wall
$C_{w,out}$	$J.K^{-1}$	Heat capacity of the exterior wall
$Q_{cor}$	$m^3.s^{-1}$	Air flow with the room next door
$Q_D$	$m^3.s^{-1}$	Air flow by the door

## Greek symbols

Symbol	Units	Description
$Q_{out}$	$m^3.s^{-1}$	Air flow with the outside
$Q_W$	$m^3.s^{-1}$	Air flow by the window
$R_D$	$K.W^{-1}$	Thermal resistance of the door
$R_i$	$K.W^{-1}$	Thermal resistance of the fictive wall
$R_n$	$K.W^{-1}$	Thermal resistance of the interior wall
$R_{out}$	$K.W^{-1}$	Thermal resistance of the exterior wall
$R_W$	$K.W^{-1}$	Thermal resistance of the window
$R_{wn1}$	$K.W^{-1}$	Half thermal resistance of the interior wall
$R_{wn2}$	$K.W^{-1}$	Half thermal resistance of the interior wall
$R_{wout1}$	$K.W^{-1}$	Half thermal resistance of the exterior wall
$R_{wout2}$	$K.W^{-1}$	Half thermal resistance of the exterior wall
$S_{CO_2}$	$ppm.m^3.s^{-1}.per^{-1}$	Average production of CO <sub>2</sub> per person
$T_{in}$	°C	Indoor temperature
$T_n$	°C	Temperature of the room next door
$T_{out}$	°C	Outdoor temperature
$T_w$	°C	Temperature of the wall
$V$	$m^3$	Volume of the room

# Introduction



In the current context of ecological transition, buildings represent a major pool of energy savings since they account for more than 40% of the final energy consumption. In the field of building energy management, researches firstly focused on the envelope of the building or on its energy systems. However, various feedbacks highlight the discrepancies between the energy consumption estimated during the design phase and the real consumption during the operating phase. Occupants play an important role in that gap since feedbacks showed that significantly different consumptions can be observed between 2 similar apartments. The first "low-consumption" buildings or "green" buildings implemented a high degree of automation regarding the energy systems: heating systems, ventilation systems... everything was centralized and automated. The degree of freedom allowed to occupants has been decreasing over several years. Yet, the gap between design and effective energy consumption remained unchanged. To understand the reasons of this phenomenon, sociology researchers started different experiments and surveys in order to understand the reasons of this phenomena. They explained that decreasing the controls available to occupants on their systems could bring a feeling of discomfort and of frustration. The challenge is then to find a way to involve and empower users.

The way explored here consists in involving users by giving them the keys to understand the impact they can have on the building, in terms of energy consumption and comfort. This can be done via a web or mobile application providing different energy services such as explanations of what happened, suggestions on improvements, simulations to measure the impact of a different behaviour or a different control strategy. All these services must access the thermal dynamics of the buildings. Usual simulations such as Dynamic Thermal Simulation (DTS) cannot be used in that context because end-users are not heating engineers and most of the time have no access to the detailed building intrinsic information. It is then necessary to implement simplified models based on data and not only on expert knowledge. The

## 2 | INTRODUCTION

challenge is to determine a model structure able to forecast the temperature and CO<sub>2</sub> evolutions with a sufficient accuracy and a low computational time. The model must be able to adapt while unknown events occur and to predict on a daily horizon basis.

In this thesis, different kinds of models are implemented and tested on two study cases: a mono-zone office and a multi-zone apartment. In the first chapter a literature review is performed on the sociology of energy as well as the different types of models for building energy management: black-box models and grey-box models. Then, in chapter 2, the mono-zone study case is presented as well as the different grey-box models used in the following chapters. Chapter 3 highlights the advantages and limits of the modelling by ARX (Auto Regressive model with eXternal inputs) models or multi regressions models and the gain brought by the introduction of some limited physical knowledge. Then, are described the two parameter estimation methods implemented and their application to the mono-zone office study case. A selection methodology is implemented in order to identify and validate the best model structure for the energy services. Finally, chapter 4 investigates the behaviour of the selected model structure in boundary conditions: lack of sensors and multi-zone environment for example. It details as well an automatic procedure to generate the model and services from only the informations an end-user can provide.

# Chapter I

---

## Sociology of energy and models for new energy services

### Contents

<b>1</b>	<b>Context: Sociology of energy</b>	4
<b>2</b>	<b>Energy management services</b>	7
2.1	Objective	7
2.2	Description of the services	8
2.3	Inherent constraints	10
<b>3</b>	<b>Literature review on models</b>	11
3.1	Black-box models	11
3.2	Physical models	13
3.3	Parameter estimation	18
<b>4</b>	<b>Identifiability</b>	20
4.1	Literature review	21
4.2	Tools	22

## 1 Context: Sociology of energy △

As the energy efficiency of buildings increases, so does the influence of human practices on their global consumption. Henceforth, users can no longer be considered as disturbances in a building. It is needed to take them into consideration and thus to make them aware of the energy issues if the goal is to guaranty low energy consumption in the operation phase of the buildings. Indeed, the importance of human practices impacts was highlighted: modifying only heating set-points and the air change rate can make vary the energy consumption of a building by a factor 3 while the technical systems installed remained the same [21]. To compensate for this effect, one of the first reactions was to go towards more automation in the “green” or “smart” buildings. Despite this automation, the gap between the predicted and the effective consumption still exists as has been proven by the feedbacks of sociologists after surveys on the first office buildings HQE (High Environnemental Quality) [18] and those on residential buildings [84]. A feedback on 6 dwellings with the label “BBC Effinergie” (BBC: Low consumption building) also brought to light that unless occupants are aware of the specificities of the BBC buildings and of the recommended usages, the performances of buildings were strongly damaged [19]. When the installed systems are misunderstood, not properly controlled or simply unaccepted, behaviours may appear which bypass the systems installed and can have a severe impact on the global energy consumption of the building [17]. Energy efficiency of buildings will be reached only if the inhabitants accept to become “smart inhabitants” [95]. More precisely, Bartiaux et al. [11] explain that the energy consumption is the result of the convergence of standards, practices and technological developments which participate to the construction of a socially acceptable standard in terms of comfort. Sociologists also studied the feelings of efficient-building occupants and the efficiency of implemented measures. Dujin et al. [29] furthermore explained that in these approaches, users are seen as passive actors of the buildings, having to learn everything anew and suffering from systems whose rhythms and cri-

teria are not suitable to their needs. The challenge is to involve users so that they change their practices. However, this requires facing several difficulties, for example to make users accept technical systems in buildings or to have interests in energy savings. To answer these questions, these last years have seen the emergence of the sociology of energy focusing on the relationships between humans and energy and inspired by concepts of social psychology. Studies realised are done according to different scales [27]:

- Micro-individual : Scale of individuals
- Micro-social : Scale of domestic spaces and small groups (strategy of actors in the domestic sphere)
- Meso-social : Scale of the organizations and action systems (co-ownership, operator,...)
- Macro-social : Scale of social allegiances (social strata, sex, culture,...)

In the context of energy consumption at building scale, only the macro-social and micro-social scales are considered since it is a matter of understanding the courses of action and consumption habits of occupants, operators and managers in a particular situation: in an efficient building.

From here, two main trends can be distinguished within the sociology of energy: the approach focusing on the behaviours and the more recent approach focusing on the practices. The approach focusing on the behaviours bets on the behavioural incentives to lead up to change. These incentives can either be financial or appealing to the community or individual's values. About the decrease of energy consumption, Moser [69] estimates that it is possible to act for behaviour change in several ways but recommends to appeal to 3 in particular which are complementary: incentives (such as financial rewards), awareness and education campaigns, feedbacks. However, Zélem [96] tempers the impact of the financial rewards, specifying that in order to commit the change, the expected profits need to be higher than the costs either financial or social and that the induced consequences must be quickly visible. Regarding the feedback, it is a subject which appears in various projects and studies such as the project "Affichéco" where 28 dwellings were equipped during 15 to 28 months with a system enabling the display of the energy consumption both global and by function (real-time and history). It did not bring a clear impact on energy consumptions, but it contributed to the development of an energy efficiency

culture [8]. Two cross-analyses were conducted to take advantage of these experiments. The first one covers 26 studies conducted worldwide between 1987 and 2008 and stipulates that the savings realized vary from 1.1 to 20%, that the average is around 5 to 12% and that in a few cases no savings were obtained (3 studies) [35]. The second analysis compiles a dozen studies mostly North-American, led in the same field and explains that energy savings also reach 3 to 13% with an average at 7% [32]. Research was carried out also on different types of incentives. One can quote an experiment done in California aiming to study the impacts of the social standard and of the community. Occupants were informed that "the most popular choice within their community" was to use a fan instead of the air conditioning. This brought a decrease of electric consumptions by almost 10% [22]. In the same way, Zélem [96] refers to the "Hawthorne effect" highlighted by Elton Mayo in the 1930s, explaining that individuals are less concerned with decreasing their energy consumption and more with the feeling to be part of a community. They are looking for a "social recognition". It brings to light that valuing the behaviours of persons who are familiars can lead to a deep modification of practices.

Another essential element to take into consideration is the importance of the rebound effect which consists in the reinvestment of savings realized thanks to a new technology in other usages. Thus, on the basis of their study of the existing literature (in the US), Greening et al. [44] confirm that the rebound effect has a significant impact: up to 50% of the savings realized are then lost in the residential sector, i.e. the energy is spent on new usages. Barreau et al. [10] for their part estimate to one or two years the viability of a feedback-type solution before a return to the original behaviour.

The important point is that the financial incentive in itself is not a sufficient motivation to induce changes in the behaviour on the long term and that users must be integrated within their dwelling and not anymore considered as disturbances. Indeed, occupants should no longer be considered as an external input that needs to be controlled but as an actor of the energy management system that needs to be informed. It is necessary to take into account their need for information on the installed systems in order to enable a real appropriation.

## 2 Energy management services



### 2.1 Objective

The goal pursued in this research work is to develop a thermal building model and to implement it in energy management services towards the different actors of the buildings including occupants but also operators and managers. The idea then is to make available an “e-consultant” which could be materialised for instance by an application for tablets, smart-phones or web services. This “e-consultant” will have to interact with the users of the building via different services in order to help users understand their building and to better communicate on the way designers thought the building. This work is led in collaboration with the ANR project named Involved (ANR-14-CE22-0020). In this context, an example of the type of the considered interface can be seen on figure I.1.

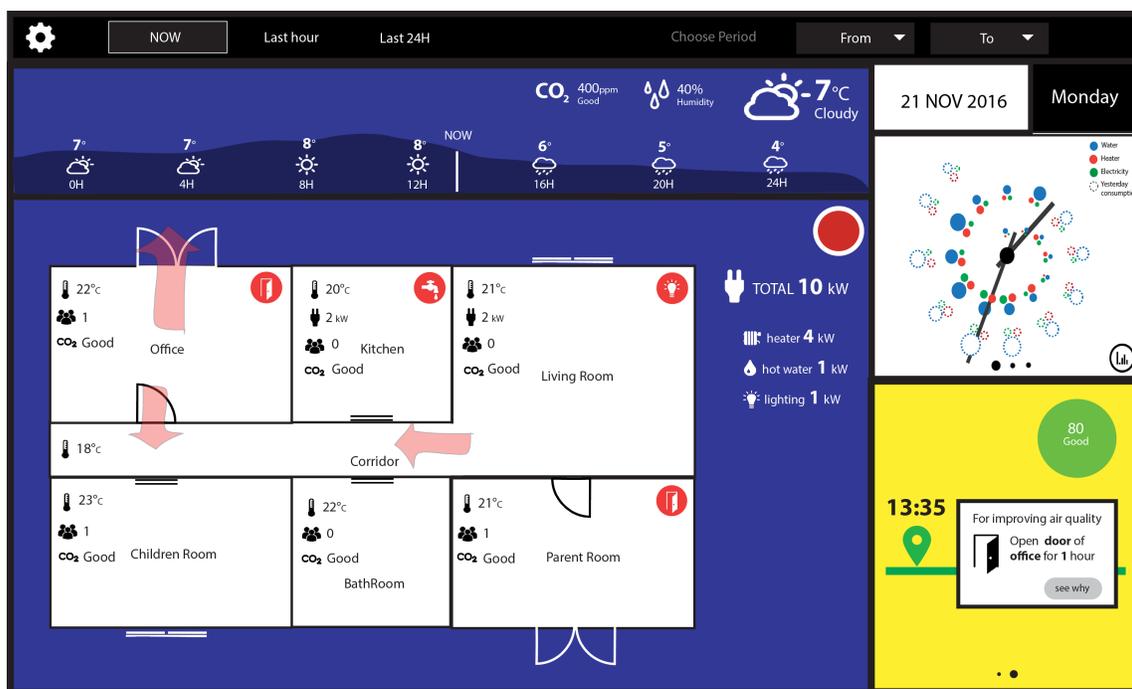


Figure I.1: Persuasive interface

The models developed in this thesis are meant to serve the end-users energy services described in the next section. That will induce several constraints in terms of time scale and complexity of the models described in section 2.3.

## 2.2 Description of the services

The analysis of the sociological literature was made in order to implement services relevant for the users and to induce a real and durable change in human practices. Therefore, with the objective to help users understand the operation of their building and their impact in terms of energy, different services are considered to be implemented:

- “What-if” : What are the consequences of this action?
- “Replay” : What happened yesterday? last month? last year?
- “Mirror” : What is the current state of my dwelling / building?
- “What-for” : How to reach my goal? (Generation of anticipative plans)
- “Suggest-and-adjust” : Adjustment of the anticipative plans by users
- “Explain” : Qualitative representation of physical phenomena
- “How-to” : Explanation of usages expected by the building designers

For each of these services match a list of relevant indicators and physical scales, time scales and rules about the frequency of solicitations of the user. The implementation of these services relies on the hypothesis that the users have previously defined their own objectives: minimizing the energy consumption, minimizing the financial bill or maximizing the comfort, ...

### 2.2.1 What-if

As explained in section 1, one of the main expectations of the end-users is to understand how their dwelling operates and what are the consequences of one action or another. It enables the occupant to raise questions such as: “If I open my window 30 minutes now, what is going to happen in the next hour?”. The e-consultant will run a simulation of the thermal model to bring quantitative answers to that question. The What-if function can take as input different time scales and actions.

### 2.2.2 Replay

The Replay function analyses what happened in the past. It relies on the data history in terms of context data, controls, answers and estimations. Research is led to decide against which time scales each sequence of data should be plot: hourly, daily, monthly, billing periods to enable a factual comparison with the bills received by households.

### 2.2.3 Mirror

The mirror function responds to different motivations; it relies first on the cognitive dissonance theorized by Festinger [34]. He explains that when somebody is subjected to two opposite stimuli, he will act in order to solve the conflict. An example could be the acknowledgement that his actions are not in agreement with his values. In that case, the person will act to match both actions and values. Following that idea, the user will be asked to define his commitment as an individual and in relation to his neighbours regarding a certain indicator. Then, the Mirror function will display a feedback on how well the actions taken match the announced objective. It could also generate notifications under the form of alerts in case of:

- Drift regarding the objective
- Drift regarding data history

It can be noticed that this function can be seen in many applications or prototypes linked to energy under the form of simple feedbacks of energy consumption for example (*cf.* Kjeldskov et al [56], Gamberini et al [39]).

### 2.2.4 What-for

The What-for function generates anticipative plans of actions in order to reach the objective previously defined by users. It is done via an optimization of the thermal model taking into account the user objective previously defined and the possible actions in the considered dwelling.

### 2.2.5 Suggest-and-adjust

Once these anticipative plans are generated and proposed to the end-user, the Suggest-and-adjust function enables him to adapt them to constraints not previously

considered. A really simple example would be to propose a plan for a very hot day in summer and then to incite users to close their shutters in the afternoon in order to maximize the comfort and avoid too high temperatures in the dwelling. It may happen, however, that the user having planned to work home is not willing to spend his afternoon in the dark. The Suggest-and-Adjust function will then make it possible for him to update the generated anticipative plan with this new constraint. In addition, it can also generate a summary of the consequences of the changes operated.

### 2.2.6 Explain

This function will be transparent for the users and transversal to all the other functions. Indeed, its role is to translate the quantitative information returned by the other functions into terms that the end-user can understand. Thus, every function will have to appeal to that Explain function before reaching the end-user or to translate user enquiries to the system.

### 2.2.7 How-to

The How-to function is the keeper of the design of the building. It consists partly in indexing the views of the designer in terms of design and uses for the buildings, explaining which systems are installed and their optimal use. This function is even more important when the dwelling is occupied by tenants and not owners because of the turnover. The use of this function is highlighted by the report "Vivre dans un logement BBC" (Living in a low-consumption dwelling) [19] studying the human practices and energy consumptions of several new dwellings. It explains that the tenants who maximize the capacity of energy savings are those who know the systems and the way to use them properly.

## 2.3 Inherent constraints

These different services require simulation, optimisation, sometimes both and impose different specifications for the models. Among them, one can focus on the "HOW-TO" function. The user will have to define his own objective: minimizing the electric consumption, minimizing his bill, maximizing the comfort, maximizing the air quality or compromising between some of these objectives... Regarding the indoor air quality (IAQ), it is affected by gases (including carbon monoxide, radon,

volatile organic compounds), particulates, microbial contaminants (mold, bacteria). Properly measuring this aspect would require many different sensors and / or experiments. For this reason, it was decided to consider only the measure of CO<sub>2</sub> in the air. Indeed, CO<sub>2</sub> has the double interest to be a good indicator of occupancy and to be a gas. Then, even if a high concentration of CO<sub>2</sub> does not constitute a great danger to health, it reveals a lack of ventilation and so a potential decrease of the IAQ. These different objectives imply different time resolutions but globally, daily end-user services follow a 24 hour time horizon with a 30 or 60 minutes time resolution to avoid bothering people with frequent advice or numerous data. For instance, a 24-hour horizon with a minimum time step of one hour is used for the generation of anticipative plans. For the function "WHAT-IF", one must be able to represent the impact of the opening of the door and the window for example which involves phenomena with a daily dynamic. In order to provide advice and explanations of the physical phenomena, models must enable to determine the causality, that is why in a first place physical models are going to be implemented. However, it is needed to take into account that physical models such as "grey-box" models require a calibration of parameters which arises in turn a number of requirements. Zayane, [94], Armstrong [7] and Alaoui El Azher [4] studied that identification problem. On the other hand, the project aims to implement the technical solution in any dwelling. This then implies that the model requires no expert knowledge about the building in which the system is installed; that is why knowledge models can not be implemented to reach that goal.

## 3 Literature review on models

In order to set up the services presented in the section 2.3, it is necessary to resort to thermal building models in order to forecast the temperature and CO<sub>2</sub> trends in the dwelling. This section presents a literature review on the different types of models used for building energy management.

### 3.1 Black-box models

Universal models or black-box models are based on a general purpose structure suitable for parameter estimation and thus do not require any expertise in physics. However, although black-box models are often criticized for their lack of physical

interpretation [33], Richalet [79] demonstrated that certain kinds of these models allow to recover some physical information. Universal models can be either linear or non linear and accept one or several inputs. Most of the time, universal models have been used for modelling specific systems or walls rather than the whole building [23], [89]. Extending these models for describing a whole building system would be very costly in terms of number of variables. Besides, since most of the time the description of the building systems is unknown, this approach cannot be applied with this objective.

More recently, several authors modelled an entire building with different objectives: predicting the indoor temperature [66], predicting the thermal load [93] or recovering u and g values of the building [52]. Even if the phenomena involved in building physics such as heating, ventilation or occupancy are non-linear, authors often chose linear models to represent the building dynamics. However, models taking only one input tend to disappear due to the increased complexity of building integrated systems. On the other hand MISO models (Multi Inputs Single Output) or MIMO (Multi Inputs Multi Output) are becoming more and more common. These two categories can be broken down into different formulations: ARX (Auto Regressive with exogenous inputs), OE (Output Error), ARMAX (Autoregressive moving average with exogenous inputs), BJ (Box-Jenkins). OE was found to provide the best prediction compared to ARX and BJ models when seeking the response of internal temperature in an office building [64]. Mustafaraj et al [71] conclude that BJ models outperform ARX and ARMAX models. Rabl [76] presented an overview of different methods, focusing especially on their physical interpretation. This study, applied to building models, revealed that even with wrong parameters a model may give a good fit to the data due to compensations in errors. It confirms thus that estimated parameters should be interpreted carefully.

ARX models are the most common in thermal modelling, Jimenez and Madsen [52] and Mustafaraj et al. [71] also show that they are sufficiently efficient to model building thermal behaviour. ARX models were also used by Armstrong [7] and Naveros et al. [73] for building applications. Thus, it is this structure which is implemented in here.

The general formulation of an ARX model with one output is illustrated in equation I.1 below for the temporal description.

$$T_{i,n} = - \sum_{l=1}^p a_l T_{i,n-1} + \sum_k \sum_{j=0}^q b_{k,j} u_{k,n-j} + e_n \quad (\text{I.1})$$

where  $b_k$  are the coefficients for the respective inputs  $u_k$  and  $e$  represents the error which is assumed to be white noise. The z-form is represented in equation I.2:

$$A(z^{-1})T_k = \sum_k B_i(z^{-1})u_k + e_k \quad (\text{I.2})$$

Lowry and Lee [64] implemented different autoregressive models such as ARX or OE models in order to predict the response of internal temperature in a building under passive conditions. Having studied different orders for their models, they concluded that even if outdoor and indoor air temperatures are highly correlated, simple linear-regression models have a variance of at least 44%. Thus, auto-regressive models are interesting. They tested different possibilities of higher-order with respect to any combination of the input and output variables and concluded that no improvement was noticed beyond second-order. The models were able to predict with a good accuracy the indoor air temperature for more than one week.

## 3.2 Physical models

In this section, there will be presented a literature review on the models in the field of energy management from prevalent structures.

### 3.2.1 Different types of models

Two main categories of physical models are usually considered:

- Pure knowledge-based models or “white-box” models
- Semi-physical models or “grey-box” models

**“White-box” models.** Knowledge-based models rely on more general models grounded on physical laws. They require much information on the physical characteristics of the buildings and on the phenomena involved. They can compensate a lack of measurements.

**“Grey-box” models.** Semi-physical models combine a part of physical knowledge formalized by structures of equations and a part of parameters learnt thanks to an estimation procedure. These models represent an interesting compromise since they can be more easily extrapolated than black-box models but they also keep a high capacity of adaptation with adjustable parameters. The parameters of these models are however difficult to estimate because unlike pure data models, their structure is not suited for calibration (non-linearity towards parameters). Compared to white-box models, the tuned parameters are degrees of freedom to fit the measurements.

White-box models are not retained for this study mainly because they require too much data on the building itself and so their deployment on different dwellings in a short time is difficult. Grey-box and black-box models have the advantage to model a dynamic system from the measurement of the inputs and outputs of the model. Mathematical functions with or without a link to the physics of the phenomena modelled are then chosen to minimise the forecast error.

### 3.2.2 Models for energy management in buildings

In the field of the energetics of buildings, most models were designed to perform analysis of energy needs in the design phase as a guide in the choice of the architecture, materials and then to ensure the respect of thermal regulations. Therefore, most studies compute annual needs. Indeed, the French regulation for example, the RT2012, limits the average annual energy consumption in order to prevent the construction of buildings with too high consumption. However, the computation is based on standard scenarios of occupancy, heating set-points... which are most of the time very far from the reality as can be seen in the appendix of the French “Official bulletin n°201114” explaining the computation method of usage scenarios. In consequence, big gaps are observed between predicted consumptions and real ones. On the other hand, the more efficient a building, the more significant the impact of human practices. As it is very difficult to forecast human practices on a long term, it is necessary to adapt the model variables and to have short time scales. The objectives of building energy management are different: they aim at producing set-points suitable for use on a short time scale. The granularity required is thus very different from that necessary for the performance evaluation of buildings, as it ranges from an annual horizon to one day. Forecasts are usually about heating set-points, indoor temperature and indoor air quality (IAQ) such as for example

CO<sub>2</sub> concentration. It is neither necessary to provide a high accuracy in measurements nor a time step of one minute. However, some phenomena neglected for an annual simulation but which impact on both the energy consumption and the occupants' comfort on a daily basis need to be taken into consideration, as for example, the ventilation habits. According to Fabi et al. [31], the opening of windows is one of the user-accessible adaptive actions most impacting energy consumption, comfort and air quality. Castillo et al. [20] also reached the conclusion that taking into account the opening of the door on the corridor led to significant improvements in their results. That's why the impact of door or window opening on thermal or CO<sub>2</sub> balances will be more specifically studied in this thesis.

In this study, the objective is to identify structures of models allowing to provide advice for users in both old or new buildings. It is then necessary to be able to free itself from an accurate knowledge of the structure of the building (insulation, materials,...) while remaining able to access some physical parameters (window and door openings for example) in order to generate relevant explanations and advice. Thus, semi-physical models are then a priori suitable for energy management needs. The literature review reveals that the use of electric analogy modelling is prevalent. The concept of the analogy is summarized in figure I.2. These models offer a good representation of the different components of the buildings and a high ease of use. It is thus this option which is first investigated. However, when deciding to deal with physical or semi-physical modelling, it is important to tackle the issue of the required granularity, i.e. the physical phenomena to take into consideration but also to question which elements of an occupied area need to be modelled. As previously seen although, this choice is strongly impacted by the objective of the model, some elements are recurrent in the literature. Two main points can be identified about which authors have divergent opinions:

- the localisation of the inertia
- the computation and impact of solar gains

### 3.2.3 Working hypothesis

Based on the literature review presented and the objective of the models developed, choices were made regarding the localisation of the inertia and the computation and impact of solar gains.

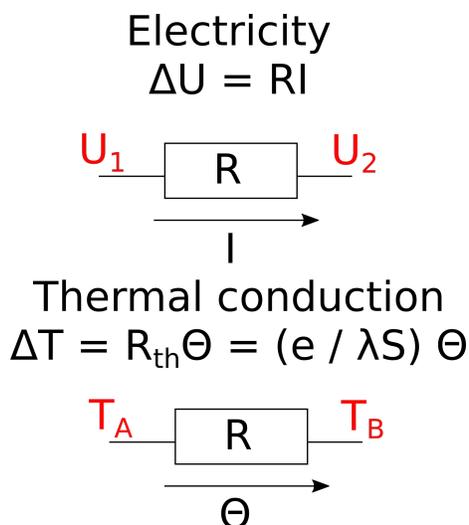


Figure I.2: Principle of electric analogy

### 3.2.3. i) Inertia and models order

Regarding the modelling of inertia in building for semi-physical models, most authors consider that it is located only in the exterior walls and the air with only a few exceptions. However, Hazyuk et al. [47] and Mathews et al. [65] also take into consideration the impact of the slab in their modelling. Bacher and Madsen [9] carried out a study aiming at comparing models with different orders. To do so, they generated seventeen models including two to five capacitors. Starting from the model of the smallest order, they moved on to determine different models of superior order, computing for each one the likelihood function. Then, they selected only the model with the bigger value of the likelihood ratio to pursue the iteration. The selection procedure stopped when no significant superior order could be found, which was the order five for that study. In conclusion, it was demonstrated that the order five brought less than a 5% improvement, in consequence, the retained model is a fourth-order model. Mejri et al [67] applied models varying from first-order to fifth-order on a tertiary building and their results show in particular that exceeding second-order not only brings no further improvement but can even lead to damaged results. When the objectives are to forecast the temperature and the power of a complex building, authors agree on the fact that a second-order model is the most simple order to reach these objectives. Thus, Zayane [94], Fraisse et al. [37] and Alaoui El Azher [4] all studied models of type R4C2 (4 resistances and two capacitors). The study conducted by Del Barrio et al [26] highlights the fact that the choice of the order also depends on the objective of the model and so on

the requirements in terms of accuracy but also of the controller used. It seems thus difficult to choose a model order a priori.

### *3.2.3. ii) Solar gains impact*

Regarding the impact of solar gains, it has been observed more heterogeneity in the proposed solutions. Teichmann et al. [87], Achterbosch et al. [2] as well as Bacher and Madsen [9] consider that solar gains contribute directly to heat the ambient air but do not take into consideration the warming of the exterior and inner walls. Others, like Berthou et al. [14] or Hazyuk et al. [47] include the impact of solar gains on the exterior walls whereas Nielsen [74] favours the impact on inner walls. Kämpf and Robinson [54] and Fouquier et al. [36] are going even further combining these three places of impact of the solar radiation. Mathews et al. [65] conduct one of the few studies neglecting the impact of solar gains on ambient air and so the radiative component; they only take into account the impact on inner and exterior walls. These different approaches can be explained partly by the diversity of the associated goals: some aim at quantifying building performance, others seek to create models for energy management systems and still others want to launch simulations at a district level. Castillo et al. [20] compared different approaches without coming to any positive conclusion about which was the most suitable. However, they confirm that it is necessary to take solar gains into account.

### *3.2.3. iii) Solar gains computation*

For the services based on simulation or prediction, it will be necessary to compute or forecast the solar gains for the next 24h or more. For this reason, a choice needs to be made on how the computation is done. Different research organizations developed their own model for solar radiation, such as the European Solar Energy Research (EUFRAAT) in the European Solar Radiation Atlas based on the work of Kasten and Czeplak [55] and ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) who developed another one which is among other things used in the Energy+ software. They all have their own specifications about the input data required and their accuracy. The data commonly used gather the type, amount and distribution of clouds, cumulated rainfall or fractional sunshine. Usually, they are sorted in two main categories:

1. parametric models
2. decomposition models

Parametric models require several inputs including detailed information on the atmospheric conditions as explained by Wong and Chow [91]. Decomposition models require information only on global radiation to predict the beam and sky components. Wong and Chow [91] proposed different comparisons between solar models. First, two parametric models are compared: the Iqbal model [51] offering a better accuracy than more conventional models [45] and the ASHRAE one, simpler but widely used by the engineering community. Then, ten decomposition models are compared and finally a comparison is made between parametric and decomposition models. The conclusions are diverse: first it must be noted that at a zero-zenith inclination, the difference between the Iqbal model and the ASHRAE one did not exceed 7 %, the ASHRAE model being though less accurate for diffuse radiation predictions due to some of its hypotheses. About the decomposition models, three of them agreed with the measured data, while the others provided reasonable predictions. Lacking precise atmospheric information, decomposition models based on measured hourly global radiation would be a good choice.

Since the objective here is to limit the number of sensors required for an easier implementation of our system, a decomposition model will be preferred.

### 3.3 Parameter estimation

This being said, it is important to realize that the choice of the model (order and structure) will depend on the possibility or not to estimate its parameters. Indeed, as the number of parameters to estimate increases, so does the cost of the estimation in terms of computation time and necessary data.

#### 3.3.1 Principle

According to Ljung [63], the estimation consists in giving a class of models, or model structures, for finding within such class the model that fits a given data-set with the minimum loss, according to a given criterion. Then, this process puts in relation three different elements: the data set, the model structure and the parameter identification procedure. This procedure can be described in different steps:

1. Measurements are conducted on the inputs and outputs of the system using or not specific excitations of the systems
2. A model is developed to process the recorded input and output sequence
3. A suitable structure of model is determined
4. A statistical method is used to estimate the unknown parameters of the model.

Usually, step 3 is done iteratively: a simple structure is defined and evaluated; if the representation of the system does not reach the required accuracy, a more complex structure is then tested. The identification processes aroused a lot of research works (Jimenez and Madsen [52]) and comparisons (Androutsopoulos [6]) as well as international competitions (1994, 1996). The software and tools implement different estimation methods and different kinds of models but the bulk of them are based on deterministic differential equations or state space models. During the first competition, 20 participants had to solve a specific parameter estimation problem. At the end, the estimated results were compared according to the overall values for R and C with the values used in the simulation model for the creation of the data along with statistical tests. The conclusions drawn were that estimating the individual parameters of a lumped model was more difficult than estimating an overall value and that researchers needed an expertise on the models they used in order to obtain good results. Rabl [76] provides a description of the different type of models together with the different estimation methods.

### 3.3.2 Illustration

Figure I.3 shows a practical example of the estimation process. During the training period, the input and output data are used to identify the parameters of the model. In that example the input data are the outside temperature and the state of the window, and the output chosen is the indoor air temperature. It can be observed in figure I.3 that the estimated temperature is close to the reference except on one or two points. During the validation period, the identified parameters are recovered, the input data used and the error between the predicted and measured outputs is computed. It can be seen in figure I.3 that the gap between estimated and measured temperatures is bigger, but this can be explained by the appearance of some new physical phenomena. So, it is possible to obtain the parameters measuring the outputs without a need for expert knowledge on the physical characteristics of the

building, which is perfectly adapted to the needs of this study. In the following chapters, different parameter estimation methods will be investigated in order to ensure a maximum agreement between the prediction of the model and the measures.

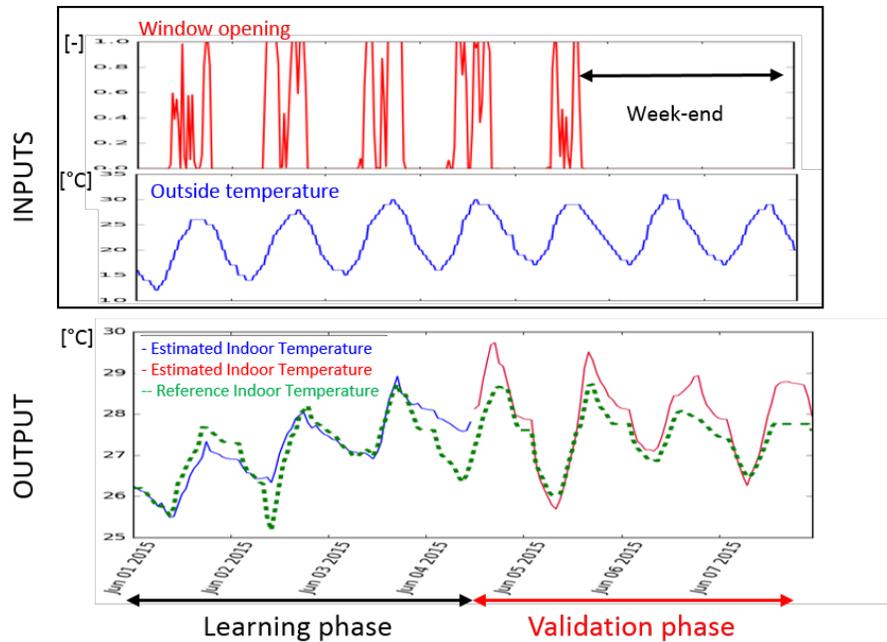


Figure I.3: Learning and validation process

### 3.3.3 Data used for parameter estimation

Another aspect of major importance on both the results accuracy and the computational time required for the learning phase is the specifications of the training phase: its duration and the data used. Castillo et al. [20] also performed tests on periods from 1 to 20 days and noticed that a longer training phase led to a better model accuracy. Thus, they finally obtained interesting results for periods of 7 days.

## 4 Identifiability

The objective of this research work is to identify a structure of model able to accurately forecast indoor CO<sub>2</sub> concentration and air temperature in different study cases in order to help the occupants to reach their comfort with the less energy possible. The physical meaning of the parameters is not required for providing the energy services described in section 2 of chapter I. However, this possibility is investigated in order to well identify the limits of this modelling and the hypothetical

Authors	Learning duration	Order
Bacher and Madsen [9]	6 days	2
Bouache et al. [15]	5 days	7
Coley and Penman [24]	10 days	2
Braun et al [16]	7 to 14 days	8
Hazyuk et al. [47]	60 days	2
Freire et al. [38]	15 days	1
Harb et al. [46]	39 to 110 days	1 to 3
Jimenez and Madsen [52]	10 to 12 days	0 to 4
Le Mounier et al. [60]	37 days	1
Liao and Dexter [61]	14 days	2
Mustfaraj et al. [71]	1.9 days	1 to 2

**Table I.1:** Learning duration and orders of the models

other applications. In this context, the structural identifiability does not constitute a requirement but an information on the possibility of recovering physical knowledge from the parameters.

Methods to determine identifiability are very rarely applied in the field of building models but very common in other fields such as biosciences. Identifiability can be divided into two types: structural identifiability and practical identifiability.

**Structural identifiability (*a priori*).** It deals with the model structure in itself and gives information on the ability to obtain a unique set of parameters for a given output.

**Practical identifiability (*a posteriori*).** It evaluates the input data available and ensures that it is sufficient to reliably uniquely estimate the parameters.

## 4.1 Literature review

The structural identifiability is a necessary condition to the practical identifiability. Several techniques exist to perform an identifiability analysis, either algebraic or analytic. Most of them were implemented for linear models and cannot be easily applied to non-linear models due to computational cost. Some techniques such as that based on the Markov parameters, the transfer matrix or the minimum realisation of the state space use different forms of the model. In order to apply those methods to non-linear models, different techniques can be implemented:

- Linearisation around an equilibrium point (Godfrey and Distefano III [42])
- Development in series (Godfrey and Distefano III [42] and Walter [90]);
- Method of similarities (Vajda et al. [88])
- Differential algebra (Ollivier [75])

Only some of these approaches allow to determine the *global* identifiability while others address *local* identifiability, holding around a point in the parameter space. The complexity in the number of variables and parameters of these methods grows too fast for them to be generally applicable to models of large dynamic systems. Sedoglavic developed a probabilistic semi-numerical algorithm for testing observability and global structural identifiability of systems even for large models with a few hundred state variables and parameters and also yielding the functional relations between parameters [82]. This tool has been developed under Maple, a software for symbolic computation.

The practical identifiability will not be studied in this work since the objective is to validate the selected structure regardless the data available.

## 4.2 Tools

In order to test the structural identifiability, two tools were used:

- The Differential Algebra Identifiability of Systems (DAISY) proposed by Bellu et al [12]
- The algorithm developed by Sedoglavic under Maple [82]

Unfortunately, if the tool DAISY returns quickly interesting results for simple models such as Reference model, the computation time rockets for the selected Model4C. After more than 11 days of computations, it did not yield any conclusions. Raue et al. [77] compared three different approaches for identifiability including DAISY and reached the same conclusion about the ability of the tool to deal with complex models. In contrast, the approach chosen by Sedoglavic is based on Exact Arithmetic Rank (EAR) which presents the advantage to handle large and complex systems. For these reasons, DAISY was not used in this study since the selected parameter cannot be tested via this method.

## 5 Problem statement



The global objective of the ANR project Involved (ANR-14-CE22-0020) is the development of a persuasive e-consultant for energy management services. Focusing on residential sector, the objective is to allow any end-user to configure its own system in order to enjoy dedicated energy management services. The goal is not to provide real-time control of the dwelling but to give explanations of past events (replay) or anticipative plans for the day to come. The implementation of energy services implies to develop thermal models to forecast the temperature and CO<sub>2</sub> to assess the thermal comfort and the air quality. Targeting the end-users and the existing buildings implies to free the model from building detailed characteristics. It requires also to study the modalities of interaction with the end-user. Indeed, generating anticipative plans with actions to do every 10 minutes is not realistic: end-users are not willing to do this investment. For this reason, the time-step defined is of 1 hour and the horizon to 24 hours. The general objective pursued and the methodology adopted in the present thesis are summarized in figure I.4. Upstream of this work, a team is working on the persuasive interface and the interaction with the end-user. They are, in particular, in charge of defining how the end-user can describe its dwelling and how to store the information into a standardized format (xml form).

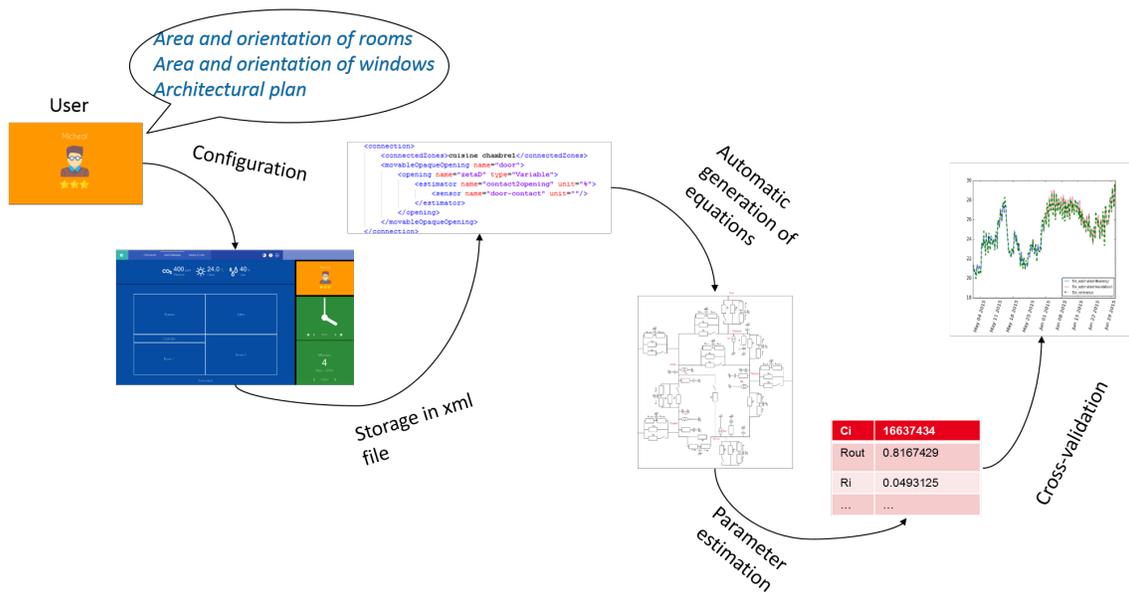


Figure I.4: General objective and process

The aim of this thesis is to pass from this description of the dwelling on an xml form to an estimated model able to forecast the temperature and CO<sub>2</sub> with a time-step of 1 hour and a horizon of 24 hours. This objective is decomposed in different steps: the first one is to select a suitable model structure. Different types of models are implemented and compared: first, data-based models such as ARX or multiple regressions. Then, physics will be taken into consideration in a wider model with the development of models based on thermal circuits (called RC-models or semi-physical models in the following). The detailed building characteristics being unknown, all these models require a step of training to estimate the values of the resistances and capacitances or the coefficients of the regressions. Semi-physical models are generally more difficult to estimate than regressions, this is why two different parameter estimation methods are applied: a gradient-based and a meta-heuristic algorithms. Once all these models defined, they are applied first to a mono-zone study case in order to compare their performance. The second step of the work consists in ensuring the implementation of this model to more complex study case. The different types of models are then applied to a multi-zone study case and compared again. This new study case brings issues regarding the generation of the model. Defining manually the semi-physical model structure for an entire flat is hazardous. Then, it has been necessary to implement an automatic process to pass from the simple description of the building to the equations of the model. To finish, additional studies have been led regarding the physical interpretation of the parameters or the identifiability. The physical meaning of the parameters is not considered as an objective in this research work but studied as an interesting additional aspect.

## Chapter II

---

# Study cases and candidate self-tuning model structures

### Contents

<b>1</b>	<b>Presentation of the case study</b>	26
1.1	Architecture and instrumentation	26
1.2	Solar gains	27
1.3	Estimators	28
<b>2</b>	<b>Description of model structures</b>	29
2.1	Structural description	29
2.2	Building blocks	30
2.3	Reference model	34
2.4	Model with no capacitor	36
2.5	Model with two capacitors	37
2.6	Models with 3 capacitors	41
2.7	Model with 4 capacitors	44
2.8	Summary	44
<b>3</b>	<b>Acceptability and selection criterion</b>	45
3.1	Tools	46
3.2	Test of acceptability	46
3.3	Selection method	47

3.4 Sensitivity analysis of the parameters . . . . .	47
4 Conclusion . . . . .	49

# 1 Presentation of the case study △

The study case used to test the models and compare the different estimation methods is an office of the G-SCOP laboratory in Grenoble.

## 1.1 Architecture and instrumentation

A first wall of the office, with two windows, is in direct contact with the outdoor. Two walls overlook a corridor, one with a door and the other with a glass partition. The last wall is in contact with stairs and an emergency exit. The floor and the ceiling are both in contact with offices or classrooms. This office is occupied by 1 to 4 people depending on the periods in the year. In terms of instrumentation, the office is on a 1500 m<sup>2</sup> site equipped with almost 200 ENOCEAN sensors. The office is equipped with 26 sensors including: temperature sensors, CO<sub>2</sub> and COV sensors, door and windows contacts, motion sensor, one illuminance sensor, one humidity sensor, several power consumption meters and a weather station (*cf.* figure II.1). As the problem posed generates constraints in terms of the number of sensors, only part of the measurements available will be used to estimate the parameters of the models presented. The useful data are then coming from air temperature, CO<sub>2</sub>, window and door contacts sensors and from the weather station for the outdoor temperature and the nebulosity. As the behaviour of a building is very different in winter and summer, it has been chosen to validate the models on two different datasets covering these two periods.

## 1.2 Solar gains

For the solar gains, according to what was explained in section 3.2.3. iii) in chapter I, the model used is a decomposition model. First of all, solar gains are the sum of the direct, the diffuse and the reflected radiations. The direct radiation is based on:

$$\phi_{direct} = \tau_M \phi_{Atm} e^{-M E_{rayleigh} L} \quad (W.m^{-2}) \quad (II.1)$$

where  $\tau_M$  is the transmissivity coefficient,  $\phi_{Atm}$  the solar radiation coming to the atmosphere,  $M$  the actual crossed mass of air with regard to the vertical mass of air,  $E_{rayleigh}$  the Rayleigh optic thickness (diffusion due to molecules) and  $L$  represents the atmospheric blur (steam, fog, dusts) of Linke. These coefficients depend on the day of the year, the atmospheric pressure, the temperature and the humidity.

The solar power collected by an inclined surface depends on the incidence angle  $\theta$  defined as:

$$\theta = \cos^{-1}(\cos(\alpha)\sin(\beta)\cos(\psi - \gamma) + \sin(\alpha)\cos(\beta)) \quad (II.2)$$

where  $\alpha$  is the altitude,  $\beta$  is the slope,  $\psi$  the azimuth and  $\gamma$  the exposure.

The next step consists in determining the diffuse radiation which can be considered isotropic according to Kumar et al [58]. Based on this hypothesis, Liu and Jordan [62] developed an empirical relationship between the transmissivity coefficients for direct and diffuse radiation for clear days:

$$\tau_{diffuse} = 0.271 - 0.294\tau_M \quad (II.3)$$

Then, the extraterrestrial radiation incident on the plane normal to the radiation on the  $n$ th day of the year is defined according to (*cf.* Spencer [86]):

$$G_{on} = G_{sc} \left(1 + 0.033 \cos\left(\frac{360n}{365}\right)\right) \quad (II.4)$$

where  $G_{sc}$  is the solar constant considered equal to  $1367 W.m^{-2}$

which yields to:

$$\phi_{diffuse} = G_{on} \tau_{diffuse} \sin(\alpha) \quad (W.m^{-2}) \quad (II.5)$$

Finally, the reflected solar radiation is computed according the Gate's formula as expressed below:

$$\phi_{reflected} = G_{on}(0.271 + 0.706\tau_M)\sin(\alpha)\sin^2\left(\frac{\beta}{2}\right) \quad (W.m^{-2}) \quad (II.6)$$

### 1.3 Estimators

The models will need some input information about occupancy and heating power which will cannot be deduced directly from measurements. It is thus needed to implement what is called here "Estimators" of these variables. They will first be applied to this case study and, in chapter IV, it will be investigated whether or not they can be generalized. The occupancy is deduced from power meters measurements on the different desks. According to the number of plugs delivering power, the number of people in the office is computed.

Due to the global heating system configuration, it is impossible to measure the heating power injected in the office. An alternative was found by placing a temperature sensor on the surface of the heater and by applying a factor  $K_{heat}$  to the difference of temperature  $T_{heater} - T_{in}$ . The  $K_{heat}$  factor is part of the parameters which need to be learnt.

To complete these estimators, an internal gain is also computed, as the sum of the solar gains, the electric gains due to the appliances, the gains due to occupancy and a constant heat flow for including non-modelled phenomena. The solar gains have already been described, the electric gains are also based on the sum of the power measurements of the desks. The gains due to occupancy are the results of an average of a power due to the body metabolism of 80W multiplied by the occupancy. And the constant heat flow is one of the parameters requiring to be estimated and representing all the phenomena neglected here. Then, the internal gains are defined as the sum of solar gains, heating gains, electric gains and occupancy gains:

$$\phi_{in} = \phi_{solar} + \phi_{heat} + \phi_{electric} + \phi_{occupancy} \quad (II.7)$$

which yields to:

$$\phi_{in} = \phi_{solar} + K_{heat}(T_{heater} - T_{in}) + \phi_{electric} + B_m n \quad (II.8)$$

where  $\phi_{solar}$  is the solar gains described in the previous section,  $T_{heater}$  is the temperature on the surface of the heater,  $T_{in}$  is the indoor air temperature,  $K_{heat}$  is a parameter which is learnt,  $\phi_{electric}$  is the heat emitted by the laptops,  $B_m$  is the average production of heat per person and  $n$  is the number of occupants.

Obviously, some of these estimators are suitable for an office building but not at all for a residential one. For example, the occupancy forecast estimated from power-meters sensors will very likely not be valid for dwellings. It can be noticed also that these estimators require data from sensors, which seems in contradiction with the objective of making the developed solution easy to spread. That is why, different estimators will be applied in chapter IV aiming both at decreasing the need for sensors, while allowing to match the different functions of the building.

## 2 Description of model structures

### 2.1 Structural description

The structures of the models studied in the present work rely both on the previous literature review and on a dwelling ontology. Establishing such a “dwelling ontology” involved the listing of all components present in a dwelling along with their characteristics impacting the energy management. Each zone is characterized by its geographical location, its inertia, its volume, its area, its appliances and its sensors. The links between the zones can be of different types: openings or walls. Walls are characterized by their thermal resistance whereas openings can be broken down into several categories. Openings may have a glass or opaque surface. Openings with a glass surface are characterized by their ability to be opened or not, their area, their orientation and the sensor informing on their state. Connexions with an opaque surface are characterized only by their ability to be opened or not.

Then, as described in the ontology, it was necessary that the models be able to represent different specific characteristics, among them the impact of window and door openings. In order to define which structure would be simplest to implement, a literature review was first performed, as presented in section 3. But the structures identified neither matched the requirements of the ontology nor the needs of users advice generation. Thus, it has been chosen to extract from this state-of-the-art the

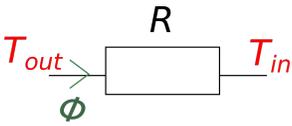
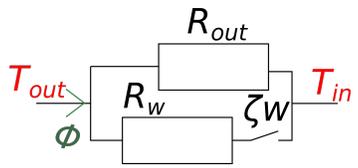
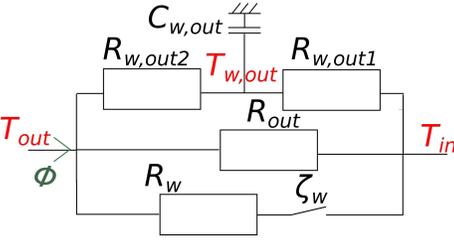
different hypotheses on the inertia modelling in order to quantify the improvements, if any, on the results (*cf.* table II.1).

Localization of the inertia			
Exterior walls + Ambient air	Exterior walls + Inner walls	Exterior walls + Inner walls + Ambient air	Inner walls + Ambient air

**Table II.1:** Approaches to be tested

## 2.2 Building blocks

Based on the literature review on the physical models and the “dwelling ontology” established in the previous section, different building blocks were defined for each part of the ontology. They are distinguished in 3 categories: the exterior wall represented in table II.2, the inner wall which only differs by taking into consideration a door and not a window represented in table II.3 and the zone represented in table II.4. These different building blocks will then be combined to create different structures in order to select the most suitable structure for end-users energy services. To guarantee a low computational time (especially regarding the parameter estimation), the structures are going to be tested from the most simple (with no capacitance) to the most complex (with 4 capacitances). The main difference between the different structures tested in the location of the inertia as described in the previous section. The four possible inertia locations are: the indoor air, the inner wall, the outer wall or what is called here a “fictive” wall. The “fictive” wall intervenes when the whole inertia in a unique equivalent capacitance.

Exterior wall	
Building block	Equation
	$T_{in} - T_{out} = R\phi$
	$T_{in} - T_{out} = \left( \frac{1}{\frac{1}{R_{out}} + \frac{\zeta_w}{R_w}} \right) \phi$
	$\frac{dT_{w,out}}{dt} = \frac{1}{C_{w,out}} \left[ \frac{T_{out}}{R_{w,out2}} + \frac{T_{in}}{R_{w,out1}} - T_{w,out} \left( \frac{1}{R_{w,out1}} + \frac{1}{R_{w,out2}} \right) \right]$ $T_{in} = \frac{R_{eq}}{R_{w,out1}} T_{w,out} + R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_w}{R_w} \right) T_{out}$ $\frac{1}{R_{eq}} = \frac{1}{R_{w,out1}} + \frac{1}{R_{out}} + \frac{\zeta_w}{R_w}$

where  $\phi$  represents the heat flow rate between the indoor air temperature  $T_{in}$  and the outdoor temperature  $T_{out}$

**Table II.2:** Building blocks for exterior wall

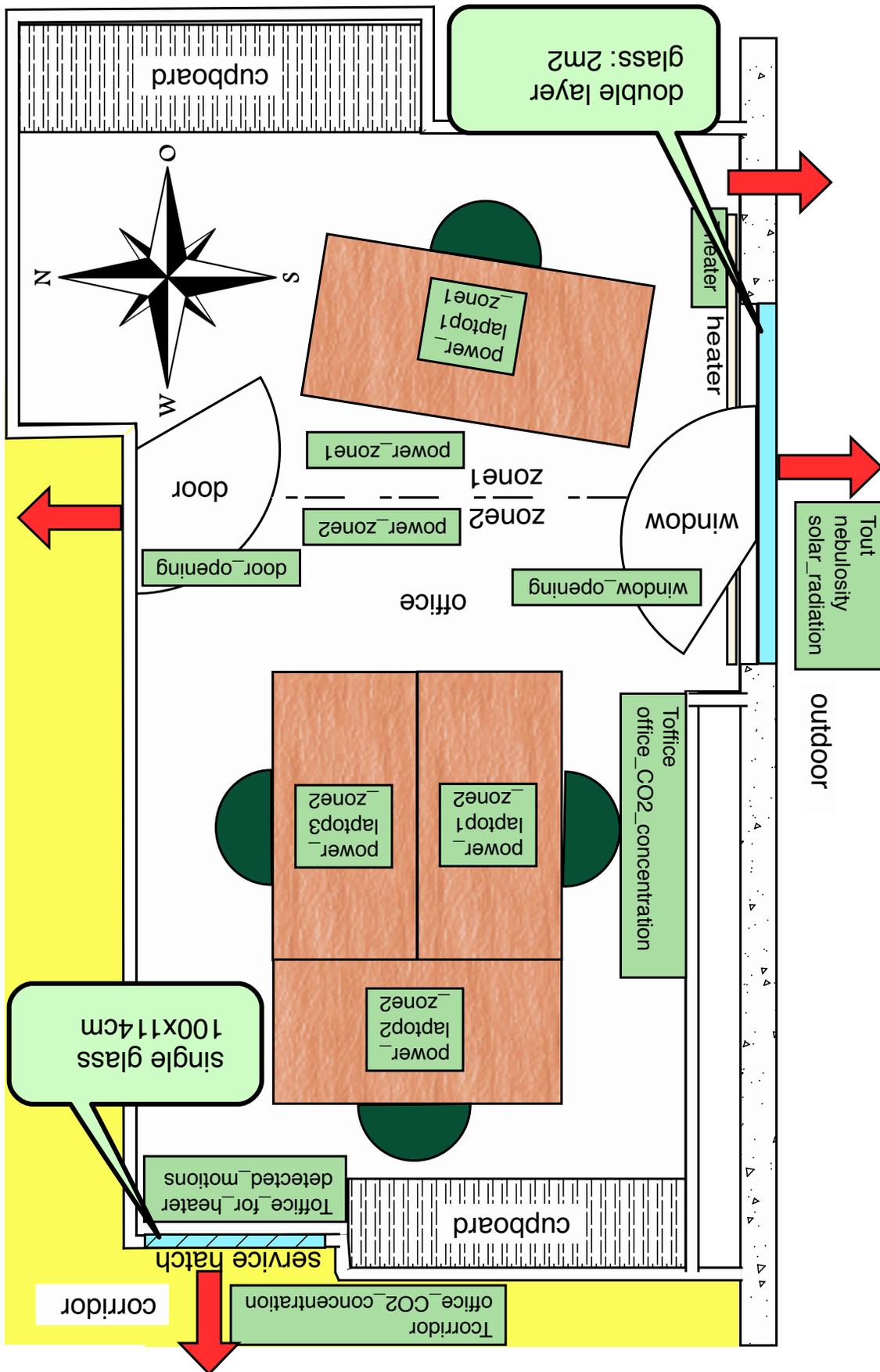


Figure II.1: Study case: researcher office

Inner wall	
Building block	Equation
	$T_{in} - T_n = R\phi$
	$T_{in} - T_n = \left( \frac{1}{\frac{1}{R_n} + \frac{\zeta_D}{R_D}} \right) \phi$
	$\frac{dT_{w,n}}{dt} = \frac{1}{C_{w,n}} \left[ \frac{T_n}{R_{w,n2}} + \frac{T_{in}}{R_{w,n1}} - T_{w,n} \left( \frac{1}{R_{w,n1}} + \frac{1}{R_{w,n2}} \right) \right]$ $T_{in} = \frac{R_{eq}}{R_{w,n1}} T_{w,n} + R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_W}{R_W} \right) T_n$ $\frac{1}{R_{eq}} = \frac{1}{R_{w,n1}} + \frac{1}{R_n} + \frac{\zeta_W}{R_W}$

where  $\phi$  represents the heat flow rate between the indoor air temperature  $T_{in}$  and the neighbouring air temperature  $T_n$

Table II.3: Building blocks for inner wall

Zone			
Building block	Equation	Building block	Equation
	$\frac{d\tau}{dt} = \frac{1}{C_i} \left( \frac{T_{in}}{R_i} - \frac{\tau}{R_i} \right)$		$\frac{d\tau}{dt} = \frac{1}{C_i} \left( \frac{T_{in}}{R_i} - \frac{\tau}{R_i} \right)$ $\frac{dT_{in}}{dt} = \frac{1}{C_{air}} \phi_{in}$

Table II.4: Building blocks for zone

## 2.3 Reference model

From these building blocks, a first simple structure is built locating the whole inertia in a unique equivalent capacitance (called later “fictive” wall). This simple structure is represented in figure II.2.

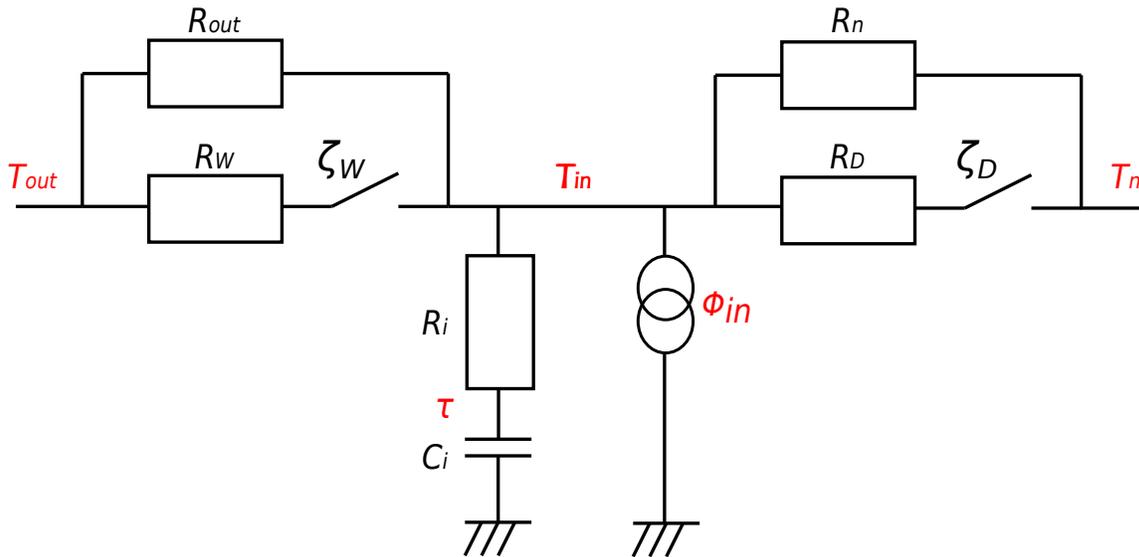


Figure II.2: Reference model

As a reminder, the notations used in the reference model in figure II.2 are explained below:

$T_n$ ,  $T_{in}$  and  $T_{out}$  represent respectively the temperature of the neighbouring room, the room considered and the outdoor

$R_n$  and  $R_{out}$  represent the resistance of the wall between the room and respectively the corridor and the outdoor

$R_D$  and  $R_W$  represent respectively the resistance of the door and the window

$\zeta_D$  and  $\zeta_W$  represent respectively the opening rate of the door and the window

$R_i$  and  $C_i$  represent respectively the resistance and the capacity of a fictive wall lumping the whole inertia of the building in the reference model

$\phi_{in}$  represents the internal gains: solar gains, heating gains, electric gains and occupancy gains

The thermal resistance of the door  $R_D$  and the thermal resistance of the window  $R_W$  are defined as below:

$$R_D = \frac{1}{\rho_{air} c_{p,air} Q_D} \quad (\text{II.9})$$

$$R_W = \frac{1}{\rho_{air} c_{p,air} Q_W} \quad (\text{II.10})$$

where  $\rho_{air}$  is the air density,  $c_{p,air}$  the specific heat capacity of the air and  $Q_D$  and  $Q_W$  respectively the air flow by the door and the window.  $Q_D$  and  $Q_W$  are part of the parameters which are estimated during the training phase.

To ease the reading of equations, an equivalent conductance is defined below:

$$\frac{1}{R} = \frac{1}{R_i} + \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} + \frac{1}{R_n} + \frac{\zeta_D}{R_D} \quad (\text{II.11})$$

It is then possible to define the general differential equation of the model Reference to determine  $\tau(t)$ :

$$\frac{d\tau(t)}{dt} = \frac{R}{C_i R_i} \left[ \left( \frac{1}{R_i} - \frac{1}{R} \right) \tau(t) + \phi_{in}(t) + \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) T_{out}(t) + \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) T_n(t) \right] \quad (\text{II.12})$$

As the model Reference includes only one state variable, it can be described by only one differential equation, but later the resolution of the state-space system will be detailed. The evolution of the indoor temperature in this case is then governed by the equation below:

$$T_{in}(t) = \frac{R}{R_i} \tau(t) + R \left( \frac{1}{R_{out}} + \frac{\zeta_W(t)}{R_W(t)} \right) T_{out}(t) + R \left( \frac{1}{R_n} + \frac{\zeta_D(t)}{R_D(t)} \right) T_n(t) + R \phi_{in}(t) \quad (\text{II.13})$$

with  $R_n$ ,  $R_{out}$ ,  $R_i$  and  $C_i$  time invariant

Solving the system of equations composed with equation II.12 et II.13 allows to forecast the evolution of the indoor temperature. To this thermal model is coupled a mass balance model presented in figure II.3. This model will be used for every model structure studied in this work, thus it will no be detailed again in the following descriptions.

From this simple aeraulic model, the following equations can be readily deduced:

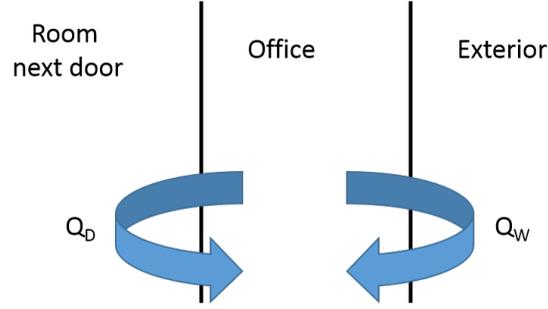


Figure II.3: Aeraulic model

$$V \frac{d\Gamma_{in}(t)}{dt} = -(Q_W(t) + Q_D(t))\Gamma_{in}(t) + Q_W(t)\Gamma_{out}(t) + Q_D(t)\Gamma_n(t) + S_{CO_2}^{body}n(t) \quad (\text{II.14})$$

where  $\Gamma_{in}, \Gamma_{out}$  and  $\Gamma_n$  represent the  $CO_2$  concentrations of the office, the outdoor and the neighbourhood,  $V$  is the volume of the room considered,  $S_{CO_2}^{body}$  is the average production of  $CO_2$  per person,  $Q_W$  and  $Q_D$  are the air flows from respectively the window and the door and  $n$  the number of occupants in the room.  $Q_W$  and  $Q_D$  are part of the parameters learnt during the training phase.

## 2.4 Model with no capacitor

The next step consisted in modifying the inertia modelling in order to measure the impact on the accuracy and computational time needed for the parameter estimation. Therefore, in a first time the capacity of the fictive was deleted. Considering that the aim is to predict the air temperature and  $CO_2$  concentration trends in the office (or dwelling) on a very short time period, it has been considered relevant to test a model without any capacitance (*cf.* figure II.4).

The model is thus linear, as shown below:

$$T_{in}(t) = R_{eq} \left[ \left( \frac{\zeta_D(t)}{R_D(t)} + \frac{1}{R_n} \right) T_n(t) + \left( \frac{\zeta_W(t)}{R_W(t)} + \frac{1}{R_{out}} \right) T_{out}(t) + \phi_{in}(t) \right] \quad (\text{II.15})$$

with:

$$\frac{1}{R_{eq}} = \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} + \frac{1}{R_n} + \frac{\zeta_D}{R_D} \quad (\text{II.16})$$

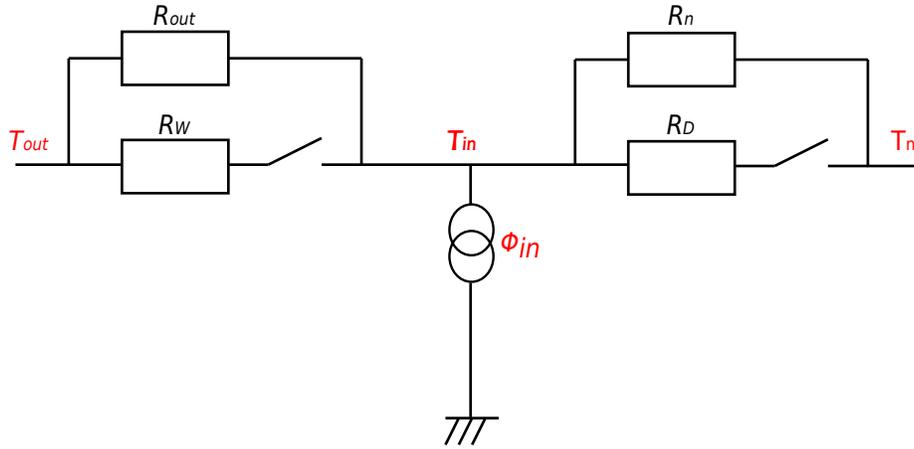


Figure II.4: Model with no capacitor

## 2.5 Model with two capacitors

Next, different structures with two capacitors have been defined to bring a second dynamics. In the first instance, a capacitor was added to model the air capacitance (quick dynamics) and then another capacitor to model either the inertia of the exterior wall or that of an inner wall.

### 2.5.1 Models with a wall capacitor

Therefore, the reference model has first been completed with a wall capacitor inserting a slow dynamics in the model (*cf.* figures II.5 et II.6).

#### 2.5.1. i) Exterior wall

In this case, the expression of the equivalent conductance is:

$$\frac{1}{R_{eq}} = \frac{1}{R_i} + \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} + \frac{1}{R_n} + \frac{\zeta_D}{R_D} + \frac{1}{R_{w1}} \quad (\text{II.17})$$

The resolution is done using a state-space representation which can be written in the following form:

$$\dot{X} = AX + BU \quad (\text{II.18})$$

$$Y = CX + DU \quad (\text{II.19})$$

where  $A$ ,  $B$ ,  $C$  and  $D$  are the matrix of the state-space system,  $U$  the input vector,  $X$  the state vector,  $\dot{X}$  the derivative state vector and  $Y$  the output vector.

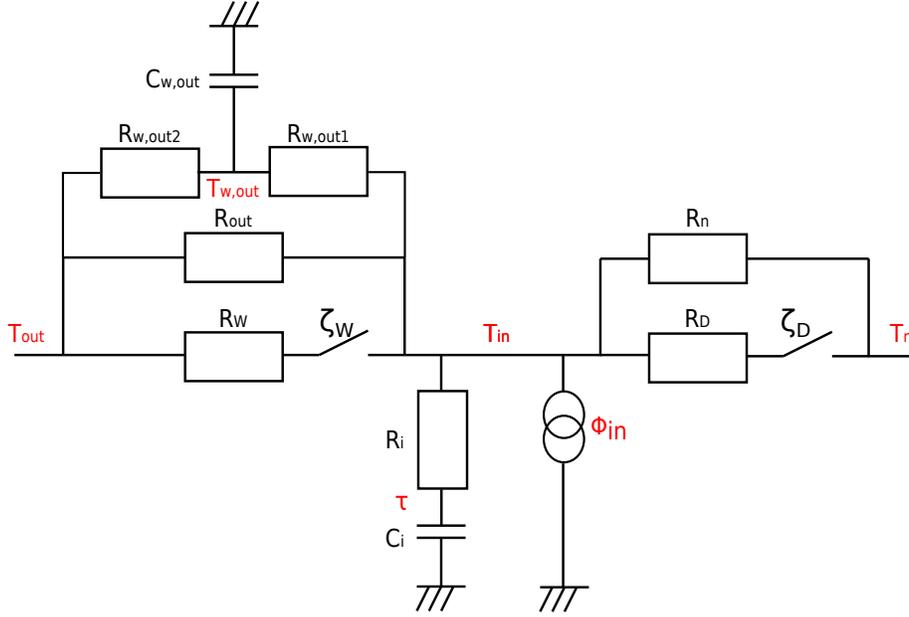


Figure II.5: Model with an exterior wall capacitor

In this case, the input vector is:

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (\text{II.20})$$

The state vector is constituted of all the temperatures directly linked to a capacitance:

$$X = \begin{bmatrix} \tau \\ T_{w,out} \end{bmatrix} \quad (\text{II.21})$$

and the output  $Y$  is the indoor air temperature  $T_{in}$ .

In these conditions, the state-space matrix  $A$ ,  $B$ ,  $C$  and  $D$  are:

$$A = \begin{bmatrix} \frac{R_{eq}}{C_i R_i^2} - \frac{1}{C_i R_i} & \frac{R_{eq}}{C_i R_i R_{wout1}} \\ \frac{R_{eq}}{C_{wout} R_{wout1} R_i} & \frac{R_{eq}}{C_{wout} R_{w1}^2} - \frac{1}{C_{wout}} \left( \frac{1}{R_{wout1}} + \frac{1}{R_{wout2}} \right) \end{bmatrix} \quad (\text{II.22})$$

$$B = \begin{bmatrix} \frac{R_{eq}}{C_i R_i} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & \frac{R_{eq}}{C_i R_i} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & \frac{R_{eq}}{C_i R_i} \\ \frac{R_{eq}}{C_{wout} R_{wout1}} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & \frac{R_{eq}}{C_{wout} R_{wout1}} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & \frac{R_{eq}}{C_{wout} R_{wout1}} \end{bmatrix} \quad (\text{II.23})$$

$$C = \begin{bmatrix} \frac{R_{eq}}{R_i} & \frac{R_{eq}}{R_{wout1}} \end{bmatrix} \quad (\text{II.24})$$

$$D = \begin{bmatrix} R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & R_{eq} \end{bmatrix} \quad (\text{II.25})$$

Yielding to:

$$T_{in} = \frac{R_{eq}}{R_i} \tau + \frac{R_{eq}}{R_{wout1}} T_{w,out} + R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) T_{out} + R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) T_n + R_{eq} \phi_{in} \quad (\text{II.26})$$

### 2.5.1. ii) Inner wall

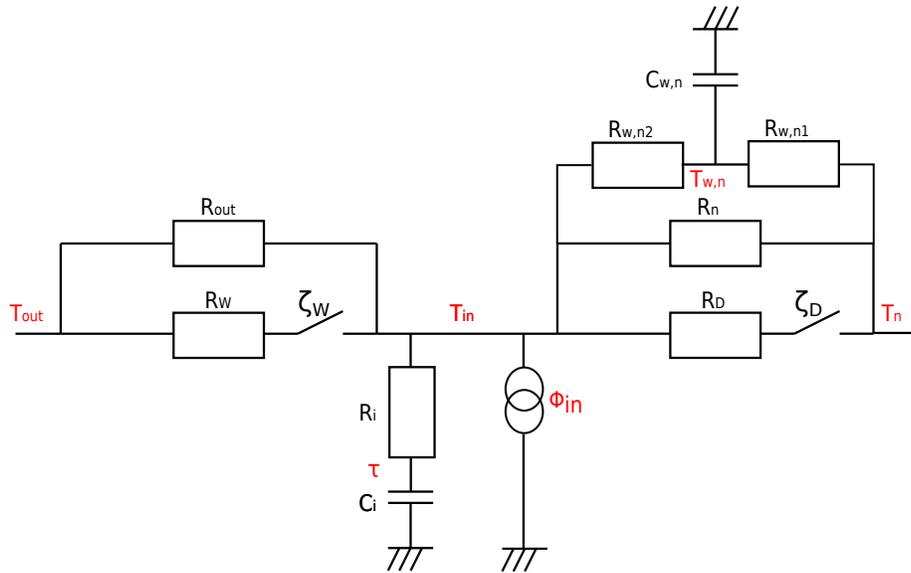


Figure II.6: Model with inner wall capacitor

For this structure, the equivalent conductance is:

$$\frac{1}{R_{eq}} = \frac{1}{R_i} + \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} + \frac{1}{R_n} + \frac{\zeta_D}{R_D} + \frac{1}{R_{wn1}} \quad (\text{II.27})$$

The structure is perfectly symmetric to the previous structure. Then, the input vector is:

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (\text{II.28})$$

The state vector is constituted of all the temperatures directly linked to a capacitance:

$$X = \begin{bmatrix} \tau \\ T_{w,n} \end{bmatrix} \quad (\text{II.29})$$

and the output  $Y$  is the indoor air temperature  $T_{in}$ .

In these conditions, the state-space matrix  $A$ ,  $B$ ,  $C$  and  $D$  are:

$$A = \begin{bmatrix} \frac{R_{eq}}{C_i R_i^2} - \frac{1}{C_i R_i} & \frac{R_{eq}}{C_i R_i R_{wn1}} \\ \frac{R_{eq}}{C_{wn} R_{w1} R_i} & \frac{R_{eq}}{C_{wn} R_{wn1}^2} - \frac{1}{C_{wn}} \left( \frac{1}{R_{wn1}} + \frac{1}{R_{wn2}} \right) \end{bmatrix} \quad (\text{II.30})$$

$$B = \begin{bmatrix} \frac{R_{eq}}{C_i R_i} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & \frac{R_{eq}}{C_i R_i} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & \frac{R_{eq}}{C_i R_i} \\ \frac{R_{eq}}{C_{wn} R_{wn1}} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & \frac{R_{eq}}{C_{wn} R_{wn1}} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & \frac{R_{eq}}{C_{wn} R_{wn1}} \end{bmatrix} \quad (\text{II.31})$$

$$C = \begin{bmatrix} \frac{R_{eq}}{R_i} & \frac{R_{eq}}{R_{wn1}} \end{bmatrix} \quad (\text{II.32})$$

$$D = \begin{bmatrix} R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) & R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) & R_{eq} \end{bmatrix} \quad (\text{II.33})$$

Yielding to:

$$T_{in} = \frac{R_{eq}}{R_i} \tau + \frac{R_{eq}}{R_{w1}} T_{w,n} + R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) T_{out} + R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) T_n + R_{eq} \phi_{in} \quad (\text{II.34})$$

### 2.5.2 Model with an air capacitor

An extra capacitor representing the dynamics of the air allows to add fast dynamics to the model. The corresponding structure is represented in figure II.7.

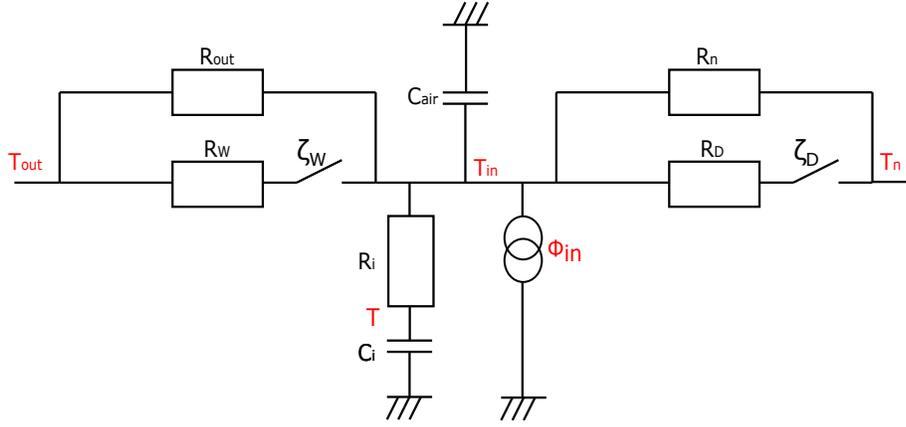


Figure II.7: Model with an air capacitor

Equations of the model are described below:

$$\frac{d\tau}{dt} = \frac{1}{C_i} \left( \frac{T_{in}}{R_i} - \frac{\tau}{R_i} \right) \quad (\text{II.35})$$

$$\frac{dT_{in}}{dt} = \frac{1}{C_a} \left[ \frac{1}{R_i} \tau + \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) T_{out} + \left( \frac{1}{R_n} + \frac{\zeta_D}{R_D} \right) T_n - \frac{1}{R_{eq}} T_{in} + \phi_{in} \right] \quad (\text{II.36})$$

Here again, the indoor temperature  $T_{in}$  is one of the state variables, so it follows naturally the resolution of the state space.

## 2.6 Models with 3 capacitors

The next step is to define structures with three capacitors, which has been done in two ways: combining the capacitance of the indoor air with a capacitor for either the inner wall or the exterior wall or combining both the capacitors of the inner and exterior walls. The modelling and the resolution of these models are done in the same way as for those with two capacitors detailed in section 2.5.

### 2.6.1 Models with an air capacitor

#### 2.6.1. i) Exterior wall

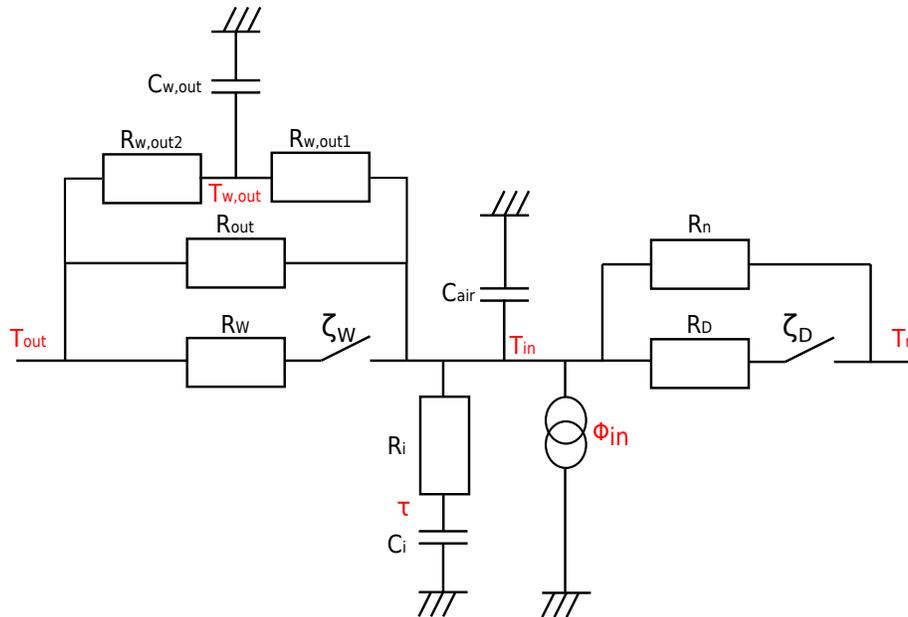


Figure II.8: Model with air and exterior wall capacitors

So we have:

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (II.37)$$

$$X = \begin{bmatrix} \tau \\ T_{w,out} \\ T_{in} \end{bmatrix} \quad (II.38)$$

#### 2.6.1. ii) Inner wall

So it yields to:

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (II.39)$$

$$X = \begin{bmatrix} \tau \\ T_{w,n} \\ T_{in} \end{bmatrix} \quad (II.40)$$

### 2.6.2 Model with wall capacitors

So we have:

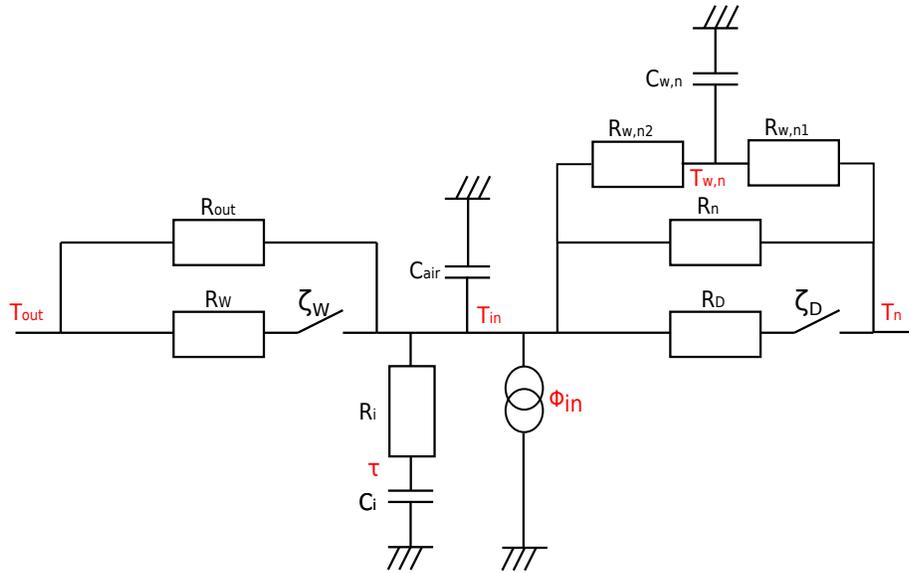


Figure II.9: Model with air and inner wall capacitors

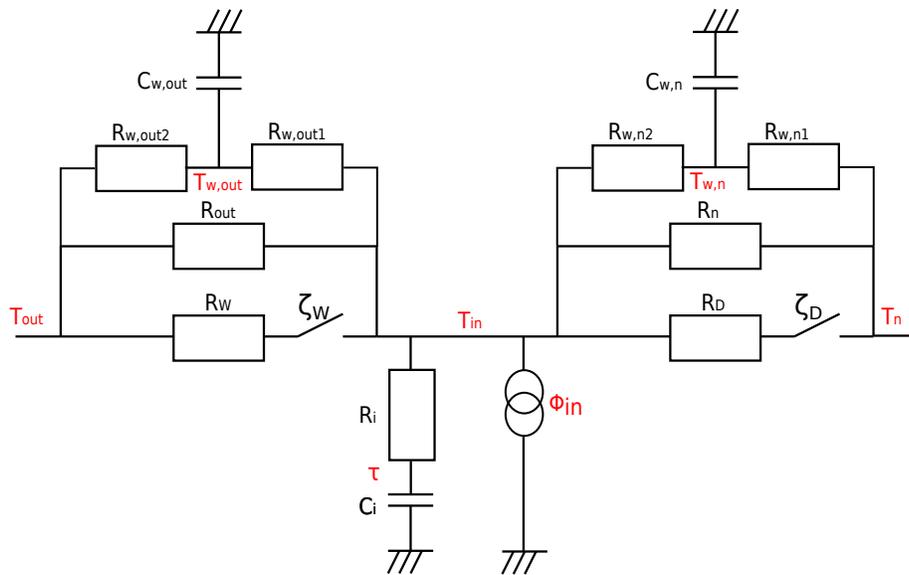


Figure II.10: Model with a capacitor for the inner and the exterior wall

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (II.41)$$

$$X = \begin{bmatrix} \tau \\ T_{w,out} \\ T_{w,n} \end{bmatrix} \quad (II.42)$$

## 2.7 Model with 4 capacitors

The model with 4 capacitors is the most complex of those appropriate for this study taking into account the data available and the willingness not to increase the number of sensors required for estimating the parameters. It is represented in figure II.11. It includes simultaneously the capacitor of the “fictive wall”, the ones of the inner and exterior walls and the one of the indoor air, in order to represent the most of the building dynamics.

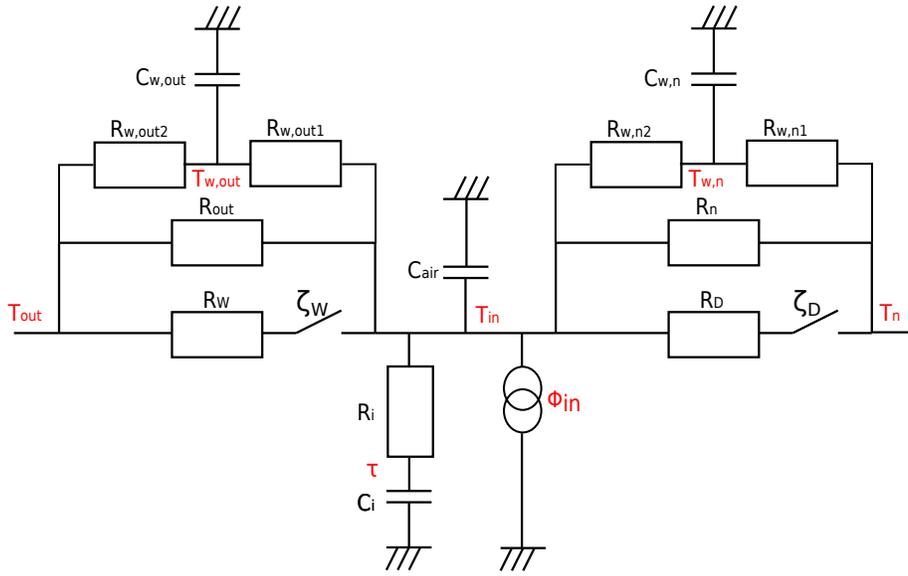


Figure II.11: Model with 4 capacitors

The input and the state variables for this model are described below:

$$U = \begin{bmatrix} T_n \\ T_{out} \\ \phi_{in} \end{bmatrix} \quad (II.43)$$

$$X = \begin{bmatrix} \tau \\ T_{w,out} \\ T_{w,n} \\ T_{in} \end{bmatrix} \quad (II.44)$$

## 2.8 Summary

Table II.5 summarizes the different structures studied with their name, the state variables, the data required and the number of parameters to estimate  $N$ . For each of the models detailed in section 2 of chapter II and the associated estimation methods, tests will be conducted to study the decrease of the number of sensors with the

Models	States	Observations	N
Reference	$\tau T_{in}$	$T_n T_{out} \zeta_D \zeta_W$	12
ReferenceNo $\zeta$	$\tau T_{in}$	$T_n T_{out}$	14
Model0C	-	$T_n T_{out} \zeta_D \zeta_W$	10
Model0CNo $\zeta$	-	$T_n T_{out}$	12
Model2Cair	$\tau T_{in}$	$T_n T_{out} \zeta_D \zeta_W$	13
Model2CairNo $\zeta$	$\tau T_{in}$	$T_n T_{out}$	15
Model2Cwall $T_{out}$	$\tau T_{w,out}$	$T_n T_{out} \zeta_D \zeta_W$	15
Model2Cwall $T_{out}$ No $\zeta$	$\tau T_{w,out}$	$T_n T_{out}$	17
Model2Cwall $T_n$	$\tau T_{w,n}$	$T_n T_{out} \zeta_D \zeta_W$	15
Model2Cwall $T_n$ No $\zeta$	$\tau T_{w,n}$	$T_n T_{out}$	17
Model3CT $_{out}$	$\tau T_{in} T_{w,out}$	$T_n T_{out} \zeta_D \zeta_W$	16
Model3CT $_{out}$ No $\zeta$	$\tau T_{in} T_{w,out}$	$T_n T_{out}$	18
Model3CT $_n$	$\tau T_{in} T_{w,n}$	$T_n T_{out} \zeta_D \zeta_W$	16
Model3CT $_n$ No $\zeta$	$\tau T_{in} T_{w,n}$	$T_n T_{out}$	18
Model3Cwall	$\tau T_{w,out} T_{w,n}$	$T_n T_{out} \zeta_D \zeta_W$	18
Model3CwallNo $\zeta$	$\tau T_{w,out} T_{w,n}$	$T_n T_{out}$	20
Model4C	$\tau T_{in} T_{w,out} T_{w,n}$	$T_n T_{out} \zeta_D \zeta_W$	19
Model4CNo $\zeta$	$\tau T_{in} T_{w,out} T_{w,n}$	$T_n T_{out}$	21

Table II.5: Summary of studied models

elimination of window and door contacts (*cf.* section 2.3 of chapter III). These models are designated by adding No $\zeta$  to the name of the model.

### 3 Acceptability and selection criterion △

Once the parameters of the models identified, several procedures of selection and validation of models are presented in the literature. They differ according to the objective(s) of the model. In this work, the procedure chosen is derived from Bacher and Madsen [9] and the PhD thesis of Reynders [78]. These steps allow to validate the robustness and the accuracy of the model identified but also to select the most suitable order.

### 3.1 Tools

To do so, different tools are needed: the standardized version of the Root Mean Square Error (sRMSE) and a sensitivity analysis. The RMSE enables to quantify the forecast error between the indoor simulated temperature and the measured one. It is defined by:

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{y}_k - y_k)^2} \quad (\text{II.45})$$

where  $y_k$  represents the measured value and  $\hat{y}$  the predicted one of the indoor temperature.

From here, the sRMSE is computed to allow the comparison of the model on different datasets:

$$sRMSE = \frac{RMSE}{y_{max} - y_{min}} \quad (\text{II.46})$$

A sensitivity analysis is also performed in order to ensure that every parameter is useful for forecasting the air temperature and the CO<sub>2</sub> concentrations in the case study as described in section 3.4 of chapter II.

### 3.2 Test of acceptability

The validation method consists in different steps:

- Validity of the model on the estimation phase
- Capacity of the model to forecast future behaviour
- Stability of the model along seasonal variations

For the first two steps, the sRMSE is computed and a limit to 0.1 is set to accept a model. Then, for the stability, the higher value on the validation phase is chosen between both summer and winter scenario. A range of  $\pm 20\%$  of this value is defined and if the other value does not fit in that range, the model is rejected.

### 3.3 Selection method

If the models verify every condition of the validation, they are then compared to each other in order to select the most suitable one. For that purpose, the sum of the sRMSE obtained for both summer and winter is computed and then compared. The selected model will be the one with the lowest value.

### 3.4 Sensitivity analysis of the parameters

To complete the validation process, a sensitivity analysis is performed in order to ensure that every parameter is useful for forecasting the air temperature and CO<sub>2</sub> concentrations in the case study. Otherwise, that would lead to a simplification of the model structure.

#### 3.4.1 Choice of the method

Different methods exist such as the Sobol's indices [85] which rely only on the hypothesis that the variance and the expectation of the output are finite. However, despite the few hypotheses needed, it requires a high number of simulations so it is a very interesting method for models with low computation cost. Another method was developed by Saltelli et al [80] in 1999; it is based on the principles of the Fourier analysis. It enables to write the variance as a Fourier series. Quicker and more stable than the Sobol method, it still requires a lot of simulations:  $N \cdot d$  with  $N$  the number of samples and  $d$  the number of factors. Also available, the method of Morris [68] relies on a discretization of the space of variables. Thus, the sensitivity of the output to one of the factors  $X_i$  is measured comparing results where only this parameter  $X_i$  varied. This is why this method is classified amongst the "One at a Time" (OAT) methods. The number of simulations required is equal to  $(N + 1)^2$  where  $N$  is the number of samples. It is then faster than the Sobol and FAST (Fourier amplitude sensitivity testing) methods. However, contrary to the other methods, the Morris method is qualitative which means that it is possible to classify the different parameters but not to ensure their relevance. As the goal of the sensitivity analysis in this research is to ensure that the structure of the model is coherent and that each parameter has an impact on the output, it is not suitable. Therefore, despite its computational cost, the Sobol method will be implemented.

### 3.4.2 Principle

The method of Sobol is a global and model-independent sensitivity analysis method based on variance decomposition. It determines the contribution of each input parameter and their interactions to the overall model output variance. In the general framework of a non-linear and non-monotonous model, if the output can be deduced from the set of parameters  $x = (x_1, x_2, \dots, x_d)$  by the function  $f$ , Sobol suggested to decompose that function into summands of increasing dimensionality:

$$f(x_1, x_2, \dots, x_d) = f_0 + \sum_i^d f_i(x_i) + \sum_{i<j}^d f_{ij}(x_i, x_j) + \dots + f_{12\dots d}(x_1, x_2, \dots, x_d) \quad (\text{II.47})$$

where  $f_0$  is a constant and this condition is verified:

$$\int_0^1 f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0 \forall k = 1, \dots, s, \forall \{i_1, \dots, i_s\} \subseteq \{1, \dots, d\} \quad (\text{II.48})$$

Sobol demonstrated then that this decomposition is unique. From here, the total variance decomposition can be obtained:

$$\text{Var}[Y] = \sum_{i=1}^d V_i(Y) + \sum_{i<j}^d V_{ij}(Y) + \sum_{i<j<k}^d V_{ijk}(Y) + \dots + V_{12\dots d}(Y) \quad (\text{II.49})$$

where:  $N$  is the number of samples,  $d$  the number of factors,  $Y$  the output of the system and  $V_i(Y) = \text{Var}[\mathbb{E}(Y | x_i)]$ ,  $V_{ij}(Y) = \text{Var}[\mathbb{E}(Y | x_i x_j)] - V_i(Y) - V_j(Y)$  and so on.

This writing of the variance allows to easily obtain the sensitivity indices or Sobol's indices:

$$S_i = \frac{\text{Var}[\mathbb{E}(Y | x_i)]}{\text{Var}(Y)} = \frac{V_i(Y)}{\text{Var}(Y)}, \quad S_{ij} = \frac{V_{ij}(Y)}{\text{Var}(Y)}, \quad S_{ijk} = \frac{V_{ijk}(Y)}{\text{Var}(Y)}, \quad \dots \quad (\text{II.50})$$

The Sobol's indices are normalized, and so very easily interpreted. The second-order contribution illustrates the sensitivity of the model to the interaction between the parameters  $x_i$  and  $x_j$ . If the number of parameters increases and exceeds 10, the computation of these indices becomes really costly. This is why Homma and

Saltelli [48] introduced the total sensitivity indices such as:

$$S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i, k \neq i, j < k} S_{ijk} + \dots = \sum_{l \in \#i} S_l \quad (\text{II.51})$$

where  $\#i$  represents all the subsets of indices containing the index  $i$ . In our study, according to the number of parameters to estimate in our models, only the first-order and the total indices will be computed and analysed.

## 4 Conclusion



In this chapter, a general methodology to generate models has been introduced. First, the mono-zone office study case is presented with an architectural point of view as well as different useful tools: estimators and the computation of the solar gains. Then, a focus is done on the structure of the semi-physical models studied in chapter III. To do so, an ontology for dwelling is proposed followed by a decomposition into different building blocks developed in agreement with both this ontology and the literature. From here, the different blocks are assembled into the 8 model structures, which will be analysed considering our purpose: models for end-user services. To conclude, the acceptability and selection methods used to elect the most relevant structure are described.



## Chapter III

---

# Learning behavioural models for one day time horizon



### Contents

<b>1</b>	<b>Introduction</b>	52
<b>2</b>	<b>End-user energy services requirements</b>	53
<b>3</b>	<b>Multi-regression models for energy services</b>	54
3.1	Implementation	54
3.2	The case of June	60
3.3	Multi-regression inspired from physics	62
3.4	Limits	66
3.5	Conclusion	67
<b>4</b>	<b>Meta-optimization</b>	67
4.1	Principle of the meta-optimization	67
4.2	Application to the different structures	69
4.3	Limits	75
4.4	Conclusion	78
<b>5</b>	<b>Meta-heuristic</b>	79
5.1	Principle	80
5.2	Configuration	81
5.3	Test of sensibility to the initial point	83

5.4	Application to the different model structures . . . . .	84
5.5	Sensitivity analysis . . . . .	87
5.6	Observability of the model . . . . .	96
5.7	Reliability of the model along the seasons . . . . .	99
5.8	Physical interpretation . . . . .	101
6	Conclusion . . . . .	105

## 1 Introduction



As explained in section 5 of chapter I, two different categories of models are implemented in this work: black-box models such as multi-regression models and semi-physical models such as RC-models. Indeed, recovering physical values of the model parameters is not an objective in itself. Then, it is logical to first implement easily estimated non-physical models such as black-box models. Their application is not common for forecast purposes their performance in prediction is arguable. But, as shown in section 3 of chapter I, they have gained in popularity for their low computational costs for the estimation process.

Semi-physical models, on the other hand, are a common solution in the building physics community, then this option is explored as well. But, using physical models for energy management in buildings requires a calibration step since there is no expert information (insulation, materials,...) available on the specific building in which the solution will be implemented. This step will be realized in this study with measured data. Different parameter estimation methods exist: some of them demand expert knowledge to select a relevant initial dataset such as usual descent algorithms in a rough parameter space while others require a linear model such as linear regressions. Optimization methods can be not ergodic (which means that the algorithm does not systematically converge towards the same optimized dataset). The choice of the parameter estimation method is impacted by the linearity with respect to model parameters of the problem. Considering the reference model with

one capacity, it can be seen than if it is linear with regards to its variables, it is not with regard to its parameters. Then, a certain number of optimization methods are not suitable for estimating parameters. Optimization algorithms can be sorted into two main categories: the stochastic ones and the deterministic ones. The descent algorithms, part of the deterministic category, are based on the computation of the gradient of the objective function. Thus, according to the initial point, it can be stuck in a local optimum but it is less costly in terms of computational time than genetic algorithms. That is why a "meta optimization" procedure is implemented to try to avoid that problem and keep a low computational cost.

In this chapter, are tested whether these two kinds of models as well as the two different estimation methods described above for semi-physical models are suitable for end-user energy services or not. First, the requirements induced by the implementation of end-user energy services are detailed. Then, the meta-optimization approach is described and applied on the office mono-zone study case, followed by the meta-heuristic approach.

## 2 End-user energy services requirements

The final objective here is to allow the implementation for end-user energy services in order to help them understand the thermal phenomenon in their dwellings. The services considered are not about control of the house but generation of anticipative plans for the next 24 hours for advice and explanations of past behaviour. Occupants having very little knowledge about the intrinsic building physical parameters, the thermal models developed must be mainly based on data. The main target of the project is the residential sector and the goal is to obtain a model easily extrapolated to other kind of buildings. Then, it is aimed to implement these services in every house. This leads to another constraint: the instrumentation must be as minimal as possible. In order to ease the implementation and to allow a calibration for each dwelling without requiring an expert, models must be self-tuning. Most of the services based on predictions are built on a 24-hour horizon so the model can be daily set-up to measurements to avoid any drifts. Indeed, as end-users are targeted and services considered are not about control, the 24 hour horizon appears to be the most relevant for providing advice, anticipative plans of energy management and explanations. The resolution considered is 1 hour time-step as there is no point in giving advice of actions every 10 minutes, hence the quick dynamics do not signifi-

cantly impact the models. The most important here is to properly model the slow thermal dynamics of the building. As explained in the introduction of this chapter, models have to be suitable for learning process. This means that the computational time required for estimating the parameters the first time must be less than 24 hours and the length of data required as small as possible since it cannot be expected from occupants to wait until they have one-year of data before being able to use the energy services. It also means that the estimation process must be ergodic: the estimation provided by the services needs to be always as accurate as ever. Otherwise, the user's trust in the e-consultant would be definitely damaged.

To meet all these technical specificities, the methodology implemented consists first in defining the right kind of model and the right model structure to properly represent the building thermal dynamics. And, in parallel, to ensure that this model structure can be coupled with a parameter estimation method which is not too costly in computational time and is ergodic.

## 3 Multi-regression models for energy services △

Firstly, multi-regression models have been applied to the mono-zone office described in section 1 of chapter II for both summer and winter periods. The first step consists in defining the training and cross-validation datasets. The inputs and outputs of the model are then determined as well as the error criteria considered.

### 3.1 Implementation

#### 3.1.1 Training and validation datasets

The summer scenario consists in estimating the model for the month of May 2015 and to validate it for the month of June 2015. June is a fairly hot month in Grenoble, with maximum outdoor temperatures oscillating between 20 and 32°C as can be seen in figure III.1a. Measured data take into account a great number of openings of windows and door, with a clear impact on the office temperature and CO<sub>2</sub> concentrations. The winter scenario consists in estimating the models using the month of October 2015 and in validating it with the month of November 2015. Here again the separation is materialized by a change of colours on the graphs: blue for the training phase and red for the cross-validation phase. This period is characterised

by a relatively high indoor temperature for the season: around 22°C for the month of October and by an abrupt fall of temperatures in November, down to 16°C. The outdoor temperature oscillates between 0 and 15°C (*cf.* figure III.1b), which suggests a significant usage of the heaters.

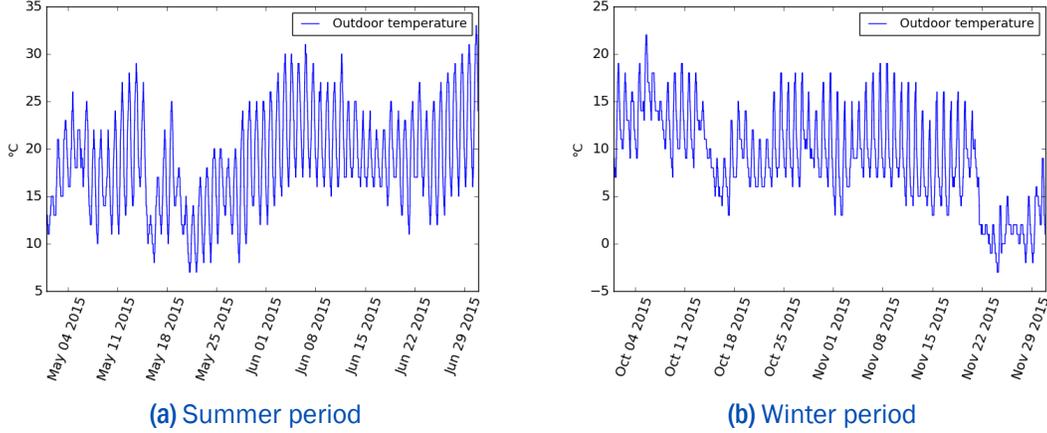


Figure III.1: Outdoor temperature

### 3.1.2 Choice of the inputs

For both the temperature and CO<sub>2</sub> models, inputs need to be defined. For the temperature model, the inputs considered are: the outdoor temperature ( $T_{out}$ ), the temperature of the corridor ( $T_{corridor}$ ), the hourly rate of the door opening ( $\zeta_D$ ) varying from 0 to 1, the hourly rate of the window opening ( $\zeta_W$ ) varying from 0 to 1 and the internal gains: the total electric power ( $P_{elec}$ ) and the solar gains ( $\phi_{solar}$ ). For the CO<sub>2</sub> model, the variables of interest are: corridor CO<sub>2</sub> concentration ( $\Gamma_{corridor}$ ), the hourly rate of the door opening ( $\zeta_D$ ) varying from 0 to 1, the hourly rate of the window opening ( $\zeta_W$ ) varying from 0 to 1 and the CO<sub>2</sub> sources ( $S_{CO_2}^{body}n$ ). The internal sources are not considered as inputs since they are not directly measured in this study case and should then be estimated. The structure of the temperature model is:

$$\widehat{T_{in}(t)} = b_0 T_{out}(t) + b_1 T_{corridor}(t) + b_2 \zeta_D(t) + b_3 \zeta_W(t) + b_4 \phi_{solar}(t) + b_5 P_{elec}(t) \quad (III.1)$$

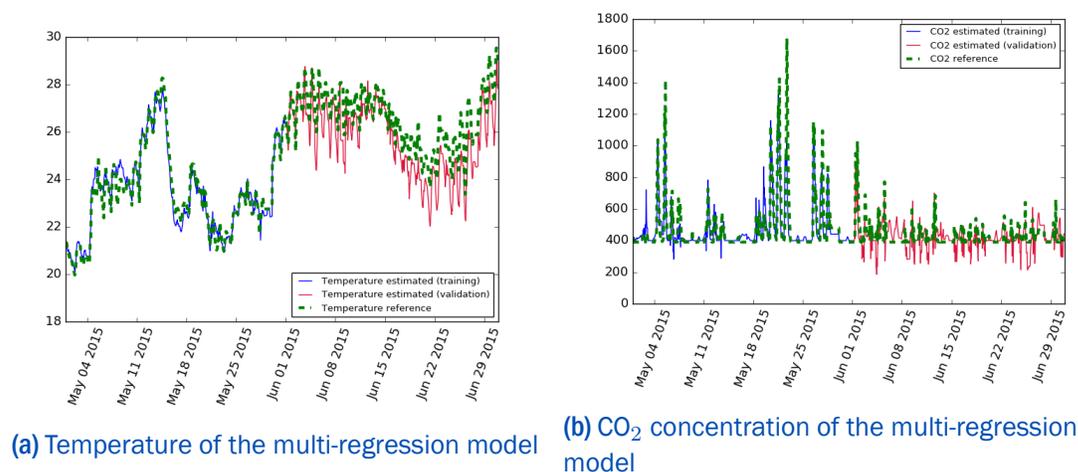
and the one of the CO<sub>2</sub> model:

$$\widehat{\Gamma_{in}(t)} = b_0 \Gamma_{corridor}(t) + b_1 \zeta_D(t) + b_2 \zeta_W(t) + b_3 S_{CO_2}^{body}n(t) \quad (III.2)$$

where  $b_i$  represent the coefficients to estimate and the  $\hat{\phantom{x}}$  represents the estimation.

### 3.1.3 Results

Results of the models for the temperature and the CO<sub>2</sub> concentration in the mono-zone study case can be seen in figure III.2.



**Figure III.2:** Results of prediction for the multi-regression models including the internal gains

It can be observed that the CO<sub>2</sub> concentration values obtained during the validation are significantly different from the measurements and most of all unrealistic. Indeed, the outdoor CO<sub>2</sub> concentration is around 400 ppm and cannot be lower. With this model, CO<sub>2</sub> concentrations reach almost 2000 ppm. The validation is performed by recovering the coefficients  $b$  estimated based on the training data and launch the simulation for the following month. The order of the multi-regression models is zero, then there is no daily calibration. Indeed, the prediction being independent of the past events, there is no risk of drifts.

In order to generalize the result, the same protocol has been applied to the months of April and May 2015 instead of May and June. The results are presented in figure III.3. It can be observed that the CO<sub>2</sub> concentration estimation is greatly improved: the number of unrealistic values is close to zero and the amplitude is correct. The same thing is observed for the temperature estimation. It is more accurate and there is less fluctuations during the validation phase than previously.

Regarding the winter scenario, results are presented in figure III.4. It can be seen that results are pretty accurate for both CO<sub>2</sub> concentration and air temperature. Even if there is a gap of temperature between the month of September and the month

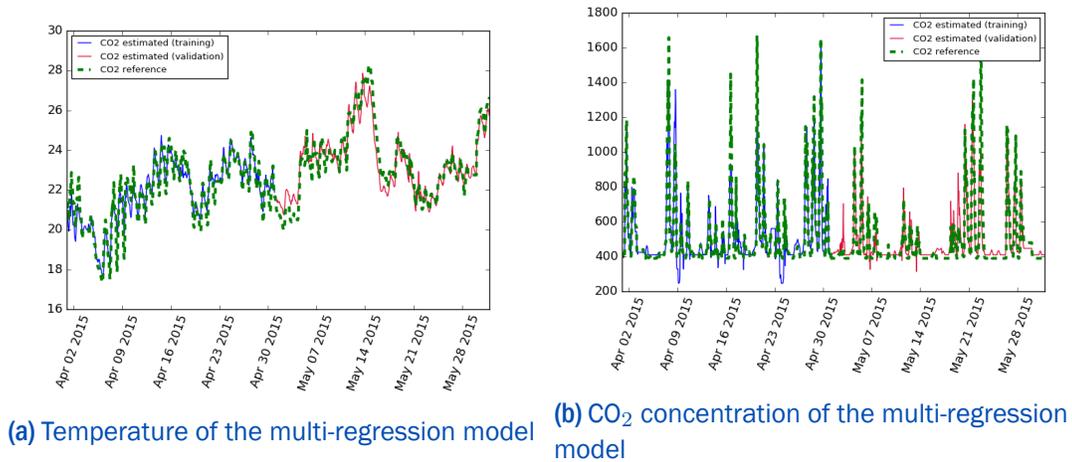


Figure III.3: Results of the multi-regression model during April and May

of October, the model successfully forecast the evolution of temperature during the validation phase.

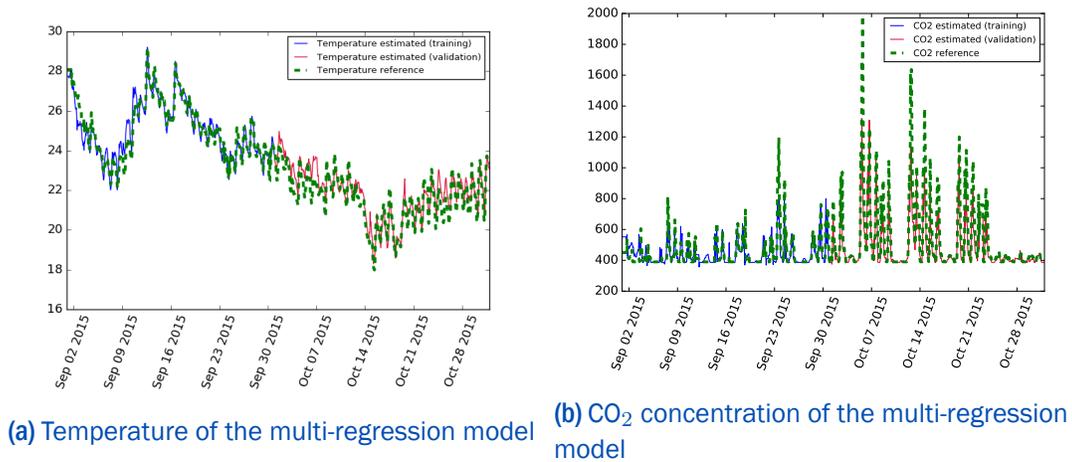


Figure III.4: Results of the multi-regression model during September and October

From a quantitative point of view, two elements are observed: the absolute average error between the measured and the estimation values and the standard deviation. They are observed for both the training phase and the the validation phase. This information is summarized in table III.1 where the absolute average is expressed in °C for the temperature and in ppm for the CO<sub>2</sub> concentration.

Months	Data	Absolute Average Error	Standard Deviation
May - June	Temperature (°C)	<i>0.29</i> 0.95	<i>0.38</i> 0.58
	CO <sub>2</sub> (ppm)	<i>46</i> 59	<i>88</i> 80
April - May	Temperature (°C)	<i>0.39</i> 0.47	<i>0.52</i> 0.55
	CO <sub>2</sub> (ppm)	<i>57</i> 50	<i>90</i> 89
Sept - Oct	Temperature (°C)	<i>0.32</i> 0.44	<i>0.44</i> 0.49
	CO <sub>2</sub> (ppm)	<i>34</i> 58	<i>57</i> 103

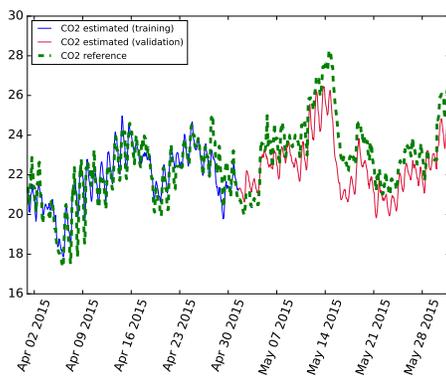
**Table III.1:** Comparison of absolute average error and standard deviation for the reference model and the multi-regression inspired by physics on both training (in italics) and validation phases

As a test, other models are implemented, deleting the internal sources (of heat and CO<sub>2</sub>) as shown below:

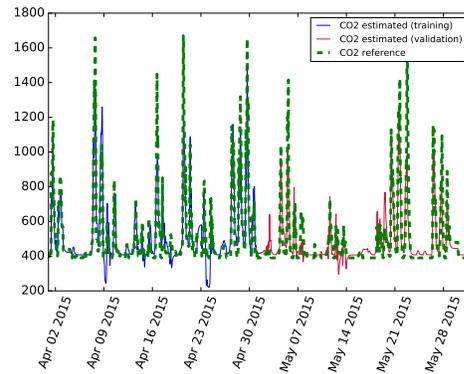
$$\widehat{T_{in}(t)} = b_0 T_{out}(t) + b_1 T_{corridor}(t) + b_2 \zeta_D(t) + b_3 \zeta_W(t) \quad (III.3)$$

$$\widehat{\Gamma_{in}(t)} = b_0 \Gamma_{corridor}(t) + b_1 \zeta_D(t) + b_2 \zeta_W(t) \quad (III.4)$$

where  $b_i$  represent the coefficients to estimate and the  $\widehat{\phantom{x}}$  represents the estimation.



(a) Temperature of the multi-regression model



(b) CO<sub>2</sub> concentration of the multi-regression model

**Figure III.5:** Results of the multi-regression model without the internal gains

Results can be seen in figure III.5 and are very similar to the ones obtained with the internal gains. This result is quite surprising and requires further investigation. Then, the coefficients of the different terms are observed in order to see their respective impact. For the temperature model with the internal gains, the impacts are expressed below in percentage:

- $T_{out}[k]$ : 2.10%
- $T_{corridor}[k]$ : 92.4%
- $\zeta_D[k]$ : 0.157%
- $\zeta_W[k]$ : 0.225%
- $\phi_{solar}[k]$ : 4.84%
- $P_{elec}[k]$ : 0.309%

Then, it can be observed that the temperature of the corridor is the main contributor to the indoor air temperature of the office. This highlights that there is a strong correlation between these two temperatures that can hide the impacts of the different actions. In this case, multi-regression models are not suitable for end-user energy services in a mono-zone study case. To confirm this, the estimated model represented by the equation III.1 has been used for the simulation of the month of April considering that the window stay open all the time. This simulation has been compared with the simulation of the model taking the measured window opening. The difference of these two different models is plot in figure ???. It highlights that opening the window has no impact on the indoor air temperature despite the low outdoor temperatures which can reach negative temperatures.



Figure III.6: Difference of temperatures between the two scenarios tested

This study shows that multi-regression models do not seem to be suitable for end-user energy services. Despite this observation, the case of June is further investigated in the next section in order to understand the specific behaviour observed for this month. Later, other models are implemented integrating more physics in order to study if it helps to correct the importance of the temperature of the corridor.

## 3.2 The case of June

Multi-regression models have been proved to deliver acceptable performance for forecasting air temperature and CO<sub>2</sub> concentration in some cases but to present significant errors in others. Regarding the CO<sub>2</sub>, the case of June requires further investigation to understand the observed behaviour. For that purpose, environmental inputs are plotted in order to identify whether any specific phenomena happened during this month. Then, window and door openings, outdoor and indoor air temperatures as well as CO<sub>2</sub> concentrations in the corridor are observed (cf. figure III.7). It can be seen that while the door behaviour is quite similar: graphically the behaviour is quite the same during May, June and July; there is a gap between the window openings in May and those in June which is consistent with the significant increase in outdoor temperatures. Indeed it can be observed in figure III.7b that there are very few window openings between the 7th of May and the 4th of June, whereas the months of June and July are characterized by a high rate of openings. The opening rate is three times higher in June than in May. This can explain why the multi-regression model is not able to forecast the CO<sub>2</sub> concentrations in June.

Then, another test is launched, estimating the model on the month of July which also presents many window openings and similar outdoor conditions with validation on the month of June. Results in figure III.8 reveal that the prediction is still not as accurate as it was for winter forecasts but there is a significant improvement with an average absolute error of 34.54 ppm against 72.62 ppm and a standard deviation of 64.52 ppm against 96.60. This study highlights that multi-regression models require rich datasets for learning the coefficient of the variables of interest and that they cannot ensure a systematic accuracy in the estimations.

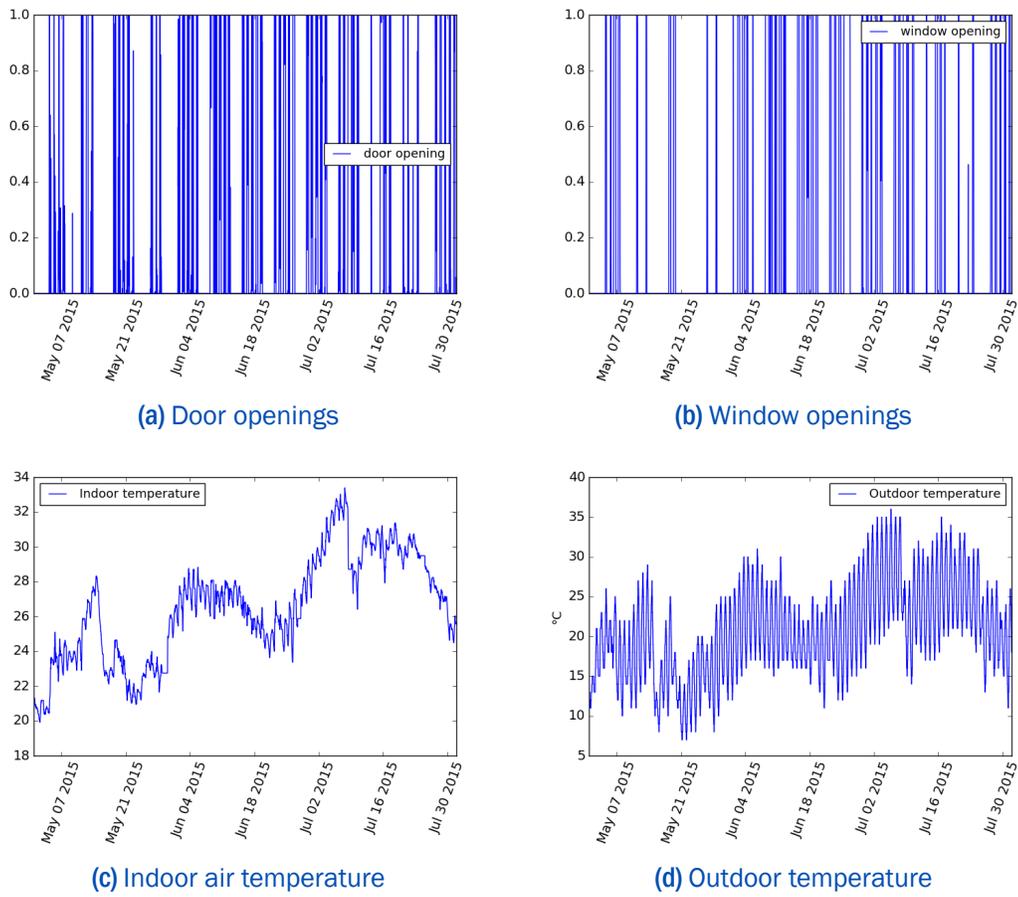


Figure III.7: Environmental inputs during the month of June

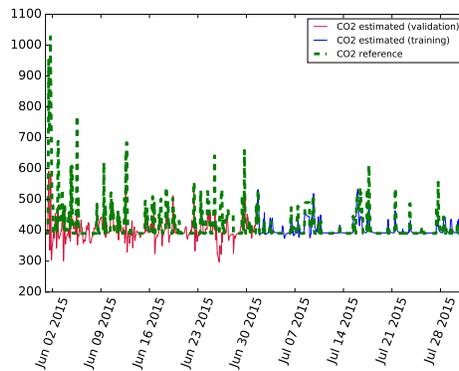


Figure III.8: CO<sub>2</sub> concentrations of the multi-regression model estimated during the month of July and cross-validated in June

### 3.3 Multi-regression inspired from physics

In order to solve problems such as the case studied in the previous section, an issue is explored: how to integrate an *a priori* physical knowledge in the multi-regression model structure and how it can improve the accuracy of the estimations. The low performances observed for the month of June may indeed be due to new physical phenomena (absent from the training data) but these phenomena may be accounted for by physical equations.

#### 3.3.1 Principle

The main idea is to improve the prediction ability of multi-regression models when facing new dynamics or behaviours and to develop an automatic method for determining the inputs to consider and the model order. Indeed, if black-box models have the advantage of not requiring any specific prior knowledge of the building, their structure needs to be properly selected [71]. For that purpose, physical equations can be extracted either from grey-box models such as electric analogy models (RC models) presented in figure III.9 for forecasting the temperature or from mass balance equations for forecasting the CO<sub>2</sub> concentration.

In the RC-model,  $T_n$ ,  $T_{in}$  and  $T_{out}$  represent respectively the temperature of the neighbouring room, the room considered and the outdoor  
 $R_n$  and  $R_{out}$  represent the resistances of the wall between the room and respectively the corridor and the outdoor  
 $R_D$  and  $R_W$  represent respectively the resistances of the door and the window  
 $\zeta_D$  and  $\zeta_W$  represent respectively the opening rate of the door and the window  
 $R_i$  and  $C_i$  represent respectively the resistance and the capacity of the fictive wall  
 $\phi_{in}$  represents the internal gains : solar gains, electric gains, heating gains and occupancy gains.

Wu and Sun [92] tackled the same questions but using thermodynamics equations to enrich ARMAX models. They predicted the indoor air temperature in the Science and Engineering (SE) building of UC Merced. They succeeded in precisely forecasting the room temperature over both short and long-term periods from models trained with the data over a relatively short time.

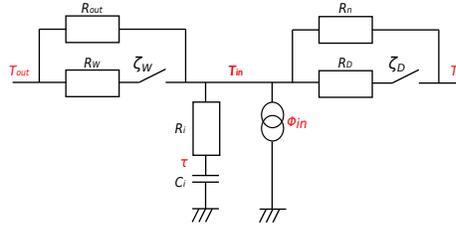


Figure III.9: RC model

The simple mass balance model used for forecasting the  $CO_2$  concentration is presented schematically in figure III.10 and mathematically in the three following equations:

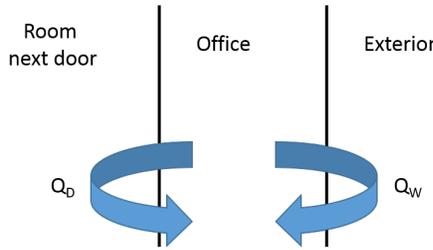


Figure III.10: Aeraulic model

$$V \frac{d\Gamma_{in}}{dt} = - (Q_W + Q_D)\Gamma_{in} + Q_W\Gamma_{out} + Q_D\Gamma_n + S_{CO_2}^{body}n \quad (III.5)$$

$$Q_{out}(t) = Q_{out}^0 + \zeta_W(t)Q_W \quad (III.6)$$

$$Q_{corridor}(t) = Q_{corridor}^0 + \zeta_D(t)Q_D \quad (III.7)$$

where  $\Gamma_{in}, \Gamma_{out}$  and  $\Gamma_n$  represent the  $CO_2$  concentrations of the office, the outdoor and the neighbourhood,  $V$  is the volume of the room considered,  $S_{CO_2}^{body}$  is the average production of  $CO_2$  per person,  $Q_W$  and  $Q_D$  are the air flows from respectively the window and the door and  $n$  the number of occupants in the room.  $Q_W$  and  $Q_D$  are part of the parameters learnt during the training phase.

### 3.3.2 Application

To apply this principle to the multi-regression model as defined in section 3, one had first to identify those variables from the mass balance model that were relevant

to the physics-inspired multi-regression. The training and validation periods had then to be properly defined to guarantee a rich enough dataset.

### 3.3.2. i) *Determination of coefficients*

Equation III.5 represents a simplified physical CO<sub>2</sub> behaviour of the mono-zone office. This equation is parsed in order to recover the variables of interest. Firstly, the variables  $Q_W$  and  $Q_D$  are replaced in equation III.5 according to equations III.6 and III.7. The parameters  $Q^0$  are part of the parameters to estimate. This induces the appearance of terms including  $\zeta_D$  and  $\zeta_W$ . As the structure of a multi-regression model is linear, terms containing  $\Gamma_{in}$  are not considered here. Only the coefficients of measurable inputs are kept:  $\Gamma_{cor}$ ,  $n$ ,  $\zeta_D$  and  $\zeta_W$  for the estimation of the CO<sub>2</sub> and  $T_{out}$ ,  $T_{corridor}$ ,  $n$ ,  $\zeta_D$  and  $\zeta_W$  for the temperature estimation. This yields the variables listed below for the CO<sub>2</sub> model:

- $\Gamma_{corridor}$ : the CO<sub>2</sub> concentration of the corridor
- $\zeta_D$ : the rate of door openings
- $\zeta_W$ : the rate of window openings
- $\Gamma_{corridor}\zeta_D$ : the product between the rate of door openings and the CO<sub>2</sub> concentration of the corridor
- $n$ : the occupancy

The term  $\Gamma_{out}\zeta_W$  has not been taken in consideration as  $\Gamma_{out}$  is constant. It has then been replaced by the term  $\Gamma_{corridor}\zeta_W$ . and the variables listed below for the temperature model:

- $T_{corridor}$ : the air temperature of the corridor
- $\zeta_D$ : the rate of door openings
- $\zeta_W$ : the rate of window openings
- $T_{corridor}\zeta_D$ : the product between the rate of door openings and the air temperature of the corridor
- $T_{out}\zeta_W$ : the product between the rate of window openings and the outdoor temperature

- $n$ : the occupancy

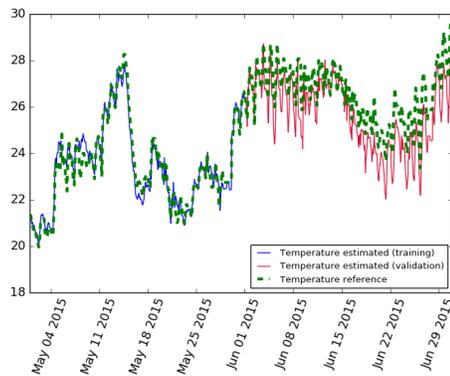
These analyses led to the following models :

$$\Gamma_{in} = b_0\Gamma_{corridor} + b_1\zeta_D + b_2\zeta_W + b_3n + b_4\Gamma_{corridor}\zeta_D + b_5\Gamma_{corridor}\zeta_W \quad (III.8)$$

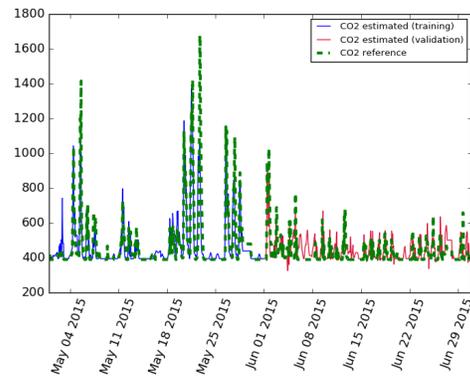
$$T_{in} = b_0T_{out} + b_1T_{corridor} + b_2\zeta_D + b_3\zeta_W + b_4T_{corridor}\zeta_D + b_5T_{out}\zeta_W + b_6\phi_{solar}(t) + b_7P_{elec}(t) \quad (III.9)$$

### 3.3.2. ii) Implementation

The coefficients of the variables of interest in the multi-regression model inspired by physics can be estimated and the resulting model can be simulated. Results can be seen in figure III.11 and values of the absolute average error and standard deviation are shown in table III.2. Regarding the temperature model, the new structure does not improve significantly the results obtained. But, regarding the CO<sub>2</sub> model, it can be seen that both the absolute average error and the standard deviation are decreased which improves the accuracy of the model and increases the trust that end-users can put in the forecast. Especially, it can be seen that the estimations with the multi-regression model inspired by physics stay most of the time in the realistic ranges which represents a great advantage in terms of trust.



(a) Air temperature of the multi-regression model inspired by physics



(b) CO<sub>2</sub> concentration of the multi-regression model inspired by physics

Figure III.11: Results obtained for the multi-regression models inspired by physics

Models	Data	Average Error	Standard Deviation
Standard multi-regression	T (°C)	0.29 0.95	0.38 0.58
	CO <sub>2</sub> (ppm)	46 59	88 80
Inspired by physics	T (°C)	0.29 0.94	0.38 0.59
	CO <sub>2</sub> (ppm)	41 42	84 58

**Table III.2:** Comparison of absolute average error and standard deviation for the reference model and the multi-regression inspired by physics during the validation

### 3.4 Limits

Multi-regression models have proved to deliver good performance for forecasting air temperature and CO<sub>2</sub> concentration in some specific cases. However, several limits need to be raised. The first one was described in section 3.2 of chapter III: when forecasting CO<sub>2</sub> concentrations, multi-regression models require a sufficiently rich training dataset to reach a satisfactory precision. It was also highlighted that for the mono-zone study case there is a strong correlation between the indoor air temperature and the temperature of the corridor. Indeed, keeping only the outdoor temperature and the corridor temperature is enough to well forecast the temperature of the office. This reveals issues regarding the possibility to develop end-users energy services since the impact of openings or internal gains is negligible. In the next sections, semi-physical models are implemented in order to assess whether introducing some physics can solve this problem and estimate properly the respective impacts of the actions and environmental inputs. Then, it is also important to further investigate the ability of the model to be generalized to other study cases (especially multi zone study cases) and configurations. For that purpose, the impact of the lack of contact sensors on the accuracy of the output is investigated.

Indeed, it would ease the implementation of the model on other study cases with less sensors and reduce the instrumentation needed. For that purpose, the information of contacts of the door and window used in the thermal model is no longer used. In fact, if temperature sensors and to some extent CO<sub>2</sub> sensors are currently spreading over the residential and tertiary sector it is less so for contact sensors. Although some can be found in security packages for detection of intrusion,

it is very unlikely to find any on indoor openings. Besides, they can raise issues of social acceptability because they are perceived as more intrusive. Then, it would be very interesting to be able to free the models from the corresponding data. On the other hand, the rest of the data currently used to forecast CO<sub>2</sub> concentrations comes from temperature and CO<sub>2</sub> sensors which cannot be avoided. Some other sensors are used especially for estimating the occupancy in the office but those are not common to all case studies since this logic cannot be applied for residential buildings.

This raises in turn the question of the adaptation of this mono-zone model to a multi-zone one. In fact, how will the model behave when the information of the adjacent room needs also to be estimated?

### 3.5 Conclusion

In this section, the possibility of modelling a mono-zone office thanks to multi-regression structure has been explored. Despite the great accuracy obtained in some cases, several issues have been raised. The main one being the structure of the multi-regression obtained and the weight of the corridor temperature in the model. Besides, the month of June has also highlighted some limits especially regarding the CO<sub>2</sub> concentration which reached unrealistic values. Indeed, multi-regression models require rich datasets to forecast the temperature with a sufficient accuracy. If multi-regressions fit the needs of some applications, they do not seem to suit the requirements of energy services for mono-zone study cases at least.

## 4 Meta-optimization



### 4.1 Principle of the meta-optimization

The approach chosen here was presented by Audrey Le Mounier in her PhD thesis [59]. The goal is to guarantee the physical validity of parameters and to guide the optimization in order to improve the results. It requires an *a priori* knowledge of the values of the parameters in order to initialize them. As the goal is to free the system of any expert knowledge, parameters are initialized with some vague values which are only of the right order of magnitude. Then, low and high bounds are defined very broadly to simultaneously give more freedom to the algorithm and limit the impact of the choice of initial values. Giving freedom to the algorithm is done

with the aim of finding the more suitable and accurate result. But, knowing that the goal is to guarantee the physical validity of the parameters and that parameters are restricted by their limited values in terms of physical acceptability, a parameter reaching its bounds can difficultly be justified on a physical level. Thus, the optimization is guided towards zones where the objective function is not monotonous in the directions of the parameters. The procedure consists in the implementation of a series of successive optimizations and in the selection of set of parameters minimizing the errors without letting them reach their bounds. The optimization used in this process is a non-linear optimization called Sequential Quadratic Programming (SQP) [57] which is an iterative method. This algorithm requires that both the objective function and the constraints can be differentiated twice. Then, it divides the problem into sub-problems and optimizes the quadratic model of the objective function subject to the linearisation of the constraints. To explain more precisely the whole process of this method, two terms need first to be defined: adjustable and current value.

**Adjustable.** A parameter is said adjustable in the neighbourhood of a given parameter value if the error criterion admits a minimum for a variation of the parameter within its definition range starting from a given parameter value set.

**Current value.** The current value of a parameter is the last value estimated by the optimization process which did not reach the bounds of the parameter domain. In case the parameter is out of the acceptable domain and never was determined as adjustable, it is reset to its initial value defined a priori. In case the parameter is out of the acceptable domain and never was determined as adjustable, it is estimated to the value reached at the last step when it was adjustable.

At each step, the following are performed:

- An analysis to determine if the parameter is adjustable or not in the neighbourhood of the current value of the parameters.
- An optimization of the adjustable parameters

A summary is presented in figure [III.12](#).

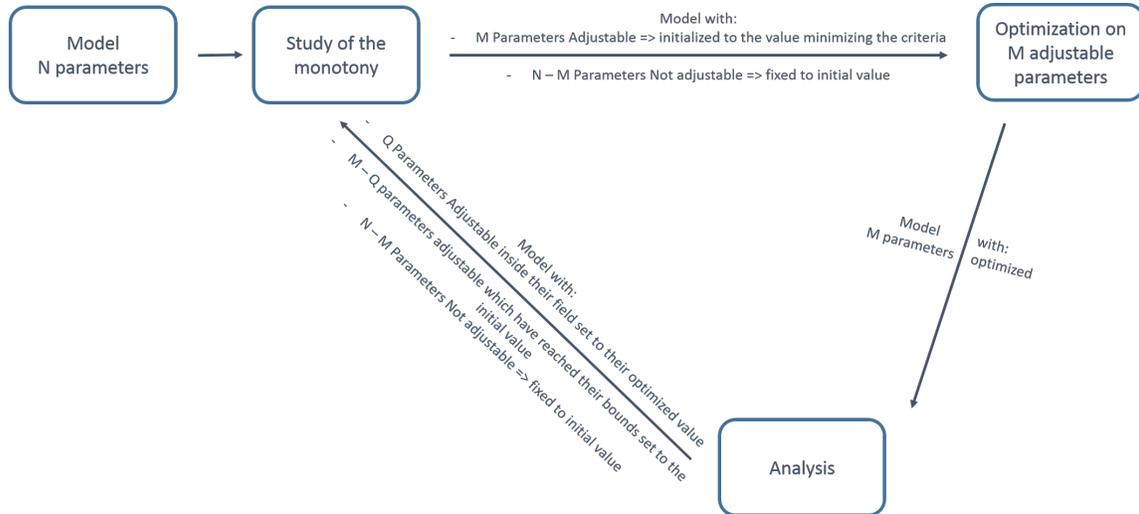


Figure III.12: Diagram of the meta-optimization principle

#### 4.1.0. i) Study of the monotony of the error criteria

In order to determine whether a parameter is adjustable or not, an estimation of the error criterion is performed by varying the concerned parameter in its definition range, all other parameters remaining at their current value. The monotony provides information on the capacity of the parameter to be identified. Indeed, if the error criterion is strictly monotonous, then the value of the parameter minimizing the error criteria will slide outside of the physically acceptable domain. Because the monotony study is implemented while all the other parameters are frozen to their current value, the adjustable characteristic of a parameter can change according to the considered current parameter values.

#### 4.1.0. ii) Optimization

For the optimization phase, any algorithm can be used. The one used in this study is a truncated Newton algorithm with is a conjugate gradient method. It minimizes a scalar function of one or several variables and is based on the computation of the gradient and the Hessian.

## 4.2 Application to the different structures

To apply this method to the case study presented in chapter II section 1, it is necessary to define for each structure initial values to the set of parameters to be

identified but also a research interval. In this case, some of the building physics was accessible. Then, it allowed to define the physical variables of the reference model according to the insulation, surface and type of walls. For the other parameters which required a more detailed knowledge, it was set to the right order of magnitude. This information is summarized in table III.3. In general building behaviours are significantly different in summer and winter as illustrated in the French translation of the European directive 2002 / 91 / CE on energy efficiency of existing buildings [1]. In this technical report the goal is to make appear the behaviour of buildings according to the seasons. To do so, two different periods were selected to validate the models: from May to July and from October to December. Datasets consist in data collected during occupation of the building, so they include the impact of users on their dwelling: opening of windows and doors or heating set-points.

Parameters	Minimum value	Initial value	Maximum value
$Q_{0out}$ ( $m^3.s^{-1}$ )	$2,77.10^{-5}$	0,010779	0,00277
$Q_W$ ( $m^3.s^{-1}$ )	0,00556	0,041031	1,38889
$Q_D$ ( $m^3.s^{-1}$ )	0,00556	0,040493	1,38889
$Q_{0cor}$ ( $m^3.s^{-1}$ )	$2,77.10^{-5}$	0,006958	0,00277
$R_{out}$ ( $K.W^{-1}$ )	$10^{-4}$	0,246071	1
$R_n$ ( $K.W^{-1}$ )	$10^{-3}$	0,010291	$10^{-1}$
$S_{CO_2}$ ( $ppm.m^3.s^{-1}.per^{-1}$ )	3	4,117725	12
Solar factor (-)	0	0,154460	1
$\phi_{constant}$ (W)	-10000	0	10000
$K_{heat}$ (-)	0	25	10000
$R_i$ ( $K.W^{-1}$ )	$10^{-4}$	0,005284	$10^{-1}$
$C_i$ ( $J.K^{-1}$ )	3600	720000	86400000
$C_a$ ( $J.K^{-1}$ )	3600	967,04	86400000
$C_w$ ( $J.K^{-1}$ )	483,52	967,04	4835200
$R_{w1}$ ( $K.W^{-1}$ )	$10^{-4}$	0,001	1
$R_{w2}$ ( $K.W^{-1}$ )	$10^{-4}$	0,001	1
$\zeta_D$ (-)	0	0,001	1
$\zeta_W$ (-)	0	0,001	1

**Table III.3:** Initial values and range of parameters to be estimated

### 4.2.1 The reference model

The training and validation datasets used are the same as those used in section 3 of chapter III. End-users energy services requiring a 24-hour time horizon, models variables can be daily set to the measured value in order to avoid the drifts in the model. This is possible as long as the model remains observable. The temperature and CO<sub>2</sub> concentration responses of the reference model are represented on figure III.13. It illustrates the first phase of validation of the model which consists in the simulation of the model on a dataset not used for estimating the parameters. The separation between the training phase and the cross validation one is materialized by a vertical black bar on the graphs. The graphs enable in a first time to perform a qualitative evaluation of the results of the model. It can be observed also a missing data period for the CO<sub>2</sub> concentration in the month of November due to a failure of the corresponding sensor. Then, for a quantitative analysis of the estimation results, the standardized RMSE is computed (*cf.* tables III.4 and III.5), the numerical values confirm the relevance and accuracy of the model and also highlight that the model is more accurate in winter than in summer.

	<b>Estimation</b>	<b>Validation</b>
<b>sRMSE</b>	0.06931	0.08726

**Table III.4:** Errors in summer

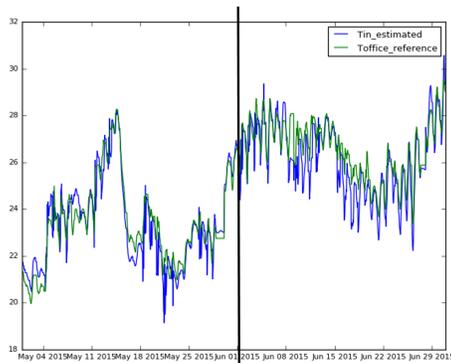
	<b>Estimation</b>	<b>Validation</b>
<b>sRMSE</b>	0.04279	0.06099

**Table III.5:** Errors in winter

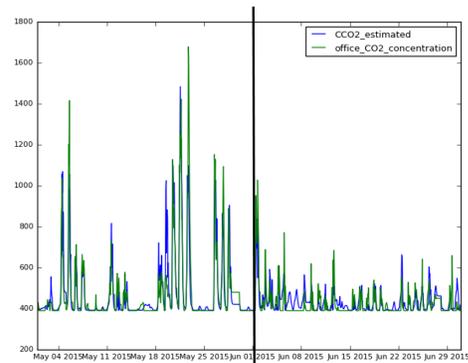
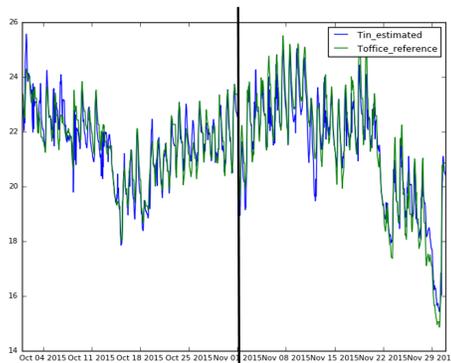
This improvement in the performances of the model in winter can be explained by the strong diminution of the doors and windows openings during that period (*cf.* figure III.14).

### 4.2.2 Modification of the inertia

Displaying all the temperature profiles of the models for the same periods underlines, important differences between these models (*cf.* figure III.15). This significant



(a) Summer Temperature

(b) CO<sub>2</sub> in summer

(c) Winter Temperature

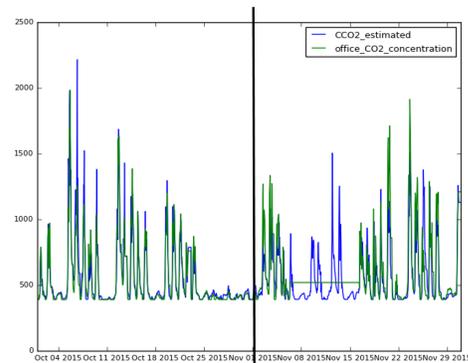
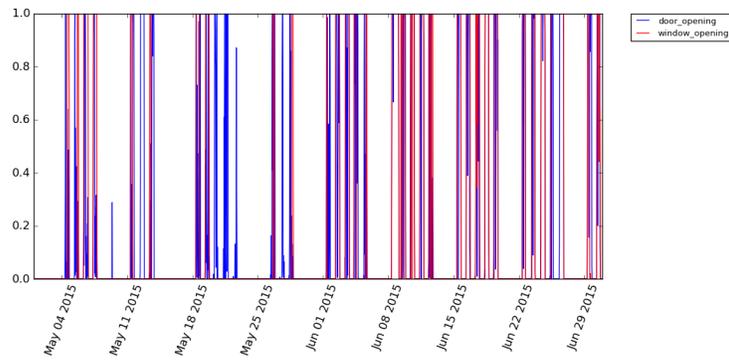
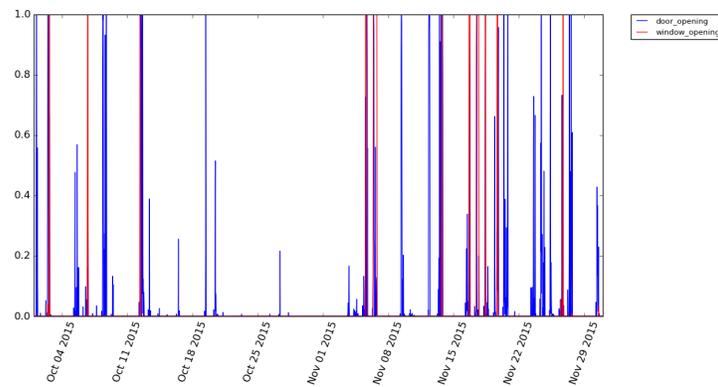
(d) CO<sub>2</sub> in winter

Figure III.13: Results of the reference model

gap cannot be found in figures III.16a and III.16b, which display all the values of sRMSE of the different structures in both the validation and learning phases. It confirms the need of different steps in the validation process. Observing more precisely the sRMSE graphs, what should first draw the attention is the high value of sRMSE for the "Model2CwallT<sub>out</sub>" in Summer. Indeed, this way of computing the error penalized significantly the outliers in temperatures. Observing the temperature profile obtained for that model (*cf.* figure III.17), it can be seen that the fluctuations are very important and furthermore present a high offset. For the winter behaviour, it can be noticed that all the models with two capacities present higher values for sRMSE than the reference model which is not consistent with the theory. On the other hand, except for the "Model2CwallT<sub>n</sub>", there is a big gap between the results in estimation and validation, apparently ruling out these models for prediction purposes.

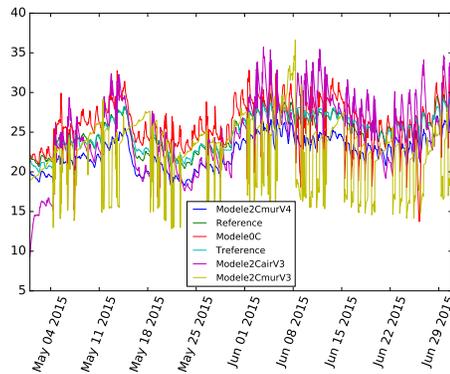


(a) Summer

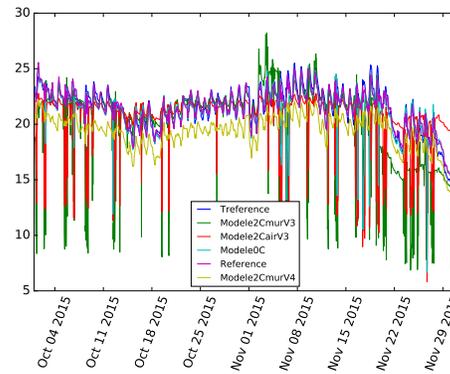


(b) Winter

Figure III.14: Sensor measurements of door and window openings



(a) Summer



(b) Winter

Figure III.15: Temperature profiles obtained with the different models

To understand what happens for the models with two capacitors, the optimization process was studied in respect of how the parameters evolve along with the iterations

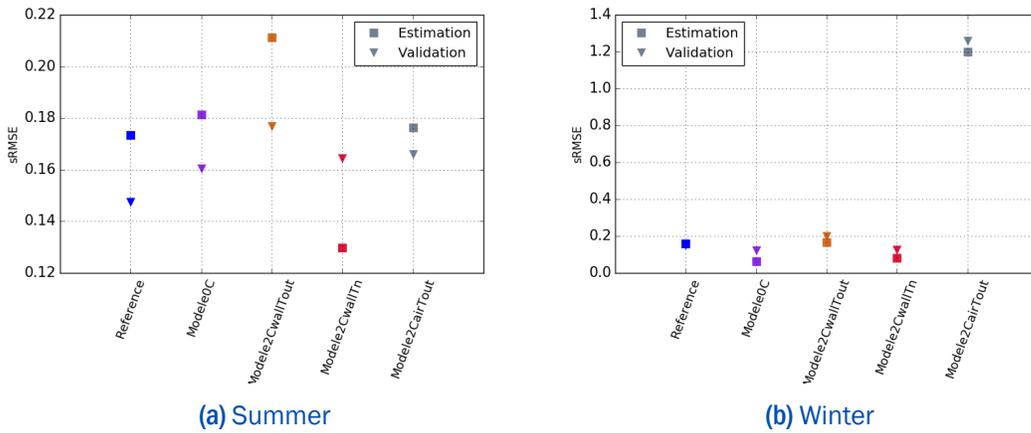


Figure III.16: Results of the different models in estimation and validation phase

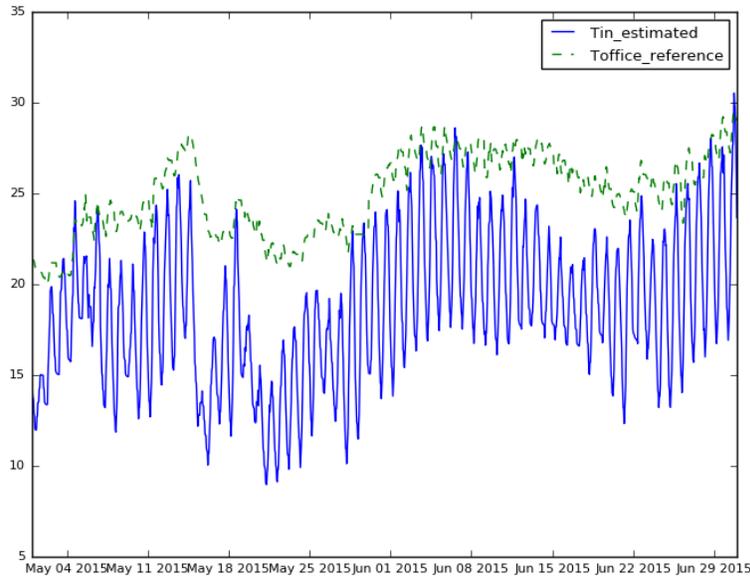


Figure III.17: Estimated and measured temperatures for the model "Model2CwallT<sub>out</sub>"

(cf. figure III.18). It can be observed that the optimization seems to find it difficult to converge: indeed for some parameters such as  $R_{out}$  the values oscillate several times between the upper and lower bound. This phenomenon is symptomatic of a poorly stated problem: it fails to converge during the optimization because it remains stuck in a zone where the objective function does not admit an obvious minimum. On the other hand, it can be seen in figure III.17 that dynamics are not well represented.

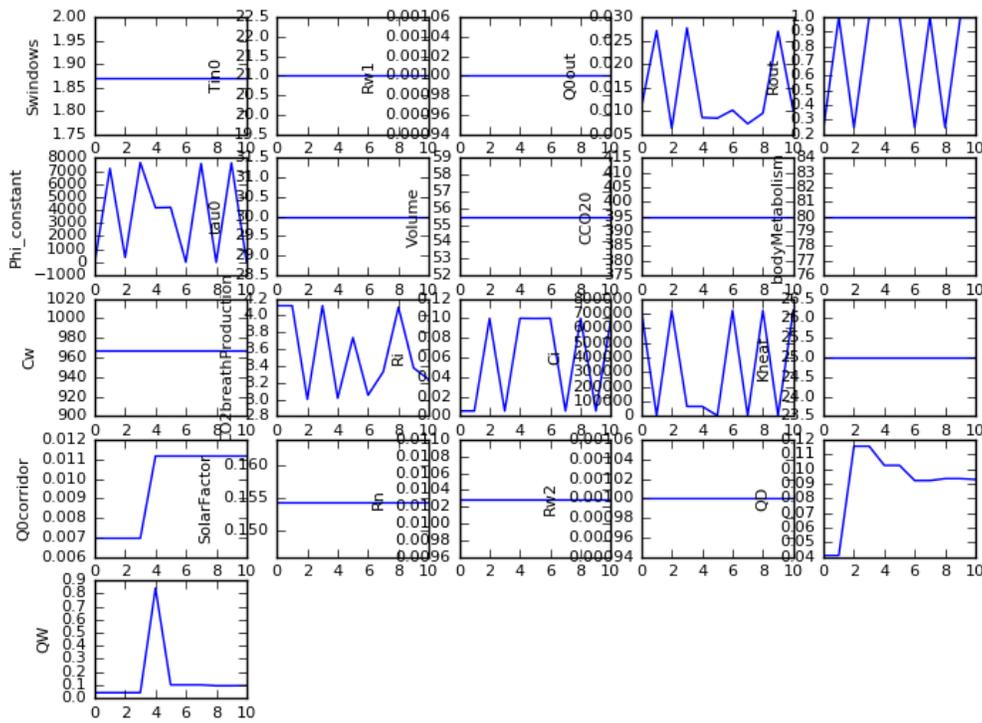


Figure III.18: Values of parameters to estimate along the iterations for the model `Model2CwallTout`

These observations raise real issues stressing the limits of this method. Indeed, it seems that the meta-optimization does not succeed in guiding the optimization towards the global optimum and thus the algorithm remains stuck in some local optimum. This is why, the validation procedure was not continued and first limits of the optimisation process were investigated (*cf.* section 4.3 of chapter III).

### 4.3 Limits

Since the results obtained for models with two-capacitors were not those expected: a better accuracy despite a higher computational time, several analyses were performed in order to explore the limits of the procedure presented in that chapter. The objective was to investigate the ability of the parameter estimation method to return parameter values quickly and with an acceptable accuracy.

### 4.3.1 Time of computation

In section 2 of chapter II, structures with up to 4 capacitors were introduced. However due to both the rapid deterioration of the results and the exponential increase of computation time, the meta-optimization described in this chapter was applied only to the models up to 2 capacitors. Indeed, where the computation time for the reference model was around 300 seconds, it rose to 7240 seconds for the Model2CwallT<sub>n</sub>. This exponential increase for a mono-zone model combined with the fact that the convergence of the algorithm is questionable was not in favour of proceeding with the development of this method for more complex models. Moreover, digging into the process of the optimization algorithm (algorithm SQP from Python library scipy) revealed that for all the models with two capacitors, the algorithm returned the error “Unable to progress” for some optimizations. This problem of convergence is partly responsible of the increase of computational time.

### 4.3.2 Training period

In order to decrease the computational time needed, some tests were conducted to quantify the impact of a shorter learning period on the accuracy of the results. Besides, the length of the training period also impacts the attractiveness of the solution. Indeed the longer it is, the harder the implementation in a real context is in terms of acceptability of the solution. Thus, it would be interesting to determine the minimum length of training period bringing the required accuracy. Results are presented in figure III.20. It can be seen that, as expected, dividing the length of the training period by 2 damages the accuracy for all the models except the reference model for which the impact is negligible. For Model0C, “Model2CwallT<sub>out</sub> and “Model2CwallT<sub>n</sub> the error rises by around 20% which remains significant. Then, it can be seen that for Model2Cair, the error almost doubles to reach 0.32. Considering the gain in computational time, it was not really significant and depended a lot on the convergence or not of the algorithm as described in the previous section.

### 4.3.3 Sensitivity to the initial point

As the convergence problem seems to appear only occasionally, a further analysis is led to investigate the sensitivity to the initial point. A high Sensitivity to the initial point jeopardizes the user’s trust in the e-consultant and would lead to the opposite of the expected result: a disengagement of the user regarding its dwelling energy

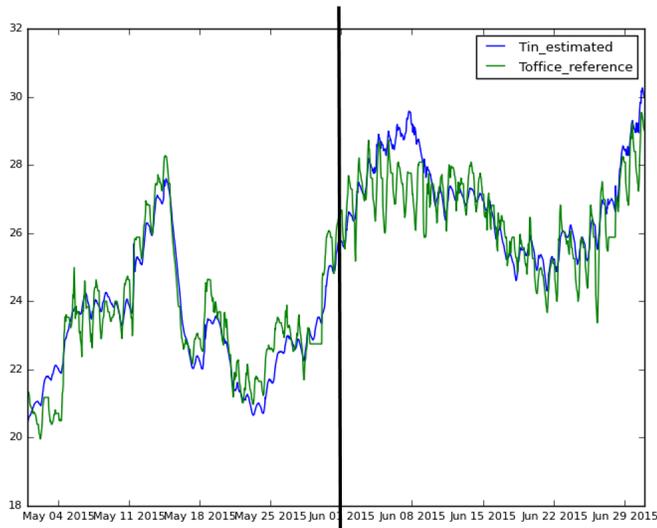


Figure III.19: Estimated and measured temperatures for the model "ReferenceNoζ"

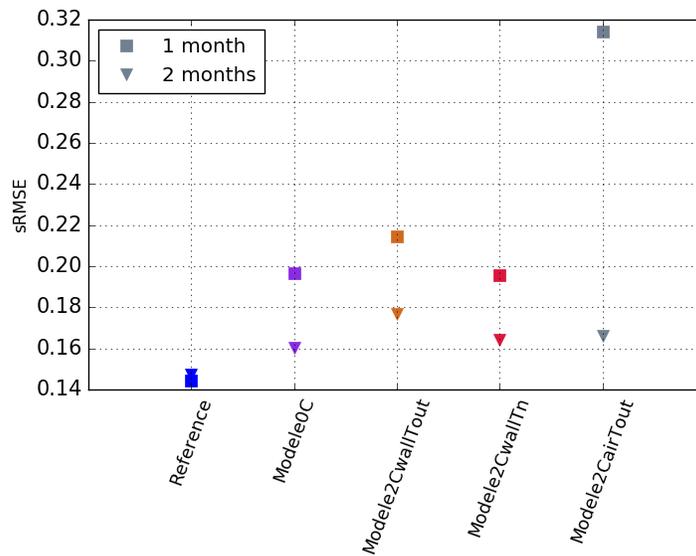


Figure III.20: Impact of the duration of the training data

management. The idea here is to evaluate if a good initialization could solve the limits developed above. Indeed, passing from the reference model to two-capacitor models means taking into consideration the inertia of walls or the ambient air. Due to the lack of available data for the study case (insulation, materials...), it was difficult to define relevant initial points for  $C_{wout}$ ,  $C_{wn}$ ,  $C_{air}$  or the new resistances. These parameters have therefore been initialized with vague values on the right order of magnitude. The stated hypothesis is that this optimization method requires accurate

initial values for the parameters. To investigate this hypothesis, an ergodicity test has been launched: 406 simulations were performed each of them initialized with a different dataset. The results can be seen in figure III.21. It can be noticed that according to the initial point, the absolute average error can vary strongly, with values up to 20°C which is obviously widely off the acceptable limit for a energy management model. It seems that as soon as parameters are added with vague initial values, the algorithm fails to converge. It is then impossible to guarantee the end-users any model performances.

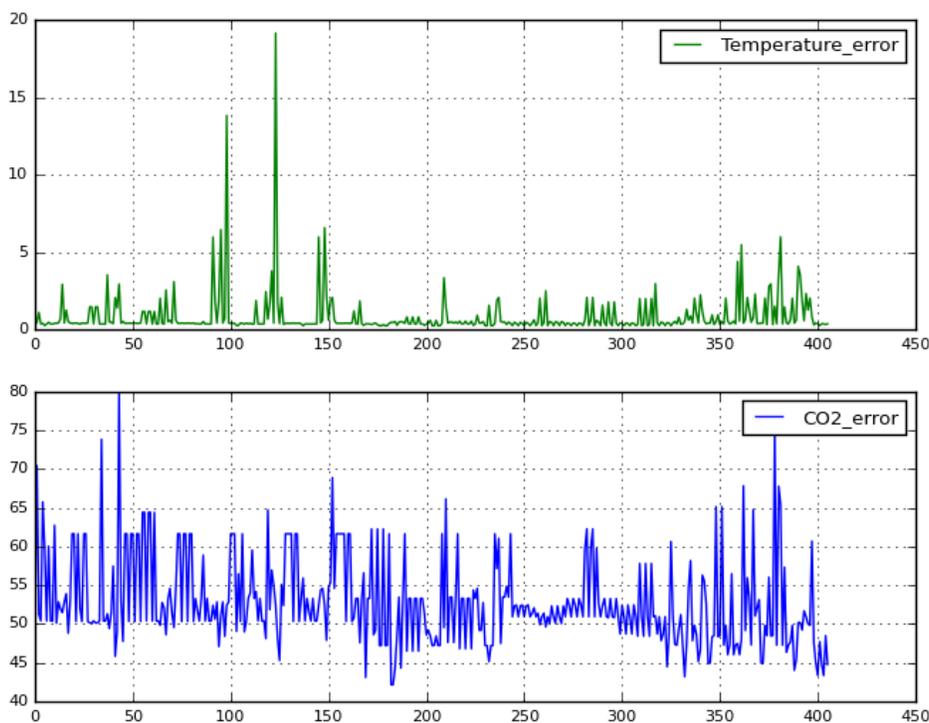


Figure III.21: Absolute average error for 406 simulations

## 4.4 Conclusion

This section presented a study of a gradient-based deterministic optimisation algorithm guided by an exploratory process. This methodology has been applied to the reference model and to the models up to 2 capacitors. At this point, some issues have been raised since the accuracy of the models obtained was damaged compared to the reference model. Then, different investigations were led to understand the

origins of this result. It has been demonstrated that the procedure is highly sensitive to the initial parameter values with an error that can reach 20°C. Besides, some convergence issues arose for the more complex models leading also to an exponential increase of the computational time. All those elements illustrate that the procedure implemented is not sufficient to guide the optimization algorithm towards the global optimum. Consequently, this method does not suit the implementation of energy services. Another deterministic algorithm than SQP might solve this issue but as the parameter estimation methods should be adapted to more complex study cases and to more complex models, the risks of non-convergence seem too high. Another method of parameter estimation based on a stochastic algorithm will then be investigated in order to allow a comparison between models and a choice of a suitable thermal structure.

All these results highlight that this estimation method is not suitable for end-users energy services. Indeed, occupants will not be able to build a trust in a service which cannot guarantee them any performances. Besides, the model structure implemented and for which the convergence is ensured is not able to properly model the slow and quick dynamics as required by the energy services.

## 5 Meta-heuristic



To remedy the problems of convergence and sensitivity to the initial point and better satisfy the technical specificities implied by the energy services described in section 2 of chapter III, it has been considered to use genetic algorithms which enable to cope with non linear models with regard to parameters. On the other side, genetic algorithms are usually more costly in terms of computational time. Furthermore, the choice was made of a multi-objective algorithm in order to deal with the CO<sub>2</sub> and temperature objectives separately. The objective of the models remains the ability to forecast the temperature and CO<sub>2</sub> evolution with good accuracy. Energy management end-user services also require from the models to be able to be self-tuned but the physical meaning of the estimated parameters is not mandatory. Several algorithms exist in the literature, such as the VEGA algorithm (Vectorial Evaluation Genetic Algorithm) [81], the NPGA (Niche Pareto Genetic Algorithm) which resorted mainly to a selection based on Pareto domination [49]. Later, it was superseded by the algorithm NPGA2 which uses the degree of domination of an individual as a deciding factor for the tournament selection [30]. Also available the

algorithm NSGA-II (Non Dominated Sorting Genetic Algorithm) is an elitist genetic algorithm introduced by Deb et al. in the 2000s [25] and based on a classification of individuals in several levels. In the field of buildings and energy optimization for different purposes, are also used the algorithm MIGA (Multi-island Genetic Algorithm) by Huang et al. [50] or the algorithm NSGA-II by Mozer [70] or Ghisi and Tinker [41]. In this work, the NSGA-II algorithm was chosen for several reasons: the elitist approach implemented enables to accelerate the sorting process in comparison with the algorithm NSGA but also to preserve the diversity of populations by saving the best found solutions during the previous generations on one hand and on the other hand by a comparison operator based on the computation of the Crowding distance [83].

## 5.1 Principle

The basic principle of this algorithm involves four main steps: random creation of the initial population, selection of individuals, mutations and crossover operations and computation of the selection criteria.

Figure III.22 summarizes the NSGA-II mechanism for a given population size of 6 [72]. First, an initial population is randomly generated. Then, while the number of generations is inferior to the maximum set by the user (termination criterion of the algorithm iterations), each individual is assigned its non-domination rank and crowding distance based on the evaluation of the objective function. During the reproduction step (or crowded tournament selection), the solution with a lower rank value is allowed to win a tournament. If two solutions have the same rank, then the solution with the higher crowding distance is allowed to win in order to preserve the diversity of the population. After a complete generation, a maximum of 6 non-dominated solutions can be obtained. After each generation, both the parent ( $P_t$ ) and offspring ( $Q_t$ ) populations are mixed to form a combined population,  $R_t$ . The next step consists in sorting all the solutions according to their rank. Solutions with the same rank are grouped in the same front. In the example presented in figure III.22, 3 different fronts ( $F_1, F_2, F_3$ ) are obtained after the non-dominated sorting of  $R_t$  corresponding to rank values. From these fronts, the new population  $P_{t+1}$  is filled with six individuals by the best non-dominated solutions. The filling process then starts with the front  $F_1$ , if the number of solutions in this front is inferior to 6, it continues with the front  $F_2$ . If they are too many solutions in the front  $F_2$  to fit in the new population, then solutions of this front are passed through a niche-preserving

operator which sorts them according to their crowding distance. Then, the solutions with the higher crowding distance are preferred. If two solutions present the same crowding distance, the choice is made randomly. This new population  $P_{t+1}$  is now considered as the parent population  $P_t$ . The offspring generation  $Q_t$  is then created using genetic operators like crowded tournament selection, crossover and mutation in the next generation. This cycle is continued until the termination criterion is reached.

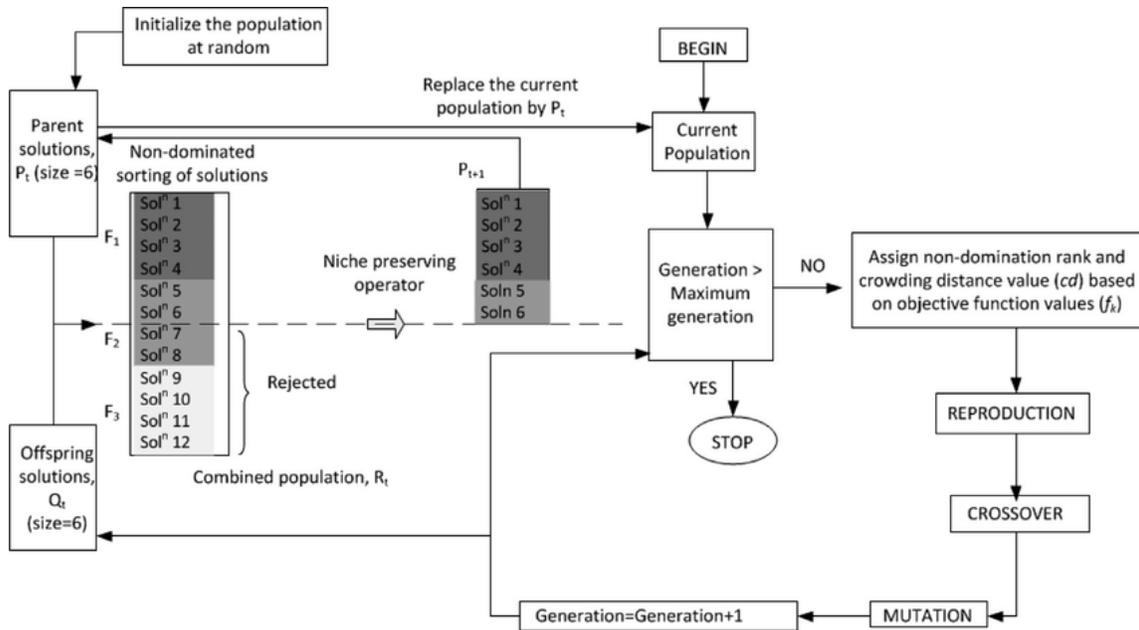


Figure III.22: Scheme of principle of the algorithm NSGA-II [72]

## 5.2 Configuration

The implementation of the algorithm requires a certain amount of manual configurations. Indeed, several parameters need to be set up: the number of generations, the number of individuals amongst every generation, the rate of mutation and cross over and so on.

### 5.2.1 Number of generations and individuals

The number of generations was first set to a very large number and then the evolution of the error criteria along the generations was plotted in order to see when the improvement was no longer significant. As it is a two-objective optimization, the error criteria could be based either on the temperature criteria, on the CO<sub>2</sub> criteria

or even on a recombined total objective. A two-objective optimization is considered because the models of temperature and CO<sub>2</sub> are linked by the air change rate which improves the air quality and can damage the thermal comfort. In this case, it was decided to focus on the temperature criteria as the CO<sub>2</sub> forecasts did not seem to vary a lot among the optimizations and mainly because actual occupants in buildings are more interested in a good forecast of temperatures than in CO<sub>2</sub>. As the objective is to implement energy services for end-users it is important to answer their needs. The first simulation was done with 200 individuals and 400 generations, the evolution of the error can be seen in figure III.23a. It seems that the error decreases a lot in the beginning and reaches its minimum after approximately 45 generations. As this operation required 64 minutes of computation, the number of individuals was divided by two, in order to see the impact on the error trend. For this second optimization, the required computational time was around 50 minutes. It can be seen that after 100 generations, the error has almost reached its minimum. Thus, it was decided to launch the optimizations for 100 individuals and 100 generations which required only 5 minutes for the Reference model.

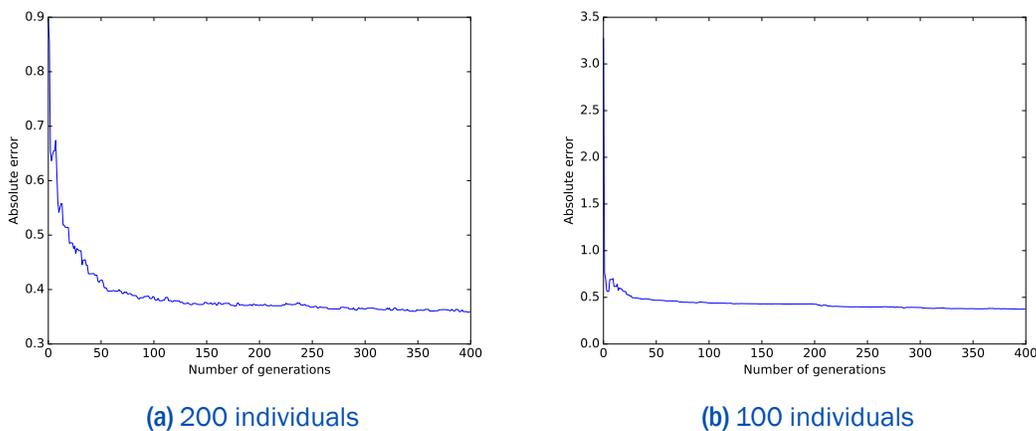
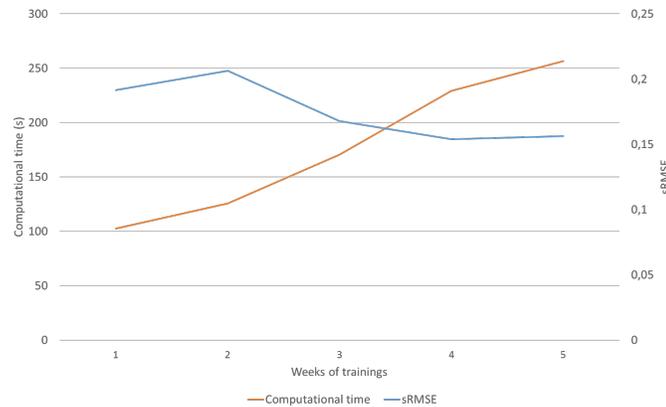


Figure III.23: Evolution of the temperature error along the generation

## 5.2.2 Impact of the length of the learning phase

As it is known that genetic algorithms are more costly than descent ones, it was tried to reduce the computational time as much as possible. One way to do it was to study the impact of the length of the learning phase on computational time and on errors in order to choose the best compromise. Figure III.24 shows the trends of the computational time and the sRMSE according to the number of weeks used in the training dataset. It can be seen that the computational time increases quickly

with the number of weeks used to learn the models but also that using more than 4 weeks of data does significantly impact the sRMSE. This is why a period of 4 weeks is used in the rest of the study.



**Figure III.24:** Impact of the length of the learning phase on the temperature error

### 5.3 Test of sensibility to the initial point

A high sensibility to the initial point jeopardizes the user's trust in the e-consultant and would lead to the opposite of the expected result: a disengagement of the user regarding the energy management of its dwelling. Indeed, if for some days, the convergence is not reached and the model returns a prediction with low accuracy, users will find it difficult to trust the prediction the following day. Since, the meta-optimization developed in section 4 of chapter III suffered from problems of robustness, i.e the algorithm did not reach the same optimum according to the chosen initial parameter values, the same test was performed for the genetic algorithm. In that case, the initial point is chosen randomly, so around 160 simulations were performed for the Reference model and the values of sRMSE were recovered. It can be seen in figure III.25 that all the values are in a narrow range. They are not all equal because a genetic algorithm is less accurate than a descent one but it seems that the algorithm is robust with regard to initial parameter values.

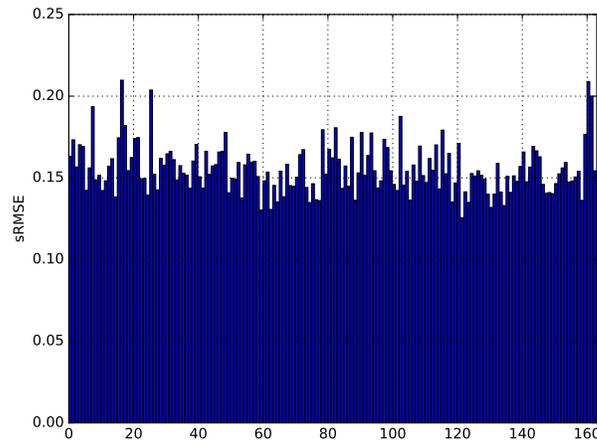


Figure III.25: Results of the robustness test

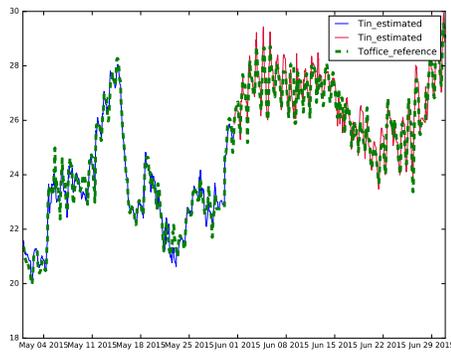
## 5.4 Application to the different model structures

### 5.4.1 The Reference model

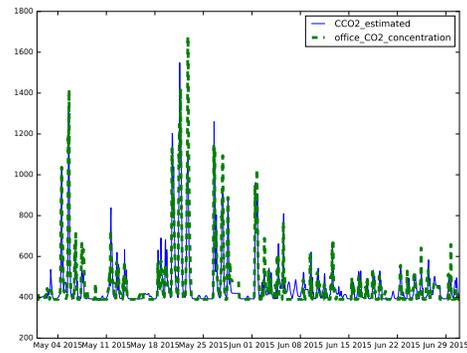
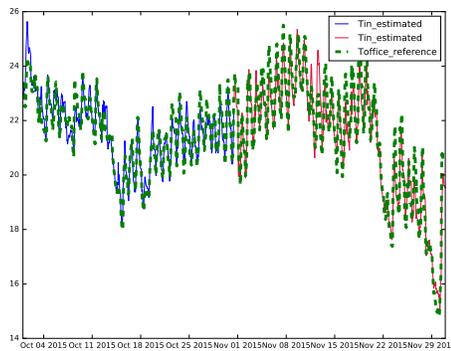
The training and validation datasets used with the genetic algorithm are the same than those used for multi-regression models in section 3.1.1 of chapter III. The temperature and CO<sub>2</sub> concentration responses of the reference model are represented in figure III.26. This illustrates the first phase of validation of the model which consists in the simulation of the model on a dataset not used for estimating the parameters. The separation between the training phase and the cross validation one is materialized by a change of colours in the graphs (blue : training / red : validation). The graphs allow firstly to perform a qualitative evaluation of the results. Then, for a quantitative analysis of the estimation results, the standardized RMSE is computed (*cf.* table III.6), the numerical values confirm the relevance and accuracy of the model.

	Training	Validation
<b>Summer</b>	0.13949	0.11774
<b>Winter</b>	0.14327	0.167634

Table III.6: sRMSE values



(a) Summer Temperature

(b) CO<sub>2</sub> in summer

(c) Winter Temperature

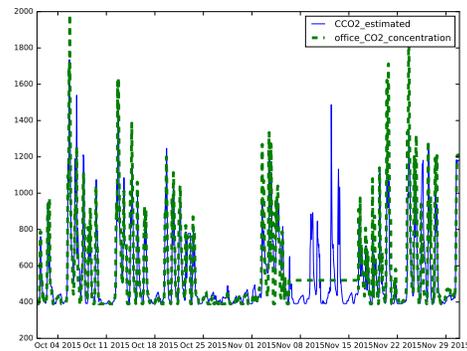
(d) CO<sub>2</sub> in winter

Figure III.26: Results of the reference model

### 5.4.2 Acceptability of the models

According to the results of the sRMSE presented in figure III.27, it can be seen that the “Model2Cair” and “Model0C” respect neither the sRMSE threshold nor the robustness along the seasons. Then, the “Reference” model exceeds the sRMSE threshold set to 0.1 and the “Model3CT<sub>out</sub>” is not robust: its performances are too different between the summer and winter scenarios.

The models considered as valid after that first step are: “Model3CT<sub>n</sub>”, “Model2CwallT<sub>out</sub>”, “Model3Cwalls” and “Model2CwallT<sub>n</sub>”. As a reminder, the descriptions of the model structures can be found in section 2 of chapter III. Figure III.28 shows their performances in terms of sRMSE for both winter and summer. In this graph, only the validation phase is considered as the estimation process does not require further investigation. Indeed, it has already proven to be ergodic and requiring low computational time.

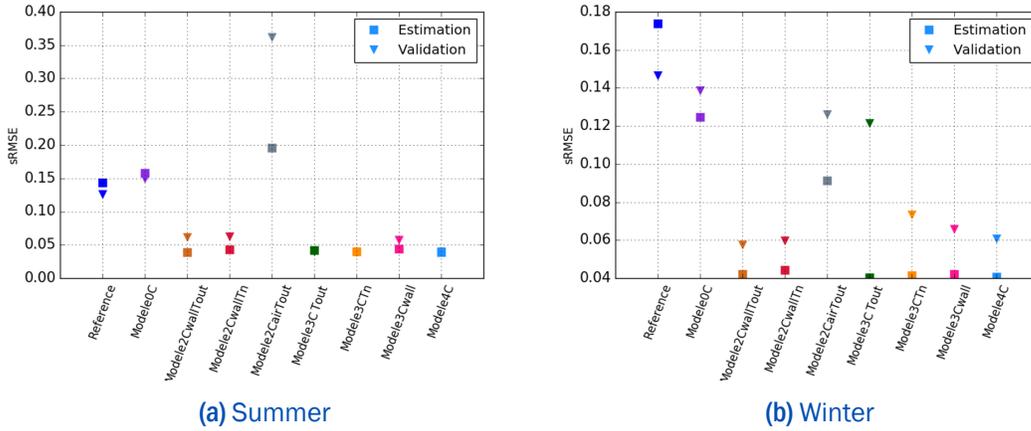


Figure III.27: Results of the different models in estimation and validation phase

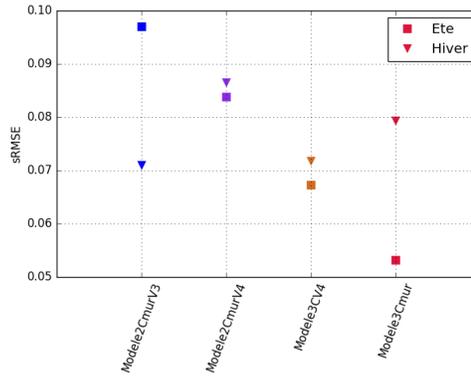


Figure III.28: sRMSE values of the valid models in validation phase

### 5.4.3 Selection of the most suitable structure

The next step consists in selecting the most suitable model between “Model3CT<sub>n</sub>”, “Model2CwallT<sub>out</sub>”, “Model3Cwalls”, “Model2CwallT<sub>n</sub>” and “Model4C” which will be done by looking at the sum of the sRMSE values obtained for the different seasons. The result can be seen in figure III.29. It appears directly that ”Model4C” is the best according to the criterion described above. From here, a sensitivity analysis is performed to ensure the necessity of each parameter involved.

In order to ensure that thermal circuits are suitable for end-user energy services, the same test applied to multi-regression models has been applied here. Two different scenarios have been compared in simulation with the estimated model for the month of April: the standard simulation with measured window openings and one considering that the window is opened all the time. Then, the difference of temper-

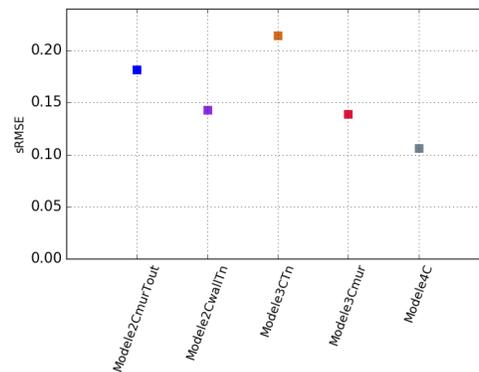


Figure III.29: Sum of the sRMSE values for both seasons

ature between these two scenario are compared (*cf.* figure III.30) and it appeared that this time modifying the window opening impact the indoor air temperature.

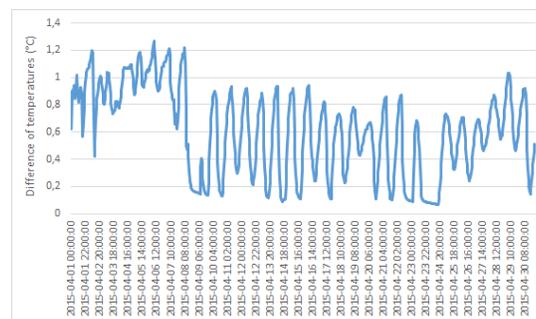


Figure III.30: Difference of indoor air temperature between the two scenario simulated

## 5.5 Sensitivity analysis

After the selection, several analyses have been performed on the selected model in order to ensure the relevance of the intrinsic model parameters. Firstly, sensitivity analyses are applied in order to evaluate the relevancy of every parameter and to seek some simplifications in the model structure. Later, in chapter IV, an analysis on the identifiability of the model is led to assess the ability or not of the parameters to support physical knowledge.

### 5.5.1 Morris

The Morris analysis shown in figure III.31 reveals that three main structural parameters of the model  $C_{w,out}$ ,  $R_{w,n2}$  and  $R_{w,out2}$  seem not to have a negligible

impact on the output. Yet, deleting the parameters  $C_{w,out}$  and  $R_{w,out2}$  would mean considering the structure with three capacities for which performances (according to the criteria of accuracy) are less interesting. Hence, this result needs to be further investigated, especially studying the estimated values of the parameters. However, contrary to the other methods, the Morris method is qualitative which means that it is possible to classify the different parameters but not to ensure their relevance. The goal of the sensitivity analysis in this research being to ensure that the structure of the model is consistent and that each parameter has an impact on the output, the Morris method is not suitable. Therefore, despite its computational cost, the Sobol method will be further investigated in order to confirm these results.

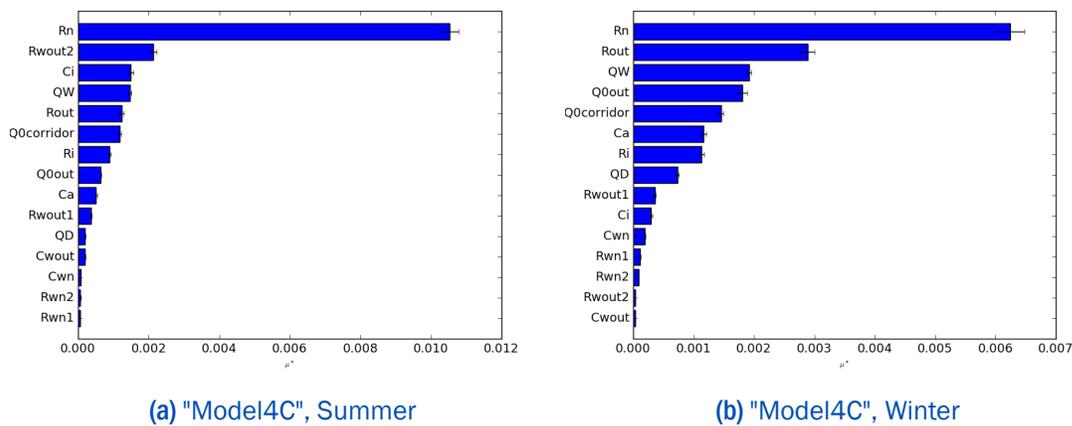


Figure III.31: Morris analysis

## 5.5.2 Sobol

As a reminder, the Sobol method is detailed in chapter II section 3.4.2. The sensitivity analysis is used to validate the relevance of the model structure. First, the evolution of the residuals according to the variability of these parameters is examined. For that purpose, wide intervals have been adopted for each parameter in accordance to what was done in the meta-optimisation approach (*cf.* table III.3). This step will show whether the model has the same behaviour in all these intervals and thus allow the definition of the variability intervals of each parameter required for the configuration of the analysis. Goffart [43] in her PhD thesis explained the importance of determining these variability intervals. Indeed, if for some values of one parameters in the wide interval initially defined, the model returns a high error, then this parameter will be estimated as having a major impact on the output. However, in this analysis, the convergence of the model has been previously ensured, the

question then is: ‘Has each parameter an impact on the output considering that the returned error of the model is acceptable?’. Considering parameter domains where the residuals are important would lead to overestimate the impact of parameters on the output as the indices are normalized.

### 5.5.2. i) *Sobol analysis over parameter domains leading to low residuals*

In consequence, the residuals have been plotted for each parameter varying in their wide initial interval, all others remaining still. In figure III.32, it can be seen that the different parameters do not have the same impact on the residuals. Indeed, the capacitances representing the dynamics of the inner and outer walls do not significantly impact the residuals contrary to the air capacitance. The most impacting parameters are  $R_n$  and  $R_{out}$ . It can be seen especially that  $R_n$  can only vary on a narrow range without greatly impacting the residuals. It validates and explains the results of the Morris method in which the impact of this parameter was leading. Then,  $R_n$  was kept out of the sensitivity analysis. These conclusions are model dependent and do not necessarily reflect the real behaviour since the models are simplified.

In figure III.33, it can be seen that  $R_i$  and  $R_{wn1}$  seem to have a minor impact on the output. However, deleting  $R_{wn1}$  would mean deleting also  $R_{wn2}$  and  $C_{wn}$  because they are intrinsically linked and this does impact the output. Besides, deleting  $R_i$  would lead to have 2 air capacitances the  $C_i$  and  $C_a$  and it can be seen that  $C_i$  Sobol index is low: under 0.05. It would be then interesting to consider a new structure corresponding to the Model4C without the branch with  $C_i$  and  $R_i$ . The corresponding model is represented in figure III.34.

However, this approach is valid only if the operating point chosen for this analysis is illustrative of the global operation of the system. Do the estimated parameter values always reach the same range of values or can they take significantly different values? To answer this question an analysis of the dispersion of parameters has been led over 50 parameter estimations.

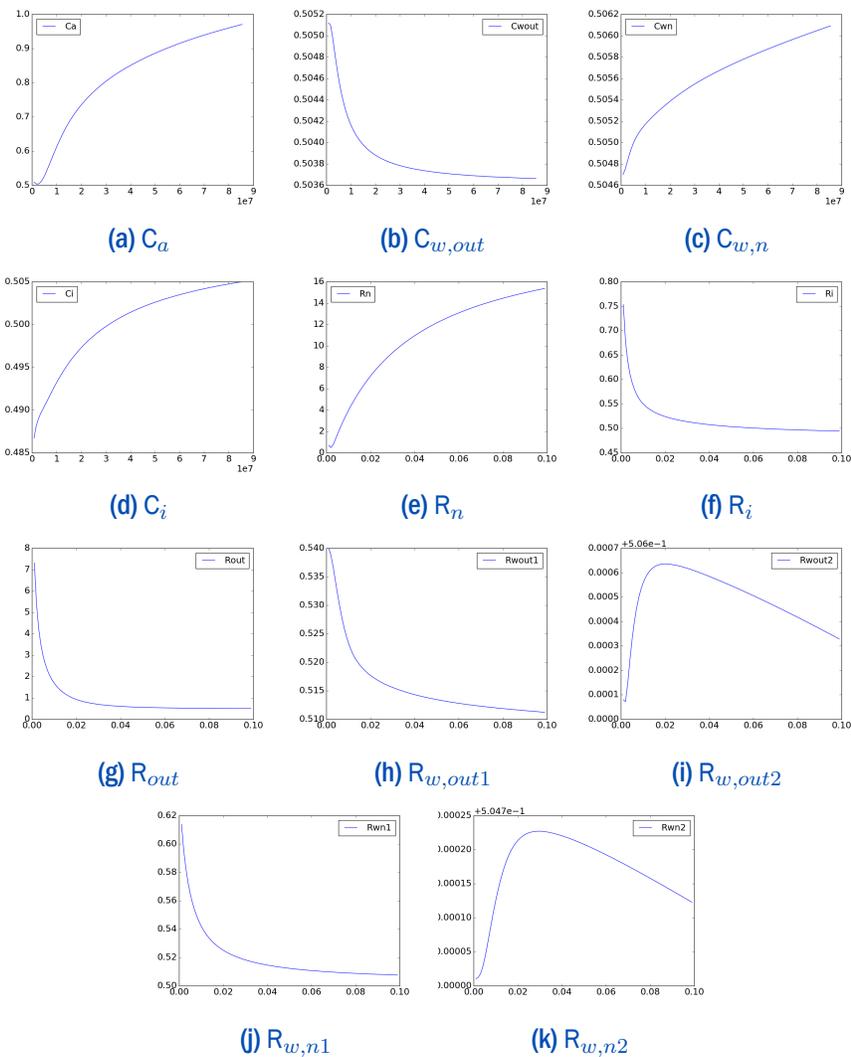


Figure III.32: Sum

Structural parameters

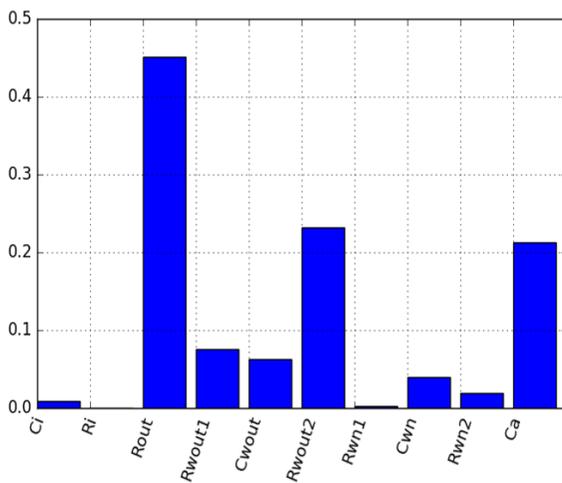


Figure III.33: Sobol index

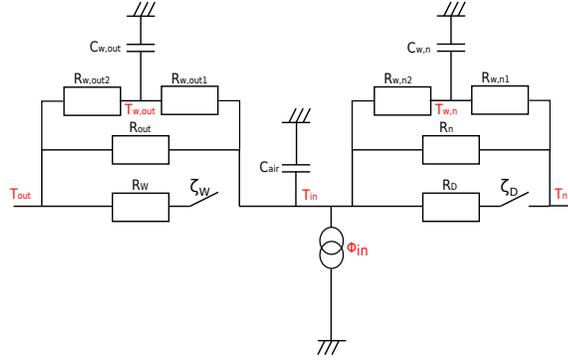


Figure III.34: Model3Cfinal

### 5.5.2. ii) Dispersion of parameters

In order to verify the results obtained with the Sobol sensitivity analysis, 50 parameter estimations have been performed and the dispersion of the parameter values obtained is displayed in figure III.35. It can be seen that for  $C_a$  or  $R_n$ , the values obtained among the 50 parameter estimations are concentrated around the same final value. For the other parameters, the estimated values obtained are quite distributed on the whole parameter field. However, since the scales of the different parameters are significantly different, this simple graphic observation is not enough to conclude on the dispersion of the parameters. This is why, for each parameter, the standard deviation is computed (*cf.* table III.7). It can be observed that except for  $C_a$ , the standard deviation is at least of 0.33 and can reach 85%. Then, it seems difficult to draw any physical meaning from the parameter values obtained but that was suspected already. Indeed, most of the time, dealing with simplified models reduces the meaningfulness of the considered parameters.

Different conclusions can be drawn from these results. The first thing that can be observed is that the more the parameters have an impact on the output, the lower the standard deviation is, which is consistent. The two analysis led then to the same conclusion.

However, the high dispersion of parameters informs that starting from a specific configuration to determine the intervals of Sobol introduces a significant bias in the results. Indeed, the set of parameters obtained after a parameter estimation could be very different from the one obtained previously and so the residual computed for every parameter could vary completely differently. To outreach this, a first Sobol analysis is performed with very wide ranges and then, for each parameter set evalu-

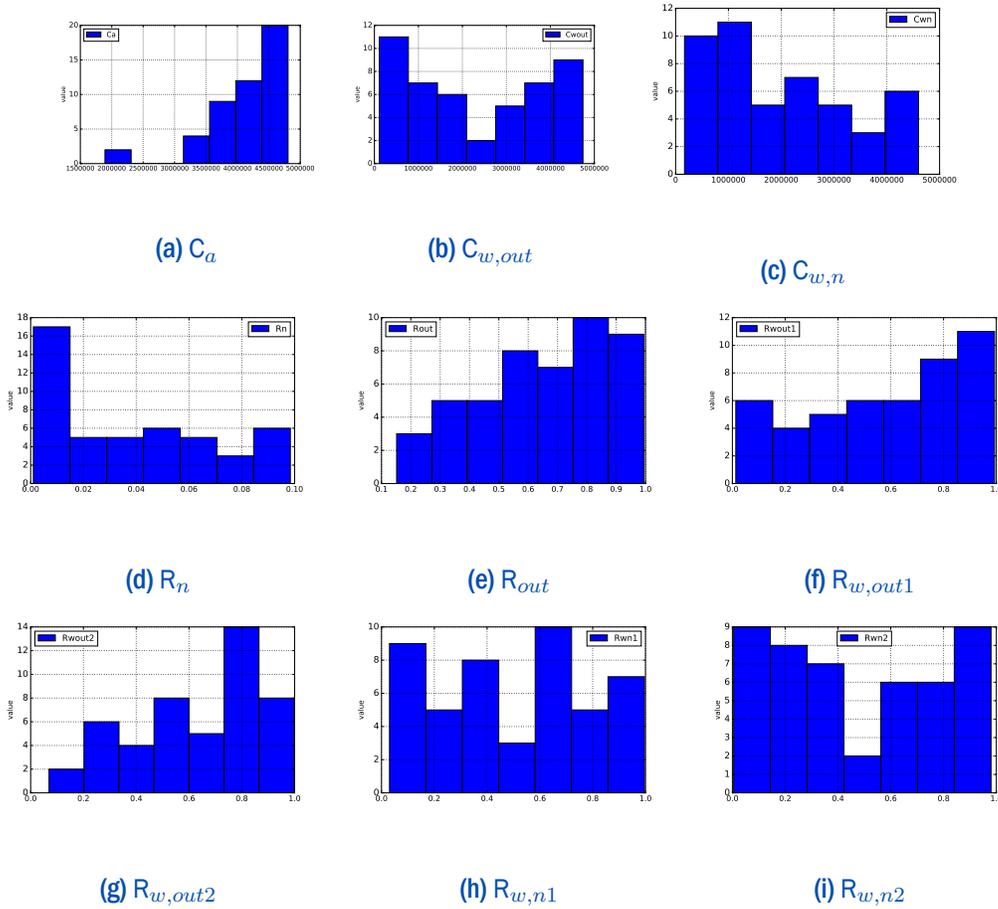


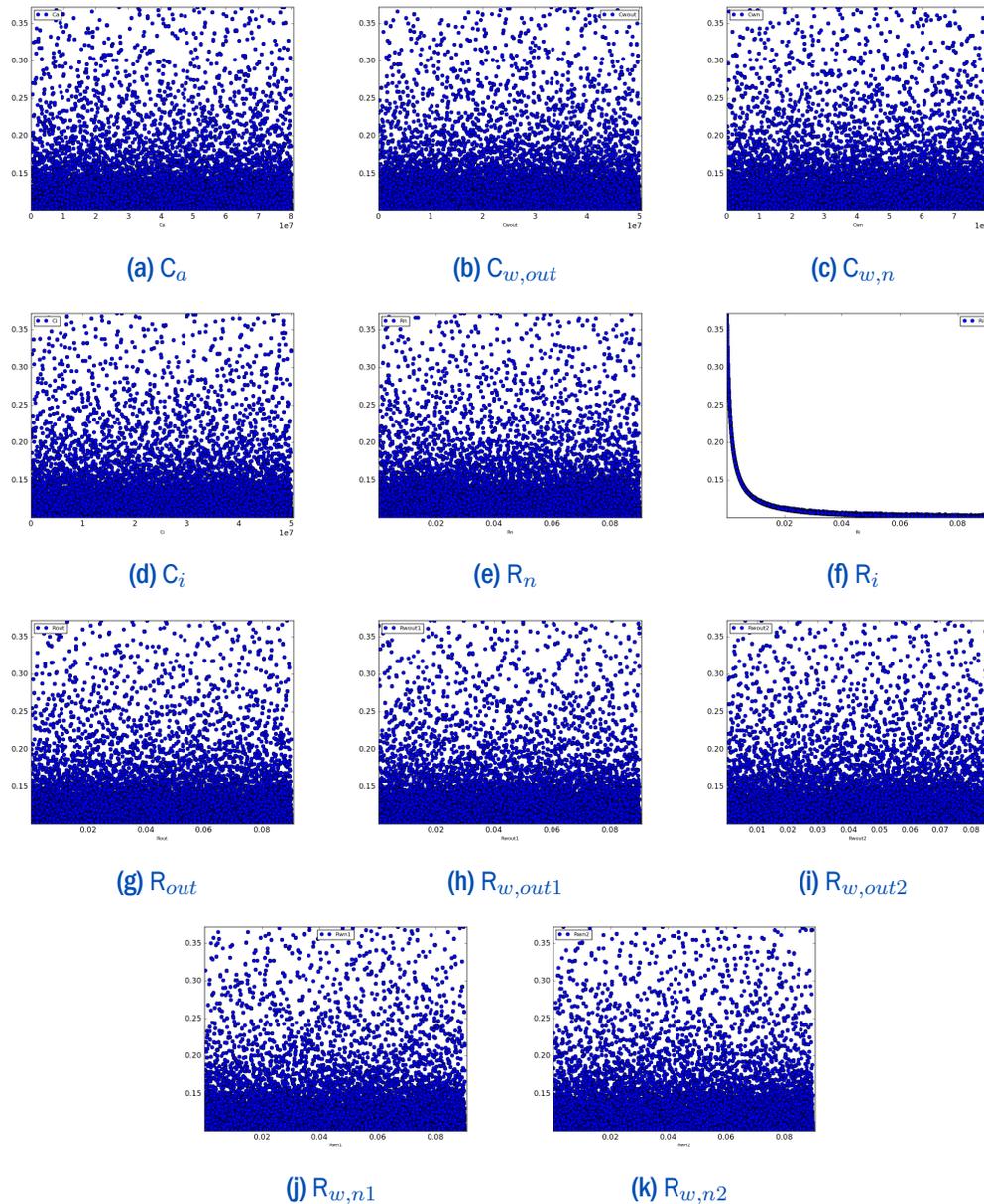
Figure III.35: Dispersion of the parameters over 50 estimations for the Model4C

Parameter	Standard deviation (%)
$C_a$	0.15
$C_{w,n}$	64.40
$C_{w,out}$	0.66
$R_n$	0.86
$R_{wout,1}$	0.53
$R_{wout,2}$	0.39
$R_{out}$	0.34
$R_{wn,1}$	0.59
$R_{wn,2}$	0.65

Table III.7: Standard deviation of model estimated parameters for the Model4C

ated via Sobol to plot the residuals of the model according to the parameter values. This will allow to have a more global vision of the model behaviour. Then, this will allow to determine the intervals to consider for the Sobol analysis. The Sobol anal-

ysis with wide ranges returns that only one parameter has an impact on the output:  $R_i$ . Then the 240 000 values of residuals are plot for each parameter as shown in figure III.36. These figures confirm that in this context only  $R_i$  has a clear impact on the residuals. Then, the Sobol's analysis can be launched again keeping the wide intervals for every parameter except  $R_i$  which will be kept between  $[0.02,1]$ .



**Figure III.36:** Sum of the residuals according to the values of the different structural parameters

The dispersion analysis has been launched for the “Reference” model in order to see whether the number of parameters has an impact on their dispersion (*cf.* figure III.37 and table III.8). It can be noted that standard deviations are all around 50%.

Then, even with fewer parameters, it is difficult to recover any physical knowledge from the estimated parameters.

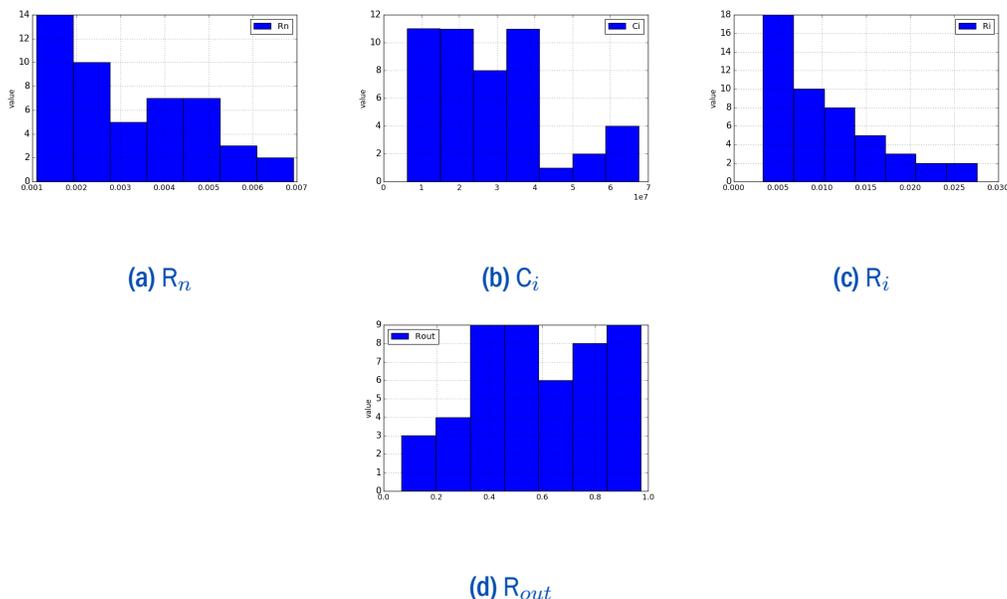


Figure III.37: Dispersion of the parameters over 50 estimations for the Reference model

Parameter	Standard deviation (%)
$C_i$	54.51
$R_n$	49.55
$R_i$	59.22
$R_{out}$	39.93

Table III.8: Standard deviation of model estimated parameters for the "Reference" model

### 5.5.2. iii) Discussion

The dispersion of parameters observed for the parameters of the Model4C highlights that the first analysis performed in order to define the range for Sobol sensitivity analysis is not sufficient. Starting from a specific configuration to determine the intervals of Sobol introduces a significant bias in the results. Indeed, the set of parameters obtained after a parameter estimation could be very different from the one obtained previously and so the residual computed for every parameter could vary completely differently. To outreach this, a first Sobol analysis is performed with very wide ranges and then, for each parameter set evaluated via Sobol to plot the residuals of the model according to the parameter values. This will allow to have a more

global vision of the model behaviour. Then, this will allow to determine the intervals to consider for the Sobol analysis. The Sobol analysis with wide ranges returns that only one parameter has an impact on the output:  $R_i$ . Then the 240 000 values of residuals are plot for each parameter as shown in figure III.38. These figures confirm that in this context only  $R_i$  has a clear impact on the residuals. Then, the Sobol's analysis can be launched again keeping the wide intervals for every parameter except  $R_i$  which will be kept between  $[0.05,0.1]$ .

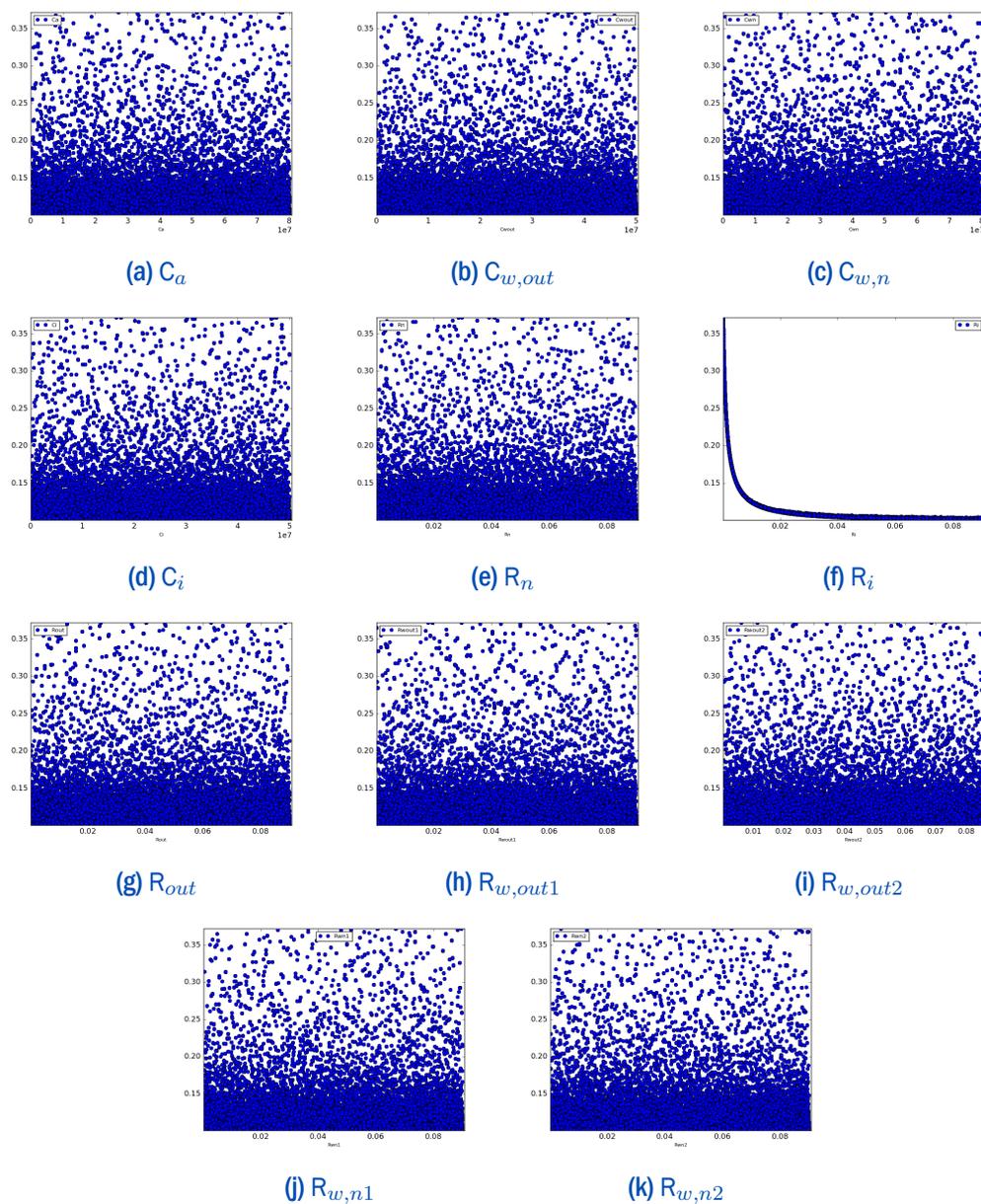


Figure III.38: Sum of the residuals according to the values of the different structural parameters

The results returned by the Sobol's analysis are displayed in figure III.39. The impact of the parameter  $R_i$  can be seen again. It can be noticed that only two other parameters impact the output and significantly less than  $R_i$ . However, considering only these parameters does not really make sense in this context. The hypothesis studied is then that the impact of  $R_i$  is too significant that it hides the impact of other parameters. Another sensibility analysis is thus launched excluding  $R_i$  from the varying parameters. The problem of this approach is that it is highly dependent on the value set for  $R_i$ . And indeed, this analysis returned only three significant parameters which differ from the previous significant ones.

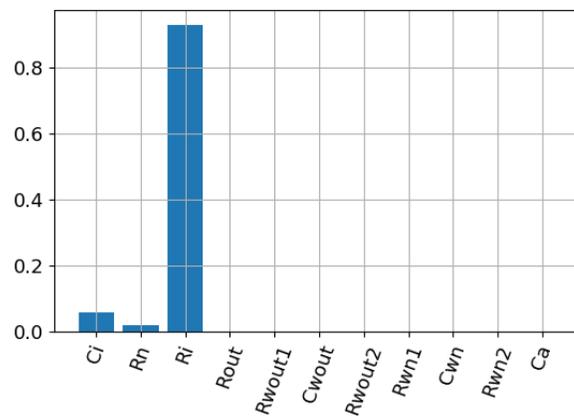


Figure III.39: Sobol index over the new parameter ranges

Finally, the analyses returning no meaningful results regarding the objective, the model structure stays unchanged.

## 5.6 Observability of the model

Considering the 24-hour horizon required by the energy services, an important issue is to initialize the model. Until now, models have been launched for periods of about 1 month, in which case initial values do not impact significantly the results. On a 24-hour simulation with a hourly step, initial values take on more importance. In order to allow this initialisation, models must be observable. Indeed, a linear system is considered completely observable if, given the control and the output over the interval  $t_0 < t < T$ , one can determine any initial state  $x(t_0)$ . It can be ensured through the rank of the observability matrix defined as:

$$O = \begin{bmatrix} C \\ CA \\ \dots \\ CA^{n-1} \end{bmatrix} \quad (\text{III.10})$$

where  $A$  and  $C$  are the matrix of the state space and  $n$  the number of state variables.

Computing the observability matrix for the Model4C has highlighted that two states are not observable. Further investigations have been led to quantify the impact of this non-observable state on the output and so on the initialization process. For this analysis, the state space has been transformed into the Kalman observability decomposition form [53]. It is a mathematical way to decompose a representation of any linear time-invariant (LTI) control system into a standard form which makes clear the observable and controllable components of the system. The Kalman observability decomposition form states that if the state space is non-observable, then there exist a similarity transformation such as  $\hat{A}$ ,  $\hat{B}$ ,  $\hat{C}$ ,  $\hat{D}$  defined as below:

$$\hat{A} = P^{-1}AP \quad (\text{III.11})$$

$$\hat{B} = P^{-1}B \quad (\text{III.12})$$

$$\hat{C} = CP \quad (\text{III.13})$$

$$\hat{D} = D \quad (\text{III.14})$$

where  $P$  is an invertible matrix. For real-values matrix state space,  $P$  can be chosen as an orthogonal matrix.

In this context, matrix  $\hat{A}$ ,  $\hat{B}$  and  $\hat{C}$  can be put under the following form:

$$\hat{A} = \begin{bmatrix} A_o & 0 \\ A_{\bar{o}o} & A_{\bar{o}} \end{bmatrix} \quad (\text{III.15})$$

$$\hat{B} = \begin{bmatrix} B_o \\ B_{\bar{o}} \end{bmatrix} \quad (\text{III.16})$$

$$\hat{C} = \begin{bmatrix} C_o & 0 \end{bmatrix} \quad (\text{III.17})$$

where  $A_o \in \mathbb{R}^{n_o \times n_o}$ ,  $C_o \in \mathbb{R}^{n_o \times n}$ ,  $n$  being the rank of the state variables vector and  $n_o$  the rank of the observability matrix. In this configuration, the pair  $(C_1, A_{11})$  is observable.

The next step consists in expressing the state-space in the new space:

$$\widehat{\dot{X}} = \widehat{A}\widehat{X} + \widehat{B}U \quad (\text{III.18})$$

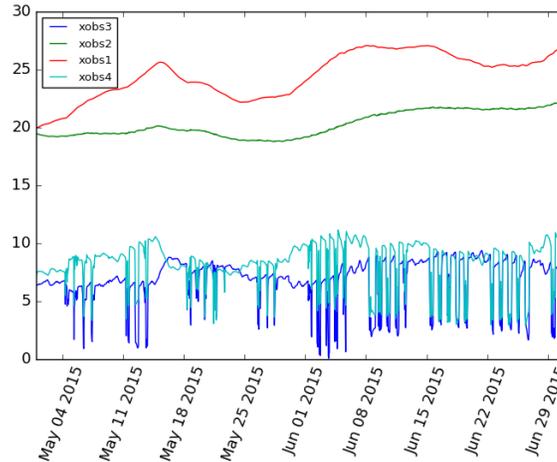
and the state vector can be written:

$$X = P\widehat{X} \quad (\text{III.19})$$

and decomposed in the new base according to:

$$X = \widehat{X}_1\widehat{P}_1 + \widehat{X}_2\widehat{P}_2 + \widehat{X}_3\widehat{P}_3 \quad (\text{III.20})$$

Then, matrix  $P$  being normalized, the magnitude of  $\widehat{X}_3$  and  $\widehat{X}_4$  directly illustrates the impact of the non observable states on the output. It can be observed that although they are less important than the two first components, they cannot be considered as negligible.



**Figure III.40:** Evolution of the different components of  $\widehat{X}$  over 60 days

To complete the analysis, a test of the observer has been performed with a simulation with wrong initial values and a 24 hour horizon. The 1st of October has been chosen, the indoor air temperature initialized at 40°C and the observer used for the daily calibration set at 6 AM. The temperature profiles obtained can be seen in

figure III.41. It can be observed that at 6 AM the estimated temperature is indeed equal to the measure which reflects the relevancy of the observer.

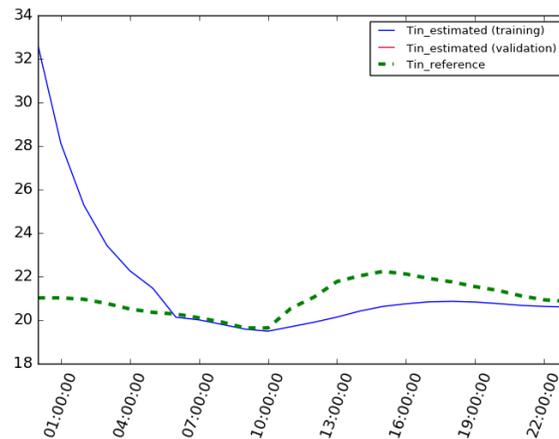


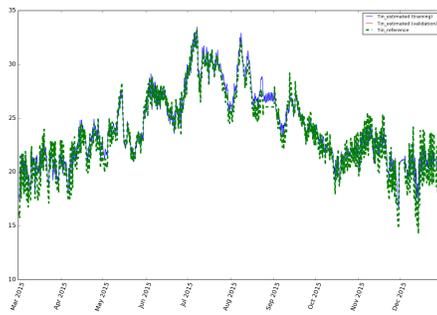
Figure III.41: Temperature profiles for the 1st of November

## 5.7 Reliability of the model along the seasons

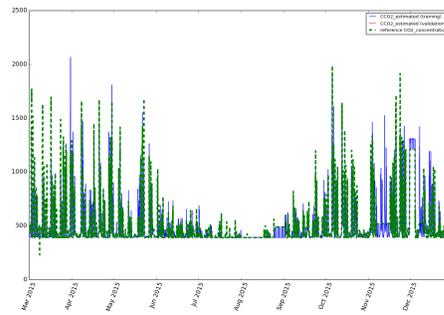
The reliability of the estimated model along the year is an important issue to investigate. Can a model estimated during the month of May predict with sufficient accuracy the thermal behaviour during the month of October? Is the model requires to be estimated several times along the year? How does the accuracy of the estimation evolve?

In figure III.42, yearly estimations of the Model3Cfinal learnt during the month of May and during the month of October respectively are displayed. It can be observed that the model succeeds in estimating the thermal dynamics in every scenario. It can be reminded that during the training month the heating was off, the window and door openings numerous. Nevertheless, the model successfully forecasts the CO<sub>2</sub> and temperature trends during winter. In turn, the model estimated during the heating period adapts successfully to a period with heating systems off and numerous openings. These results demonstrate how strong the semi-physical models are compared to the multi-regression which failed in this same exercise.

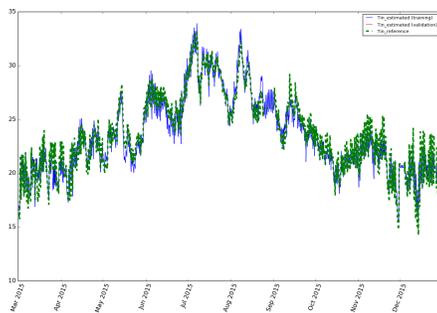
To confirm quantitatively these graphical results and to determine their accuracy, sRMSE values for the yearly validation are computed. Table III.9 shows that the errors obtained are surprisingly lower. This can be explained by the choice of the



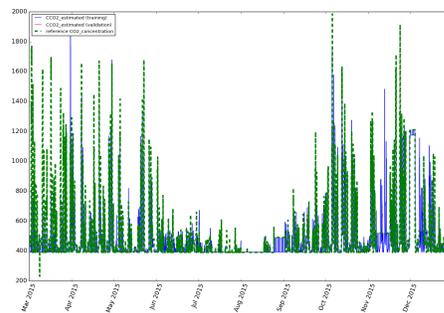
(a) Yearly temperature of the Model4C estimated in May



(b) Yearly CO<sub>2</sub> concentration of the Model4C estimated in May



(c) Yearly temperature of the Model4C estimated in October



(d) Yearly CO<sub>2</sub> concentration of the Model4C estimated in October

**Figure III.42:** Yearly forecasts of CO<sub>2</sub> and temperature of the multi-regression models estimated either in a month in winter or a month in summer

standardized version of the sRMSE since it is obtained from the RMSE divided by the higher gap in temperature. And obviously this gap is significantly higher over a year than a month. In order to bypass this difficulty, the sRMSE has been calculated for each month separately and the worst value has been kept to compare with previous results (*cf.* table III.10). It can be observed that in this case, values somewhat exceed the sRMSE limit defined in the acceptability process while remaining reasonable according to the objectives.

Training month	Data	sRMSE value
May	Temperature	0.0423
October	Temperature	0.0393

**Table III.9:** sRMSE values for the validation on the whole year for the Model4C estimated in either summer or winter

Training month	Data	sRMSE value
May	Temperature	0.1472
October	Temperature	0.1252

**Table III.10:** sRMSE values for the validation on the whole year for the Model4C estimated in either summer or winter

## 5.8 Physical interpretation

The objective of this research work has been to identify a structure model able to accurately forecast indoor CO<sub>2</sub> concentration and air temperature in different study cases. The physical meaning of the parameters is not required for providing the energy services described in section 2 of chapter I. In this section, this possibility is investigated in order to well identify the limits of this modelling and the hypothetical other applications. It is done in two ways: firstly the study of the structural identifiability of the model and then the study of the parameters dispersion over an important number of parameter estimations. In this context, the structural identifiability does not constitute a requirement but an information on the possibility of recovering physical knowledge from the parameters.

### 5.8.1 Identifiability

As a starting point, the identifiability of models is evaluated considering that window and door contacts are inputs of the model. It was the case for example in the study case of the mono-zone office. For the implementation of the Sedoglavic's algorithm, the formulation of Model4C is adapted according to the following equations:

$$\begin{aligned}
F := & \frac{-x_1}{C_i R_i} + \frac{x_4}{C_i R_i}, \\
& \frac{-1}{C_{wout}} \left( \frac{1}{R_{wout1}} + \frac{1}{R_{wout2}} \right) x_2 + \frac{x_4}{C_{wout} R_{wout1}} + \frac{u_2}{C_{wout} R_{wout2}}, \\
& \frac{-1}{C_{wn}} \left( \frac{1}{R_{wn1}} + \frac{1}{R_{wn2}} \right) x_3 + \frac{x_4}{C_{wn} R_{wn1}} + \frac{u_1}{C_{wn} R_{wn2}}, \\
& \frac{x_1}{C_a R_i} + \frac{x_2}{C_a R_{wout1}} + \frac{x_3}{C_a R_{wn1}} \\
& - \frac{1}{C_a} \left( \frac{1}{R_i} + \frac{1}{R_{wn1}} + \frac{1}{R_{wout1}} + \frac{u_4}{R_D} + \frac{u_5}{R_W} + \frac{1}{R_n} + \frac{1}{R_{out}} \right) x_4 \\
& + \frac{1}{C_a} \left( \frac{u_4}{R_D} + \frac{1}{R_n} \right) u_1 + \frac{1}{C_a} \left( \frac{u_5}{R_W} + \frac{1}{R_{out}} \right) u_2 + \frac{u_3}{C_a}
\end{aligned} \tag{III.21}$$

$$X := [x_1, x_2, x_3, x_4] \tag{III.22}$$

$$G := [x_4] \tag{III.23}$$

$$\Theta := [C_i, R_i, C_{wn}, R_{wn1}, R_{wn2}, C_{wout}, R_{wout1}, R_{wout2}, C_a, R_D, R_{out}, R_W, R_n] \tag{III.24}$$

$$U := [u_1, u_2, u_3, u_4, u_5] \tag{III.25}$$

where  $F$  represents the state variables derivatives,  $X$  the state variables,  $G$  the output,  $\Theta$  the estimated parameters and  $U$  the inputs.

The output returned by the Sedoglavic's algorithm indicates that the Model4C is locally structurally identifiable according to the definition of Sedoglavic, i.e. "the model is structurally identifiable almost everywhere".

The identifiability analysis is first performed on the model developed in section 2.3 of chapter III. Following the progress previously detailed of the models, firstly it has been decided to consider that door and window openings are unknown and therefore need to be estimated. If these openings are estimated by a unique value then the system cannot be identified. Indeed, the variables  $\zeta_D$ ,  $\zeta_W$ ,  $R_D$  and  $R_W$  appear in the equations only under the form  $\frac{\zeta_W}{R_W}$  and conversely  $\frac{\zeta_D}{R_D}$ , thus it becomes impossible to estimate these two parameters independently on a unique way. They are considered as variables for identifiability problem because they are part of the parameters to be estimated. From this structure, simplifications need to be performed in order to obtain an identifiable model. Obtaining a single value for  $\zeta_D$  (or  $\zeta_W$ ) is the same than estimating an equivalent resistance  $Rz_D$  which would gather the values of  $\zeta_D$

and  $R_D$ . Then, it will lead to two resistances in parallel :  $R_{z_D}$  and  $R_n$  which once again can be simplified in one new resistance  $R_{z_{Dn}}$  as explained in equation below:

$$\frac{1}{R_{z_{Dn}}} = \frac{1}{R_{z_D}} + \frac{1}{R_n} \tag{III.26}$$

All these changes lead to the structure represented in figure III.44.

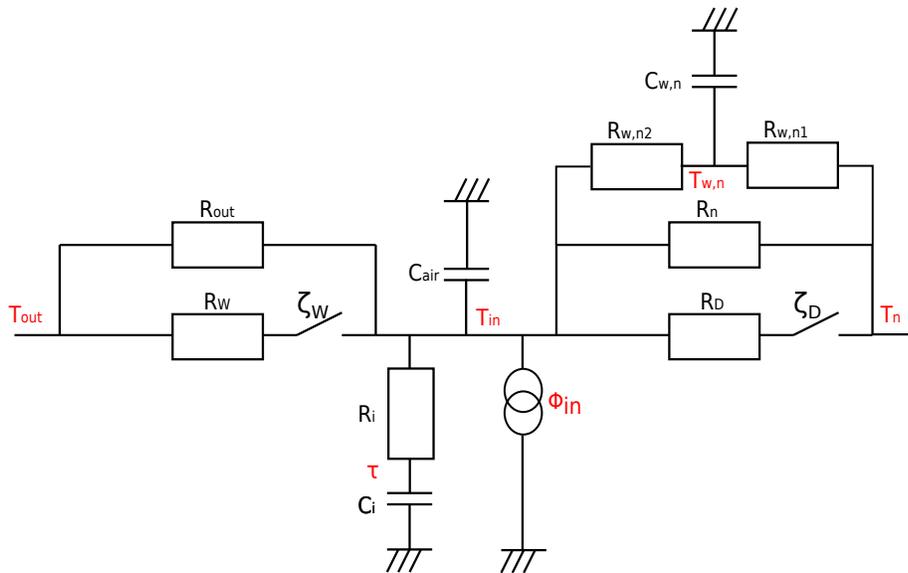


Figure III.43: Model with an air capacitor and one for the inner wall

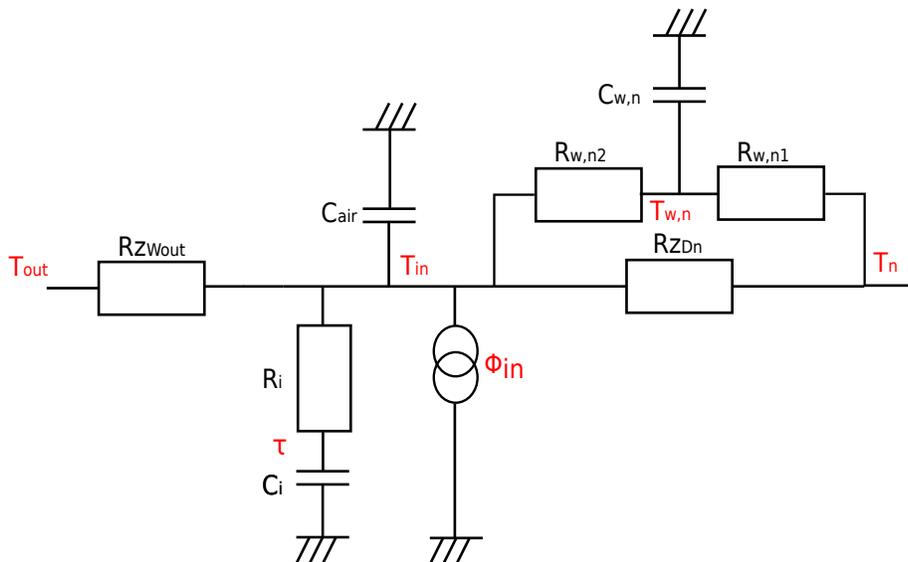


Figure III.44: Model with an air capacitor and one for the inner wall

This new structure of model represented in figure III.44 illustrates well why it does not allow anymore to provide advice on window and door openings. Indeed,

the values of door or window openings have been directly integrated in the values of resistances and cannot be accessed independently. The same logic can be applied to all the model structures. From this statement, the objective is to see whether introducing the window and door opening logit model (binomial regression) can make the model structurally identifiable since the  $\zeta_D$  (or  $\zeta_W$ ) will then be time-dependent.

### 5.8.2 Physical values

In order to confirm the results previously obtained regarding identifiability, the physical values of the different parameters of the study case were computed for the mono-zone office. Indeed, the access to the building information on materials and insulation is still preserved. From this physical information about insulation and composition of walls, physical knowledge is used to compute theoretical values of the different equivalent resistors and capacitors. Nevertheless, the available information is not sufficient to compute all the model parameters. Then, the estimated values obtained via the genetic algorithm for the selected model are compared with those. The theoretical and estimated values are summarized in table III.11.

Parameter	Physical value	Estimated value
$C_a$ (J.K <sup>-1</sup> )	4.62e4	3.72e6
$C_i$ (J.K <sup>-1</sup> )	5.6e6	2.2e7
$C_{w,n}$ (J.K <sup>-1</sup> )	-	7.2e4
$C_{w,out}$ (J.K <sup>-1</sup> )	3.5e6	3.9e5
$R_D$ (K.W <sup>-1</sup> )	0.25	0.011
$R_i$ (K.W <sup>-1</sup> )	0.04	0.01
$R_n$ (K.W <sup>-1</sup> )	0.04	4.3e-3
$R_{wout}$ (K.W <sup>-1</sup> )	7.1e-2	0.77
$R_W$ (K.W <sup>-1</sup> )	0.16	5.1e-3
$R_{wn}$ (K.W <sup>-1</sup> )	0.07	0.63

**Table III.11:** Physical values of model parameters

It can be seen that some values are quite accurate such as  $C_{w,out}$  but a factor 100 can appear for some others ( $R_W$ ). As a single optimization may not be enough to conclude on the physical meaning of the parameters, a dispersion study has been performed.

## 6 Conclusion



In this chapter, multi-regression models of indoor air temperature and CO<sub>2</sub> concentration for energy services have first been explored. They have been applied to a mono-zone study case located in Grenoble. They have been proved to be quite accurate for most situations, their reliability however suffering some exceptions such as demonstrated by the example of the month of June. Indeed, when new physical phenomena appear in the cross-validation period, it can be seen that the multi-regression model finds it difficult to forecast CO<sub>2</sub> concentration. To solve this problem, a new structure has been examined using physical equations to determine the variables of interest of the multi-regression structure. Results were mixed: the CO<sub>2</sub> prediction markedly improved, but the temperature prediction deteriorated. In a last section, the limits of application of such a model have been explored. The ability of the model to adapt in case of missing sensors was first investigated. Results showed that deleting such information improved the results of CO<sub>2</sub> concentration, which can be explained by the specificities of the study case. Indeed, the door of this office appeared to stay open most of the time during summer. Using the CO<sub>2</sub> concentration of the corridor as an input therefore amounts to forecasting something known since the two concentrations are very close. This is why multi-regression models do not seem to fit the requirements for implementing energy services intended for the end-user. To overcome these limitations, new structures of models are investigated including more physical knowledge in order to better forecast new phenomena.

In this chapter, were presented two parameter estimation methods of semi-physical models: a study of a gradient-based deterministic optimisation algorithm guided by an exploratory process (called meta-optimization) and a genetic algorithm. The objective is to accurately forecast the indoor CO<sub>2</sub> concentration and air temperature of the study case. The physical meaning of the estimated parameters obtained is not an objective in this research work since the energy services do not require access to the physical values of insulation, materials or other building specifics.

These methodologies have been applied to different model structures from no capacitance up to 4 capacitances and to a mono-zone study case. Concerning the meta-optimization, the increase in complexity was stopped after 2 capacitances due to different issues. Indeed, the accuracy of the models with 2 capacitances obtained was damaged compared to the reference model. Then, different investigations were led to understand the origins of this result. It has been demonstrated that the

procedure is highly sensitive to the initial parameter values with an error around  $5^{\circ}\text{C}$  for most models and which can reach  $20^{\circ}\text{C}$ . Besides, some convergence issues arose for the more complex models leading also to an exponential increase of the computational time. All of those elements illustrate that the meta-optimization process implemented is not sufficient to guide the optimization algorithm towards the global optimum. Another deterministic algorithm than SQP might solve this issue but as the parameter estimation methods are intended be adapted to more complex study cases and to more complex models, the risks of non-convergence seem too high. Then, another method of parameter estimation based on a stochastic algorithm is investigated in order to allow a comparison between models and a choice of a suitable thermal structure.

With this purpose a parameter estimation method based on a genetic algorithm was implemented. It was aimed at overcoming the robustness problems of the descent algorithm. It has been applied to all eight model structures studied in this work with no problem of convergence or computational time. For each model, the algorithm succeeded to converge in a reasonable time. It was proved that this method is stable and robust and that it does not require a great computational effort. Then, a selection method could be implemented in order to compare all these models according to the criteria presented in chapter II and technical specificities detailed in section 2. The winning model was the Model4C which is both stable and the most accurate with a sRMSE around 0.1 for both summer and winter scenarios i.e. less than  $0.3^{\circ}\text{C}$  of error. Then, a sensitivity analysis has been implemented to ensure that each structural parameter is relevant and do impact the output of the model. It has revealed that in this mode, no simplification of the structure was obvious.

This last method meets all the requirements implied by the implementation of end-users energy services. The model structure is sufficient to describe the slow dynamics while neglecting the quick dynamics. It is as minimal as possible as shown by the sensitivity analysis which allows to keep an acceptable computational time for the estimation process. The estimation method implemented is both ergodic and requiring a low computational time which confers a good reactivity to the end-users services. The period of data required for the estimation is about 1-month which is acceptable for the end-user. It can be compared to some off-the-shelf smart-thermostat that requires 2 weeks to learn the occupants behaviour; period during which users have to be pro-active and to feed information to the system. Here,

models are designed as self-tuning with a minimal set of information required from the end-user.



## Chapter IV

---

# From building description to end-user energy services



### Contents

<b>1 Objectives</b>	110
<b>2 Study case</b>	110
2.1 Architecture and instrumentation	111
2.2 Estimators	112
2.3 Removal of contact sensors	112
<b>3 Mono-zone model applied to a multi-zone study case</b>	116
3.1 Application to the "garden" bedroom	116
3.2 Two-neighbouring zones model applied to multi-zone study case	118
3.3 Impact of a false temperature.	119
<b>4 Generation of the model</b>	120
4.1 Objectives and methodology	121
4.2 Configuration: Gathering information from users	122
4.3 Storage of the information	122
4.4 Model configuration: Generating equations	123
<b>5 Implementation</b>	129
5.1 Configuration of the genetic algorithm	130
5.2 Choice of the right configuration.	130

5.3 Study of the convergence . . . . .	132
6 Application to an energy service . . . . .	134
7 Conclusions . . . . .	137

## 1 Objectives △

Once the model structure selected and validated on a mono-zone office, it is important to change the scale and to evaluate the adaptability of the model to a whole flat. In order to verify the extensibility of the approach, a new study case is introduced involving a three-room apartment. Introducing a multi-zone case study implies a significant increase in the complexity of the model, both in the equation definition and in the parameter estimation problem. For that reason and in the perspective to ease the implementation in the e-consultant, a general framework is set up. Thus, further work had been led on the automatic generation of the model from the simple descriptive information given by the end-user. In this chapter, the study case will be described as well as the estimators used, then the different steps to implement in the process of automatic generation are discussed and finally the genetic algorithm will be applied to estimate the parameters of the new generated structure. In this section, firstly the study case is described. Then, the ARX structure described in chapter III is applied to the multi-zone study case. Finally, the general process for automatically generating the physical model, estimating it and applying it on the multi-zone study case is detailed.

## 2 Study case △

In order to test the model developed in a multi-zone environment, the case studied is a three-room flat located in Aix-les-Bains in France.

## 2.1 Architecture and instrumentation

The architectural plan can be seen in figure IV.1. The flat is equipped with 5 NetAtmo stations, a main one in the Living Room providing measurements of noise level, relative humidity, temperature, pressure and CO<sub>2</sub> concentration and three auxiliaries in each room delivering only measurements on humidity, temperature and CO<sub>2</sub> concentration. The device in the bathroom will not be used in a first time, since the significant amount of extra complexity needed for the model to advice on energy management in the bathroom was deemed irrelevant considering the common use of this particular room. There is also an outside auxiliary measuring temperature, humidity and atmospheric pressure. For a preliminary study, the outside station will not be taken into consideration since the relevancy of the measures really depends on where the user choose to place it. So, it was tried to use more general information coming from more distant weather stations and test whether this information is sufficient. One-year of data was available.

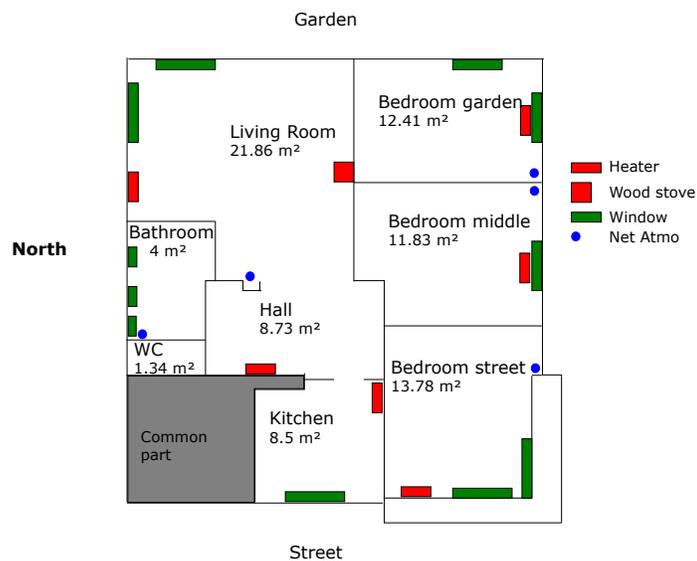


Figure IV.1: Architectural plan of the flat

The flat is occupied by a couple and a baby. Both parents are working on a conventional schedule: the whole week from 9 am to 6 pm. They do not come back home for lunch. About the appliances, each room has a heater represented by a red rectangle on figure IV.1 and a wood stove in the living room represented by a red square. None of these heating sources were measured or precisely modelled. The windows and glazed doors are represented in green.

## 2.2 Estimators

The first study case used in this research was a mono-zone office where occupancy was estimated from the electric consumptions of the different plugs and where users had only a limited number of actions available regarding energy management. In this new case study, the context is completely different. Indeed, occupants can have a major impact on the energy consumption of the building and electric consumption has a remote correlation with occupancy. Then, estimators must be adapted. As a starting point, occupancy in the complex case study was estimated according to some limits set on CO<sub>2</sub> level for the bedrooms and a limit on noise level for the living room. Indeed, CO<sub>2</sub> level has extensively been studied in the literature. Aglan [3] established a physical model to reduce energy use and improve comfort based on CO<sub>2</sub> measurements. Dong et al [28] used CO<sub>2</sub> data combined with a camera to estimate occupancy.

Regarding, the estimation of the heating power, it was previously estimated thanks to a temperature sensor placed on the surface of the heater. In the flat, there is just one air temperature sensor in each room so the heating power estimator has to be adapted. In order to obtain a variable power without introducing too many variables to estimate, the heating power was defined based on occupancy as below:

$$P_{heat} = P_1 + P_2 occupancy \quad (IV.1)$$

## 2.3 Removal of contact sensors

This study case is not equipped with contact sensors. Then, attempts have been made to quantify the impact of the loss of data on the accuracy of the results. More generally, it would ease the implementation of the model on other study cases with less sensors and reduce the instrumentation required. For that purpose, the information of the door and window contacts used in the thermal model is left out. Indeed, if temperature sensors and to some extent CO<sub>2</sub> sensors are currently spreading over the residential and tertiary sector it is less the case for contact sensors. Whereas some can be found in security packages for detection of intrusion, it is very unlikely to find any on indoor openings. Besides, it can raise issues on the field of social acceptability. Then, it would be very interesting to be able to free the models from such data. On the other hand, the rest of data currently used comes from

temperature and CO<sub>2</sub> sensors which cannot be avoided. Some other sensors are used especially for estimating the occupancy in the office but those ones are not common to all case studies since this logic cannot be applied to residential buildings.

The Model4C first applied to the office mono-zone study case to quantify the impact of the loss of data, was then implemented on the multi-zone study case.

### 2.3.1 Application to the mono-zone study case

For this study, data from sensors are then no longer used and the states of opening are added to the parameters to be estimated as a unique value for the whole period. The results can be seen in figure IV.2. Contrary at expectations, deleting data does not necessarily lead to a decrease of performance.



Figure IV.2: Comparison between contact data and not

### 2.3.2 Limits

Including parameters  $\zeta_D$  and  $\zeta_W$  in the parameters to estimate presents two main limitations: first it can be observed in table IV.1 that the values cannot really be interpreted in a physical way since 34% of window openings in winter is highly unlikely and above all it becomes impossible to provide advice based on opening doors and windows. Indeed, in that configuration,  $\zeta_D$  or  $\zeta_W$  are two estimated parameters as are the resistances and capacities of the structure, and as such they participate in the overall estimation of the model. Modifying their value in order to simulate one action or another would then amount to simulating a non-calibrated model. In that situation, it is impossible to predict the model response and to ensure the relevance of the advice provided.

	$\zeta_D$	$\zeta_W$
Winter	0,45	0,34
Summer	0,08	0,41

$1 = open, 0 = closed$

**Table IV.1:** Estimated values of  $\zeta_D$  and  $\zeta_W$

### 2.3.3 Modelling of window openings

In order to bypass this main problem, another approach has been implemented to model the window and door openings.

#### 2.3.3. i) Implemented approach

In order not to maintain the complexity of the model to a similar level and to allow advice on the openings, window and door opening models have been developed intuitively according to the difference of temperature as described in the two following equations:

$$\zeta_D = \frac{\exp(a_D 0 + b_D * (T_{in} - T_n))}{1 + \exp(a_D 0 + b_D * (T_{in} - T_n))} \quad (IV.2)$$

$$\zeta_W = \frac{\exp(a_W 0 + b_W * (T_{in} - T_{out}))}{1 + \exp(a_W 0 + b_W * (T_{in} - T_{out}))} \quad (IV.3)$$

where  $a_D, b_D, a_W$  and  $b_W$  are the parameters to estimate.

It is a highly simplified model but as window or door openings depend on human behaviour and many other environment variables, it is unrealistic to have a model both simple and reliable. The objective here is to obtain a model with a variable rate of openings for window and door.

#### 2.3.3. ii) Application to the office mono-zone study case

This opening model has been integrated in the thermal model selected (Model4C) and applied to the same mono-zone study case presented in chapter II section 1. The temperature profiles can be seen in figure IV.3 and the sRMSE values in table IV.2 for both summer and winter scenarios. It can be seen that the behaviour and the accuracy of both training and validation phases are still accurate and efficient.

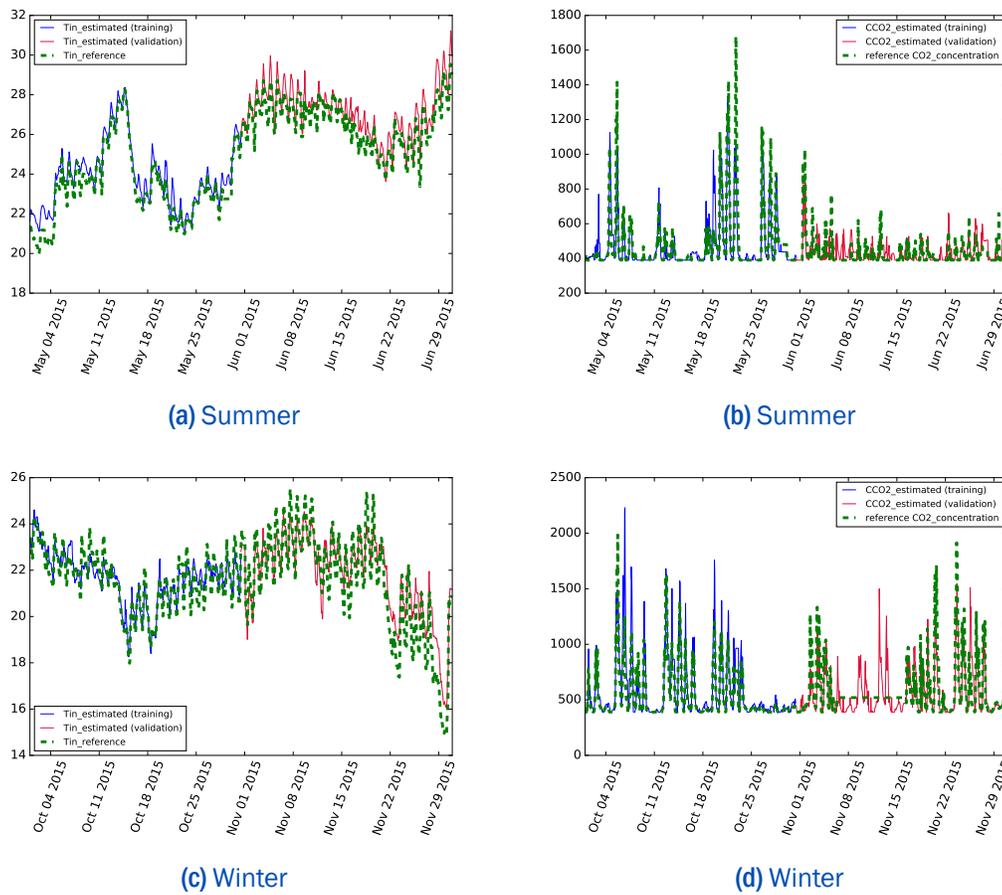


Figure IV.3: Temperature and CO<sub>2</sub> profiles for Model4C with the new model for openings

	Estimation	Validation
Summer	0.06602	0.07468
Winter	0.043061	0.06940

Table IV.2: sRMSE values

However, if the specific behaviour of the opening model is compared to the measured data available in the office, some inconsistencies can be observed (*cf.* figure IV.4). This was not further investigated here since the objective is not to ensure the physical interpretation of the estimated parameter but to obtain a model able to accurately forecast the indoor temperature.

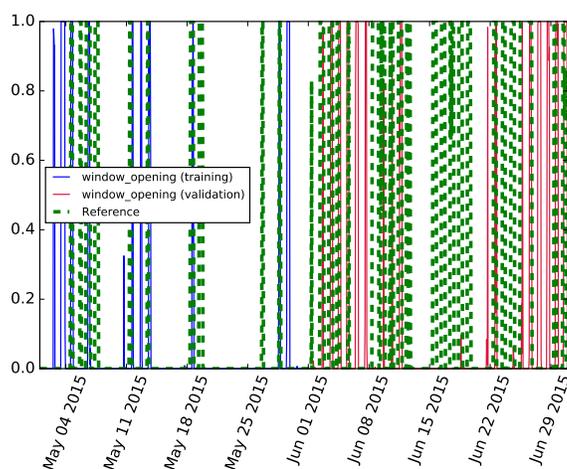


Figure IV.4: Measured and simulated window openings

### 3 Mono-zone model applied to a multi-zone study case △

Firstly, the same mono-zone structure has been applied to a room of the apartment described in the section above in order to validate the model.

#### 3.1 Application to the "garden" bedroom

The “garden” bedroom has first been considered in contact with the exterior and the living room. The selected model is represented in figure IV.5 where  $T_{in}$  is the indoor air temperature of the bedroom “garden” and  $T_n$  the indoor air temperature of the living room, all other parameters remaining unchanged. As previously,  $\phi_{in}$  represents the internal gains : solar gains, heating gains and occupancy gains according to the estimators defined in section 2.2 of chapter IV.

This model has been applied in both summer and winter. Results can be seen in figure IV.6 with the same configuration of the genetic algorithm than before, i.e. 100 generations of 100 individuals. It can be observed that results are less accurate than those of the office mono-zone study case. This is even more obvious in winter since the variations are larger. The indoor air temperature observed during the third week of October reveals a period without occupants. The model fails to predict this fall of temperature with sufficient accuracy: a gap of around 2°C can be noticed. During summer, the model succeeds in modelling the dynamics of the model but the

### 3. MONO-ZONE MODEL APPLIED TO A MULTI-ZONE STUDY CASE | 117

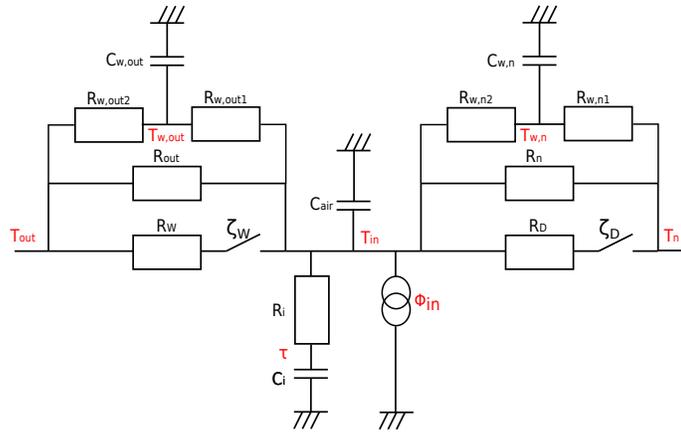


Figure IV.5: Model4C

range is not well represented. The same issue can be observed regarding the CO<sub>2</sub> concentration estimations.

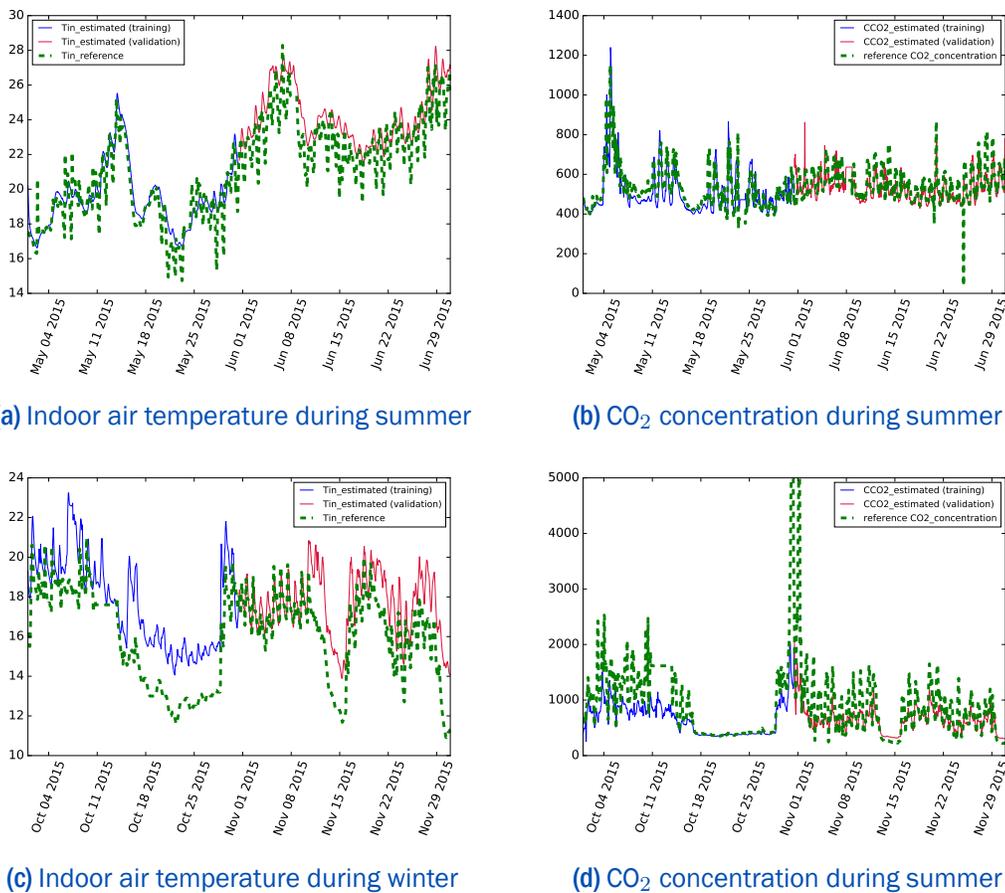


Figure IV.6: Temperature and CO<sub>2</sub> concentration of the bedroom "garden" with a mono-zone model

These results can be explained in different ways. First of all, the variety of states is higher in an apartment than in an office: less automated control, more variability in occupancy. Besides, the “garden” bedroom is not connected to only one neighbouring zone but at least to two, so this model does not well represent the interactions between the rooms. An hypothesis could be that introducing the other zones as other branches impacting the indoor air temperature of the living room and improve the accuracy of the estimation.

### 3.2 Two-neighbouring zones model applied to multi-zone study case

In order to verify this hypothesis, a new model has been implemented taking into consideration the two neighbouring zones surrounding the “garden” bedroom. The new model implemented is represented in figure IV.7. It can be seen that the added connection with the “middle” bedroom does not contain a door opening since only a wall separates these two rooms.

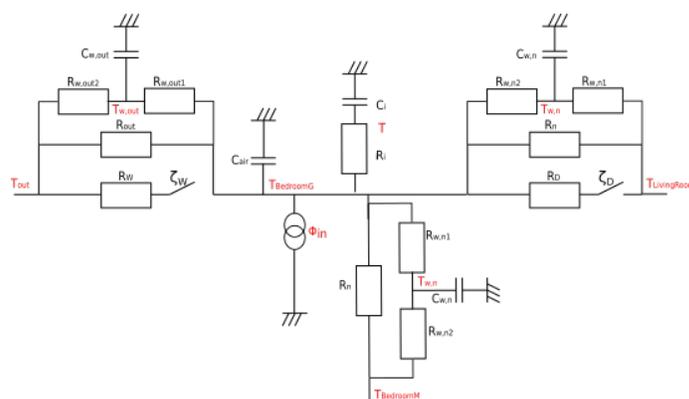
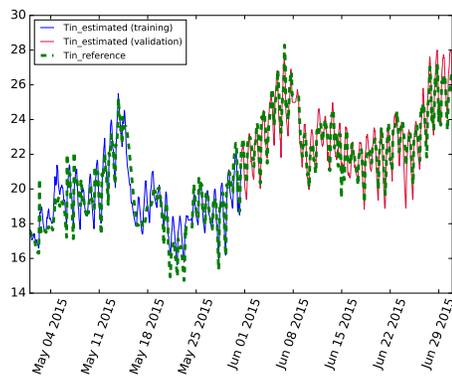


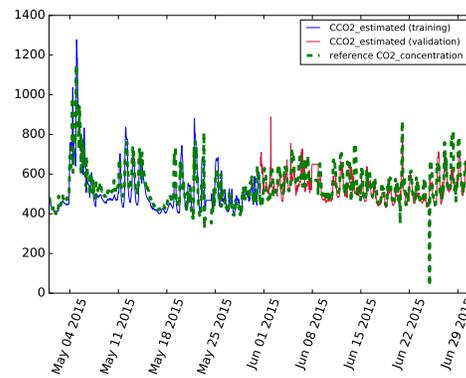
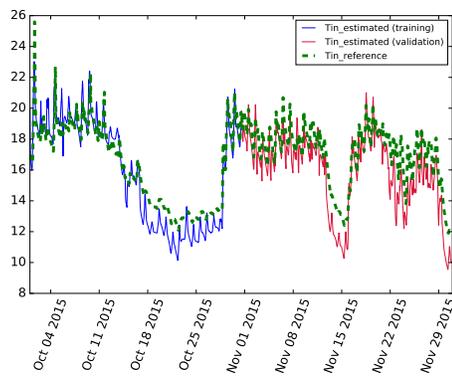
Figure IV.7: Model with 2 neighbouring zones

Consequently, as can be seen in the figure IV.8, the results for CO<sub>2</sub> concentrations have not changed. However, as expected, it can be noticed that the results for indoor air temperature are significantly improved.

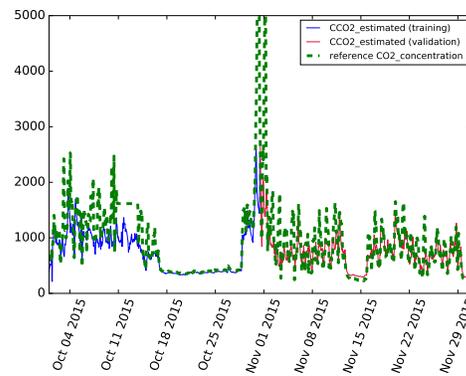
To confirm the graphical results, as previously, sMRSE values are computed for the mono-zone model applied to a multi-zone study case and the two-neighbouring-zones model for both the training and validation phases. Results are presented in table IV.3. It can be observed that taking into consideration the “middle” bedroom improves the sRMSE values by at least 50%. Referring to the selection procedure presented in chapter II, it can be noted that the acceptability criteria (each sRMSE value below 0.1 and stability of the model along the seasons) is not always respected.



(a) Indoor air temperature during summer

(b) CO<sub>2</sub> concentration during summer

(c) Indoor air temperature during winter

(d) CO<sub>2</sub> concentration during winter

**Figure IV.8:** Temperature and CO<sub>2</sub> concentration of the bedroom "garden" with a two-neighbouring zones model

This raises issues concerning the relevancy of the model. Besides, it must be noted that in these two examples it is required to know the neighbouring temperatures. Nevertheless, in a real multi-zone study case, these temperatures must be estimated as well. Although, this way of modelling enables to reproduce the thermal dynamics of the zone considered its accuracy is not sufficient and it does not take all the constraints into consideration.

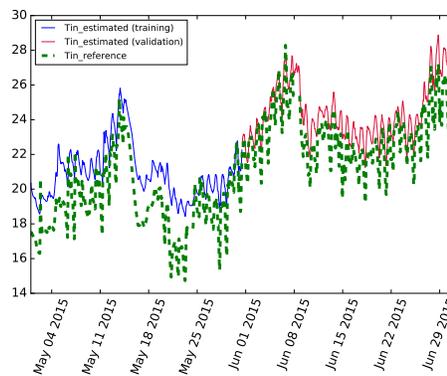
### 3.3 Impact of a false temperature

In parallel, it was tried to quantify the impact on the temperature estimation of a false input, i.e. an error on the estimated neighbouring temperature. To do so, an error is introduced in the neighbouring zone indoor air temperature by adding with a random term ranging from 0 to 3°C during the training phase. It can be seen

Model	Period	sRMSE
Mono-zone	Summer	<i>0.1513</i>
		0.1031
	Winter	<i>0.3524</i>
		0.3592
2 neighbouring zones	Summer	<i>0.0588</i>
		0.0467
	Winter	<i>0.1377</i>
		0.1087

**Table IV.3:** Comparison of sRMSE values for the two different approaches during both training (in italics) and validation phases

in figure IV.9 that the estimated temperature during the training phase presents great differences with the measured temperature. As a consequence, the model encounters difficulties to forecast the indoor air temperature during the validation phase. As expected, the estimated temperature is higher than the measured one. This illustrates well the sensitivity of the output to the input.



**Figure IV.9:** Bedroom "garden" indoor air temperature with an error on the input

## 4 Generation of the model



Then semi-physical models are faced with the constraints of a multi-zone study case. The first issue is to determine the global model structure including every room of the study case and the equations matching it. The problem is that as the size of the study case increases, so does the complexity of the model. In order to properly

model flats or even buildings, making the generation of the equations automatic becomes essential.

## 4.1 Objectives and methodology

The objectives here are to generate the complete model structure of an entire flat only based on the information that an occupant could give through the e-consultant interface. This information should be as small as possible not to bother the occupants. This requires first to structure the information recovered from the end-user on a machine-readable form. Then, to extract this information for a modelling purpose and to generate the equations relative to a specific dwelling. From here, the same methodology as previously can be applied for estimating the parameters and providing the end-user energy services.

In the global project, discussions made appear that different teams required information from the occupant. Then, it was decided to define a tool which can be useful for everyone. For that purpose, it was first needed to establish the list of all the information required for the Human Machine Interface (MHI), the modelling part and the explanation part. From here, a general structure is defined pooling all this information which happened to be presented structurally as xml (Extensible Markup Language) file. This allows to define all the information required from the end-user and ensure that every end-user can access it. Then, the xml file will be used by the different parts of the e-consultant in order to finalize the configuration by itself as described in figure IV.10.

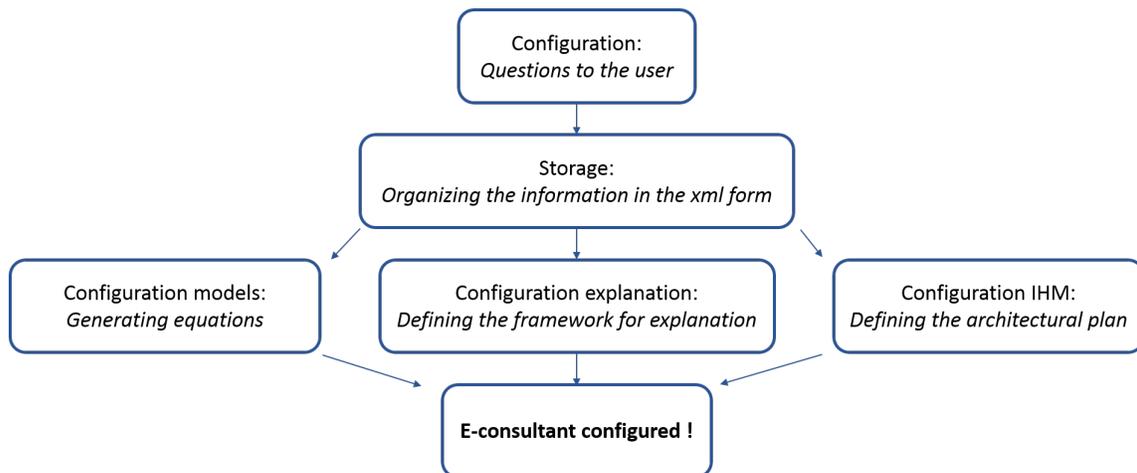


Figure IV.10: Methodology implemented

## 4.2 Configuration: Gathering information from users

The following is the list of information needed from the user in order to properly model the flats:

- The number and the orientation of the rooms,
- The connections between them,
- The surface and the orientation of the windows for each room.

## 4.3 Storage of the information

The minimal structure of the xml file results from the gathering of required information on one hand and on the ontology approach developed in section 2.1 of chapter I on the other hand. The root element of the description file is the architecture and contains two main pieces of information: the latitude and longitude of the study case which are important for the computation of solar gains. Then different child elements can be identified:

- Outdoor
- LivingZone
- Connection

The general organization of the xml can be seen in figure IV.11.

The elements “Outdoor” and “LivingZone” represent respectively the exterior and the rooms of the study case and are described with a list of the sensors present in these different zones. Each sensor is itself described by the physical quantity it measured and the unity of the value returned. The role of the “Connexion” element is to transcribe the architectural plan filled out by the user into usable information. For that purpose, it was proposed to describe it connection by connection where “connection” designates a surface between two rooms. Then, the element “Connexion” contains the references of the two rooms connected and the description of the connection: its type (glazed or opaque surface), its ability to be opened, its orientation and its surface. An example of the xml for a “LivingZone” is given in figure IV.12.

```

<?xml version="1.0" encoding="UTF-8"?>
<description xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="file:description.xsd">
  <architecture latitude="45.1666700" longitude="5.7166700">
    <outdoor name="outdoor" type="Zone">
    <livingZone name="Chambre3" type="Zone">
    <livingZone name="Chambre2" type="Zone">
    <livingZone name="Chambre1" type="Zone">
    <livingZone name="Sejour Entree Cuisine" type="Zone">
    <connection>
    <connection>
    <connection>
    <connection>
    <connection>
    <connection>
    </architecture>
  </description>

```

Figure IV.11: Structure of the xml file

```

<livingZone name="Chambre3" type="Zone">
  <temperature name="Chambre3_Temperature" type="Variable"><sensor name="Chambre3_Temperature" unit="Celsius"/></temperature>
  <CO2concentration name="Chambre3_CO2" type="Variable"><sensor name="Chambre3_CO2" unit="ppm"/></CO2concentration>
  <humidity name="Chambre3_humidity" type="Variable"><sensor name="Chambre3_humidity" unit="%"></humidity>
  <occupancy name="occupancy" type="Variable">
    <estimator name="Ch3occupancy" unit="occupants">
      <sensor name="Chambre3_CO2" unit="ppm"/>
    </estimator>
  </occupancy>
</livingZone>

```

Figure IV.12: Structure of an element "LivingZone" of the xml file

The ways to recover this information from the occupant are studied by other partners in the project and will not be detailed here.

## 4.4 Model configuration: Generating equations

### 4.4.1 Principle

These information are used in different parts of the model. The number of rooms and connexions between them are used to generate the equations and identify the heat exchanges between the rooms. The number and area of windows are used to compute the internal gains and more specifically the solar gains.

From only these information, a generic function has been developed to generate the state-space system of the flat model. To do so, it was decided to pass by symbolic computations even though they are known to be very time consuming. This method

has been developed and applied to the selected model: Model4C. However, as it was developed in a very generic way, every model presented in section 2 and even more could be automatically tested.

The symbolic equation generation is done thanks to the library SymPy implemented in Python 3.4. This library aims to become a full computer algebra system. The first step consists in generating all the symbolic variables required according to the number of rooms and connections between them as well as the state and input vectors. From here, the different derivatives of the state variables need to be expressed. They are then two different situations: models which present an air capacitance in each zone and models which do not.

For every model without an air capacitance, it can be noticed that consequently the indoor air temperatures of the different rooms are not state variables. Besides, they all depend on each other according to the connections between the rooms. In order to express the equations of the derivatives of the state variables in function of only the state variables, it becomes necessary to express all the indoor air temperature in function of the state variables. This step requires to solve a system of  $n$  equations and  $n$  unknown variables (the indoor air temperatures of each room), where  $n$  is the number of rooms, before generating the matrix of the state-space system. Unfortunately, this kind of models introduce a problem due to the computational time required. Indeed, the next step consists in parsing the equations of the derivatives of the state-space to recognize the state variables and the inputs and to extract their coefficients in order to fill the matrix. To do so, the library SymPy implemented in Python expands the expression at its maximum level. Then, replacing the indoor air temperatures by their expression in function of the state variables increases significantly the complexity of the equations. Scanning all the expression of the derivatives of the state variables to identify each one of the coefficients of the state-space matrix becomes then very costly. As an example, for a study-case with only 4 rooms, three days were not enough to achieve the generation of the state-space system. The simulation was launched on a classic laptop with 8 Go of active memory and a Pentium Core i5 with 4 cores. Besides, the computational time required increases exponentially with the number of rooms.

For models with an air-capacitance, this computational time is significantly decreased. Indeed, for the same study case with the same computer, the generation of the state-space system takes less than 1 second. If the initial model does not have an air capacitance, a solution could then be to consider one but adapt its range

to minimize its impact on the system behaviour. A perspective would be either to investigate further this issue in order to decrease the computational time or to use another software specialized in dealing with symbolic computations such as R or Maple. However, for our current problem as the selected model does contain an air capacitance, this problem was not further investigated.

#### 4.4.2 State variables derivatives

The inputs of this automatic process are issued from the xml and stored in two different lists. The first one pools the information about the number and kind of zones:

$$list\_rooms = [(0, 0), (1, 0), (2, 0), (3, 0)] \quad (IV.4)$$

where the first number of each tuple (list between brackets) is the number of the room and the second is to indicate if there is an air capacitance (0) or no (1).

The second one pools the information about the number and kind of connections between the rooms and is expressed as a list of lists:

$$list\_connections = [[(0, 1), 2], [(1, 2), 1], [(0, 2), 2], [(0, 3), 2], [(2, 3), 1]] \quad (IV.5)$$

where, in a sub-list, the two elements of the tuple are the numbers of the rooms concerned by the connection described, and the last element is the type of the connexion: 2 means there is a door and 1 a wall.

The two last equations are applied to the study case presented in section 2 of chapter IV. In this case, the living room is numbered 0, the “garden” bedroom 1, “middle” bedroom 2 and the “street” bedroom 3. Each room is considered to be connected to the exterior by a wall with a window and modelled by a capacitance. Otherwise, it will just be necessary to add the exterior as a zone and to add the connections with their kind to the list below.

The two lists are then used to create all the symbolic variables required for the generation of the state-space system. First, the state variables are created. The number of “fictive” wall and outer wall temperatures to be considered is directly linked to the number of rooms if there is an air capacitance. Otherwise, the temperatures of the “fictive” walls do not intervene. The number of inner wall temperatures is based on the *list\_connections* and one is created for each connection of type 2 using the numbers in the tuple to specify the wall temperature considered. And finally,

the indoor air temperatures are added thanks to the *list\_rooms* if there is an air capacitance. Regarding the input vector it is constituted of the outdoor temperature and all the internal gains for each room.

Once, this is defined, the next step consists in building the derivatives of the state variables. For this step, for each building blocks defined in section 2 of chapter II, the relative equations have been written. Their generation is also based on the lists defined on the beginning of the process.

For each room present in *list\_rooms*, if there is an air capacitance, two derivatives of the state variables can be written due to the modelling of the zone itself:

$$\frac{d\tau_i}{dt} = \frac{1}{C_i} * \left( \frac{T_i}{R_i} - \frac{\tau_i}{R_i} \right) \quad (\text{IV.6})$$

$$\frac{dT_i}{dt} = \frac{1}{C_{air,i}} * \left( \frac{\tau_i}{R_i} - \frac{T_i}{R_i} + \phi_i \right) \quad (\text{IV.7})$$

where  $i$  is the number of the room considered,  $T_i$  is the indoor air temperature of the room  $i$ ,  $\tau_i$  is the “fictive” wall temperature of the room  $i$ ,  $R_i$  is the resistance of the “fictive” wall of the room  $i$ ,  $C_{air,i}$  is the capacitance of the indoor air of the room  $i$ ,  $C_i$  is the capacitance of the “fictive” wall of the room  $i$  and  $\phi_i$  is the internal gains of the room  $i$ .

Then two other equations can be written since it is considered that each room is connected to the outdoor thanks to a window and a wall with a capacitance:

$$\frac{dT_i}{dt} = \frac{1}{C_{air,i}} * \left[ \left( \frac{\zeta_{Wi}}{R_{Wi}} + \frac{1}{R_{outi}} \right) * T_{out} + \frac{T_{wouti}}{R_{wouti_a}} - T_i * \left( \frac{\zeta_{Wi}}{R_{Wi}} + \frac{1}{R_{outi}} + \frac{1}{R_{wouti_a}} \right) \right] \quad (\text{IV.8})$$

$$\frac{dT_{wouti}}{dt} = \frac{1}{C_{wouti}} * \left[ \frac{T_i}{R_{wouti_a}} + \frac{T_{out}}{R_{wouti_b}} - T_{wouti} * \left( \frac{1}{R_{wouti_a}} + \frac{1}{R_{wouti_b}} \right) \right] \quad (\text{IV.9})$$

where the index  $i$  identifies the room considered and the different variables are represented on figure IV.13.

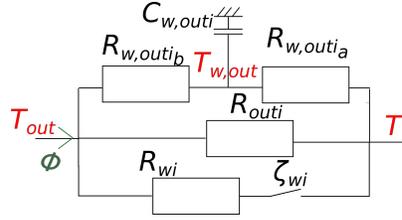


Figure IV.13: Representation of an outer wall

Finally, for each element of the *list\_connections* of type 2, the following equations can be defined:

$$\frac{dT_i}{dt} = \frac{1}{C_{air,i}} * \left[ \left( \frac{\zeta_{Dij}}{R_{Dij}} + \frac{1}{R_{nij}} \right) * T_j + \frac{T_{wnij}}{R_{wnij_i}} - T_i * \left( \frac{\zeta_{Dij}}{R_{Dij}} + \frac{1}{R_{nij}} + \frac{1}{R_{wnij_j}} \right) \right] \quad (IV.10)$$

$$\frac{dT_j}{dt} = \frac{1}{C_{air,j}} * \left[ \left( \frac{\zeta_{Dij}}{R_{Dij}} + \frac{1}{R_{nij}} \right) * T_i + \frac{T_{wnij}}{R_{wnij_j}} - T_j * \left( \frac{\zeta_{Dij}}{R_{Dij}} + \frac{1}{R_{nij}} + \frac{1}{R_{wnij_i}} \right) \right] \quad (IV.11)$$

$$\frac{dT_{wnij}}{dt} = \frac{1}{C_{wnij}} * \left[ \frac{T_i}{R_{wnij_i}} + \frac{T_j}{R_{wnij_j}} - T_{wnij} * \left( \frac{1}{R_{wnij_i}} + \frac{1}{R_{wnij_j}} \right) \right] \quad (IV.12)$$

where the index *i* and *j* identify the rooms involved in the connection and the different variables are represented on figure IV.14.

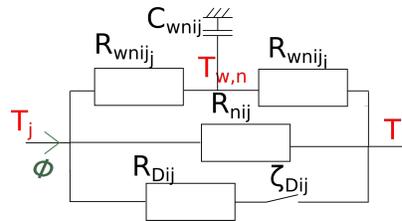


Figure IV.14: Representation of an inner wall with a door

If the type of the connection is different, the terms relative to the branch with the switch must not be considered. It can be noticed that the derivative of the indoor air temperature appears several times in the equations below. This is simply because every segment added is linked to the indoor temperature and so impacts its

evolution. Combining these different equations allows to reconstruct the derivatives of every state variables of the global model.

From these equations, parsing methods have been used to find the different coefficients of the state space matrix and build them. As explained in the previous section, the first step the library Sympy requires is to expand every terms of these equations. Then, it is possible to recover the coefficients of the state-space matrix and to build them. The more complex the equations are, the longer this step will be. To decrease this computational time, all the equations have been simplified. Each derivative has been written under the form:

$$\frac{dX_i}{dt} = \sum_i X_i + \sum_k u_k \quad (\text{IV.13})$$

where  $X$  are the state variables,  $i$  the number of the state variables,  $u$  the inputs and  $k$  their number.

It is then easier for SymPy to identify the coefficients of the state-space matrix. This operation in that configuration takes less than 1 second. Once the system is defined, it is not longer necessary to appeal to symbolic computation. The system is then translated in a non-symbolic form. To do so, the process of the genetic algorithm is called. Indeed, the library Deap in Python 3.4 requires to register every parameter which needs to be estimated on a certain form in a toolbox. From here, the library can generate the population 0 with random values for the different parameters. Each one of the values are registered in a dictionary under their name as key. The variables can then be handled as usual with a very low computation cost.

Once the state-space system is well defined, the model can be estimated thanks to the genetic algorithm presented in the chapter III. The first step consists in replacing the global variables defined to gain some time by their expression according to the different parameters.

In order to illustrate, the thermal circuit of the flat is represented in figure IV.15. It is a combination of all the different building blocks defined in section 2 of chapter II and applied to the new study case respecting the selected structure: Model4C. This means that each room has been represented by an air capacitance, a “fictive” wall and internal gains. Each connection between rooms is represented with a resistance and two resistances and a wall capacitance in parallel. If there is a door between

these rooms is added in parallel a resistance and a opening rate. Each room is connected to the exterior in the same way but with a window opening rate.

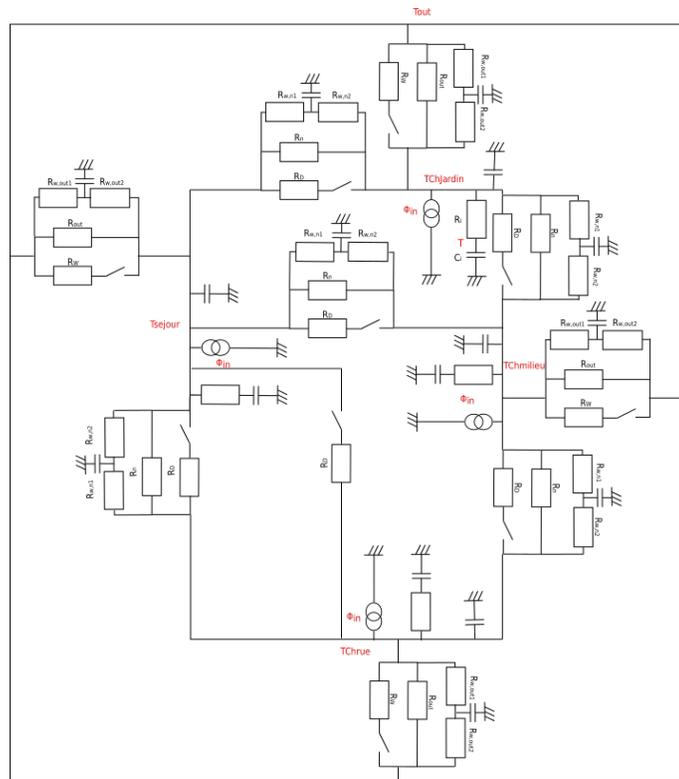


Figure IV.15: RC model of the flat

## 5 Implementation



Once the structure defined (*cf.* figure IV.15) and automatically generated thanks to the proposed process, the next step consists in estimating the parameters and validating the results. If the mono-zone parameter estimation model was made of 12 parameters, the number of parameters reach 69 for the multi-zone study case. Then the model of openings presented in section 2.3.3 was not implemented at first since it implies a high increase in terms of computational time. The same genetic algorithm presented in section 5 is used for that purpose.

As a reminder, the measured inputs are the data from the Netatmo stations in the flat (*cf.* section 2 of chapter IV) which provide: indoor air temperature, CO<sub>2</sub> concentration and relative humidity. The one in the living room measures in addition the noise level. Data from a weather station are also used to forecast the

solar gains. Occupancy and heating power are estimated according to the estimators explained in section 2.2 of chapter IV.

## 5.1 Configuration of the genetic algorithm

### 5.1.1 Starting configuration

For the mono-zone study case, 100 generations of 100 individuals were simulated and it was enough to reach a sufficient accuracy. However, if the same configuration is applied for the multi-zone study case, convergence is not reached (i.e. the accuracy obtained is too low) due to the increased complexity of the problem (*cf.* figures IV.16 and IV.17). This is especially true for the CO<sub>2</sub> estimations which are measured via the Netatmo sensors as explained in section 2 of chapter IV. Actually, for three rooms out of four, the CO<sub>2</sub> profiles obtained seem likely but estimated with a great offset while the last one is accurate. As for temperature, it can be noticed that though the results are not as good as they were for the mono-zone study case, they remain promising. Indeed, the dynamics are well represented. Still, the gaps between predicted and measured values can reach 2°C which can damage user's trust. Indeed, there are already discussions concerning the 19°C required by the French regulation whether it is sufficient or not. Many voices rise claiming that 21°C would be more relevant. This illustrates that a gap of 2°C is considered as significant by most (at least) French people and researchers.

It can be observed that for the estimated temperature in the living room, results seem to be less accurate during the training phase. This is due to a daily calibration during the cross-validation phase: every day at the same hour the estimated state spaces are set to the observed ones. This allows to significantly improve the results and is in agreement with the energy services considered. Indeed, the daily horizon is the most relevant for most services.

## 5.2 Choice of the right configuration

The next step is then to determine the right configuration to reach the accuracy required for the energy services. Different tests have been launched in order to reach the target: considerably decreasing the number of individuals and increasing the number of generations on one hand, and increasing both individuals and generations in a second hand.

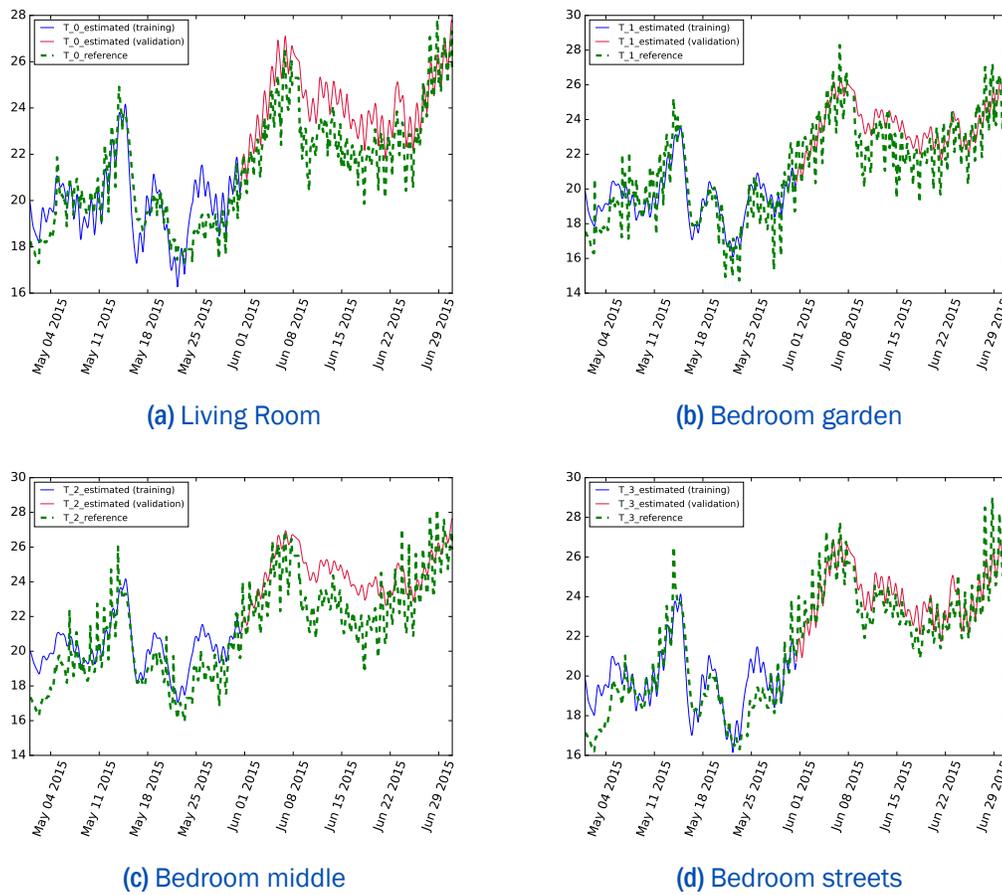


Figure IV.16: Temperature profiles for the summer scenario

	Training	Validation
“garden” bedroom	0.2046	0.1780
“middle” bedroom	0.1825	0.1906
“street” bedroom	0.1536	0.2003
Living room	0.2763	0.2148

Table IV.4: sRMSE values for the simulation of 100 generations of 100 individuals

Then, another test was launched with 1200 generations of 800 individuals (*cf.* figure IV.18). The results confirm that the previous configuration was not sufficient to ensure the convergence of the algorithm. Indeed, if no significant improvement

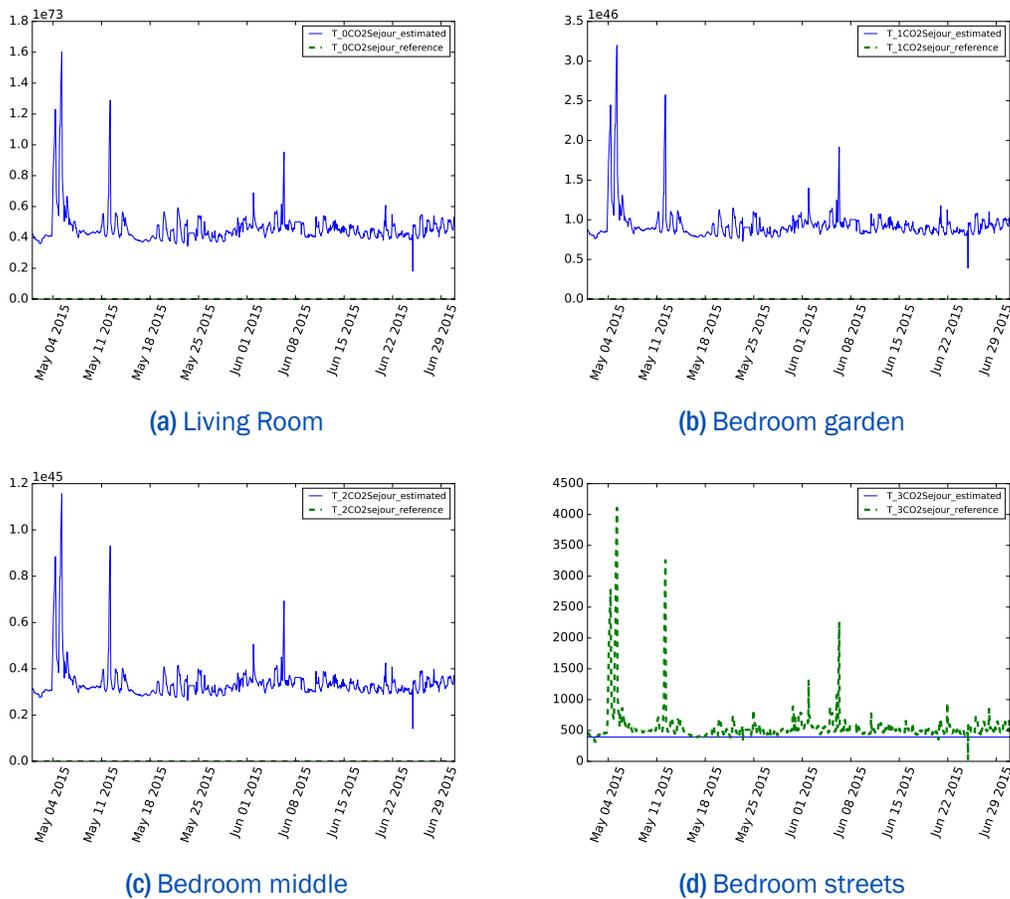


Figure IV.17: Temperature profiles for the summer scenario

can be noticed on the prediction of the temperature, the  $\text{CO}_2$  trends are greatly improved. Nonetheless, the convergence does not seem to be reached. The same tools are applied in order to validate and quantify the accuracy of the model (*cf.* table IV.5). As expected, the accuracy is not as high as it was for the mono-zone office.

### 5.3 Study of the convergence

In order to see whether or not the lack of accuracy results from a convergence problem, the evolution of the error along the optimisations have been plot. Figure IV.20 shows an healthy behaviour with an error decreasing all along the generations. However, it must be noted that from the generation number 600, the diminution of the error is really slow down. This confirms that the results obtained are not due to a convergence problem.

	Training	Validation
“garden” bedroom	0.2310	0.2285
“middle” bedroom	0.1950	0.1937
“street” bedroom	0.1874	0.1950
Living room	0.3476	0.2042

Table IV.5: sRMSE values for the simulation of 1200 generations of 800 individuals

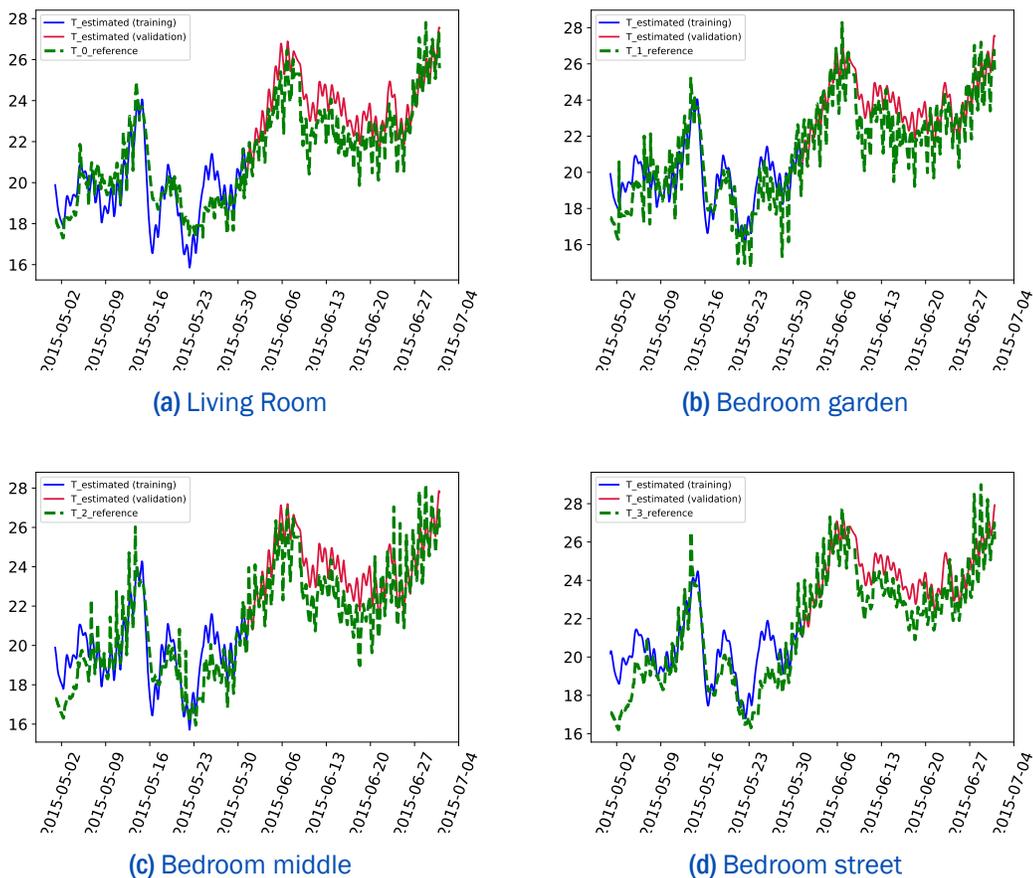


Figure IV.18: Temperature profiles for the winter scenario (1200 generations of 800 individuals)

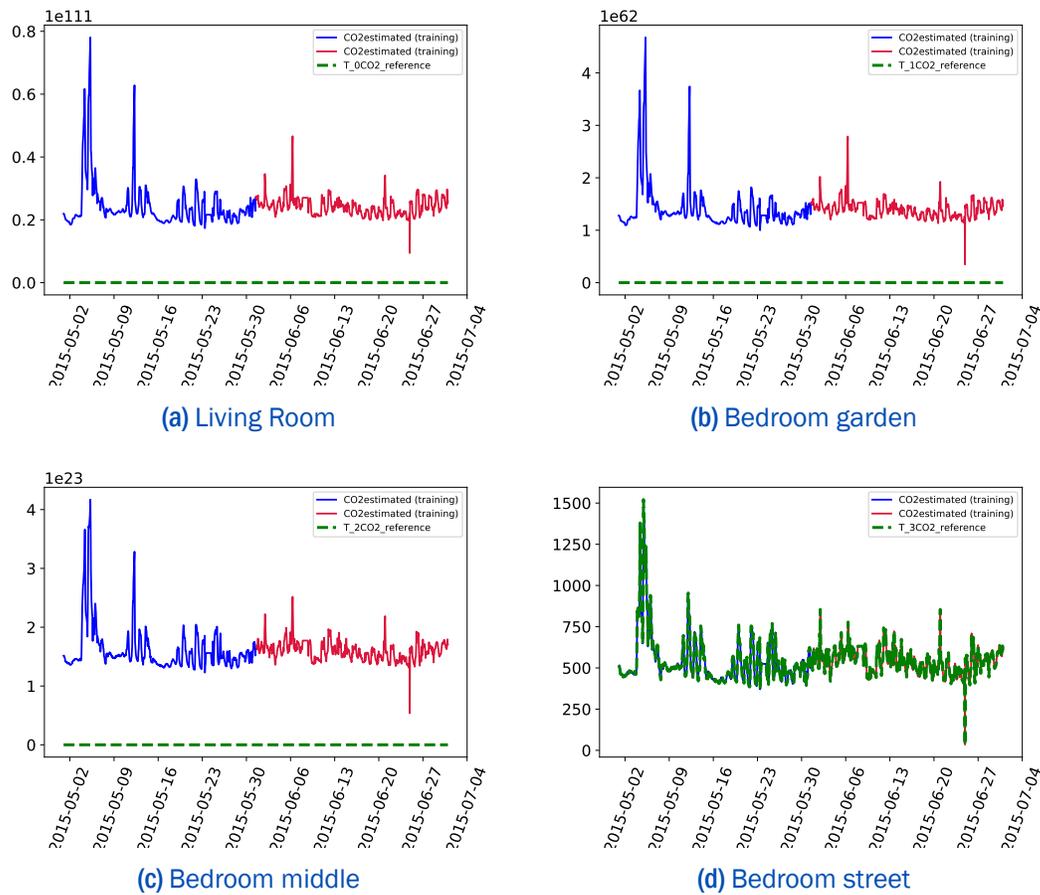


Figure IV.19: CO<sub>2</sub> profiles for the winter scenario (1200 generations of 800 individuals)

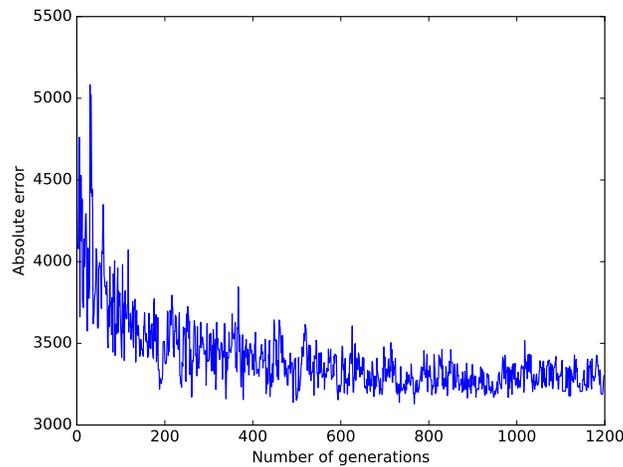


Figure IV.20: Evolution of the absolute error along the generations

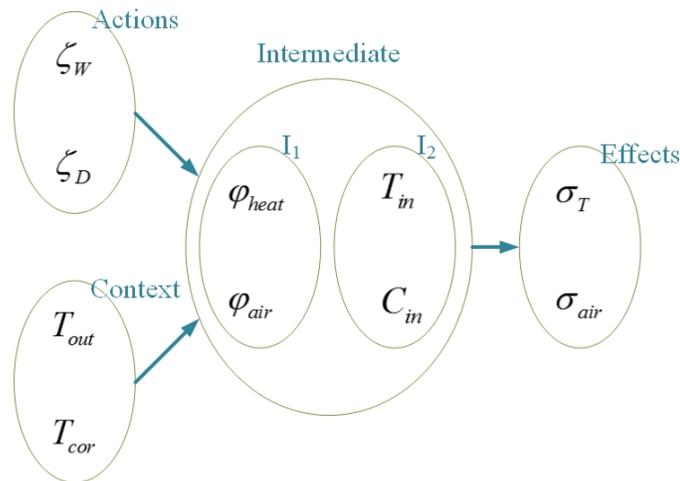
## 6 Application to an energy service



As explained all along this thesis, these models have been developed to serve end-user energy services. However, the implementation of these services are not part

of this research work. Amr Alyafi, part of INVOLVED project, is in charge of the generation of explanations. In this section, a part of his work is presented [5] in order to illustrate how the models developed in this thesis will be used. For the moment, it has been applied to the mono-zone office study case described in chapter II.

The semi-physical model developed in this thesis is used for generating differential explanations. The goal of his work is to create a contextual causal explanation and so to identify the causes and consequences of the different actions available. In this context the causes are the occupant's actions or the contextual phenomena and the consequences are divided between the final effects and the intermediate effects. It is summarized in figure IV.21 where the arrows represent the cause-effect relation between the groups and  $\sigma$  the dissatisfaction felt by the user in terms of temperature or air quality. This dissatisfaction criteria is determined based on the preferences defined by the end-user. The distinction between the intermediate and effects groups is whether or not the variable considered is experienced directly by the end-user or not. The intermediate effects allow also to build a complete causal chain and to help users understand why their comfort or energy consumption have been affected.



**Figure IV.21:** General schema of explanations

To generate the explanations, a concept of a qualitative distance is defined by comparing the actual scenario to the optimal one.

This mapping is done to show to the occupants the impact of their actions on the comfort criteria, and to convince them why do they need to change their behaviour when it is far from optimality. First, the semi-physical model is used to determine the optimal set of actions to reach the objective defined by the end-user regarding the temperature and the CO<sub>2</sub> concentration. From here, the following

step consists in determining the distance between the actual occupant's behaviour and the optimal one. From the optimization results, occupants can understand their position according to the possible set of the optimal solutions for the entire day. Differential explanations will help them to understand what needs to be done to reach the chosen optimal solution according to their objectives. This explanation is done by differentiating explanations, i.e. based on the comparison of two scenario on different levels: the set of actions, the intermediate variables and the effects. It is then possible to define an indicator representing the impact of the actions taken on the optimality. The cause-effect relations must be highlighted to improve the occupant's understanding. This step leads to the construction of the table represented in figure IV.22 where the cause-effect relations are represented on both the final and intermediate effects. In this table, it can be observed that different levels of impact are modelled. This is a direct consequence of the qualitative distance computation to the optimal scenario. Resorting to the physical model including capacitances also allow to evaluate the delayed impact of an action on the different variables.

hour	$\Delta$ actions	$\Delta$ effects	$\Delta$ intermediates
08:00			OUT
09:00			COR
10:00			COR
11:00			COR
12:00			COR
13:00			COR
14:00			COR
15:00			COR
16:00			
17:00			COR
18:00			
19:00			
ALL			COR

The average gain for the entire day if occupant followed the recommended actions

- (Opening / closing \) the door
- (Opening / closing \) the window
- Heat flow (to / from \) the office
- Air flow (to / from \) the office
- (Enhance / decrease \) Air quality in the office
- (Enhance / decrease \) Thermal comfort in the office
- The amount of change

**Figure IV.22:** Differential explanations help the inhabitants to understand their actions like at 4pm, the user behaves correctly (its action was similar to the proposed optimal plan); at 12am, the inhabitant should have opened the door for much longer time, this would helped him getting a high improvement in his thermal comfort and slight enhancement in the air quality.

Still, this table amounts too complex to be well understood by the end-user. That is why the last step of this work consists in generating automatically a well-written statement synthesizing the information included in the table. For that purpose, a first model was developed in ARIANE-HELOISE [13] for a feasibility demonstration: the GRA-FRA ("GRAphe vers texte en FRAn ais": graph to French text) model. It was specifically designed for generating messages from the tables containing differential explanations. In figure IV.23, it can be observed an example of the output generated by the automatic text generation process.

**Traduction**

Dans le créneau horaire 17h-18h, si vous aviez laissé la porte et la fenêtre ouvertes beaucoup plus longtemps, la qualité de l'air aurait diminué un petit peu, il y aurait eu un courant d'air sensible vers le couloir et un flux thermique léger venant de l'extérieur ;

Figure IV.23: Output from Ariane-Heloise GRA-FRA GM phase

## 7 Conclusions △

In this chapter, the application of models to a multi-zone study case have been explored. This new study case has been challenging for different reasons. Despite, the simple increase in the complexity of the model, it presented also a great loss of data compared to the office study case. In the same way, estimators are more difficult to define considering the change of function of the study case: from an office to a dwelling. All of these modifications increase greatly the complexity of the estimation process. The relevance of the mono-zone semi-physical model applied to each room of a more complex study case has been tested. It revealed that the performances were acceptable but that this model was very sensitive to the inputs and so that it could not be applied to all the rooms since each of them is dependent on one another. The next step consisted then in implementing a multi-zone semi-physical model. But, the implementation of such a model requiring to define complex equations and structure, an automatic generation has been implemented. The developed automatic process is able to determine the state space of the model based only on the information provided by the end-user such as: orientation and area of windows and rooms and connections between the different rooms. From here, equations are generated and the estimation process is launched on the data available. Regarding the parameter estimation process based on the genetic algorithm, it needed to be adapted since the number of parameters to estimate increased greatly for the multi-zone model. Then from 100 generations of 100 individuals, 1200 generations of 800 individuals were generated and the probability of mate and mutations were increased as well to improve the exploration. However, despite these modifications, it has been shown that predicting CO<sub>2</sub> concentration with a good accuracy remained a challenging topic. Regarding the temperature, the results were good but still less accurate than for a mono-zone study case. It can be noted nonetheless that the trends obtained for the CO<sub>2</sub> suggest that the convergence is not reached. Different simulations have been run with higher number of generations and individuals but the same trends were obtained.



## Conclusions and future works



### Conclusion



This thesis presented a new methodology to select a suitable thermal and aerodynamic model for end-user energy services. The objective was to determine a model structure able to predict temperature and CO<sub>2</sub> evolutions on a 24-hour horizon based on little expert knowledge. These requirements induce the different constraints regarding the model structure and its configuration. The structure might neglect fast dynamics and sufficiently recover slow dynamics to fit a daily prediction with 1-hour time step. The learning phase must be done within a reasonable amount of data (1 month maximum) and based on the fewest sensors as possible. It has been performed with only 6 sensors for the mono-zone study case and one Netatmo station per room (providing three measurements) for the multi-zone study case.

The first step consisted in a state-of-the-art regarding different fields. The sociology of energy helped us to determine the relevant energy services which will engage end-users towards habit changes. The study of the different model structures used within the community of building energy analyses enabled to identify which one that could answer the energy services needs as well as the parameter estimation methods.

Then, different kind of models have been implemented and validated on two different study cases: an office and a flat. This enabled to validate the model on buildings with different purposes and different scales and then to open many opportunities of implementation. The first kind of models implemented are multi-regression models; structure only based on data which do not require expert knowledge because they are easy to estimate. These models proved to forecast with a good accuracy the temperature and CO<sub>2</sub> concentration on the mono-zone office. However, for instance, the month of June which differs a lot from the month of May regarding window openings and occupancy presented poor performances in particular for the

CO<sub>2</sub> estimation. This behaviour raised issues regarding the ability of the model to guarantee performances along the year. On top of that, the implementation on a multi-zone study case returned low performance. In order to outreach these limits, another model structure was introduced embedding physical knowledge: the RC-models dedicated to runtime end-user energy services. These are commonly used among the community of building modellers and based on thermal circuits. It was added switches in order to take into consideration door and window openings which was not found in the literature. From a simple 1-C structure, different models of higher order were defined in order to test the different ways of inertia modelling found in the state-of-the-art. Acceptability and selection method were set up to compare the performances of these different models and to identify the one answering the best to all the requirements of end-users energy services. For these models, the estimation process is more complex and computationally costly than for multi-regression structures. Hence, firstly a classic descent-algorithm called SQP has been implemented for its low computational cost. The research space being wide and complex, a framework has been set to improve the exploration of the research space and avoid to be stuck in local optimum. Unfortunately, this method requires initial values and its convergence was very dependant on it. It has been shown that as soon as the complexity of the model structure increased, the algorithm did not reach the convergence. In the following, another method has been tested: genetic algorithms which are known for implying higher computational costs. Among the choice of genetic algorithm, was chosen NSGA-II which preserves the diversity of the offspring generated thanks to the computation of the crowding distance. The combination of RC-models and NSGA-II has been proven to answer the end-users energy services requirements: stability along the seasons, good accuracy and learning phase of 1 month. The 8 RC-structures have been applied to the mono-zone office and compared according their accuracy and prediction stability along the seasons. The final selected structure has then been further studied. A sensitivity analysis has been performed to ensure that every structural parameters had a significant impact on the model output. This analysis did not reveal an obvious set of variables without an impact on the output. Then, the structure stayed unchanged. To complete the investigation of this final structure, a state observer is set. If the model structure is not directly observable, it has been demonstrated that the non-observable states have a negligible impact on the output. It is then possible to implement a state observer and so to initialize the model for daily-horizon optimizations.

Finally, the scale of the study case has been increased from a mono-zone office to a whole flat. On top of that, the flat is equipped with very few sensors: temperature, humidity and CO<sub>2</sub> concentration in the bedrooms and sound level measure in the living room. The challenge of parameter estimation was then doubly increased. To do so, a complete framework has been implemented to allow the automatic generation of the model from end-users configuration information. In the current state, from the very simple information given by the end-user about architectural plan, orientation and size of windows and rooms, the complete RC-model of the flat is automatically generated and estimated via the NSGA-II algorithm. The configuration of the algorithm must have been adapted in order to take into consideration the increase of parameters to estimate. The CO<sub>2</sub> estimation has not succeed probably because of a lack of sensors. But, the temperature estimation reached a satisfactory level considering the complexity of the model.

To conclude, an illustration of how the model will be used in future works for implementing energy services, the work of Alzhouri Alyafi has been presented. It highlights the gains of the semi-physical models for the generation of explanation as well as the process implemented.

## Future works



This work appeals extensions.

Regarding the results on the CO<sub>2</sub> estimation on the multi-zone study case, an hypothesis is that the lack of sensors is partly responsible of the estimation results. To validate or not this hypothesis, it would be interesting to use the modelling generation approach to other study cases with more sensors and especially door and windows contacts.

Following the same idea, there is for sure a link between the instrumentation available and the possible accuracy of the model. Indeed, the more sensors are installed, the less it is necessary to build estimators. For instance, estimating with a good accuracy and even measuring the effective heating power or occupancy would significantly impact the accuracy of the prediction. Then, investigating the evolution of the accuracy of the model according to the instrumentation as well as the impact on the structure would give a meaningful insight.

Clustering methods could be used to segment the days into clusters of similar days. One challenge of the parameter estimation is the length of data necessary to estimate the model. A clustering process could give information on the similarity of the day considered compared to the precedent ones it has seen. If there is a match, it would then allow to learn the model on fewer days and enhance the accuracy of the model. During this thesis, it was shown that one model estimated whether in summer or in winter could predict the evolution of temperature and CO<sub>2</sub> all along the year. But, perhaps these clustering methods could enhance the accuracy of the prediction by estimating several different set of parameters for the different clusters.

The proposed validation methodology has focused on the estimation error for temperature and CO<sub>2</sub> but end-user energy services use sometimes estimation of energy needs. Therefore, the validation at service level should be done: maybe with poor estimations the services could yield good results. A first example of an energy service has been presented through the work of Amr Alzhouri Alyafi but many others are still to be developed. As the specifications of the end-user energy services are different from one another, it should be tested that the structure is resilient to all the services.

Another issue regarding energy management is the detection of drifts or abnormal behaviours. Indeed, flats or offices are evolving along the time and some systems can be replaced or added. In this case, the learnt model will not be adequate anymore. Implementing an automatic detection of such phenomenon and identify the causes could enhance the robustness of the model and ensure that the accuracy of the estimation can be guaranteed any time.

# Résumé en français



## 1 Introduction



Dans le contexte de transition énergétique actuel, les bâtiments représentent un enjeu majeur puisqu'en France comme à l'étranger ils sont responsables de près de 40% de la consommation d'énergie finale. Les recherches dans le domaine de l'énergétique du bâtiment se sont dans un premier temps concentrées sur les systèmes thermiques et l'enveloppe bâtiment. Nous savons donc aujourd'hui construire des bâtiments très performants. Toutefois, il subsiste toujours un écart important entre la consommation prévue lors de la phase de conception et la consommation réellement observée en phase d'exploitation. Dans un premier temps, le bâtiment a évolué vers toujours plus d'automatisation en se basant sur l'hypothèse que l'occupant était la cause de cet écart. L'objectif était ainsi d'assurer le fonctionnement optimal des différents systèmes CVC installés. Malheureusement, cela n'a pas réduit l'écart entre consommation prévue et consommation réelle. Suite à ce constat, des sociologues sont allés à la rencontre des occupants de ces nouveaux bâtiments économes pour identifier les causes de ce phénomène. Les premiers retours ont mis en lumière qu'un occupant en situation d'inconfort allait ressentir son inconfort de manière décuplée s'il n'avait pas de contrôle sur son bâtiment. Pour atteindre une réduction effective des consommations énergétiques dans le bâtiment, les sociologues suggèrent de rendre le pouvoir à l'occupant. Ce travail de thèse s'inscrit dans le projet INVOLVED financé par l'Agence Nationale de la Recherche (ANR) dont l'objectif est d'impliquer l'utilisateur dans sa gestion énergétique via une Interface Homme - Machine (IHM). Cette IHM permettrait de fournir différents services énergétiques à l'utilisateur final pour lui permettre de comprendre son impact sur le bâtiment et l'aider à atteindre au mieux ses objectifs.

Ces différents services énergétiques ont besoin de prédire l'évolution de la température et du CO<sub>2</sub> dans les prochaines 24 heures à pas de temps horaire. Ils ont

donc pour cela besoin de se baser sur un modèle thermique du bâtiment. Toutefois, comme l'IHM s'adresse au grand public, il n'est pas envisageable de faire appel à des connaissances expertes du bâtiment comme par exemple les caractéristiques des matériaux d'enveloppe. La Simulation Thermique Dynamique (STD) n'est donc pas ici un outil pertinent. Il est alors nécessaire de recourir à des modèles simplifiés pouvant se baser en partie sur des données capteurs pour combler le manque de données expertes. Dans cette thèse, différents types et typologies de modèles ont été étudiés et testés sur deux cas d'études réels : un bureau et un appartement. Dans le chapitre 1, est conduit un état de l'art sur la sociologie de l'énergie afin d'identifier les services énergétiques pertinents ainsi que les différents types de modèles utilisés pour la gestion énergétique des bâtiments. Le chapitre 2 présente le cas d'étude mono-zone ainsi que les modèles "boîte grise" utilisés par la suite. Le chapitre 3 détaille les différents types de modèles implémentés ainsi que les méthodes d'estimation paramétrique liées. Il met tout d'abord en lumière les avantages et limites des modèles de type ARX (Auto Regressive model with eXternal inputs) et l'intérêt d'un apport de connaissance physique. Puis, il détaille l'approche par des modèles semi-physiques de type analogie électrique (ou modèles RC) et les deux méthodes d'estimation paramétrique implémentées. Ces différents modèles sont soumis à des tests d'acceptabilité et à un processus de sélection pour identifier la structure la plus adéquate à la mise en œuvre des services énergétiques. Pour finir, le chapitre 5 explore la capacité du modèle à s'adapter à des cas plus larges : absence de capteurs de contacts, modélisation d'un cas d'études multi-zones. Est également présentée une méthodologie de génération automatique du modèle à partir des seules informations accessibles à l'utilisateur final.

## 2 Sociologie de l'énergie et modèles pour les services énergétiques

### 2.1 Sociologie de l'énergie

Plus les bâtiments deviennent performants, plus l'activité humaine a un impact considérable sur la gestion énergétique des bâtiments. Les occupants ne peuvent donc de fait plus être considérés comme des perturbateurs du bâtiment mais doivent être impliqués dans leur gestion énergétique. Dujin et al. [29] expliquent que dans les

approches actuelles, les occupants sont vus comme des acteurs passifs du bâtiment devant tout apprendre et subissant les systèmes dont les rythmes et caractéristiques ne correspondent pas à leur besoins. Des études ont été menées dans les premiers bâtiments performants HQE (Haute Qualité Environnementale) et BBC Effinergie (Bâtiments Basse Consommation) (*cf.* [11] [19] and [84]). Les observations font remonter que dès lors que les systèmes installés sont mal appréhendés, la performance énergétique du bâtiment s'en trouve fortement dégradée. A partir de là, deux tendances se distinguent dans le domaine de la sociologie : l'approche par les comportements et l'approche par les pratiques, plus récente. Concernant l'approche par les comportements, Moser [69] estime que le changement de comportement vis à vis de la consommation énergétique peut être amenée de trois manières différentes: incitations, sensibilisation et éducation. En effet, différentes études ont porté sur les impacts d'un simple affichage temps réel de la consommation. Des analyses croisées de ces études ont été réalisées sur 26 études menées mondialement entre 1987 et 2008 et stipulent que les économies réalisées peuvent varier de 1.1% à 20% avec une moyenne entre 5 et 12%.

### 2.2 Services énergétiques

L'objectif de cette thèse est de développer un modèle thermique du bâtiment permettant de fournir des conseils et explications à l'utilisateur final. Les différents services envisagés sont listés ci-dessous :

- “What-if” : Quelles sont les conséquences de telle action?
- “Replay” : Que s'est-il passé hier? le mois dernier? l'année dernière?
- “Mirror” : Quel est l'état actuel de mon logement?
- “What-for” : Comment atteindre mon objectif?
- “Suggest-and-adjust” : Ajustement de plans anticipatifs par l'utilisateur
- “Explain” : Formalisation qualitative des phénomènes physiques
- “How-to” : Explication des usages prévus par les concepteurs

A chacun de ces services correspond ou ou plusieurs indicateurs pertinents, des échelles temporelles et des règles quand à la fréquence de sollicitation de l'utilisateur.

## 2.3 Etat de l'art sur les modèles

Afin de mettre en place les services énergétiques précédemment cités, il est nécessaire de faire un état de l'art des modèles actuellement utilisés en gestion énergétique. Ils se classent en trois grandes catégories : les modèles “boîte noire” basés exclusivement sur de la donnée, les modèles “boîte blanche” ou modèles de connaissance qui reposent sur les lois de la physique et enfin les modèles “boîte grise” s'appuyant sur quelques connaissances physiques mais dont certains paramètres sont appris à partir de données. Dans cette étude, les modèles “boîte blanche” ne seront pas étudiés car ils requièrent trop de connaissance experte en entrée pour être adaptés à notre cas.

Les modèles “boîte noire” ou modèles universels présentent l'avantage d'être facilement estimés et de ne se baser sur aucune connaissance experte. La plupart des modèles universels ont été développés pour modéliser des systèmes spécifiques du bâtiment ou des parois plutôt qu'un bâtiment dans son ensemble [23], [89]. Étendre ces modèles au système bâtiment serait très coûteux en nombre de variables. De plus, comme la plupart du temps, la description complète du bâtiment n'est pas accessible, ces modèles ne sont pas adaptés à nos besoins. Plus récemment, plusieurs auteurs ont modélisé un bâtiment complet avec différents objectifs: prédiction de la température intérieure [66], prédiction de la demande thermique [93] ou récupération des valeurs U et G du bâtiment [52]. Si différentes structures de modèles universels cohabitent, les modèles ARX (Auto-Regressive with eXogenous inputs) sont les plus répandus dans la communauté. Jimenez et Madsen [52] et Mustafaraj et al. [71] ont mis en évidence qu'ils permettaient de mettre en équation le bâtiment dans son ensemble. La formulation générale d'un modèle ARX à une sortie est décrite dans l'équation N.1.

$$T_{i,n} = - \sum_{l=1}^p a_l T_{i,n-l} + \sum_k \sum_{j=0}^q b_{k,j} u_{k,n-j} + e_n \quad (\text{N.1})$$

avec  $b_k$  les coefficients des entrées respectives  $u_k$  et  $e$  l'erreur qui sera assumée égale au bruit blanc.

En ce qui concerne les modèles “boîte grise”, il est tout d'abord important de noter que dans le domaine de la gestion énergétique des bâtiments, la plupart des modèles ont été pensés pour des phases de conception. Ils s'intéressent donc aux besoins annuels des bâtiments et se basent sur des scénarios types concernant

l'occupation ou les températures de consigne du chauffage. Ces scénarios sont bien souvent très loin des scénarios réels comme l'indique le "Bulletin Officiel n°201114". D'autre part, ces objectifs différents conduisent à une granularité et des hypothèses différentes. L'état de l'art a mis en évidence que les modèles les plus utilisés dans la communauté sont des modèles de type analogie électrique (ou modèles RC) qui offrent à la fois une bonne représentation des différents éléments du bâtiment ainsi qu'une grande facilité d'utilisation. Toutefois, les avis divergent quant à la modélisation de l'inertie, la prise en compte des apports solaires et les phénomènes négligés. Un des facteurs prédominants dans le choix du modèle va donc être l'objectif du modèle. La plupart des auteurs considèrent que l'inertie du bâtiment se situe dans les murs extérieurs et dans l'air. Toutefois, Hazyuk et al. [47] et Mathews et al. [65] prennent également en considération l'inertie de la dalle. Bacher et Madsen [9] ont mené une étude visant à comparer différents ordres de modèles. Ils ont généré 17 modèles allant de 2 à 5 capacités et concluent qu'à plus de 4 capacités, l'amélioration apportée en termes de précision ne dépasse pas les 5%.

Ceci étant dit, le choix du modèle et de la structure dépend également de la capacité ou non d'estimer ses paramètres. L'estimation paramétrique consiste à la calibration des paramètres du modèle afin que la sortie simulée du modèle corresponde à la sortie mesurée. S'enchaîne alors une phase de validation où le modèle estimé est lancé sur une période n'ayant pas servi à la calibration des paramètres. Cette étape permet de caractériser la capacité de prédiction du modèle.

### 2.4 Problématique

L'objectif général de cette thèse et la méthodologie mise en place sont résumés dans la figure N.1. Tout d'abord, les utilisateurs finaux décrivent la configuration de leur logement. Le système stocke alors les informations dans un fichier xml. Ce fichier servira de base à la génération automatique du modèle adapté à ce logement en particulier et lancera l'estimation paramétrique et la validation. Cette thèse s'intéresse plus précisément à la prédiction de l'évolution de la température d'air intérieur et de la concentration de CO<sub>2</sub>. Les services énergétiques envisagés requièrent une bonne précision dans la prédiction à plus 24 heures. L'accent est donc mis ici sur la sortie du modèle et non sur l'interprétation physique de la valeur des paramètres estimés. Différentes structures de modèles ont été développées et testées sur deux cas d'études distincts. Tout d'abord des modèles de type ARX puis des modèles RC. Pour ces derniers, deux méthodes d'estimation paramétrique ont été mis en oeuvre et com-

parées : une optimisation de descente et un algorithme génétique. Finalement, on s'intéresse au passage d'un cas d'étude mono-zone à un cas d'étude plus complexe incluant 4 pièces.

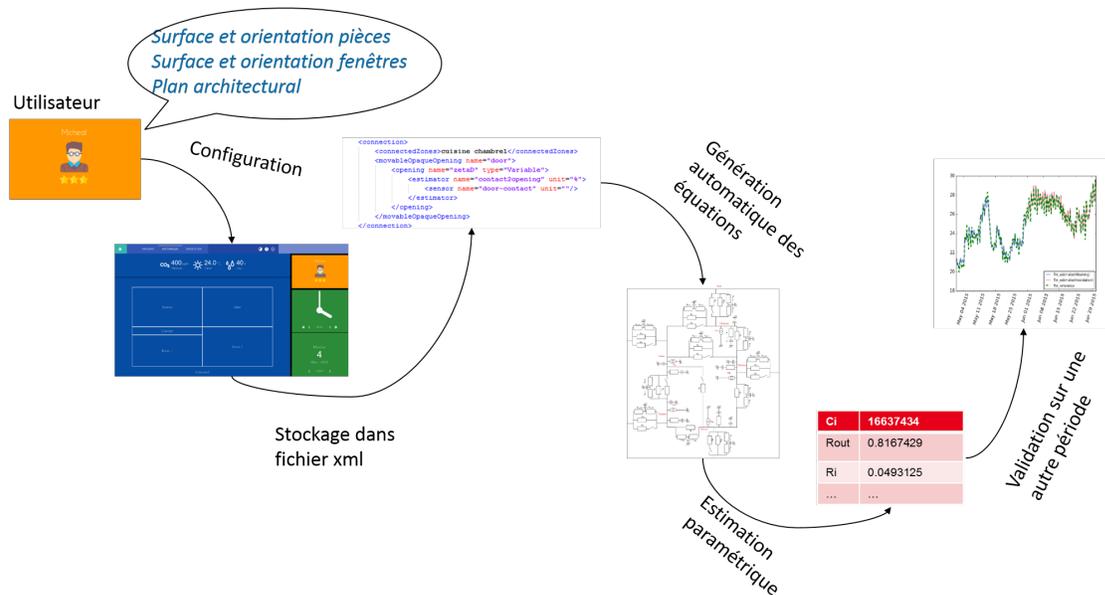


Figure N.1: Objectif général et méthodologie mise en place

## 3 Cas d'études et structures de modèles auto-configurables

### 3.1 Cas d'études mono-zone

Le premier cas d'études considéré est un bureau accueillant 1 à 4 personnes selon les périodes. Il est équipé de 26 capteurs dont : capteurs de température, contacts portes et fenêtres, détecteur de mouvement, éclairement, capteur d'humidité et station météo (cf. figure N.2). Afin de favoriser le déploiement de la solution, un set minimum de données de capteurs a été défini comme tel : température d'air intérieur, température de surface du radiateur, concentration de CO<sub>2</sub>, contacts fenêtres et portes ainsi que la nébulosité et la température extérieure.

Les apports solaires sont basés sur un modèle de décomposition. Ils ont été calculés à partir de la littérature en se basant sur la formule de Gate [40] pour le flux réfléchi et les travaux de Spencer [86] pour le flux diffus. D'autre part, des estimateurs sont définis pour prédire l'occupation et la puissance de chauffage car ces informations sont inaccessibles par la mesure. Pour cela, la puissance de chauffage

est définie par l'équation N.2 où  $K_{chauffage}$  est un paramètre appris. La présence quant à elle est estimée à partir de la consommation électrique des prises de chaque bureau.

$$P_{chauffage} = K_{chauffage}(T_{chauffage} - T_{int}) \quad (\text{N.2})$$

## 3.2 Blocs élémentaires et structures de modèles

Pour définir les structures pertinentes de modèles à appliquer, une ontologie de bâtiment a été définie. Il a donc été nécessaire de lister tous les composants du bâtiment impliqués dans la gestion de l'énergie ainsi que leurs caractéristiques vis à vis de l'énergie. Par la suite, chaque partie du bâtiment a été modélisée de différentes manières selon la prise en compte de l'inertie considérée (*cf.* tableaux N.1 to N.3). Ces blocs élémentaires seront ensuite assemblés pour définir les différentes structures de modèles qui seront implémentées dans cette thèse.

A partir de ces différents blocs, un premier modèle à 1 capacité a été créé (*cf.* figure N.3) puis la complexité du modèle a été progressivement augmentée jusqu'à obtenir un modèle à 4 capacités; précision maximale possible avec les données disponibles (*cf.* figure N.4).

$T_n$ ,  $T_{in}$  et  $T_{out}$  représentent respectivement la température de voisinage, d'air intérieur et d'air extérieur

$R_n$  et  $R_{out}$  représentent la résistance du mur entre la pièce et respectivement le voisinage et l'extérieur

$R_D$  et  $R_W$  représentent respectivement la résistance de la porte et de la fenêtre

$\zeta_D$  et  $\zeta_W$  représentent respectivement les ouvertures de portes et fenêtres

$R_i$  et  $C_i$  représentent respectivement la résistance du mur fictif et la capacité du mur fictif

$\phi_{in}$  représentent les gains intérieurs dus à l'occupation, aux équipements électriques, au chauffage et aux apports solaires.

$R_D$  et  $R_W$  sont définis ci-dessous :

$$R_D = \frac{1}{\rho_{air} c_{p,air} Q_D} \quad (\text{N.3})$$

$$R_W = \frac{1}{\rho_{air} c_{p,air} Q_W} \quad (\text{N.4})$$

Mur extérieur	
Bloc élémentaire	Équation
	$T_{in} - T_{out} = R\phi$
	$T_{in} - T_{out} = \left( \frac{1}{\frac{1}{R_{out}} + \frac{\zeta_W}{R_W}} \right) \phi$
	$\frac{dT_{w,out}}{dt} = \frac{1}{C_{w,out}} \left[ \frac{T_{out}}{R_{w,out2}} + \frac{T_{in}}{R_{w,out1}} - T_{w,out} \left( \frac{1}{R_{w,out1}} + \frac{1}{R_{w,out2}} \right) \right]$ $T_{in} = \frac{R_{eq}}{R_{w,out1}} T_{w,out} + R_{eq} \left( \frac{1}{R_{out}} + \frac{\zeta_W}{R_W} \right) T_{out}$ $\frac{1}{R_{eq}} = \frac{1}{R_{w,out1}} + \frac{1}{R_{out}} + \frac{\zeta_W}{R_W}$

où  $\phi$  représente le flux entre  $T_{in}$  et  $T_{out}$

**Table N.1:** Bloc élémentaire pour le mur extérieur

A ces modèles est couplé un modèle aéraulique pour modéliser l'évolution de la concentration de CO<sub>2</sub> présenté en figure N.5. La mise en équation est détaillée dans l'équation N.5.

$$V \frac{d\Gamma_{in}}{dt} = -(Q_W + Q_D)\Gamma_{in} + Q_W)\Gamma_{out} + Q_D)\Gamma_n + S_{CO_2}^{body}n \quad (N.5)$$

où  $\Gamma_{in}$ ,  $\Gamma_{out}$  et  $\Gamma_n$  représentent les concentrations en CO<sub>2</sub> de l'intérieur, de l'extérieur et du voisinage,  $Q_W$  et  $Q_D$  les flux d'air via respectivement la fenêtre et la porte,  $S_{CO_2}^{body}$  le volume de CO<sub>2</sub> expiré par une personne et  $n$  le nombre d'occupants dans le bureau.

Mur intérieur	
	$T_{in} - T_n = R\phi$
	$T_{in} - T_n = \left( \frac{1}{\frac{1}{R_n} + \frac{\zeta_D}{R_D}} \right) \phi$
	$\frac{dT_{w,n}}{dt} = \frac{1}{C_{w,n}} \left[ \frac{T_n}{R_{w,n2}} + \frac{T_{in}}{R_{w,n1}} - T_{w,n} \left( \frac{1}{R_{w,n1}} + \frac{1}{R_{w,n2}} \right) \right]$ $T_{in} = \frac{R_{eq}}{R_{w,n1}} T_{w,n} + R_{eq} \left( \frac{1}{R_n} + \frac{\zeta_W}{R_W} \right) T_n$ $\frac{1}{R_{eq}} = \frac{1}{R_{w,n1}} + \frac{1}{R_n} + \frac{\zeta_W}{R_W}$

Table N.2: Bloc élémentaire pour le mur intérieur

Zone			
Bloc élémentaire	Equation	Bloc élémentaire	Equation
	$\frac{d\tau}{dt} = \frac{1}{C_i} \left( \frac{T_{in}}{R_i} - \frac{\tau}{R_i} \right)$		$\frac{d\tau}{dt} = \frac{1}{C_i} \left( \frac{T_{in}}{R_i} - \frac{\tau}{R_i} \right)$ $\frac{dT_{in}}{dt} = \frac{1}{C_{air}} \phi_{in}$

Table N.3: Bloc élémentaire pour une zone

### 3.3 Méthodes de sélection et d'acceptabilité des modèles

Une fois les modèles développés, une méthode d'acceptabilité et de sélection ont été développées afin de garantir une précision minimum et de définir des critères de comparaison entre les modèles. Le critère d'acceptabilité repose sur un seuil de précision minimum basé sur la sRMSE ("Root Mean Square Error" standardisée) ainsi que sur un devoir de stabilité au long des saisons. En effet, il est essentiel de pouvoir garantir les mêmes performances en hiver et en été malgré le fait que le comportement du bâtiment soit substantiellement différent. Les modèles qui respecteront ces critères seront ensuite comparés en termes de précision en se basant à sur la somme de la sRMSE en été et en hiver.

Pour finir, le modèle sera soumis à une analyse de sensibilité afin de s'assurer que chaque paramètre structurel a effectivement un impact sur la sortie et que la structure ne peut donc pas être simplifiée. En effet, il est important de minimiser au maximum le nombre de paramètres du modèle afin de permettre le déploiement sur des cas d'études plus complexes puisque les estimations paramétriques peuvent devenir très coûteuses en temps de calcul. De nombreuses méthodes d'estimation paramétrique existent. Dans le domaine de la gestion énergétique des bâtiments, la méthode de Morris est la plus couramment utilisée [68] du fait de son faible coût en temps de calcul. Malheureusement, cette méthode ne donne qu'une réponse qualitative sur l'impact des paramètres et non pas quantitative. C'est donc la méthode de Sobol qui sera utilisée dans ces travaux [85]. Il s'agit d'une méthode d'analyse de sensibilité globale indépendante du modèle basée sur une décomposition de variance. Elle détermine la contribution de chaque paramètre d'entrée et leurs interactions sur la variance de sortie du modèle. Puis, à partir de la décomposition de la variance totale, les indices de Sobol sont calculés comme indiqué dans l'équation N.6. Les indices de Sobol sont normalisés et donc facilement interprétables.

$$S_i = \frac{\text{Var}[\mathbb{E}(Y | x_i)]}{\text{Var}(Y)} = \frac{V_i(Y)}{\text{Var}(Y)}, \quad S_{ij} = \frac{V_{ij}(Y)}{\text{Var}(Y)}, \quad S_{ijk} = \frac{V_{ijk}(Y)}{\text{Var}(Y)}, \quad \dots \quad (\text{N.6})$$

où:  $V_i(Y) = \text{Var}[\mathbb{E}(Y | x_i)]$ ,  $V_{ij}(Y) = \text{Var}[\mathbb{E}(Y | x_i x_j)] - V_i(Y) - V_j(Y)$  et ainsi de suite.

## 4 Apprentissage de modèles à l'horizon 24 heures

Comme l'objectif ne nécessite pas de donner un sens physique aux paramètres du modèle, il est logique d'implémenter dans un premier temps des modèles de type "boîte noire". En effet, ces modèles présentent le grand avantage de ne se baser que sur des données et d'être très peu coûteux en temps de calcul. Leur application n'est pas commune en énergétique du bâtiment mais comme expliqué précédemment ils ont gagné en popularité ces dernières années. A l'opposé, les modèles semi-physiques sont très répandus dans la communauté de la physique du bâtiment. Toutefois, ils nécessitent une étape d'apprentissage des paramètres puisque l'on n'a pas accès aux caractéristiques intrinsèques du bâtiment. Cette étape est réalisée dans cette thèse à partir d'informations capteurs présents dans les deux cas d'études considérés. Différentes solutions existent pour estimer les paramètres d'un modèle. Certaines demandent une expertise pour déterminer un point initial à partir duquel le gradient de la fonction objectif est étudié pour se diriger vers le minimum. Ces optimisations peuvent par construction aboutir à des optimum locaux. D'autres méthodes appelées méta-heuristiques peuvent pallier à ce problème mais sont significativement plus coûteuses en temps de calcul. Ces différentes options seront étudiées dans cette section après avoir rappelé les pré-requis imposés par la mise en place de services énergétiques.

### 4.1 Pré-requis pour les services énergétiques

L'objectif final est de permettre l'implémentation de services énergétiques à destination de l'utilisateur final pour l'aider à comprendre l'impact de ses actions. Si les données sont essentielles à l'apprentissage des modèles afin de pallier au manque de données expertes, l'instrumentation se doit toutefois de rester minimale et peu intrusive si l'on veut pouvoir déployer largement la solution. D'autre part, les modèles doivent pouvoir se calibrer de manière automatique sans nécessiter l'intervention d'un expert à chaque déploiement. La plupart des services ont un horizon 24 heures avec un pas de temps horaire. D'autre part, la quantité de données nécessaire à l'apprentissage du modèle doit rester dans le domaine de l'acceptable car on ne peut pas attendre d'un utilisateur final qu'il installe le système s'il ne peut l'utiliser que six mois plus tard. Enfin, le modèle doit être fiable. En effet, si d'un jour à l'autre, la précision de la prédiction ne peut être garantie, l'utilisateur perdra très vite confiance en ces services et cessera de les utiliser.

## 4.2 Modèles ARX

Dans un premier temps, les modèles ARX ont été appliqués au cas d'étude mono-zone. Les périodes d'entraînement et de validation sont définies puis les structures de modèles identifiées. Pour le scénario hiver, le modèle est appris sur le mois d'Octobre et validé sur le mois de Novembre et pour le scénario d'été le modèle est appris sur le mois de Mai et validé sur le mois de Juin. Deux modèles ARX sont donc définis : un pour la température et un pour la concentration en CO<sub>2</sub> comme défini dans les équations N.7 et N.8.

$$\widehat{T_{in}(t)} = b_0 T_{out}(t) + b_1 T_{corridor}(t) + b_2 \zeta_D(t) + b_3 \zeta_W(t) + b_4 \phi_{solar}(t) + b_5 P_{elec}(t) \quad (N.7)$$

$$\widehat{\Gamma_{in}(t)} = b_0 \Gamma_{corridor}(t) + b_1 \zeta_D(t) + b_2 \zeta_W(t) + b_3 S_{CO_2}^{body} n(t) \quad (N.8)$$

### 4.2.1 Résultats

Les résultats de ces 2 modèles appliqués au cas d'étude mono-zone peuvent être visualisés dans la figure N.6. On peut constater que si les résultats en termes de température présentent une bonne précision, il n'en va pas de même pour l'estimation de la concentration en CO<sub>2</sub>. Les valeurs prédites sont non seulement très éloignées des mesures mais plus grave n'ont pas de sens physiquement puisque la concentration extérieure en CO<sub>2</sub> (c'est à dire la plus faible) est aux alentours de 400ppm. Pour comprendre les causes de ce comportement, la même procédure a été relancée cette fois sur les mois d'Avril et Mai et sur le scénario d'hiver (*cf.* figure N.7). On peut cette fois constater que les prédictions en CO<sub>2</sub> sont meilleures.

Par la suite, la fiabilité du modèle tout au long de l'année a été évaluée et s'est révélée problématique. Si les dynamiques générales se retrouvent bien et les erreurs moyennes en valeur absolue ne sont pas trop élevées, le modèle semble instable selon les mois. Le cas de Juin particulièrement pose problème. Les performances à la fois en termes de température et de CO<sub>2</sub> sont fortement dégradées.

### 4.2.2 Le cas de Juin

Afin de comprendre les raisons des performances dégradées sur le mois de Juin, les variables environnementales ont été observées. Cette analyse a révélé que le

mois de Juin se différencie de manière significative par rapport au mois précédent sur plusieurs aspects : le nombre d'ouvertures de fenêtres a fortement augmenté de même que les températures intérieures et extérieures. L'évolution de la concentration de CO<sub>2</sub> étant très sensible aux ouvertures de fenêtres, elles peuvent expliquer les performances dégradées de la prédiction.

Pour pallier à ce problème, une nouvelle structure de modèle ARX a été mise en place. L'origine du problème pouvant être physiquement expliquée, l'hypothèse est posée qu'une insertion de signification physique dans la structure du modèle pourrait améliorer ses performances. Cette hypothèse a déjà été étudiée par Wu and Sun [92] qui se sont inspirés des équations de la thermodynamique pour enrichir des modèles ARMAX. Dans cette thèse, se sont les équations du modèle RC à une capacité ainsi que le modèle aéralique qui ont servi de base pour identifier les variables d'intérêt à prendre en compte dans la nouvelle structure. Les équations physiques ont donc été décomposées pour mettre en évidence toutes les variables d'entrée intervenant ainsi que leurs interactions afin de s'en servir pour déterminer les différents termes des modèles ARX. Cela a donc mené à l'élaboration des modèles ARX inspirés de la physique détaillés dans les équations suivantes :

$$\Gamma_{in} = b_0\Gamma_{corridor} + b_1\zeta_D + b_2\zeta_W + b_3n \quad (\text{N.9})$$

$$T_{in} = b_0T_{out} + b_1T_{corridor} + b_2\zeta_D + b_3\zeta_W + b_4n \quad (\text{N.10})$$

Les résultats obtenus avec ces modèles inspirés de la physique peuvent être visualisés qualitativement sur la figure N.8 et quantitativement dans le tableau N.4 grâce au calcul de l'erreur absolue moyenne et de l'écart type. Cela illustre la capacité de cette nouvelle structure à améliorer la prédiction du modèle et donc la confiance que l'utilisateur peut avoir dans l'estimation et les services énergétiques.

Modèle	Erreur absolue moyenne	Écart-type
ARX standard	57.2248	78.3289
ARX inspiré de la physique	48.9711	66.0473

**Table N.4:** Comparaison de l'erreur absolue moyenne et de l'écart-type pour le modèle ARX classique et celui inspiré de la physique pendant la phase de validation

### 4.2.3 Limites

Les modèles ARX ont démontré de bonnes performances quant à la prédiction de la température et pour une moindre mesure de la concentration en CO<sub>2</sub>. Toutefois, plusieurs limites ont été observées. En effet, le cas du mois de Juin met en lumière la nécessité pour le modèle de disposer d'un jeu de données d'apprentissage riche, c'est-à-dire incluant la plupart des phénomènes que le modèle est amené à rencontrer dans la phase de validation. La capacité du modèle à être généralisé à d'autres cas d'études nécessite d'être plus précisément étudiée, et particulièrement sa capacité à se passer de certaines données capteurs.

## 4.3 Méta-optimisation

L'approche choisie ici a été développée par Audrey Le Mounier dans sa thèse [59] dont l'objectif est d'explorer le domaine de recherche et de garantir la cohérence physique des paramètres. Elle se base sur un algorithme de descente de type SQP (Sequential Quadratic Programming). Le principe de la méthode est résumé dans la figure N.9.

### 4.3.1 Application aux différentes structures

Cette méthode nécessite d'être initialisée. Des valeurs ont donc été définies pour les différentes capacités, résistances et estimateurs. Disposant d'informations sur la structure du bâtiment, certains paramètres ont été initialisés avec des valeurs initiales réalistes alors que d'autres ont été initialisés avec des valeurs du bon ordre de grandeur. L'application au modèle Référence a démontré des résultats intéressants avec une sRMSE variant de 0.06 à 0.09 selon les saisons. Cela reste donc dans les critères d'acceptabilité.

Toutefois, lors de la modification de l'inertie et donc de la montée en complexité du modèle, des phénomènes de divergence ont été observés. On peut observer sur la figure N.10, les profils de températures obtenus pour les différents modèles. On ne peut que constater que les modèles à 2 capacités présentent des erreurs conséquentes et ne retrouvent pas les dynamiques du modèle.

Une étude plus poussée pour comprendre l'origine du problème a été menée. L'évolution de la valeur des différents paramètres au cours des itérations a été ob-

servée. Cela a révélé un problème de convergence avec des paramètres présentant des itérations incessantes entre leurs bornes minimum et maximum.

### 4.3.2 Limites

Ce problème de convergence est associé à une augmentation exponentielle du temps de calcul avec un temps requis de 300 secondes pour le modèle Référence et un temps de 7240 secondes pour le modèle à 2 capacités. Une telle augmentation n'est pas acceptable car cela remet en question l'extension du modèle à des cas d'études plus complexes.

D'autre part, une étude de l'ergodicité du modèle a été faite pour s'assurer que le résultat de l'estimation paramétrique garantissait toujours les mêmes performances. Pour cela, 406 simulations du modèle Référence ont été lancées à partir de points initiaux différents et l'erreur a été récupérée. La figure N.11 illustre l'impossibilité de garantir une quelconque performance de l'estimation finale avec cette méthode puisque l'erreur obtenue varie fortement et au-delà des valeurs acceptables puisque près de 25% des simulations ont une erreur supérieure à 2°C.

## 4.4 Optimisation génétique

Pour remédier aux problèmes de convergence et de sensibilité au point initial, une nouvelle méthode a été développée, basée sur un algorithme génétique. De nombreux algorithmes de ce type existent mais le plus courant et performant est l'algorithme NSGA-II (Non Dominated Sorting Genetic Algorithm), algorithme élitiste introduit par Deb et al. [25] dans les années 2000. Dans le domaine de l'énergétique du bâtiment, il a été utilisé notamment par Mozer [70] et Ghisi et Tinker [41].

### 4.4.1 Principe et configuration

Le principe de base de cet algorithme repose sur quatre grandes étapes : génération aléatoire d'une population initiale, sélection des individus, opérations de mutation et croisement entre individus et évaluation des individus. L'intérêt de l'algorithme NSGA-II réside dans le fait que les individus sont triés en fonction de la distance de "crowding" afin de préserver une diversité lors de la création de la génération suivante.

Contrairement à un algorithme de descente comme celui développé précédemment, l'algorithme ne s'arrête pas lorsqu'il a atteint son critère de convergence. Le point d'arrêt est ici défini par l'expérimentateur qui doit décider du nombre de générations. Des tests ont donc été menés pour étudier l'évolution de l'erreur au cours des générations avec différentes tailles de population. Finalement, pour les modèles mono-zone l'algorithme sera configuré avec 100 générations de 100 individus.

#### 4.4.2 Ergodicité

Au vu des résultats obtenus précédemment, une analyse de l'ergodicité de l'estimation paramétrique est menée d'emblée. Les résultats peuvent être visualisés sur la figure N.12 et confirme la constance de l'estimation pour 50 simulations.

#### 4.4.3 Application aux différentes structures

Une fois la phase de configuration effectuée, l'algorithme NSGA-II a été appliqué aux différentes structures de modèles semi-physiques. Pour le scénario hiver, le modèle est appris sur le mois d'Octobre et validé sur le mois de Novembre et pour le scénario d'été le modèle est appris sur le mois de Mai et validé sur le mois de Juin. Les résultats en termes de sRMSE pour les deux scénarios sont représentés sur la figure N.13. A partir de là on peut appliquer les méthodes d'acceptabilité et de sélection détaillées précédemment. On peut donc constater que le Model2CairTout, le modèle Référence et le Model0C ne respectent pas le seuil de précision fixé à 0.1. Le Model3CTout lui ne respecte pas la condition de stabilité entre les saisons. Les autres modèles respectent les critères de validité et peuvent donc passer à la phase de sélection.

Pour cela, les valeurs de sRMSE en validation des scénarios hiver et été sont sommées et le modèle sélectionné correspond à celui qui affiche la meilleure précision : le Model4C.

#### 4.4.4 Analyse de sensibilité

Une fois la structure de modèle la plus adéquate identifiée, une analyse de sensibilité est lancée pour s'assurer que chacun des paramètres structurels a un réel impact sur la sortie. Comme expliqué précédemment, c'est l'analyse de Sobol qui est choisie. La difficulté ici est que l'on cherche à identifier l'impact des paramètres sur

la sortie en partant de l'hypothèse que la structure a été bien identifiée et présente de bonnes performances. Il s'agit donc de bien définir les intervalles de variation des paramètres structurels sur lesquels va être lancée l'analyse de sensibilité. On lance donc dans un premier temps une analyse de Sobol avec des bornes très larges pour un grand nombre d'échantillons. Cela génère ainsi 240 000 simulations pour lesquelles on récupère à la fois le résidu calculé et les valeurs prises par les différents paramètres et on observe leur évolution (*cf.* figure N.14). Cela permet d'identifier que le paramètre  $R_i$  a une forte influence sur le résidu et donc de restreindre ses bornes. L'analyse de sensibilité de Sobol peut ensuite être lancée avec  $R_i$  compris entre  $[0.05, 0.1]$ .

L'analyse de Sobol relancée sur ces intervalles ne renvoie malheureusement pas de résultat utilisable puisque seuls les paramètres  $R_i$ ,  $C_i$  et  $R_n$  sont détectés comme impactant, ce qui n'a pas de sens au vu du contexte. En conséquence, la structure du modèle n'est pas simplifiée.

#### 4.4.5 Observabilité du modèle

La dernière validation mise en place a consisté à s'assurer de l'observabilité du modèle, élément essentiel à la gestion énergétique. En effet, dans les simulations précédentes, l'objectif était d'identifier le modèle et donc le modèle était lancé sur un mois. Dans ce contexte, l'initialisation n'a alors que peu d'impact sur les performances du modèle.

Toutefois, lorsqu'il s'agit de fournir du conseil et de l'explication à l'utilisateur, le problème est inversé et il s'agit alors d'optimiser des actions selon les objectifs de l'utilisateur. Le nombre de variables à optimiser augmente donc fortement puisqu'il faut considérer la variation de la valeur de ces actions pour chaque pas de temps. De plus, cela n'aurait pas de sens de conseiller l'utilisateur sur les actions à mettre en oeuvre la semaine suivante. Or, sur une simulation à horizon 24 heures, l'initialisation prend alors toute son importance.

Dans le cas du Model4C, deux états sur quatre seulement sont observables. Il a donc fallu quantifier l'impact des deux états non observables pour évaluer si oui ou non il s'agissait d'un frein. Pour cela, l'observateur d'état a été exprimé dans une nouvelle base choisie grâce à une matrice de passage orthogonale et les composantes de cet observateur d'état ont été tracées. Comme on peut le voir sur la figure N.15,

si les deux composantes  $x_3$  et  $x_4$  sont plus faibles que les deux premières, on ne peut pas pour autant les considérer comme négligeables.

Pour compléter cette analyse, l'observateur a donc été vérifié en pratique sur une simulation de 24h le 1er Novembre 2017. La température initiale a été volontairement mal initialisée et on a fixé le recalage via l'observateur à 6h du matin. La figure N.16 montre que l'observateur est effectivement capable de recalculer avec précision la température.

## 4.5 Conclusion

Dans ce chapitre, différentes topologies de modèles et différentes structures ont été implémentées. Si les modèles boîte noire de type ARX ont montré de bonnes précisions, ils sont toutefois dans l'incapacité de garantir une performance comme l'a démontré leur application au mois de juin. Par la suite, ce sont donc des modèles boîte grise de type analogie électrique qui ont été implémentés. Ces modèles nécessitent une phase d'apprentissage plus complexe que ne le sont les régressions linéaires notamment à cause du coût en temps de calcul. Dans un premier temps, c'est donc une estimation paramétrique basée sur un algorithme de descente qui a été implémentée pour son faible coût en appel de fonction objectif. Toutefois, cette méthode s'est révélée non robuste et trop sensible au point d'initialisation des paramètres. Pour finir, il a été fait appel à un algorithme génétique. Il a été appliqué aux différentes structures de complexité croissante présentées ci-dessus. Les méthodes d'acceptabilité et de sélection ont été appliquées afin de garantir une précision minimum ainsi qu'une stabilité au long des saisons. A partir de là, la structure de modèle sélectionnée a été le Model4C; structure la plus complexe. Pour finir, une analyse de sensibilité a été menée afin d'évaluer si certains groupes ou paramètres pouvaient être simplifiés mais cela n'a pas amené à simplifier la structure.

# 5 De la description du bâtiment aux services énergétiques

## 5.1 Objectifs

L'objectif de cette section est d'étudier la capacité de la structure sélectionnée à être déployée à plus grande échelle sur un cas d'étude plus complexe. Le cas d'étude

multi-zones implique une augmentation de la complexité du modèle, à la fois de la mise en équation et de l'estimation paramétrique.

## 5.2 Cas d'études multi-zones

Le cas d'étude considéré dans cette section est un appartement comportant 3 chambres et habité par un couple avec bébé. L'instrumentation est ici bien plus sommaire que dans le cas d'étude précédent et se compose de capteurs grands publics (stations Netatmo). Chaque station fournit la mesure de la température d'air intérieur, de l'humidité et de la concentration en CO<sub>2</sub>. La station présente dans le séjour fournit également une mesure du niveau sonore. Le tout est complété par des prévisions météorologiques. Le plan architectural du logement ainsi que le positionnement des capteurs peut être visualisé sur la figure [N.17](#).

Du fait du changement de fonction du cas d'études, certains estimateurs nécessitent d'être adaptés. En effet, si dans un bureau la consommation électrique des ordinateurs est un bon indicateur de l'occupation, ce n'est plus vrai dans le cas d'un logement. L'occupation est donc estimée à partir d'un seuil de CO<sub>2</sub> dans les chambres et d'un seuil de niveau sonore dans le salon.

## 5.3 Modélisation des ouvertures de fenêtres

Une information importante jusqu'ici disponible et qu'on ne retrouve pas dans ce cas d'études est l'ouverture des portes et fenêtres. En effet, l'ouverture de portes et fenêtres est une action qui impacte fortement à la fois la concentration en CO<sub>2</sub> et la température intérieure en fonction des conditions extérieures.

Deux solutions ont été mises en place, la première a consisté à simplement ajouter les paramètres de taux d'ouverture dans les paramètres à estimer. Les performances du modèle obtenue sont restées constantes mais cela enlève toute possibilité de pouvoir développer des services énergétiques conseillant sur les actions à avoir sur les portes et fenêtres. Cela n'est donc pas une solution très propice au déploiement de services pertinents pour l'utilisateur final. Estimer des profils d'ouverture de portes et de fenêtres dépendant du temps n'était pas non plus envisageable d'un point de vue de l'estimation paramétrique puisque cela impliquerait d'augmenter de manière considérable le nombre de paramètres à estimer. Afin de trouver un entre-deux, un modèle intuitif a été développé comme présenté dans les équations [N.11](#) et [N.12](#). Il

s'agit d'un modèle très simplifié mais limitant le nombre de paramètres à estimer tout en permettant de pouvoir influencer sur les ouvertures de portes et fenêtres et ainsi pouvoir fournir du conseil.

$$\zeta_D = \frac{\exp(a_D 0 + b_D * (T_{in} - T_n))}{1 + \exp(a_D 0 + b_D * (T_{in} - T_n))} \quad (\text{N.11})$$

$$\zeta_W = \frac{\exp(a_W 0 + b_W * (T_{in} - T_{out}))}{1 + \exp(a_W 0 + b_W * (T_{in} - T_{out}))} \quad (\text{N.12})$$

## 5.4 Génération automatique du modèle

La structure de modèle RC identifiée est alors confrontée aux contraintes d'un cas d'études multi-zones. Le premier défi consiste à identifier les équations correspondant au modèle global de l'appartement. En vue d'un déploiement à plus grande échelle et parce que la mise en équation manuelle est très accidentogène pour des modèles aussi complexes, une procédure de mise en équation automatique est mise en place. Elle s'organise en plusieurs étapes :

- Récupérer les informations nécessaires à la configuration des utilisateurs finaux
  - Nombre et orientation des pièces
  - Connexions entre pièces
  - Surface et orientation de chaque pièce
- Stockage de l'information dans un fichier xml basé sur l'ontologie d'un appartement
- Génération des équations
- Estimation paramétrique

Cette dernière étape a tout d'abord généré des problématiques des temps de calcul puisqu'il a été nécessaire de passer par du calcul symbolique. Toutefois, une fois ces difficultés contournées il a été possible de définir pour chaque connexion et pièce identifiées les équations correspondantes et de générer automatiquement les équations du système d'état représentant la mise en équation du système global. Le modèle RC de l'appartement complet est représenté dans la figure [N.18](#).

## 5.5 Implémentation

Le passage à un cas d'étude multi-zones a généré une forte augmentation de la complexité du problème d'estimation paramétrique du fait de l'augmentation du nombre de paramètres à estimer. De ce fait, il est nécessaire d'adapter la configuration de l'algorithme génétique en augmentant le nombre de générations et d'individus. L'estimation a donc été lancée pour 500 générations de 300 individus. Les résultats présentés par les figures N.19 et N.20 révèlent toutefois que cela ne suffit pas pour atteindre la convergence. En effet, si les estimations en terme de température semblent bonnes même si moins précises que pour le bureau, les résultats obtenus pour la concentration en CO<sub>2</sub> sont eux loin de représenter l'évolution réelle.

Afin de savoir si ces résultats sont dus à un problème de convergence, d'autres simulations ont été lancées. Le plus haut critère d'arrêt a été fixé à 1200 générations de 800 individus et l'évolution de l'erreur tout au long des générations a été tracée (*cf.* figure IV.20). On observe bien une diminution de l'erreur tout au long des générations mais à partir de la 400<sup>ième</sup> génération, l'erreur semble atteindre un plateau. Il ne s'agit donc pas d'un problème de convergence.

## 5.6 Application aux services énergétiques

Dans cette section est présentée une partie des travaux de thèse d'Amr Azhour Alyafi [5], prenant part également au projet de recherche INVOLVED. Son objectif se situe en aval de mes travaux et consiste à générer de l'explication pour l'utilisateur final. Il s'intéresse donc à générer une explication contextualisée à l'utilisateur final pour lui expliquer l'impact de ses actions sur le bâtiment et plus particulièrement sur son confort et sa consommation énergétique. Ces travaux sont appliqués au cas d'études mono-zone présenté dans la section 2. L'objectif est de retrouver les relations de cause à effets où les causes sont soit les actions de l'utilisateur soit les variables environnementales et les effets sont catégorisés entre les effets ressentis directement par l'utilisateur (inconfort thermique ou dû à la qualité de l'air) et les effets intermédiaires. Tout cela est résumé dans la figure N.22.

Pour cela, il se sert dans un premier temps de la structure de modèle sélectionnée pour déterminer le jeu d'actions optimales à réaliser afin d'atteindre les objectifs définis par l'utilisateur.

Par la suite, il s'attache à définir la distance entre deux scénarios: le scénario passé et le scénario optimal pour expliquer un comportement passé par exemple et le poids des différentes causes. Cela lui permet alors de construire le tableau représenté figure N.23.

Toutefois, ce tableau reste encore difficilement appréhendable par l'utilisateur final. La dernière étape consiste alors à formaliser ce tableau sous forme de phrases le mieux construite possible et le tout de manière automatique. Un premier modèle a donc été développé dans ARIANE-HELOISE [13] pour une étude de faisabilité : le modèle GRA-FRA ("GRAPhe vers texte en FRANçais"). Un exemple de résultat obtenu peut être observé figure N.24.

## 6 Conclusions et perspectives



Ce travail de thèse a consisté en l'implémentation d'une nouvelle méthodologie pour sélectionner une structure de modèle thermique et aéraulique adéquate pour les services énergétiques. L'objectif a été de déterminer une structure de modèle capable de prédire l'évolution de la concentration en CO<sub>2</sub> et de la température avec un horizon 24 heures au pas de temps horaire. Ces spécifications induisent différentes contraintes concernant la structure de modèle et sa configuration. La première étape a consisté en un état de l'art dans différents domaines. L'étude de la littérature concernant la sociologie de l'énergie a permis d'identifier les services pertinents. Puis, une étude plus classique des modèles utilisés pour la gestion énergétique des bâtiments a été faite pour identifier les types de modèles pertinents et élaborer la gamme de modèles testés.

Ainsi, des modèles de type boîte noire comme les modèles ARX ont été développés et implémentés. Si leurs performances en termes de précision étaient globalement bonnes, il est toutefois apparu qu'il n'était pas possible de garantir une précision tout au long de l'année. En effet, on a pu voir avec une application sur le mois de Juin que le modèle n'arrivait pas à délivrer une bonne performance. Par la suite, ce sont donc des modèles de type RC (analogie électrique) qui ont été choisis. Ces modèles présentant de plus grands défis concernant l'estimation paramétrique, deux méthodes ont été comparées dans un objectif de garder des temps d'apprentissage raisonnables. C'est l'algorithme génétique qui l'a finalement remporté car l'algorithme de descente auparavant testé s'est avéré trop sensible aux conditions initiales et donc non ergodique.

Pour finir, la capacité du modèle a s'adapter à d'autres conditions a été explorée et plus particulièrement le changement d'échelle. Pour cela, un nouveau cas d'études a été utilisé : un appartement de quatre pièces. Afin de s'affranchir des erreurs de mise en équation relatives à la montée en complexité du modèle, un mécanisme d'automatisation a été développé. De l'entrée des informations accessibles à l'utilisateur concernant son logement (surface et nombres d'ouvertures ou plan architectural), le système est ensuite capable de générer automatiquement les équations correspondantes et de lancer l'estimation paramétrique. L'application à un cas d'études plus large a toutefois soulevé des limites notamment sur la prédiction de l'évolution de la concentration en  $\text{CO}_2$ . Cela est probablement dû également à un manque de capteurs.

Ce travail de recherche appelle des extensions. Concernant la faible précision des résultats obtenus pour l'estimation de la concentration en  $\text{CO}_2$  du cas d'études multi-zones, une première hypothèse est d'en imputer la responsabilité au manque de capteurs. Il serait donc intéressant d'appliquer le modèle à un autre cas d'études multi-zones mais disposant de plus de capteurs et particulièrement de contacts de portes et de fenêtres. Cela permettrait alors de valider ou non cette hypothèse.

Dans la même lignée, il est évident qu'il existe un lien entre le nombre de capteurs et la précision du modèle. En effet, plus on a de mesures moins il est nécessaire de recourir à des estimateurs. Par exemple, estimer avec précision ou même mesurer des informations telles que la puissance effective de chauffage ou l'occupation permettrait alors d'augmenter de manière significative la précision du modèle. Ainsi, explorer l'évolution de la précision du modèle en fonction de l'instrumentation disponible apporterait une information précieuse.

Des méthodes de clustering pourraient être utilisées pour classer les jours en différents clusters de jours similaires. En effet, un défi de l'estimation paramétrique est la quantité de donnée nécessaire à l'apprentissage. Un processus de clustering pourrait renvoyer une information qualifiant la similarité du jour considéré par rapport aux jours précédents. Si le système a déjà rencontré des jours similaires, il serait alors possible d'estimer le modèle sur une période plus courte. Dans cette thèse, il a été montré qu'un modèle estimé en hiver ou en été était capable de prédire l'évolution de la température tout au long de l'année. Toutefois, ces méthodes de clustering pourraient améliorer la prédiction en générant plusieurs jeux de paramètres selon les clusters.

La méthodologie de validation proposée s'est concentrée sur l'erreur d'estimation entre la température et le CO<sub>2</sub> mais les services énergétiques pour l'utilisateur utilisent parfois l'estimation des besoins en énergie. Ainsi, la validation au niveau du service devrait être menée: peut-être qu'avec des estimations non précises, les services pourraient mener à de bons résultats. Un premier exemple de service énergétique a été présentée via le travail de Amr Alzhouri Alyafi mais plusieurs restent à développer. Les pré-requis des différents services énergétiques variant fortement, il serait pertinent de s'assurer que la structure de modèle est robuste à l'ensemble des services envisagés.

Un autre défi de la gestion énergétique est la capacité à détecter des modifications dans le comportement thermique du bâtiment. En effet, les bureaux ou logements évoluent au long du temps et certains systèmes peuvent être ajoutés ou remplacés. Dans ce cas là, le modèle appris n'est alors plus adapté à la situation. Déployer une détection automatique de tels phénomènes et en identifier les causes pourrait alors améliorer la robustesse du modèle et garantir sa performance en toutes circonstances.

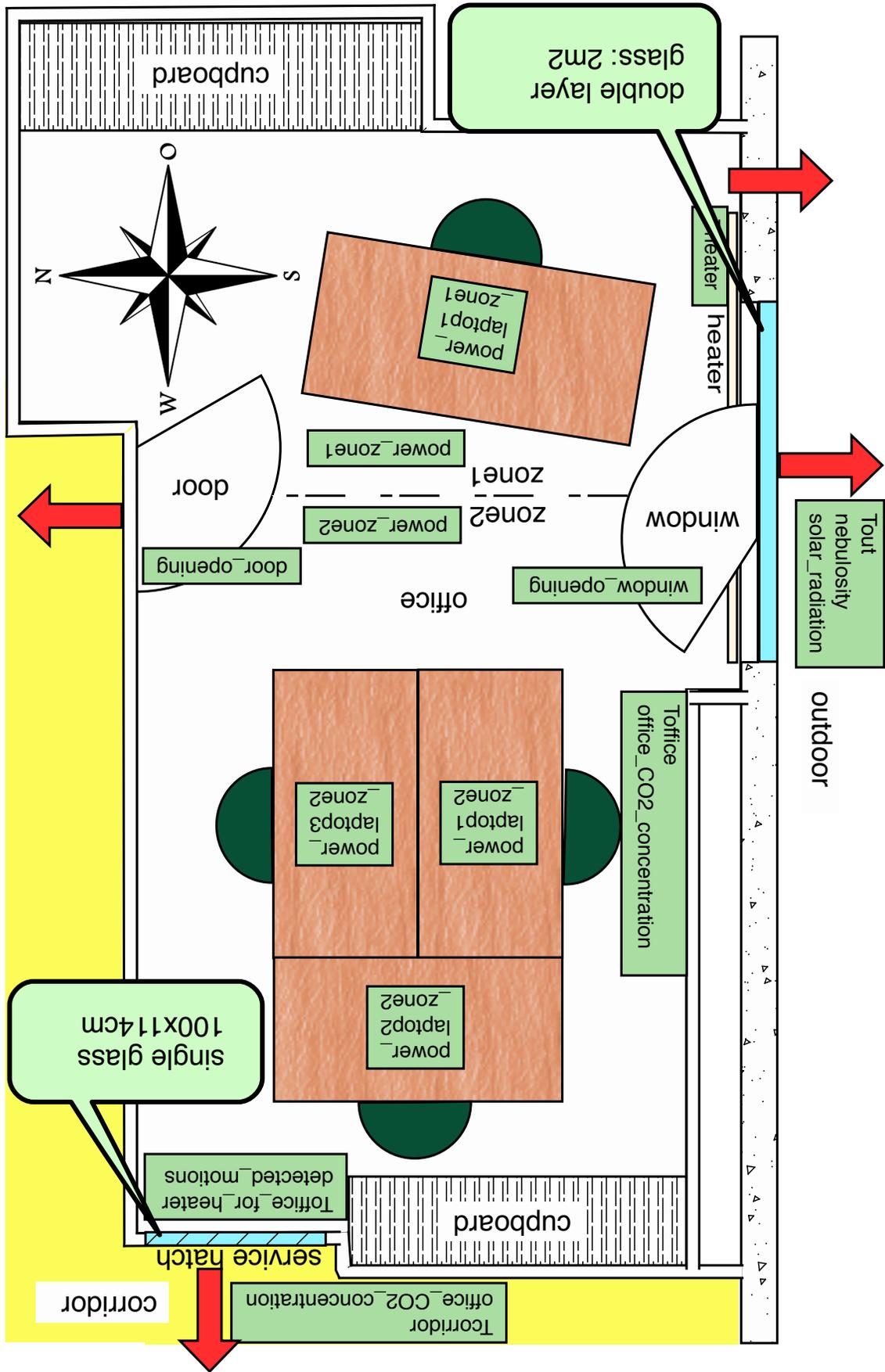


Figure N.2: Cas d'études : Bureau de chercheurs

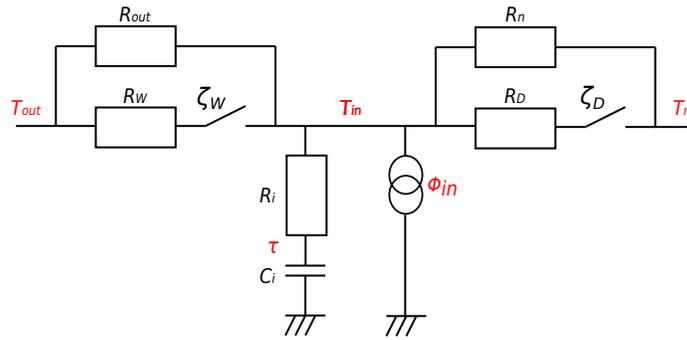


Figure N.3: Modèle "Référence"

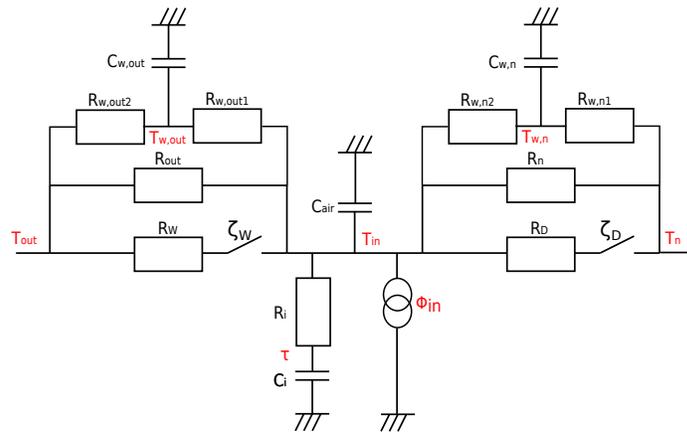


Figure N.4: Modèle à 4 capacités

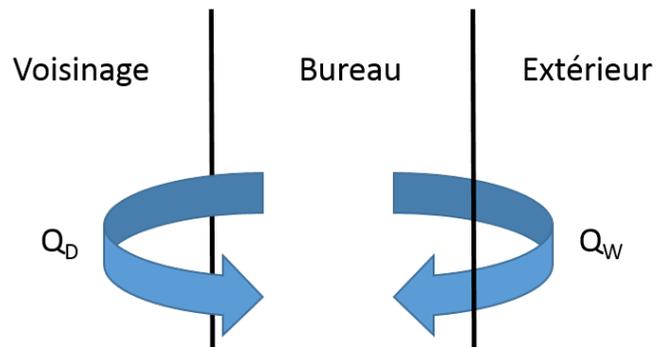
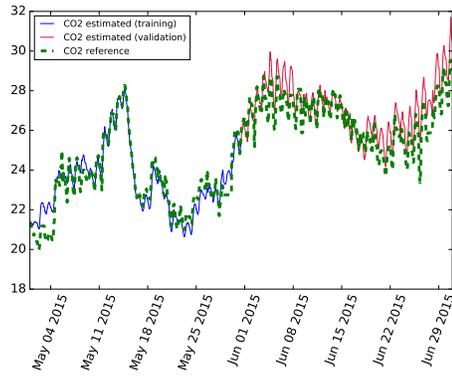
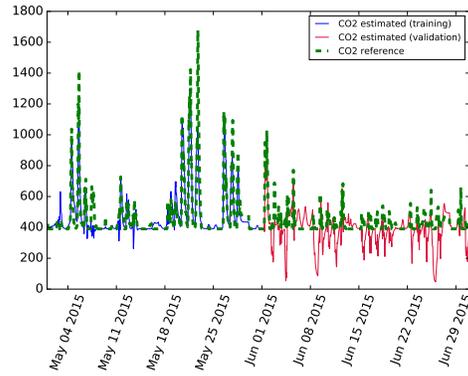


Figure N.5: Modèle aéralique

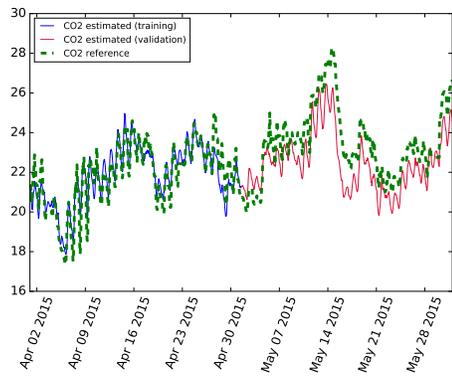


(a) Température prédite par le modèle ARX pour les mois de Mai / Juin

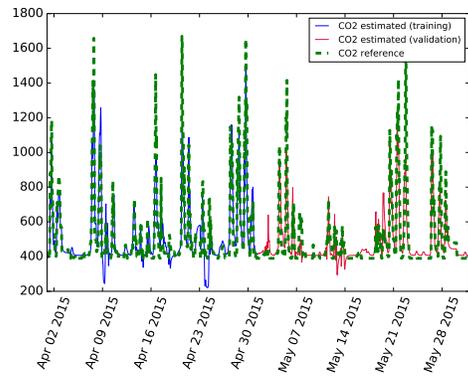


(b) Concentration de CO<sub>2</sub> prédite par le modèle ARX pour les mois de Mai / Juin

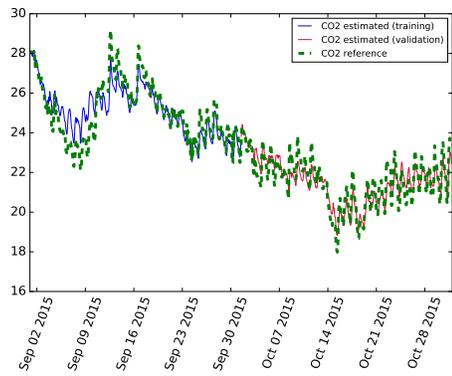
Figure N.6: Résultats du modèle ARX pour la période Mai / Juin



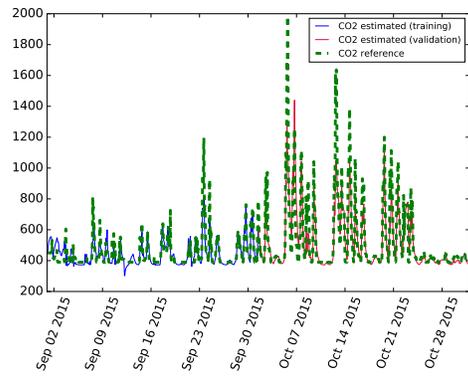
(a) Température prédite par le modèle ARX pour les mois de Avril / Mai



(b) Concentration de CO<sub>2</sub> prédite par le modèle ARX pour les mois de Avril / Mai



(c) Température prédite par le modèle ARX pour les mois de Septembre / Octobre



(d) Concentration de CO<sub>2</sub> prédite par le modèle ARX pour les mois de Septembre / Octobre

Figure N.7: Résultats du modèle ARX

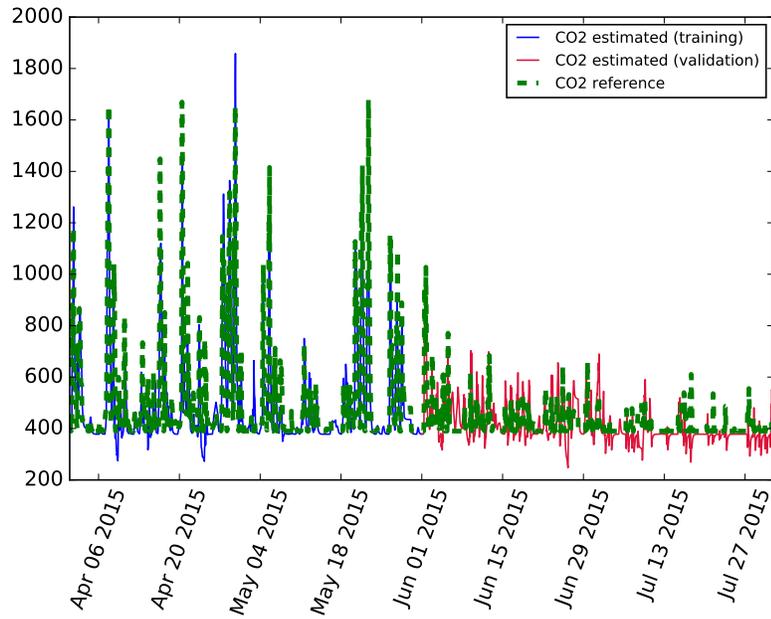


Figure N.8: Concentration de CO<sub>2</sub> prédite par le modèle ARX inspiré de la physique

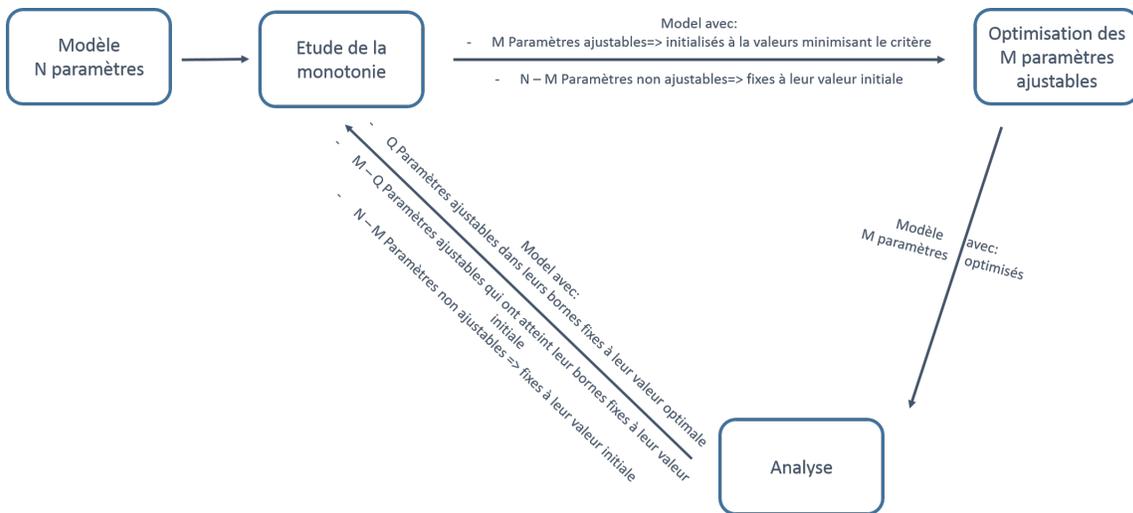


Figure N.9: Schéma de principe de la méthode de méta-optimisation

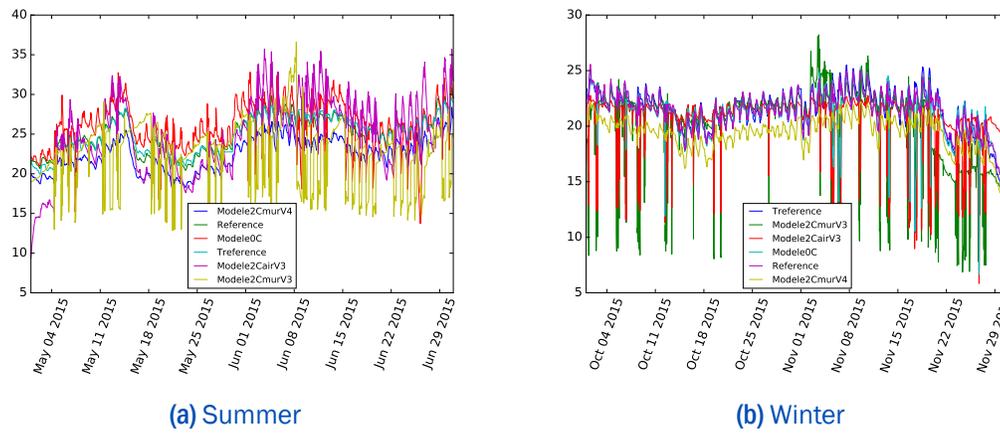


Figure N.10: Temperature profiles obtained with the different models

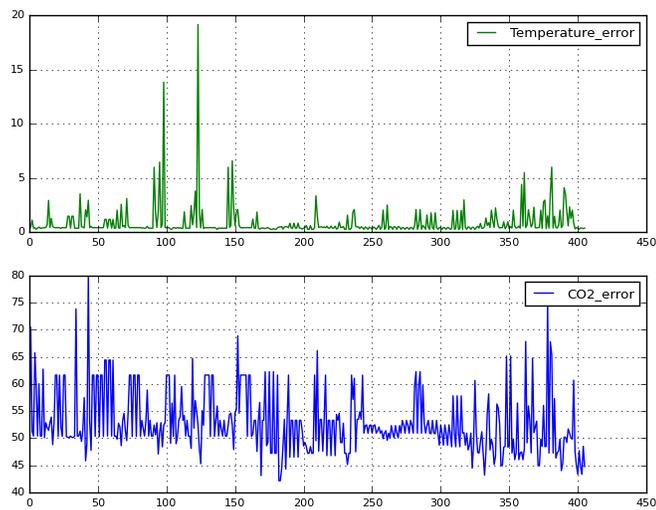


Figure N.11: Erreur moyenne absolue obtenue pour 406 simulations différentes.

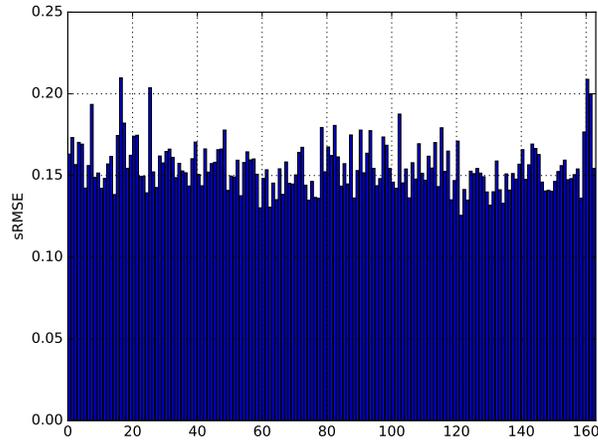
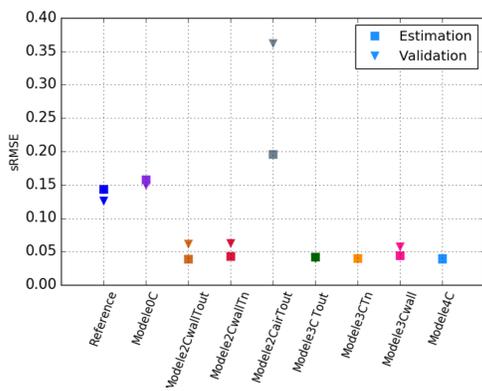
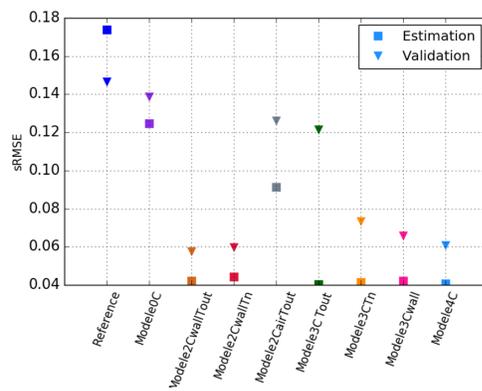


Figure N.12: Résultats du test d'ergodicité



(a) Summer



(b) Winter

Figure N.13: Results of the different models in estimation and validation phase

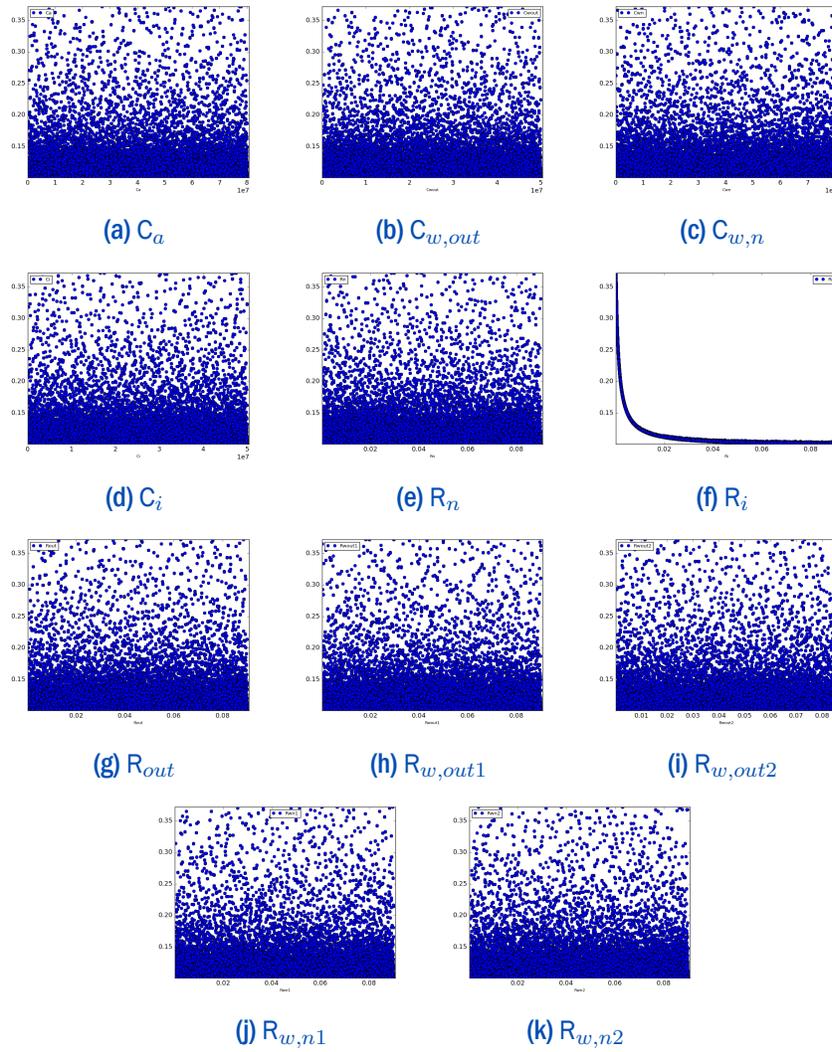


Figure N.14: Sum of the residuals according to the values of the different structural parameters

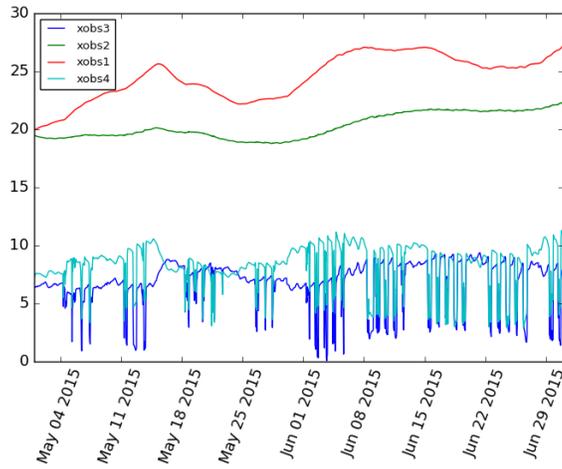


Figure N.15: Evolution of the different components of  $\hat{X}$  over 60 days

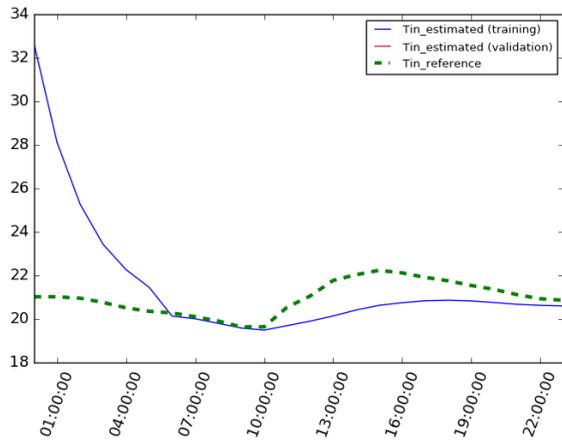


Figure N.16: Temperature profiles for the 1st of November

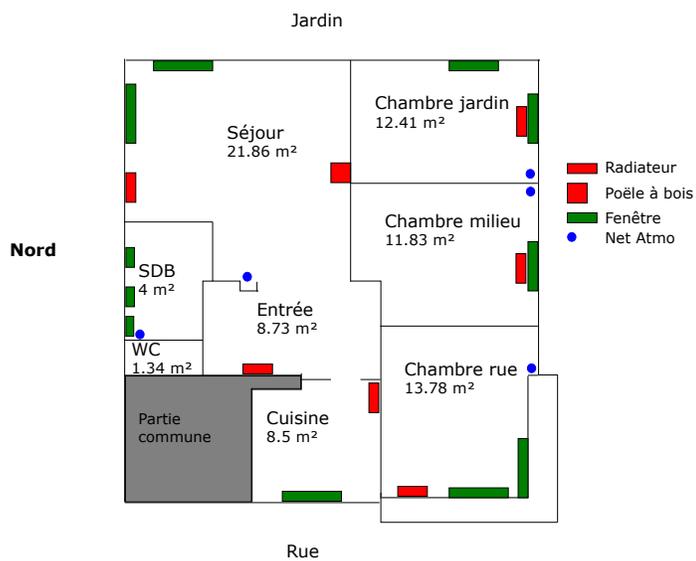


Figure N.17: Plan architectural de l'appartement

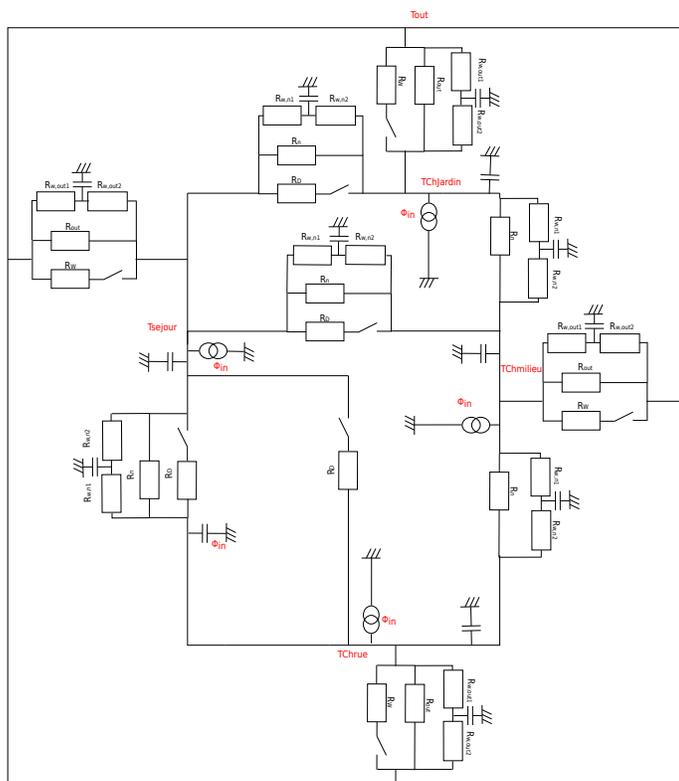
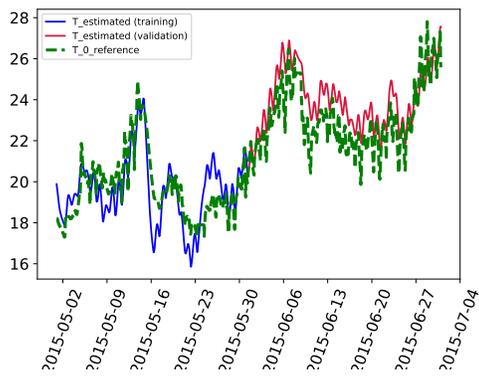
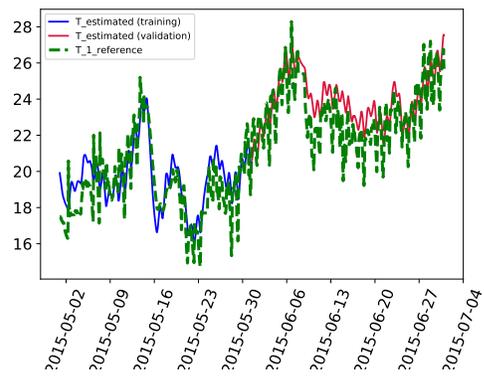


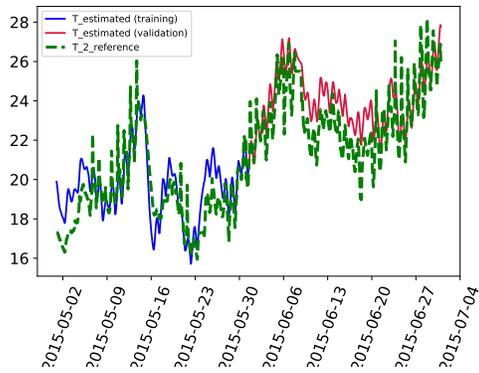
Figure N.18: RC model of the flat



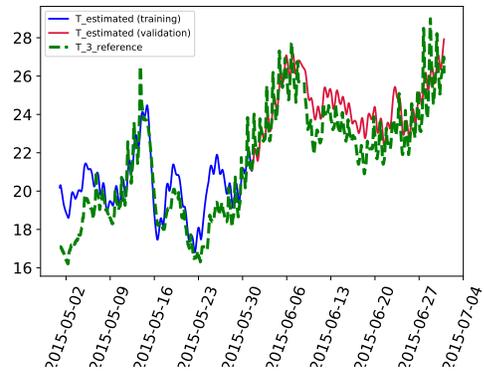
(a) Séjour



(b) Chambre jardin

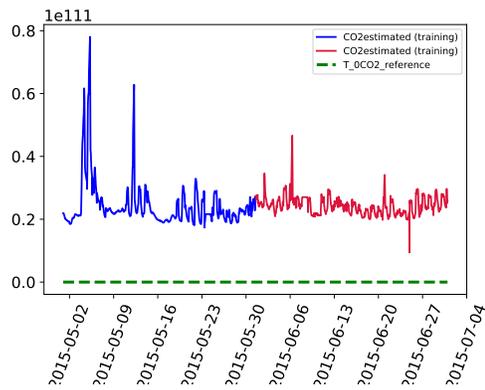


(c) Chambre milieu

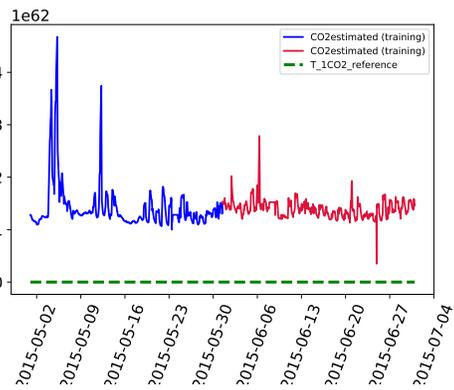


(d) Chambre rue

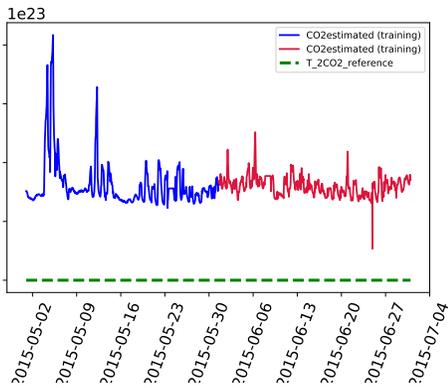
Figure N.19: Profils de température pour le scénario été



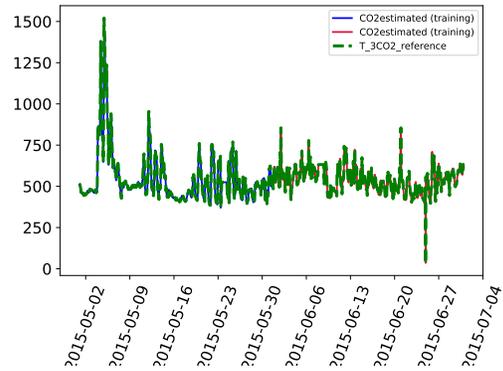
(a) Séjour



(b) Chambre jardin



(c) Chambre milieu



(d) Chambre rue

Figure N.20: Profil de concentration en CO<sub>2</sub> pour le scénario été

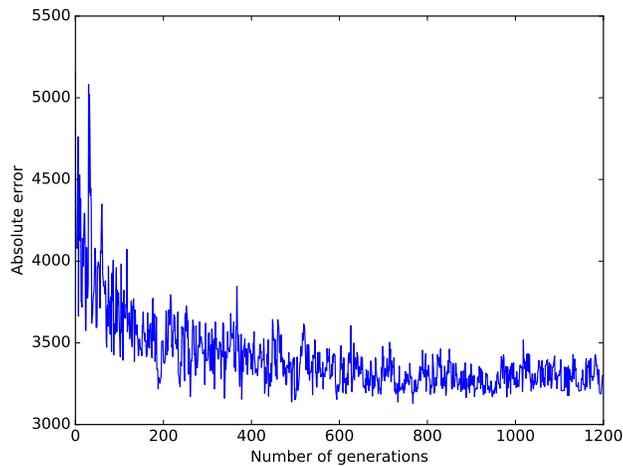


Figure N.21: Evolution of the absolute error along the generations

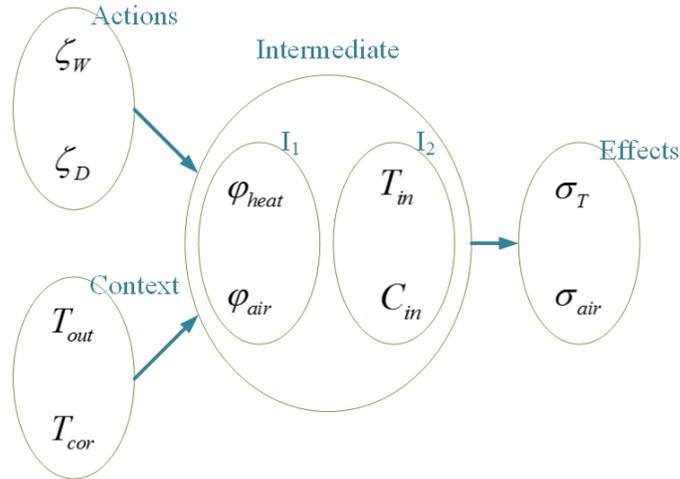


Figure N.22: Schéma général d'explications

hour	Δ actions	Δ effects	Δ intermediates
08:00	[Window icon] ↑		OUT [Door icon] ↓
09:00	[Door icon] ↑ [Window icon] ↑		COR [Heat icon] ↑ [Air icon] ↓
10:00	[Door icon] ↓ [Window icon] ↓		COR [Heat icon] ↓ [Air icon] ↓
11:00	[Door icon] ↓		COR [Heat icon] ↓
12:00	[Door icon] ↑	[Heat icon] ↑ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓
13:00	[Door icon] ↑ [Window icon] ↑	[Heat icon] ↑ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓
14:00	[Door icon] ↑		COR [Heat icon] ↓ [Air icon] ↓
15:00	[Door icon] ↓ [Window icon] ↓	[Heat icon] ↓ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓
16:00		[Heat icon] ↓ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓
17:00	[Door icon] ↑ [Window icon] ↑	[Heat icon] ↓ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓
18:00	[Window icon] ↑		
19:00	[Window icon] ↓		
ALL		[Heat icon] ↓ [Air icon] ↓	COR [Heat icon] ↓ [Air icon] ↓

The average gain for the entire day if occupant followed the recommended actions

- [Door icon] (Opening / closing) the door
- [Window icon] (Opening / closing) the window
- [Heat icon] Heat flow (to / from) the office
- [Air icon] Air flow (to / from) the office
- [Air quality icon] (Enhance / decrease) Air quality in the office
- [Thermal comfort icon] (Enhance / decrease) Thermal comfort in the office
- [Change icon] The amount of change

Figure N.23: Explications différentielles pour aider les occupants à comprendre l'impact de leurs actions. par exemple à 16h l'occupant se rapproche du comportement "optimal"; à midi il aurait du ouvrir la porte plus longtemps pour améliorer fortement son confort thermique et sensiblement la qualité d'air.

### Traduction

Dans le créneau horaire 17h-18h, si vous aviez laissé la porte et la fenêtre ouvertes beaucoup plus longtemps, la qualité de l'air aurait diminué un petit peu, il y aurait eu un courant d'air sensible vers le couloir et un flux thermique léger venant de l'extérieur ;

Figure N.24: Exemple de phrase obtenue en sortie de Ariane-Heloise GRA-FRA GM

## Bibliography



- [1] *Connaissance des bâtiments anciens et économie d'énergie*, tech. rep., CEREMA, 2007.
- [2] G. G. J. ACHTERBOSCH, P. P. G. DE JONG, C. E. KRIST-SPIT, S. F. VAN DER MEULEN, AND J. VERBERNE, *The development of a convenient thermal dynamic building model?*, *Energy and Buildings*, 8 (1985), pp. 183–196.
- [3] H. A. AGLAN, *Predictive model for co 2 generation and decay in building envelopes*, *Journal of applied physics*, 93 (2003), pp. 1287–1290.
- [4] A. ALAOUI EL AZHER, *Identification de systèmes à deux échelles de temps et application au chauffage optimal de bâtiments*, PhD thesis, Paris, ENMP, 1992.
- [5] A. ALZHOURI ALYAFI, M. PAL, S. PLOIX, P. REIGNIER, AND S. BANDY-OPADHYAY, *Differential explanations for energy management in buildings*, Computing Conference London, (2017).
- [6] A. ANDROUTSOPOULOS, J. BLOEM, H. VAN DIJK, AND P. H. BAKER, *Comparison of user performance when applying system identification for assessment of the energy performance of building components*, *Building and environment*, 43 (2008), pp. 189–196.
- [7] P. R. ARMSTRONG, *Model identification with application to building control and fault detection*, PhD thesis, Massachusetts Institute of Technology, 2004.
- [8] C. ASSEGOND, *Affichage des consommations et réflexivité des ménages : construire une culture domestique de l'énergie par l'information*, in *les sociétés contemporaines l'épreuve des transitions énergétiques*, 2015.
- [9] P. BACHER AND H. MADSEN, *Identifying suitable models for the heat dynamics of buildings*, *Energy and Buildings*, 43 (2011), pp. 1511–1522.

- [10] B. BARREAU, A. DUJIN, AND M. VEDIE, *Comment limiter l'effet rebond des politiques d'efficacité énergétique dans le logement ? l'importance des incitations comportementales*, in Centre d'analyse stratégique, vol. 320, 2013.
- [11] F. BARTIAUX-UCL, G. VEKEMANS-VITO, F. BARTIAUX, G. VEKEMANS, K. GRAM-HANSEN, D. MAES, M. CANTAERT, B. SPIES, J. DESMEDT, O. M. JENSEN, ET AL., *influencing residential energy consumption*, (2006).
- [12] G. BELLU, M. P. SACCOMANI, S. AUDOLY, AND L. D'ANGIÒ, *Daisy: A new software tool to test global identifiability of biological and physiological systems*, *Computer methods and programs in biomedicine*, 88 (2007), pp. 52–61.
- [13] C. . BERMENT, V. ET BOITET.
- [14] T. BERTHOU, P. STABAT, R. SALVAZET, AND D. MARCHIO, *Comparaison de modèles linéaires inverses pour la mise en place de stratégies d'effacement.*, in XXXe Rencontre AUGCIBPSA Chambéry, Savoie,, 2012.
- [15] T. BOUACHE, K. LIMAM, AND W. BOSSCHAERTS, *New thermal parameters identification approach applied to the thermal renovation of buildings*, *Energy and Buildings*, 104 (2015), pp. 156–164.
- [16] J. E. BRAUN AND N. CHATURVEDI, *An inverse gray-box model for transient building load prediction*, *HVAC&R Research*, 8 (2002), pp. 73–99.
- [17] G. A. BRISEPIERRE, *Pratiques de consommation d'énergie dans les bâtiments performants : consommations théoriques et consommations réelles - monographie citée de l'environnement*, tech. rep., leschantiersleroymerlinsource, 2013.
- [18] J. CARASSUS, *Les immeubles de bureaux «verts» tiennent-ils leurs promesses?*, *Performances réelles, valeur immobilières et certification*, (2011).
- [19] J. CARASSUS, C. LAUMONIER, B. SESOLIS, D. JANVIER, AND R. WRONA, *Vivre dans un logement basse consommation d'énergie - une approche socio-éco-technique.*, tech. rep., CERQUAL, 2013.
- [20] L. CASTILLO, R. ENRÍQUEZ, M. JIMÉNEZ, AND M. HERAS, *Dynamic integrated method based on regression and averages, applied to estimate the thermal parameters of a room in an occupied office building in madrid*, *Energy and Buildings*, 81 (2014), pp. 337–362.

- [21] J.-M. CAYLA, B. ALLIBE, AND M.-H. LAURENT, *From practices to behaviors: Estimating the impact of household behavior on space heating energy consumption*, in ACEEE Summer Study on Energy Efficiency in Buildings, 2010.
- [22] CENTRE D'ANALYSE STRATÉGIQUE, *Nudges verts : De nouvelles incitations pour des comportements écologiques.*, Cahier de recherche CREDOC, (2011).
- [23] T. CHEN, *Real-time predictive supervisory operation of building thermal systems with thermal mass*, Energy and Buildings, 33 (2001), pp. 141–150.
- [24] D. COLEY AND J. PENMAN, *Second order system identification in the thermal response of real buildings. paper ii: recursive formulation for on-line building energy management and control*, Building and Environment, 27 (1992), pp. 269–277.
- [25] K. DEB, S. AGRAWAL, A. PRATAP, AND T. MEYARIVAN, *A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii*, in International Conference on Parallel Problem Solving From Nature, Springer, 2000, pp. 849–858.
- [26] E. P. DEL BARRIO, G. LEFEBVRE, P. BEHAR, AND N. BAILLY, *Using model size reduction techniques for thermal control applications in buildings*, Energy and Buildings, 33 (2000), pp. 1–14.
- [27] D. DESJEUX, *La question des échelles d'observation en sciences humaines appliquées au domaine de la santé*, Recherche en soins infirmiers, (2006), pp. 14–21.
- [28] B. DONG, B. ANDREWS, K. P. LAM, M. HÖYCNCK, R. ZHANG, Y.-S. CHIOU, AND D. BENITEZ, *An information technology enabled sustainability test-bed (itest) for occupancy detection through an environmental sensing network*, Energy and Buildings, 42 (2010), pp. 1038–1046.
- [29] A. DUJIN, I. MOUSSAOUI, AND M. B. MORDRET, X AND, *Les usages de l'énergie dans les entreprises du secteur tertiaire. des systèmes techniques aux pratiques*, CREDOC, Cahier de recherche, (2012).
- [30] M. ERICKSON, A. MAYER, AND J. HORN, *The niched pareto genetic algorithm 2 applied to the design of groundwater remediation systems*, in International Conference on Evolutionary Multi-Criterion Optimization, Springer, 2001, pp. 681–695.

- [31] V. FABI, R. V. ANDERSEN, S. CORGNATI, AND B. W. OLESEN, *Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models*, *Building and Environment*, 58 (2012), pp. 188–198.
- [32] A. FARUQUI, S. SERGICI, AND A. SHARIF, *The impact of informational feedback on energy consumption? a survey of the experimental evidence*, *Energy*, 35 (2010), pp. 1598–1608.
- [33] L. FERKL AND J. ŠIROKÝ, *Ceiling radiant cooling: Comparison of armax and subspace identification modelling methods*, *Building and Environment*, 45 (2010), pp. 205–212.
- [34] L. FESTINGER, *A theory of cognitive dissonance*, vol. 2, Stanford university press, 1962.
- [35] C. FISCHER, *Feedback on household electricity consumption: a tool for saving energy?*, *Energy efficiency*, 1 (2008), pp. 79–104.
- [36] A. FOUCQUIER, A. BRUN, G. A. FAGGIANELLI, AND F. SUARD, *Effect of the wall merging on a simplified building energy model : accuracy vs number of equations*, in *Proceedings of BS2013*, 2013.
- [37] G. FRAISSE, C. VIARDOT, O. LAFABRIE, AND G. ACHARD, *Development of a simplified and accurate building model based on electrical analogy*, *Energy and Buildings*, 34 (2002), pp. 1017 – 1031.
- [38] R. Z. FREIRE, G. H. OLIVEIRA, AND N. MENDES, *Development of regression equations for predicting energy and hygrothermal performance of buildings*, *Energy and Buildings*, 40 (2008), pp. 810–820.
- [39] L. GAMBERINI, A. SPAGNOLLI, N. CORRADI, G. JACUCCI, G. TUSA, T. MIKKOLA, L. ZAMBONI, AND E. HOGGAN, *Tailoring feedback to users? actions in a persuasive game for household electricity conservation*, in *International Conference on Persuasive Technology*, Springer, 2012, pp. 100–111.
- [40] D. M. GATES, *Biophysical ecology*, Courier Corporation, 2012.
- [41] E. GHISI AND J. TINKER, *Optimising energy consumption in offices as a function of window area and room size*, in *Seventh International Ibpsa Conference—International Building Performance Simulation association*, 2001.

- [42] K. GODFREY AND J. DiSTEFANO III, *Identifiability of model parameters*, Identifiability of parametric models, 1 (1987), pp. 1–20.
- [43] J. GOFFART, *Impact de la variabilité des données météorologiques sur une maison basse consommation. Application des analyses de sensibilité pour les entrées temporelles.*, PhD thesis, Université Grenoble Alpes, 2013.
- [44] L. A. GREENING, D. L. GREENE, AND C. DIFIGLIO, *Energy efficiency and consumption?the rebound effect?a survey*, Energy policy, 28 (2000), pp. 389–401.
- [45] C. GUEYMARD, *Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data*, Solar Energy, 51 (1993), pp. 121–138.
- [46] H. HARB, N. BOYANOV, L. HERNANDEZ, R. STREBLOW, AND D. MÜLLER, *Development and validation of grey-box models for forecasting the thermal response of occupied buildings*, Energy and Buildings, 117 (2016), pp. 199–207.
- [47] I. HAZYUK, C. GHIAUS, , AND D. PENHOUE, *Optimal temperature control of intermittently heated buildings using model predictive control*, Building and Environment, Part I?Building Modeling (2012), p. 379?87.
- [48] T. HOMMA AND A. SALTELLI, *Importance measures in global sensitivity analysis of nonlinear models*, Reliability Engineering & System Safety, 52 (1996), pp. 1–17.
- [49] J. HORN, N. NAFPLIOTIS, AND D. E. GOLDBERG, *A niched pareto genetic algorithm for multiobjective optimization*, in Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on, Ieee, 1994, pp. 82–87.
- [50] H. HUANG, S. KATO, AND R. HU, *Optimum design for indoor humidity by coupling genetic algorithm with transient simulation based on contribution ratio of indoor humidity and climate analysis*, Energy and Buildings, 47 (2012), pp. 208–216.
- [51] M. IQBAL, *An introduction to solar radiation*, Elsevier, 2012.
- [52] M. J. JIMENEZ AND H. MADSEN, *Models for describing the thermal characteristics of building components*, Building and Environment, 43 (2008), pp. 152–162.

- [53] R. E. KALMAN, *Canonical structure of linear dynamical systems*, Proceedings of the National Academy of Sciences, 48 (1962), pp. 596–600.
- [54] J. KÄMPF AND D. ROBINSON, *A simplified thermal model to support analysis of urban resource flows.*, Energy and Buildings, 39 (2007).
- [55] F. KASTEN AND G. CZEPLAK, *Solar and terrestrial radiation dependent on the amount and type of cloud*, Solar energy, 24 (1980), pp. 177–189.
- [56] J. KJELDSKOV, M. B. SKOV, J. PAAY, AND R. PATHMANATHAN, *Using mobile phones to support sustainability: a field study of residential electricity consumption*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2012, pp. 2347–2356.
- [57] D. KRAFT, *A software package for sequential quadratic programming*, Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt, (1988).
- [58] L. KUMAR, A. K. SKIDMORE, AND E. KNOWLES, *Modelling topographic variation in solar radiation in a gis environment*, International Journal of Geographical Information Science, 11 (1997), pp. 475–497.
- [59] A. LE MOUNIER, *Méta-optimisation pour la calibration automatique de modèles énergétiques bâtiment pour le pilotage anticipative*, PhD thesis, Communauté Universit Grenoble Alpes, 2016.
- [60] A. LE MOUNIER, B. DELINCHANT, AND S. PLOIX, *Choix de structures de modèles pertinentes pour l'identification des systèmes de gestion d'énergie*, in Conférence IBPSA France-Arras, 2014.
- [61] Z. LIAO AND A. DEXTER, *A simplified physical model for estimating the average air temperature in multi-zone heating systems*, Building and environment, 39 (2004), pp. 1013–1022.
- [62] B. Y. LIU AND R. C. JORDAN, *The interrelationship and characteristic distribution of direct, diffuse and total solar radiation*, Solar energy, 4 (1960), pp. 1–19.
- [63] L. LJUNG, *System identification: Theory for the user*, prentice hall information and system sciences series, 1999.

- [64] G. LOWRY AND M.-W. LEE, *Modelling the passive thermal response of a building using sparse bms data*, Applied energy, 78 (2004), pp. 53–62.
- [65] E. MATHEWS, P. RICHARDS, AND C. LOMBARD, *A first-order thermal model for building design.*, Energy and Buildings, (1994), pp. 133–145.
- [66] A. MECHAQRANE AND M. ZOUAK, *A comparison of linear and neural network arx models applied to a prediction of the indoor temperature of a building*, Neural Computing & Applications, 13 (2004), pp. 32–37.
- [67] O. MEJRI, E. P. DEL BARRIO, AND N. GHRAB-MORCOS, *Energy performance assessment of occupied buildings using model identification techniques*, Energy and Buildings, 43 (2011), pp. 285–299.
- [68] M. D. MORRIS, *Factorial sampling plans for preliminary computational experiments*, Technometrics, 33 (1991), pp. 161–174.
- [69] G. MOSER, *Psychologie environnementale: les relations homme-environnement*, Armando Editore, 2009.
- [70] M. C. MOZER, *The neural network house: An environment hat adapts to its inhabitants*, in Proc. AAAI Spring Symp. Intelligent Environments, vol. 58, 1998.
- [71] G. MUSTAFARAJ, G. LOWRY, AND J. CHEN, *Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office*, Energy and Buildings, 43 (2011), pp. 1452–1460.
- [72] A. K. NANDI, D. CHAKRABORTY, AND W. VAZ, *Design of a comfortable optimal driving strategy for electric vehicles using multi-objective optimization*, Journal of Power Sources, 283 (2015), pp. 1–18.
- [73] I. NAVEROS, C. GHIAUS, D. RUÍZ, AND S. CASTAÑO, *Physical parameters identification of walls using arx models obtained by deduction*, Energy and Buildings, 108 (2015), pp. 317–329.
- [74] T. K. NIELSEN, *Simple tool to evaluate energy demand and indoor environment in the early stages of building design.*, Solar Energy, 78 (2005).
- [75] F. OLLIVIER, *Le probl me de l'identifiabilité structurelle globale: approche théorique, thode effectives et bornes de complexité*, theses, Ecole Polytechnique, 1990.

- [76] A. RABL, *Parameter estimation in buildings: methods for dynamic analysis of measured energy use*, Journal of Solar Energy Engineering, 110 (1988), pp. 52–66.
- [77] A. RAUE, J. KARLSSON, M. P. SACCOMANI, M. JIRSTRAND, AND J. TIMMER, *Comparison of approaches for parameter identifiability analysis of biological systems*, Bioinformatics, (2014), p. btt006.
- [78] G. REYNDERS, *Quantifying the Impact of Building Design on the Potential of Structural Storage for Active Demand Response in Residential Buildings*, PhD thesis, 2015.
- [79] J. RICHALET, *Pratique de l'identification*, Hermes, 1991.
- [80] A. SALTELLI, S. TARANTOLA, AND K.-S. CHAN, *A quantitative model-independent method for global sensitivity analysis of model output*, Technometrics, 41 (1999), pp. 39–56.
- [81] J. D. SCHAFFER, *Multiple objective optimization with vector evaluated genetic algorithms*, in Proceedings of the 1st international Conference on Genetic Algorithms, L. Erlbaum Associates Inc., 1985, pp. 93–100.
- [82] A. SEDOGLAVIC, *A probabilistic algorithm to test local algebraic observability in polynomial time*, in Proceedings of the 2001 international symposium on Symbolic and algebraic computation, ACM, 2001, pp. 309–317.
- [83] P. H. SHAIKH, N. B. M. NOR, P. NALLAGOWNDEN, I. ELAMVAZUTHI, AND T. IBRAHIM, *A review on optimized control systems for building energy and comfort management of smart sustainable buildings*, Renewable and Sustainable Energy Reviews, 34 (2014), pp. 409–429.
- [84] O. SIDLER, *De la conception à la mesure, comment expliquer les écarts?*, Evaluer les performances des bâtiments basse consommation, CSTB/CETE de l'OUEST, (2011).
- [85] I. M. SOBOL, *Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates*, Mathematics and computers in simulation, 55 (2001), pp. 271–280.
- [86] J. SPENCER, *Fourier series representation of the position of the sun*, Search, 2 (1971), pp. 172–172.

- [87] J. TEICHMANN, M. LAUSTER, M. FUCHS, R. STREBLOW, AND D. MÜLLER, *Validation of a simplified building model used for city district simulation*, in Proceedings of BS2013, 2013.
- [88] S. VAJDA, K. R. GODFREY, AND H. RABITZ, *Similarity transformation approach to identifiability analysis of nonlinear compartmental models*, *Mathematical biosciences*, 93 (1989), pp. 217–248.
- [89] G. VIRK, J. CHEUNG, AND D. LOVEDAY, *Practical stochastic multivariable identification for buildings*, *Applied mathematical modelling*, 19 (1995), pp. 621–636.
- [90] E. WALTER, *Lecture notes in biomathematics*, (1982).
- [91] L. WONG AND W. CHOW, *Solar radiation model*, *Applied Energy*, 69 (2001), pp. 191–224.
- [92] S. WU AND J.-Q. SUN, *A physics-based linear parametric model of room temperature in office buildings*, *Building and Environment*, 50 (2012), pp. 1–9.
- [93] K. YUN, R. LUCK, P. J. MAGO, AND H. CHO, *Building hourly thermal load prediction using an indexed arr model*, *Energy and Buildings*, 54 (2012), pp. 225–233.
- [94] C. ZAYANE, *Identification d'un modèle de comportement thermique de bâtiment à partir de sa courbe de charge*, PhD thesis, École Nationale Supérieure des Mines de Paris, 2011.
- [95] M. ZÉLEM, R. GOURNET, AND C. BESLAY, *Pas de «smart cities» sans «smart habitants»*, *Urbia. Les cahiers du développement urbain durable*, (2013), pp. 45–60.
- [96] M.-C. ZÉLEM, *Politiques de maîtrise de la demande d'énergie et résistances au changement: une approche socio-anthropologique*, Editions L'Harmattan, 2010.



## List of Figures



I.1	Persuasive interface . . . . .	7
I.2	Principle of electric analogy . . . . .	16
I.3	Learning and validation process . . . . .	20
I.4	General objective and process . . . . .	23
II.1	Study case: researcher office . . . . .	32
II.2	Reference model . . . . .	34
II.3	Aeraulic model . . . . .	36
II.4	Model with no capacitor . . . . .	37
II.5	Model with an exterior wall capacitor . . . . .	38
II.6	Model with inner wall capacitor . . . . .	39
II.7	Model with an air capacitor . . . . .	41
II.8	Model with air and exterior wall capacitors . . . . .	42
II.9	Model with air and inner wall capacitors . . . . .	43
II.10	Model with a capacitor for the inner and the exterior wall . . . . .	43
II.11	Model with 4 capacitors . . . . .	44
III.1	Outdoor temperature . . . . .	55
III.2	Results of prediction for the multi-regression models including the internal gains . . . . .	56
III.3	Results of the multi-regression model during April and May . . . . .	57

III.4 Results of the multi-regression model during September and October	57
III.5 Results of the multi-regression model without the internal gains . . . .	58
III.6 Difference of temperatures between the two scenarios tested . . . . .	59
III.7 Environmental inputs during the month of June . . . . .	61
III.8 CO <sub>2</sub> concentrations of the multi-regression model estimated during the month of July and cross-validated in June . . . . .	61
III.9 RC model . . . . .	63
III.10Aeraulic model . . . . .	63
III.11Results obtained for the multi-regression models inspired by physics .	65
III.12Diagram of the meta-optimization principle . . . . .	69
III.13Results of the reference model . . . . .	72
III.14Sensor measurements of door and window openings . . . . .	73
III.15Temperature profiles obtained with the different models . . . . .	73
III.16Results of the different models in estimation and validation phase . .	74
III.17Estimated and measured temperatures for the model “Model2CwallT <sub>out</sub> ” . . . . .	74
III.18Values of parameters to estimate along the iterations for the model “Model2CwallT <sub>out</sub> ” . . . . .	75
III.19Estimated and measured temperatures for the model “ReferenceNoζ”	77
III.20Impact of the duration of the training data . . . . .	77
III.21Absolute average error for 406 simulations . . . . .	78
III.22Scheme of principle of the algorithm NSGA-II [72] . . . . .	81
III.23Evolution of the temperature error along the generation . . . . .	82
III.24Impact of the length of the learning phase on the temperature error .	83
III.25Results of the robustness test . . . . .	84
III.26Results of the reference model . . . . .	85
III.27Results of the different models in estimation and validation phase . .	86

III.28	sRMSE values of the valid models in validation phase . . . . .	86
III.29	Sum of the sRMSE values for both seasons . . . . .	87
III.30	Difference of indoor air temperature between the two scenario simulated	87
III.31	Morris analysis . . . . .	88
III.32	Sum of the residuals according to the values of the different structural parameters . . . . .	90
III.33	Sobol index . . . . .	90
III.34	Model3Cfinal . . . . .	91
III.35	Dispersion of the parameters over 50 estimations for the Model4C . .	92
III.36	Sum of the residuals according to the values of the different structural parameters . . . . .	93
III.37	Dispersion of the parameters over 50 estimations for the Reference model . . . . .	94
III.38	Sum of the residuals according to the values of the different structural parameters . . . . .	95
III.39	Sobol index over the new parameter ranges . . . . .	96
III.40	Evolution of the different components of $\hat{X}$ over 60 days . . . . .	98
III.41	Temperature profiles for the 1st of November . . . . .	99
III.42	Yearly forecasts of CO <sub>2</sub> and temperature of the multi-regression mod- els estimated either in a month in winter or a month in summer . . .	100
III.43	Model with an air capacitor and one for the inner wall . . . . .	103
III.44	Model with an air capacitor and one for the inner wall . . . . .	103
IV.1	Architectural plan of the flat . . . . .	111
IV.2	Comparison between contact data and not . . . . .	113
IV.3	Temperature and CO <sub>2</sub> profiles for Model4C with the new model for openings . . . . .	115
IV.4	Measured and simulated window openings . . . . .	116

IV.5 Model4C . . . . .	117
IV.6 Temperature and CO <sub>2</sub> concentration of the bedroom “garden” with a mono-zone model . . . . .	117
IV.7 Model with 2 neighbouring zones . . . . .	118
IV.8 Temperature and CO <sub>2</sub> concentration of the bedroom ”garden” with a two-neighbouring zones model . . . . .	119
IV.9 Bedroom ”garden” indoor air temperature with an error on the input	120
IV.10Methodology implemented . . . . .	121
IV.11Structure of the xml file . . . . .	123
IV.12Structure of an element ”LivingZone” of the xml file . . . . .	123
IV.13Representation of an outer wall . . . . .	127
IV.14Representation of an inner wall with a door . . . . .	127
IV.15RC model of the flat . . . . .	129
IV.16Temperature profiles for the summer scenario . . . . .	131
IV.17Temperature profiles for the summer scenario . . . . .	132
IV.18Temperature profiles for the winter scenario (1200 generations of 800 individuals) . . . . .	133
IV.19CO <sub>2</sub> profiles for the winter scenario (1200 generations of 800 individuals)	134
IV.20Evolution of the absolute error along the generations . . . . .	134
IV.21General schema of explanations . . . . .	135
IV.22Differential explanations help the inhabitants to understand their actions like at 4pm, the user behaves correctly (its action was similar to the proposed optimal plan); at 12am, the inhabitant should have opened the door for much longer time, this would helped him getting a high improvement in his thermal comfort and slight enhancement in the air quality. . . . .	136
IV.23Output from Ariane-Heloise GRA-FRA GM phase . . . . .	137
N.1 Objectif général et méthodologie mise en place . . . . .	148

N.2	Cas d'études : Bureau de chercheurs . . . . .	167
N.3	Modèle "Référence" . . . . .	168
N.4	Modèle à 4 capacités . . . . .	168
N.5	Modèle aéraulique . . . . .	168
N.6	Résultats du modèle ARX pour la période Mai / Juin . . . . .	169
N.7	Résultats du modèle ARX . . . . .	169
N.8	Concentration de CO <sub>2</sub> prédite par le modèle ARX inspiré de la physique	170
N.9	Schéma de principe de la méthode de méta-optimisation . . . . .	170
N.10	Temperature profiles obtained with the different models . . . . .	171
N.11	Erreur moyenne absolue obtenue pour 406 simulations différentes. . .	171
N.12	Résultats du test d'ergodicité . . . . .	172
N.13	Results of the different models in estimation and validation phase . .	172
N.14	Sum of the residuals according to the values of the different structural parameters . . . . .	173
N.15	Evolution of the different components of $\hat{X}$ over 60 days . . . . .	174
N.16	Temperature profiles for the 1st of November . . . . .	174
N.17	Plan architectural de l'appartement . . . . .	175
N.18	RC model of the flat . . . . .	175
N.19	Profils de température pour le scénario été . . . . .	176
N.20	Profil de concentration en CO <sub>2</sub> pour le scénario été . . . . .	177
N.21	Evolution of the absolute error along the generations . . . . .	177
N.22	Schéma général d'explications . . . . .	178
N.23	Explications différentielles pour aider les occupants à comprendre l'impact de leurs actions. par exemple à 16h l'occupant se rapproche du comportement "optimal"; à midi il aurait du ouvrir la porte plus longtemps pour améliorer fortement son confort thermique et sensi- blement la qualité d'air. . . . .	178

N.24 Exemple de phrase obtenue en sortie de Ariane-Heloise GRA-FRA GM178

## List of Tables



I.1	Learning duration and orders of the models . . . . .	21
II.1	Approaches to be tested . . . . .	30
II.2	Building blocks for exterior wall . . . . .	31
II.3	Building blocks for inner wall . . . . .	33
II.4	Building blocks for zone . . . . .	33
II.5	Summary of studied models . . . . .	45
III.1	Comparison of absolute average error and standard deviation for the reference model and the multi-regression inspired by physics on both training (in italics) and validation phases . . . . .	58
III.2	Comparison of absolute average error and standard deviation for the reference model and the multi-regression inspired by physics during the validation . . . . .	66
III.3	Initial values and range of parameters to be estimated . . . . .	70
III.4	Errors in summer . . . . .	71
III.5	Errors in winter . . . . .	71
III.6	sRMSE values . . . . .	84
III.7	Standard deviation of model estimated parameters for the Model4C .	92
III.8	Standard deviation of model estimated parameters for the "Reference" model . . . . .	94

III.9 sRMSE values for the validation on the whole year for the Model4C estimated in either summer or winter . . . . .	100
III.10 sRMSE values for the validation on the whole year for the Model4C estimated in either summer or winter . . . . .	101
III.11 Physical values of model parameters . . . . .	104
IV.1 Estimated values of $\zeta_D$ and $\zeta_W$ . . . . .	114
IV.2 sRMSE values . . . . .	115
IV.3 Comparison of sRMSE values for the two different approaches during both training (in italics) and validation phases . . . . .	120
IV.4 sRMSE values for the simulation of 100 generations of 100 individuals	131
IV.5 sRMSE values for the simulation of 1200 generations of 800 individuals	133
N.1 Bloc élémentaire pour le mur extérieur . . . . .	150
N.2 Bloc élémentaire pour le mur intérieur . . . . .	151
N.3 Bloc élémentaire pour une zone . . . . .	151
N.4 Comparaison de l'erreur absolue moyenne et de l'écart-type pour le modèle ARX classique et celui inspiré de la physique pendant la phase de validation . . . . .	155

I



## Abstract



In the current context of ecological transition, buildings represent a major pool of energy savings since they account for more than 40% of the final energy consumption. Involving end-users in their energy consumption could make a difference, but it requires to properly model the building to be able to forecast the temperature and CO<sub>2</sub> evolutions with a sufficient accuracy and a low computational time.

In this thesis, different kinds of models are implemented as well as different estimation methods and tested on two study cases: a mono-zone office and a multi-zone apartment. A selection methodology is set up in order to identify and validate the best model structure for the energy services. Finally, an automatic procedure to generate the model and services from only the informations an end-user can provide is developed for a multi-zone study case.

## Résumé



Dans le contexte de transition énergétique actuel, les bâtiments représentent un enjeu majeur puisqu'en France comme à l'étranger ils sont responsables de près de 40% de la consommation d'énergie finale. Impliquer l'utilisateur final dans sa consommation énergétique peut faire la différence mais cela nécessite de modéliser le bâtiment de manière adéquate afin de prédire l'évolution de la température et du CO<sub>2</sub> avec une précision suffisante et un temps de calcul faible.

Dans cette thèse, différents modèles et méthodes d'estimation paramétriques sont implémentées et testées sur deux cas d'études : un bureau et un appartement. Une méthode de sélection est mise en place pour identifier et valider la meilleure structure de modèle pour les services énergétiques. Enfin, est développée pour un cas d'étude multi-zones une procédure automatique pour générer les modèles et les services à partir uniquement des informations fournies par l'utilisateur final.