



HAL
open science

Assessment of the performances of building energy retrofit using sensor networks for the monitoring of energy consumption and usages

Mathieu Bourdeau

► **To cite this version:**

Mathieu Bourdeau. Assessment of the performances of building energy retrofit using sensor networks for the monitoring of energy consumption and usages. Electronics. Université Gustave Eiffel, 2022. English. NNT: 2022UEFL2031 . tel-04301338

HAL Id: tel-04301338

<https://theses.hal.science/tel-04301338v1>

Submitted on 23 Nov 2023

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.

**Ecole doctorale Mathématiques, Sciences et Technologies de
l'Information et de la Communication (MSTIC)**

THESE DE DOCTORAT CIFRE

Spécialité en génie électrique, électronique, photonique et systèmes (CNU 63)

Présentée et soutenue le 21 Novembre 2022, par:

Mathieu BOURDEAU

**Evaluation des performances de la rénovation
énergétique du bâtiment à l'aide de réseaux de
capteurs pour la supervision des consommations
énergétiques et des usages**

*Assessment of the performances of building energy
retrofit using sensor networks for the monitoring of
energy consumption and usages*

Thèse dirigée par :

M. Philippe BASSET, Professeur, ESYCOM, Université Gustave Eiffel

et co-encadrée par :

M. Elyes NEFZAOU, Professeur associé, ESYCOM, Université Gustave Eiffel

M. Julien WAEYTENS, Directeur de Recherche, COSYS, Université Gustave Eiffel

Composition du jury

Marjorie MUSY	Directrice de Recherche	CEREMA /IRSTV	Rapporteur
Brice TREMEAC	Maître de conférences	Lafset, CNAM	Rapporteur
Xiaofeng GUO	Professeur	LIED, Université Paris Cité	Président du jury
Benedict SCHWEGLER	Professeur associé	Université de Stanford	Invité
Philippe Basset	Professeur	ESYCOM, Université Gustave Eiffel	Directeur de thèse
Elyes NEFZAOU	Professeur Associé	ESYCOM, Université Gustave Eiffel	Co-encadrant
Julien WAEYTENS	Directeur de Recherche	COSYS, Université Gustave Eiffel	Co-encadrant

Acknowledgments

Ce travail est le résultat de trois années, neuf mois et quatre jours de travail très précisément au jour de sa soumission, une période riche d'enseignements à de nombreux niveaux. Je tiens donc à remercier ici toutes les personnes qui m'auront permis de mener ce projet à bien.

En premier lieu, mes remerciements d'adressent aux membres du jury qui ont accepté de bien vouloir évaluer mes travaux de recherche. Je remercie notamment les rapporteurs Mme. Marjory Musy, Directrice de Recherche au CEREMA, et M. Brice Tremeac, Maître de Conférences au CNAM. Je remercie également M. Xiaofeng Guo en sa qualité d'examineur.

Ce travail étant le fruit d'une collaboration entre le laboratoire ESYCOM et CAMEO SAS au sein du consortium ANDRE, j'adresse mes remerciements aux nombreux acteurs ayant activement participé à sa réalisation. J'adresse également mes remerciements à l'Initiative I-SITE FUTURE, financeur du projet ANDRE.

Côté CAMEO, je remercie M. David Werner pour son encadrement sur les débuts de ma thèse et M. Clément Gohet qui a accepté de reprendre et suivre ce projet pour me permettre de continuer à travailler dans les meilleures conditions possibles, et ce malgré les importants remaniements internes que nous avons connu. Je remercie également mes collègues Serge, Thibault, Alex, Alice et Yoann pour ces moments de détente au travail, indispensables au maintien de la motivation et du moral.

Côté projet, je remercie Mme. Margaux Happel qui a toujours pris le temps et a eu la patience de répondre à mes nombreuses questions. Je remercie également les participants à cette étude qui ont rendu ce travail possible et m'ont toujours accueilli avec la plus grande sympathie, et ce malgré mes (trop) nombreuses interventions chez eux.

Côté université, je remercie tout d'abord M. Philippe Basset, Professeur à l'Université Gustave Eiffel, d'avoir accepté de diriger cette thèse. Je remercie également M. Julien Waeytens, Directeur de recherche à l'Université Gustave Eiffel, pour sa grande disponibilité, ses questions et remarques constructives et ses encouragements. Je remercie également sincèrement M. Elyes Nefzaoui, enseignant, maître de stage et encadrant de thèse. Après ces années de collaboration, et malgré parfois des divergences d'opinion – sur plus de sept ans à se côtoyer, on ne peut pas toujours être d'accord. Je lui dois une partie importante de ma discipline de travail et de ma jeune carrière (et même celle à venir). Je le remercie de m'avoir donné ma chance après l'épisode difficile que j'ai traversé avant

de revenir en France en 2018 et de commencer ce travail. Je remercie également mes stagiaires, Mamadou et James, pour leur aide sur les analyses de données et pour m'avoir donné l'occasion de m'essayer en tant qu'encadrant de stage. Je remercie sincèrement M. Amine Bouzidi dont l'expertise et l'aide m'ont été très précieuses pour gagner un temps considérable. Enfin, je remercie chaleureusement mes collègues et amis Naïda et Sreyash. *Together through the maze of PhD studies, ups and (many) downs, it seems we finally made it!*

Ces remerciements ne peuvent évidemment pas se faire sans citer toutes les personnes qui ont largement contribué à mon équilibre personnel et sans qui je ne serai pas allé aussi loin. Je m'adresse en premier lieu à M. Benedict Schwegler, Professeur à Stanford et également membre invité du jury. Un remerciement en anglais, même s'il a la parfaite maîtrise de la langue de Molière. *Thank you for being such a source of inspiration, for boosting my ideas and following up my work, and thank you for opening the gates of DRC when I was just a young engineering student and absolutely no match to the people you used to work with. It was a life-changing opportunity (and I keep on thinking that the fact I was French played a major role in my recruitment).* Merci à Zach, *my friend from the New World, deeply missed in the forest (and also in life in general).* *Your time in Fontainebleau definitely was a huge help to blow off some steam!* Merci à ma tante, Cathy, d'avoir pris régulièrement de mes nouvelles, de m'avoir écouté et conseillé sur de nombreux sujets : il est difficile aujourd'hui d'avoir une oreille attentive et compréhensive accompagnée d'une parole honnête et franche. Merci également à mon père, terrorisé à l'idée que je ne finisse pas cette thèse et que je revienne vivre à la maison – comme si j'allais pouvoir le supporter – mais qui a toujours su être disponible malgré les épisodes difficiles traversés ces dernières années. Merci également à ma mère qui a su convaincre mon entourage aux moments opportuns que mon bien-être personnel devait passer avant certaines opportunités. Un grand merci également à Jo, mon grand-père, qui malheureusement n'est plus là aujourd'hui pour assister à cette réalisation. Malgré son caractère bien à lui, je lui dois une grande partie de la personne que je suis aujourd'hui.

Enfin, sur le podium des remerciements, ma petite famille. A égalité, la deuxième place pour Ayla et Moïra. Des dizaines d'heures de sommeil à récupérer, d'autres heures perdues et des crises de nerfs à réparer vos bêtises, une fortune dépensée en soins vétérinaires, mais un amour inconditionnel comme seuls un chat – qui n'est pas vraiment un chat dans sa tête – et un chien – qui n'est pas non plus vraiment un chien – peuvent en donner. Même si vous ne savez pas lire, merci d'être entrées dans ma vie.

Enfin, en haut de la liste et sans concurrence, Clara. Merci à toi de me supporter au quotidien – et je n’imagine pas à quel point ça doit être parfois difficile, moi-même je n’y arriverais pas. Merci pour ton soutien indéfectible depuis le début, de croire en moi, et malgré nos différences de m’accepter comme je suis. Ce furent (presque) quatre années très riches en émotions et malgré tout tu as toujours été là. Alors, tout simplement, merci.

Résumé

Le secteur du bâtiment, et sa part résidentielle notamment, figurent parmi les secteurs les plus énergivores. La rénovation énergétique est un des principaux axes de réduction des consommations d'énergie. Sa réalisation repose sur un outil clé : la modélisation énergétique du bâtiment, largement employée notamment au travers de calculs règlementaires mais qui fait face à un enjeu majeur que sont les écarts de performance énergétique. Cette thèse porte sur l'étude des écarts de performance énergétique dans le cadre de la rénovation d'un cas d'étude de trois bâtiments de logements sociaux. L'objectif est l'amélioration de la calibration des modèles énergétiques par une meilleure caractérisation des consommations d'énergie et des comportements à retombées énergétiques via la collecte de données de terrain. Cette collecte est effectuée sur plus de trois ans par un réseau de capteurs comprenant 170 objets connectés, installés dans les parties communes et dans un échantillon de huit logements représentatifs. La collecte et l'analyse des données portent sur les consommations d'énergie, la qualité de l'environnement intérieur, les comportements des occupants et la météo locale. Des scénarios typiques d'exploitation des bâtiments sont extraits des données collectées et intégrés aux modèles énergétiques. Ils mettent en évidence les différences importantes entre l'exploitation réelle des bâtiments et les scénarios standardisés des études règlementaires. La calibration des modèles énergétiques permet de quantifier le poids de leurs différents paramètres sur les consommations d'énergie et contribue à améliorer leur précision au service d'une rénovation énergétique performante.

Mots clés: Bâtiment, Consommations énergétiques, Réseaux de capteurs, Modélisation, Usages, Rénovation énergétique

Résumé substantiel

Le secteur du bâtiment, et sa part résidentielle notamment, sont encore aujourd’hui parmi les secteurs les plus énergivores. En France, les bâtiments neufs bénéficient de l’application la Règlementation Thermique (RT), désormais appelée Règlementation Environnementale (RE). Cependant, une large partie du parc résidentiel français est constitué de bâtiments anciens, construits avant ou au tout début des premières réglementations thermiques et qui sont aujourd’hui loin de correspondre aux attentes en termes d’efficacité énergétique. Bien que la RE s’applique depuis 2007 aux actions de rénovations énergétiques menées sur des bâtiments existants, il subsiste un important potentiel de réduction des consommations d’énergie et des émissions de gaz à effet de serre dans le secteur du bâtiment. C’est dans ce contexte qu’un plan massif de réhabilitation énergétique a été lancé au niveau national afin d’encourager et de démultiplier les actions de rénovation performante.

Pour être implémentées, les actions de rénovation reposent sur un outil clé : la modélisation énergétique du bâtiment. Cet outil permet d’étudier dans un environnement numérique la consommation d’énergie d’un bâtiment, entre autres paramètres, afin d’en comprendre les principaux aspects, et notamment de tester des actions de rénovation énergétique pour quantifier leur impact. Néanmoins, la principale difficulté de la modélisation reste les écarts de performance entre les résultats des modèles et les consommations mesurées. En effet, tout l’enjeu repose sur l’implémentation d’un modèle qui reproduit fidèlement le comportement énergétique d’un bâtiment dans son ensemble : consommations d’énergie des différents postes, températures intérieures, confort thermique, niveau de détail des différents paramètres constituant le modèle, etc. C’est ce processus de fidélisation des modèles énergétiques, appelé calibration, qui est le sujet de recherche cette thèse.

L’objectif de ce travail est donc d’étudier les écarts de performance de bâtiments résidentiels existants dans un contexte de rénovation énergétique lourde, et de proposer une solution d’amélioration de la calibration des modèles énergétiques au travers de la collecte de données in situ via un réseau de capteurs.

La modélisation énergétique du bâtiment dans son ensemble est le sujet du premier chapitre de ce manuscrit. Une revue de la littérature est proposée sur trois aspects de la modélisation : les méthodes de modélisation, les écarts de performance et la calibration des modèles. Parmi les méthodes de modélisation de la consommation d’énergie des bâtiments, on s’intéresse aux méthodes dites *”bottom-up”*. On distingue alors les techniques basées sur la physique du bâtiment, celles basées sur les données et les méthodes

hybrides, qui combinent les deux approches précédentes. La première partie de ce chapitre décrit les principales techniques utilisées dans la littérature, leurs avantages et leurs inconvénients. Au vue des objectifs de la thèse, on attend des modèles énergétiques qu'ils soient les plus détaillés et les plus interprétables possible. Le niveau de détail dans un modèle garanti une flexibilité dans l'investigation des sources d'incertitude qui peuvent entraîner des écarts de performance. L'interprétabilité permet de s'assurer que les résultats des simulations et des tests sur différents paramètres mettront clairement en évidence les liens entre les variables d'entrée des modèles et les performances énergétiques des bâtiments étudiés. C'est pour ces raisons que les modèles basés sur la physique du bâtiment sont sélectionnés comme la solution la plus adaptée au projet. La seconde partie du chapitre présente le concept d'écart de performance énergétique et ses origines dans le contexte du bâtiment. On identifie communément trois sources d'écarts de performance qui correspondent aux phases de vie du bâtiment : la conception, la construction et l'exploitation. Dans le cas de la rénovation énergétique, les sources potentielles d'écarts de performance peuvent venir de ces trois aspects, aussi bien avant que pendant la rénovation. Dans le contexte de la thèse, on se focalise sur la phase d'exploitation post-rénovation. En effet, peu d'informations sont disponibles sur les bâtiments avant leur réhabilitation et le chantier de rénovation est conduit indépendamment du travail de recherche. Il n'est donc pas possible d'étudier les écarts de performance dans les phases de conception et de construction. Pour tenter de réduire les écarts de performance, on s'oriente alors vers une double stratégie : la calibration des modèles énergétiques au travers d'une meilleure connaissance de l'exploitation des bâtiments garantie par la collecte de données de terrain. La calibration est le sujet de la troisième partie de ce chapitre. La revue de littérature met en évidence deux aspects importants que sont les sources d'incertitudes que la calibration peut viser et les techniques de calibration. Dans le cas présent, sans négliger l'ensemble des sources d'incertitudes dans un modèle énergétique, on s'intéresse principalement aux incertitudes dites de scénarios, qui décrivent l'exploitation des bâtiments au travers de différents paramètres tel que l'occupation, les températures de consigne de chauffage ou la caractérisation des apports de chaleur internes, par exemple.

La stratégie de calibration des modèles énergétiques proposée repose ainsi sur deux points importants : un terrain d'expérimentation existant et une solution d'instrumentation exhaustive pour obtenir suffisamment de données et mener la calibration des modèles à bien. Le second chapitre de la thèse décrit le cas d'étude et l'expérimentation. Il s'agit de trois bâtiments résidentiels comprenant 63 logements sociaux et construits en 1974. Les trois bâtiments ont été entièrement rénovés entre l'été 2020 et l'été 2021. Le réseau de capteurs sans fil déployé dans ces bâtiments est une solution

d'instrumentation IoT ("*Internet of Things*") reposant sur le protocole de communication radio LoRaWan ("*Long Range Wide Area network*") et le protocole GPRS ("*General Packet Radio Service*"). Le réseau de capteurs s'intéresse à trois aspects en lien avec les performances énergétiques des bâtiments : les consommations d'énergie (thermique, électrique et gas naturel), la qualité de l'environnement intérieur (température des parois froides, température intérieure, humidité relative, concentration de CO₂, luminosité), et les comportements des occupants (ouverture des fenêtres, présence). Les mesures sont effectuées dans les parties communes des bâtiments, sur les systèmes énergétiques et dans un échantillon de huit logements ayant accepté de participer à la campagne d'instrumentation. Au total, 170 capteurs sont installés sur site, dont 26 dans les parties communes et sur les systèmes énergétiques, et 144 dans les logements, soit une moyenne de 18 capteurs par logement. Une station météo est également déployée sur le campus à trois kilomètres du cas d'étude. Elle mesure la température de l'air, l'humidité relative, la pluviométrie, l'irradiation solaire, la vitesse et la direction du vent, la pression atmosphérique, la température radiante et la température de rosée. L'instrumentation du cas d'étude permet la collecte d'une grande quantité de données et la caractérisation de paramètres variés. L'implémentation et la supervision de la solution sur une période de plus de trois ans apporte également un retour d'expérience conséquent sur l'utilisation des solutions IoT pour le suivi des consommations énergétiques des bâtiments résidentiels occupés. Plusieurs solutions sont finalement mises en lumière sur des aspects tels que les mesures et les technologies des capteurs utilisés, la calibration de ces capteurs, et les données manquantes, en vue d'une amélioration et d'une répliquabilité de la solution pour de futurs projets.

L'analyse des données de terrain est le sujet du troisième chapitre de la thèse. Une quantité importante de données a pu être collectée via le réseau de capteurs. L'accent est mis sur certains paramètres spécifiques qui caractérisent l'exploitation des bâtiments, à savoir les températures intérieures, l'occupation des logements, les puissances électriques dissipées, les consommations d'eau chaude sanitaire et l'ouverture des fenêtres. La première étape est la préparation des données, avec le nettoyage, la mise en forme et l'évaluation de la qualité des données, leur reconstruction partielle, leur agrégation et la transformation de certains paramètres. Ces étapes sont importantes pour obtenir des données qui puissent être utilisées pour des analyses fiables. Elles fournissent également une vision d'ensemble sur les données collectées et les performances du réseau de capteurs. Les analyses portent sur l'extraction de tendances qui peuvent résumer l'évolution de ces principaux paramètres représentatifs de l'exploitation des bâtiments. Les analyses s'appuient sur plusieurs outils graphiques ainsi que sur une application de classi-

fication de profils journaliers. Plusieurs résultats marquants sont mis en lumière. Il apparaît notamment que les températures intérieures sont particulièrement élevées, avec des moyennes entre 21,4°C et 24,9°C sur l'échantillon instrumenté. Il n'y a pas non plus de régulation apparente du système de chauffage, les températures étant constantes sur la journée et en fonction des différentes typologies de jours. Les puissances dissipées, les scénarios d'occupation et l'utilisation d'eau chaude sanitaire mettent en avant la diversité des profils d'usages. Il apparaît clairement que chaque appartement montre des profils spécifiques à ses occupants. Il en résulte une importante différence entre les scénarios issus de données et les scénarios standardisés extraits des logiciels de modélisation énergétique. Ces scénarios sont notamment ceux utilisés pour les études thermiques réglementaires qui sont à l'origine de l'évaluation avant-projet de l'efficacité des actions de rénovation énergétique.

Le quatrième et dernier chapitre de la thèse présente l'étude des écarts de performance énergétique sur le cas d'étude, et qui repose sur les résultats et conclusions des trois précédents chapitres. On s'intéresse à la consommation de chauffage qui est le premier poste de consommation énergétique des bâtiments résidentiels en France. La calibration des modèles est une calibration manuelle de niveau quatre (utilisation de données collectées sur plusieurs mois), effectuée en trois étapes. Dans un premier temps, des modèles énergétiques non calibrés sont développés pour chaque bâtiment. Ces modèles sont développés comme pour une étude thermique réglementaire avec la description de l'enveloppe et des systèmes énergétiques après rénovation. Les scénarios d'exploitation sont standardisés. Des études de sensibilité sont également menées pour mettre en évidence et quantifier l'impact des paramètres d'entrée des modèles. L'écart de performance initial est évidemment important et les modèles montrent une sous-consommation de chauffage dans les bâtiments. Si la description des bâtiments est sensée être réaliste, leur exploitation réelle est différente de ce que les scénarios de modélisation initiaux proposent. Les études de sensibilité montrent également l'impact majeur des températures de consigne, assimilées aux températures intérieures, sur la consommation de chauffage. Dans un second temps, les paramètres d'entrée ayant le plus d'impact sur les consommations énergétiques et caractérisant l'exploitation des bâtiments sont intégrés aux modèles énergétiques, à partir des résultats d'analyse des données de terrain du Chapitre 3. Cette intégration permet une calibration paramètre par paramètre. Les températures intérieures, les puissances dissipées et l'occupation des logements sont traitées séparément, puis combinés. Au vue de l'impact des températures intérieures sur les consommations d'énergie, ce paramètre est également extrapolé pour les logements non instrumentés. Les consommations d'eau chaude sanitaire sont quant à elles éliminées car elles n'ont

pas d'impact sur les consommations de chauffage. Enfin, l'ouverture des fenêtres est traitée séparément. Les données collectées pour ce paramètre permettent de quantifier les temps d'ouverture. Néanmoins il manque un paramètre important et non mesuré, qui est l'amplitude d'ouverture des fenêtres. Les différents tests de calibration montrent des résultats qui ne sont pas tous à la hauteur des attentes initiales. La calibration paramètre par paramètre confirme les résultats des analyses de sensibilité et le poids de paramètres spécifiques sur les consommations de chauffage. L'étude des ouvrants montre aussi l'importance majeure de ce paramètre et du comportement des occupants sur les performances énergétiques des bâtiments. L'intégration des données apporte donc bien une meilleure connaissance des bâtiments étudiés. Cependant, l'écart de performance pour la calibration combinant tous les paramètres ajustés et l'extrapolation des températures intérieures n'est réellement réduit que pour le plus petit des bâtiments. Les autres bâtiments montrent un écart de performance quasi équivalent à celui avant calibration, mais cette fois en surconsommation. Ces résultats mettent en lumière plusieurs points importants et notamment sur les nombreuses incertitudes résiduelles dans les modèles. Ces incertitudes concernent en particulier les hypothèses utilisées pour remplacer les données manquantes et les caractéristiques des bâtiments pour lesquelles les informations n'ont pas pu être obtenues. Elles se rapportent aussi à la diversité des usages mis en évidence dans les analyses de données et qui interrogent la stratégie d'extrapolation employée. La représentativité des mesures effectuées est également remise en question et souligne certaines limites de la solution d'instrumentation. Néanmoins, ces résultats ouvrent des perspectives prometteuses sur les futurs axes de recherche et d'amélioration de la solution proposée, notamment sur l'utilisation d'objets connectés pour la collecte de données de terrain au service d'une rénovation énergétique performante.

Summary

Buildings, and specifically residential buildings, still remain today among the most energy-consuming sectors. Building energy retrofit is one of the major line of actions to improve building energy efficiency. It relies on a key tool, that is building energy modeling. Building energy modeling is widely used, particularly for regulatory calculations. However, it faces a significant challenge with the energy performance gap, that is the difference between simulation results and measured energy consumption. The present thesis focuses on the study of the energy performance gap over in a context of deep energy retrofit of three existing buildings of social housings. The goal is to improve the calibration of building energy models through an enhanced knowledge of building energy consumption and energy-driving behaviors, using field data collection. Data collection is performed over more than three years, using a wireless sensor networks comprising 170 connected objects. Sensors were installed in common areas, on energy systems and in an eight-housing sample. The electrical and thermal energy consumption, the indoor environment quality, the occupants' behavior and the local weather were monitored. Data analyses are performed to extract typical building operation patterns, and are integrated into building energy models. These operation scenarios highlight the significant difference between field data and standard building operation as in regulatory studies. The calibration of energy models underlines the impact of their different input parameters, and paves the way towards an efficient building energy retrofit.

Key words: Building, Energy consumption, Sensor networks, Modeling, Usages, Energy retrofit

Table of contents

- List of Figures** **xxi**
- List of Tables** **xxvii**
- Acronyms** **xxxix**
- Glossary** **xxxv**
- General introduction** **1**
 - Building energy efficiency in a climate change context 2
 - Thermal renovation and building instrumentation: addressing the gaps of thermal regulations 4
 - Thesis goals and content 9
 - Bibliography 11
- Chapter 1 Building energy modeling: methods and challenges** **13**
 - Introduction 14
 - 1.1 A few definitions 15
 - 1.2 Approaches and techniques for building energy modeling 15
 - 1.2.1 Classification of bottom-up modeling techniques 15
 - 1.2.2 Metrics for modeling performance assessment 17
 - 1.2.3 Physics-based building energy modeling 19
 - 1.2.4 Data-driven building energy modeling 22
 - 1.2.5 Hybrid modeling 29
 - 1.2.6 Conclusions on the applications of approaches and methods for building energy modeling 30
 - 1.3 Performance gap in building energy modeling 33
 - 1.3.1 Context and definitions 33

TABLE OF CONTENTS

1.3.2	Origins of performance gaps	34
1.3.3	Strategies to address the energy performance gap	36
1.3.4	Conclusions on energy performance gap	40
1.4	Calibration of building energy models	40
1.4.1	Identifying uncertainties	40
1.4.2	Calibration methods	41
1.4.3	Conclusions on the calibration of building energy models	43
	Conclusions	43
	Bibliography	44
Chapter 2	A sensor network for building energy monitoring	59
	Introduction	60
2.1	Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects	61
2.1.1	Industrial programs	62
2.1.2	Academic research projects	64
2.1.3	Discussions	67
2.2	Case study: existing, occupied and retrofitted residential buildings of social housings	68
2.2.1	General description	68
2.2.2	Summary of building characteristics and retrofit actions	70
2.2.3	Description of housings	71
2.3	Wireless sensor network	71
2.3.1	Targets and deployment	71
2.3.2	Preparing the measurement campaign	73
2.3.3	Data communication and storage	75
2.3.4	Measurements	76
2.4	Discussions	89
2.4.1	Feedback on the implementation and operation of the sensor network	89
2.4.2	Limitations and potential improvements	90
	Conclusions	95
	Bibliography	96
Chapter 3	Processing and analyses of field data for building energy modeling enhancement	101
	Introduction	102
3.1	Materials for data processing and analysis	103

TABLE OF CONTENTS

3.2	Data pre-processing	103
3.2.1	Cleaning and formatting	103
3.2.2	Data quality assessment, reconstruction and aggregation	104
3.2.3	Processing of datasets	108
3.3	Data analyses for building energy modeling enhancement and calibration .	109
3.3.1	Analysis methods	109
3.3.2	Towards a better understanding of building operation and usages .	116
	Conclusions	134
	Bibliography	139
	Chapter 4 Energy modeling, calibration and performance gap	141
	Introduction	142
4.1	Used modeling methods	143
4.1.1	Used tools: software and documentation	143
4.1.2	Process for building energy modeling, calibration and performance gap assessment	144
4.1.3	Sensitivity analysis	145
4.1.4	Description of implemented building energy models	146
4.2	Modeling results and discussions	152
4.2.1	Heating energy consumption: measurements	152
4.2.2	Non-calibrated building energy models	155
4.2.3	Performance gap and calibration	160
4.2.4	Practical conclusions on heating energy performances for the stud- ied buildings	173
	Conclusions	175
	Bibliography	177
	General conclusions	179
	Appendix A Tools for building energy modeling	185
A.1	Physics-based energy modeling	185
A.2	Data-driven energy modeling	189
	Bibliography	190
	Appendix B Sensor network for building energy monitoring	193
B.1	Connected objects and communication protocols for building energy mon- itoring	193

TABLE OF CONTENTS

B.1.1	The Internet of Things	193
B.1.2	Communication protocols	194
B.2	Calibration checking of IEQ sensors for housings	198
	Bibliography	200
Appendix C Data processing and analyses		203
C.1	Data quality assessment	203
C.2	Details of clusters for IAT in apartments	205
C.3	Density curves for IAT in apartments	208
C.4	Details of clusters for dissipated electric power in apartments	211
C.5	Details of clusters and mapping for window opening	215
C.6	Details of cluster mapping for DHW in apartments	220

List of Figures

- 1 Distribution of final the energy consumption in France by sector from 1990 to 2016 – Translated and adapted from [1]. 2
- 2 Distribution of energy sources in the final energy consumption for residential buildings (a) and tertiary buildings (b) from 2000 to 2017 – Translated and adapted from [2]. 3
- 3 Distribution of greenhouse gas emissions in the different sectors in France in 2016 – Translated and adapted from [1]. 4
- 4 Share of end-uses in energy consumption for residential buildings (a) and tertiary buildings (b) in France in 2020 [3, 4] 5

- 1.1 Classification of building energy modeling approaches and techniques. . . . 17
- 1.2 Summarized process for physics-based building energy modeling. 20
- 1.3 Data-driven building energy modeling process. 23

- 2.1 Neighborhood plan and pictures of the facades of buildings before the retrofit. 69
- 2.2 Deployment schedule of the instrumentation solution. 73
- 2.3 Description of the sensor network from sensors to data storage with three layers: the sensing layer includes all meters and sensors, the communication and collection layer relates to gateways and data processing platforms from Objenious [25] and The Things Network [26], and the storage layer groups all FTP and HTTP storage servers. 77
- 2.4 Pulse sensor installed on a Linky meter (a) and sensor with clamp ammeters for electricity sub-metering on a electrical switchboard (b). 82
- 2.5 Schematic of a thermal energy meter for heating or ‘DHW energy metering – Adapted from [33]. 84
- 2.6 Installation of temperature sensor for DHW (a) and a heater (b) in apartments. 85
- 2.7 Pulse sensor for natural gas Gazpar meter. 86

LIST OF FIGURES

2.8	Temperature and humidity sensors installed in the buildings shared areas. .	87
2.9	Infrared presence sensor in buildings shared areas.	88
3.1	Processing of DHW temperature measurements into volume consumption DP.	110
3.2	Tools for descriptive statistical analyses: (a) time series graph, (b) box-plot, (c) density plot, (d) cumulative frequency curve, (e) scatter plot, and (f) heatmap – displayed data are IAT measurements in B2/0 from 2021/10/20 to 2022/05/10.	111
3.3	Schematic of the clustering process for DP classification.	114
3.4	Temperature profiles (°C) from Th-BCE calculations in Pléaides software. .	117
3.5	Statistical description of temperature measurements in instrumented apartments.	118
3.6	Cumulative frequency curves of measured temperatures in instrumented apartments.	118
3.7	Results of clustering on IAT in B2/0, with daily profiles in cluster 1 (a) and cluster 2 (b), and calendar mapping of clusters (b) – the black bold line is the mean profile of each cluster; the white cells in the heatmap are missing DP.	119
3.8	Density curves for B2/0, comparing IAT measurements for day time and night time (a), and during weekdays and weekends (b).	120
3.9	Mean clustered occupancy profiles and calendar mapping for B1/2 (a) and B1/3 (b).	122
3.10	Mean clustered occupancy profiles and calendar mapping for B2/0 (a) and B2/5 (b).	123
3.11	Mean clustered occupancy profiles and calendar mapping for B3/0 (a) and B3/2 (b).	124
3.12	Occupancy profiles from Th-BCE calculations in Pléaides software.	125
3.13	Initial and corrected occupancy profiles for night time in B1/3 – Dotted lines in the morning and the evening show the initial occupancy profiles, red lines figure the applied correction.	126
3.14	Clustering results for electric dissipated power in B1/2 – the mean profile of each cluster is shown with a black bold curve.	127
3.15	Clustering results for electric dissipated power in B2/1.	128
3.16	Window opening profiles and calendar mapping in the living room of B2/2 – The averaged window opening DP illustrate between summer (in orange) and winter (in blue).	130

LIST OF FIGURES

3.17	Window opening profiles and calendar mapping in the bedroom 1 of B2/0 – Regardless of the time of the year, the window is always opened.	131
3.18	Window opening profiles and calendar mapping in the living room of B2/1 – The opening peak at the end of the day in summer (in orange) may be related to the orientation of the room or the working hours of the occupant.	132
3.19	Window opening profiles and calendar mapping in the bedroom of B1/3 – These window opening DP depict the housing-specific patterns.	132
3.20	Mean profiles for DHW use in B1/2 (a), B1/3 (b), B2/0 (c) and B3/0 (d) – B1/2 and B1/3 exhibit one main DP (in blue) while other profiles are outliers – B2/0 and B3/0 have two DP that are quite similar to each other.	134
3.21	Comparison of DHW consumption daily profiles from Pléaides software for a two-bedroom apartment (a) and B1/2 (b) – The comparison highlights the difference between measured field data and regulatory scenarios in terms of peak event timing and amplitude.	136
4.1	Three-step building energy modeling and calibration process with: Step 1 NC-BEM, Step 2 the integration of field-collected data, and Step 3 the calibration of energy models.	145
4.2	Three-dimension models of B1 (a), B2 (b) and B3 (c) extracted from Pléaides, compared with a photo of B3 after retrofit extracted from Google Earth (d).	147
4.3	Floor plans of building B1 – all buildings share similar architectural features for each floor.	148
4.4	Heating and DHW network for B1, B2 and B3 with thermal energy metering points (M1, M2 and M3).	150
4.5	Comparison of heating energy consumption from measurements and NC-BEM – On the left is the monthly evolution of heating consumption for B1 (a-1), B2 (b-1) and B3 (c-1), from measurements in red and from NC-BEM in brown – On the right is the relative error between models and measurements, for B1 (a-2), B2 (b-2) and B3 (c-2): the bar graph shows the monthly relative error in kWh and the purple line shows monthly relative error in %.	157
4.6	Process of data modification to test window opening configurations in building energy models – A maximum opening width (%) is used with a daily profile figuring window opening duration (a) to provide daily profiles with window opening width: here are examples with 20% (b), 40% (c) and 80% (d) maximum opening width.	162

LIST OF FIGURES

4.7	Manual calibration and window opening tests for B1, B2 and B3 – Five calibration tests are performed for each building with the tuning of the IAT, the occupancy, the dissipated electric power, the combination of the three parameters, and the additional extrapolation of IAT in non-instrumented housings – Three tests are conducted for each building on window opening with 20%, 40% and 80% maximum opening width – Results are compared with measurements and simulations from NC-BEM.	163
4.8	Comparison of heating energy consumption from measurements and calibration process for B1 (a), B2 (b) and B3 (c) over the heating season 2021/2022 – Measurements are figured in red, non-calibrated models are in brown and the tuning of inputs is depicted in grey, yellow, blue, green and dark blue for the tuning of IAT, dissipated power, occupancy, all parameters for instrumented housings, and the extrapolation of the IAT for non-instrumented housings, respectively – The purple line shows the relative error in % between simulations and measurements.	166
4.9	Monthly evolution and comparison of heating energy consumption from measurements and tuning of inputs for B1 (a), B2 (b) and B3 (c) – Non-calibrated models are in brown and the tuning of inputs is depicted in grey, yellow, blue, green and dark blue for the tuning of IAT, dissipated power, occupancy, all parameters for instrumented housings, and the extrapolation of IAT for non-instrumented housings, respectively – The red dotted line shows the measurements from our sensor network.	167
4.10	Tests of maximum opening width of windows with 20% (in beige), 40% (in orange) and 80% (in brown) maximum opening widths, compared with NC-BEM (in purple) for B1 (a), B2 (b) and B3 (c) – A total of eight, twelve and three sensors, for two, four and one instrumented housings are considered in B1, B2 and B3, respectively.	170
B.1	Measurements from IEQ sensors over five days for calibration checking. . .	199
C.1	IAT clusters for B1/2 (a) and B1/3 (b).	205
C.2	IAT clusters for B2/0 (a) and B2/5 (b).	206
C.3	IAT clusters for B3/0 (a) and B3/2 (b).	207
C.4	Density curves for IAT in B1/2 (a) and B1/3 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.	208

LIST OF FIGURES

C.5	Density curves for IAT in B2/0 (a) and B2/5 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.	209
C.6	Density curves for IAT in B3/0 (a) and B3/2 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.	210
C.7	Clusters and cluster mapping for dissipated electric power in B1/2 (a) and B1/3 (b).	211
C.8	Clusters and cluster mapping for dissipated electric power in B2/0 (a), B2/1 (b) and B2/2 (c).	212
C.9	Clusters and cluster mapping for dissipated electric power in B3/0 (a) and B3/2 (b).	213
C.10	Clusters for dissipated electric power from the second round of clustering, in B1/2 (a), B2/0 (b), B2/2 (c), B3/0 (d) and B3/2 (e).	214
C.11	Mean profile from clusters and cluster mapping for window opening in B1.	215
C.12	Mean profile from clusters and cluster mapping for window opening in B2.	217
C.13	Mean profile from clusters and cluster mapping for window opening in B3.	219
C.14	Cluster mapping for DHW use in B1/2 (a), B1/3 (b), B2/0 (c) and B3/0 (d).	220

LIST OF FIGURES

List of Tables

- 1 Evolution of the French thermal regulation from 1974 to 2020. 6
- 1.1 Review papers on building energy modeling published in the past 25 years. 32
- 2.1 Summary reviewed studies implementing sensor networks for data collec-
tion in operating buildings. 66
- 2.2 Summary of apartment features in the three considered buildings. 69
- 2.3 Description of the household sample that took part in the instrumentation
campaign. 70
- 2.4 Description of occupants in the instrumented households. 72
- 2.5 Details on the numbers and types of sensors in the different instrumented
households (with one apartment per floor). 78
- 2.6 Summary of the accuracy and operating range of the deployed sensors in
apartments and common areas. 79
- 2.7 Summary of accuracy and operating range of the weather station. 81
- 3.1 Sensors, data and thresholds for reconstruction. 106
- 3.2 Impact of data reconstruction and aggregation on daily profile completion
– statistics relate to the whole instrumentation period. 107
- 3.3 Averaged DHW daily needs per occupant and household (unit: liters, at
40°C temperature) – Source: [7]. 109
- 3.4 Used timeframes for data processing and analyses – In Bi/j, *i* stands for
the building number and *j* for the floor number. 115
- 3.5 Cluster summary for occupancy patterns in instrumented housings. 122
- 3.6 Cluster summary for dissipated electric power in instrumented housings. . 129
- 3.7 Summary of sensors for window opening detection. 129
- 3.8 Cluster summary for DHW in instrumented housings. 135
- 3.9 Comparison of DHW daily averaged consumption from ADEME and
Pléaides software. 135

LIST OF TABLES

4.1	Details of the number of thermal zones in B1, B2 and B3.	149
4.2	Measurements and assessments for heating energy consumption of B1, B2 and B3 and the heat loss between B2 and B1 and between B2 and B3, over the heating season from 2021/10/20 to 2022/05/10 – the gross floor areas of B1, B2 and B3 are 1,157 m ² , 1,806 m ² and 2,255 m ² , respectively.	155
4.3	Comparison between the results of the NC-BEM, and measurements and assessments from collected data, over the heating season 2021/2022 – errors are calculated over the whole heating season.	156
4.4	Results of local sensitivity analyses on NC-BEM for B1, B2 and B3 with relative impact (in kWh and kWh/m ²) and ranking of input parameters – The relative impact is presented in absolute value.	159
4.5	Results of step-by-step input calibration on heating energy consumption for B1, B2 and B3, over the heating season 2021/2022 – For each building, IAT, occupancy and dissipated power are tuned separately from each other, then they are tuned at the same time, and finally, the tuning of all three parameters is complemented with the extrapolation of IAT to non-instrumented housings.	165
4.6	Impact of window opening on the thermal energy consumption of B1, B2 and B3 over the heating season 2021/2022: 2, 4 and 1 instrumented housings with available window opening data are considered in these tests, out of total of 13, 21 and 29 apartments in B1, B2 and B3, respectively – Results show the heating energy consumption of the whole building – Window opening tests are conducted only for instrumented housings with available window opening data.	169
A.1	Summary of software for physics-based energy modeling – Technical characteristics.	186
A.2	Summary of software for physics-based energy modeling – Applications: case studies.	187
A.3	Summary of software for physics-based energy modeling – Applications: modeling outputs.	188
A.4	Summary of coding tools, packages and applications for data-driven energy modeling.	189
C.1	Summary of data collection rates over the instrumented buildings and housings.	204
C.2	Clustering summary for window opening in B1.	216

LIST OF TABLES

C.3 Clustering summary for window opening in B2.	218
C.4 Clustering summary for window opening in B3.	219

LIST OF TABLES

Acronyms

ANN artificial neural network

APC *Agence Parisienne du Climat*

AQC *Agence Qualité Construction*

ARIMA autoregressive integrated moving average

B1 Building 1

B2 Building 2

B3 Building 3

BBC *Bâtiment Basse Consommation*

BEC building energy consumption

BEM building energy models

BEMS building energy management system

BEST Building Energy Software Tools

BIM building information model

CEP *coefficient d’Energie Primaire* – primary energy factor

CFD computational fluid dynamics

CMV controlled mechanical ventilation

CNIL *Commission Nationale Informatique et Libertés*

CPCU *Compagnie Parisienne de Chauffage Urbain*

CSTB *Centre Technique et Scientifique du Bâtiment*

CSV comma-separated values

CVI cluster validation indices

Acronyms

DevEUI	device extended unique identifier
DHW	domestic hot water
DNN	deep neural network
DP	daily profiles
DT	decision tree
EDF	<i>Electricité De France</i>
FTP	file transfer protocol
GHG	greenhouse gas
GPRS	general packet radio service
HDD	heating degree days
HPE	<i>Haute Performance Energétique</i>
HTTP	hypertext transfer protocol
HVAC	heating ventilation air conditioning
IAQ	indoor air quality
IAT	indoor air temperature
IEQ	indoor environment quality
IoT	Internet of Things
JSON	javascript object notation
k-NN	k nearest neighbors
LoRaWan	long range wide area network
MAE	mean absolute error
MAPE	mean absolute percentage error
MARS	multivariate adaptive regression splines
MBE	mean bias error

MRE mean relative error

MSE mean square error

Mtoe Mega ton of oil equivalent

NC-BEM non-calibrated building energy models

NILM non intrusive load monitoring

NMBE normalized mean bias error

OAT outdoor air temperature

PCA principal component analysis

PG performance gap

PIR passive infrared

QSE *Qualité Sanitaire et Energétique*

RC resistance-capacitance

RMSE root mean square error

SEREINE *Solution d'Evaluation de la performance Energétique Intrinsèque des bâtiments*

SVM support vector machine

SVR support vector regression

THPE *Très Haute Performance Energétique*

TIC *Télé-Information Client*

Tic *Température intérieure conventionnelle*

TRS thermal regulatory study

VOC volatile organic compounds

Acronyms

Glossary

HDD_{local} The number of heating degree days from a given day and calculated from OAT data collected by the weather station.

$HDD_{simulation}$ The number of heating degree days from a given day and from the weather file used in the energy simulation.

$H_{simulated}$ The heating energy consumption over a given day and simulated from building energy models.

$H_{corrected}$ The heating energy consumption over a given day, simulated from building energy models and corrected with the $HDD_{weather,local}$.

\hat{y}_{hs} The energy consumption from simulations over the heating season.

y_{hs} The measured energy consumption over the heating season.

y_t In a given time series, y_t is a data point from the time series at time-step t .

\hat{y}_t In a given time series, \hat{y}_t is the corresponding predicted output to the data point y_t at time-step t .

A_{B1} The total heated area of B1.

A_{B2} The total heated area of B2.

A_{B3} The total heated area of B3.

H_{Bi} The energy consumption of the building Bi over the whole heating season.

$H_{day, Bi}$ The heating energy consumption of building Bi (with $i=\{1,2,3\}$) over a missing day of data.

H_{B1} The heating energy consumption of B1, measured in M1 on Figure 4.4.

H_{B2} The heating energy consumption of B2.

H_{B3} The heating energy consumption of B3, measured in M3 on Figure 4.4.

H_{tot} The total heating energy consumption of B1, B2, B3 and including the heat loss

Glossary

from B2 to B1 and from B2 to B3, measured in M2 on Figure 4.4.

HDD_{day} The amount of heating degree days over a missing day of data.

HDD_{heating season} The amount of heating degree days over the whole heating season.

L The heat loss through the piping network between B2 and B1 and between B2 and B3.

RE The relative error in %.

n In a given time series, *n* the number of data points in the time series.

t In a given time series, *t* is the time-step of the time series.

General introduction

This introduction gives an overview of contextual elements to put the research work presented in this manuscript into perspective. First, the position of the building sector as a major end-use for energy consumption and greenhouse gas emissions is described to highlight challenges and opportunities. Thereafter, key energy levers for building energy efficiency – namely, the thermal regulations, the thermal retrofit and building instrumentation plans – and synergies are discussed. Finally, the goals of this thesis and the frame of the manuscript are presented.

Building energy efficiency in a climate change context

The building sector has always been one of the main energy-consuming sectors in France. In 1990, it already accounted for 43% of the overall final energy consumption and 27% of the primary energy consumption [1]. It even increased up to 46% of the final energy consumption in 2017 (Figure 1) to reach 67 Mtoe (Mega ton of oil equivalent). Tertiary buildings (i.e. related to health services, teaching, hotels, restaurants, commercial activities, office spaces, sport and leisure, transport, and telecommunications) accounted for 25 Mtoe. Residential buildings alone consumed up to 42 Mtoe [2] and currently holds the second place in final energy consumption, after transportation (46 Mtoe, 32%), and followed by the industry (27 Mtoe, 19%), tertiary buildings (25 Mtoe, 16%) and agriculture (4 Mtoe, 3%).

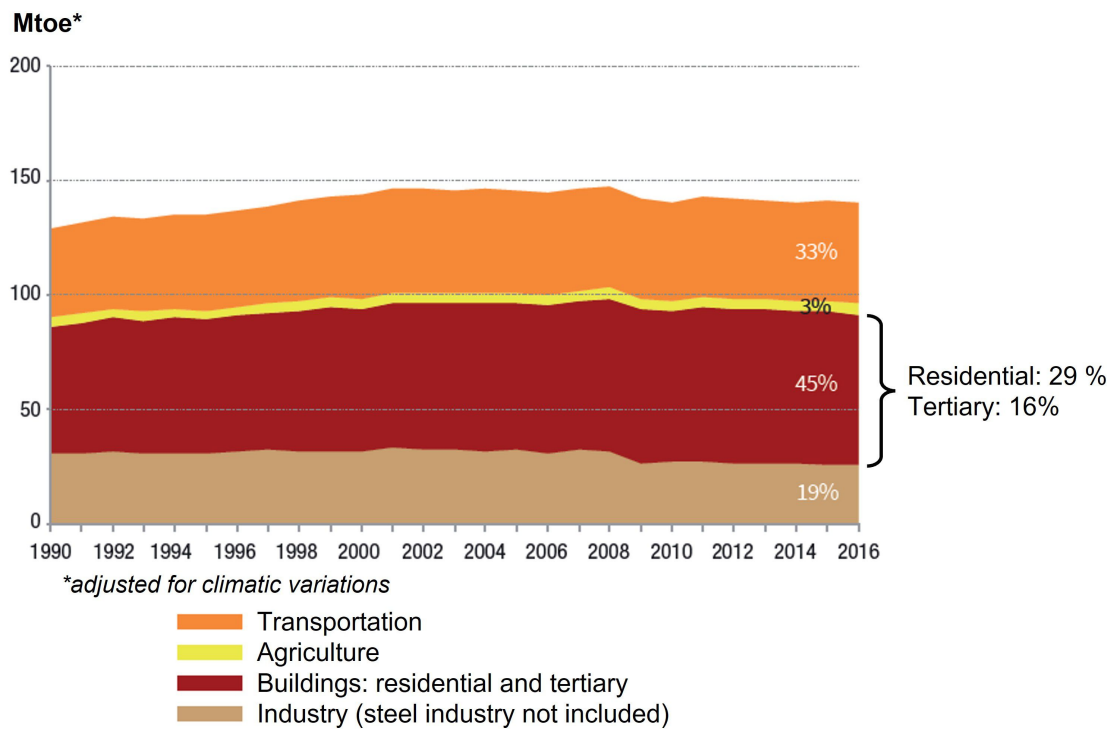


Figure 1: Distribution of final the energy consumption in France by sector from 1990 to 2016 – Translated and adapted from [1].

In terms of energy mix, for both residential and tertiary buildings in France, electricity is the first source of energy consumption (33% and 49%, in residential and tertiary buildings, respectively), followed by natural gas (28% and 31%, respectively) [2]. Residen-

Building energy efficiency in a climate change context

tial buildings have a much larger part of renewable (including electricity from renewable energy sources) and waste energy consumption than tertiary buildings (Figure 2), with 22% against 4%, respectively. Oil accounts for 13% in both sectors, followed by heating networks and co-generation plants (4%). Comparing the evolution of energy sources and their respective share in the energy consumption balance, oil consumption decreased in residential buildings to the benefit of renewable energy consumption. Nevertheless, natural gas and electricity consumption remained stable until today.

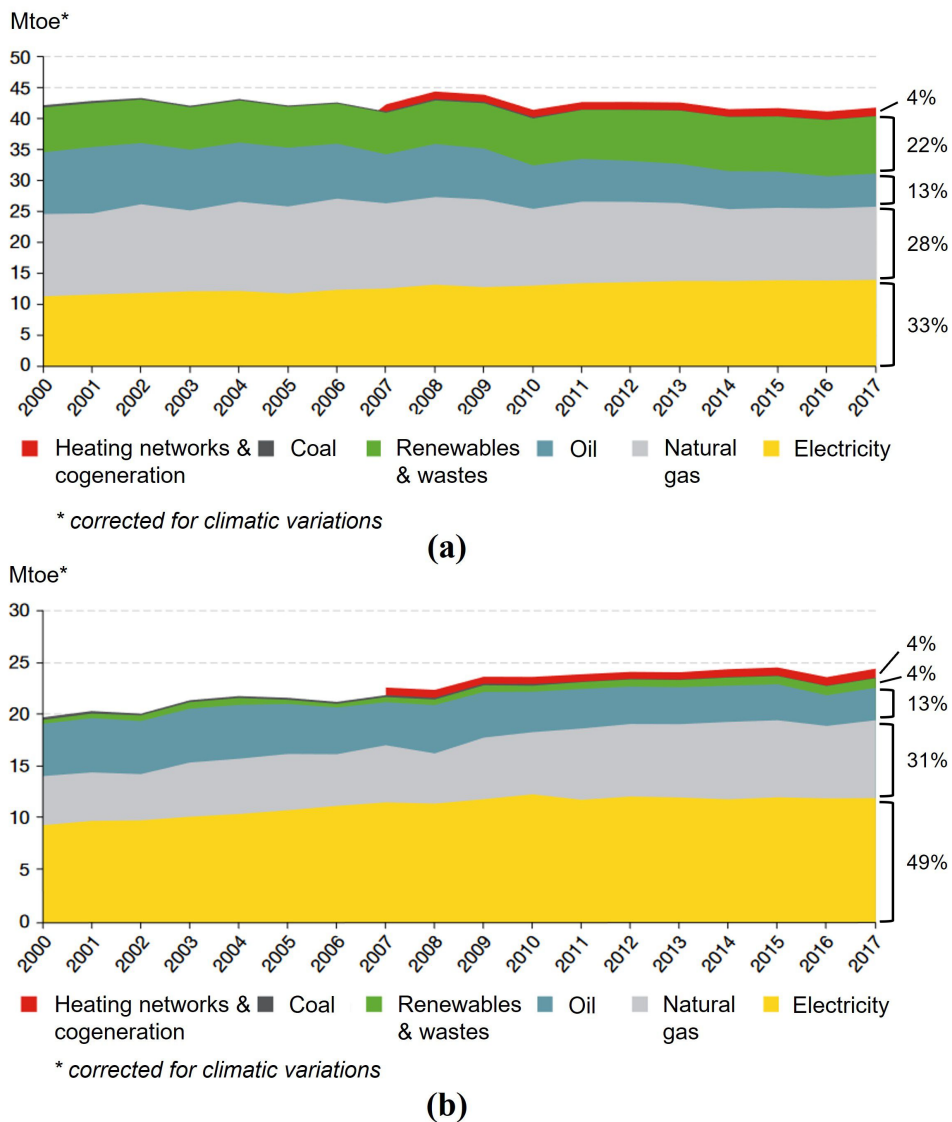


Figure 2: Distribution of energy sources in the final energy consumption for residential buildings (a) and tertiary buildings (b) from 2000 to 2017 – Translated and adapted from [2].

GENERAL INTRODUCTION

With respect to the large share of electricity in building energy consumption, and the origin of electricity production in France (63% from nuclear plants, 12% from hydro-electricity [2]), the building sector exhibits a low carbon energy mix that results in 15% and 7% of overall greenhouse gas (GHG) emissions in France [1], for residential and tertiary buildings, respectively (Figure 3).

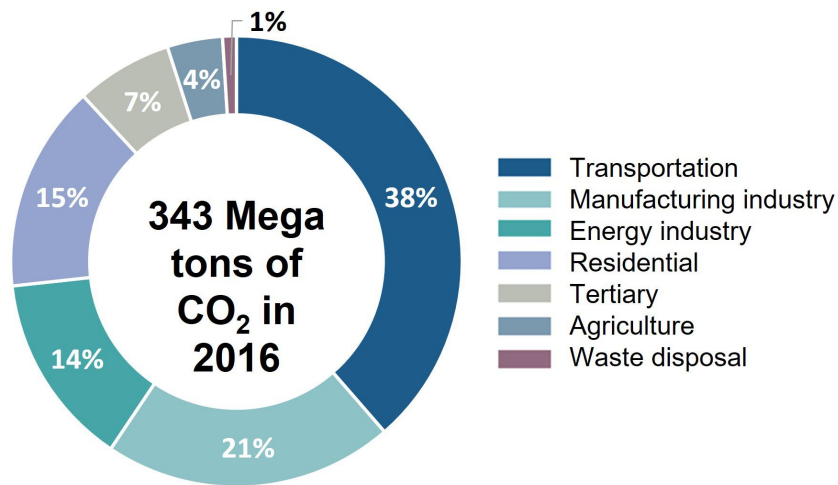


Figure 3: Distribution of greenhouse gas emissions in the different sectors in France in 2016 – Translated and adapted from [1].

The distribution of end-uses underlines differences between residential and tertiary buildings [3, 4] (Figure 4). In both sectors, heating is the main end-use (64% and 43% for residential and tertiary buildings, respectively), followed by specific electricity usage (18% and 26%) and domestic hot water (12% and 20%). However, cooling is almost nonexistent in residential buildings with 0.4% of the energy consumption, opposed to 11% tertiary buildings. Cooking accounts for 6% and 5%, respectively.

Considering the heavy weight of buildings in the energy and GHG emission balance, there are massive opportunities to improve their environmental footprint. Specifically, thermal energy consumption holds the largest share of consumption in both residential and tertiary buildings, and residential buildings alone account for almost a third of the total final energy consumption in France.

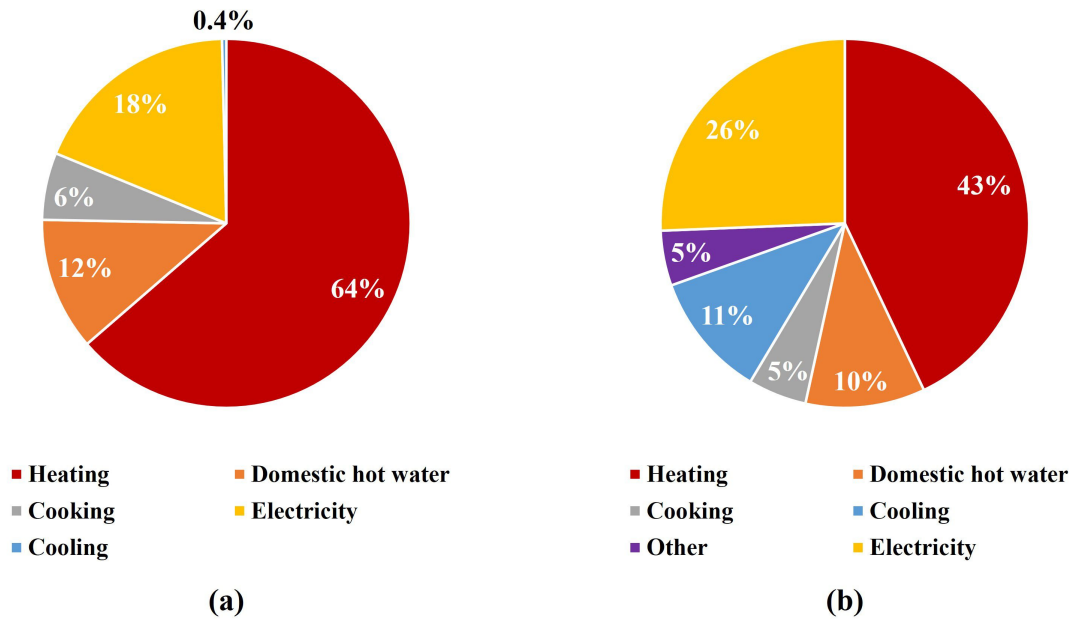


Figure 4: Share of end-uses in energy consumption for residential buildings (a) and tertiary buildings (b) in France in 2020 [3, 4]

Thermal renovation and building instrumentation: addressing the gaps of thermal regulations

Historically, the main plan of action towards building energy efficiency is the national thermal regulation – *“Règlementation thermique”* in French. It first started in 1974, followed by five regulation changes in 1988, 2000, 2005, 2012 and the latest in 2020, to gradually improve building energy efficiency and to enforce strict targets on different aspects of the building energy consumption [5]. Table 1 provides an overview of the milestones in the evolution of the French thermal regulation. Nowadays, the latest environmental regulation (RE2020) sets very strict standards not only in terms of energy consumption, but for greenhouse gas emissions and building carbon balance as well.

Nevertheless, the effects of thermal regulations are limited, first in terms of targets: new buildings. It is only from 2007 that existing buildings are considered through retrofit actions. Retrofit actions are not mandatory but must be conducted when an existing building is modified or an energy system is changed. In that case, the energy efficiency and thermal characteristics of retrofitted building parts and systems must meet the latest standards. Furthermore, focusing on residential buildings, in 2017, 51% of the 28.5 millions of main residences were built before 1975, then before the application of the first

GENERAL INTRODUCTION

Table 1: Evolution of the French thermal regulation from 1974 to 2020.

Date	Targeted building	Goal	Means of action	Coefficients	Labels
1974	Residential buildings (new and on-going projects)	Decrease BEC by 25% (225 kWh/(m ² .year))	Thermal insulation, improved heating management	G (W/m ³ /K) for building heat loss – G=[2.3;2.9] for housings under 150 m ²	/
1975	Residential buildings (new and on-going projects)	/	/	G=[1.6;2.0] for housings under 150 m ²	/
1976	All new heated buildings	Thermal insulation mandatory	/	/	/
1980	All new heated buildings	/	/	G=[1;1.1] if half of heating energy is electrical; G=[1.2;1.4] otherwise	<i>Haute isolation thermique</i>
1982	–	Decrease heating energy consumption by 20% (170 kWh/m ² /year)	Energy systems, thermal characteristics, ventilation	B (W/m ³ /°K) – includes solar and internal heat gains	/
1988	New residential and tertiary buildings	Including tertiary buildings; decrease heating and consider the whole energy consumption	Include DHW consumption and efficiency of energy systems	C (W/m ³ /°K) for consumption	
2000	New residential and tertiary buildings	Decrease global BEC by 20 % in residential buildings (130 kWh/m ² /year) and by 40% in tertiary buildings	Concept of summer comfort; setting of reference building characteristics and energy consumption	/	/
2005	New buildings and parts of buildings	Decrease global BEC by 15% (90 kWh/m ² /year)	Including heating, cooling, ventilation, DHW, lighting in energy balance; new reference values for coefficients; definition of climate zones; reference performances of building characteristics	U _{bat} (W/m ² /K); CEP; summer Tic (°C)	HPE; THPE; BBC
2007	Existing retrofitted buildings	Mandatory retrofit actions when construction work is done	Mandatory improvement of thermal insulation (partial or total) and efficiency of energy systems if any change occur	/	/
2012	All new buildings and existing retrofitted buildings	Aim for energy consumption of 50 kWh/m ² /year; improve air quality	Removing thermal bridges, improving air tightness, using at least one renewable energy source	/	/
2020	All new and existing retrofitted buildings	Improve thermal insulation, limit GHG emissions, zero energy consumption balance	New reference values for thermal characteristics, include life cycle analysis	/	/

Thermal renovation and building instrumentation: addressing the gaps of thermal regulations

thermal regulation [1]. Another 29% were built before 2000 and the renewal rate of the building housing stock today is of less 1% [6]. In other words, there are more old buildings that do not fit any thermal efficiency standard than new energy-efficient buildings. Therefore, the key to catch on with building energy efficiency goals is to enforce massive energy retrofit plans of actions.

To address this need, the Housing Energy Retrofit Plan (*"Plan de Rénovation Energétique de l'Habitat"*) [7] was launched in 2013. This national strategy targeted an ambitious planning of retrofit actions in the residential sector relying on several financial supports, with:

- Over 2014-2017: an annual rate of 180,000 private housings including 38,000 housings in an energy poverty situation and 90,000 social housings;
- From 2017: an annual rate of 380,000 private housings including 50,000 housings in an energy poverty situation and 120,000 social housings;
- Within ten years: the retrofit of 1.5 millions of the most poorly insulated buildings.

Today, the national retrofit plan exhibits mixed results. Over 5 millions individual houses have reported energy retrofit actions [1]. Although it looks like a significant number, it does not mean that the full potential of renovation has been achieved. Indeed, there is a large diversity of retrofit actions, and they do not all have the same impact on building energy efficiency. Also, the review of preliminary results highlighted several issues [8]. Overall, the renovation did increase energy savings in the residential sector. However, in some cases, specific actions did not have the intended effect or even had a negative impact on building energy consumption – for instance, when a partial retrofit is conducted without considering the full building picture in the energy balance. Finally, benefits of the national retrofit plan mostly affected individual housings and private housing owners. Therefore, there is yet a lot of work to be done to target these objectives and keep on supporting the retrofiting effort.

Aside from their targets in terms of buildings, thermal regulations also only progressively took into account the energy systems. At first, they focused on thermal insulation. Although the energy balance is now more detailed, several appliances are still excluded, such as leisure equipment, small and middle-size domestic appliances, or additional portable heaters. These appliances result in significant heat gains but can only be considered if they are accurately surveyed and characterized. Then, it becomes an enormous task among many others in building energy studies – assuming that the full and detailed list of appliances can be available [9]. This also relates to the integration of

GENERAL INTRODUCTION

usages and occupants' behaviors. The management of building energy system is usually automated based on specific physical parameters (e.g. the weather, a theoretical occupancy, the time of the day). However, there is often a large difference between standard building operation and the actual occupants' behavior, including window opening for natural ventilation, manual changes over specific systems, or the use of additional appliances. Such a difference impacts on building energy efficiency in unexpected ways [10].

Hence, alongside of the national retrofit campaign, two directives were implemented. The "*Dispositif Eco Efficacité Tertiaire*" also called "*Décret Tertiaire*" – tertiary ordinance, in English – was enacted in 2019 to target an ambitious 60%-decrease of energy consumption in tertiary buildings (for buildings with an area greater than or equal to 1000 m²) [11]. It also enforced energy consumption data collection to create national databases. Consequently, this directive significantly boosted the building instrumentation market, specifically building energy management systems (BEMS), to centralize and manage all aspects of building operation through dedicated systems and platforms [12]. It was preceded by a complementary European directive that aspired to a massive implementation of communicating meters for electricity and natural gas consumption in residential and small tertiary buildings. In France, these meters are called Linky for electricity [13] and Gazpar for natural gas [14]. Linky meters rely on power line carriers for data communication from 10-minute to one hour measurement time-step. Gazpar meters use radio frequencies (169 MHz) for data communication, with two measurements per day. Initial deployment goals were ambitious, with a target of 200 millions meters for electricity and 45 millions meters for natural gas by 2020 – covering 72% and 40% of European energy end-users, respectively [15]. However, only 99 millions electricity meters and 16 millions gas meters were installed by that time [16]. It led to a target review: 123 millions meters by 2020, 225 millions meters by 2024 (77% of the metering points in Europe) for electricity, and 31 millions meters by 2020 and 51 million meters by 2024 (44% of the metering points in Europe) for natural gas. In France, Linky meters were expected up to 35 millions units by 2021 (the number of corresponding metering points at that time): 31.2 millions were actually installed. Gazpar program aimed for 11 millions units by 2022 and it seems to be on track due to a different deployment strategy. Indeed, the installation of Gazpar meters is mandatory unlike for Linky, that recently opted for this strategy as well.

Despite expected deployment delays, smart metering systems bring in a significant benefit over traditional metering systems. Data collection is easier to conduct and with a much smaller time-step. It provides a national overview, almost in real time, of the electricity and natural gas consumption. Energy production management is eased and the

prediction of energy needs is more accurate. Smart metering plans are a great asset for retrofit actions as well [17]. Instead of collecting monthly billing information, Linky and Gazpar meters offer a much more detailed source of data regarding energy consumption and its dynamics with a fine data temporal granularity. Moreover, massive national smart metering plans are complemented with the widespread use of new information and communication technologies, including the Internet of Things (IoT) connected objects, to collect other types of data such as indoor temperature, humidity or electricity use of appliances [18].

Thesis goals and content

As highlighted in the previous sections, the building sector, and specifically residential buildings, remain among the largest energy-consuming sectors. Several national and European tools have been implemented over the past five decades to face this challenge, starting with thermal regulations and followed by national energy retrofit and instrumentation plans. Considering that the largest part of the building housing stock was built before the most recent thermal regulations and that the actual renewal rate is extremely low, massive retrofit actions are a priority. However, the efficiency of thermal renovation conducted so far does not fully match the expectations because of a significant lack of knowledge on building operation and the many parameters involved. Then, national smart metering plans and the recent development of new information and communication technologies, with the IoT, can significantly contribute to ensure an efficient thermal retrofit of buildings.

Based on these observations, the ANDRE research project – *“modèles pour l’ANalyse, la Décomposition et la REconstruction de données de consommations énergétiques”* / models for the analysis, disaggregation, and reconstruction of energy consumption data, in English – was jointly launched by six founding members: four laboratories from Université Gustave Eiffel with ESYCOM, COSYS LISIS, GRETTIA and Lab’URBA to respectively provide knowledge and skills on building energy modeling and instrumentation, calibration of energy models, statistical data analyses and socioeconomic aspects of building energy consumption. They were complemented by a social landlord, *Marne-et-Chantereine Habitat*, to provide a case study undergoing retrofit actions, and a company from the energy efficiency sector, CAMEO SAS, to bring in experimental skills and field expertise. The research goals were to combine building instrumentation, building energy modeling and data analyses from physics, statistical and socioeconomic perspectives, in order to assess the performance of retrofit actions over an existing residential case study. The research

GENERAL INTRODUCTION

presented in this thesis focuses on the three aspects of the project from a physics point-of-view: building energy modeling and calibration of models based on the deployment of a dedicated sensor network for data collection and analyses over an existing building case study.

Consequently, the first chapter of the manuscript reviews the topic of building energy modeling. The first part of the chapter is dedicated to building energy modeling approaches and techniques. It is followed by a study of the related challenges, with discussions on the energy performance gap – i.e. the difference between the results of energy models and the measured energy consumption – and the calibration of energy models.

The second chapter introduces the case study on which the thesis relies and the experimental set-up for field data collection. Three social apartment buildings undergoing a deep energy renovation are considered. A wireless sensor network of 170 sensors is deployed in common areas, on energy systems and in an eight-apartment sample to characterize the local weather, the energy consumption, the indoor environment quality and occupants' behavior. The architecture of the sensor network is described along with specifications of the measurements, data communication and storage. Then, a feedback regarding the management of the instrumentation solution summarizes its benefits and limitations, and discusses its replicability for future similar experimental projects.

The third chapter is based on the results of the sensor network. In the first place, collected data are pre-processed. Pre-processing includes six steps: cleaning (the identification and correction of errors), formatting, data quality assessment, reconstruction, aggregation and dedicated processing of specific parameters. Pre-processing is followed by data analyses presented in the second part of the chapter. Results of data analyses intend to be used to enhance building energy models. Hence, analyses focus on data collected in apartments and specific drivers of the building energy consumption. The goal is to characterize the dynamics of five parameters, including indoor air temperature, occupancy, dissipated electric power, window opening and domestic hot water consumption. These parameters are representative of the three aspects addressed in this thesis: energy consumption, indoor environment quality and usages. Parameters are processed using descriptive statistical tools and automatic classification tools to extract typical patterns.

In the fourth and final chapter, we investigate the performance of the conducted energy retrofit actions through the implementation and calibration of building energy models. To this end, a three-step process is applied focusing on heating energy consump-

tion that accounts for the largest share of energy consumption in residential buildings. The chapter starts with a description of the overall modeling process. Then, we discuss simulation results from non-calibrated building energy models that are implemented with standard information on building operation as for a regulatory thermal study. The initial performance gap – the difference between non-calibrated energy simulations and measured heating energy consumption – is investigated. Then, in the third part of the chapter, we propose a strategy to close the performance gap. Energy-driving parameters are selected out of a sensitivity analysis and tuned with the results of data analyses from the third chapter. Different tests are conducted to picture the impact of input tuning on the energy models. This chapter is concluded with a discussion on the final energy performance gap to further address the remaining uncertainties in building energy models.

Bibliography

- [1] ADEME, *Climat, air et énergie — Chiffres clés*. 2018. xxi, 2, 4, 7
- [2] Ministère de la Transition Ecologique et Solidaire, “Chiffres clés de l’énergie. Édition 2019,” pp. 1–80, 2019. xxi, 2, 3, 4
- [3] Ministère de la Transition Ecologique et de la Cohésion des Territoires, “Consommation d’énergie par usage du résidentiel — Données et études statistiques.” <https://www.statistiques.developpement-durable.gouv.fr/consommation-denergie-par-usage-du-residentiel>. xxi, 4, 5
- [4] Ministère de la Transition Ecologique et de la Cohésion des Territoires, “Consommation d’énergie par usage du tertiaire — Données et études statistiques.” <https://www.statistiques.developpement-durable.gouv.fr/consommation-denergie-par-usage-du-tertiaire>, 2021. xxi, 4, 5
- [5] EDF, “Il était une fois les réglementations thermiques...” <https://particulier.edf.fr/fr/accueil/economies-d-energie/construction-et-renovation/la-reglementation-thermique-2012/il-etait-une-fois-les-reglementations-thermiques>. 5
- [6] PACTE, *Rapport - Analyse détaillée du parc résidentielle existant*. 2017. 7
- [7] ADEME, “Le plan de rénovation énergétique de l’habitat (PREH),” tech. rep., 2018. 7
- [8] Ministère de la Transition Ecologique et de la Cohésion des Territoires, “La rénovation énergétique des logements : bilan des travaux et des aides entre 2016 et 2019 — Résultats définitifs — Données et études statistiques,” 2022. 7

BIBLIOGRAPHY

- [9] P. Roques, *La question de la consommation d'énergie dans les logements sociaux réhabilités - Pratiques et identité*. PhD thesis, Université Côte d'Azur, 2016. 7
- [10] A. L. Pisello and F. Asdrubali, "Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies," *Applied Energy*, vol. 133, pp. 224–235, nov 2014. 8
- [11] Ministère de la Transition Écologique et Solidaire, "Décret n° 2019-771 du 23 juillet 2019 relatif aux obligations d'actions de réduction de la consommation d'énergie finale dans des bâtiments à usage tertiaire," 2019. 8
- [12] ADVIZEO, "Réduire la consommation du parc tertiaire : un enjeu pour la transition." <https://www.advizeo.io/blog/decret-tertiaire/reduire-la-consommation-du-parc-tertiaire-un-enjeu-majeur-de-la-transition-energetique/>, 2022. 8
- [13] ENEDIS, "Linky, le compteur communicant." <https://www.enedis.fr/linky-compteur-communicant>. 8
- [14] GRDF, "Compteur gaz communicant Gazpar : comment ça marche?." <https://www.grdf.fr/particuliers/fonctionnement-compteur-gaz-communicant-grdf>. 8
- [15] European Commission — Joint Centre — Smart Electricity Systems and Interoperability, "Smart Metering deployment in the European Union." <https://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union>, 2014. 8
- [16] F. Tounquet and C. Alaton, *Benchmarking Smart Metering Deployment in EU-28*. No. December, 2020. 8
- [17] L. Jankovic, "Lessons learnt from design, off-site construction and performance analysis of deep energy retrofit of residential buildings," *Energy and Buildings*, vol. 186, pp. 319–338, mar 2019. 9
- [18] G. Pasolini, C. Buratti, L. Feltrin, F. Zabini, C. De Castro, R. Verdone, and O. Andrisano, "Smart City Pilot Projects Using LoRa and IEEE802.15.4 Technologies," *Sensors*, vol. 18, p. 1118, apr 2018. 9

Chapter 1

Building energy modeling: methods and challenges

The first chapter of this manuscript aims to provide an exhaustive summary and a technical framework over building energy modeling, as a key tool to improve building energy efficiency. A review of the main modeling techniques – physics-based, data-driven and hybrid – is provided. It is completed with a discussion on the best-fitted approach considering the targets of this research work. Challenges in building energy modeling are studied in the second part of this chapter, specifically the energy performance gap. The calibration of building energy models is stressed as a relevant strategy to close the performance gap.

Introduction

Building energy modeling is one of the key tools to assess building energy consumption. Building energy modeling makes it possible to formulate energy consumption – among other outputs – as a function of a diversity of parameters describing a given case study as precisely as possible. Various applications come out of building energy modeling. The most common is the regulatory calculations for new or existing buildings, and used prior to retrofit planning [1]. The literature review highlights a variety of modeling techniques, but not all techniques are equivalent to each other. Types and amount of input data, reliability, flexibility, interpretability of the modeling process: these are some of the many aspects to consider to select the fittest energy modeling technique for a given case study and a given purpose.

The first part of this chapter introduces the basics of building energy modeling. It provides a review and a classification of building energy modeling techniques, a description of the modeling process, a presentation of pros and cons of the different methods, and an overview of modeling tools. The scope of the review is limited to building-scale energy modeling in post-occupancy stage. Building energy modeling is a large research topic that can be tackled with various strategies. Most techniques can be transposed to different case studies and different spatial modeling scales, from an energy system to city-scale. However, from one scale to the other, the inputs, the specificity of case studies or the dynamics of energy consumption differ [2]. Therefore, conclusions on the applicability of modeling techniques are also different. In the present work, we focus on bottom-up modeling techniques, based on the combination of sub-building-level inputs to achieve the final modeling output [3]. The building life stage also matters. Between the design, the construction and the building operation, available data and modeling goals can be quite disparate. Therefore, considering the goals of this research work, the scope of the review is set with respect to our case study: existing and operated buildings.

Hence, the selection of a specific building energy modeling technique is crucial. However, aside from this specific aspect, building energy modeling entails other challenges. A model aims to give a picture of an existing system – a building, in the present case – as close as possible to the reality, and to test on various scenarios and process changes. Therefore, the challenge lies in the accuracy and the reliability of the simulations. A performance gap between the models and the actual building performances, meaning a difference between simulation results and the monitored building energy consumption, is often reported [4]. The goal is to minimize this performance gap but also to identify its origins. Consequently, building energy models must be calibrated. The different input

parameters of the models must be checked and tuned for the energy models to match the existing case studies [5]. The second part of this chapter discusses the origins of the performance gap with a focus on those related to the energy modeling process. Strategies to reduce the performance gap are also discussed, specifically the calibration process of energy models.

1.1 A few definitions

Prior to the review of the literature, this section defines a few key concepts to ease the reading and understanding of the chapter.

In the classification and description of modeling methods, we discriminate approaches from methods and techniques. Approaches are defined as the broad strategies to implement building energy models. Methods and techniques refer to the specific algorithms and processes that are implemented [1].

Through the comparison of building energy modeling methods, there is a concept of performance that arises. From a modeling point of view, the performances can refer to the accuracy of models (the closeness between simulation results and measured energy consumption), but also to computing performances, such as computing speed. From an energy point-of-view, the meaning of modeling performances is limited to modeling accuracy since computing aspects are not the focus of the present research work.

Modeling interpretability is the expression used to discuss how the modeling process links input variables to simulated outputs. It means that the impact of an input variable on the output of a model can be clearly understood from a physics point-of-view [6].

Finally, a distinction is made between modeling and forecasting, as many reviewed research papers use both terms. Modeling refers to a situation that can be compared to measured field data. Forecasting refers to a future situation for which forecasting performance cannot be assessed. Both modeling and forecasting depend on the same process, and a forecasting is based on a model.

1.2 Approaches and techniques for building energy modeling

1.2.1 Classification of bottom-up modeling techniques

The review of the literature highlights a significantly large panel of approaches and methods to implement building energy modeling. There are two global strategies: bottom-up modeling, opposed to top-down modeling. Both aim to identify and characterize the driving components of building energy consumption. However, these two categories of models differ by their hierarchical construction. Swan and Ugursal [2] refer to bottom-up and top-down models at the residential sector level and their definition can be transposed to building-scale. Bottom-up modeling strategy is based on the combination of constitutive elements of a building and its environment (i.e. the characteristics of the envelope, the energy systems, the operation schedules, the local weather) to achieve the final building-scale energy consumption. Top-down modeling works the other way around. It starts from a larger scale (city-scale or country-scale, for instance) to get down to building-scale energy consumption using a disaggregation strategy, and cross-referencing various meta-data (e.g. the age of a building, the location, the architectural style, or socioeconomic data on occupants) [3]. Considering the objectives of the present research work, in this review, we focus on bottom-up techniques.

To provide a summary of the many existing modeling techniques, the present section gives a classification of bottom-up energy modeling methods, based on the research work we previously conducted on the review of data-driven energy modeling techniques [7].

Three approaches are highlighted: physics-based, data-driven and hybrid building energy models (BEM) (Figure 1.1). Physics-based methods, also called white-box models [8] solve energy and mass-transfer equations in a building to describe its energy consumption and thermal behavior. Several classifications have been suggested in the literature, depending on the source of input data [2], the level of detail in the modeling process [1] or the calibration methods [9, 10]. The sub-classification of Fouquier et al. [1] is preferred in this review. It underlines differences regarding the level of detail and implemented calculations, with computational fluid dynamics (CFD), zonal and multizone techniques. On the opposite of physics-based approach, data-driven models, also called black-box models, rely on the sole combination of measured or simulated datasets to deduce building energy consumption [11]. Data-driven methods are based on machine learning algorithms. A distinction is suggested between unique and combined models. The former is built upon one single algorithm. The second combines two or more machine learning methods. Finally, hybrid models are the combination of physics-based and data-driven models. They are also called grey-box models [8]. For hybrid models, there is no defined classification. The literature review shows that the structure of hybrid BEM is mostly case-dependent. The goal usually is to replace parts of the physics-based structure with data-driven algo-

rithms.

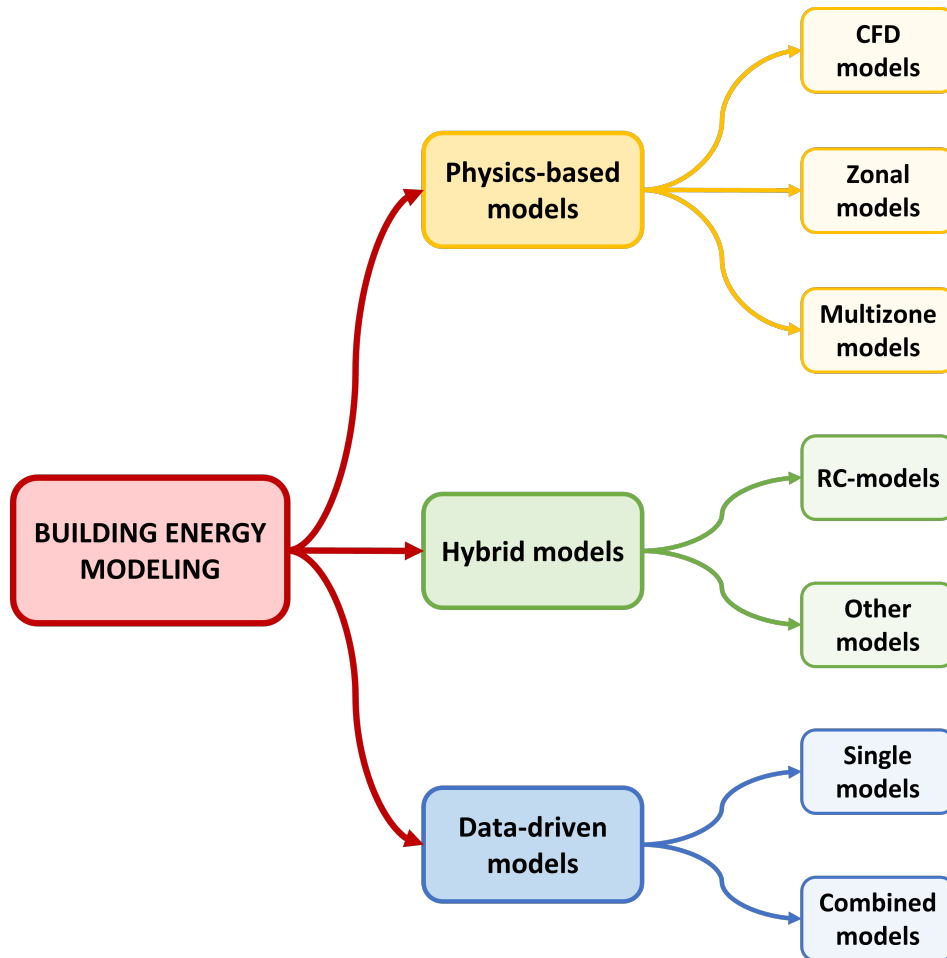


Figure 1.1: Classification of building energy modeling approaches and techniques.

1.2.2 Metrics for modeling performance assessment

Many performance assessment metrics are reported in the literature. For each considered time step, the difference (i.e. the error) between the simulation results and the comparative reference dataset (based on measured or simulated data) is computed and summarized into one representative number for the studied timeframe. There is a large variety of metrics in the literature. The most implemented metrics are the mean absolute error (MAE – Eq. 1.1), the mean absolute percentage error (MAPE – Eq. 1.2) and the root mean square error (RMSE – Eq. 1.3). Other metrics are implemented as well, such as the coefficient of determination, R^2 , the mean square error (MSE), the mean relative error (MRE), the mean bias error (MBE) and the normalized mean bias error (NMBE).

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (1.1)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (1.2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (1.3)$$

Where, for a given time series, n is the number of data points in the time series, t is the time-step, y_t is a data point from the time series at time-step t , and \hat{y}_t is the corresponding predicted output at time-step t .

The choice of a given metric offers specific insights on performances of a BEM[12]. Comparing RMSE, MSE and MAE, the latter is easier to understand due to the simple calculation involved and it gives an equal weight to large and small modeling errors. The former two penalize larger errors and the RMSE is usually preferred to the MSE since it has the same units as the data. However, most reviewed studies in the literature seem to choose the metrics that looks like the "best" performance, meaning that the difference between simulation results and data is minimized. For instance, in the work of Liu, Chen and Mori [13], there are only two metrics – MSE and R^2 – while a comparison of more metrics such as in the work of Zhang et al. [14] – MAE, MAPE, RMSE, mean percentage error, and mean error – provides a more complete assessment of modeling performances. Also, two types of metrics should be distinguished: scale-dependent and scale-independent metrics. The former, such as MAE, is useful to compare several methods over a given dataset but cannot be used to provide a comparison of different case studies. In that case, scale-independent metrics (percentage or relative error metrics), such as MAPE, should be selected.

Performance assessment metrics are mostly found in data-driven applications, especially as input and output data can easily be compared and have the same time-step. It may be implemented as well for physics-based and hybrid models. The only limitation is the possibility to compare the physics-based or hybrid simulations with detailed field data.

1.2.3 Physics-based building energy modeling

1.2.3.1 Overall principle

Physics-based models compute energy and mass-transfer equations to assess the building energy consumption [1]. They mostly focus on thermal energy – heating and cooling – but also domestic hot water (DHW) and electricity demand. Depending on the modeling tools, level of details and input parameters, other outputs can be obtained. Aside from energy consumption and demand, physics-based models may provide outputs such as building load profiles including peak loads, energy system efficiency, parameters to characterize the thermal comfort (indoor temperature, humidity, solar gains) or life cycle analysis [15].

We focus on building-scale energy modeling. Therefore, techniques such as CFD and zonal approaches are not considered. Due to a very high level of details required with these techniques, specifically in the definition of the geometrical mesh, CFD and zonal techniques are mostly implemented on parts of the building for flow modeling (air flow, thermal flux, etc.).

The process of physics-based building energy modeling process can be summarized considering four categories of input parameters, as illustrated in Figure 1.2:

- The passive system includes a description of the envelope, with the building architecture, glazing system, walls, thermal bridges or crawl space, for instance;
- The energy systems with specifications over heating, cooling and DHW generation, as well as ventilation, electric appliances, lighting and renewable energy systems if any are considered;
- The operation scenarios that describes all building usages and including, for example, occupancy, temperature setpoints, DHW demand, ventilation schedules, dissipated power, window opening or shading;
- The outdoor environment, meaning the local weather and the surroundings such as other buildings and vegetation that have a cooling or shading effect on buildings [16].

Physics-based modeling is performed using specific software to build a digital twin of a building focusing on energy aspects. A review of these tools is provided in Appendix A.1.

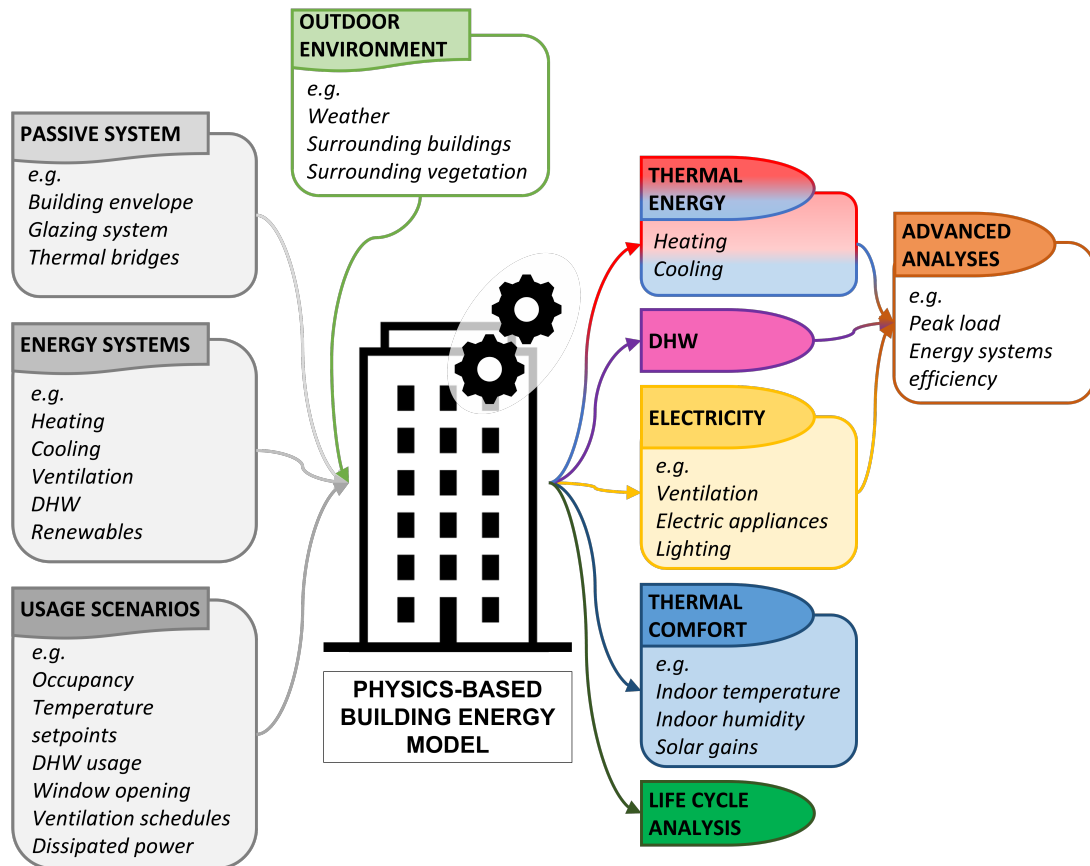


Figure 1.2: Summarized process for physics-based building energy modeling.

1.2.3.2 Multizone modeling

The multizone technique, also called nodal technique, is the method commonly applied for building energy modeling. The overall principle is the same as in any physics-based building energy model described in the previous section 1.2.3.1.

However, the key point of multizone building energy modeling is the definition building thermal zones. A thermal zone is a volume of a building showing a unique thermal behavior and with dedicated operation scenarios. This volume can be the whole building, a floor, an apartment, a room, or part of room such as in an open-space office with several air-conditioning units, for instance. Building orientation can also be considered in the definition of a thermal zone if the orientation induces different thermal conditions. Then, energy consumption and thermal features are computed for each zone. The selection of the most adapted number of thermal zones depends on the objective of the study and the available information to implement the model. Because of the large number of required input data, a thermal zone can only be considered if all features are known. Then, with

1.2 Approaches and techniques for building energy modeling

a few large thermal zones, the computation time of the software is fast but there is little available detail in the model. On the opposite, with several smaller thermal zones, the picture becomes more accurate, but the calculation time and complexity significantly increase, which is also a source of modeling mistakes. In practice, a commonly adopted process is to first model a building as a single thermal zone and divide the building into several thermal zones up to the highest expected level of detail. Such a process helps identifying modeling errors and uncertainties on a large scale before the complexity of the BEM is increased.

Multizone technique describes building the energy behavior on a large time-scale (up to one year) and at full building-scale with customizable level of detail [17]. This is a significant advantage over other modeling techniques described in the following sections. There is a wide range of applications in the literature. Pan et al. [18] implemented a typical BEM of a high-rise building in China to assess energy conservation measures. They use DOE-2 software to perform hourly time-step simulations, calibrated with monthly billing data. Simplified applications have also been proposed such as the HC-method developed by Zhu et al. [19]. They used a limited description of the building characteristics to assess its heating and cooling needs. Specific aspects of building energy performance are also targeted. Two specific topics are highlighted by the literature review. The first the characterization of the impact of energy retrofit actions, such as the work of Jankovic [20], who compared building characteristics before and after retrofit to compute the performance of retrofit actions through BEM. The second aspect is the study of the impact of occupants' behavior on building energy consumption. Dar et al. [21] looked at the effect of occupants' and their behavior on heating energy consumption for a detached well-insulated housing. Ahn et al. [22] focused on predicting the occupants' presence and its effect on modeled building energy consumption, and compared with measured energy consumption in a two laboratories.

The level of modeling detail in physics-based models is an advantage as well as a major drawback. The accuracy and reliability of modeling performances entirely depends on the available information to describe the different thermal zones. This information are often quite difficult to collect and are often replaced with assumptions that lead to uncertainties [23]. Although specific methods have been implemented for the collection of data to provide modeling inputs [24], a lower level of detail necessarily implies a lower modeling performance and less interpretability of energy models. Furthermore, physics-based modeling software are limited regarding the modeling of thermal characteristics and flows in and between different thermal zones. Indeed, it is not possible to model the temperature gradient in a room, as only the temperature for a whole thermal zone is

computed. Having several thermal zones then provides a compensation for an analysis of the temperature gradient at building scale. Also, air flow between thermal zones, figuring natural air flow or related to window/door opening is not considered.

1.2.4 Data-driven building energy modeling

We reported on data driven techniques for building energy modeling in a previous detailed literature review. Therefore, the present section summarizes the main characteristics of data-driven techniques and more details can be found in [7].

1.2.4.1 Overall approach

Unlike physics-based models, data-driven modeling techniques do not rely on physics laws but on the links between sets of data collected in buildings, using statistical data analyses methods and machine learning algorithms. A majority of methods focus on time series data analyses [11], although a physical description of building may also be considered as input variables [25].

The overall modeling concept, regardless of the data-driven method, is to solve a regression problem. This means that the goal is to analyze the causal link between one (or more) output(s) – usually the building energy consumption – and inputs, also called explanatory variables. The combination of inputs aims to achieve the output(s), as for physics-based modeling, but without physical information and using instead a function deduced from measured data describing the building operation.

The data-driven modeling process follows four steps: data pre-processing, training, validation, and testing (Figure 1.3). Data pre-processing prepares an initial dataset with data cleaning and input data selection, that is then divided into three distinctive datasets for each following step of the process. Datasets must include the same types of inputs variables and the same corresponding outputs but with different combinations of values. For instance, considering time series data, the three datasets should have the same inputs but over different timeframes [26].

Training is the first step. The algorithm provides an initial function to formulate the targeted output based on a combination of input variables. Simulation results are compared with measured data from the training dataset. Parameters of the algorithm tuned to reduce the error [27]. Validation is the second modeling step. The tuned algorithm from training performs a second round of simulations using the validation dataset that differs from the training dataset. The modeling error is assessed, and parameters of

1.2 Approaches and techniques for building energy modeling

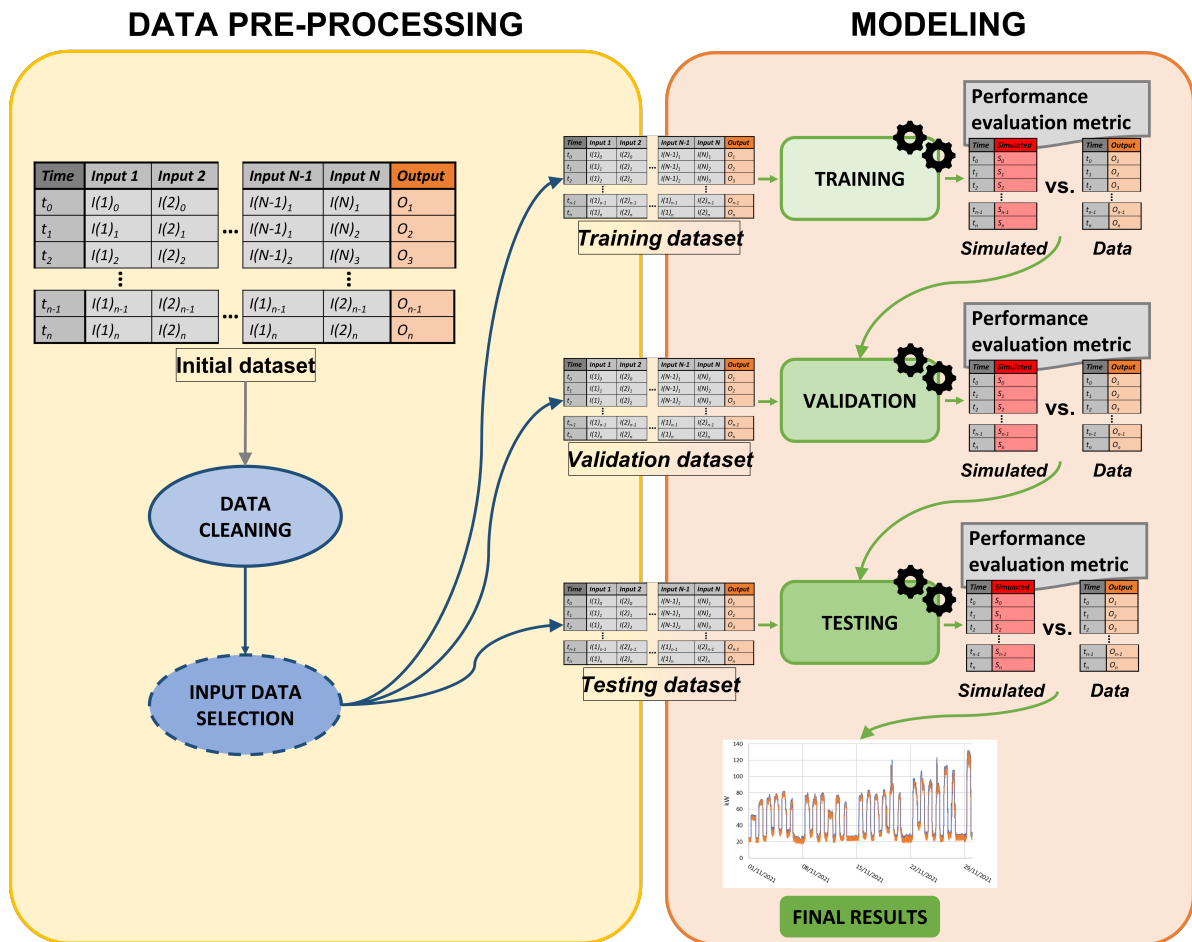


Figure 1.3: Data-driven building energy modeling process.

the algorithm are tuned once more, if necessary. It is common in the literature to find that this step is sometimes merged with the training step [7]. However, using a different dataset with the validation step offers two advantages: a larger variety of situations in building operation can be taken into account in the energy model, and it helps avoiding over-fitting – the model would be too specific to a given combination of data and would poorly model other events in building operation [14]. Finally, the testing step uses the tuned model from validation step to provide an unbiased comparison between simulation results and data. It is conventional not to tune the algorithm at this step, for the testing results to be considered as the modeling performances of the implemented technique [28].

Regarding the distribution of data in training, validation and testing datasets, the majority of datasets use the largest part of the data for training or training and validation – up to 60-80% of the data [7]. Then the rest of the data is evenly distributed between validation and testing. Data distribution can be performed manually (e.g. the selection of three distinct timeframes) or randomly, such as with cross-validation technique [29].

Finally, a key aspect of building energy modeling is the availability of data, and even more so for data-driven techniques whose process entirely depends on data, as the name suggests. There is no generic conclusion regarding the amount of data that is required to implement data-driven techniques. It usually works the other way around: the amount of available data drives on the choice of a modeling technique. There is also a balance to observe. Small amounts of data for a few specific parameters may provide reliable models if the measured parameters are related energy drivers and if the application range does not drift much from the building operation depicted in the dataset. On the other hand, a very large amount of data gives a larger picture of building operation but significantly complicates the energy model. First, it requires a long and time-consuming data checking and cleaning process. Then, more data does not necessarily mean better performances. On the contrary, using too many or unsuitable inputs can decrease modeling performances [30].

1.2.4.2 Data pre-processing

As for any data analysis, data pre-processing is an essential step to assess the quality of collected data and decide what would be the optimal inputs, time-steps and timeframes to consider for modeling.

Data pre-processing can be dividing into two steps: cleaning and data selection of explanatory variables. Cleaning includes data checking, assessment of missing data, removal

1.2 Approaches and techniques for building energy modeling

and/or substitution of outliers to ensure an optimal training of data-driven models. Data cleaning can be implemented manually [31] or automated such as in the work of Fan et al. [28] who proposed an automated detection of outlying values to be removed before the modeling process.

Input data selection is not mandatory. However, it is clear from the literature review and the many implemented techniques that it is highly recommended. The first aspects to consider are the time-step [32, 33], since it has an impact on modeling results [34], but also the distribution of data for training, validation and testing [27, 35]. Then, the physical meaning of the data can be taken into account to highlight specific building operation characteristics and optimize building energy modeling: different models for different building operation situations. Newsham and Birt [36] proposed to separate input variables based on the typology of days (worked days versus days off), Ma et al. [37] and Mena et al. [38] looked at weather conditions, and Tang et al. [39] considered seasonal effects. However, a manual process might be biased because it relies on the experience of the user. Classification, based on clustering algorithms, provides an automated alternative although the results of input data distribution might be difficult to analyze if there is not enough available knowledge on the case studies [39, 40].

The choice of specific input variables is the most adopted strategy. It aims to only keep the most relevant input parameters, those that are the most impacting in terms of building energy consumption [17, 41, 42, 43]. The most common methods are sensitivity analyses [44] and principal component analysis (PCA) as implemented in the work of Li et al. [45] and Nilashi et al. [46]. The literature also highlights other techniques. Deb et al. [47] chose clustering analyses based on building variables, Paudel et al. [48] used wavelet analysis for the clustering of days with similar climatic and building operation conditions.

1.2.4.3 Single models

Single data-driven models relate to all data-driven methods using a single algorithm.

Among the easiest algorithms to implement are the moving average, the exponential smoothing and the autoregressive models [11]. These types of algorithms are based on iterative combinations of past observation of the output time series [49]. For building energy consumption, it draws on the past data points of the energy consumption. Moving average weights all terms of the combinations equally. Exponential smoothing gives a larger weight to the most recent observations. Autoregressive models, whose most common variation is autoregressive integrated moving average (ARIMA) models [50] adds to

CHAPTER 1 – BUILDING ENERGY MODELING: METHODS AND CHALLENGES

the autoregressive part a linear combination (moving average) of the last modeling errors – i.e. the difference between simulated and observed data. Although basic techniques can be improved (considering seasonal aspects or exogenous variables such as building occupancy [51, 52]) and are simple to handle, they have several major drawbacks. They do not handle non-linear phenomena – although a common situation in building energy modeling [53]. They show poor modeling results when the energy demand has significant variations over a short period of time [54, 55], and they can only use a limited number of explanatory variables, which means a poor modeling interpretability.

A second category of popular models are the statistical regressions. These models group several different techniques combining explanatory variables into a mathematical expression to give an output (building energy consumption in the present case). Statistical regressions are often used as reference methods to compare with more complex techniques [53]. They offer a simple implementation, a reasonable interpretability and relevant performances. Their major drawback is their inability to handle non-linear modeling (except for multivariate adaptive regression splines – MARS) and the large amount of data – in term of data points – required to implement a model.

The k nearest neighbors (k-NN) technique stands out of the other types of regression algorithms. k-NN aims to identify and group similar trends within time series [56]. For instance, building energy consumption can be related to similar trends in occupancy, weather conditions and other relevant parameters. Unlike other data-driven algorithms, k-NN does not follow a training-validation-testing process. The whole set of available data is used, and modeling is performed by comparing new inputs to other available values of a given variable. The k-NN technique has two key parameters. The first is the k parameter, which is the number of neighbors to which a new datapoint is compared to be grouped. The second parameter is the distance metric to assess how close a new datapoint is to its neighbors. The final energy consumption is deducted from the averaged energy consumption of the k neighbors. Although they provide an intuitive and simple implementation [37, 57] and reasonable modeling performances [27, 58], the latter highly depends on the comprehensiveness of the dataset to depict the operation of building case studies.

Three other categories of single models are the decision trees (DT), support vector regressions (SVR) and artificial neural networks (ANN). The former, DT, is inspired from the structure of trees, from roots to leaves. It consists in a set of consecutive tests checking on values of explanatory variables and the final corresponding output. Each test corresponds to a tree node. The number of tests, dealing with categorical or

1.2 Approaches and techniques for building energy modeling

binary explanatory variables, depends on the number of combinations of values. Tests are conducted until the final output is obtained. Using DT, the interpretation of results is relatively simple [59]. This method offers significant modeling performances and is adapted to non-linear modeling [30]. However, DT show issues in adapting to large sets of data that would depict a large diversity of building operation conditions. It implies a larger number of combinations and longer computation, which makes it quite difficult to optimize the models.

SVR are based on SVM (support vector machine) process. It follows a similar idea to a statistical regression but with a more complex approach, using several hyper-parameters in order to fit to training time series. One key parameter is the kernel function. This parameter transforms the data representation space to account for non-linearity in a dataset: if the dataset cannot be accurately modeled by a linear expression, another mathematical expression can be found in another dimension to fit to the dataset [48]. SVR are especially performing for non-linear regression problems [60]. They rank among the most used data-driven models [13, 61, 62]. However, the calibration of SVR is particularly difficult because of the selection of hyper-parameters. Tuning hyper-parameters plays a decisive role in modeling performances. Optimisation of SVR has even become a specific research topic for building energy modeling [63, 64].

Finally, ANN algorithms are inspired by neural networks in the human brain and are built on a similar concept of information propagation. Typical ANN are composed of three types of layers. Input data are introduced through the first layer of neurons and are given a random weight. The intermediate layers combine the different inputs, transformed through a transfer function. The final layer provides the output result. The combination-transformation-output process is reiterated until there is a convergence of the error rate (the difference between simulation and measured data) or until the maximum number of iterations is reached. Key parameters can be optimized: the number of intermediate neurons and layers, the connections between neurons, the transfer functions [65] and the back-propagation of the error rate for the model to these parameters. ANN basic algorithms offer a simple grasp of the mathematical model, performing for non-linear phenomena, adapted to many types of input data and case studies. More advanced algorithms offer a range of improvements. ANN is an umbrella term and numerous different algorithms are implemented in the literature, from the simple three-layer neural networks to deep neural networks [66]. Nevertheless, the modeling process is not interpretable and the more advanced the algorithm, the more complex the implementation, with a risk of over-fitting [61, 67].

1.2.4.4 Combined models

Unlike single models, combined models rely on several algorithms combined to take advantage of the strengths of each constituting technique. There are two types of combined models: ensemble models, and single models improved with an optimization technique.

Ensemble models combine single models together to improve modeling performance [28]. It is a two-step process with i) the creation of several single sub-models called base-learners or base-models and ii) the comparison of the performances of sub-models. Results of each sub-model are weighted according to their performance and are combined to get the fittest results. There are two types of ensemble models [30]. Homogeneous ensemble models use one single modeling technique trained on several sub-datasets, such as random forest [68, 69] or boosting decision trees [28, 30]. Heterogeneous models use different single models trained on a unique dataset. Examples of heterogeneous models can be found in the work of Fan et al. [28] who combined eight single models – statistical regressions, auto-regressive models, SVR, k-NN and ANN – or the work of Alobaidi et al. [70] who used a combination of ANNs. Ensemble models offer significant modeling performances. However, they are particularly difficult to implement and they require advanced coding skills and knowledge.

Improved models are the second category of combined models. There is no clear definition in the literature. Some are called "hybrid models" [71] but this expression also refers to a combination of physics-based and data-driven techniques, as in this manuscript. In the present case, improved models use optimization techniques to improve the performance of single models. Common optimization techniques are swarm intelligence algorithms. Most improved models are implemented with SVR [14] and ANN [72, 73].

1.2.4.5 Machine learning tasks

The modeling techniques presented above can be implemented with different tasks, depending on how input data are taken into account. There are four main types of machine learning tasks applied for data-driven BEM: supervised, unsupervised, reinforcement and transfer learning.

Supervised learning is the most common task, found in the majority of reviewed papers. With supervised learning, input data are known and labelled. The physical meaning of each input data is clearly identified. All data, possibly pre-selected, are used in the model and keep their label over the modeling process. All techniques introduced above can be implemented with supervised learning approach. On the opposite, with

1.2 Approaches and techniques for building energy modeling

unsupervised learning, input data are not labelled. The goal is to automatically identify links between inputs that would not be obvious before the modeling process, and maybe not even considered with a biased supervised modeling – since the user performs the selection of inputs in supervised modeling. Hence, common tools aim to perform input data characterization independently from their physical meaning. Unsupervised learning is particularly adapted to very large sets of data when labelling would become difficult and time-consuming [74]. It significantly improves modeling performances. However, because input data are unlabelled, the interpretability of the models is particularly poor.

The literature review also highlights two other machine learning tasks. The first task is reinforcement learning [75]. Unlike supervised and unsupervised machine learning, the modeling process does not extract input data characteristics. Reinforcement learning defines an artificial decision-making agent. A goal (the targeted energy consumption) is defined and given to the agent and it must reach it. Each decision brings a reward based on how close decisions lead the agent to the final result. The artificial agent must maximize the reward to find the optimal way to achieve the final result. A frequent analogy is the decision-making process of a mouse finding its way towards food in a maze. Reinforcement learning algorithms are specific to this type of machine learning task but rely on artificial neural networks.

Finally, transfer learning, as the name suggests, transfers knowledge from a given case study to another case study for which there is a lack of information to model building energy consumption [76]. For instance, Mocanu et al. [33] used transfer learning to train a model using inputs from a specific commercial building and to model the energy consumption of different commercial and residential buildings. Algorithms try and fit data and modeling techniques from a building to the other by, for instance, adjustments over seasonal or recurrent events. Transfer learning has been scarcely implemented for building energy modeling applications.

1.2.5 Hybrid modeling

Hybrid building energy modeling, also called grey-box modeling [77], is the third category of building energy models. It combines both previous approaches, physics-based and data-driven methods: a simplified white-box basis complemented by the use of data with data-driven models.

1.2.5.1 Resistance-capacitance models

Resistance-capacitance models (RC-models) are the most implemented hybrid models [78]. They are implemented upon an electrothermal analogy with resistances and capacitances. The former represents thermal resistances for the property of building materials to resist or conduct heat flow. The latter mimics the thermal capacity of materials to store and release heat over time. The order of a RC-model is given by the number of RC nodes, related to the number of thermal zones in a building and its constituting elements (e.g. the envelope, the energy systems, the occupants) [79]. RC-models mostly focus on thermal energy consumption. Li et al. [78] highlighted the main RC-modeling applications including heat dynamics analysis, thermal load calculation and thermal studies for building control optimization, district/urban energy modeling and building grid integration. There are two types of RC-models: the forward approach and the inverse approach. The former is mostly physics-based. Resistance and capacitance components usually figure physical components of a building calculated using a physics-based modeling software. The latter is mostly data-driven-based: data-driven techniques are implemented to assess resistance and capacitance components.

1.2.5.2 Other hybrid models

Other hybrid models follow a similar idea to inverse RC-modeling technique: the use of data-driven models to assess and tune specific parameters or parts of a physics-based model. Siddarth et al. [80] used a genetic algorithm to highlight key energy-driving parameters in a physics-based model and to create realistic sets of values for these parameters to be tested in their model. The second hybrid modeling scheme aims to replace part of physics-based models, for instance with a given energy system or energy driver. Focusing, on HVAC (heating ventilation air conditioning) systems, Collinge et al., Dong et al. and Ratz et al. [81, 82, 83] combined physics-based building models with data-driven models of a HVAC.

1.2.6 Conclusions on the applications of approaches and methods for building energy modeling

Building energy modeling is one of the key tools to assess building energy consumption and target energy savings. Building energy modeling has been a major research topic, well documented and studied. In the past twenty-five years, at least twenty-four review papers have been published on the topic of building-scale energy modeling techniques alone,

1.2 Approaches and techniques for building energy modeling

reviewing hundreds of original research papers (Table 1.1). The review of approaches and techniques for building energy modeling highlighted the three main categories of bottom-up techniques: physics-based, data-driven and hybrid modeling. The former relies on physics equation solving, the second uses combinations of building energy-related measured data and the later combines both strategies.

Each approach embeds various methods. Data-driven models offer the largest diversity of techniques. They have been a main research topic for building energy modeling in the past twenty years. Twenty-three of the twenty-four reviewed review papers targeted data-driven BEM, with fourteen articles that exclusively focused on these techniques (Table 1.1). Data-driven techniques stand out for their modeling performance, their accessibility thanks to several open-source tools (Appendix A.2) and the diversity of applications (electricity, thermal energy, etc.). Also in their favour, they do not require advanced knowledge on building science. However, they provide a poor interpretability. Data-driven modeling is a purely statistical and mathematical process that prevents any physics-oriented analysis of the modeling results. The main goal is modeling accuracy using time series data. Therefore, it is almost impossible – or at least particularly complex – to interact with physical characteristics of the modeled buildings and their operation to quantify the impact of specific changes on the energy consumption. Furthermore, without sufficiently diverse and reliable building operation data, there is no model. It should also be noted that, if there is an easy access to many pre-coded algorithms, a complete understanding of modeling techniques and of the tuning of parameters still requires advanced coding skills and knowledge on machine learning.

On the opposite, physics-based methods solve physics equations that link the description of a building (structure, usage, environment, and energy systems) to its energy consumption. Although, some models are more detailed than others, it is usually a complex modeling process to perform. Physics-based BEM require specific software and tools, with dedicated skills from the users. Nevertheless, physics-based models are well-known and reliable methods, often used as references to compare with data-driven techniques [77], for industrial energy analysis and diagnoses to enquire about energy efficiency and potential savings over retrofit actions [97]. A significant benefit is the interpretability of the models. The many inputs provide a large line of possibilities to optimize building energy consumption, although the significant amount of information required can be a drawback as well. Information are difficult to collect while they are a must to implement an energy model and then have to be replaced with assumptions: the more assumptions, the more uncertainty in the modeling results.

CHAPTER 1 – BUILDING ENERGY MODELING: METHODS AND CHALLENGES

Table 1.1: Review papers on building energy modeling published in the past 25 years.

Authors	Publishing year	Reviewed techniques		
		Data-driven	Physics-based	Hybrid
Pedersen [84]	2007	X	X	
ASHRAE [9]	2009	X	X	
Swan and Ugursal [2]	2009	X		
Zhao and Magoulès [23]	2012	X		
Foucquier et al.[1]	2013	X	X	X
Fumo [10]	2014	X	X	X
Tardioli et al. [8]	2015	X		
Chalal et al. [67]	2016	X	X	
Deb et al. [11]	2017	X		
Mat Daut et al. [71]	2017	X		
Wang and Srinivasan [85]	2017	X		
Yildiz et al. [86]	2017	X		
Ahmad et al. [87]	2018	X		
Amasyali and El-Gohary [88]	2018	X		
Wei et al. [89]	2018	X		
Wei et al. [90]	2019	X		
Bourdeau et al. [91]	2020	X		X
Sun et al. [92]	2020	X		
Fu et al. [93]	2021	X		
Li et al. [78]	2021			X
Zhang et al. [94]	2021	X		
Al-Shargabi et al. [95]	2022	X		
Chen et al. [77]	2022	X	X	X
Lu et al. [96]	2022	X		

1.3 Performance gap in building energy modeling

Finally, the third approach, is the combination of data-driven and physics-based modeling. Hybrid models are a simplification of the latter tuned with the former. Then, hybrid models offer a relevant balance between modeling interpretability, performances, and ease of implementation. Most models are RC-models, focusing on the thermal needs of building case studies as for physics-based models. However, in comparison with the other two categories of techniques, hybrid modeling has been less implemented in the literature for building energy modeling applications.

The implementation of different modeling techniques answers different questions. Data-driven models are the most fitted to target the highest modeling performance. As the meeting of data-driven and physics-based techniques, hybrid models offer the best compromise between an interpretable and performing modeling process. However, if descriptive knowledge of the building case studies is sufficient, physics-based modeling is the most adapted approach to investigate energy consumption and identify energy-saving strategies. The optimization of energy consumption usually targets building operation, energy systems or the envelope. However, usages and occupants tend to be under-considered while they play a significant role over building energy consumption. Due to the many involved parameters, physics-based modeling allows a dedicated investigation on a fully understandable model which would be the most fitted choice in the present case.

Nevertheless, aside from the choice of a modeling technique a significant challenge remains with all types of building energy models: the gap of modeling performance between the model and the existing case study [4]. Identifying the gap, its origins and closing the gap is a key research topic as well, discussed in the following section.

1.3 Performance gap in building energy modeling

1.3.1 Context and definitions

With BEM, the goal is the performance: to improve – reduce and optimize – energy consumption of buildings, to accurately match modeling results with measured energy consumption, and to comply with standards. However, when it comes to the comparison with the measured building energy consumption, there is often a gap with the simulation output [4].

This gap is called *performance gap* (PG [4, 98]). It is a discrepancy in building energy consumption modeling due to various factors such as the building design, its operation or modeling assumptions. This concept is of paramount importance and affects many

building construction and energy efficiency stakeholders. The identification of origins and improvement strategies is essential to take the full benefit of BEM.

1.3.2 Origins of performance gaps

Various origins to the PG have been identified in the literature. A recent review work suggests a classification of origins following the building life cycle [4, 98]: design, construction and post-occupancy. It applies for new buildings as well as retrofitted buildings, since retrofit actions follow the same process of design, construction, and operation.

1.3.2.1 Building design

Issues in the building design lead to significant PG. It can originate from communication issues, which is a very common situation within large projects when many actors do not understand each other because of their different technical backgrounds [99]. This leads to a poor definition of the design goals and a significant lack of understanding regarding the final use of the building and its characteristics [36]. Also, a lack of perspective on the construction project induces poorly executed work with many modifications and then significant PG afterwards [100]. One of the most common challenges in building design is also the lack of operational data. Because it is a design, there is no feedback on how the building will actually operate [101]. Many assumptions are made to assess the performances of the buildings and these assumptions may end up being quite different from the real future operation.

Within the design goals, there are common misconceptions about how the design fits to the needs of the project [102] and its feasibility. [4]. Over-original designs can be counter-intuitive for users and too difficult to operate [103, 104]. This is even reinforced because technical testing is conducted in controlled environments while the actual use in the built environment reveals a very different operation perspective [105]. Then, one of the keys is to provide a clear documentation regarding design choices, intended uses and modifications which often lacks afterwards.

BEM in the design stage is a significant source of performance gaps as well. It is very common to find differences between building operation and design modeling results [106]. These may be related to the tools [107], the actors' skills in design [108] or, as previously stated, the challenges in assessing the many input parameters because there is no feedback on building operation yet – for instance the occupancy or indoor heat gains. Finally, an underestimated cause of PG is the testing of models performed during the design phase [102]. As there are many parameters with uncertainties, designs and models

1.3 Performance gap in building energy modeling

must be adaptable enough to minimize the impact of these parameters [99]. Furthermore, as part of building life cycle, short and long-term deterioration of systems and materials must be considered in the design, specifically in case of retrofit actions [109].

1.3.2.2 Building construction

Construction is a second origin of PG, mostly because it is almost entirely manually conducted, with very little automation. The first source of performance gap comes from the checking of the quality of the construction that may not meet the initial goals – specifically regarding thermal insulation and air tightness [4, 107, 110].

There are also common issues in project construction management: a lack of directives regarding technical details and leading to interpretations [111], construction defects (intended or not) such as replacing materials for cost reduction [112], a lack of technical skills from the construction teams, last minute modifications and tight schedules to follow [4].

Construction issues are particularly difficult to survey because they would require a full supervision of the building construction site from day one, which is hardly possible. Therefore, most of the time, there is only a visual inspection when the building is delivered. However, most issues may only be visible after a few months of operation, which is why most strategies now focus on soft landing processes [113]. Soft landing processes provide a longer handover period and energy performance assessment. Nevertheless, most of the time, reception tests remain pretty simple and must be improved: ISABELLE [114] and PROFEEL programs in France [97] aim to ensure the quality of construction and retrofit actions with standard test processes.

1.3.2.3 Building operation

Building operation is the third origin of building energy PG. The first and most reported driver is related to building users and their behavior [4, 98]. Depending on the building, its size or the activities it hosts, parameters such as occupancy and usages may be difficult to assess. Hence, this information is over-simplified using a few observations or standard operation scenarios. They can significantly differ from measured occupancy and monitored building operation [115, 116]. Behavior-related examples may include window opening that impacts on energy consumption for heating and cooling, or the use of appliances with an effect on internal heat gain assessment [110, 115, 116, 117]. Also, cultural and social aspects can impact residents' behavior and trigger a chain reaction on building energy consumption [21, 108, 118].

The human factor also impacts on energy management when it is manually performed, modified by hand, or influenced by the occupants of buildings. It creates a PG related to unexpected energy management strategies, because of a lack of expertise and knowledge, interference with adjusted settings or unsuitable management strategies. Enhancing residents comfort with respect to the optimal regulation of energy systems can be a potential solution [119].

Overall, it is also necessary to acquire a more complete knowledge about considered buildings. Most of the time, there is a lack of detailed information and data to optimize energy performances. Data collection through instrumentation can be highly beneficial. Many nation-wide intensive instrumentation plans have been set accordingly [120, 121, 122]. Mixed public and company-led initiatives are quite common as well to implement massive sensor networks [97, 123]. However, the deploying a sensor network is not a straightforward task and it faces many challenges. Sensors and meters frequently focus only on energy consumption [20] while there are many other aspects to consider over the study of building energy dynamics. As previously discussed, occupants are a major energy driver. However, their presence and impact is difficult to monitor [124].

Another aspect to take into account is the spatial and temporal granularity of data collection. The number and location of sensors is important to provide an exhaustive picture of the building operation. Data collection time-step impacts on the quality of analyses and the results that can be obtained: too few data lack of relevant information and precious details that may alter the efficiency of energy monitoring [125] and result in a PG. Finally, building instrumentation faces many challenges including installation, calibration, data collection, or long-term network maintenance that can spearhead PG as well[91].

1.3.3 Strategies to address the energy performance gap

Strategies have been developed to think ahead the PG, and try and close it [98]. Strategies are divided into three categories: design strategies [100, 126, 127], technological and methodological strategies, and *soft measures* [128, 129].

Many issues originate from the initial design, construction and operation of case studies, both during the construction of studied buildings but also during energy retrofits. The scope of this thesis focuses on building post-occupancy. It relies on a case study of three existing buildings built in 1974 and presented in the following chapter. Buildings underwent a deep energy retrofit, but the thesis was conducted without any control over retrofit design and construction management. Therefore, the review of strategies to reduce

1.3 Performance gap in building energy modeling

PG is limited to strategies for operated buildings.

Post-occupancy strategies target a better understanding of the building operation. They provide the tools to assess the impact of modifications of operation over building energy consumption. These strategies are divided into three categories: energy modeling, data collection and analysis, and characterisation of occupants' behavior.

1.3.3.1 Building energy modelling

Reducing the performance gap starts with the improvement BEM performances, considering the needs for a model, the modeling process, and the reliability of simulation results. There are many means of implementing a BEM, as detailed in Section 1.2. Such a pool of techniques renders a difficult choice to select the best fitted method. The selection of a method must be driven by the availability of data to provide modeling inputs, the goal of the study, and the required level of detail and interpretability in a model.

Besides the choice of the modeling method, calibration is an essential aspect to tackle the PG. Even though the modeling technique is suited for the application, if the model is not calibrated, results cannot be relied on. Calibration is the process of using an existing BEM and tuning various input parameters in the model so that the simulated energy consumption matches with observed energy consumption within a given margin defined by the user as acceptable [130]. In other words, a calibrated energy model provides results that must be similar to the reality, in terms of cumulative consumption, energy dynamics (peak demand, typical demand profiles) but also regarding the different end-uses and parameters of the model. Considering a detailed physics-based energy model, the overall energy consumption must be accurate, but heating, cooling, DHW and the different characteristics of the building model (passive systems, operation scenarios, energy systems) should also be as realistic as possible. Hence, if some assumptions are made in the modeling process, they must be carefully considered to find the most suitable value to fit to the measured energy consumption. Many methods are reported in the literature and are further discussed in Section 1.4.

1.3.3.2 Data collection and analysis

Field data collection provides opportunities to improve the description of the building energy consumption and to add specific relevant knowledge into an energy model to replace assumptions. Data collection can be performed either during pre-occupancy or post-occupancy stages.

Pre-occupancy data collection is usually straightforward. Buildings are not occu-

ped. Hence, there are fewer potential sources of disturbance. However, for the same reason, data do not necessarily reflect the actual building energy behavior. Parameters such as occupancy and occupants' behavior cannot be fully captured but only guessed. Most collected data in pre-occupancy phase are related to the design of the buildings and parameters that are not affected by building usage: the building envelope, the glazing system, the technical description of energy systems. If the documentation on these features is not available, processes can be applied to recreate specific conditions and characterize building parameters [131], such as the thermal resistance of walls. Strategies have been reported to improve pre-occupancy data collection. For instance, virtual reality solutions can provide an immersive experience of a building to highlight design failures and collect usage data [132]. Also, another promising opportunity lies in transfer learning [33, 133] that relies on data from one building to another similar one, to gain knowledge that cannot be accessed otherwise. Even though it is not a data collection method properly speaking, it may be used for such purpose.

Post-occupancy data are suitable to accurately describe the actual building operation. The most significant challenge, that is also the main interest, is the building occupancy [91]. As the building is occupied, collected data are representative of how the building is operated but there are many sources of disturbance. Occupants can disrupt data collection (moving sensors, disconnecting gateways) and they can also be disturbed by the sensors and adapt their behavior, which results in biased data collection. Intrusiveness is also a key aspect of building instrumentation. There are many existing technologies and techniques to collect data in the best possible conditions, including wireless sensor networks [134], smart meters [135, 136], building energy management system (BEMS), Wi-Fi or camera systems [124], for instance.

Once data are collected, data analysis is performed. There are two classical approaches: descriptive statistical analyses and data mining. A descriptive statistical analysis provides a summary on a given dataset using descriptive quantities (mean, median, standard deviation and variance, for instance) [98] and statistical tests [31, 137]. It relies on a sample of the collected data which is a representative population of the full data set. Data mining uses different tools to analyse data from various perspectives, to summarize useful information, to highlight the links between different parameters, to extract patterns [34] and to eventually implement predictive models. Tools include machine learning algorithms, artificial intelligence, database management and clustering [98]. Data mining techniques are designed to process very large amounts of data, even the full amount of collected data, unlike descriptive statistical analysis. However, it also reduces data interpretability: it is not possible anymore to understand data analysis results from a

1.3 Performance gap in building energy modeling

physics-based point-of-view.

1.3.3.3 Integration of occupants' behavior

Occupants' behavior modeling and integration is the third major strategy to tackle PG reduction in operated buildings. Occupants' behavior has been reported many times as a source of uncertainties and major energy driver [138]. It is often narrowed to a few assumptions [7, 98]. However, a deeper understanding and characterisation of the residents' impact on building energy behavior is essential.

There are two means of considering occupants' behavior, either with models or data collection. The former relies on the latter. Data collection can be considered separately as a source of information to input in BEM without going through occupancy modeling and prediction. There is a wide variety of sensors for occupant-related data collection [124]. Basic measurements target the detection of occupants. However, the difficulty is to count the number of occupants in a given area and not only to detect the occupants' presence. Occupants' counting requires cameras or to use wireless local area networks, for instance, which needs more complex, more expensive, and also more intrusive instrumentation set ups. Other types of data collection strategies try and quantify the actions of occupants. Carlucci et al. [139] reported on six types of occupant-related actions, including window, solar shading and lighting operation, thermostat and clothing adjustment, and appliance use. These parameters are directly related to human activity in a building and impact on building energy consumption.

The modeling of occupants' behavior depends on the collection of occupants-related data. It goes further than a descriptive analysis and investigates the mechanics of occupants' behavior to predict occupants' actions and their effect. Most models use occupants' presence and counting, activity and movement between building zones [139]. Two types of modeling strategies are highlighted in the literature. Implicit models focus on the actions of occupants and the link between building features and occupants. Explicit models capture and predict the decision-making process of occupants. The details of occupants' behavior modeling falls out of the scope of the present work. Further information and details on this topic can be found in [139, 140].

The main challenges in an accurate model of occupants' behavior are the reliability of the models and the difficulty to handle both behaviors and their resulting actions at the same time. Because they are models, they simplify the reality [141] while the reality of occupants' behavior is not as rational as building physics and depends on many inter-dependent parameters that are difficult to quantify [142].

1.3.4 Conclusions on energy performance gap

A major challenge in building energy modeling is to provide an accurate and reliable picture of the reality. As a model, BEM remains a simplification of a real case study and the PG, the difference between simulations and measurements can be significant. The PG originate in the design, construction, and operation of buildings.

In building post-occupancy, origins of PG mostly come from the energy modeling process and over-simplification of models, the characterization of the human impact on energy consumption, the lack of general knowledge on studied buildings that requires a more extensive data collection and the consecutive data collection means.

Most strategies to reduce the PG in building post-occupancy focus on enhancing BEM with a calibration process. Strategies consider i) the modeling process itself to select the fittest modeling technique for a given application and case study, ii) the collection and analysis of more operation data, iii) the accurate characterization of occupants' behaviors and their impact on building energy consumption. These aspects play a major role in providing a reliable – then calibrated – BEM. For these reasons, the next section focuses on BEM calibration, and specifically the benefits of the integration of measured data in BEM.

1.4 Calibration of building energy models

1.4.1 Identifying uncertainties

The calibration of energy models addresses the reduction of the PG. However, in order to properly calibrate a model, it is important to understand the parameters to target through the calibration and the origins of the modeling uncertainties. There are four commonly reported categories of uncertainties that can be the focus of a calibration [134]: specification uncertainties, modeling uncertainties, scenario uncertainties and numerical uncertainties. Reddy et al. [130] also suggested a classification and grouped specification, modeling and scenario uncertainties under the name, "external errors", opposed to "internal errors" with numerical uncertainties.

Uncertainties can be summarized as follow. Specification uncertainties relate to building-specific parameters and the way they are filled in a model. They are caused by incorrect parameters or incompletely information in a model because of limited knowledge on a case studies or the user's limited skills. Building-specific parameters include, for instance, the glazing system, the envelope, heating and cooling systems, or HVAC

1.4 Calibration of building energy models

units. Modeling uncertainties are linked to the building energy modeling process: the selected method, the level of detail and the simplifications. It is up to the user to decide what is the most appropriate modeling technique, and how its choices drive on modeling uncertainties. Scenario uncertainties are about building usage and operating conditions. They relate to all the aspects of building operation, meaning everything else than the physical and technical description of the building: weather, occupancy, occupants' behavior, lighting schedules, for instance. Some are easily accessible (weather files are widely available), some can be simply adjusted (equipment and system operation can be checked onsite) but others (such as occupancy and energy-driving behaviors) are much more difficult to assess. It results in many assumptions and therefore large uncertainties. Finally, numerical uncertainty are the errors introduced when the calculations are discretised and due the simplifications of simulations with assumptions in the calculations performed by the modeling tool [143].

1.4.2 Calibration methods

Calibration is the process to better fit building energy simulation results to the real energy consumption. As presented in section 1.3.3.1, it tunes parameters for which there is an uncertainty and to find the value for these parameters that gives the most accurate simulation results.

Reddy et al. [130] first highlighted four categories of calibration methods: i) a "manual, iterative and pragmatic intervention", ii) a "suite of informative graphical comparative display", iii) "special tests and analytical procedure" and iv) "analytical/mathematical methods". The first category uses data from bills and short-term monitoring complemented with audits. The second is based on the comparison of graphical representations of monthly energy consumption from simulations and bills. The third approach considers specific tests: intrusive blink-tests (triggering on and off specific equipment with very-short-term monitoring to disaggregate building power demand), short-term energy monitoring of specific end-uses over a three-to-five-day timeframe, signature analysis method to compare more detailed measurements with simulations. Finally, analytical and mathematical methods are based on specific algorithms for optimization problem solving.

Fumo [10] proposed a broader picture, with a classification between the calibrated approach relying on data collected onsite, opposed to the forward approach for which data are unavailable. It depends on the comparison with tests from national standards such as ASHRAE or the *Règlementation Thermique* in France. In the recent research, the focus

has been given to the forward approach, further divided between manual and automated methods [143]. The former involves all methods with a predominant intervention from the user, and including the first two categories and part of the third category of methods from Reddy et al. [130]. The latter groups methods that are not user-driven, then mostly relying on statistical and mathematical tools.

A key aspect of calibration is the complexity level of the calibration process, based on the available data. According to Fabrizio and Monetti [144], there are five complexity levels. The first level only uses bills and as-built data. Complexity gradually increases with the integration of site visits, detailed audits, short-term monitoring, and long-term monitoring for level two, three, four and five, respectively. Level four and five figure the most detailed processes. Because they rely on instrumentation set ups, they can target other parameters than the sole energy consumption. The difference between the two complexity levels is that level four uses several-month-long data acquisition, usually with hourly to sub-hourly acquisition time-step of different energy-related parameters, while Level five uses permanently installed metering systems with a significantly larger number of equipment. With a level five complexity level, the instrumentation is more likely to be managed through building energy management system (BEMS) [5].

A significant number of studies have reported on calibration methods and applications. Detailed reviews on the different applied techniques can be found in the work of Coackley et al. [143] and Chong et al. [145]. Regardless of the implemented method, the main idea remains the same: to perform the variation of specific parameters in a BEM and observe the impact of this variation on the output of the model. Procedural manual methods, specifically sensitivity analyses, are effective on offering an overview of the impact of input parameters on BEM outputs [44]. However, the point is not to fudge a model with over-tuning and to lose credibility in the results. The identification of energy-driving parameters with large uncertainties must be carefully performed and modifications should be implemented according to reliable information [143].

Furthermore, regardless of the calibration technique, it remains a complex process with several issues. These issues include the identification of uncertainties and level of detail in available input data, as previously reported. It also depends on the complexity of the building energy model, the selected method and, mostly, and the user's experience [144].

1.4.3 Conclusions on the calibration of building energy models

The calibration of building energy models follows a two-step process. The first step identifies uncertainties in a model out of four categories: specification, modeling, scenario, and numerical uncertainties.

Once the uncertainties are identified comes the choice of a calibration method. The goal is to use additional data and information to input in the model for simulation results to become more realistic. The complexity of the calibration depends on the level of detail in the model and the type of available input data. Calibration methods can either be manual or automated. Manual methods heavily rely on the user whose skills and expertise are driving the choices when the model is implemented, and when changes and comparison are performed in order to calibrate the energy model. Automated techniques rely on algorithms to identify and tune specific parameters in the model, and to fit the simulation output to collected energy consumption data.

Regardless of the calibration method, the user's judgment has a significant impact on the results since physics-based building energy models are implemented by hand. Then, the skills and experience of the user are also the key to avoid discrepancies in the models, to save time and to provide a clear interpretation of the simulation results.

Conclusions

To implement a building energy model, three aspects must be considered: the modeling process, the closeness between the simulation results and the measured energy consumption, and the way to bridge the gap between models and measured data.

The literature review highlighted a variety of modeling techniques related to three main categories of models: physics-based, data-driven and hybrid models. All three approaches and specific methods have pros and cons. However, considering the context of the present study, physics-based modeling can be better fitted. The goal is to understand, quantify and check the impact of retrofit actions on building energy consumption. Because these were conducted on almost all aspects of our building case study, there are many different parameters involved. Therefore, there is a need for a detailed knowledge, understanding and control over these parameters that data-driven techniques cannot offer. Hybrid modeling also could be a relevant option. However, it still blacks out part of the building energy model and cannot be considered for a fully interpretable energy

BIBLIOGRAPHY

analysis.

Once the choice of a modeling approach is defined, the performance gap must be considered. It is expected to observe a difference between simulations and the measured building energy consumption. The two key aspects of the performance gap are its origins and the strategies to close it. Because the case study of this research project was recently targeted by energy retrofit actions, the performance gap may come from all of three stages of the building life-cycle – design, construction and operation. However, there is no opportunity during the project for control nor a close follow up of the retrofit work, apart from the collection of documentation. Therefore, considered strategies to close the performance gap can only be related to building operation: building energy modeling, data collection and integration of occupants' behavior. The former is decided with respect to the literature review of building energy modeling applications, as discussed above. However, there is a gap in research for the collection of sufficient operation data, to gain substantial knowledge over the building operation, specifically considering occupants' behavior. Such a data collection aims for the calibration of BEM with the addition of more realistic information on building operation and to replace initial assumptions.

The need of building operation data leads to the implementation of an exhaustive instrumentation solution to target and characterize building energy consumption, indoor environment quality (IEQ), the local weather and occupants' behavior, described in the following chapter.

Bibliography

- [1] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, “State of the art in building modelling and energy performances prediction: A review,” *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 272–288, jul 2013. 14, 15, 16, 19, 32
- [2] L. G. Swan and V. I. Ugursal, “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques,” *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 1819–1835, oct 2009. 14, 16, 32
- [3] C. Deb and A. Schlueter, “Review of data-driven energy modelling techniques for building retrofit,” jul 2021. 14, 16
- [4] P. de Wilde, “The gap between predicted and measured energy performance of buildings: A framework for investigation,” *Automation in Construction*, vol. 41, pp. 40–49, may 2014. 14, 33, 34, 35

BIBLIOGRAPHY

- [5] D. Guyot, F. Giraud, F. Simon, D. Corgier, C. Marvillet, and B. Tremeac, “Detailed monitoring as an essential tool for achieving energy performance targets in operation conditions: The HIKARI case study,” *E3S Web of Conferences*, vol. 172, 2020. 15, 42
- [6] The MathWorks Incorporation, “Mastering Machine Learning — A Step-by-Step Guide with Matlab,” tech. rep., 2019. 15
- [7] M. Bourdeau, X.-Q. X. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, “Modeling and forecasting building energy consumption: A review of data-driven techniques,” *Sustainable Cities and Society*, vol. 48, p. 101533, jul 2019. 16, 22, 24, 39
- [8] G. Tardioli, R. Kerrigan, M. Oates, J. O’Donnell, and D. Finn, “Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level,” *Energy Procedia*, vol. 78, pp. 3378–3383, nov 2015. 16, 32
- [9] ASHRAE, *2009 ASHRAE HANDBOOK FUNDAMENTALS I-P Edition — Energy estimating and modeling methods [chapter 19]*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2009. 16, 32
- [10] N. Fumo, “A review on the basics of building energy estimation,” *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 53–60, mar 2014. 16, 32, 41
- [11] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, “A review on time series forecasting techniques for building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017. 16, 22, 25, 32
- [12] R. J. Hyndman and A. B. Koehler, “Another look at forecast-accuracy metrics for intermittent demand,” *International Journal of Applied Forecasting*, no. 4, pp. 43–46, 2006. 18
- [13] D. Liu, Q. Chen, and K. Mori, “Time series forecasting method of building energy consumption using support vector regression,” in *2015 IEEE International Conference on Information and Automation*, pp. 1628–1632, IEEE, aug 2015. 18, 27
- [14] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, “Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique,” *Energy and Buildings*, vol. 126, pp. 94–103, aug 2016. 18, 24, 28
- [15] IZUBA, “Pleiades — Des outils logiciels pour l’optimisation énergétique et environnementale des bâtiments.” <https://www.izuba.fr/logiciels/outils-logiciels/>. 19

BIBLIOGRAPHY

- [16] C.-M. M. Hsieh, J.-J. J. Li, L. Zhang, and B. Schwegler, “Effects of tree shading and transpiration on building cooling energy use,” *Energy and Buildings*, vol. 159, pp. 382–397, jan 2018. 19
- [17] A. H. Neto and F. A. S. Fiorelli, “Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption,” *Energy and Buildings*, vol. 40, pp. 2169–2176, jan 2008. 21, 25
- [18] Y. Pan, Z. Huang, and G. Wu, “Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai,” *Energy and Buildings*, vol. 39, pp. 651–657, jun 2007. 21
- [19] P. Zhu and D. Yan, “Adapting LT-Method for Building Energy Prediction in China,” *Procedia Engineering*, vol. 205, pp. 3–10, jan 2017. 21
- [20] L. Jankovic, “Lessons learnt from design, off-site construction and performance analysis of deep energy retrofit of residential buildings,” *Energy and Buildings*, vol. 186, pp. 319–338, mar 2019. 21, 36
- [21] U. I. Dar, L. Georges, I. Sartori, and V. Novakovic, “Influence of occupant’s behavior on heating needs and energy system performance: A case of well-insulated detached houses in cold climates,” *Building Simulation*, vol. 8, pp. 499–513, oct 2015. 21, 35
- [22] K.-U. Ahn, D.-W. Kim, C.-S. Park, and P. de Wilde, “Predictability of occupant presence and performance gap in building energy simulation,” *Applied Energy*, vol. 208, pp. 1639–1652, dec 2017. 21
- [23] H.-x. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 3586–3592, aug 2012. 21, 32
- [24] A. Rice, S. Hay, and D. Ryder-Cook, “A limited-data model of building energy consumption,” in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building - BuildSys ’10*, (New York, New York, USA), p. 67, ACM Press, 2010. 21
- [25] A. Tsanas and A. Xifara, “Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools,” *Energy and Buildings*, vol. 49, pp. 560–567, jun 2012. 22
- [26] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006. 22
- [27] F. Wahid and D. Kim, “A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings,” *International Journal of Smart Home*, vol. 10, no. 2, pp. 97–108, 2016. 22, 25, 26

BIBLIOGRAPHY

- [28] C. Fan, F. Xiao, and S. Wang, “Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques,” *Applied Energy*, vol. 127, pp. 1–10, aug 2014. 24, 25, 28
- [29] K. A. Ross, C. S. Jensen, R. Snodgrass, C. E. Dyreson, and L. Chen, “Cross-Validation,” in *Encyclopedia of Database Systems*, pp. 532–538, Boston, MA: Springer US, 2009. 24
- [30] Z. Wang, Y. Wang, and R. S. Srinivasan, “A novel ensemble learning approach to support building energy use prediction,” *Energy and Buildings*, vol. 159, pp. 109–122, jan 2018. 24, 27, 28
- [31] M. Bourdeau, X. Guo, and E. Nefzaoui, “Buildings energy consumption generation gap: A post-occupancy assessment in a case study of three higher education buildings,” *Energy and Buildings*, vol. 159, 2018. 25, 38
- [32] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, “Deep learning for estimating building energy consumption,” *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016. 25
- [33] E. Mocanu, P. H. Nguyen, W. L. Kling, and M. Gibescu, “Unsupervised energy prediction in a Smart Grid context using reinforcement cross building transfer learning,” *Energy and Buildings*, vol. 116, pp. 646–655, mar 2016. 25, 29, 38
- [34] M. Bourdeau, P. Basset, S. Beauchêne, D. Da Silva, T. Guiot, D. Werner, and E. Nefzaoui, “Classification of daily electric load profiles of non-residential buildings,” *Energy & Buildings*, vol. 233, p. 110670, 2021. 25, 38
- [35] F. Massa Gray and M. Schmidt, “A hybrid approach to thermal building modelling using a combination of Gaussian processes and grey-box models,” *Energy and Buildings*, vol. 165, pp. 56–63, apr 2018. 25
- [36] G. R. Newsham, B. J. Birt, C. Arsenault, L. Thompson, J. Veitch, S. Mancini, A. Galasiu, I. Macdonald, and G. Burns, “Do green buildings outperform conventional buildings? Indoor environment and energy performance in North American offices (Report RR-329),” tech. rep., 2012. 25, 34
- [37] Z. Ma, J. Song, and J. Zhang, “Energy consumption prediction of air-conditioning systems in buildings by selecting similar days based on combined weights,” *Energy & Buildings*, vol. 151, pp. 157–166, sep 2017. 25, 26
- [38] R. Mena, F. Rodríguez, M. Castilla, and M. Arahál, “A prediction model based on neural networks for the energy consumption of a bioclimatic building,” *Energy and Buildings*, vol. 82, pp. 142–155, oct 2014. 25

BIBLIOGRAPHY

- [39] F. Tang, A. Kusiak, and X. Wei, “Modeling and short-term prediction of HVAC system with a clustering algorithm,” *Energy and Buildings*, vol. 82, pp. 310–321, oct 2014. 25
- [40] D. Toffanin, *Generation of customer load profiles based on smart-metering time series , building-level data and aggregated measurements*. PhD thesis, Swiss Federal Institute of Technology (ETH), 2016. 25
- [41] P. Dagnely, T. Ruetten, T. Tourwé, and E. Tsiporkova, “Predicting hourly energy consumption. Can you beat an autoregressive model?,” *Proceeding of the 24th Annual Machine Learning Conference of Belgium and the Netherlands, Benelearn, Delft, The Netherlands*, 2015. 25
- [42] C. Fan, F. Xiao, and Y. Zhao, “A short-term building cooling load prediction method using deep learning algorithms,” *Applied Energy*, vol. 195, pp. 222–233, jun 2017. 25
- [43] K. Yun, R. Luck, P. J. Mago, and H. Cho, “Building hourly thermal load prediction using an indexed ARX model,” *Energy and Buildings*, vol. 54, pp. 225–233, nov 2012. 25
- [44] M. H. Kristensen and S. Petersen, “Choosing the appropriate sensitivity analysis method for building energy model-based investigations,” *Energy and Buildings*, vol. 130, pp. 166–176, oct 2016. 25, 42
- [45] K. Li, X. Xie, W. Xue, X. Dai, X. Chen, and X. Yang, “A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction,” *Energy and Buildings*, vol. 174, pp. 323–334, 2018. 25
- [46] M. Nilashi, M. Dalvi-Esfahani, O. Ibrahim, K. Bagherifard, A. Mardani, and N. Zakuan, “A soft computing method for the prediction of energy performance of residential buildings,” *Measurement*, vol. 109, pp. 268–280, oct 2017. 25
- [47] C. Deb and S. E. Lee, “Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data,” *Energy and Buildings*, vol. 159, pp. 228–245, jan 2018. 25
- [48] S. Paudel, M. Elmitri, S. Couturier, P. H. Nguyen, R. Kamphuis, B. Lacarrière, and O. Le Corre, “A relevant data selection method for energy consumption prediction of low energy building based on support vector machine,” *Energy and Buildings*, vol. 138, pp. 240–256, mar 2017. 25, 27
- [49] P. R. Winters, “Forecasting Sales by Exponentially Weighted Moving Averages,” *Management Science*, vol. 6, pp. 324–342, apr 1960. 25
- [50] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis : forecasting and control*. John Wiley, 2008. 25

BIBLIOGRAPHY

- [51] K. Jeong, C. Koo, and T. Hong, “An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network),” *Energy*, vol. 71, pp. 71–79, 2014. 26
- [52] G. R. Newsham and B. J. Birt, “Building-level occupancy data to improve ARIMA-based electricity use forecasts,” in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building - BuildSys '10*, (New York, New York, USA), p. 13, ACM Press, 2010. 26
- [53] S. Sekhar Roy, R. Roy, and V. E. Balas, “Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 4256–4268, feb 2018. 26
- [54] R. Bindiu, M. Chindriú, and G. Pop, “Day-Ahead Load Forecasting Using Exponential Smoothing,” *Scientific Bulletin of the Petru Maior University of Tirgu Mures*, vol. 6, 2009. 26
- [55] S. Ariffin, A. Karim, and S. A. Alwi, “Electricity Load Forecasting in UTP Using Moving Averages and Exponential Smoothing Techniques,” *Applied Mathematical Sciences*, vol. 7, no. 80, pp. 4003–4014, 2013. 26
- [56] E. Fix and J. L. Hodges, Jr, “Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties,” 1951. 26
- [57] O. Valgaev and F. Kupzog, “Building power demand forecasting using K-nearest neighbors model - initial approach,” in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1055–1060, IEEE, oct 2016. 26
- [58] D. Lachut, N. Banerjee, and S. Rollins, “Predictability of Energy Use in Homes,” in *International Green Computing Conference*, pp. 1–10, IEEE, nov 2014. 26
- [59] J.-S. Chou and D.-K. Bui, “Modeling heating and cooling loads by artificial intelligence for energy-efficient building design,” *Energy and Buildings*, vol. 82, pp. 437–446, oct 2014. 27
- [60] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995. 27
- [61] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, “Short-term load forecasting in a non-residential building contrasting models and attributes,” *Energy and Buildings*, vol. 92, pp. 322–330, apr 2015. 27

BIBLIOGRAPHY

- [62] J. Q. Zhang, “Sampling for building energy consumption with fuzzy theory,” *Energy and Buildings*, vol. 156, pp. 78–84, dec 2017. 27
- [63] X. Chen and H. Yang, “Integrated energy performance optimization of a passively designed high-rise residential building in different climatic zones of China,” *Applied Energy*, vol. 215, pp. 145–158, apr 2018. 27
- [64] Y. Fu, Z. Li, H. Zhang, and P. Xu, “Using Support Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices,” *Procedia Engineering*, vol. 121, pp. 1016–1022, jan 2015. 27
- [65] F. Magoulès and H.-X. Zhao, *Data Mining and Machine Learning in Building Energy Analysis*. Wiley-ISTE, 2016. 27
- [66] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, may 2015. 27
- [67] M. L. Chalal, M. Benachir, M. White, and R. Shrahily, “Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review,” *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 761–776, oct 2016. 27, 32
- [68] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, “Random Forest based hourly building energy prediction,” *Energy and Buildings*, vol. 171, pp. 11–25, 2018. 28
- [69] S. Papadopoulos, E. Azar, W.-L. Woon, and C. E. Kontokosta, “Evaluation of tree-based ensemble learning algorithms for building energy performance estimation,” *Journal of Building Performance Simulation*, pp. 1–11, jul 2017. 28
- [70] M. H. Alobaidi, F. Chebana, and M. A. Meguid, “Robust ensemble learning framework for day-ahead forecasting of household based energy consumption,” *Applied Energy*, vol. 212, pp. 997–1012, feb 2018. 28
- [71] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, and F. Hussin, “Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review,” *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1108–1118, apr 2017. 28, 32
- [72] M. Castelli, L. Trujillo, L. Vanneschi, and A. Popovič, “Prediction of energy performance of residential buildings: A genetic programming approach,” vol. 102, pp. 67–74, 2015. 28
- [73] K. Li, C. Hu, G. Liu, and W. Xue, “Building’s electricity consumption prediction using optimized artificial neural networks and principal component analysis,” *Energy and Buildings*, vol. 108, pp. 106–113, dec 2015. 28

BIBLIOGRAPHY

- [74] C. Fan, F. Xiao, Z. Li, and J. Wang, “Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review,” *Energy and Buildings*, vol. 159, pp. 296–308, jan 2018. 29
- [75] L. Busoniu, D. Ernst, B. De Schutter, and R. Babuska, “Approximate reinforcement learning: An overview,” in *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pp. 1–8, IEEE, apr 2011. 29
- [76] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, oct 2010. 29
- [77] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, “Physical energy and data-driven models in building energy prediction: A review,” *Energy Reports*, vol. 8, pp. 2656–2671, nov 2022. 29, 31, 32
- [78] Y. Li, Z. O’Neill, L. Zhang, J. Chen, P. Im, and J. DeGraw, “Grey-box modeling and application for building energy simulations - A critical review,” *Renewable and Sustainable Energy Reviews*, vol. 146, p. 111174, aug 2021. 30, 32
- [79] S. Hassid, “A linear model for passive solar calculations: Evaluation of performance,” *Building and Environment*, vol. 20, no. 1, pp. 53–59, 1985. 30
- [80] V. Siddharth, P. Ramakrishna, T. Geetha, and A. Sivasubramaniam, “Automatic generation of energy conservation measures in buildings using genetic algorithms,” *Energy and Buildings*, vol. 43, pp. 2718–2726, oct 2011. 30
- [81] W. O. Collinge, J. C. DeBlois, A. E. Landis, L. A. Schaefer, and M. M. Bilec, “Hybrid Dynamic-Empirical Building Energy Modeling Approach for an Existing Campus Building,” *Journal of Architectural Engineering*, vol. 22, p. 04015010, mar 2016. 30
- [82] B. Dong, Z. Li, S. M. Rahman, and R. Vega, “A hybrid model approach for forecasting future residential electricity consumption,” *Energy and Buildings*, vol. 117, pp. 341–351, 2016. 30
- [83] M. Rätz, A. P. Javadi, M. Baranski, K. Finkbeiner, and D. Müller, “Automated data-driven modeling of building energy systems via machine learning algorithms,” nov 2019. 30
- [84] L. Pedersen, “Use of different methodologies for thermal load and energy estimations in buildings including meteorological and sociological input parameters,” *Renewable and Sustainable Energy Reviews*, vol. 11, pp. 998–1007, jun 2007. 32

BIBLIOGRAPHY

- [85] Z. Wang and R. S. Srinivasan, “A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models,” *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 796–808, aug 2017. 32
- [86] B. Yildiz, J. Bilbao, and A. Sproul, “A review and analysis of regression and machine learning models on commercial building electricity load forecasting,” *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 1104–1122, jun 2017. 32
- [87] T. Ahmad, H. Chen, Y. Guo, and J. Wang, “A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review,” *Energy and Buildings*, vol. 165, pp. 301–320, apr 2018. 32
- [88] K. Amasyali and N. M. El-Gohary, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, jan 2018. 32
- [89] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, feb 2018. 32
- [90] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, “Conventional models and artificial intelligence-based models for energy consumption forecasting: A review,” *Journal of Petroleum Science and Engineering*, vol. 181, no. March, p. 106187, 2019. 32
- [91] M. Bourdeau, D. Werner, P. Basset, and E. Nefzaoui, “A Sensor Network for Existing Residential Buildings Indoor Environment Quality and Energy Consumption Assessment and Monitoring - Lessons Learnt from a Field Experiment,” no. *Sensornets*, pp. 105–112, 2020. 32, 36, 38
- [92] Y. Sun, F. Haghghat, and B. C. Fung, “A review of the-state-of-the-art in data-driven approaches for building energy prediction,” *Energy and Buildings*, vol. 221, p. 110022, aug 2020. 32
- [93] H. Fu, J. C. Baltazar, and D. E. Claridge, “Review of developments in whole-building statistical energy consumption models for commercial buildings,” *Renewable and Sustainable Energy Reviews*, vol. 147, p. 111248, sep 2021. 32
- [94] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, and W. Livingood, “A review of machine learning in building load prediction,” *Applied Energy*, vol. 285, p. 116452, mar 2021. 32
- [95] A. A. Al-Shargabi, A. Almhafdy, D. M. Ibrahim, M. Alghieth, and F. Chiclana, “Buildings’ energy consumption prediction models based on buildings’ characteristics: Research

BIBLIOGRAPHY

- trends, taxonomy, and performance measures,” *Journal of Building Engineering*, vol. 54, p. 104577, aug 2022. 32
- [96] C. Lu, S. Li, and Z. Lu, “Building energy prediction using artificial neural networks: A literature survey,” *Energy and Buildings*, vol. 262, p. 111718, may 2022. 32
- [97] PROFEEL, “Programme PROFEEL — Innover pour la rénovation énergétique.” <http://programmeprofeel.fr/>, 2022. 31, 35, 36
- [98] P. X. Zou, X. Xu, J. Sanjayan, and J. Wang, “Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives,” *Energy and Buildings*, vol. 178, pp. 165–181, 2018. 33, 34, 35, 36, 38, 39
- [99] Carbon Trust, “Closing the Gap : Lessons learned on realising the potential of low carbon building design,” 2011. 34, 35
- [100] S. Niu, W. Pan, and Y. Zhao, “A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap,” *Sustainable Cities and Society*, vol. 27, pp. 275–286, nov 2016. 34, 36
- [101] P. van den Brom, A. Meijer, and H. Visscher, “Performance gaps in energy consumption: household groups and building characteristics,” *Building Research & Information*, vol. 46, pp. 54–70, jan 2018. 34
- [102] Zero Carbon Hub, “Closing the Gap Between Design and As-built Performance,” no. July, p. 44, 2014. 34
- [103] R. Gupta and M. Gregg, “Empirical evaluation of the energy and environmental performance of a sustainably-designed but under-utilised institutional building in the UK,” *Energy and Buildings*, vol. 128, pp. 68–80, sep 2016. 34
- [104] A. Oti, E. Kurul, F. Cheung, and J. Tah, “A framework for the utilization of Building Management System data in building information models for building design and operation,” *Automation in Construction*, vol. 72, pp. 195–210, dec 2016. 34
- [105] P. America, P. van de Laar, and G. Muller, “Experiences in evolvability research,” *Advanced Engineering Informatics*, vol. 26, pp. 478–486, aug 2012. 34
- [106] C. Gorse, M. Brooke-Peat, J. Parker, and F. Thomas, “Building Simulation and Models: Closing the Performance Gap,” in *Building Sustainable Futures* (Springer, ed.), pp. 209–226, Springer International Publishing, 2016. 34

BIBLIOGRAPHY

- [107] A. C. Menezes, A. Cripps, D. Bouchlaghem, and R. Buswell, “Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap,” *Applied Energy*, vol. 97, pp. 355–364, sep 2012. 34, 35
- [108] T. Dwyer, “Knowledge is Power: Benchmarking and prediction of building energy consumption,” *Building Services Engineering Research and Technology*, vol. 34, pp. 5–7, feb 2013. 34, 35
- [109] T. J. Williamson, “Predicting building performance: The ethics of computer simulation,” *Building Research and Information*, vol. 38, no. 4, pp. 401–410, 2010. 35
- [110] G. R. Newsham, S. Mancini, and B. J. Birt, “Do LEED-certified buildings save energy? Yes, but...,” *Energy and Buildings*, vol. 41, pp. 897–905, aug 2009. 35
- [111] M. Bell, J. Wingfield, D. Miles-Shenton, and J. Seavers, “Low carbon housing: lessons from Elm Tree Mews,” tech. rep., nov 2010. 35
- [112] C. Turner and M. Frankel, “Energy Performance of LEED for New Construction Buildings,” tech. rep., 2008. 35
- [113] R. Bunn, “Soft Landings for Sustainable Buildings — Designing For Operational Outcomes,” *Building Services Research and Information Association and Usable Building Trust*, 2010. 35
- [114] CSTB, “ISABELE mesure la performance énergétique intrinsèque d’un logement neuf.” <https://www.cstb.fr/fr/actualites/detail/isabele-performance-energetique-logement-neuf-0916/>, 2016. 35
- [115] F. Haldi and D. Robinson, “On the behaviour and adaptation of office occupants,” *Building and Environment*, vol. 43, pp. 2163–2177, dec 2008. 35
- [116] A. Molin, P. Rohdin, and B. Moshfegh, “Investigation of energy performance of newly built low-energy buildings in Sweden,” *Energy and Buildings*, vol. 43, pp. 2822–2831, oct 2011. 35
- [117] A. Dasgupta, A. Prodromou, and D. Mumovic, “Operational versus designed performance of low carbon schools in England: bridging a credibility gap,” *HVAC&R Research*, vol. 18, no. 1-2: Indoor Air Quality, Ventilation and Energy Conservation in Buildings: Innovation and Integration (Part 2), pp. 37–50, 2012. 35
- [118] C. Peng, D. Yan, R. Wu, C. Wang, X. Zhou, and Y. Jiang, “Quantitative description and simulation of human behavior in residential buildings,” *Building Simulation*, vol. 5, pp. 85–94, jun 2012. 35

BIBLIOGRAPHY

- [119] C. Demanuele, T. Tweddell, and M. Davies, “Bridging the gap between predicted and actual energy performance in schools,” *Low Energy Architecture (LEA)*, no. September, pp. 1–6, 2010. 36
- [120] Ministère de la Transition Ecologique et Solidaire, “Décret 2010-1022 du 31 août 2010 relatif aux dispositifs de comptage sur les réseaux publics d’électricité.” <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000022765140/>, 2016. 36
- [121] European Commission, “Smart Metering deployment in the European Union,” 2019. 36
- [122] European Union, “Directive of 2009/72/EC of the European Parliament and of the Council of 13 July 2009 Concerning Common Rules for the Internal Market in Electricity and Repealing Directive 2003/54/EC,” *Official Journal of the European Union*, vol. L211, no. August, pp. L 211/55 – L 211/93, 2009. 36
- [123] Agence Parisienne du Climat, “Performance énergétique en copropriété - Retour sur deux ans d’instrumentation,” tech. rep., 2020. 36
- [124] J. Yang, M. Santamouris, and S. E. Lee, “Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings,” *Energy and Buildings*, vol. 121, no. January, pp. 344–349, 2016. 36, 38, 39
- [125] L. B. Saad and B. Tourancheau, “Multiple mobile sinks positioning in wireless sensor networks for buildings,” in *Proceedings - 2009 3rd International Conference on Sensor Technologies and Applications, SENSORCOMM 2009*, pp. 264–270, IEEE Computer Society, 2009. 36
- [126] J. K. Day and D. E. Gunderson, “Understanding high performance buildings: The link between occupant knowledge of passive design systems, corresponding behaviors, occupant comfort and environmental satisfaction,” *Building and Environment*, vol. 84, pp. 114–124, jan 2015. 36
- [127] T. S. Blight and D. A. Coley, “Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house dwellings,” *Energy and Buildings*, vol. 66, pp. 183–192, nov 2013. 36
- [128] P. G. Tuohy and G. B. Murphy, “Closing the gap in building performance: Learning from BIM benchmark industries,” *Architectural Science Review*, vol. 58, pp. 47–56, jan 2015. 36
- [129] R. Cohen and B. Bordass, “Mandating transparency about building energy performance in use,” *Building Research and Information*, vol. 43, pp. 534–552, jul 2015. 36

BIBLIOGRAPHY

- [130] T. A. Reddy, “Literature review on calibration of building energy simulation programs: Uses, problems, procedure, uncertainty, and tools,” in *2006 Winter Meeting of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE* (ASHRAE Transactions, ed.), (Chicago, IL, United States), pp. 226–240, 2006. 37, 40, 41, 42
- [131] P. Biddulph, V. Gori, C. A. Elwell, C. Scott, C. Rye, R. Lowe, and T. Oreszczyn, “Inferring the thermal resistance and effective thermal mass of a wall using frequent temperature and heat flux measurements,” *Energy and Buildings*, vol. 78, pp. 10–16, 2014. 38
- [132] S. Shin, S. Jeong, J. Lee, S. W. Hong, and S. Jung, “Pre-Occupancy Evaluation based on user behavior prediction in 3D virtual simulation,” *Automation in Construction*, vol. 74, pp. 55–65, feb 2017. 38
- [133] M. Ribeiro, K. Grolinger, H. F. ElYamany, W. A. Higashino, and M. A. Capretz, “Transfer learning with seasonal and trend adjustment for cross-building energy forecasting,” *Energy and Buildings*, vol. 165, pp. 352–363, apr 2018. 38
- [134] D. Guyot, F. Giraud, F. Simon, D. Corgier, C. Marvillet, and B. Tremeac, “Building energy model calibration: A detailed case study using sub-hourly measured data,” *Energy and Buildings*, vol. 223, p. 110189, sep 2020. 38, 40
- [135] GRDF, “Compteur gaz communicant Gazpar : comment ça marche?.” <https://www.grdf.fr/particuliers/fonctionnement-compteur-gaz-communicant-grdf>. 38
- [136] ENEDIS, “Linky, le compteur communicant.” <https://www.enedis.fr/linky-compteur-communicant>. 38
- [137] M. Jia, R. S. Srinivasan, and A. A. Raheem, “From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency,” feb 2017. 38
- [138] A. L. Pisello and F. Asdrubali, “Human-based energy retrofits in residential buildings: A cost-effective alternative to traditional physical strategies,” *Applied Energy*, vol. 133, pp. 224–235, nov 2014. 39
- [139] S. Carlucci, M. De Simone, S. K. Firth, M. B. Kjærgaard, R. Markovic, M. S. Rahaman, M. K. Annaqeeb, S. Biandrate, A. Das, J. W. Dziedzic, G. Fajilla, M. Favero, M. Ferrando, J. Hahn, M. Han, Y. Peng, F. Salim, A. Schlüter, and C. van Treeck, “Modeling occupant behavior in buildings,” *Building and Environment*, vol. 174, no. February, p. 106768, 2020. 39

BIBLIOGRAPHY

- [140] T. Hong, S. C. Taylor-Lange, S. D'Oca, D. Yan, and S. P. Corgnati, "Advances in research and applications of energy-related occupant behavior in buildings," mar 2016. 39
- [141] W. O'Brien and H. B. Gunay, "The contextual factors contributing to occupants' adaptive comfort behaviors in offices - A review and proposed modeling framework," *Building and Environment*, vol. 77, pp. 77–87, 2014. 39
- [142] P. Roques, *La question de la consommation d' énergie dans les logements sociaux réhabilités - Pratiques et identité*. PhD thesis, Université Côte d'Azur, 2016. 39
- [143] D. Coakley, P. Raftery, and M. Keane, "A review of methods to match building energy simulation models to measured data," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 123–141, sep 2014. 41, 42
- [144] E. Fabrizio and V. Monetti, "Methodologies and Advancements in the Calibration of Building Energy Models," *Energies 2015, Vol. 8, Pages 2548-2574*, vol. 8, pp. 2548–2574, mar 2015. 42
- [145] A. Chong, G. Augenbroe, and D. Yan, "Occupancy data at different spatial resolutions: Building energy performance and model calibration," *Applied Energy*, vol. 286, p. 116492, mar 2021. 42

**CHAPTER 1 – BUILDING ENERGY MODELING: METHODS
AND CHALLENGES**

Chapter 2

A sensor network for building energy monitoring

The following chapter is dedicated to the experimental part of the research work. To collect data for the calibration of energy models, a large wireless instrumentation solution is implemented over an existing case study. At first, a description of the building case study is given. It is followed by a review of studies using sensor networks as a mean to collect and target energy efficiency enhancement. A full overview of the sensor network is provided, with details on the types of measurements, the communication system and data storage. Finally, the chapter is concluded with a feedback on the operation and management of the network.

Introduction

Given the insights from the previous chapter, a relevant way to improve building energy models and specifically the calibration process, is to replace the assumptions in standard building operation data by measured field data. It implies that many different parameters should be monitored over a large observation timeframe. Then, the field data collection must be automated, and the fittest solution is a sensor network. Considering the building case study of the research project and its characteristics such a sensor network should solve two problems:

- *The comprehensiveness of the solution:* the sensor network must cover all or most aspects related to building energy consumption, with a greater depth of details than the usual sources of data (onsite inquiries, energy diagnoses, smart meters).
- *The inclusion of field constraints:* these constraints include the nature of the experimentation site – pre-existing, occupied and renovated buildings – and the project management – cost of the sensor network, maintenance and human resources.

To answer the first problem, there two aspects to address: the types of measurements, and the spatial and temporal granularity of the measurements. Four types of measurements can be targeted. The first and obvious category of measurements is the energy consumption. Heating, DHW, ventilation, specific electricity and cooking are the usual end-uses in residential buildings [1], considering there is no air cooling system for the studied buildings. Other appliances should be included as well, such as electrical and leisure goods, and any equipment that could result in indoor heat gains and depict occupants' behavior. Indoor environment quality, to characterize the indoor comfort, is the second type of measurements, including (but not limited to) the indoor air temperature (IAT), the relative humidity, the luminosity, solar gains, the radiant temperature, the CO₂ concentration or volatile organic compounds (VOC). Occupants' behaviors monitoring is the third target of measurements. It is by far the most complex aspect to characterize and it has a significant impact on building energy consumption, as discussed in Chapter 1. Finally, weather conditions are essential for building thermal energy behavior analysis. They usually drive thermal energy consumption, and building energy management is traditionally based on outdoor conditions. It is highlighted that physical building characteristics, such as the building envelope and the energy systems, can also be instrumented. In the present case, it is not part of the instrumentation campaign. Related information are retrieved from technical documentation. Then, spatial and temporal granularity of measurements depends on the target of the energy analysis and the types of acquired measurements. For example, with a very small electrical energy consumption acquisition

2.1 Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects

time-step (one-minute or lower), it is possible to deduce relevant information on energy-driving behaviors. Thermal energy consumption, on the opposite, is not expected to have sudden variations. Half-hour or hourly measurements would be sufficient. Spatial granularity comprises measurements at neighborhood scale (for the weather), building scale, housing scale and appliance scale. This strategy provides a variety of options and back-up solutions for data analyses.

The second problem that impacts on the sensor network relates to the choice of data collection technologies. Wireless or wired sensor network, in-house instrumentation or commercial solution, IoT or fully embedded systems (such as BEMS): these are some examples of technological options that impact on the cost, the maintenance and the management process of the sensor network. In this project, our solution is a wireless sensor network using IoT connected objects available on the professional market, and managed between our research team and contractors.

This chapter describes our wireless sensor network deployed in three residential buildings for field data acquisition over more than three years. The first section provides a review of instrumentation solutions and projects for building operation monitoring in a retrofit context. It highlights the gaps in the current research, to position our instrumentation solution. The second section introduces our case study. The third section details the deployment of our instrumentation solution, the measurement targets – energy consumption, indoor environment quality (IEQ), occupants' behavior, and local weather, the data communication and storage. It is complemented by Appendix B to provide more context content regarding the concepts of IoT, wireless communication protocols for building instrumentation and the calibration of the sensor network. Section four gives a feedback on the management of our sensor network to discuss achievements, limitations and future replicability.

2.1 Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects

This section provides a review on instrumentation solutions for academic research and industrial applications, to collect building operation data, including energy consumption, occupants' behavior and IEQ, preferably in a context of thermal renovation. It aims to highlight the gaps in the current research on building instrumentation, to also compare

with our strategy. It should be noted that this section does not aim a measurement-by-measurement review of the literature, but it targets studies with a broader instrumentation objective.

The review is divided between industrial and academic research projects. Academic and industrial research context may not have the same objectives. The former targets a feedback for professional applications, to disseminate broader knowledge to the public and acquire in-house expertise to develop commercial solution. The later tends to dig deeper into the details and to develop innovative methods for onsite data collection. It is also quite likely that the former may access to a larger financial support, that would impact on the instrumentation solution and data analysis results.

2.1.1 Industrial programs

2.1.1.1 PROFEEL: characterizing the impact of retrofit actions

PROFEEL [2] is a nine-project research program focusing on field-operative solutions to prepare and optimize energy retrofit planning in buildings, to assess energy savings and the performance of implemented retrofit actions. It stands for "*Programme de la Filière pour l'innovation en faveur des Economies d'Énergie dans le bâtiment*" (translated to "program of the building sector for innovation towards building energy savings"). PROFEEL started in 2019 to support massive energy renovation plans for existing buildings as targeted by the national building retrofit plan [3]. The program was funded by CSTB (*Centre Technique et Scientifique du Bâtiment*) and AQC (*Agence Qualité Construction*). It is supported by five major actors of the energy sector in France and in Europe: EDF (*Electricité De France*), ENGIE, TOTAL, ENI and CPCU (*Compagnie Parisienne de Chauffage Urbain*). It also involves several public institutes: ADEME, *Ministère de la transition écologique et solidaire* and *Ministère de la cohésion des territoires et des relations avec les collectivités territoriales*.

Two specific projects from the PROFEEL program have similar objectives to those of our research project.

The first project is called SEREINE (*Solution d'Évaluation de la performance Énergétique Intrinsèque des bâtiments*). It focuses on the development of innovative and efficient methods and tools to assess the performance of retrofit actions, based on the study and instrumentation of newly retrofitted individual and small-size collective residential buildings. Three outputs are expected. New methods for the characterization of the performance of the building envelope at building-scale are designed and adapted

2.1 Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects

from existing techniques to assess the thermal resistance of walls [4]. It requires a fully controlled environment (no occupants) and measurements of the indoor temperature, heating power demand and outdoor meteorological conditions. The goal is also to implement measurement campaigns over twenty-four hours and less, to provide accurate results. Similarly, innovative techniques are developed for the characterization of energy systems. The expected result should be a guideline for building energy experts to identify all energy systems in a given building and perform the adequate measurements for energy performance assessment. However, data acquisition would either result from punctual measurements or be extracted from technical documentation. There is no long-term instrumentation procedure reported in the method. The number of case studies ranges between forty and fifty houses to develop and optimize the proposed methods. Then, all collected data and information will be gathered in an open-access database.

The second project of interest is called QSE (*Qualité Sanitaire et Énergétique*). The goal is to develop and simplify in situ measurements methods to assess the impact of building energy retrofit actions, using IoT (Internet of Things) technologies and to target the optimal balance between health, comfort, and energy performance. Case studies include residential, commercial, office and school buildings. Up to date, the investigation includes twenty-one individual houses, six collective residential buildings, five schools and four office buildings. Unfortunately, there are no available details so far on the type of sensors and instrumentation solutions deployed, nor on the characteristics of case studies.

2.1.1.2 *Agence Parisienne du Climat*: energy performance in residential condominiums

An industrial instrumentation project has been conducted between 2017 and 2019 by *Agence Parisienne du Climat* (APC) to study the energy performance of condominiums in a context of energy retrofit planning [5]. The project was funded by six members – Energetech consulting firm, ADEME, *Ville de Paris*, *Métropole du Grand Paris*, *ALEC Plaine Commune* – together with APC. The project was supported by the ACE-Retrofitting European project.

The project aimed a large instrumentation campaign in eight residential condominiums located in Paris and its suburb. Considered buildings were built between 1914 and 1978. Therefore, most of them did not fit to any specific construction standards regarding energy consumption and efficiency. They comprised from 23 to 199 housings with a floor area between 2,662 m² and 13,385 m².

Measurements were conducted using two strategies. A "light" instrumentation program lasted for fourteen months over two heating seasons and in four condominiums. An "advanced" program lasted for twenty months over two heating seasons in the remaining four buildings. Measurements were conducted in a sixty-four-housing sample over the eight buildings. Measurements included electricity consumption – with optical-reading sensors on electricity meters, indoor temperature in living rooms and bedrooms, and punctual measurements of ventilation air flow rates at extraction units. The "advanced" instrumentation program added measurements of CO₂ and formaldehyde concentrations. Seven out of eight boiler rooms were instrumented as well with pulse counting on energy meters (the energy meters were installed prior to the project), hydraulic circuit temperatures and water flow measurements in the pipping circuits (using ultrasonic probes), electricity consumption data acquisition for the boiler room systems and punctual measurements of air flow in HVAC units. Finally, electricity consumption of common areas was monitored using optical-reading sensors on electricity meters and electrical switchboards supervision. Measurements were complemented using energy bills, reports from energy and architectural diagnoses, reports from regulatory calculations to assess the impact of energy retrofit actions, and project management portfolios.

The results of the instrumentation campaign provided four types of insights:

- A feedback on the monitoring systems;
- An analysis of the performances of retrofit actions: the comparison between measurements and energy models, and the assessment of performances of heating energy systems, DHW production, ventilation and electricity in common area of the buildings;
- A discussion on IEQ and comfort related the thermal renovation;
- Energy saving strategies and energy management optimization solutions.

2.1.2 Academic research projects

This section provides an overview of academic research projects for building instrumentation. Prior review work has been conducted on instrumentation solutions applied to building energy consumption, IEQ and occupants' behavior characterization. Ahmad et al. [6] highlighted the most implemented technologies of sensors and communication protocols recovered. They reported on sensor technologies for electricity and gas metering in the energy consumption category; air temperature, mean radiant temperature, relative humidity, indoor air velocity, IAQ (indoor air quality, with volatile organic compounds,

2.1 Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects

CO₂, CO and other oxides) and daylight for IEQ characterization; and occupancy for occupants' behavior assessment. They summarized common data communication protocols including Zigbee, power line carriers, M-Bus, the Ethernet, Modbus, GPRS, Wi-Fi and Bac net. It is a valuable review of technologies involved in sensor networks. However, applications provide other perspectives on the typical implementation of instrumentation strategies to collect field data. They also bring the light on practical issues that may be encountered.

Non-intrusive load monitoring (NILM), that aims to characterize electricity demand and electrical appliance signature [7], is voluntarily left apart from the review. There is a significant amount of research work on the topic [8]. However, instrumentation techniques are too specific. Although they give relevant perspectives on electricity monitoring, they target extremely fine data granularity (usually sub-minute) to observe electric appliance triggering which is beyond the focus of the data collection strategy implemented in this thesis. A focus is given on studies with a broader application range, targeting energy consumption, IEQ and occupants' behavior at apartment and building scales.

Studies can be divided into two categories: field tests and in-lab tests. The main topic of interest is the characterization of building energy performances. However, in-lab sensor network implementations and applications bring in a design perspective that may be relevant. A summary of reviewed studies is provided in Table 2.1. The work of Martín-Garín et al.[9], Karami et al. [10] and Frei et al. [11] relate to the latter category of studies. Authors describe the development of wireless Arduino-based sensor kits using open-source storage technologies and platforms. The former two studies focused on IEQ while the second focused on U-value measurements. Martín-Garín et al. designed a low-cost IEQ wireless sensor network system based on Wi-Fi communication. They tested their solution over seven days of one-minute time-step data in an apartment in Spain. Karami et al. differed with their testing set-up in a computer lab, and they added PM2.5 and volatile organic compounds (VOC) measurements with occupancy assessment. Frei et al. [11, 12] tested a sensor kit with eighteen sensors, considering energy consumption measurements, in 8 single-family houses in Switzerland.

Studies that focus on the implementation rather than on the design of development of sensor networks have larger instrumentation goals to survey energy performance [15, 16, 18] and on the impact of energy retrofit actions [13, 17]. For energy performance characterization, Jnat et al. [16] used only two measurements (indoor air temperature and relative humidity), Deb et al. [15] based their work on the solution from Frei et al., while Guyot et al. [14] used an existing building energy management system (BEMS)

CHAPTER 2 – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

Table 2.1: Summary reviewed studies implementing sensor networks for data collection in operating buildings.

Studies	Frei et al.[11, 12]	Martín-Garín et al. [9]	Karami et al. [10]	Jankovic [13]	Guyot et al. [14]	Deb et al. [15]	Jnat et al. [16]	Jacopo et al. [17]
Country	Switzerland	Spain	N.C.	United-Kingdom	France	Switzerland	France	Italy
Case study	8 single-family individual houses	1 apartment	1 computer lab	2 half-attached houses	1 mix-used building (mostly instrumented office spaces)	1 single house	3 social apartments	20 apartments
Sensor network	Arduino-based sensor kits	Arduino-based	Arduino-based	Wireless	BEMS system	c.f. Frei et al.[11]	Raspberry-based	LoRaWan-based
Measurements	Air temperature, relative humidity, volumetric oil flow, light pulses for electricity meters, CO ₂ concentration, luminosity, windows opening	Temperature, humidity, barometric pressure, CO ₂ concentration, window opening	Dry bulb temperature, elevation temperature, relative humidity, horizontal illuminance, vertical illuminance, CO ₂ concentration, VOCs, PM2.5, occupancy	Natural gas, electricity, indoor temperature, outdoor temperature, humidity, solar irradiation, rainfall, wind speed and direction	Measurements on air handlers, radiant panels, indoor temperature, humidity, lighting, occupancy, window opening, shading, energy systems setups	Air temperature, CO ₂ concentration, inlet heating temperature, outdoor temperature, heat flow for walls and windows, window opening	Indoor air temperature, relative humidity	Indoor air temperature, pressure, relative humidity, indoor illuminance, VOC
Acquisition time-step	5 minutes	1 minute	N.C.	Daily energy consumption, 15 minutes for other data	5 minutes	5 minutes	30 minutes	1 hour
Acquisition period	5 months	7 days	10 days	2 years and a half	N.C.	1 heating season	6 months	1 year and 1 month
Number of sensors	144 (18 per apartment)	N.C.	9	12	6,000	16	6	40 (2 per apartment)
Communication	GPRS, Zigbee	Wi-Fi	Zigbee	N.C.	N.C.	Zigbee	Radio frequencies	LoRa
Storage	Php, MySQL	Google Drive excel sheets	VOLTRON open-source software	N.C.	N.C.	MySQL	SD card	N.C.

2.1 Instrumentation solution for onsite building operation data collection: a review of industrial and academic projects

system in three mix-used buildings with over 6,000 measurements points. For the study of retrofit actions, Jankovic [13] relied on a measurement campaign over the retrofit of two semi-attached houses in the United-Kingdom, to calibrate building energy models and improve the assessment of the effect of energy retrofit actions. Jacopo et al. [17] highlighted changes in IEQ after a deep energy retrofit of twenty apartments in Italy.

2.1.3 Discussions

The review of instrumentation projects for the characterization and monitoring of building operation shows two distinct types of projects. Industrial projects deploy a large amounts of sensors on a large number of buildings. They target the development of operative methods and broadcast the acquired knowledge for professionals from construction and energy sectors. Because of the significant budget and human resources involved, monitoring and data analysis opportunities are superior to academic projects, which are focusing on specific smaller-scale case studies. Academic research objectives are to dig deeper into the specificity and details of a given case study and instrumentation method.

Indoor air temperature and relative humidity are the most common measurements, supposedly because of the availability of simple components for such measurements. CO₂ concentration and window opening detection are also popular. However, a significant absent is the monitoring of heating/cooling energy consumption, specifically in academic research. Some related measurements are used. Nevertheless, data are mostly collected from bills or large time-step manual readings. Furthermore, occupants' behavior characterization is scarcely investigated. Yang et al. have reported a diversity of occupancy monitoring techniques in their review [19]. However, only three studies focused on window-opening and only one was monitoring the occupancy. Hence, there is a significant challenge in gathering the many aspects of building energy efficiency – energy consumption, IEQ and occupants' behavior – in a single sensor network.

Between the design and implementation of sensor networks, regardless of the acquisition time-step, a significant difference is the monitoring period. Design and calibration of a set-up is made over a few days of monitoring. The study of building energy performance requires several months of good-quality data. However, the question of ensuring the calibration over time is not raised in reviewed studies focusing on sensor network designs. Calibration is not mentioned at all in application studies while this should be an important point to address, especially for long-term monitoring campaigns, when some sensors may exhibit measurement drifts over time.

It is also important to highlight that the number of deployed sensors is usually quite

small, which is evidently related to the fact that most solutions are developed by research teams with limited human resources and because of the cost of commercial solutions. This raises two questions regarding the replicability of the work, since technologies might not be inter-operable; but also on the conclusions of the results, because of the small number of measurements and case studies.

This review provided relevant perspectives on the instrumentation set-up. There are three possible options: in-lab development, commercial solutions using IoT and portable connected objects, and BEMS-like networks. A BEMS is quite likely to be installed in buildings upon their construction. Otherwise, it is a costly and permanent solution, which would not fit the objectives of our project. On the other hand, in-lab implementation is time-consuming and it restricts the types of measurements that can be performed. The inter-operability of the devices with other systems is also an issue. Hence, considering the needs and goals of the thesis, the most affordable solution would be a commercial IoT instrumentation solution, such as the one use by APC.

2.2 Case study: existing, occupied and retrofitted residential buildings of social housings

This section provides a short overview of the main building features. More details are given along with the description of the energy models, in Chapter 4.

2.2.1 General description

The case study is a group of three existing and occupied residential buildings located in Seine-et-Marne, in Paris greater area. Buildings were built in 1974. Together, they group sixty-three social housings, for a total living area of 3,825 m². Through the manuscript, buildings are referred as B1, B2 and B3 (Figure 2.1). They respectively have thirteen, twenty-one and twenty-nine housings, and a living area of 765 m², 1,275 m² and 1,785 m². A description of the characteristics of housings is given in Table 2.2.

The instrumentation campaign at building-scale is deployed on energy systems and in shared areas. Sensors are also deployed in an eight-apartment sample recruited to take part in the instrumentation campaign. The distribution of instrumented housings is presented in Table 2.3

2.2 Case study: existing, occupied and retrofitted residential buildings of social housings



Figure 2.1: Neighborhood plan and pictures of the facades of buildings before the retrofit.

Table 2.2: Summary of apartment features in the three considered buildings.

Housing type	B1	B2	B3	Total number of apartments	Area per apartment (m ²)
1 living room/bedroom (T1)	2	2	2	6	36
1 bedroom (T2)	4	5	7	15	50-53
2 bedrooms (T3)	4	10	14	28	63-68
3 bedrooms (T4)	3	4	6	13	74-79
TOTAL	13	21	29	63	/

Table 2.3: Description of the household sample that took part in the instrumentation campaign.

Building	Floor number	Orientation	Size	Surface	Number of occupants
B1	2	South-East	T3	63 m ²	2
B1	3	North-East	T2	50 m ²	1
B2	Ground floor	North-West	T3	64 m ²	1
B2	1	South-East	T2	53 m ²	1
B2	2	South-East	T2	53 m ²	1
B2	5	South-East	T2	50 m ²	1
B3	Ground floor	North-East	T4	74 m ²	2
B3	2	North-West	T3	70 m ²	1

2.2.2 Summary of building characteristics and retrofit actions

All three buildings in our case study have undergone a deep energy retrofit from July 2020 to July 2021. The landlord, *Marne-et-Chantereine Habitat*, also built at the same time a new building located nearby the existing buildings and where the current boiler is now located.

The building envelope was fully insulated, including outdoor walls, and walls and ceilings that separates heated areas from non-heated areas. The roof was not modified since its thermal features were already up to the current energy efficiency requirements. The ground floor, above the crawl space, was not insulated since it could not be properly accessed during the retrofit. All windows, glazed doors, outdoor doors and apartment entrance doors were changed. An entrance air lock was created at the main entrance of the buildings.

Heating and DHW production system was already central. Prior to retrofit actions, heating and domestic hot water were produced by a neighborhood furnace using natural gas. The furnace served the three buildings as well as dozens of other social collective housing buildings owned by another landlord. During the retrofit *Marne-et-Chantereine Habitat* opted for a more cost-effective solution. A furnace was built in their new building and connected the extended geothermal heating network of the city. Ventilation was replaced by a new humidity-sensitive simple-flow CMV (controlled mechanical ventilation)

in each building and new extraction units were installed in apartments.

Common areas of the building do not have any heaters nor ventilation. Appliances include lifts (only for B2 and B3) and occupancy-driven lighting with a timer. These were retrofitted up to required security standards. In housings, heaters were replaced with new models including a thermostatic valve, and the main hot water pipes were insulated, as well as part of the secondary piping circuit. Electric systems and switchboards were rehabilitated.

2.2.3 Description of housings

There are four types of apartments in the three studied buildings, with a main bedroom/living room or one, two and three bedrooms. Architectural features of housings are similar within and between buildings. There are four apartments per floor, except for the ground floor – three apartments – and the last floor – two apartments. Each housing is located in a corner of a floor and has two orientations: North-West, North-East, South-East or South-West.

Energy meters, including electricity and natural gas, are located outside of the apartment but on the same floor. Natural gas is only used for cooking, but most apartments have electrical cooking appliances and do not use natural gas at all. Heaters are located in each room, except in the bathroom and water closet – apartments of the last floor have a heater in the bathroom. CMV extraction units are located in the bathroom, the water closet and the kitchen. Water meters (tap water and hot water) are located in the water closet. They serve the bathroom nearby and the kitchen on the other side of the apartment.

The instrumented eight-housing sample shows an interesting diversity of sizes, locations in the buildings, orientations and residents' profiles, summarized in Table 2.3. A few socioeconomic information also highlight the variety of occupants' profiles in instrumented housings, and summarized in Table 2.4

2.3 Wireless sensor network

2.3.1 Targets and deployment

The wireless sensor network focuses on different energy end-uses and parameters, both at building and household levels. Measurements are divided into four categories: i) the energy consumption – with thermal energy, electricity, and natural gas, ii) the IEQ, iii)

CHAPTER 2 – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

Table 2.4: Description of occupants in the instrumented households.

Building	Floor	Residents	Occupancy	Children	Pets
B1	2	2	Public servant	1	None
B1	3	1	Public servant	None	None
B2	Ground floor	1	Private sector employee	None	6
B2	1	1	Public servant	None	None
B2	2	1	Teacher	None	2
B2	5	1	Unemployed	None	2
B3	Ground floor	2	Retired	None	None
B3	2	1	Retired	None	1

the occupants’ behaviors and iv) the local weather.

Energy consumption characterization is the first objective. As described in Section 2.2, energy sources include electricity, thermal energy for heating and domestic hot water, and natural gas. IEQ focuses on parameters that depict the occupants’ comfort [20] and the thermal characteristics of the building. Both are related to building energy consumption [21, 22]. Occupants’ behavior is also a significant energy driver [23]. Finally, the local weather impacts on building energy needs, and specifically thermal energy consumption.

Because of the diversity and the number of sensors, the deployment of the instrumentation solution is divided into two steps and four batches. The final schedule is the following:

- Step 1: instrumentation in common areas and building-scale measurements, including:
 - Batch 1: the IEQ and the occupants’ behavior in common areas, the building-scale electricity and thermal energy consumption monitoring;
 - Batch 2: the weather station;
- Step 2: instrumentation in households, including:
 - Batch 1: the IEQ and occupants’ behavior;
 - Batch 2: the energy consumption (thermal, electricity and natural gas).

The complete installation took over two years and a half (Figure 2.2). Sensors and

2.3 Wireless sensor network

communication protocols from the first step, were also used as a test phase to highlight practical issues and conclude on the use of sensors for specific data acquisition. In the following sections, preparation of the sensor network is briefly summarized? and the sensor network is detailed for each type of measurement. Common areas and housings are discussed separately. Data communication and storage are described in a separate section.

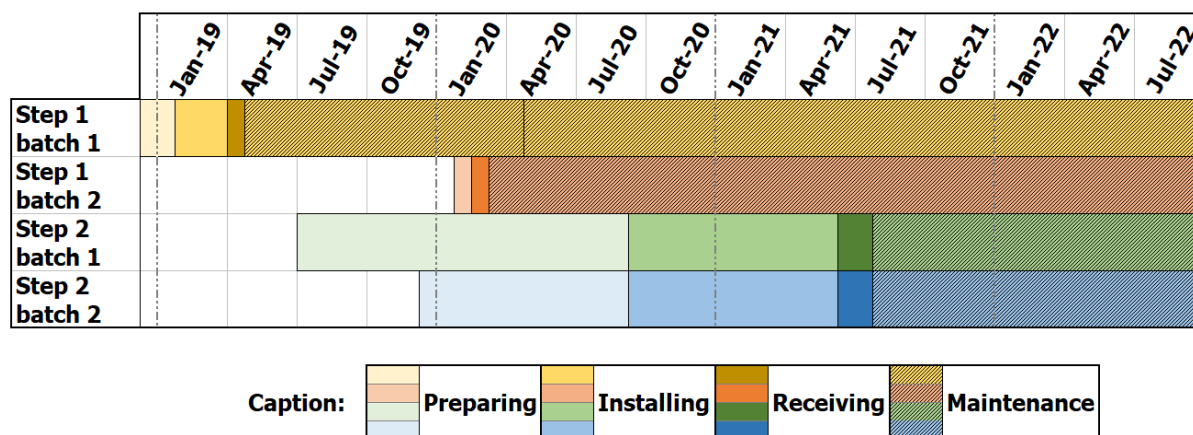


Figure 2.2: Deployment schedule of the instrumentation solution.

2.3.2 Preparing the measurement campaign

Before the selection and installation of the sensor network, there were a few non-technical but necessary steps to consider.

A design brief was created to specify the needs and the purpose of the instrumentation. It is a reference document for all actors of the sensor network deployment, either from the research team or contractors providing equipment and sensors. It evolved through discussions with the contractors, to fit to the reality of the IoT market, regarding existing instrumentation solutions and costs.

Once the design brief was ready, three actions were led simultaneously. First, project volunteers were recruited. It was one of the major challenges of the research project since participation of residents was voluntary. The entire data collection and research project depended on a sufficient number of instrumented households. In exchange to their participation, collected data are returned to residents with a dedicated overview over their apartment and energy consumption. During the recruitment campaign, the sixty-three households were considered as potential participants with an initial goal of eighteen apartments for the instrumentation campaign:

CHAPTER 2 – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

- Only nine housings agreed to participate, later reduced to eight housings since one family moved out of their apartment during the measurement campaign and was not replaced;
- Four households agreed at first and later declined their participation;
- Eighteen households strictly refused to be part of the experiment;
- Thirty-one households never answered any of our contact attempts.

Recruitment has highlighted that most residents did not pay much attention and efforts to energy efficiency, either because they were not aware on the topic or because they were mostly focused on the retrofit of their apartment in terms of comfort. A few households were also concerned with the use and security of their data and electro-sensitivity issues.

Jointly with the recruitment of housings, the call for commercial proposals was led, to select one or several contractors to provide sensor network solutions. The strategy was to keep the number of involved contractors to a minimum, to minimize potential project management issues. After a large survey, thirteen contractors provided a commercial offer. Four contractors were disqualified straightaway due to budget constraints. Seven contractors went to a second round to refine their offer based on the detailed needs of the project. Two contractors were finally selected. The first selected contractor provided the supply, installation? and partial management of the sensor network for step 1 batch 1 and step 2 batch 1 of the instrumentation solution. The weather station (step 1 batch 2) and step 2 batch 3 were provided by the second contractor. For these steps, part of the setup, the installation and the supervision were performed by our research team.

The final step to prepare the instrumentation campaign focused on the communication about the project with CNIL organism (*Commission Nationale Informatique et Libertés*). It was required to provide all details of the project related to data collection, analyses and future use, to protect the residents' privacy. CNIL advised on potential warnings and provided recommendations. For similar projects, if a declaration on data usage is mandatory, the implementation of recommendations is not. For the present project, most comments were about:

- The relationship between the landlord and other project actors;
- The communication means and documents regarding the research project and for the residents;
- Collected data usage: potential use outside of the European Union (resulting in

modified regulations), storage specifications, data transmission means, residents' personal data transferred to the landlord. Both the researchers and the landlord are responsible for collected data use, since the data are related to residents living in the landlord's buildings, and even though data are processed by the research team only;

- Residents' agreement: it had to be a written agreement.

2.3.3 Data communication and storage

The sensor network is entirely wireless and it relies on two different communication protocols. The first protocol is GPRS (general packet radio service), a standard data package communication protocol. The second protocol is LoRaWan (long range wide area network). It is a radio communication protocol commonly used for smart cities applications [24] and relying on LoRa peer-to-peer technology to connect communicating objects. It offers advantages over other competing communication protocols with a long range communication and a very low energy consumption (more details can be found in Appendix B.1). Within the sensor network, LoRaWan is divided between an operated network and two private networks.

Operated LoRaWan networks are managed by national telecommunication companies. This option has a major benefit with a simple and straightforward implementation. Communicating devices are declared online and a fee is applied for each communicated data point. Data pre-processing such as identifying missing data, treating duplicates and formatting is processed on the operator's servers. However, operated LoRaWan networks are constrained regarding bandwidth usage, which limits potential applications (such as small acquisition time-steps or large numbers of sensors). For the implementation of an operated LoRaWan network, gateways are used as data communication relays to the operators' servers.

Private LoRaWan networks rely on the same technology but the network is a local network. LoRaWan communication is implemented between sensors and a dedicated gateway. The gateway decodes radio packages and transfers the data to a private server using the Internet. Hence, there is no constraint regarding data acquisition. However, each sensor is connected to a specific gateway with a higher risk of data loss in case of a malfunction, and the initial investment cost is more significant than with operated networks – specifically because of the cost of the gateways.

The main part of the sensor network relies on LoRaWan protocol as illustrated in

Figure 2.3. The LoRaWan operated network depends on two gateways for sensors installed in common areas, along with pulse sensors for gas meters and connected plugs in housings. Two different private LoRaWan networks are implemented. The first (private LoRaWan network #1 in Figure 2.3) is dedicated to sensors for IEQ monitoring (temperature, humidity, luminosity and CO₂) and presence detection in apartments. The second private network (private LoRaWan network #2) connects all remaining sensors located in apartments. It depends on two gateways due to the number of connected sensors. GPRS data communication is used for electrical measurements on electric smart meters and switchboards. All gateways are installed in technical rooms on the ground floor of the buildings.

Prior to data storage, part of the collected data are pre-processed through network servers. The GPRS and operated LoRaWan network are processed through Objenious [25], the national LoRa company from Bouygues Telecom. IEQ data are managed through the platform of The Things Network [26]. Data from private LoRaWan network #1 are processed online by our research team.

Collected data are initially stored on separate servers due to sensor technologies and deployment management during the research project. The largest part of the data is stored on FTP (file transfer protocol) servers except for the LoRaWan private network supporting window opening detection sensors. Window-related data collection is event-driven. Therefore, a HTTP (hypertext transfer protocol) server is required. Data processing is performed to store all collected data on a single FTP server (FTP server #3 in Figure 2.3), in CSV (comma-separated values) files.

2.3.4 Measurements

Measurements are divided into four categories: the local weather, the energy consumption, the IEQ and the occupants' behavior. The following sections detail the instrumentation for each type of measurements. The characteristics of the sensors with the accuracy and operating range are summarized in Table 2.6.

2.3.4.1 Local weather

2.3.4.1.1 Context

Good quality weather data – meaning complete, exhaustive data for a specific location – are a must to complete energy analyses.

Weather data can be collected through different means. The most common source of

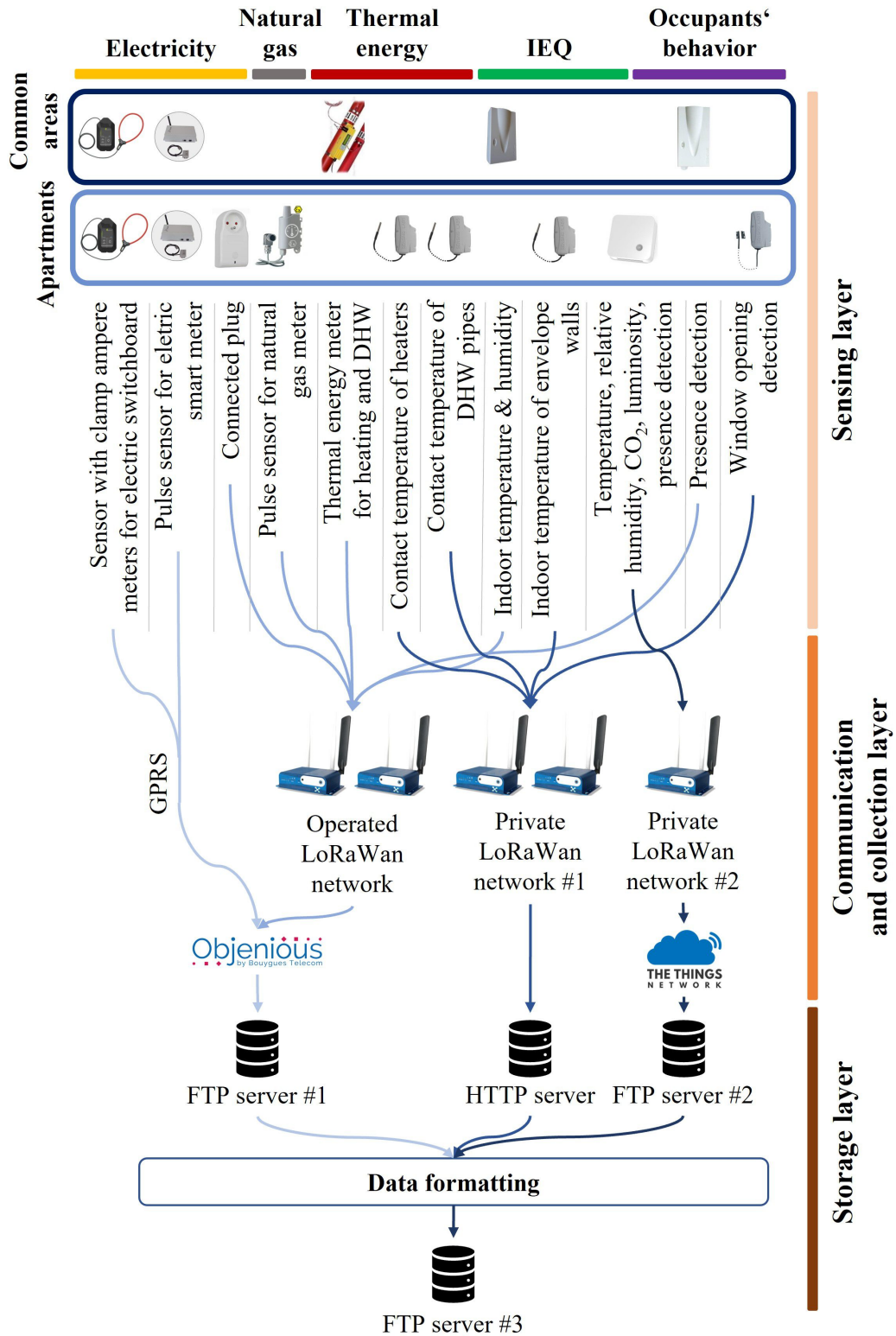


Figure 2.3: Description of the sensor network from sensors to data storage with three layers: the sensing layer includes all meters and sensors, the communication and collection layer relates to gateways and data processing platforms from Objenious [25] and The Things Network [26], and the storage layer groups all FTP and HTTP storage servers.

CHAPTER 2 – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

Table 2.5: Details on the numbers and types of sensors in the different instrumented households (with one apartment per floor).

Building	B1		B2				B3		TOTAL
Sensors in housings									
Housing floor	2F	3F	GF	1F	2F	5F	GF	2F	
Sensor with clamp am- pere meters	0	1	1	0	0	0	1	0	3
Pulse sensor for electric smart meter	1	1	1	1	1	1	1	1	8
Connected plug	4	6	4	3	2	6	6	5	36
Pulse sensors for gas meter	0	1	1	0	1	0	1	0	4
Contact temperature of heaters	4	4	4	3	3	3	4	4	29
Contact temperature of DHW pipes	1	1	1	1	1	1	1	1	8
Indoor temperature of cold walls	2	3	2	0	2	3	3	2	17
Temperature, humidity, luminosity, CO ₂ , pres- ence	1	1	1	1	1	1	1	1	8
Window opening detec- tion	4	4	4	3	3	4	5	4	31
Total housings	17	22	19	12	14	19	23	18	144
Sensors in common areas									
Sensor with clamp am- pere meters	0		0				1		1
Pulse sensor for electric smart meter	2		3				2		7
Thermal energy meter for DHW and heating	2		2				2		5
Indoor temperature and humidity	3		3				3		9
Presence detection	1		1				1		3
Total common areas	6		8				8		26
TOTAL SENSOR NETWORK	45		72				49		170

2.3 Wireless sensor network

Table 2.6: Summary of the accuracy and operating range of the deployed sensors in apartments and common areas.

Measurement targets	Sensors	Accuracy	Operating range
Electricity	Sensor with clamp ampere meters (Ewattch Tyness)	N.C.	N.C.
	Pulse sensor for electric smart meter (Fludia BelSenso FM410e)	N.A.	10 pulse/sec max
	Connected plug (NKEWatteco Smartplug)	>1% (P > 40 W) <1% (P < 40 W)	Voltage: 100...250 V Frequency: 50...60 Hz
Natural gas	Pulse sensor for gas meter (Adeunis Pulse ATEX)	N.A.	8 pulse/sec max
Thermal energy	Thermal energy meter for DHW and heating (Ultraflow U1000)	$\pm 1...3\%$ (flow > 0.3 m/s)	Flow: 0.1...10 m/s Temperature: 0...+85°C
	Contact temperature of heaters (SensingLab TEM-LAB-14NS)	$\pm 0.5^\circ\text{C}$ (-10...+85°C) $\pm 2^\circ\text{C}$ otherwise	Temperature: -45...+125°C
	Contact temperature of DHW pipes (SensingLab TEM-LAB-14NS)	$\pm 0.5^\circ\text{C}$ (-10...+85°C) $\pm 2^\circ\text{C}$ otherwise	Temperature: -45...+125°C
IEQ	Indoor temperature and humidity (SensingLab THY-LAB-41NS)	Temperature: $\pm 0.3^\circ\text{C}$ Humidity: $\pm 2\%$	Temperature: 0...+55°C Humidity 0...80%
	Indoor temperature of cold walls (SensingLab TEM-LAB-14NS)	$\pm 0.5^\circ\text{C}$ (-10...+85°C) $\pm 2^\circ\text{C}$ otherwise	Temperature: -45...+125°C
	Temperature Humidity CO ₂ Luminosity (ELSYS ERSCO2)	$\pm 0.2^\circ\text{C}$ (0...+60°C) $\pm 0.2\%$ (10...90%, T=25°C) ± 50 ppm ± 10 Lux	-40...+120°C 0...100% 0...2,000 ppm 0...65,535 Lux
Occupants' behavior	Presence detection (ELSYS ERSCO2)	N.A.	0...255 motions
	Presence detection (common areas) (SensingLab PIR-LAB-41NS)	N.A.	N.C.
	Window opening detection (SensingLab OPE-LAB-41NS)	N.C.	N.C.

weather data is open-access or private weather station networks. The access is simple but it has several major disadvantages. The weather station location can be different from the location of the case study, and it induces significant micro-climate effects – for example using data from an airport for a urbanized location [27]. Data quality and measurements are also an issue. Data may be pre-processed to remove outliers and complete missing datapoints, without details on data pre-processing steps. Data collection time-step cannot necessarily be customized, and all expected measurements may not be available either. Moreover, weather data acquisition requires a significant budget with respect to the number of measurements, temporal granularity and timeframe of the data.

A second alternative is to use energy data from energy modeling software. Software embed weather files with various measurements used to simulate the building energy consumption. Many locations can be found across the world and data are already processed in adapted file formats for the modeling software [28]. However, the location of the collected data still is an issue. Another major disadvantage is that these types of data are statistical data processed from a much earlier time than the time of the research study. Hence, it is impossible to relate statistical weather data to collected field data.

Considering the pros and cons of the two options above, a dedicated weather station was acquired for the research project. With a reasonable budget, it provides a customized data collection with little day-to-day supervision required and an autonomous power supply.

2.3.4.1.2 Technical specifications

The weather station includes eight measurements through three groups of sensors. The main group gathers outdoor air temperature, relative humidity, rainfall, wind speed, wind direction and atmospheric pressure sensors. Two sensors are added: a black ball with a PT100 temperature probe inside to measure the radiant temperature, and a pyranometer to measure the solar radiation. Specifications of the sensors are summarized in Table 2.7.

2.3.4.1.3 Set up

The weather station is set up on the roof of a university building, three kilometers away from the instrumentation site. Installing the weather station directly onsite would have been the optimal solution. However, the local building configuration could not allow a simple and secure access on a roof for occasional maintenance. The weather station data collection and transmission are entirely managed by a data logger using GPRS data communication. The data acquisition and communication time-steps are set to

Table 2.7: Summary of accuracy and operating range of the weather station.

Sensors	Accuracy	Operating range
Temperature	$\pm 0.3^{\circ}\text{C}$ ($0^{\circ}\text{C}\dots+70^{\circ}\text{C}$) 0.4°C otherwise	$-40\dots+105^{\circ}\text{C}$
Humidity	$\pm 1.8\%$ ($0\dots 85\%$, $T=+15\dots+35^{\circ}\text{C}$) $\pm 2.5\%$ ($85\dots 100\%$, $T=+15\dots+35^{\circ}\text{C}$) $\pm 2+1.5\%$ otherwise	$0\dots 100\%$
Rainfall	N.C.	N.C.
Solar irradiation	$\pm 10\text{ W/m}^2$	$0\dots 2,000\text{ W/m}^2$
Wind speed	$\pm 2\%$ ($0\dots 65\text{ m/s}$), $\pm 3\%$ otherwise	$0\dots 80\text{ m/s}$
Wind direction	$\pm 2^{\circ}$	$0\dots 359.9^{\circ}$
Atmospheric pressure	$\pm 0.5\text{hPa}$ ($800\dots 1,100$ hPa , $T=25^{\circ}\text{C}$) $\pm 1\text{hPa}$ ($300\dots 1,100$ hPa , $T=0\dots 50^{\circ}\text{C}$)	$300\dots 1,100\text{ hPa}$
Radiant temperature	$\pm 0.1^{\circ}\text{C}$	$-200\dots +650^{\circ}\text{C}$
Dew point temperature	N.C.	N.C.

five minutes. Data are stored on a dedicated FTP server and on the cloud storage of the manufacturer. The later solution provides an online visualisation platform of the collected data. The weather station has an autonomous power supply, with a photovoltaic solar panel and a battery.

2.3.4.2 Energy consumption: electricity

2.3.4.2.1 Common areas

Electric appliances at building-scale include elevators, lighting, CMV and hot water pumps. Two different types of sensor are implemented. Linky smart electricity meters are already installed onsite. Although Linky data can be collected through a specific process with the operating company Enedis [29], data granularity is thirty-minute at the lowest. A smaller acquisition time-step is expected to capture small triggering events. Hence, Linky meters are instrumented using a pulse sensor [30] (Figure 2.4(a)) to count the number of light pulses: one pulse equals one unit of electricity consumption (in Wh). Data are acquired and transferred at one-minute time-step using GPRS communication

protocol.

Seven sensors are installed in technical rooms. In all three buildings, there is one sensor for the elevator and the CMV (grouped together on the same meter), one sensor for lighting, elevator lighting and the magnetic entrance door. The seventh meter is used for the hot water pumps of the heating substation – and later removed when the heating substation is modified following energy retrofit actions. Pulse sensors are powered with batteries.

The building main electrical switchboard of B3 is instrumented with a sensor connected to six clamp ampere meters [31] (Figure 2.4(b)). It measures the current of indoor shared areas lighting and calculates the corresponding energy consumption – constant 240V voltage, single-phased current ($C = 0.9$). Data are collected and transferred with a ten-minute time-step using GPRS. The ten-minute time-step is set up instead of one-minute time-step which was initially planned. After a few months, the one-minute time-step data acquisition resulted in sensor malfunctions. A dedicated power supply and circuit breaker is required for this sensor.

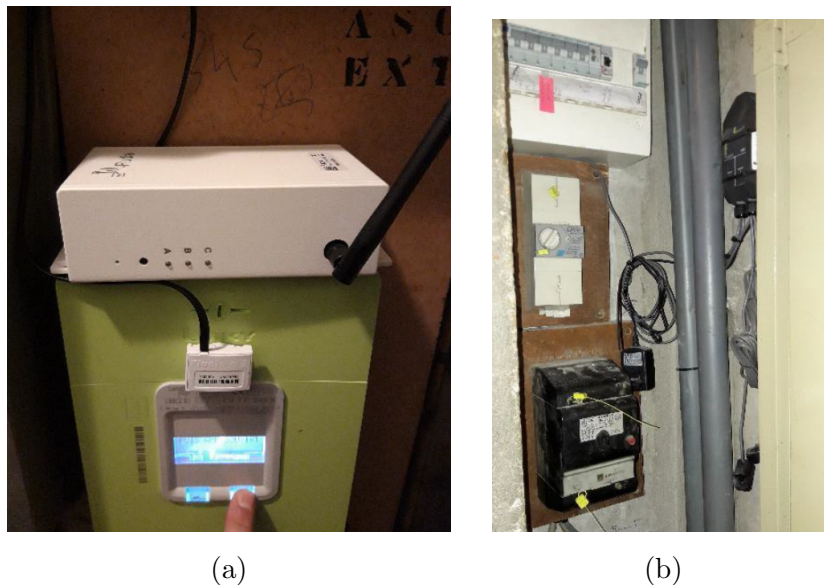


Figure 2.4: Pulse sensor installed on a Linky meter (a) and sensor with clamp ammeters for electricity sub-metering on a electrical switchboard (b).

2.3.4.2.2 Housings

In housings, electricity monitoring targets two spatial scales for data collection. Pulse-reading sensors for Linky smart meters are installed in each instrumented apartment. They monitor the overall electricity consumption of the apartment. Sensors with clamp

ammeters for electrical switchboard are used as well. Only three of these sensors are installed, in B1/3, in B2/0 and in B3/0, because of the intrusive data collection process, the difficult installation and the cost of the sensor and their installation. The supervision of the electricity switchboard focuses on room-aggregated electricity consumption, lighting, and the large appliances whose electricity consumption cannot be accessed by other means. The electricity monitoring system is complemented with connected plugs [32] that measure the energy consumption on household appliances. They are used to break the energy consumption of the apartments down to appliance-level. The number of smart plugs depends on the configuration of the households and the identified appliances.

Data collection is performed using the operated LoRaWan network with one-minute acquisition time-step and ten-minute data communication time-step. Hence, electricity consumption can be analyzed at three different scales: apartment, room, and appliance.

2.3.4.3 Energy consumption: thermal energy

2.3.4.3.1 Common areas

Thermal energy consumption refers to hot water production for heating and domestic use. It is monitored using specifically designed thermal energy meters. Thermal energy meters have three distinctive elements (Figure 2.5): i) two temperature probes: one on the inlet and one on the outlet water pipe, ii) a flow meter installed on the outlet water pipe, iii) a computer to calculate the corresponding thermal energy consumption out of temperature and flow measurements with respect to the of the piping system features.

Energy meters can be integrated to the water circuit with specific interventions (circuit cleansing). However, this installation is costly and sensors cannot be removed afterwards. Considering the needs of our sensor network, this option is not considered. Implemented thermal meters use ultrasonic water flow probes that do not require an integration into the water circuit. They are simply installed on the water pipes. Ultrasonic meters measure the time needed by an ultrasonic impulsion to move from the sending probe to the receiving probe through the water pipes. Knowing the traveling time and depending on the inner diameter, outer diameter and material of the pipe, the water flow is assessed. Temperature probes are located on the outside of the pipe and are insulated from the surrounding environment.

Ultrasonic thermal energy meters are adapted to the objectives of our sensor network, for short-term measurement campaigns and flexible applications. Nevertheless, they require a lot of attention regarding their installation and calibration. Calculations can be

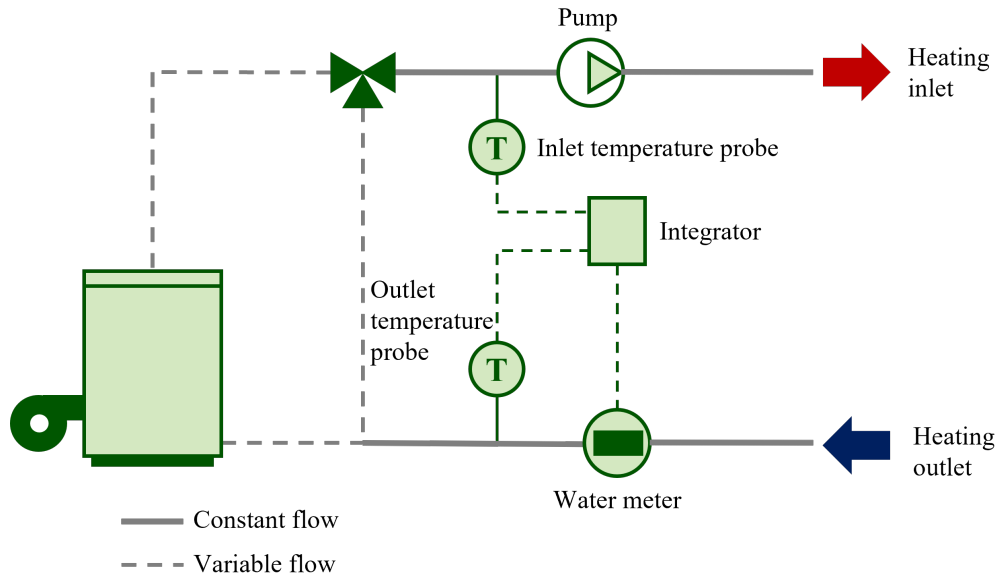


Figure 2.5: Schematic of a thermal energy meter for heating or ‘DHW energy metering – Adapted from [33].

significantly influenced by a poor set up of the ultrasonic probes or wrong piping characteristics (diameters, materials, or if the circuit is clogged). Furthermore, the position of the probes is important to accurately measure the water flow. For example, elbow pipes drastically impact on the water flow [33].

Thermal energy meters available on the IoT market are not communicating meters, and it is necessary to add a pulse sensor [34]. As for electricity meters, this sensor counts the number of pulses from the energy meter (one pulse equals one kWh of thermal energy consumption). Thermal energy meters require a power supply with circuit breaker and the pulse sensors are powered with batteries.

Thermal energy consumption data are collected for rather large time-range analyses (daily, monthly or annual). Hence, an hourly acquisition and transmission time-step is set up for all thermal energy meters using the operated LoRaWan network.

2.3.4.3.2 Housings

Thermal energy in apartments is characterized considering heaters and DHW pipes, but it is more difficult than for building-scale data acquisition. Ideally, measurements for both end-uses should be performed on the main hot water pipes with thermal energy meters. However, these pipes can hardly be accessed, and such an instrumentation set up would be too expensive to be considered.

2.3 Wireless sensor network

Contact temperature sensors [35] are installed on all heaters of each instrumented apartment (Figure 2.6(a)). Measurements are performed on a private LoRaWan network with thirty-minute data acquisition and communication time-steps.

For DHW, the same sensor ((Figure 2.6(b)) is used with one-minute acquisition time-step and twenty-minute communication time-step. The small data acquisition time-step aims to precisely capture DHW consumption through the variations of DHW temperature. The sensor is installed on the outlet of hot water meters in the apartments.



Figure 2.6: Installation of temperature sensor for DHW (a) and a heater (b) in apartments.

2.3.4.4 Energy consumption: natural gas

Natural gas is only used in a few apartments for cooking. Natural gas meters are equipped with Gazpar modules [36]. As for the Linky smart electric meter, the Gazpar meter is a smart meter implemented by the national administrator GRDF (*Gaz Réseau Distribution France*) and collecting consumption data. A pulse sensor is used on Gazpar meters (Figure 2.7). It reads every pulse corresponding to a 10 dm^3 gas consumption. Unlike other pulse sensors, it is specifically adapted to be used in explosive atmospheres [37]. A total of four sensors is implemented with a one-hour data acquisition and communication time-step on the operated LoRaWan network.



Figure 2.7: Pulse sensor for natural gas Gazpar meter.

2.3.4.5 Indoor environment quality (IEQ)

2.3.4.5.1 Common areas

In common areas of our building case study, IEQ is characterized through indoor air temperature (IAT) and relative humidity measurements. One sensor is used to provide both measurements [38]. For each building, three sensors are installed: one on the ground floor, one on the intermediate floor and one on the last floor. Preferably, these sensors would be installed at a height of 1.50 meters to give an averaged picture of the temperature and humidity level in the rooms, considering temperature gradient phenomenon. However, since sensors are installed in corridors, they need to be out of sight and reach. Sensors are located under the ceiling, on the structure of the stairs (Figure 2.8).

2.3 Wireless sensor network

Data acquisition time-step is set to one hour. No sudden variations of temperature and humidity are expected, apart from doors or windows that could be opened near the sensors. Data are communicated with hourly time-step using the LoRaWan operated network.



Figure 2.8: Temperature and humidity sensors installed in the buildings shared areas.

2.3.4.5.2 Housings

IEQ in housings is characterized with two types of sensors. A first sensor performs four IEQ measurements: IAT, relative humidity, brightness and CO₂ concentration [39]. These parameters are also combined a fifth measurement in the sensor, the occupants' presence detection, described in the following section. The sensors are located in the living room of housings. Measurements are performed with half-hour time-step on a private LoRaWan network (private LoRaWan network #2 in Figure 2.3).

A second sensor measures the contact temperature of the inner surface of cold walls. The sensor is the same as for heaters and DHW. These data aim to assess the radiant temperature of walls, which is a relevant parameter regarding indoor comfort: even though the air temperature might be high enough, if the surrounding walls are too cold it affects the feeling of comfort [40]. These data can also be combined with IAT measurements to calculate the operative temperature – the average between IAT and the indoor temperature of cold walls. These sensors are set up with hourly data acquisition and communication time-steps on the operated LoRaWan network.

2.3.4.6 Occupants' behavior

2.3.4.6.1 Common areas

Occupants' behavior characterisation in buildings common areas is performed with presence detection sensors above the main entrance door of the buildings (Figure 2.9). Sensors are infrared presence detectors [41]. This technology is commonly found in automated lighting systems linked with timers [19]. PIR sensors do not discriminate between people entering or leaving the building and they cannot count the number of detected people. Presence detection is aggregated and communicated at hourly time-step using the operated LoRaWan network.



Figure 2.9: Infrared presence sensor in buildings shared areas.

2.3.4.6.2 Housings

Occupants' behavior in households is characterized through two measurements: presence detection in the apartments (from the same multi-measurement sensor as for IEQ monitoring [39]), and window and glazed door opening detection. Presence is assessed using an infrared sensor as in common portions of the buildings. Data are aggregated at half-hour time-step. It counts the number of passing in front of the sensor.

Window opening detection is monitored with contact sensors [42]. A magnet is connected to the sensor and another magnet is installed on the window. When the two magnets are in contact, then the window is closed, otherwise it is opened. Data acquisition and communication are event-driven: data are acquired and transferred when the window opening status changes. Then, if there is no status change for over an hour, a data point is sent every hour to recall the latest opening/closing status of the window.

2.4 Discussions

2.4.1 Feedback on the implementation and operation of the sensor network

The deployment of our sensor network provides a valuable feedback on the main achievements and difficulties in the implementation of an instrumentation solution. Critical points to take into account can be summarized in three categories (ranked by decreasing impact on the project) and are further described in [43]:

- Installation conditions and environment. The user case study is a group of three existing buildings from 1974. The integration of a sensor network is more difficult than with a newly built or more recent building. In that case, the building design would be more adapted to support the monitoring solution, even with an on-going retrofit. Also, since it is the purpose of the research project, buildings are occupied. This means that the sensor network must be as little intrusive as possible. However, necessary maintenance can still disturb inhabitants and data collection may be disturbed back by inhabitants (moving or switching off sensors and gateways, for instance). Finally, retrofit actions are conducted during part of the instrumentation process, which resulted in several issues.
- Targets of the sensor network. Because the sensor network targets a large range of measurements, it is necessary to mix many different technologies of sensors, data acquisition, communication and storage. These technologies are not always fully compatible and result in additional challenges for the long-term project maintenance. Also, the sensor network relies on IoT objects. We observed that the current IoT market is more fitted for large-scale deployment strategies over a year or so, rather than for high-precision measurements and long-term monitoring.
- Project management. From the design brief to the installation and long-term maintenance, the sensor network is a whole project than needs an optimized management, specifically regarding external parties, such as volunteering housings and contractors. Participants are essential to the project, since the results entirely depend on a sufficient amount and diversity of collected field data. Contractors also play a significant role. They usually provide a "plug-and-play" management: sensors are provided, set up, installed and maintained. However, most of the IoT contractors are energy managers that delegate installation and maintenance tasks to other contractors and have limited field knowledge. Hence, it is very time-consuming to solve any technical issue. On the other hand, a direct management by our research team

eased and fasten up technical issues processing, but it required extensive technical skills and knowledge.

In addition to the published results [43], the installation of sensors in households has highlighted a few more insights on the challenges of building energy monitoring, specifically on the limitations of radio frequencies. LoRaWan is the main communication protocol of our sensor network. It uses free radio frequencies and is supposed to benefit from a "long range wide area" network. However, construction works have shown that the LoRaWan network is easily disturbed, especially with scaffolding all around the buildings that disrupted radio communication. The position of the LoRa gateways has a significant impact as well on data communication, even with sensors and gateways within the same building. Several options could be considered to try and solve this issue but all show much uncertainty regarding their efficiency. LoRa antennas located on a building roof is a simple and efficient strategy. However, it is difficult to link the antenna to the gateways, knowing that the sensors must be above or at the same level as the gateways. It could be possible to add more gateways but there is a question of cost and a lack of adapted locations onsite to secure the equipment. Switching to an operated LoRaWan network would also be a solution, but without guarantee that it could solve the problem. Moreover, some sensors would remain on the private LoRaWan network because of acquisition time-steps and bandwidth usage, and the cost of data collection fees would be significant with an operated network.

2.4.2 Limitations and potential improvements

The objective of our sensor network was to provide an exhaustive characterization of energy consumption, IEQ and inhabitants energy-driving behaviors in an occupied and retrofit group of three existing buildings. We achieved a significant challenge with the monitoring of twenty different parameters in housings, common areas and energy systems of the three studied buildings with 170 sensors collecting data for over three years in common areas and for one year in half in apartments. The implementation of the sensor network resulted in a significant amount of collected data that provide extensive knowledge on the operation of the studied buildings. Nevertheless, the sensor network still highlights several gaps that leave room for improvements.

2.4.2.1 Measurements

The characterisation of the occupants' behavior is performed with presence detection and windows/glazed doors opening detection. Presence detection is achieved using a single

sensor. One single presence detection sensor for each household limits the data acquisition strategy. The location of the sensor, in the living room, also constrains data analyses – even though volunteers declared spending most of their time in this specific room. Sensors perform presence detection but they cannot count the number of occupants in the room. To cope with these constraints, in terms of instrumentation, there would be complementary solutions, such as using additional sensors for presence detection in different rooms and a monitoring of the entrance door opening. However, additional sensors would bring up cost challenges and question inhabitant’s privacy protection. Housing occupancy can be deduced from other measurements, including CO₂ concentration, electricity consumption of appliances and window opening. This option can be considered but relates to data analyses rather than instrumentation solution.

Comfort and indoor environment characterization may be improved as well. In the present study, it includes indoor air temperature, relative humidity, CO₂ concentration, luminosity and indoor surface temperature of cold walls. Several aspects could be further explored with our sensor network, such as indoor air quality [44] (VOCs, air pollutants, PM_{2.5}), the characterization of thermal comfort [45], or visual and noise disturbances [46] that can impact on occupants’ behaviors and energy consumption.

Thermal energy consumption monitoring at apartment scale was significantly constrained in our project. Because the main hot water pipes were not accessible, DHW and heating energy consumption was assessed with contact temperature measurements on DHW pipes and heaters. These measurements are useful to understand DHW and heating patterns. However, to translate to energy consumption, it requires other information – hot water flow, heater and pipe characteristics – that cannot, or only partially, be collected in apartments. Therefore, several assumptions are necessary to assess the energy consumption. Another gap lies in the characterization of energy systems and building envelope. Energy systems are considered with their energy consumption and through information extracted from documentation. Building envelope is known thanks to the retrofit project management portfolio. Depending on the energy system, there are many existing characterization techniques [47]. For building envelope characterization, several industrial and experimental methods have been developed as well [48]. Nevertheless, most methods were too specific and costly for our research project that aimed a broader picture of building energy monitoring.

Finally, spatial and temporal data granularity are essential aspects that could be improved. More sensors in more apartments with a smaller time granularity would offer more opportunities. Nevertheless, choices were made to the best of the knowledge regarding

the needs for data and energy analyses, the specificities of the experimentation site and also keeping in mind that the larger and the more detailed the sensor network, the larger the challenge for day-to-day management, data processing and long-term maintenance. With budget constraints and difficulties for the enrollment of participating households, it was also decided to restrict the household sample and increase the number of sensors for each housing, rather than limiting the number of sensors and recruiting more volunteers. However, a limited eight-apartment sample can question the reliability of data analyses and replicability of the research work.

2.4.2.2 Technological choices

The choice of data to collect is critical to perform the expected energy analyses. Measurements are also impacted by available monitoring technologies and communication protocols. Technological choices are performed based on the IoT market, the expertise of contractors, the needs of our research project and the overall cost of the instrumentation solution.

The instrumentation solution is efficiently monitoring the targeted measurements. However, sensor technologies can be discussed. For instance, pulse technology is a simple, affordable and efficient mean to collect energy consumption data. Meters with pulse output produce one pulse for each unit of consumed energy. The aggregated number of pulses for a given time-step is the information communicated by a pulse sensor connected to an energy meter. However, this technology does not provide the precision of meter readings and rounds the energy consumption for each time-step. Hence, electricity and thermal energy consumption monitoring can be improved. For the former, TIC technology (*Télé-Information Client*, standing for remote client information) is an option. Instead of counting the number of pulses, it collects the same data as those sent to the electricity network administrator (energy index, amp demand, power demand, active power). However, this technology was not available with the expected specifications through the contractors we selected and it would have been too complicated for our research team to install and supervise these sensors on its own.

For thermal energy monitoring, because of the configuration of our building case study, ultrasonic thermal energy meters were selected. These meters can theoretically be installed and removed easily on any piping system. In practice, ultrasonic meters show strong limitations. Installation and calibration is tricky. Indeed, ultrasonic probes must be carefully set up to make sure that the ultrasonic signal and flow calculation are reliable. It needs a steady and reliable flow with a clean piping system, and meters must

be installed on pipe sections at least two-meter long, without bending, size differences or pumping systems nearby. Probes are very sensitive to dust, humidity, vibrations and any handling that can significantly disturb the ultrasonic signal. Consequently, maintenance is advised every six months to ensure the measurement quality. Finally, energy meters also embed a calculator to assess the energy consumption based on flow measurements, inlet-outlet temperature difference and characteristics of the piping system. The latter are not always fully known. Therefore, integrated meters would have offered a more reliable measurement but could not be installed. The other option would have been a Modbus [49] communication protocol to provide flow and temperature difference data along the energy consumption.

Other sensor technologies can be targeted by potential improvement, including presence detection and window opening sensors. There is a variety of occupant counting technologies already used in industrial and commercial buildings [19]. These could be investigated with a focus on the preservation of privacy protection and a limited intrusivity of the monitoring solution. Window opening detection sensors gave relevant insights on occupants' behavior. Nevertheless, a main issue is the impossibility to accurately assess the amount of missing data because of the event-based data collection process. Furthermore, window opening detection translates to window opening and closing duration. There is no information on the opening width of the window. This information impacts on the air flow between the inside and outside, that affects the IEQ, the IAT and the heating energy demand.

Finally, communication protocols and equipment are an important part of the sensor network. LoRaWan and GPRS both show pros and cons. The latter is reliable regarding punctual data loss and network dependency. However, it may have troubles accessing the network depending on the location of the sensors in the building. Moreover, because it transfers data upon collection, the life expectancy of batteries is rather short. LoRaWan is the main communication protocol for our research project, divided between operated and private networks. Operated LoRa network is a "plug-and-play" solution. It is fully managed by operators and depends on a national LoRaWan network. However, because it is operated, there are significant usage constraints regarding the amount of transmitted data. A private network is the obvious replacement solution. Nevertheless, it entirely relies on dedicated gateways. This is a precarious situation that results in data loss if there are any arising technical issue.

2.4.2.3 Calibration and long-term maintenance

Calibration and long-term maintenance are critical aspects to consider during the installation and supervision of a sensor network. They are essential to ensure the reliability of collected field data over the duration of the research project.

All sensors except for the thermal energy meters were calibrated by manufacturers before their installation. Thermal energy meters are calibrated upon installation – although temperature and flow measurements cannot be calibrated – and, as described in Section 2.4.2.2, it is recommended to regularly check on the calibration of these meters.

To account for potential measurement drifts, several options are available. Occasional measurements onsite can be made to compare with a calibrated sensor. If a difference is spotted between collected data and punctual measurements, sensors can be removed and sent back to their manufacturer for calibration. The measurement difference can also be accounted for in the following data analyses if the previous solution cannot be considered. In any case, onsite punctual measurements are challenging. They require several-hour measurements in apartments at different times over the research project, while one of the key aspects of the solution is to be as little intrusive as possible. If sensors are sent back to manufacturers, it results in a loss of data during the calibration, unless removed sensors are momentarily replaced by other sensors which have to be specifically set up for this application. This solution could not be implemented in our project because of limited number of available sensors and the consecutive sensor management. Furthermore, measurement comparison is more or less simple to implement, depending on the type of measures that are performed. IEQ, presence detection and contact temperature measurements can be checked. It is significantly more challenging for electricity-related sensors, when comparative measurements have to be acquired at the exact same time as for the sensors installed onsite.

Due to technical and time constraints, an extensive calibration study could not be conducted in our research project. Only IEQ sensors measuring IAT, relative humidity, luminosity, presence detection and CO₂ concentrations were investigated before their installation, to ensure the reliability of their measurements. The details of this study are discussed in Appendix B.2.

2.4.2.4 Data loss

A final aspect to improve in our sensor network is related to the missing data. Sensors were installed onsite and supervised to make sure they measured and communicated the

targeted data. Nevertheless, this does not mean collected data are flawless. There might be outliers (unexpected negative, zero or abnormal data, or error messages) or missing data. These are related to different identified causes, such as:

- Installation conditions and environment: walls, scaffolding, metal elements, sensor location or distance to the gateways;
- Unexpected sensors and gateways handling, from residents, construction teams or visitors, willingly or accidentally;
- Electricity shortage affecting sensors and gateways with grid power supply;
- Equipment failure: worn out batteries, crashed sensors or gateways, either because of manufacturing defect, punctual bugs, overuse or long-term usage defect;
- Network issues including difficulties for sensors or gateways to access GPRS or LoRaWan networks;
- Server issues.

Because of the nature of these issues, there is very little that can be done to prevent data loss. However, we observed that an efficient mean to reduce the loss of data to a minimum is to implement an automated data collection checking system. This was experimented on part of our sensor network, with a daily verification of the amount of collected data for a group of sensors we supervised, comparing the size of the daily data packages to the size of a normal data package, when there is no missing data. This solution could be further improved with a dedicated detailed process to check on the amount of data received for each individual sensor, and to send warning e-mails in case of a suspected data loss.

Conclusions

This thesis project depends on field data collection for the characterization of energy consumption (electricity, thermal energy, natural gas), local weather, indoor environment quality (temperature, humidity, luminosity, CO₂, temperature of indoor walls) and occupants' behavior (presence and window opening detection) in residential buildings. A case study of three retrofitted residential buildings is selected. A sensor network is deployed onsite, in common areas, on energy systems and in a eight-housing sample. The sensor is entirely wireless, using LoRaWan and GPRS communication protocols. It includes 170 sensors and meters distributed between shared areas with 26 sensors and households with 144 sensors, for an average of 18 sensors per housing. The weather station includes

BIBLIOGRAPHY

9 additional sensors. Data collection and storage are conducted using FTP and HTTP servers. The deployment is performed over two years and a half, in four different batches. Data are collected for over three years in common areas and for more than a year and half in apartments.

The deployment and operation of the sensor network provided a relevant feedback for the replicability of the project. Critical points include the installation conditions and environment, the targets of the sensor network and the project management.

Efforts were made to cover most aspects of building energy monitoring in a retrofit context. Nevertheless, some aspects were more depicted than others because of technological, human and budget constraints. Energy consumption is the most exhaustively covered category of measurements with 102 related sensors (57% of the solution). IEQ and residents' behaviors could be improved both in terms of data acquisition technologies and targeted measurements to provide a more complete overview of the case study. Calibration of the sensors network and long-term maintenance are significant topics as well. They should be further explored to ensure the reliability of the solution and of the collected data.

Despite these difficulties, the experimental set up resulted in a significant amount of collected data to deepen the knowledge and understanding of the operation of the building case study. Collected data should help assessing the efficiency of the retrofit actions and improving building energy modeling performances. Field data analysis, which is the first step towards these objectives, is presented and discussed in the following chapter.

Bibliography

- [1] ADEME, “Guide technique — Les besoins d’eau chaude sanitaire en habitat individuel et collectif,” tech. rep., 2017. 60
- [2] PROFEEL, “Programme PROFEEL — Innover pour la rénovation énergétique.” <http://programmeprofeel.fr/>, 2022. 62
- [3] ADEME, “Le plan de rénovation énergétique de l’habitat (PREH),” tech. rep., 2018. 62
- [4] C. Peng and Z. Wu, “In situ measuring and evaluating the thermal resistance of building construction,” *Energy & Buildings*, vol. 40, pp. 2076–2082, 2008. 63
- [5] Agence Parisienne du Climat, “Performance énergétique en copropriété - Retour sur deux ans d’instrumentation,” tech. rep., 2020. 63

BIBLIOGRAPHY

- [6] M. W. Ahmad, M. Mourshed, D. Mundow, M. Sisinni, and Y. Rezgui, “Building energy metering and environmental monitoring – A state-of-the-art review and directions for future research,” *Energy and Buildings*, vol. 120, pp. 85–102, may 2016. 64
- [7] R. Gopinath, M. Kumar, C. Prakash Chandra Joshua, and K. Srinivas, “Energy management using non-intrusive load monitoring techniques – State-of-the-art and future research directions,” *Sustainable Cities and Society*, vol. 62, p. 102411, nov 2020. 65
- [8] J. Kelly and W. Knottenbelt, “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes,” *Scientific Data*, vol. 2, p. 150007, mar 2015. 65
- [9] A. Martín-Garín, J. A. Millán-García, A. Bairei, J. Millán-Medel, and J. M. Sala-Lizarraga, “Environmental monitoring system based on an Open Source Platform and the Internet of Things for a building energy retrofit,” *Automation in Construction*, vol. 87, pp. 201–214, mar 2018. 65, 66
- [10] M. Karami, G. V. McMorrow, and L. Wang, “Continuous monitoring of indoor environmental quality using an Arduino-based data acquisition system,” *Journal of Building Engineering*, vol. 19, pp. 412–419, sep 2018. 65, 66
- [11] M. Frei, J. Hofer, A. Schlüter, and Z. Nagy, “An easily-deployable wireless sensor network for building energy performance assessment,” *Energy Procedia*, vol. 122, pp. 523–528, sep 2017. 65, 66
- [12] M. Frei, C. Deb, R. Stadler, Z. Nagy, and A. Schlueter, “Wireless sensor network for estimating building performance,” *Automation in Construction*, vol. 111, p. 103043, mar 2020. 65, 66
- [13] L. Jankovic, “Lessons learnt from design, off-site construction and performance analysis of deep energy retrofit of residential buildings,” *Energy and Buildings*, vol. 186, pp. 319–338, mar 2019. 65, 66, 67
- [14] D. Guyot, F. Giraud, F. Simon, D. Corgier, C. Marvillet, and B. Tremeac, “Detailed monitoring as an essential tool for achieving energy performance targets in operation conditions: The HIKARI case study,” *E3S Web of Conferences*, vol. 172, 2020. 65, 66
- [15] C. Deb, M. Frei, and A. Schlueter, “Identifying temporal properties of building components and indoor environment for building performance assessment,” *Building and Environment*, vol. 168, no. May 2019, p. 106506, 2020. 65, 66
- [16] K. Jnat, I. Shahrou, and A. Zaoui, “Impact of smart monitoring on energy savings in a social housing residence,” *Buildings*, vol. 10, no. 2, 2020. 65, 66

BIBLIOGRAPHY

- [17] J. Vivian, L. Carnieletto, and M. D. Carli, “Monitoring the change of indoor environmental conditions of refurbished buildings in Milan .,” pp. 1–8, 2022. 65, 66, 67
- [18] D. Guyot, F. Giraud, F. Simon, D. Corgier, C. Marvillet, and B. Tremeac, “Building energy model calibration: A detailed case study using sub-hourly measured data,” *Energy and Buildings*, vol. 223, p. 110189, sep 2020. 65
- [19] J. Yang, M. Santamouris, and S. E. Lee, “Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings,” *Energy and Buildings*, vol. 121, no. January, pp. 344–349, 2016. 67, 88, 93
- [20] M. A. Humphreys, “Quantifying occupant comfort: are combined indices of the indoor environment practicable?,” vol. 33, pp. 317–325, jul 2007. 72
- [21] Y. He, W. Chen, Z. Wang, and H. Zhang, “Review of fan-use rates in field studies and their effects on thermal comfort, energy conservation, and human productivity,” *Energy and Buildings*, vol. 194, pp. 140–162, jul 2019. 72
- [22] E. Biyik and A. Kahraman, “A predictive control strategy for optimal management of peak load, thermal comfort, energy storage and renewables in multi-zone buildings,” *Journal of Building Engineering*, vol. 25, p. 100826, sep 2019. 72
- [23] F. Asdrubali, F. D’Alessandro, G. Baldinelli, and F. Bianchi, “Evaluating in situ thermal transmittance of green buildings masonries — A case study,” *Case Studies in Construction Materials*, vol. 1, pp. 53–59, 2014. 72
- [24] G. Pasolini, C. Buratti, L. Feltrin, F. Zabini, C. De Castro, R. Verdone, and O. Andrisano, “Smart City Pilot Projects Using LoRa and IEEE802.15.4 Technologies,” *Sensors*, vol. 18, p. 1118, apr 2018. 75
- [25] Objenious, “Objenious by Bouygues Telecom.” <https://objenious.com/>. xxi, 76, 77
- [26] The Things Network, “The Things Network.” <https://www.thethingsnetwork.org/>, 2022. xxi, 76, 77
- [27] Iowa State University — College of Agriculture and Life Science, “Iowa Environmental Mesonet.” <https://mesonet.agron.iastate.edu/>, 2022. 80
- [28] Ladybug Tools, “EPW weather files map.” <https://www.ladybug.tools/epwmap/>, 2022. 80
- [29] ENEDIS, “Linky, le compteur communicant.” <https://www.enedis.fr/linky-compteur-communicant>. 81

BIBLIOGRAPHY

- [30] FLUDIA, “Solutions dédiées à la maîtrise de l’énergie — FM410 GPRS.” <https://www.fludia.com/BELSENSO-FM410-CAPTEUR-CONNECTE-GPRS.html>, 2022. 81
- [31] EWATTCH, “Capteur Tyness Energy.” <https://www.ewattch.com/portfolio-item/tyness-energy-lora/>, 2017. 82
- [32] NKE-Watteco, “LoRaWAN SMART PLUG.” <http://www.nke-watteco.fr/wp-content/uploads/2018/06/LoRaWAN-Smart-Plug-Fiche-Technique-v1.1.pdf>, 2022. 83
- [33] Energieplus-lesite, “Compter l’énergie thermique.” www.energieplus-lesite.be/index.php?id=17078{#}c20935101, 2022. xxi, 84
- [34] SensingLabs, “PUL-LAB-13XS Outdoor ATEX LoRa,” tech. rep., 2022. 84
- [35] SensingLabs, “TEM-LAB-14NS Module temperature LoRa,” tech. rep., 2022. 85
- [36] GRDF, “Compteur gaz communicant Gazpar : comment ça marche?.” <https://www.grdf.fr/particuliers/fonctionnement-compteur-gaz-communicant-grdf>. 85
- [37] Zone-atex.fr, “Normes ATEX et matériel ATEX.” Zone-atex.fr, 2019. 85
- [38] SensingLabs, “THY-LAB-61NS Capteur Temperature-Hygrometrie LoRa,” tech. rep., 2022. 86
- [39] ELSYS, “ERS LoRaWAN room sensor for measuring indoor environment.” <https://www.elsys.se/en/ers/>, 2022. 87, 88
- [40] I. Atmaca, O. Kaynakli, and A. Yigit, “Effects of radiant temperature on thermal comfort,” *Building and Environment*, vol. 42, pp. 3210–3220, sep 2007. 87
- [41] SensingLabs, “PIR-LAB-21NS Capteur PIR Indoor,” tech. rep., 2022. 88
- [42] SensingLabs, “OPE-LAB-13NS Transmetteur contact magnetique outdoor,” tech. rep., 2022. 88
- [43] M. Bourdeau, D. Werner, P. Basset, and E. Nefzaoui, “A Sensor Network for Existing Residential Buildings Indoor Environment Quality and Energy Consumption Assessment and Monitoring - Lessons Learnt from a Field Experiment,” no. *Sensornets*, pp. 105–112, 2020. 89, 90
- [44] W. Wei, O. Ramalho, and C. Mandin, “Indoor air quality requirements in green building certifications,” *Building and Environment*, vol. 92, pp. 10–19, 2015. 91

CHAPTER 2 – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

- [45] Y. Allab, M. Pellegrino, X. Guo, E. Nefzaoui, and A. Kindinis, “Energy and comfort assessment in educational building - Case study in a French university campus,” *Energy and Buildings*, vol. 143, pp. 202–219, may 2017. 91
- [46] G. K. Oral, A. K. Yener, and N. T. Bayazit, “Building envelope design with the objective to ensure thermal , visual and acoustic comfort conditions,” vol. 39, pp. 281–287, 2004. 91
- [47] International Organization for Standardization, “EN ISO 9869 — thermal insulation : Building elements : In-situ measurement of thermal resistance and thermal transmittance,” tech. rep., 2017. 91
- [48] F. Asdrubali and G. Baldinelli, “Thermal transmittance measurements with the hot box method: Calibration, experimental procedures, and uncertainty analyses of three different approaches,” *Energy and Buildings*, vol. 43, no. 7, pp. 1618–1626, 2011. 91
- [49] The Modbus Organization, “The Modbus Organization.” <https://www.modbus.org/>, 2019. 93

Chapter 3

Processing and analyses of field data for building energy modeling enhancement

The third chapter focuses on the analysis of field data collected through our sensor network. It starts with a description of the data pre-processing, including data cleaning, formatting, quality assessment, reconstruction, aggregation and processing of specific datasets. From the many available measurements, a set of five parameters is selected, namely the indoor air temperature, the occupancy, the electric dissipated power, window opening and domestic hot water use, to target energy consumption, indoor environment quality and usages. The second part of this chapter is dedicated to the analyses of these parameters using descriptive statistical analysis techniques and daily profile clustering. Finally, this chapter is concluded by a discussion on analyses results and remaining opportunities for a better understanding of the buildings operation.

Introduction

As a result of the sensor network presented in Chapter 2, a large amount of field data is collected. Collected data provide a variety of opportunities to acquire better knowledge on building energy consumption, building energy management, but also other aspects including the indoor environment quality and the interactions between occupants and the buildings.

Prior to any data analysis, data processing has several targets. The first is the essential cleaning and formatting, to ensure data quality, to retain the essential information and to work with a unified file format. Data are also partially rebuilt to collect enough information and to perform reliable analyses. Then specific datasets are processed depending on the intended analyses, with temporal aggregation and modifications to match modeling inputs.

In this chapter, the goal of data analyses is to use field data to improve modeling assumptions, specifically on usage scenarios and the regulation of energy systems, through the study of the impact of occupants' behavior on building energy consumption.

There is wide variety of tools and techniques for statistical analyses on occupants' behavior, energy consumption and indoor environment such as classification algorithms [1], artificial neural networks [2] or statistical analyses, onsite diagnoses and surveys [3]. Because of the many different monitored parameters and the large quantity of collected data, it is hardly possible to implement specific tools for each targeted measurement. The choice was made to build on previous work conducted by our research team on time series analyses [4, 5]. Analyses focus on the characterization of data distributions over selected timeframes such as heating periods, and the extraction of typical patterns using classification techniques.

In the present chapter, the first section briefly introduces the tools used to perform data analyses. Section 2 details the pre-processing of collected data with cleaning and formatting, data quality assessment, reconstruction and aggregation, and specific measurements processing. Section 3 presents the analysis techniques and results, focusing on indoor air temperature and heating patterns, occupancy, electric dissipated power, window opening and DHW.

3.1 Materials for data processing and analysis

Data processing and analysis are performed using Python 3.7.12. Two environments are used, with Spyder 3.8 and Google Research Colab 7.5.0.

Most of the processing rely on Pandas 1.3.5 and Numpy 1.21.5 packages, along with Statistics and Math modules. Date and time formatting is supported by Datetime module. Graphs and visualization are made with Matplotlib 3.2.2 and Seaborn 0.11.2.

Clustering is performed using R 4.1.1 in RStudio environment, with NbClust 3.0 and TSclust 1.3.1 packages.

Data storage and part of the visualisations are performed using Microsoft Excel.

3.2 Data pre-processing

3.2.1 Cleaning and formatting

Prior to advanced data pre-processing steps and data analyses, it is essential to "prepare" the data with cleaning and formatting.

Data cleaning aims the recovery of critical information: date, time, measurement and unit of measure. Remaining data such as battery level, error codes, text messages are removed. In a second time, inconsistent data are also treated. Formatting standardizes the data filing structure to help implementing unified coding scripts for processing and analysis.

Regardless of the type of measurement, data from a given sensor can be related to a unique identification number, called DevEUI (device extended unique identifier). However, because of the different types of sensors, measurements, manufacturers and communication protocols, the sensor network initially provides data in different format. Data collected from step 1 batch 1 – common areas and energy systems – and step 2 batch 2 – electricity consumption and natural gas in housings (Section 2.3.1) – are managed by a contractor. One file is generated every day for each sensor and attributed with their respective DevEUI. However, it combines the different measurements if a sensor monitors more than one parameter – for example, temperature and humidity measurements in common areas. The weather station (Step 1 batch 2) creates one file for all measurements every day. Data from sensors in apartments (step 2 batch 1) come in two different formats. Sensors for IEQ and presence detection send their data on a dedicated FTP server with limited data storing. Other sensors from this batch are treated separately. Each gateway

provides one JSON (javascript object notation) file every day that aggregates data for each connected sensor. JSON files are decoded and cleaned to retrieve the expected information.

From the different sources of data, a unified format is created: one CSV file for each measurement of each sensor, with daily update. Data are presented as time series in columns. New data are added under the existing data. Another format is also used specifically for clustering applications, with daily profiles (DP). CSV files are updated daily with one new line for each day of data.

3.2.2 Data quality assessment, reconstruction and aggregation

3.2.2.1 Data quality

After cleaning and formatting, data quality is assessed before the reconstruction and aggregation processes. Data quality assessment highlights the gaps in collected data. These gaps have various origins. They also impact on the choice of available analysis timeframes and data analysis techniques. The assessment of data quality provides two types of information:

- The amount of collected data compared to the amount of expected data – it gives an overall picture of the performance of the sensor network for different types of measurements and different locations in the studied buildings;
- The time periods when one day or more data are missing for a given sensor – it highlights significant periods when data are missing and helps investigate the causes of data shortage.

A review of the performances of data collection is summarized in Appendix C.1, with data collection rates for each type of sensor in the housings and common areas.

The main identified sources of data loss are:

- Communication issues: between sensors and gateways, between sensors and the Internet (for GPRS), between gateways and servers;
- Unexpected equipment handling: removed or damaged sensors, unplugged gateways;
- Equipment failure: discharged batteries, sensor failure, power cuts, server crashing.
- Issues regarding data accessibility with contractors, specifically for some connected plugs, pulse sensors for natural gas consumption and pulse sensors for electric smart

metering.

3.2.2.2 Data reconstruction

Data reconstruction is performed to replace small parts of the missing data. Reconstruction applies to parameters including indoor air temperature (IAT), presence detection, dissipated power (from electric power demand), DHW temperature and window opening, that are further analyzed in the following sections.

The goal is not to explore the topic of time series reconstruction nor to develop advanced reconstruction tools. In this context, data reconstruction aims to increase the amount of available data, while keeping reconstructed information as realistic as possible and so that they are representative enough of the existing building operation. The targets of reconstruction are short periods of time and punctual missing data points. Indeed, when too many consecutive data are missing, the dynamics of time series can hardly be reproduced without the implementation of more complex algorithms.

The first step towards reconstruction is to assess the randomness and distribution of missing data. Because data loss is related to unexpected events that are completely independent from the measurements – i.e. there is no relationship between the fact the data are missing and the type of data that are collected – data are missing completely at random and missing data distribution is arbitrary within the datasets.

Reconstruction and filling of time series with missing data is a specific research topic with many different available techniques. According to the literature, methods can be grouped based on their implementing complexity [6]. Straightforward techniques rely on imputation of missing data, either statistical (mean, regression) or machine-learning-based such as k-NN or ANN. More complex approaches are model-based and implement a data-driven model of time series to fill the gaps in datasets. In the present case, we aim reconstruction for pattern classification in time series data. Among the many available techniques, imputation with mean calculation of the close datapoints is selected.

A final aspect of the reconstruction is the reconstruction threshold. This parameter defines the maximum amount of consecutive missing data points that can be reconstructed, using the latest acquired measurements. The threshold is set depending on both a number of data points and time scale, relating to the selected sensor, measurement and its physical meaning. For instance, IAT in apartments is not expected to show sudden variations. However, data are already acquired at half-hour time-step. Therefore, the threshold is set to two data points, then one missing hour that can be reconstructed at most. Electric power demand at housing-scale is measured with one-minute time-step and

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

is most likely to rapidly and significantly increase or decrease. The threshold is set to 4 minutes. The summary of thresholds for the different measurements is given in Table 3.1.

Table 3.1: Sensors, data and thresholds for reconstruction.

Sensor and type of data	Acquisition time-step	Reconstruction threshold
Contact temperature of DHW pipes (°C – SensingLab TEM-LAB-14NS)	1 minute	15 data points
IAT (°C – ELSYS ERSCO2)	30 minutes	2 data points
Presence detection (number of passing – ELSYS ERSCO2)	30 minutes	2 data points
Pulse sensor for electric power demand in smart meters (Wh – Fludia BelSenso FM410e) for dissipated power	1 minute	4 data points
Window opening detection (N.A. – SensingLab OPE-LAB-41NS)	Event-based, resampled to 1 minute	60 data points

Results of data reconstruction are provided in Table 3.2. They highlight the impact of this pre-processing step over the instrumentation period. The effect of the reconstruction is significant on IAT data and presence detection data (performed by the same sensor), and on DHW temperature data. ELSYS CO₂ sensors, for the former two parameters, often lack a few data points due to poor cellular network. This can easily be corrected with the implemented reconstruction strategy. Similarly, DHW data, with a one-minute acquisition time-step, can easily be filled to improve daily profile completion. Reconstruction barely impacts on electric power demand because Fludia sensors loose very little data when they are operational. However, less than 40% of the expected profiles are available due to several large time periods of malfunction. Finally, for window opening detection, no additional data could be recovered with reconstruction because of the selected reconstruction threshold and event-based data acquisition.

3.2 Data pre-processing

Table 3.2: Impact of data reconstruction and aggregation on daily profile completion – statistics relate to the whole instrumentation period.

	Collected		After reconstruction		After aggregation		Expected
	Complete profiles	Acquired vs. expected	Complete profiles	Acquired vs. expected	Complete profiles	Acquired vs. expected	Complete profiles
IAT	430	7%	3,964	63%	4,030	65%	6,258
Presence	430	7%	3,964	63%	4,030	65%	6,258
Electric power demand	1,681	32%	1,797	35%	1,930	37%	5,197
Window opening	6,784	39%	6,784	39%	10,001	57%	17,405
DHW	202	4%	1,171	23%	3,039	59%	5,153

3.2.2.3 Aggregation

As the many different sensors collect data at different time-steps, temporal data granularity requires a unified time-step to perform data analyses. Hourly time-step appears to be the most consistent choice. For energy consumption, it directly relates with the measurement unit (kWh). Also, most of the collected data aim to be used as operation scenarios in building energy models, and modeling software rely on scenarios with hourly time-step.

Aggregation is performed with the average of measurements within the same hour, for all types of data. Energy consumption measurements, specifically electricity consumption on smart meters is modified into power demand prior to aggregation.

Aggregation also impacts on data availability for analyses. At first, within a given hour, several datapoints may be missing. After hourly aggregation, considering that data aggregation is an hourly average, the completion of daily profiles is higher than with a lower time-step (Table 3.2). It significantly enhances daily profile completion for window opening and DHW temperature data, since both have a one-minute acquisition time-step (defined acquisition time-step for the former, and resampling time-step for the latter).

3.2.3 Processing of datasets

In addition to quality assessment, reconstruction and aggregation, an additional specific process is also applied depending on the type of acquired data, either before or after the three other steps described above. This process applies to DHW temperature, window opening and presence detection data.

Window opening data collection is event-based: data are acquired only when the opening status of the window changes from opened to close or counterwise. A data point is also sent every hour after the last update if the opening status has not changed within the past hour. Hence, window opening requires a steady time-step to be analyzed as time series. Resampling is performed at one-minute time-step. Resampled missing data are replaced following the reconstruction process. Aggregation is performed at last.

Presence detection sensors are detecting and counting the number of passages and movements within the range of the sensors. The number of passages is used to assess the occupancy ratio of the apartment: if there is one passage or more, the occupancy ratio is set to 1, otherwise it is set to 0. A checking is performed during night-time when occupants are expected to be asleep and to ensure there is no faulty detection from the sensor. Reconstruction and aggregation are performed afterwards.

DHW temperature is monitored using one-minute temperature measurements. However, energy modeling scenarios rely on hot water volume consumption for a given temperature. Water consumption cannot be measured since meters are already installed and operated by the billing company, and data cannot be accessed. Daily hot water volume consumption is extracted from the study conducted by ADEME [7] on DHW needs in residential housings. Two types of data can be used: daily DHW use depending on the number of people in households, and daily DHW use depending on the size of the considered apartment. The latter is not as accurate as the former since in the present case study, large apartments are sometimes occupied by one or two people instead of a family. Hence, daily DHW use per people is the selected information, as detailed in Table 3.3. Reconstruction, aggregation and clustering are performed prior to transformation of temperatures to DHW usage.

A 40°C-temperature threshold is applied on DHW temperature measurements to match with the data extracted from the study of ADEME [7], along with binary processing: any datapoints of 40°C and higher is replaced by one, for one minute of DHW use. Other datapoints are replaced by zeros. Reconstruction is performed prior to this step and aggregation is performed after, to give a number of minutes of DHW use for each

3.3 Data analyses for building energy modeling enhancement and calibration

hour of the day. The daily DHW use from Table 3.3 is evenly distributed between each minute of DHW use. This process is depicted in (Figure 3.1).

Table 3.3: Averaged DHW daily needs per occupant and household (unit: liters, at 40°C temperature) – Source: [7].

	Number of people in housings				
	1	2	3	4	5
Average use per occupant (liters)	80 ±35	60 ±25	50 ±20	45 ±20	45 ±20
Average use per housing (liters)	80 ±35	120 ±45	150 ±50	170 ±70	220 ±105

3.3 Data analyses for building energy modeling enhancement and calibration

3.3.1 Analysis methods

As stated in the previous section, analyses methods are implemented on five types of data: IAT, presence detection, DHW use, window opening and electric power demand. Two types of data analyses are performed: descriptive statistical analyses and daily profiling through clustering.

3.3.1.1 Descriptive statistical analyses

Descriptive statistical analyses focus on the characterization of data distributions in different samples. It is the first step to provide a relevant picture of different patterns in collected field data.

Six tools are implemented on Python to provide different types of graphical and statistical analyses. These can be distributed in three categories.

At first, a simple graphical inspection of time series is performed to highlight potential missing data or long term (e.g. annual or monthly) patterns (Figure 3.2(a)). Calendar mapping (Figure 3.2(f)) can be used for a similar purpose, using a three-dimensional graphical representation with hours on abscissa axis, days on ordinate axis and a color code from the lowest to the highest value of the considered parameter.

A second category of tools provides quantitative information on data distributions. Adapted from energy balance tools, cumulative frequency curves (Figure 3.2(d)) rank the values of a parameter from the highest to the lowest value and their corresponding

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

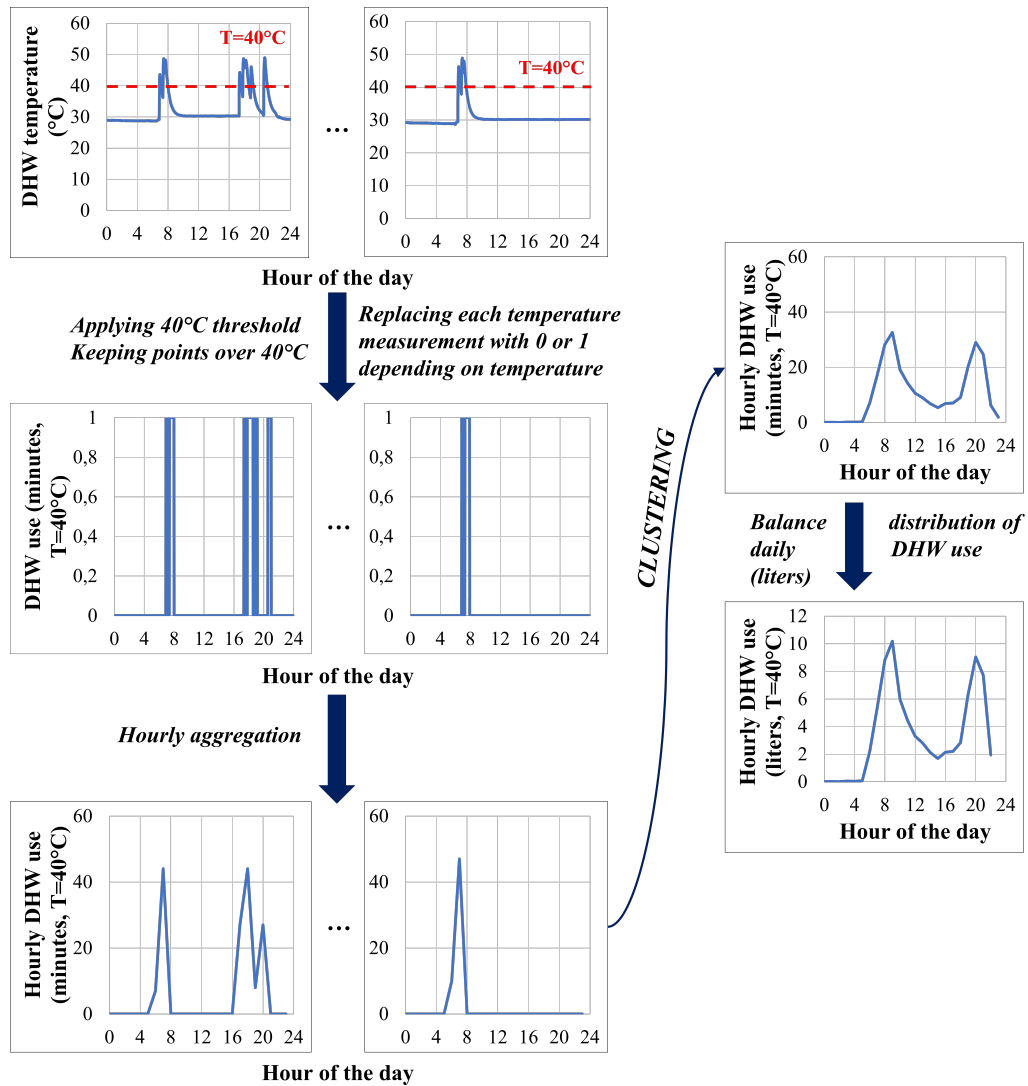


Figure 3.1: Processing of DHW temperature measurements into volume consumption DP.

3.3 Data analyses for building energy modeling enhancement and calibration

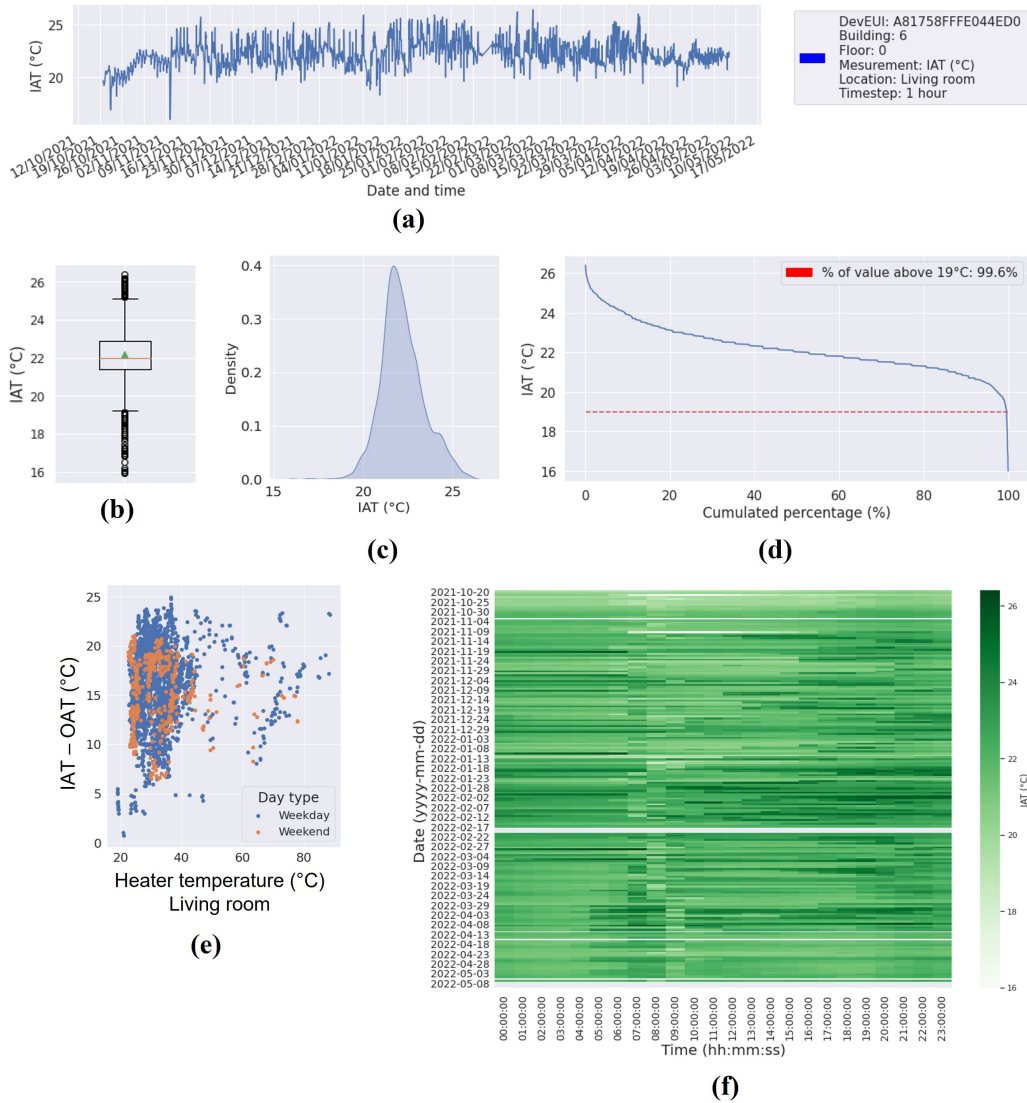


Figure 3.2: Tools for descriptive statistical analyses: (a) time series graph, (b) box-plot, (c) density plot, (d) cumulative frequency curve, (e) scatter plot, and (f) heatmap – displayed data are IAT measurements in B2/0 from 2021/10/20 to 2022/05/10.

frequency. Using different thresholds, it is possible to know the percentage of values ranging above the thresholds for a given dataset. Density curves rely on a probability density function to visualize what range(s) of values comprise most of the data (Figure 3.2(c)). Both tools provide relevant insights on the predominant patterns in the collected field data. They can also help comparing different datasets, such as day and night or weekdays and weekends. They can be complemented with box-plots that give the median, first and third quartiles (Figure 3.2(b)). They also provide two other values, the "whiskers" of the boxplot, which are the first value greater (for the higher whisker) or lower (for the lower whisker) than the third quartile plus the inter-quartile (for the higher whisker) or than the first quartile minus the inter-quartile (for the lower whisker).

Scatter plots show two different datasets on the same graph – the temperature of the heater of the living room of B2/0 as a function of the difference between the IAT minus the OAT in Figure 3.2(e) – to investigate potential correlations. However, for the analysis to be accurate, it requires that data from the different datasets must have the same time-step and must be collected at the same time.

3.3.1.2 Daily profiling

Following descriptive analyses, classification tools are implemented on daily profiles (DP) with hourly time-step. The goal of the clustering serves three purposes: to identify typical operation patterns and provide a better understanding of the building use, to embed realistic operation scenarios in energy models and to keep the use of data to a relatively simple level.

The first purpose helps highlighting the diversity of operation profiles and understanding causes of such a diversity. The second purpose underlines the difference between field collected data and standard data used in energy models. The third purpose keeps the use data to a reasonable complexity. Using collected data without clustering could be a solution, but there would not be any identification of operation patterns. Also, for the sake of replicability, it is barely a practical solution to consider. It is quite uncommon to have such an opportunity of detailed building instrumentation. Moreover, there are missing data that cannot be reconstructed. Using data for the same days over another year cannot be done, considering that these data might not be available or that they would depict behaviors from before the retrofit of buildings.

Daily profiling is performed by clustering. Clustering process is adapted a study we conducted on daily load profile characterization [5]. Although input data are different, they remain building operation time series with similar patterns. The clustering process

3.3 Data analyses for building energy modeling enhancement and calibration

provides an almost-automated identification of typical operation patterns and with low bias. Indeed, the clustering and calculations are fully automated but it requires an intervention from the user, specifically when the cluster boundaries are defined and when a tied result is obtained regarding the optimal number of clusters.

Clustering process is summarized in Figure 3.3. Following data pre-processing, daily profiles (DP) with hourly time-steps are used as inputs. The goal is to compare and group similar DP using a k-means clustering algorithm [8] and Euclidean distance calculation. Euclidean distance has proven to be an efficient distance calculation comparing DP [9]. The number of clusters – i.e. the number of groups of similar DP – is given by cluster validation indices (CVI) which are evaluation metrics to assess the optimal number of clusters in a given dataset. The R package TSClust used for clustering computation embeds thirty CVI. Five CVI commonly used in the literature for DP clustering are selected [5]: Calinski-Harabasz [10], Silhouette [8, 10], Davies-Bouldin [10, 11], gap statistic [12] and Dunn [10]. The optimal number of clusters is selected with a majority rule: the highest number of CVI that provide the same results identifies the optimal clustering results. In case of tied results, both options are analyzed manually by the user. Cluster distributions are also compared to make sure they are identical. Then the optimal distribution of DP is extracted and illustrated with two types of graphical illustrations: i) graphs of DP for each cluster with the corresponding averaged DP of the cluster (black bold profile), and ii) a calendar mapping of clusters – each cell is related to a given day and different colors are attributed to the different clusters. Averaged DP of each cluster are expected to be representative of a specific typical situation that can be further investigated through cross-analyses with other types of data (type of day or meteorological data, for example).

Daily profiles are obtained using different data timeframes, depending on available data and the purpose of the scenarios. Presence detection, DHW use, electric dissipated power (from electric power demand) and window opening are analyzed over the largest available timeframe and a maximum of one full year of data – clustering algorithms may under-perform with more data [5]. However, IAT is only taken into account over the heating season as it is assimilated to a heating setpoint temperatures in building energy models. The details of selected timeframes for the different measurements are summarized in Table 3.4.

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

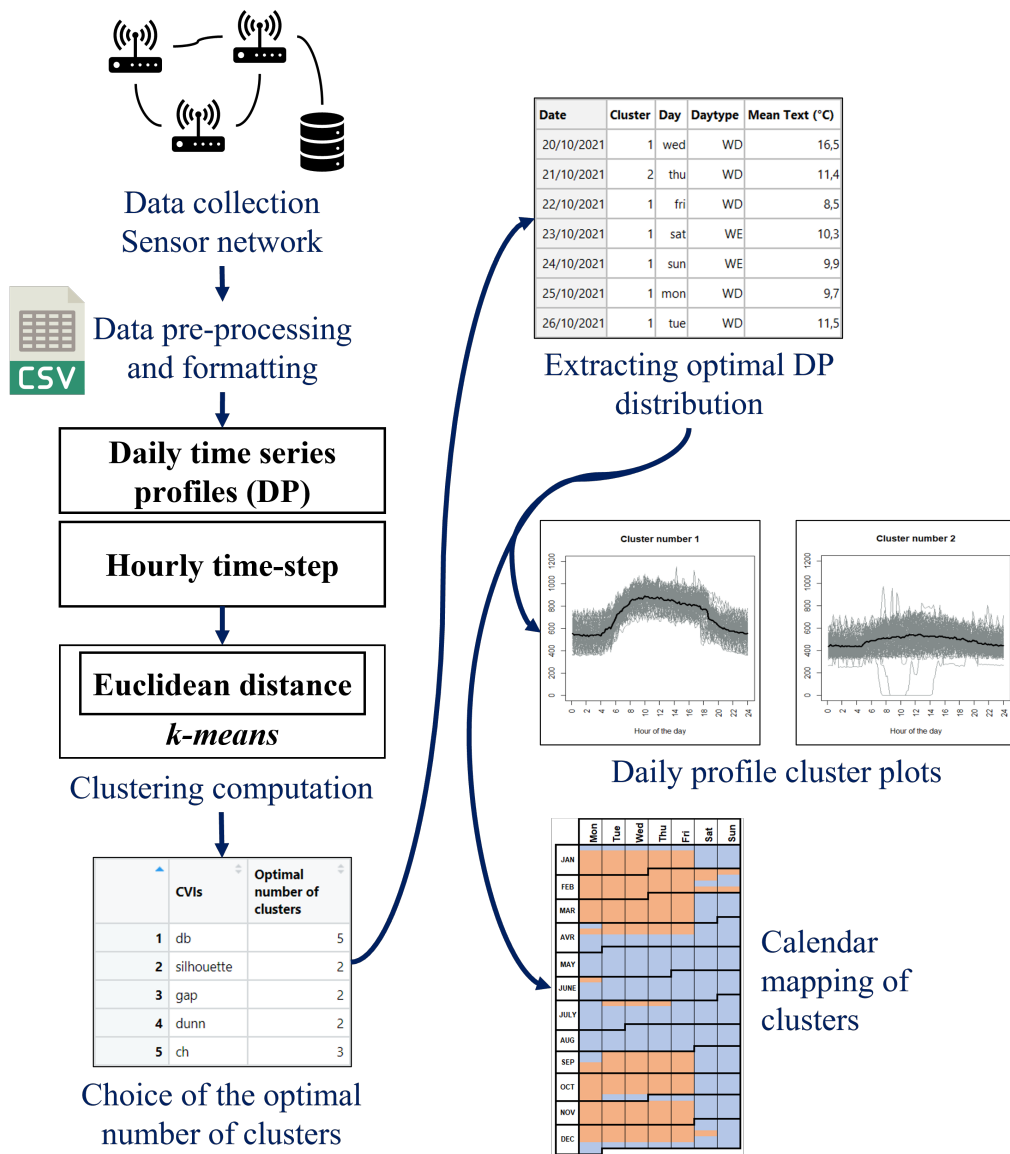


Figure 3.3: Schematic of the clustering process for DP classification.

3.3 Data analyses for building energy modeling enhancement and calibration

Table 3.4: Used timeframes for data processing and analyses – In Bi/j, i stands for the building number and j for the floor number.

		Indoor temperature	Presence	Dissipated power	Window opening	DHW
B1/2	From	2021/10/20	2021/6/10	2021/5/8	2021/6/1	2021/1/25
	To	2022/5/7	2022/5/7	2021/12/4	2022/5/31	2022/1/6
	Available profiles	80%	80%	95%	56%	56%
B1/3	From	2021/10/20	2021/6/10	2021/5/8	2021/6/1	2021/6/1
	To	2022/5/7	2022/5/7	2021/12/4	2022/5/31	2022/5/31
	Available profiles	80%	80%	97%	87%	97%
B2/0	From	2021/10/20	2021/6/10	2021/5/8	2021/6/1	2021/6/1
	To	2022/5/7	2022/5/7	2021/12/4	2022/5/31	2022/5/31
	Available profiles	81%	81%	100%	77%	93%
B2/1	From	N.A.	N.A.	2021/5/8	2021/6/1	N.A.
	To	N.A.	N.A.	2021/12/4	2022/5/31	N.A.
	Available profiles	N.A.	N.A.	98%	83%	N.A.
B2/2	From	N.A.	N.A.	2021/5/8	2021/6/1	N.A.
	To	N.A.	N.A.	2021/12/4	2022/5/31	N.A.
	Available profiles	N.A.	N.A.	95%	82%	N.A.
B2/5	From	2021/10/20	2021/6/10	N.A.	2021/6/1	N.A.
	To	2022/5/7	2022/5/7	N.A.	2022/5/31	N.A.
	Available profiles	80%	80%	N.A.	40%	N.A.
B3/0	From	2021/10/20	2021/6/10	2021/5/8	N.A.	2021/6/1
	To	2022/5/7	2022/5/7	2021/12/4	N.A.	2022/5/31
	Available profiles	76%	76%	100%	N.A.	92%
B3/2	From	2021/10/20	2021/6/10	2021/5/8	2021/6/1	N.A.
	To	2022/5/7	2022/5/7	2021/12/4	2022/5/31	N.A.
	Available profiles	80%	80%	98%	52%	N.A.

3.3.2 Towards a better understanding of building operation and usages

An improved understanding of building operation patterns depends on the analysis of key parameters and energy-drivers. Among these drivers, occupants' behavior holds a significant place [13]. Then, the characterization of IEQ provides a relevant picture to investigate building energy consumption, specifically the thermal energy consumption [3]. In building energy models, these aspects are described through building operation scenarios. Standard scenarios are extracted from Pléaides energy modeling software. However, these scenarios do not account for the characteristics of our building case study, which may result in potential PG between simulated building energy consumption and measured energy consumption.

In this section, IAT, dissipated electric power, occupancy, window opening and DHW use are analyzed. Dissipated electric power is deducted from electric power demand and occupancy results from the processing of presence detection. Standard scenarios from Pléaides modeling software are compared to scenarios obtained from clustering. The objective is to identify the differences between standard modeling scenarios and field data, but also to highlight a relevant and effort-less data analysis process to customize building energy models.

3.3.2.1 Indoor air temperature and heating patterns

Indoor temperature analysis is performed in two steps using descriptive statistical analysis and clustering.

A regulatory scenario from Pleiades modeling software would offer four temperature profiles, for work days, Wednesday, weekends and frost protection, respectively. Temperature varies from 16°C to 19 °C (Figure 3.4) and is set to 7°C for frost protection (only happening on the last week of the year in December).

Results extracted from box-plots and cumulative frequency curves, summarized in Table 3.5, show a first eye-opening observation: the mean temperature over the heating season 2021/2022 in the living rooms is much higher than what regulatory scenarios suggest. It ranges between 21.4°C and 24.9°C in average with at least 97.9% of the temperature measurements that are above the 19°C threshold. The range of temperature within an apartment is also significant, between a 4°C and 8.5°C-difference when extreme temperatures are compared over the heating period 2021/2022 (from October the 20th, 2021 to May the 10th, 2022). Comparing the instrumented apartments, there is a large

3.3 Data analyses for building energy modeling enhancement and calibration

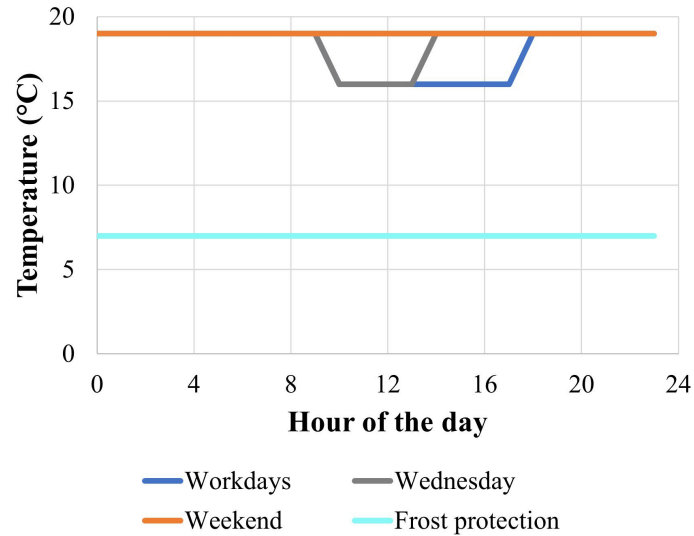


Figure 3.4: Temperature profiles (°C) from Th-BCE calculations in Pléaides software.

diversity of temperature ranges as well (Figure 3.6). Although the instrumented sample is relatively small with six housings, it seems that the higher the apartment floor, the higher the temperature. This would be expected because of a temperature gradient within the building and higher solar gains on higher floors. Nevertheless, this ought to be checked for other housings under the building ceiling which have more cold-wall surface than the rest of the housings. Furthermore, the apartment B3/0 – ground floor with North-East orientation – depicts unexpectedly high temperature levels. Possible reasons might be an unexpected handling of the sensor from the residents and the location of the sensor that might have been changed. It may also be related to the use of an additional heat source nearby the sensor. It should be specified that most sensors are placed around a height of two meters, usually on a furniture, to avoid for the sensors to be handled by occupants. Therefore, measured temperatures might be slightly higher than what should be measured at human height (around 1.5 meters) and is usually considered as the reference room temperature. Nevertheless, taking into account the measurement error reported by the manufacturer (0.2°C), these parameters cannot be the only explanations for such high IAT.

Daily temperature profile clustering highlight a second important information as temperature in instrumented apartments is constant over the day, as shown in Figure 3.7 for B2/0. Similar results are observed for other instrumented housings, and provided in Appendix C.2. A lower temperature at night would be expected, as it is recommended for optimal sleeping conditions [14]. Also, an overall decrease of indoor temperature would

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

	Q1-1.5*IQ	Q1	Median	Mean	Q3	Q3-1.5*IQ	% of data over 19°C	Orientation
B1/2	19.7	21.0	21.5	21.4	21.9	23.2	98.7%	North-East
B1/3	20.9	22.5	23.0	23.0	23.6	25.2	100%	North-West
B2/0	19.2	21.4	22.0	22.2	22.9	25.1	99.5%	North-East
B2/1	<i>No available data</i>							South-West
B2/2	<i>No available data</i>							South-West
B2/5	19.6	22.3	23.2	23.1	24.1	26.8	97.9%	South-West
B3/0	20.5	23.6	25.0	24.9	25.8	29.0	100%	North-East
B3/2	20.0	22.7	23.9	23.7	24.5	26.4	99.8%	North-West

Q1: first quartile Q3: third quartile IQ: interquartile (Q3-Q1)

Figure 3.5: Statistical description of temperature measurements in instrumented apartments.

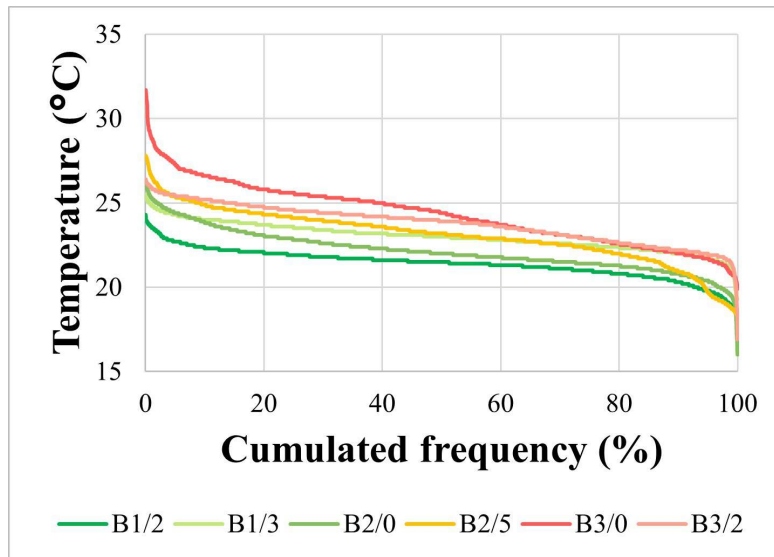


Figure 3.6: Cumulative frequency curves of measured temperatures in instrumented apartments.

3.3 Data analyses for building energy modeling enhancement and calibration

surely result in significant energy savings. However, this heating management strategy is not identified in the present case. The analysis of cumulative frequency curves and density curves also confirm observations from clustering results. There is no water logic for heating management over the day and the night or on weekdays and weekends, as depicted in B2/0 in Figure 3.8 (density curves for the remaining instrumented apartments can be found in Appendix C.3).

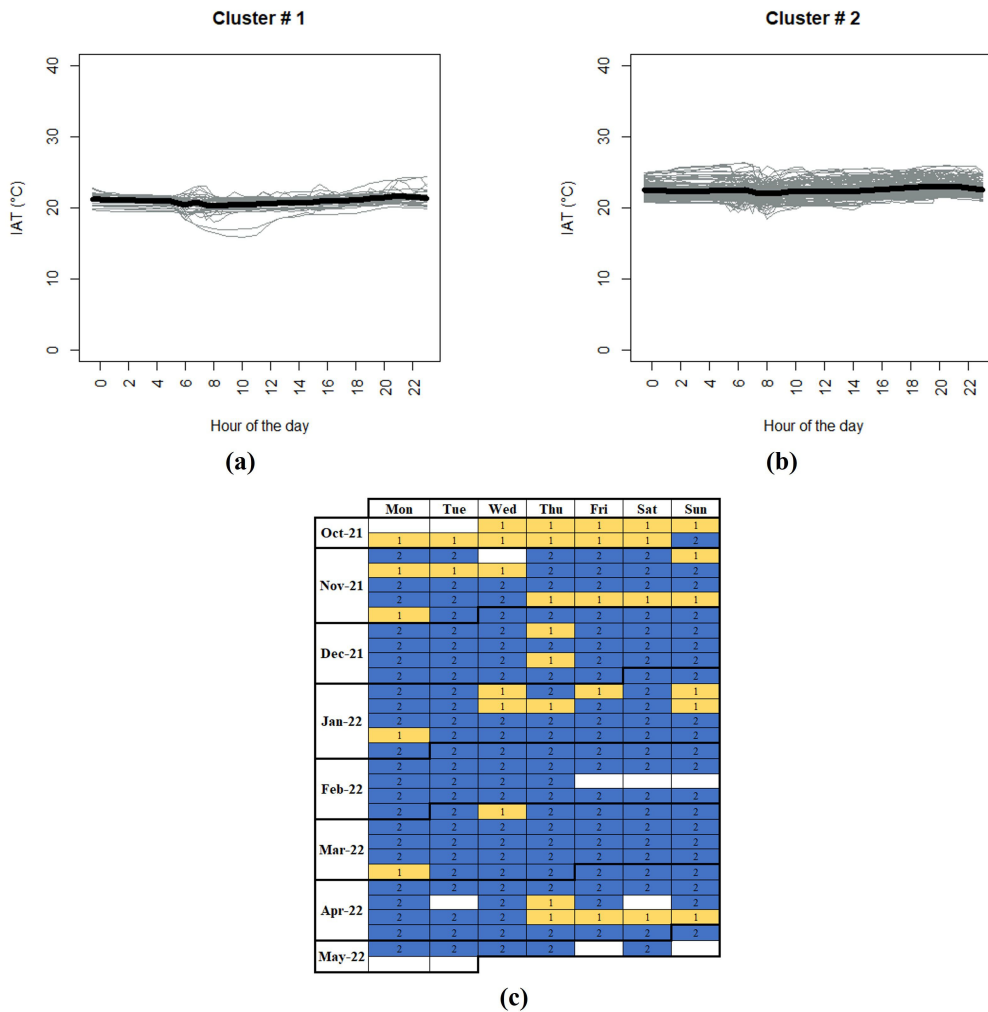


Figure 3.7: Results of clustering on IAT in B2/0, with daily profiles in cluster 1 (a) and cluster 2 (b), and calendar mapping of clusters (b) – the black bold line is the mean profile of each cluster; the white cells in the heatmap are missing DP.

For IAT measurements, clustering always results in a two-cluster DP distribution. Figure 3.7 (a) and (b) highlight the typical distribution of IAT DP in B2/0, which is a typical housing among the instrumented sample. Clusters originate from the different

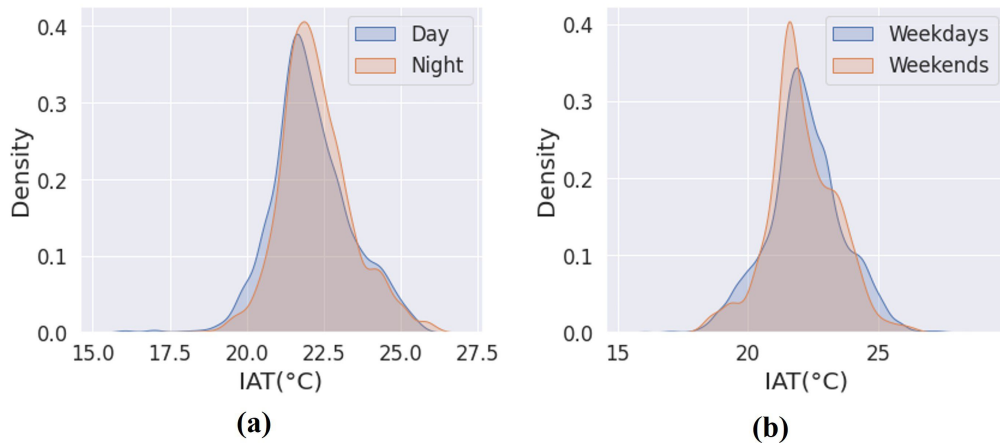


Figure 3.8: Density curves for B2/0, comparing IAT measurements for day time and night time (a), and during weekdays and weekends (b).

temperature levels over the heating period. However, there is no obvious correlation with the type of day (work day or weekend) (Figure 3.7 (c)) which is the driver to define operation scenarios in an energy modeling software. Considering the absence of water logic and the constant daily temperature profiles, IAT DP for building energy modeling are designed with constant averaged temperature measured in apartments over the heating period and extracted from Table 3.5.

3.3.2.2 Occupancy

Occupancy in apartments results from the processing of presence detection in the living rooms. Clustering is performed to extract the main occupancy profiles. Data are collected from June the 10th, 2021 to May the 7th, 2022. A full year could not be considered because of missing data in May and June 2021 and in May 2022.

Clustering results are illustrated in Figure 3.9, Figure 3.10 and Figure 3.11, for B1, B2 and B3, respectively. The results clearly depict the activity in instrumented apartments. There are three typical DP distribution: i) workdays versus weekends/days off in apartments B1/2 and B2/0, ii) workdays, half workdays and weekends in apartment B1/3, and iii) apartments with one main activity profile – B2/5, B3/0 and B3/2 – when the second activity profile is either very similar to the first or shows the complete absence of activity in the apartment. Daily profiles show the mean occupancy profile for each cluster in the different apartments.

As highlighted in the calendar mapping of Figure 3.9, Figure 3.10 and Figure 3.11, clustering results are not perfectly following a type-of-day configuration, although the

3.3 Data analyses for building energy modeling enhancement and calibration

main trends do. A rationalisation of results is necessary to depict the global picture of occupancy and remove specific events that cannot be predicted. In details, in apartments B3/0 (Figure 3.11(a)) and B3/2 (Figure 3.11(b)), the main occupancy profiles highlight an activity in the apartments all along the day with a gap at night since sensors are located in the living rooms. The other cluster shows a very small activity in the apartment that seems, according to cluster mapping, to be related to a prolonged absence of occupants. Although the second cluster provides relevant information, they are punctual events that cannot be considered to picture the typical building operation. For these two apartments, only cluster number one is considered to depict occupancy scenarios along the year. Similarly, in apartment B2/5 (Figure 3.10(b)), there is one main occupancy profile (cluster 1). The second profile is very similar to the first and only gathers fifteen DP over the observation period (4% of the sample). Hence, only the profile from cluster 1 is considered. Occupancy in all three apartments can be related to their occupants: B3/0 and B3/2 have retired occupants and B2/5 hosts an unemployed occupant (Table 2.4). Then, a steady presence over the day is not surprising.

The other three apartments – i.e. B1/2, B2/0 and B1/3 – behave differently. B2/0 (Figure 3.10(a)) shows a workday profile corresponding to the part-time professional activity of the occupants: gap of activity in the morning, occupancy at noon and from 3p.m. The second profile (cluster 2) shows a steady occupancy all along the day and is mapped as a day off, happening on weekends and on a random weekday. However, it seems that from the middle of January, the occupant of B2/0 stays home as on weekends, which is occupancy configuration retained for the energy model. B1/2 has similar occupancy patterns (Figure 3.9(a)): worked days versus days off. However, days off (cluster 2) are mostly Mondays and Sundays with the timeframe between December and January. Worked days (cluster 1) are from Tuesday to Saturday. Also, unlike other apartments, the workday profile shows a high minimum of occupancy during the day. This might be related to the occupants' schedules. One works with office hours. The second is a high school student. Finally, B2/3 has three occupancy profiles (Figure 3.9(b)): i) a worked day with continuous absence during the day (cluster 1) figuring Mondays, Tuesdays, Thursdays and Fridays; ii) a second cluster with a later peak of activity in the morning and small gap of occupancy in the afternoon, figuring weekends; iii) a third cluster, with half-day of absence on Wednesdays. As for other apartments, there are exceptions for profiles on specific days. For the sake of rationalisation, the described cluster distribution is the one retained for modeling. Selected occupancy DP distributions for building energy modeling are summarized in Table 3.5.

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

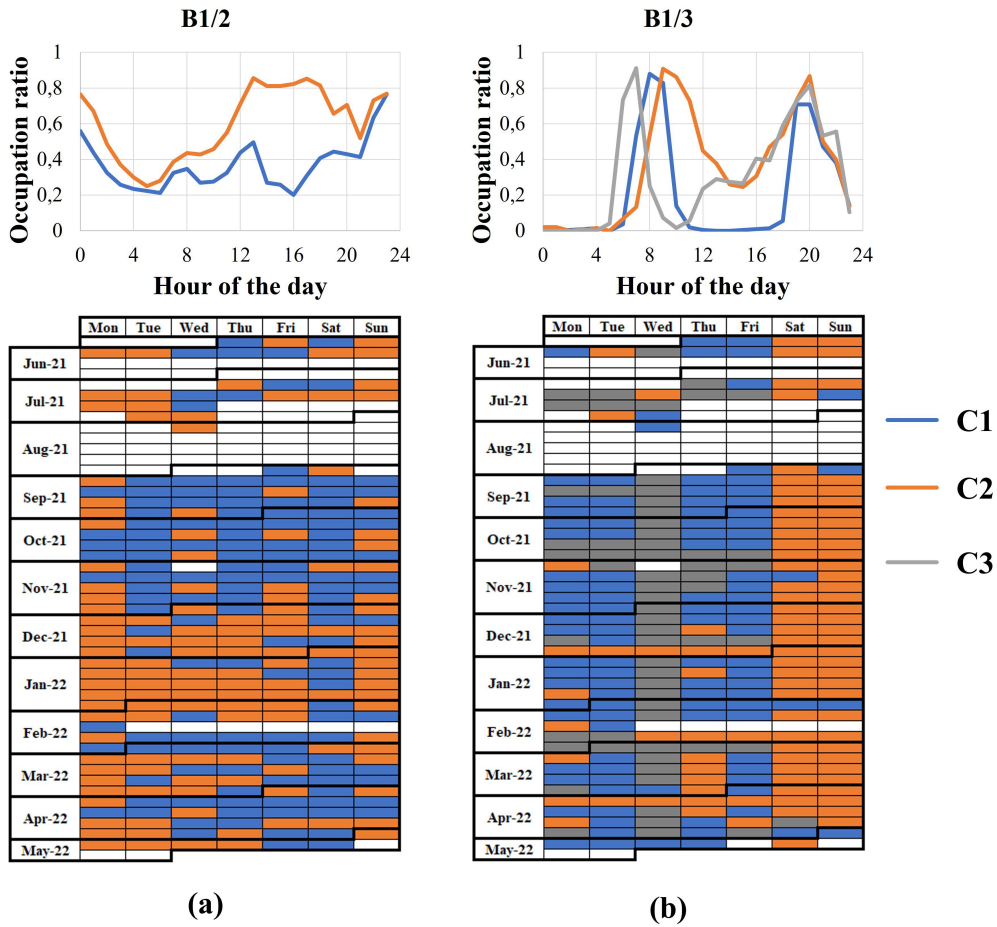


Figure 3.9: Mean clustered occupancy profiles and calendar mapping for B1/2 (a) and B1/3 (b).

Table 3.5: Cluster summary for occupancy patterns in instrumented housings.

	Cluster 1	Cluster 2	Cluster 3
B1/2	Worked days: Tuesday to Saturday	Days off: Monday and Sunday, period from January to February	/
B1/3	Worked days 1: Monday, Tuesday, Thursday, Friday	Worked days 2: Wednesday	Days off: Saturday, Sunday
B2/0	Worked days: four workdays a week from Monday to Friday	Days off: weekends, one randomly selected day from Monday to Friday, period from January to May	/
B2/5	All days	Not considered	/
B3/0	Not considered	All days	/
B3/2	Not considered	All days	/

3.3 Data analyses for building energy modeling enhancement and calibration

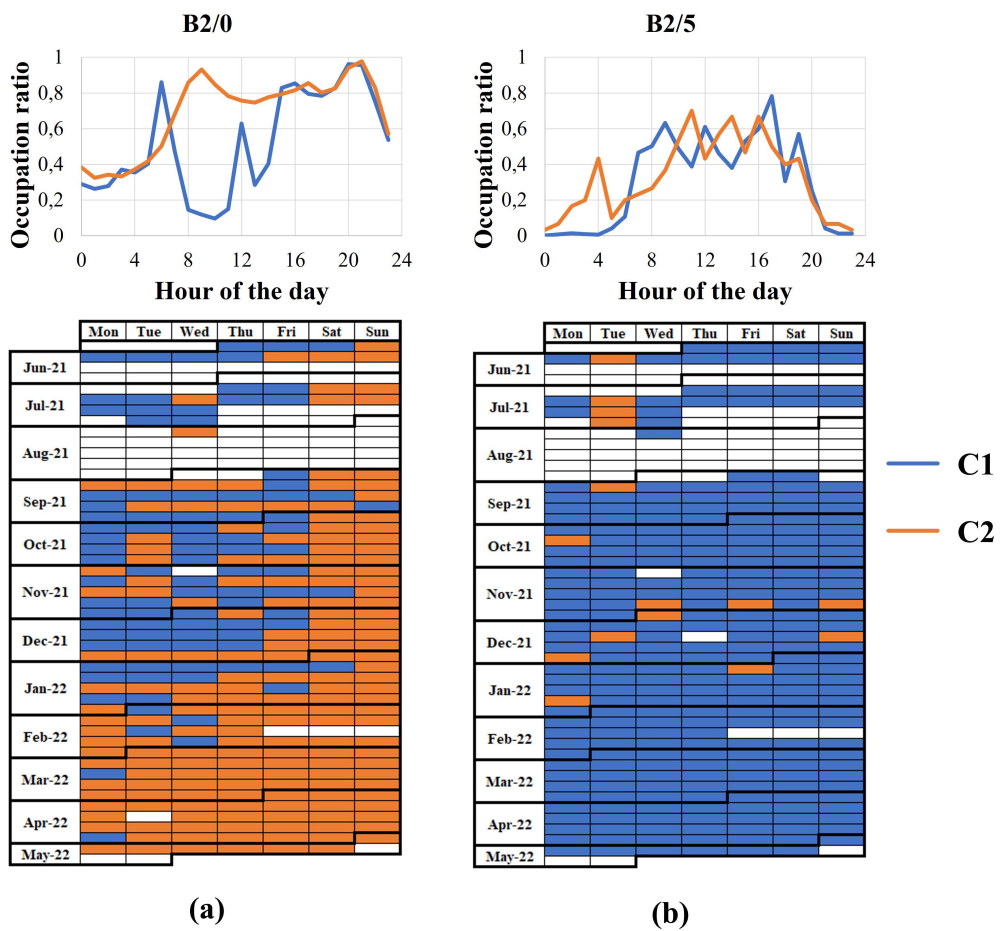


Figure 3.10: Mean clustered occupancy profiles and calendar mapping for B2/0 (a) and B2/5 (b).

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

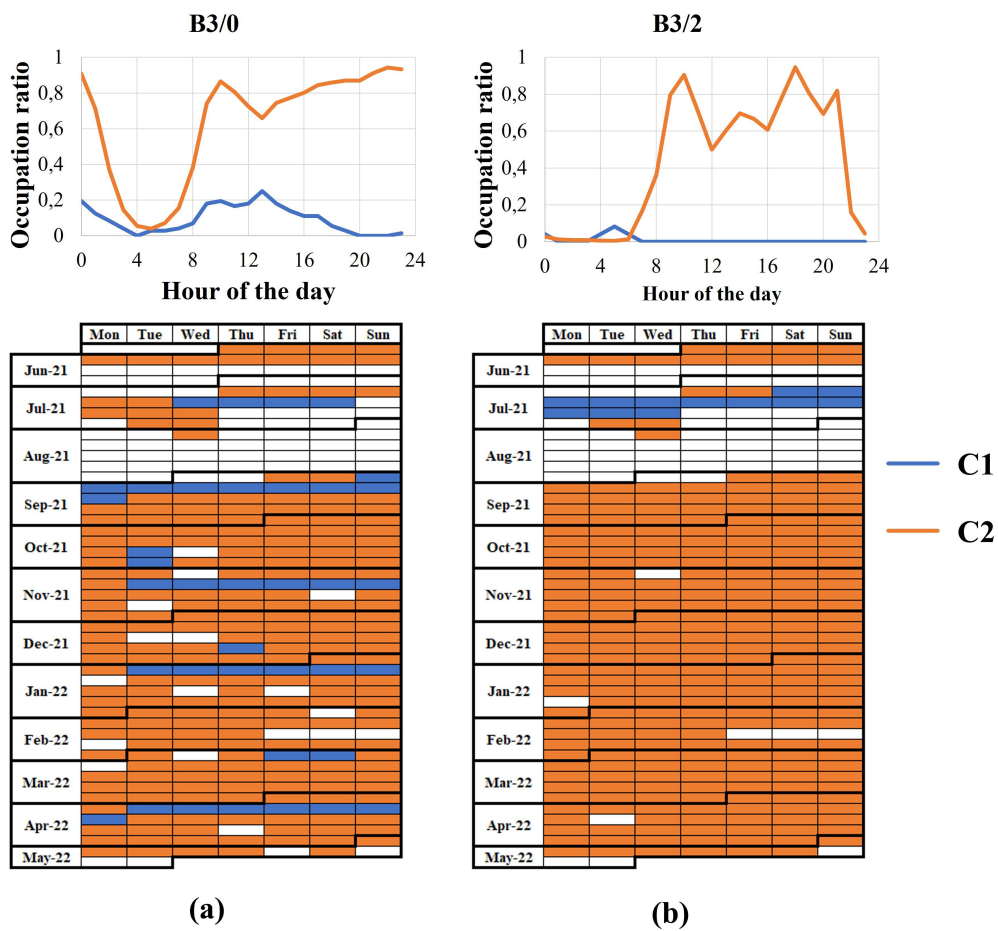


Figure 3.11: Mean clustered occupancy profiles and calendar mapping for B3/0 (a) and B3/2 (b).

3.3.2.2.1 Comparison with regulatory scenarios and correction

Clustering results highlight the diversity of occupancy profiles, depending on the number of occupants and professional activity. Over the six instrumented apartments, although they share common trends, none have the same occupancy profiles. Standard occupancy profiles, illustrated in Figure 3.12, are quite different from patterns extracted from collected data. Hence, there is a significant gain of knowledge in a more detailed characterization of occupancy profiles.

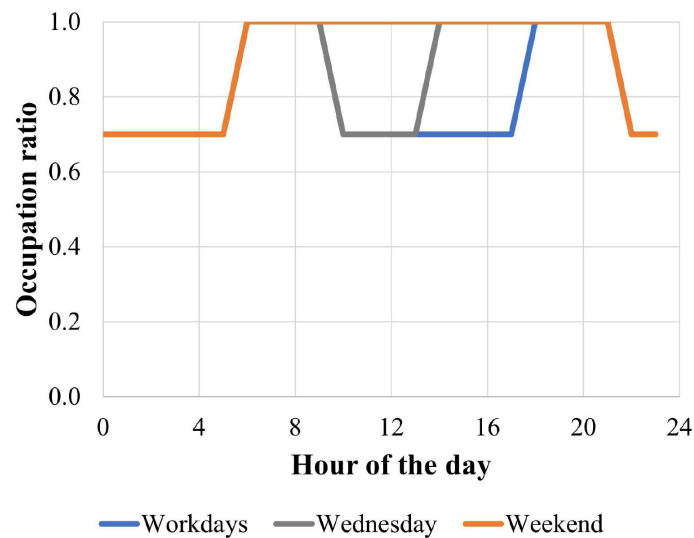


Figure 3.12: Occupancy profiles from Th-BCE calculations in Pléaides software.

Since measurements are performed in the living rooms, there is limited or no occupancy detected at night. However, it does not depict the real occupancy in instrumented housings. Furthermore, internal heat gains at night, when occupants are asleep, are lower than the standard value along the day (90 W) which is pictured in Figure 3.12. Therefore, a correction is applied over night-time on extracted profiles time. We assume that all residents are in the apartment at night and sleeping, since there is no or little presence detected in the living room. The occupancy rate is set to 70%, to match the dissipated heat level for sleeping adults of 63 W [15]. This correction is applied from midnight (or the latest activity peak in the early morning) to the first activity peak over 70% in the morning, and from the latest activity peak of the evening to midnight (Figure 3.13).

Finally, since it is not possible to count the number of people in the apartment but only to detect their presence, we assume that the apartment is occupied by all known occupants when a presence is detected by the sensor.

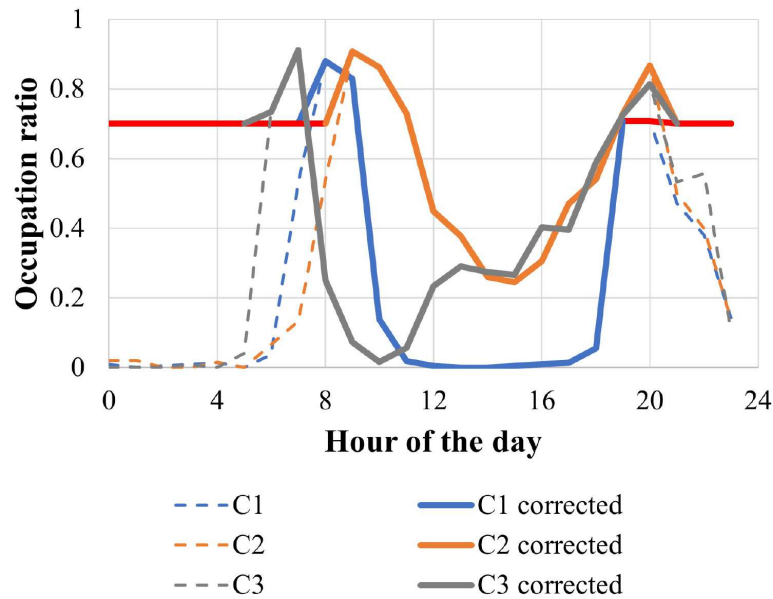


Figure 3.13: Initial and corrected occupancy profiles for night time in B1/3 – Dotted lines in the morning and the evening show the initial occupancy profiles, red lines figure the applied correction.

3.3.2.3 Electric dissipated power

Electric dissipated power is deducted from electric power demand at apartment-scale, measured on smart electricity meters. Results highlight an overall trend with two-cluster distributions. The first cluster figures a main group of DP and the second cluster highlights outlying profiles. For instance, in B1/2, the cluster 1 shows the main dissipated power pattern, while the cluster 2 groups two outliers (Figure 3.14). Although the range of power, especially the peak demand can significantly vary in amplitude, typical profiles share a similar trend over the day. Also, outliers are very different from the other DP. Similar results are observed for B2/0, B2/2, B3/0 and B3/2, and are described in Appendix C.4 along with calendar mapping. B2/1 shows comparable results, although highlighted profiles are not outliers but dissipated power profiles without activity around noon (Figure 3.15). B1/3 is different. A cluster gathers all profiles from the last two days of September 2021 to the middle of October 2021. The second cluster groups all other profiles over the observation period.

Additional tests are run after the first clustering results on B1/2, B2/0, B2/2, B3/0 and B3/2. Identified outliers are removed – since they are not considered as typical operation profiles – to try and identify additional typical patterns within the remaining

3.3 Data analyses for building energy modeling enhancement and calibration

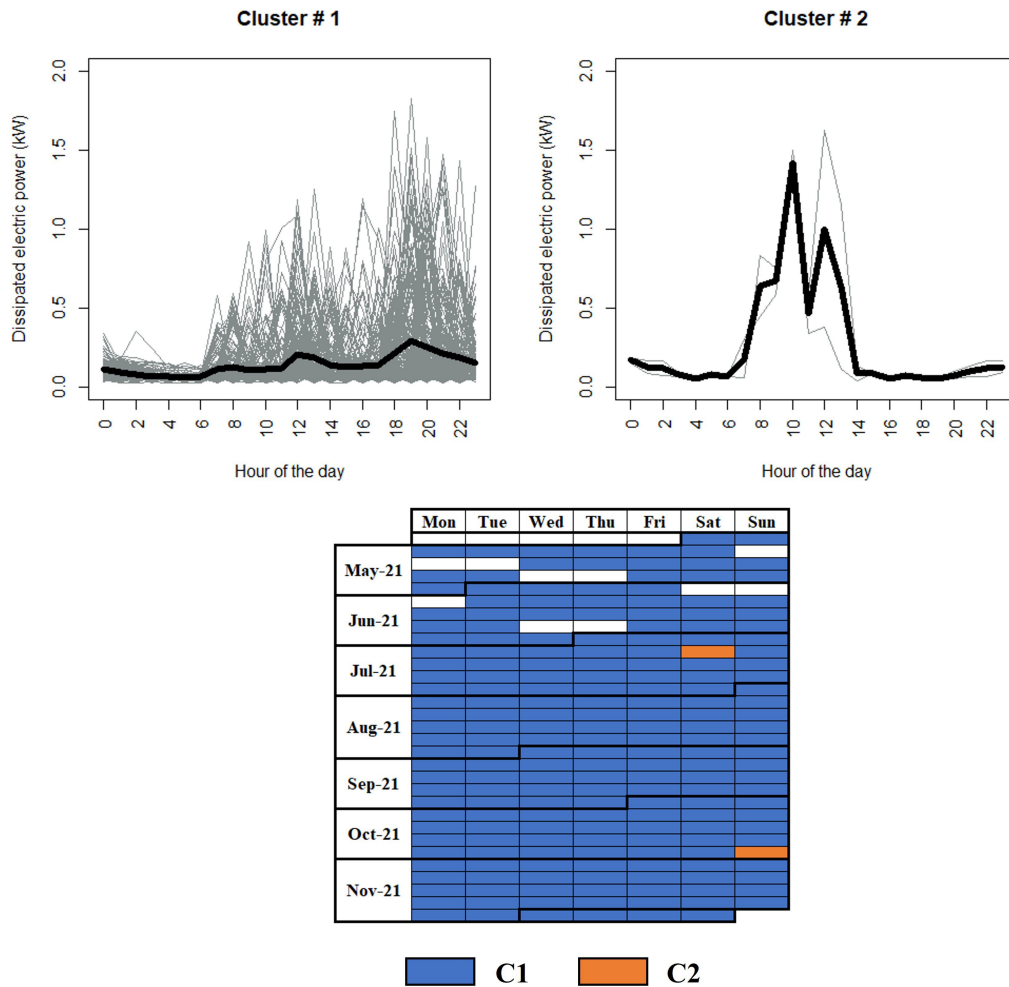


Figure 3.14: Clustering results for electric dissipated power in B1/2 – the mean profile of each cluster is shown with a black bold curve.

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

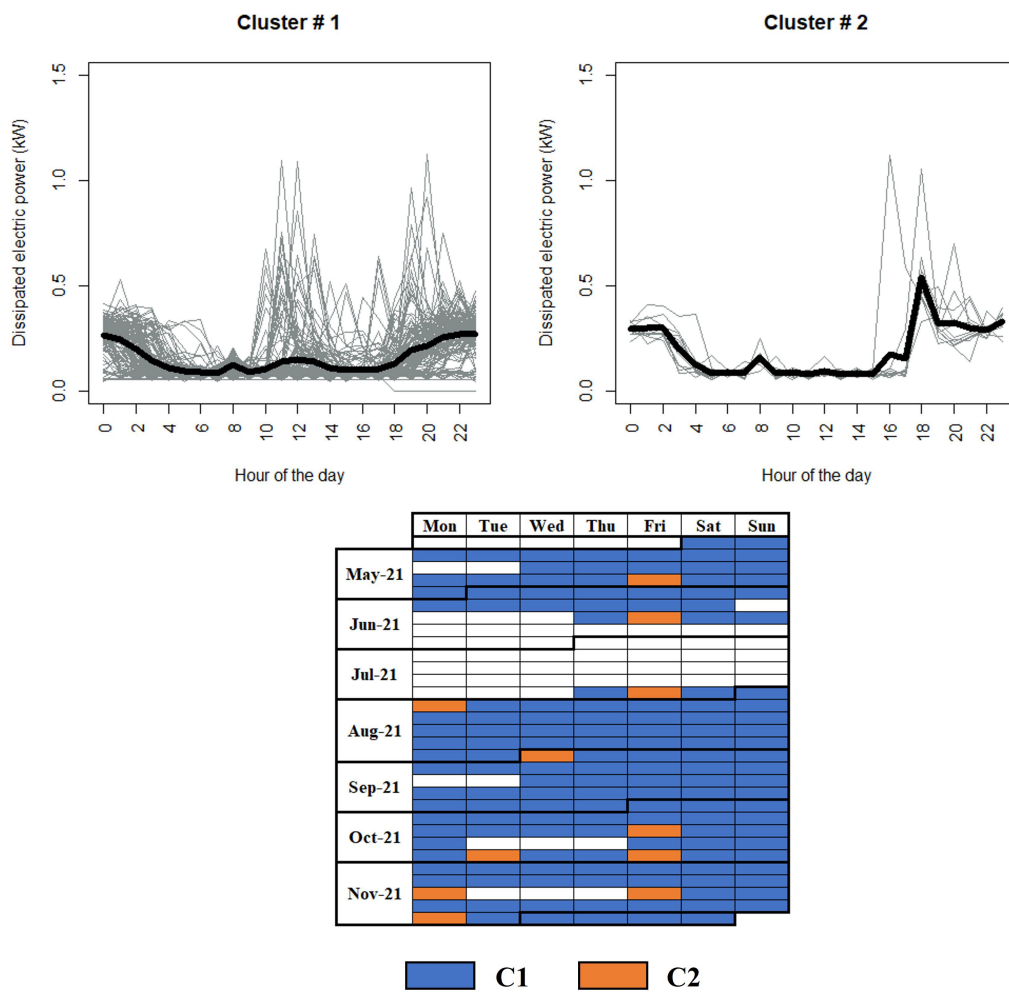


Figure 3.15: Clustering results for electric dissipated power in B2/1.

3.3 Data analyses for building energy modeling enhancement and calibration

profiles. However, the second round of clustering tests did not offer conclusive results. Details are provided in Appendix C.4. Results concluded that the clustering process either kept on highlighting outlying profiles that were not initially identified (in B2/0 and B3/0) or separated profiles that did not exhibit obvious differences with other clusters (in B1/2, B2/2 and B3/2). This stresses the limits of the classification. It would not make sense to keep on going with clustering and "force" a process that is already slightly biased by the analysis of results from the user. Therefore, the selected profile distribution is those of the first clustering round, summarized in Table 3.6.

Table 3.6: Cluster summary for dissipated electric power in instrumented housings.

	Cluster 1	Cluster 2
B1/2	All days	Not considered (2 outliers)
B1/3	All days but...	... period from last two days of September to middle of October
B2/0	All days	Not considered (1 outlier)
B2/1	All days	Not considered (similar profiles to cluster 1)
B2/2	All days	Not considered (1 outlier)
B3/0	All days	Not considered (5 outliers)
B3/2	All days	Not considered (2 outliers)

3.3.2.4 Window opening

Window opening detection is performed on all windows and glazed doors in instrumented apartments, with an average of four sensors for each housing. The details of installed and operating sensors is summarized in Table 3.7.

Table 3.7: Summary of sensors for window opening detection.

	Living room	Kitchen	Bedroom 1	Bedroom 2	Bedroom 3	TOTAL
B1/2	1	1	1	1	/	4
B1/3	2	1	1	/	/	4
B2/0	1	1	1	1	/	4
B2/1	1	1	1	/	/	3
B2/2	1	1	1	/	/	3
B2/5	N.A.	1	1	/	/	4
B3/0	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
B3/2	N.A.	1	1	1	/	4

Clustering is performed for each available sensor over an observation timeframe of one year. For the sake of legibility, all detailed results with DP clustering and calendar

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

mapping are provided in Appendix C.5, along with the summary of clustering distribution.

Among the different housings, there is a large variety of different opening profiles. These findings are particularly relevant, knowing that the default setting in building energy models is to consider that the windows are always closed. The main trend is the difference between the summer months and the rest of the year. The opening DP of the living room window in B2/2 in Figure 3.16 illustrates this pattern: in summer, the window is opened between forty and fifty minutes every hour, while in winter, the window is opened less than fifteen minutes per hour. The summer-winter contrast is highlighted in B1/2 (kitchen, living room, bedroom 1 and 2), B1/3 (kitchen, living room, bedroom 1 and 2), B2/0 (kitchen, living room), B2/1 (living room, bedroom), B2/2 (living room, bedroom), B2/5 (kitchen, bedroom), and B3/2 (kitchen, bedroom 1). Therefore, most sensors in most apartments show a much longer window opening in summer. This can be interpreted in two ways. Data were acquired starting from the end of the retrofit campaign in the studied buildings. Thermal insulation retrofit was already completed and summer thermal comfort should be close to optimal – or at least better than for the past summers before the retrofit. Therefore, it could mean that the thermal insulation and/or the ventilation would not be adapted to ensure sufficient comfort in summer. However, knowing that most residents have been living in apartments for many years prior to the retrofit, it could also be related to their habits of opening the windows to renew the air.

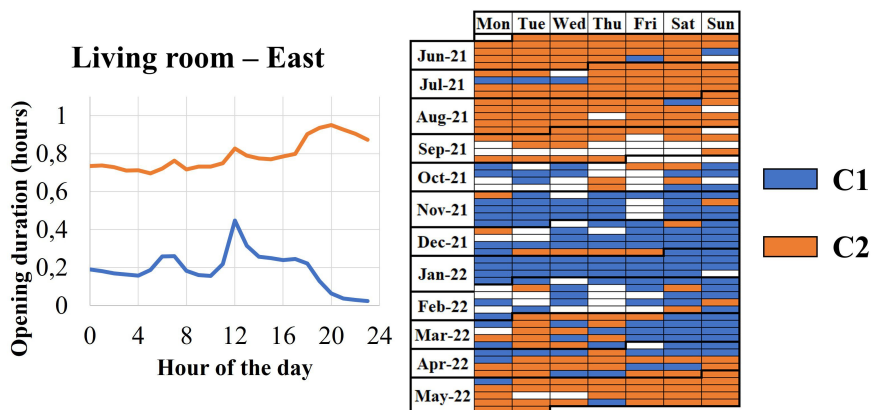


Figure 3.16: Window opening profiles and calendar mapping in the living room of B2/2 – The averaged window opening DP illustrate between summer (in orange) and winter (in blue).

Most sensors highlight a two-cluster (summer/winter) distribution of data but there

3.3 Data analyses for building energy modeling enhancement and calibration

are also a few exceptions, with up to four clusters. In most cases, they identify outlying profiles or an intermediate profile when the window is mostly closed in the first case and mostly opened in the second case. Some sensors also highlight a very specific activity, as for the first bedroom in B2/0: the window of this room is always opened for pets to come in and out (Figure 3.17).

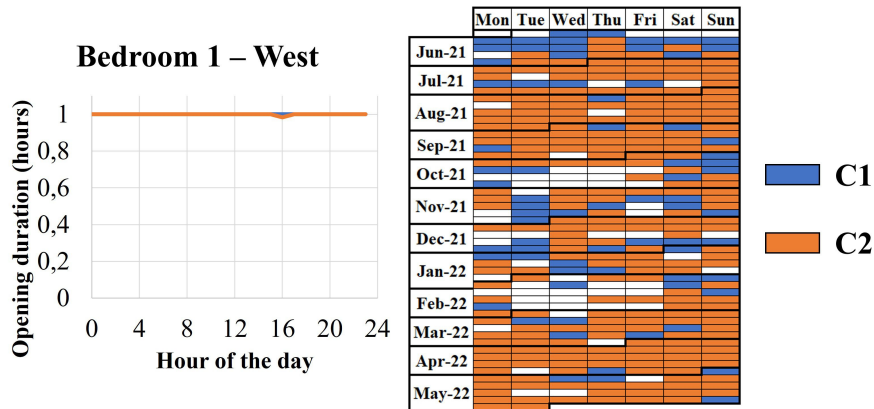


Figure 3.17: Window opening profiles and calendar mapping in the bedroom 1 of B2/0 – Regardless of the time of the year, the window is always opened.

Three other parameters may also impact on opening trends: the orientation, the type of room and the specificity of occupants in an apartment. Orientation does not seem to have a significant impact on window opening, apart from the living room of B2/1. The South-East orientation of the apartment may be related to the fact windows are opened later in the day during the summer (Figure 3.18). However, it could also be related to the professional activity of the resident (specific working hours, for instance), that cannot be checked due to the absence of occupancy data for this apartment. Taking into account the function of the rooms in instrumented housings does not highlight any specific trend. However, different apartment show a similar trend for the different rooms. B1/3 is a relevant example. Apart from the kitchen, all three other rooms (living room, bedroom 1 and 2) have very similar profiles in summer – windows opened at night and not in the day – and for the rest of the year – windows are only opened for a few minutes every day. Figure 3.19 illustrates the window opening patterns in the bedroom of B1/3 with East exposure.

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

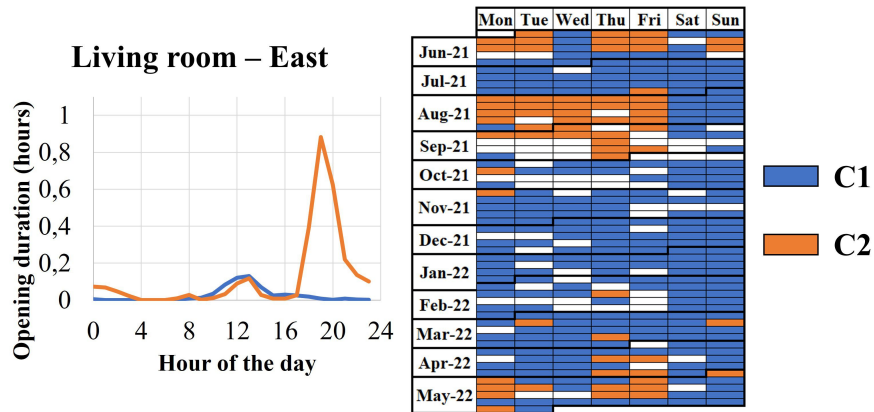


Figure 3.18: Window opening profiles and calendar mapping in the living room of B2/1 – The opening peak at the end of the day in summer (in orange) may be related to the orientation of the room or the working hours of the occupant.

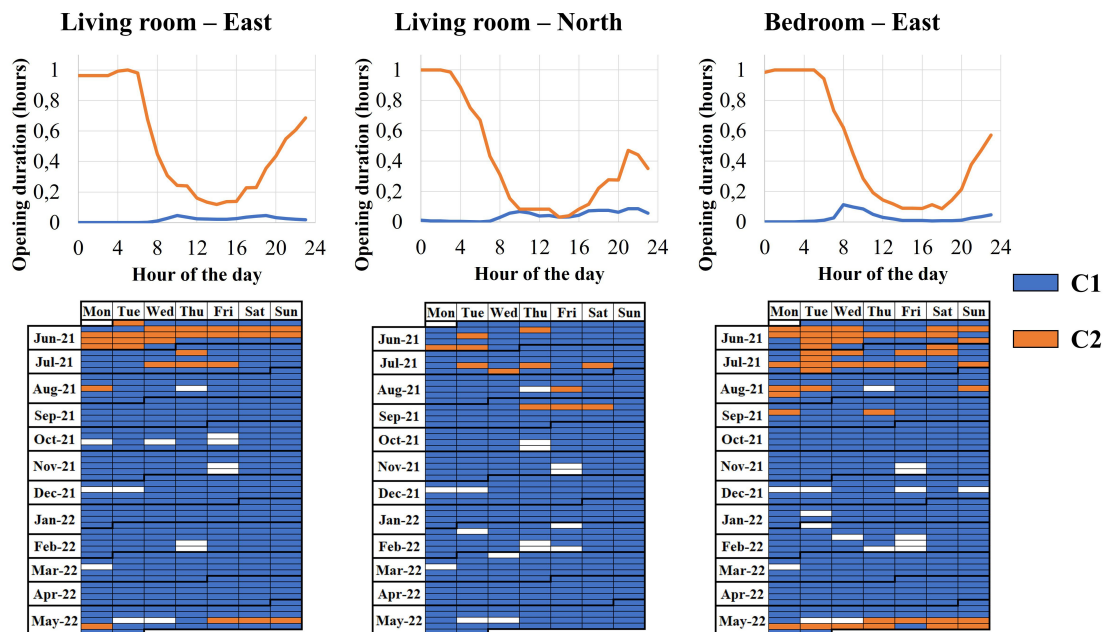


Figure 3.19: Window opening profiles and calendar mapping in the bedroom of B1/3 – These window opening DP depict the housing-specific patterns.

3.3.2.5 Domestic hot water

Domestic hot water usage (DHW) is the fifth selected parameter and that describes a direct effect of occupants' behavior on building energy consumption. As detailed in Section 3.2, DHW data are not directly analyzed. DHW consumption, in terms of hot water volume, is assessed through temperature measurements with one-minute temporal granularity, and combined with statistical data on DHW use from [7]. Clustering is performed on temperature profiles. The mean profile of each cluster is modified to obtain a volume consumption.

Over the four apartments dispatching sufficient data for analyses, B1/2 (Figure 3.20(a)) and B1/3 (Figure 3.20(b)) exhibit a main daily DHW use profile (cluster 1) while other clusters show no specific patterns and gather outlying profiles (dotted lines for clusters 2, 3 and 4 in B1/3, and for cluster 2 in B1/2). DHW profiles for B2/0 (Figure 3.20(c)) and B3/0 (Figure 3.20(d)) highlight two main profiles, that do not show much difference. For B2/0, the mean profile from cluster two describes a higher DHW consumption in the morning, found on Sundays and very occasionally on some weekdays. The mean profile from cluster 1 shows a higher activity in the evening. For B3/0, the pattern is similar with a higher DHW consumption over the day, specifically at noon, in cluster two and compared with cluster one. However, there is no well-defined patterns for the latter apartments, which makes the classification analysis more difficult to summarize for building energy modeling inputs. In a nutshell, for B3/0, two profiles from the second cluster are randomly distributed in each week. The remaining days exhibit a lower DHW use from cluster 1. Detailed calendar mapping for the analysis of the distribution of clusters are available in Appendix C.6.

Comparing the different apartments, we observe two distinctive DHW usage patterns. B1/2 (Figure 3.20(a), blue line) and B1/3 (Figure 3.20(b), blue line) highlight a typical DP that can be related to the professional activity of occupants – both have employed occupants. B1/2 shows three peak usages, in the morning, at noon and in the evening. For B1/3, the peaks are in the morning and in the evening with some less important use around noon. B2/0 and B3/0 exhibit shared trends in the day in terms of DHW usage distribution, although there is a clear sign of activity late at night for B3/0. Both housings have retired occupants. It can also be noted that among the four studied apartments, B1/2 has the highest DHW usage peaks, up to fourteen liters per hour. Although it has two occupants as well, B3/0 shows smaller activity peaks. As for B1/3 and B2/0, it can be interpreted in two different ways: either a larger diversity of profiles in a cluster, or similar profiles but with more activity during the day than for B1/2. The summary of

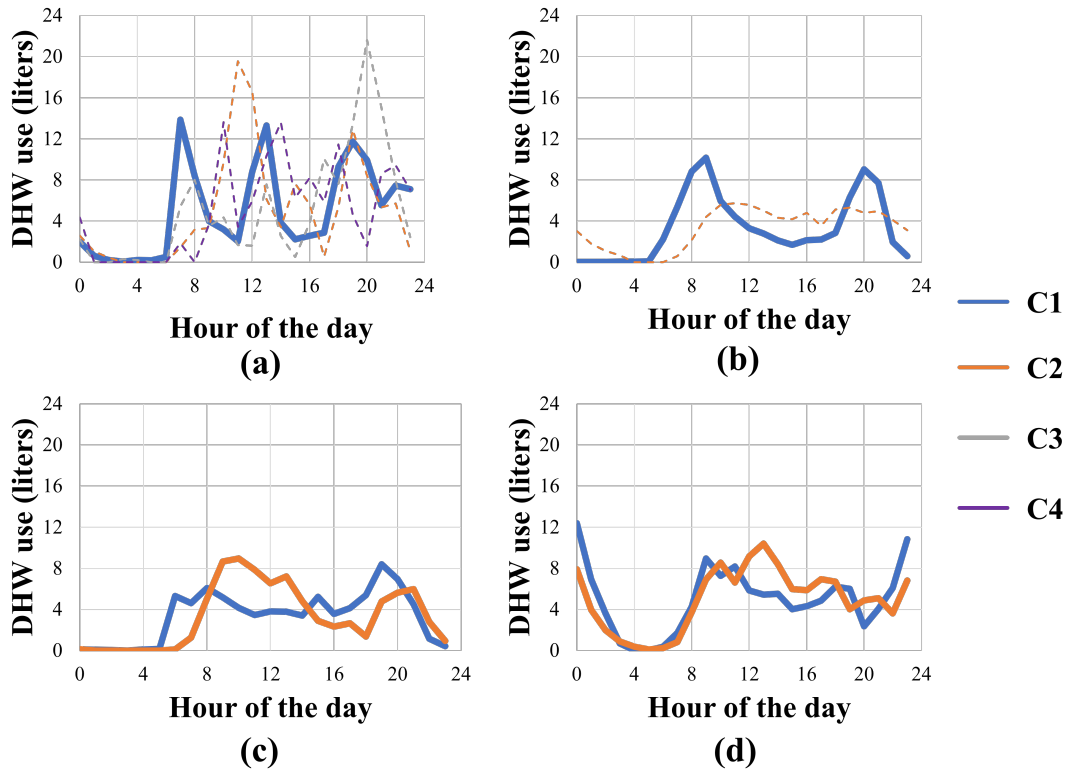


Figure 3.20: Mean profiles for DHW use in B1/2 (a), B1/3 (b), B2/0 (c) and B3/0 (d) – B1/2 and B1/3 exhibit one main DP (in blue) while other profiles are outliers – B2/0 and B3/0 have two DP that are quite similar to each other.

clustering for DHW can be found in Table 3.8.

The diversity of profiles compared with regulatory scenarios extracted from Pléaides also show the interest in field data collection. A first issue comes from the fact that regulatory scenarios are not designed depending on the number of people in the apartment in contrast with our approach, but on the size of apartments. In our building case study, some social housings (such as B2/0 and B3/0) are larger than what would be expected – one resident for a two-bedroom apartment (B2/0) and two residents for a three bedroom apartment (B3/0). Usage patterns are also quite different from a regulatory scenario. As illustrated in Figure 3.21, peak usage for B1/2 happen at different times and have a different amplitudes. For other apartment sizes, the shape of profiles from the energy modeling software remains the same but the hourly consumption changes depending on the size of the apartment. Finally, the average daily consumption from regulatory scenarios differs from data extracted from the study of ADEME (Table 3.9) and related to the water temperature of DHW (40°C in the present case, and 60°C in Pléaides software).

3.3 Data analyses for building energy modeling enhancement and calibration

Table 3.8: Cluster summary for DHW in instrumented housings.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
B1/2	All days	Not considered (9 outliers)	Not considered (14 outliers)	Not considered (6 outliers)
B1/3	All days	Not considered (7 outliers)	/	/
B2/0	Monday to Saturday	Sunday	/	/
B3/0	All days ex- cept for	Two days per week, randomly dis- tributed	/	/

Table 3.9: Comparison of DHW daily averaged consumption from ADEME and Pléaides software.

Housing type	ADEME Data [7] for social housings (liters, T=40°C)	Pléaides scenarios modeling software (liters, T=60°C)
1 living room/bed- room	75 ±60	39 ±12
1 bedroom	80 ±65	55 ±17
2 bedrooms	110 ±80	75 ±23
3 bedrooms	145 ±100	95 ±29

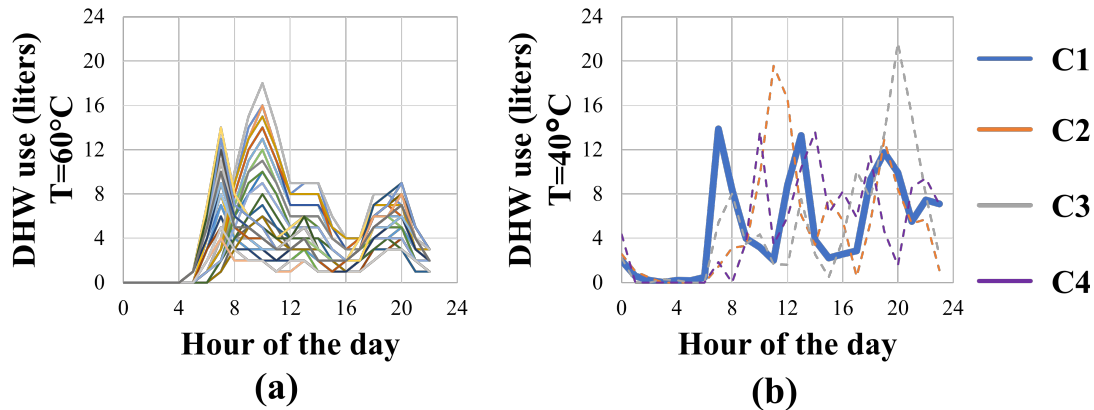


Figure 3.21: Comparison of DHW consumption daily profiles from Pléaides software for a two-bedroom apartment (a) and B1/2 (b) – The comparison highlights the difference between measured field data and regulatory scenarios in terms of peak event timing and amplitude.

Conclusions

This chapter detailed the processing and analysis of data collected through our wireless sensor network described in Chapter 2. The goal was to acquire a detailed knowledge of the operation of the case study, to refine building energy modeling assumptions.

After an overall picture of the quality of collected data, we focused on specific measurements that are among the key parameters to study building energy consumption and to implement building energy models: IAT, occupancy, dissipated electric power, window opening and DHW consumption. To improve data availability, reconstruction through mean imputation was performed. Up to 56% of additional complete daily profiles were obtained. We also implemented hourly data aggregation to match the requirements for energy modeling scenarios and it benefited the data completion.

Two types of data analyses were performed: descriptive statistical analyses and daily profile clustering. The former inquired on field data distribution features and the latter provided a classification of daily profiles to identify typical patterns.

Indoor air temperature was significantly higher than expected, as it ranged between of 21.4°C and 24.9°C in average over the heating period, with at least 97.5% of acquired data over 19°C in instrumented housings. Also, temperature profiles were constant over day. This highlighted a significant potential in heating energy savings with two solutions to focus on: the decrease of indoor air temperature and the implementation of water laws for a better fitted heating energy management. These results also lead to think that

thermal insulation following the building renovation matched expected targets.

Occupancy related to two parameters: the type of day and the professional activity of occupants. Three patterns were extracted: i) a clear weekday-weekend distribution, ii) a similar distribution of profiles with one additional profile figuring partial activity during the day, and iii) a steady housing occupancy for unemployed and retired occupants.

Electric dissipated power and DHW also provided a diversity of daily profiles. However, most apartments had a single profile to summarize both parameters, which lead to think that electricity and DHW use are steady throughout the year. Finally, window opening emphasized the major difference in occupants' habits between summertime and the rest of the year. In all apartments and most rooms, windows were largely opened in summer, while opening duration was much smaller otherwise. It raised the question of the origin of this pattern: it could either be an issue with thermal comfort in summer despite the recent retrofit, or behaviors inherited from cultural background or from living in the buildings prior to retrofit actions, when the thermal insulation was poor. It also struck that window opening patterns were apartment-specific: similar patterns could be found in different rooms within a same apartment. However, opening patterns significantly differed when apartments were compared. Finally, window orientation and room function seemed to have very little or no impact on window opening duration.

Overall, analyses results, showed that the sensor network succeeded in providing an overall yet detailed picture of building operation. With only a few sensors for the characterization of IEQ, it was possible to observe significant issues, such as with IAT in housings, that could lead to proposals for the optimization of heating management and potentially significant energy savings. Clustering results struck by the diversity of extracted patterns. Findings emphasized the differences with regulatory scenarios from Th-BCE calculations, in terms of IAT, occupancy, DHW use, dissipated power and window opening operation profiles. Despite the difficulties to collect such an amount of data and the complex analysis process required, these differences underline the relevance of detailed data collection and analysis. For specific case studies, regulatory scenarios may be too generic to be representative of building operation and lead to significant energy performance gaps.

The presented analysis also showed that several aspects in the data collection could be improved. Data quality assessment depicted large amounts of missing data due to several technical onsite difficulties. Data reconstruction and aggregation balanced part of the loss, but this process could not recover the full length of missing information. These pre-processing steps also induce uncertainties because they rebuild part of the data, despite

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

the fact it is obviously much more beneficial to have some slightly uncertain data than not enough data to perform any analysis.

Several assumptions may have an important effect on analyses results and further building energy modeling. Assumptions are related to the instrumentation solution and installation. IAT measurements may be disturbed by the surrounding environment and setup. Occupancy is assessed from presence detection. It lacks a direct occupant counting and requires assumptions on the number of people in the housing at a given moment for apartments with several inhabitants. DHW use is the most uncertain data. The small data acquisition time-step clearly depicts DHW use. However, DHW consumption is deduced from statistical studies and is not directly measured onsite. Window opening and electric dissipated power do not require many modifications. However, these parameters exhibit significant amount of missing data over observation periods.

Clustering results mostly follow a day-type-based trend. However, it requires advanced knowledge on the buildings regarding technical characteristics, operation and occupants. Part of the knowledge comes from information extracted from the retrofit portfolio and acquired through onsite visits along the project. Considering the amount of collected data and the diversity of parameters measured by the sensor network, there is a significant remaining potential for data cross-analyses that could help refining clustering results. Specifically, to improve occupancy profiles and deepen the understanding of energy-driving behaviors, cross-analyses of occupancy, CO₂ concentration, window opening and connected plug data, for instance, seems promising. The latter gives a very detailed overview of electric appliances in housings. It could lead to an occupant-by-occupant behavior characterization to accurately assess their presence and offer insights on energy management optimization (with a warning on privacy respect). Although, a significant challenge might still lie in the availability of data across apartments. As discussed above, not all installed sensors are fully functional.

Finally, in the present chapter, IAT, dissipated electric power, occupancy, window opening and DHW consumption have been processed and analyzed separately. This strategy relates to the first objective of the data analyses, to extract typical patterns out of this set of parameters and to replace assumptions in building energy models. Indeed, building energy models take these inputs into account separately, while it is necessary to specify that several of them are expected to be related in a realistic building operation. A correlation between occupancy and dissipated electric power, window opening and DHW use is quite foreseeable as, apart from some scheduled appliance triggering, all latter three parameters are highly related to the behavior of occupants. The detailed correlation be-

tween these inputs, as well as other parameters that could not be investigated in our study, should be further explored as part of the future enhancement of data analyses.

Bibliography

- [1] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, feb 2018. 102
- [2] R. Kumar, R. K. Aggarwal, and J. D. Sharma, “Energy analysis of a building using artificial neural network: A review,” *Energy and Buildings*, vol. 65, pp. 352–358, oct 2013. 102
- [3] Y. Allab, M. Pellegrino, X. Guo, E. Nefzaoui, and A. Kindinis, “Energy and comfort assessment in educational building - Case study in a French university campus,” *Energy and Buildings*, vol. 143, pp. 202–219, may 2017. 102, 116
- [4] M. Bourdeau, X. Guo, and E. Nefzaoui, “Buildings energy consumption generation gap: A post-occupancy assessment in a case study of three higher education buildings,” *Energy and Buildings*, vol. 159, 2018. 102
- [5] M. Bourdeau, P. Basset, S. Beauchêne, D. Da Silva, T. Guiot, D. Werner, and E. Nefzaoui, “Classification of daily electric load profiles of non-residential buildings,” *Energy & Buildings*, vol. 233, p. 110670, 2021. 102, 112, 113
- [6] P. J. García-Laencina, J. L. Sancho-Gómez, and A. R. Figueiras-Vidal, “Pattern classification with missing data: A review,” *Neural Computing and Applications*, vol. 19, pp. 263–282, sep 2010. 105
- [7] ADEME, “Guide technique — Les besoins d’eau chaude sanitaire en habitat individuel et collectif,” tech. rep., 2017. xxvii, 108, 109, 133, 135
- [8] S. Yilmaz, J. Chambers, and M. K. Patel, “Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management,” *Energy*, vol. 180, pp. 665–677, aug 2019. 113
- [9] A. Satre-Meloy, M. Diakonova, and P. Grünwald, “Cluster analysis and prediction of residential peak demand profiles using occupant activity data,” *Applied Energy*, vol. 260, feb 2020. 113
- [10] J. Yang, C. Ning, C. Deb, F. Zhang, D. Cheong, S. E. Lee, C. Sekhar, and K. W. Tham, “k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement,” *Energy and Buildings*, vol. 146, pp. 27–37, jul 2017. 113

CHAPTER 3 – PROCESSING AND ANALYSES OF FIELD DATA FOR BUILDING ENERGY MODELING ENHANCEMENT

- [11] A. Rajabi, M. Eskandari, M. Jabbari Ghadi, S. Ghavidel, L. Li, J. Zhang, and P. Siano, “A pattern recognition methodology for analyzing residential customers load data and targeting demand response applications,” *Energy and Buildings*, vol. 203, 2019. 113
- [12] K. A. Choksi, S. Jain, and N. M. Pindoriya, “Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset,” *Sustainable Energy, Grids and Networks*, vol. 22, p. 100346, jun 2020. 113
- [13] F. Asdrubali, F. D’Alessandro, G. Baldinelli, and F. Bianchi, “Evaluating in situ thermal transmittance of green buildings masonries — A case study,” *Case Studies in Construction Materials*, vol. 1, pp. 53–59, 2014. 116
- [14] K. Okamoto-Mizuno and K. Mizuno, “Effects of thermal environment on sleep and circadian rhythm,” *Journal of Physiological Anthropology*, vol. 31, no. 1, p. 14, 2012. 117
- [15] Centre Scientifique et Technique du Bâtiment, “ANNEXE II : Méthode de calcul détaillée « Th-BCE 2020 »,” 2020. 125

Chapter 4

Energy modeling, calibration and performance gap

The concluding chapter of this manuscript describes the building energy modeling and calibration processes after the energy retrofit of the considered buildings, based on the analysis of field data collected through our wireless sensor network. The first part of the chapter introduces the used numerical tools and the overall modeling process. It is followed by the description of the characteristics of energy models, including the envelope, the energy systems and the operation scenarios. The analysis of simulation results is divided into two parts. First, non-calibrated simulation results and the initial performance gap are analyzed. Then, the results of the calibration process lead to a discussion on the remaining energy performance gap and on further strategies to close it.

Introduction

This fourth and final chapter presents the implementation of the energy models of the considered buildings after their retrofit for the investigation of the energy performance gap (PG). In Chapter 1, we selected the most fitted building modeling approach in terms of modeling details and interpretability, namely physics-based multizone models. Several strategies were discussed as well to target the PG. We decided to focus on the improvement of the calibration of BEM, using field-collected building operation data, to enhance the knowledge on building post-occupancy operation. Data collection was performed using our wireless sensor network described in Chapter 2. From the large amount of available data, a set of parameters – including IAT, occupancy, dissipated electric power, window opening and DHW usage – was selected to address the building energy consumption, the IEQ and the occupants' behavior. These parameters were analyzed in Chapter 3 to extract typical patterns that can be used in building energy models.

The study of the PG targets heating energy consumption. Heating in apartment buildings, specifically for buildings built before 1975 (then before the first French thermal regulation) as in the present case, account for 63% of their global energy consumption, followed by electrical appliances (17%), DHW (13%) and cooking (6%) [1]. As newly retrofitted buildings, the share of heating should be reduced to 54% of the averaged building energy consumption, but it is still the first and main energy end-use.

We present in the first section the tools, the general modeling process and the description of energy models. The second section discusses the results of building energy models. It starts with the analysis of heating energy metering results, used as the reference heating energy consumption of the instrumented buildings for the comparison with simulations. Then, non-calibrated BEM (NC-BEM) are implemented using information and data from documentation and regulatory operation scenarios. Simulation results are compared to measured heating energy consumption to assess the initial PG. NC-BEM simulations are complemented by sensitivity analyses to highlight the energy-driving input parameters in the models. In the second part of the section, we focus on the tuning of modeling input parameters based on the measured data, and selected through sensitivity analyses. Finally, we conclude on the performances of the proposed strategy to reduce the PG.

4.1 Used modeling methods

4.1.1 Used tools: software and documentation

4.1.1.1 Energy modeling software

Physics-based energy modeling with multizone technique was highlighted as the fittest approach to investigate the impact of energy retrofit actions, occupants' behaviors and the optimization of building operation.

There are many available software for building energy modeling as detailed in Appendix A.1. For this thesis, Pléaides software suite [2] (version 5.21.1.2) is used. The suite provides different modules depending on the purpose of the study: assessment of heating and cooling energy needs, dynamic energy simulations, regulatory studies for new and existing buildings, optimization, sensitivity and uncertainty analyses, life cycle analysis, and BIM (building information modeling). The versatility of applications is a significant selection criterion, along with the affordability, and a responsive and efficient customer service. An embedded library of building characteristics and building operation scenarios is also available. It provides standard modeling components, that are compare with measured data from our sensor network, as presented in Chapter 3.

4.1.1.2 Documentation

Available building information are scarce regarding the description of the existing considered case study, while they should be the first source of information to implement a BEM.

The description of the building envelope depends on floor plans and the thermal regulatory study (TRS) conducted prior to the retrofit. Some of the floor plans were missing and had to be extrapolated, since the architecture of the buildings is similar. A visual diagnosis was performed onsite with CAMEO SAS to complete the documentary resources. The most detailed description of retrofit actions and characteristics of modified elements in the buildings is provided by the retrofit project management portfolio. It includes specifications on the plumbing and heating, the ventilation, electrical diagrams, the indoor and outdoor windows and doors, and the thermal insulation. Additional knowledge is brought by the many visits in housings and buildings over the deployment of the sensor network.

Missing information and data are extracted directly from Pléaides software. The embedded library provides scenarios, building and equipment characteristics that would

be used for regulatory calculations and that can then be considered for the first part of the energy modeling process.

4.1.2 Process for building energy modeling, calibration and performance gap assessment

The process for building energy modeling, calibration and PG assessment is divided into three steps. It focuses on buildings after the end of the energy retrofit.

Step 1 – Non-calibrated building energy models (NC-BEM). Non-calibrated energy simulations depend on data and information extracted from the retrofit portfolio. It aims to describe the buildings as for a TRS. Characteristics of the building envelope should reflect on the reality of retrofit actions. Part of the features of the energy systems are found in the retrofit portfolio and the TRS as well. Operation scenarios include temperature set points, window opening, occupancy, dissipated power, shading, ventilation, lighting and DHW use. These are standard scenarios from Th-BCE (*Thermique besoin-consommation-été*) library embedded in Pléiades. Any missing information is set to standard parameters suggested by the modeling software.

The NC-BEM leads to two sub-steps illustrated in Figure 4.1:

- Step 1.1: a comparison of the simulation results with the measurements from heating energy consumption meters;
- Step 1.2: sensitivity analyses to highlight the most impacting parameters in the energy models.

The comparison between simulations and measurements underlines the initial PG and sets up the goals of the calibration. The sensitivity analysis selects the parameters to tune for the calibration.

Step 2 – Integration of field-collected data. Following the non-calibrated simulation results, data from the sensor network are integrated in Pléiades library to replace standard information. Data originate from results of the analyses detailed in Chapter 3. They target parameters related to building operation and underlined by the sensitivity analyses.

Step 3 – Calibration of BEM. Building energy models are calibrated based on the NC-BEM tuned with post-operation data. The calibration is performed manually [3] with level 4 calibration [4] – using data acquired through a several-month-long monitoring campaign. Simulation results are compared with the non-calibrated results and measurements from the sensor network. The impact of input tuning is assessed, and the

remaining uncertainties are investigated to explain a potential residual PG.

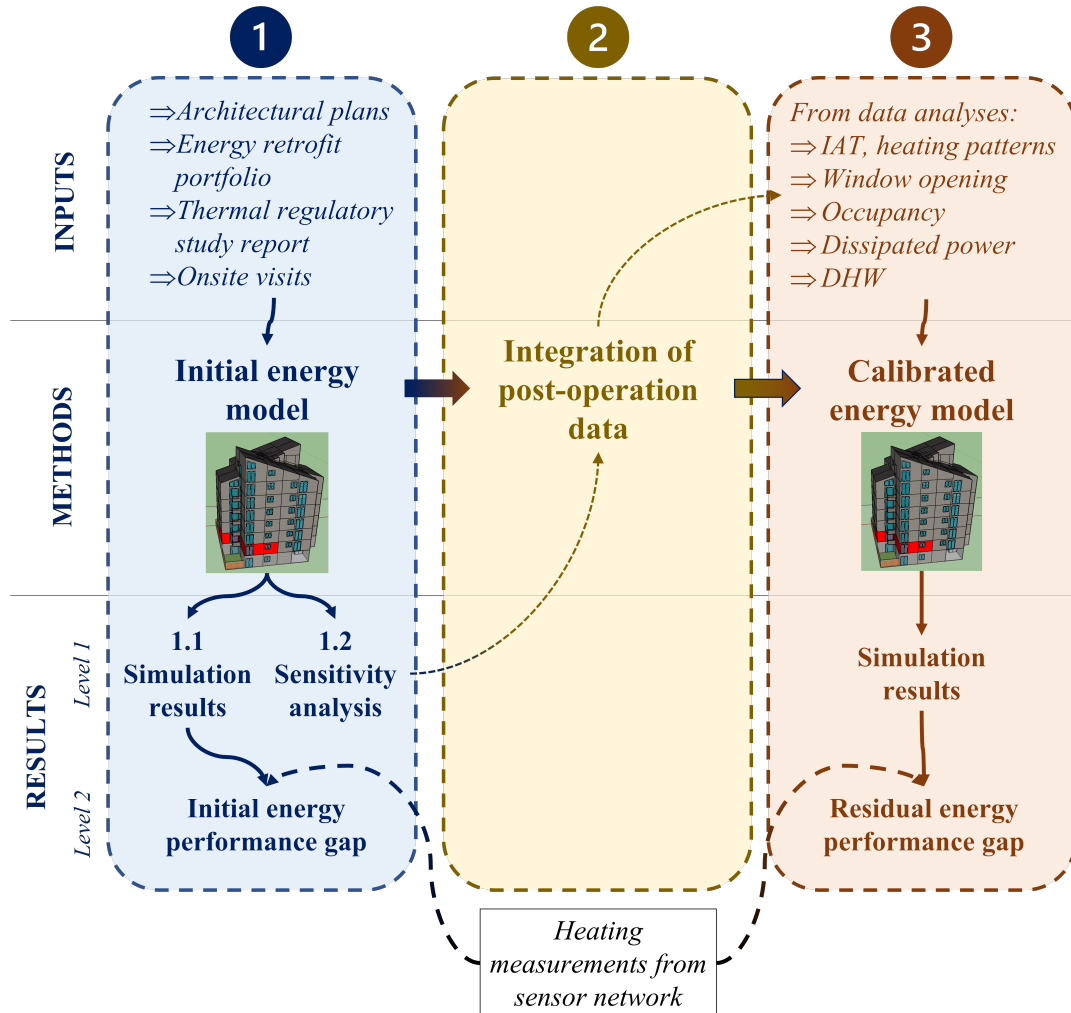


Figure 4.1: Three-step building energy modeling and calibration process with: Step 1 NC-BEM, Step 2 the integration of field-collected data, and Step 3 the calibration of energy models.

4.1.3 Sensitivity analysis

A sensitivity analysis identifies the energy-driving input parameters of a BEM. Regardless of the implemented technique, sensitivity analyses force the variation of input parameters to artificially test the weight of these inputs in the model. There are commonly two types of sensitivity analyses: local sensitivity analyses and global sensitivity analyses. In a nutshell, the former varies one parameter at a time around a set of given parameter values, and independently from other parameters; the latter varies one parameter at a

time and it also takes into account existing correlations with other input variables. A detailed review of sensitivity analysis techniques falls out of the scope of this thesis, but more details can be found in the work of Kristensen and Petersen [5], and in the work of Tian et al. [6].

In the present modeling process, sensitivity analyses are local "one-at-time" analyses. The selection of this technique is driven by its availability in the sensitivity analysis tool in Pléaides software, as well as its low computational cost and its ease to use. The list of considered parameters is the following:

- Materials and construction elements: thermal conductivity, density, specific heat;
- Walls: height, width, thickness;
- Windows, glazed doors and doors: solar factor, thermal conductivity, width, height;
- Thermal bridges: integrated, linear and window thermal bridges;
- Ventilation: ventilation scenarios, permeability;
- Operation scenarios: IAT, occupancy, dissipated power, shading, DHW consumption;
- Energy systems efficiency: heat emitters, ventilation electric power.

The number of simulations depends on the number of different of inputs for each category of parameters. In the present case, when all parameters are selected, it performs up to 95 different simulations. The variation threshold is set to 10% of the original value of all parameters. A 10%-variation provides a significant but realistic change of input values, considering that for instance for a 20°C IAT it equals to a 2°C-variation. This way, each input variation is considered in a normalized manner.

Results provide a ranking of impacting input parameters. Results are summarized for each category described above, with the global impact rated in terms of energy consumption (kWh).

4.1.4 Description of implemented building energy models

4.1.4.1 General description

A dedicated separate energy model is created for each of the three renovated buildings of the case study. Buildings are very similar regarding their respective architectural features, as depicted in Figure 4.2. The only differences come from the number of floors, the orientation of the main entrance and the distribution of apartments (in terms of size

and number) within the buildings.

Characteristics of the envelope – including materials, construction elements, walls, windows, glazed doors and doors, and thermal bridges – are the same for each building. Characteristics of energy systems are similar as well, apart from the rated electric power of the ventilation. Operation scenarios are adapted from Th-BCE scenarios embedded in the software, to fit the size of the different housings. All aspects of building energy models are processed within Pléaides modeling software.

Simulations are performed with a five-minute time-step from January the 1st to December the 31st. The heating season is from October the 20th to May the 10th.

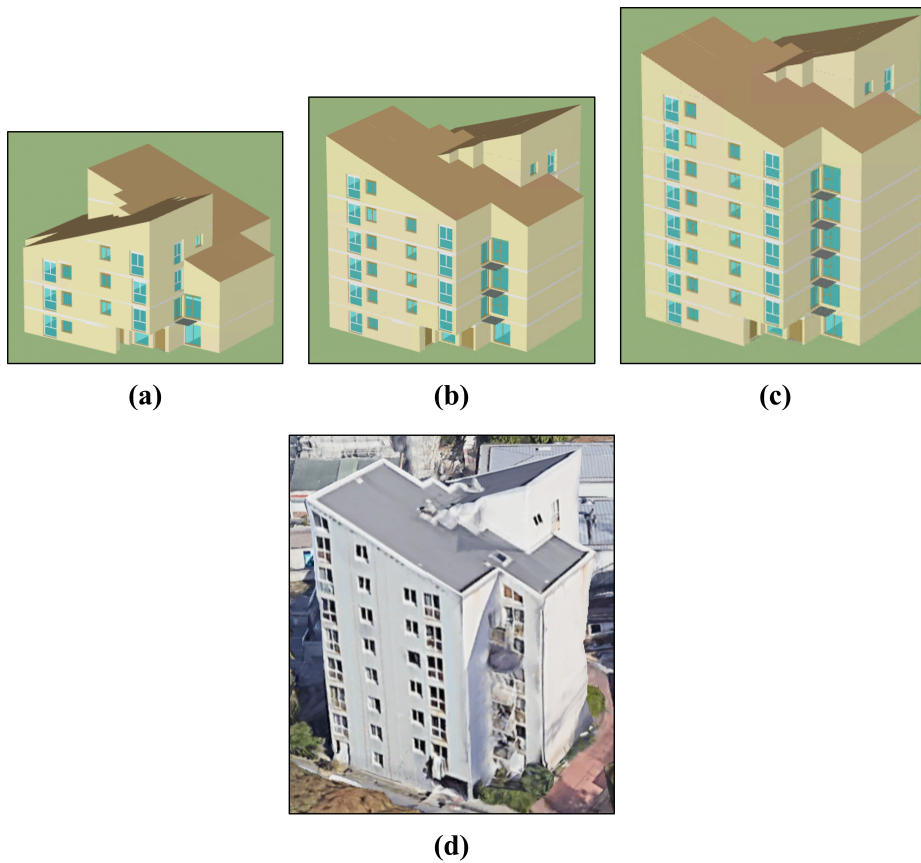


Figure 4.2: Three-dimension models of B1 (a), B2 (b) and B3 (c) extracted from Pléaides, compared with a photo of B3 after retrofit extracted from Google Earth (d).

4.1.4.2 Architectural specificity and thermal characteristics

All buildings have three apartments on the ground floor, four apartments on the intermediate floors and two apartments on the last floor (Figure 4.3). Some apartments have

distinctive architectural features, specifically in the living rooms of apartments located on the last three floors. These rooms have sloping ceilings, up to 5-meters high. They are located under the roof of the buildings without attics.

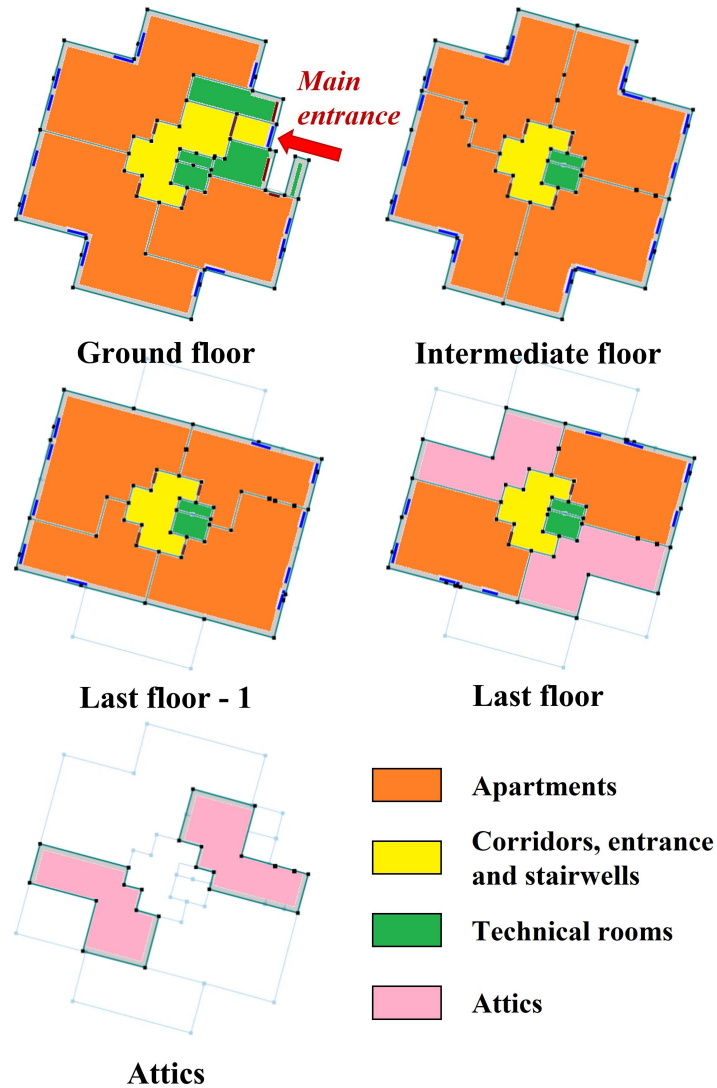


Figure 4.3: Floor plans of building B1 – all buildings share similar architectural features for each floor.

Buildings are built above a crawl space. The ground floor has a technical room where heating and DHW pipes enter the buildings from the underground network. They also host a storage room for bikes and a dustbin room. Each floor has a technical closet for electricity meters, a second one for natural gas meters and a third one for storage. Buildings B2 and B3 also have an elevator. B1 does not, but it has the space for an

4.1 Used modeling methods

elevator that has never been installed. Each floor has a U-shaped corridor serving each apartment and the stairwell.

Thermal zoning is defined as follows. Technical, storage and dustbin rooms are distinct thermal zones (green thermal zones on the ground floor in Figure 4.3). Technical closets and elevator spaces are grouped together for each floor (green thermal zones for each floor in Figure 4.3). On the two last floors, each attic is a separate thermal zone (pink thermal zones in Figure 4.3). Corridors are grouped together with stairwells (yellow thermal zones in Figure 4.3). Each stairwell is connected to the one above with a floor defined with one layer of air but they are in detached thermal zones. The energy modeling software offers the possibility to use an open floor. However, this would result in modeling conflicts regarding thermal zoning that cannot take into account a thermal zone with an opened wall. On the ground floor, the entrance airlock is also added to the thermal zone of the corridor and the stairwell. Finally, each apartment is considered as an independent thermal zone (orange thermal zones in Figure 4.3).

Other strategies could have been considered for the distribution of thermal zones. However, a larger thermal zoning – with floor distribution, for instance – would not match the goals of the research work – i.e. the detailed modeling of the energy consumption – and the spatial granularity of the collected data. A smaller thermal zoning could be intended, but there would be an issue with operation scenarios. Neither the regulatory scenarios nor the sensor network could provide a detailed and reliable room-by-room splitting. The details and numbers of thermal zones for each building are summarized in Table 4.1.

Table 4.1: Details of the number of thermal zones in B1, B2 and B3.

Type of thermal zone	B1	B2	B3	TOTAL
Apartments	13	21	29	63
Corridors, stairwells and entrance	4	6	8	18
Technical rooms	7	8	10	25
Attics	4	4	4	12
TOTAL	28	39	51	118

Thermal bridges are defined from the TRS. Characteristics of the different walls, windows and doors are acquired through the retrofit portfolio. All walls are insulated up to the most recent standards during retrofit, with an exception for the concrete slab of the ground floor, some walls between apartments and technical rooms, and the building roof. The crawl space under the buildings could not be entirely accessed and was not insulated

during the renovation. Walls between apartments and technical rooms that were not insulated are those of the former heating substation and rooms where heating and DHW pipes are coming in – the insulation would prevent a secured access to the water circuit. According to the retrofit portfolio, the roof was already insulated up to standards prior to the renovation.

Windows are double glazed. The main entrance has two double-glazed metallic doors creating an airlock. Doors for technical areas are insulated metallic doors. Entrance doors of apartments have been changed as well for secure and better insulated designs.

4.1.4.3 Heating and domestic hot water

Heating and DHW are both provided by a central heating system located in a new building built next to the three rehabilitated buildings of the case study. The boiler is connected to a local geothermal heating network. Hot water is dispatched underground through an insulated network from the new building to B2 where a heating substation was located prior to retrofit and later removed (Figure 4.4). A water pump located in B2 distributes hot water in B2, and to B1 and B3 through the existing underground water network. This part of the network (orange dotted lines on Figure 4.4) has no reported thermal insulation.

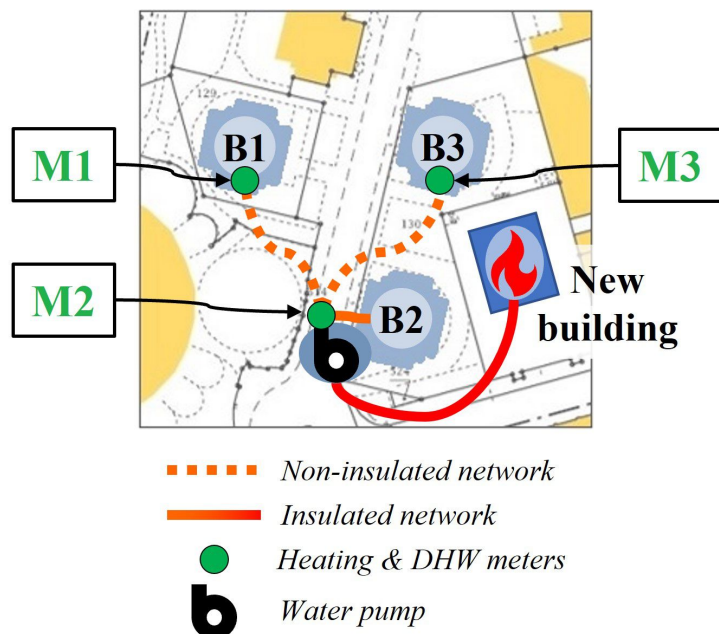


Figure 4.4: Heating and DHW network for B1, B2 and B3 with thermal energy metering points (M1, M2 and M3).

4.1 Used modeling methods

There is no available information on the characteristics of the heating and DHW production except for general knowledge presented above. Neither the TRS nor onsite visits could provide technical data on related energy systems. Therefore, in BEM, the heating and DHW systems are modeled using a pre-filled heating network component from Pléaides embedded library, that describes the existing heating network from the nearby city (whose existing local geothermal heating network connected to the studied buildings is an extension). It is a geothermal heating network with low-temperature hot water production used for both heating and DHW. Unlike the existing system, and since a distinct energy model is implemented for each studied building, hot water production in the energy models is independently considered for each model and building with a dedicated substation for B1, B2 and B3, respectively, that is directly connected to the local geothermal network. The boiler in the newly built building and the underground hot water networks are not considered due to the lack of available descriptive features. The exchanged power of the heating substation is set based on thermal need calculations for each building and depending on the temperature setpoints (that should vary along the calibration process).

The hot water circuit distributes hot water in steel heaters in each apartment. Housings have a heater in every bedroom, in the living room and in the kitchen. Heaters are also installed in bathrooms on the last floor of the buildings. There are no heaters in common areas. DHW is dispatched through the apartments from the water closet to the bathroom and to the kitchen. Following the retrofit, most hot water pipes inside the building are insulated, except for the existing hot water pipes serving heaters within the apartments and for part of the water columns that could not be accessed.

Two other parameters are considered as well and related to heating and DHW production, namely the infiltration rates and the weather conditions. Infiltration is defined with $Q_{4Pa-surf}$ and n_{50} coefficients. The former is the air leak rate with a 4-Pa-depression. A global coefficient is given and divided for each type of cold wall (outdoor walls, roof, windows and doors). The coefficient is left to default settings with $Q_{4Pa-surf}$ set to $1 \text{ m}^3/\text{h}/\text{m}^2$ (the recommended value from the latest environmental regulation for collective housings), distributed with 10%, 50%, 20% and 20% for outdoor walls, roof, windows and doors, respectively. The latter coefficient, n_{50} , is the maximum air leak rate with a 50-Pa-depression and is set to 0.6 vol/h. It is the default coefficient for all types of buildings and usages.

Climatic conditions are of paramount importance since they drive the thermal energy consumption and the building energy efficiency. As described in Section 2.3.4.1, a weather

station is installed three kilometers away from the case study site. However, for building energy simulations, an embedded weather file is selected from the same climatic zone (H1a). This choice is driven by the complexity and cost of creating a dedicated weather file based on local weather data, and because of the availability of weather data. There are up to two months and half of cumulative missing data over the summer and autumn of 2021. Since the weather station is installed in February 2020, there is not enough data to create a complete and reliable weather file. Therefore, simulations are performed using another weather station with similar climatic conditions. Then, simulation results are corrected with local weather data: daily heating energy consumption per heating degree days (HDD) from simulations is multiplied by local HDD (Eq. 4.1).

$$H_{corrected} = (H_{simulated}/HDD_{simulation}) \times HDD_{local} \quad (4.1)$$

Where $H_{simulated}$ is the heating energy consumption over a given day and simulated from building energy models; $HDD_{simulation}$ is the number of heating degree days from the same day and from the weather file used in the energy simulation; HDD_{local} is the number of heating degree days from the same day and calculated from OAT data collected by the weather station; $H_{corrected}$ is the heating energy consumption from the same day, simulated from building energy models and corrected with the HDD.

4.1.4.4 Ventilation

Ventilation in the buildings is provided by a humidity-sensitive simple-flow CMV (controlled mechanical ventilation). Each building has eight ventilation extraction columns (two per apartment). Air extraction units are located in the bathroom, water closet and kitchen of the apartments. The air extraction rate is given in the retrofit portfolio for each unit depending on the size of the apartment, room location and room use.

4.2 Modeling results and discussions

4.2.1 Heating energy consumption: measurements

Before the comparison between simulations and measurements, it is necessary to review the measurements from heating energy meters in the sensor network.

There are three heating ultrasonic meters installed onsite. B1 and B3 both have one meter counting their respective heating energy consumption (M1 and M3 metering points in Figure 4.4). A third meter is installed in B2 (M2 metering point in Figure 4.4). This

4.2 Modeling results and discussions

meter measures the aggregated heating energy consumption of B1, B2 and B3, including the heat loss from B2 to B1 and from B2 to B3.

The heating season starts on the 20th of October, 2021 and ends on the 10th of May, 2022. However, acquired data for B1 and B3 range from the 24th of November, 2021 to the end of the heating season. For B2, there are no exploitable data due to technical malfunctions and the impossibility to find a suitable replacement solution in time. Therefore, several assumptions are made to replace missing data and to perform a full-heating-season comparison for the three buildings. Considering the set up for heating energy monitoring described above, and summarized in Eq.4.2, the heating energy consumption of B2 would be deducted, taking into account the heat loss in the underground hot water network between B2 and B1 and between B2 and B3.

$$H_{tot} = (H_{B1} + H_{B2} + H_{B3}) + L \quad (4.2)$$

Where H_{tot} is the total heating energy consumption of B1, B2, B3 and including the heat loss from B2 to B1 and from B2 to B3, measured in M2 on Figure 4.4; H_{B1} is the heating energy consumption of B1, measured in M1 on Figure 4.4; H_{B2} is the heating energy consumption of B2; H_{B3} is the heating energy consumption of B3, measured in M3 on Figure 4.4; L is the heat loss through the piping network between B2 and B1 and between B2 and B3.

However, because of the lack of data from B1, another strategy is adopted. A ratio of heating energy consumption per unit of heated area is calculated for B3. This ratio is applied to B1 (Eq. 4.3) and B2 (Eq. 4.4) to assess the heating energy consumption of the two buildings.

$$H_{B1} = \frac{H_{B3}}{A_{B3}} \times A_{B1} \quad (4.3)$$

$$H_{B2} = \frac{H_{B3}}{A_{B3}} \times A_{B2} \quad (4.4)$$

Where A_{B1} , A_{B2} and A_{B3} are the total heated area of B1, B2 and B3 respectively.

The subtraction between measurements in B2, the sum of measurements in B3 and calculations for B1 and B2 results in heat loss assessment. Consequently, it is assumed that heating consumption patterns in B1 and B2 are similar to B3.

Missing data at the beginning of the heating season are processed as well, using the heating degree day method. It calculates the difference between the average daily outdoor air temperature (OAT) and the IAT set point minus one degree Celsius. In other words, it calculates "how much" the outdoor air should be heated to reach the expected indoor

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

air temperature. For a regulatory calculation, the temperature set point should be 19°C, and the selected temperature is 18°C. The 1°C difference is assumed to be related to solar gains and indoor heat gains (Eq. 4.5). In the present case however, analyses of IAT in Chapter 3 highlighted that the temperature set points are far above 19°C. Hence, the reference temperature is the daily averaged IAT of apartments within the same building and for which measurements are performed.

$$\begin{cases} HDD = 18^{\circ}C - OAT_{averaged} , & \text{if } OAT_{averaged} < 18^{\circ}C \\ HDD = 0 , & \text{if } OAT_{averaged} \geq 18^{\circ}C \end{cases} \quad (4.5)$$

Where HDD is the amount of heating degree days for a given day and $OAT_{averaged}$ is the averaged outdoor air temperature for the same day.

A ratio is applied on each building to recover missing data at the beginning of the heating season. The averaged heating energy consumption per HDD over the heating season is multiplied by the number of HDD of a given missing day of data. It results in the assessed heating energy consumption over the missing day (Eq. 4.6).

$$H_{day, Bi} = HDD_{day} \times \frac{H_{Bi}}{HDD_{heating\ season}} \quad (4.6)$$

Where $H_{day, Bi}$ is the heating energy consumption of building Bi (with $i=\{1,2,3\}$) over a missing day of data; H_{Bi} is the energy consumption of the same building Bi over the whole heating season; HDD_{day} is the amount of heating degree days over the missing day of data; $HDD_{heating\ season}$ is the amount of heating degree days over the whole heating season.

Results from measurements and calculations are summarized in Table 4.2 for the heating season 2021/2022.

Heat loss over the heating network between B2 and B1 and between B2 and B3 are assessed up to 8% of the total heating energy consumption measured in B2 over the heating season. There are no available information about the heating network that would help investigate this result, such as piping diameter, thermal insulation, pipe material and current overall state of the hot water circuit (such as leaks). However, it is assumed to be old, unmaintained and poorly insulated since it could not be accessed and was not part of the retrofit project. Considering these assumptions the overall heat loss may seem low, although not completely unrealistic looking at comparative studies on heating networks loss [7]. Nevertheless, the variations over the four months with most of the heating energy consumption (November to February), may question the reliability of the calculations:

4.2 Modeling results and discussions

Table 4.2: Measurements and assessments for heating energy consumption of B1, B2 and B3 and the heat loss between B2 and B1 and between B2 and B3, over the heating season from 2021/10/20 to 2022/05/10 – the gross floor areas of B1, B2 and B3 are 1,157 m², 1,806 m² and 2,255 m², respectively.

	B1+B2+B3+loss (measured)	B1 (assessed)		B2 (assessed)		B3 (measured)		Heat loss (assessed)	
		kWh	kWh /m ² *	kWh	kWh /m ² *	kWh	kWh /m ² *	kWh	% of total
Oct-21	15,427	3,321	4.3	4,317	3.4	6,452	3.6	1,338	9%
Nov-21	45,381	8,870	11.6	13,662	10.7	19,479	10.9	3,369	7%
Dec-21	50,822	9,130	11.9	15,222	11.9	21,313	11.9	5,157	10%
Jan-22	50,391	9,994	13.1	16,663	13.1	23,331	13.1	403	1%
Fev-22	40,137	7,909	10.3	13,185	10.3	18,462	10.3	581	1%
Mar-22	29,814	5,915	7.7	9,861	7.7	13,807	7.7	232	1%
Avr-22	21,023	2,991	3.9	4,988	3.9	6,984	3.9	6,059	29%
May-22	2,440	274	0.4	457	0.4	640	0.4	1,069	44%
TOTAL	255,435	48,405	63.3	78,354	61.4	110,468	61.9	18,207	8%

* gross floor area

for an equivalent energy consumption in December and January, the calculated heat loss dropped from 10% to 1%, which is quite unlikely to happen in reality.

4.2.2 Non-calibrated building energy models

4.2.2.1 Comparison: simulations versus measurements

A distinct NC-BEM is implemented for each building B1, B2 and B3 to model their respective heating energy consumption after the retrofit. Non-calibrated simulations are compared to measurements from our sensor network. Results of the comparison are summarized in Table 4.3. For each energy model, MAE (Eq. 1.1), MAPE (Eq. 1.2) and RMSE (Eq. 1.3) metrics are computed over the eight months of the heating season to provide the monthly averaged error rate. The relative error between the aggregated measured and simulated energy consumption over the heating season is also given by Equation 4.7.

$$RE = 100 \times \frac{\hat{y}_{hs} - y_{hs}}{y_{hs}} \quad (4.7)$$

Where RE is the relative error in %; \hat{y}_{hs} is the energy consumption from simulations over the heating season; y_{hs} is the measured energy consumption over the heating season.

The comparative study highlights that simulated heating energy consumption re-

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

Table 4.3: Comparison between the results of the NC-BEM, and measurements and assessments from collected data, over the heating season 2021/2022 – errors are calculated over the whole heating season.

		Heating energy consumption over heating season (kWh)	Comparison: simulations versus measurements			
			MAE (kWh)	MAPE (%)	RMSE (kWh)	Relative error (%)
B1	Measurements	48,400	/	/	/	/
	NC-BEM	22,182	3,277	73%	3,586	-54%
B2	Measurements	78,346	/	/	/	/
	NC-BEM	55,019	2,916	43%	3,340	-30%
B3	Measurement	110,005	/	/	/	/
	NC-BEM	73,307	4,587	47%	5,152	-33%

produces the same monthly trend as in the measurements, illustrated in Figure 4.5 that depicts the heating energy consumption from NC-BEM (brown color in the bar graphs (a-1), (b-1) and (c-1)) and measurements (red color in the bar graphs (a-1), (b-1) and (c-1)) along with the modeling error between the two (in the combined graphs (a-2), (b-2) and (c-2), with the monthly difference between measurements and simulations in kWh in the blue bar graph, and the relative error in % with the purple dotted line). However, there is an under-estimation of the heating energy consumption compared to measured heating energy consumption, with a difference -54%, -30% and -33% for B1, B2 and B3, respectively. B2 and B3 have similar error rates with monthly MAPE of 43% and 47%, respectively, over the heating season. However, the error rate of B1 is much larger with a MAPE of 73%. Then, the comparison of MAE and RMSE underlines months including February, March, October, November and December (in Figure 4.5 (a-1), (b-1) and (c-1)), for which the difference between the simulations and measurements is more significant (since RMSE gives a higher weight to larger errors). There is also a clear trend that shows that the higher the energy heating energy consumption, the lower the difference, with an exception for the month of April in B2 and B3. Indeed, January exhibits the smallest modeling error for all three buildings. In May, the relative error is significantly higher compared with other months, due to the small heating energy consumption.

In light of data analyses from Chapter 3, the initial PG resulting from NC-BEM is expected. Features of the building envelope, the surrounding environment and the energy systems match the existing buildings as realistically as possible. However, building operation is standard while results of data analyses clearly underlined the difference between standard building operation and field-measured data. Specifically, IAT is significantly higher than for regulatory scenarios. Other operation scenarios also exhibited different

4.2 Modeling results and discussions

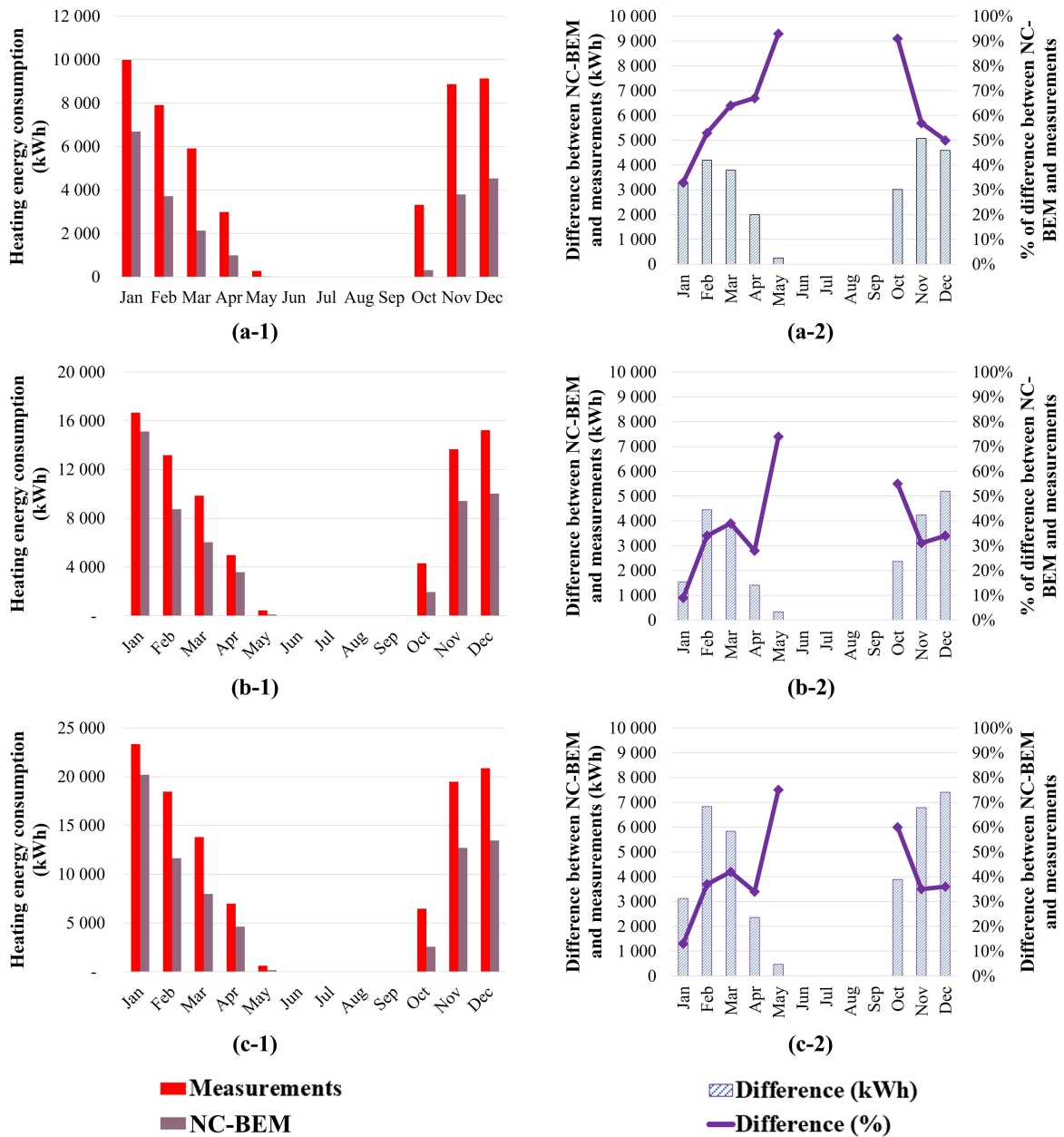


Figure 4.5: Comparison of heating energy consumption from measurements and NC-BEM – On the left is the monthly evolution of heating consumption for B1 (a-1), B2 (b-1) and B3 (c-1), from measurements in red and from NC-BEM in brown – On the right is the relative error between models and measurements, for B1 (a-2), B2 (b-2) and B3 (c-2): the bar graph shows the monthly relative error in kWh and the purple line shows monthly relative error in %.

patterns, both in terms of amplitude and timing of events.

4.2.2.2 Sensitivity analysis

A sensitivity analysis is performed for each studied building to identify the input parameters that drive the building heating energy consumption. It ranks each tested parameter with respect to its effect on the selected output. Hence, it also provides a relevant line of enquiry to identify potential sources of uncertainties in energy models. Input parameters with their respective relative impact and ranking are presented in Table 4.4 for B1, B2 and B3.

Sensitivity analyses highlight the weight of IAT scenarios in the energy models as the first-ranking energy-driving input. IAT has a relative impact of 68.3%, 55.2% and 60.3% on heating energy consumption, equivalent to an additional 13.1 kWh/m², 16.8 kWh/m² and 19.6 kWh/m² for B1, B2 and B3, respectively, considering a 10% increase of the temperature. Since heating is provided by a collective energy system without indoor thermostat and with no possible interaction from building occupants, IAT is the logical input to have the most effect on heating energy consumption. Occupancy and dissipated power relate to indoor heat gains. The latter has a non-negligible but limited predicted impact: 1.6 kWh/m², 1.6 kWh/m² and 2.2 kWh/m² additional heating energy consumption for B1, B2 and B3, respectively. Since dissipated power is electrical, it is mostly related to occupancy of housings. Occupancy is at less than 5% of relative impact on heating energy consumption. Shading standard scenarios are designed to simulate rolling shutters used only from 8 pm to 7 am all along the year. Therefore, it is expected to have very little effect: in winter it maximizes solar heat gains. There is little heat loss through the ventilation and no reported impact of DHW use. However, it should be observed that window opening cannot be selected for sensitivity analyses and its effect is not taken into account, due to software limitation.

All other inputs relate to characteristics of the building envelope and of the energy systems. They have a quite limited effect on heating energy consumption, except for the dimensions of walls – for B1 and B3 – and the thermal conductivity of materials – for B1 – with at least 5% of relative impact.

Out of sensitivity analyses, we conclude that operation scenarios, and specifically IAT scenarios are by far the most energy-driving inputs. For a 10% variation of the values of input parameters, operation scenarios have 82.5%, 62.5% and 72.0% of cumulative relative impact for B1, B2 and B3, respectively. Energy systems and the building envelope have 28.9%, 19.1% and 20.5% of cumulative impact for B1, B2 and B3, respectively.

4.2 Modeling results and discussions

Table 4.4: Results of local sensitivity analyses on NC-BEM for B1, B2 and B3 with relative impact (in kWh and kWh/m²) and ranking of input parameters – The relative impact is presented in absolute value.

	B1			B2			B3			
	Rank	Relative impact (%)	kWh /m ² *	Rank	Relative impact (%)	kWh /m ² *	Rank	Relative impact (%)	kWh /m ² *	
NC-BEM	/	/	19.2	/	/	30.5	/	/	32.5	
Measurements	/	/	41.8	/	/	43.4	/	/	48.8	
Input parameters										
Scenarios	IAT	1	68.3%	13.1	1	55.2%	16.8	1	60.3%	19.6
	Dissipated power	3	8.6%	1.6	2	5.2%	1.6	3	6.9%	2.2
	Occupancy	6	3.3%	0.6	9	1.2%	0.4	5	2.6 %	0.8
	Ventilation	8	2.2%	0.4	13	0.7%	0.2	7	2.0%	0.7
	Shading	14	0.1%	0.0	14	0.3%	0.1	14	0.2%	0.1
	DHW	17	0.0%	0.0	17	0.0%	0.0	17	0.0%	0.0
Energy systems and envelope	Dimensions of walls	2	10.1%	1.9	4	4.3%	1.3	2	8.8%	2.8
	Thermal conductivity of materials	4	6.1%	1.2	3	4.5%	1.4	4	2.8%	0.9
	Linear thermal bridge	5	3.5%	0.7	7	1.5%	0.5	10	1.1%	0.4
	Thermal conductivity of windows	7	2.9%	0.6	6	1.5%	0.5	8	1.5%	0.5
	Thickness of materials	13	0.6%	0.1	5	2.9%	0.9	6	2.4%	0.8
	Efficiency of heaters	9	1.9%	0.4	8	1.5%	0.4	9	1.6%	0.5
	Solar factor of windows	10	1.8%	0.3	10	1.0%	0.3	11	0.9%	0.3
	Dimensions of windows	11	1.0%	0.2	11	1.0%	0.3	12	0.8%	0.3
	Air infiltration	12	0.9%	0.2	12	0.8%	0.2	13	0.4%	0.1
	Density of materials	15	0.1%	0.0	15	0.1%	0.0	16	0.1%	0.0
	Specific heat of materials	16	0.1%	0.0	16	0.1%	0.0	15	0.1%	0.0

* gross floor area

Therefore, the largest uncertainties to explain the initial PG may lie in the building operation scenarios.

4.2.2.3 Conclusions on the non-calibrated building energy models

Non-calibrated building energy models provide a simulation of heating energy consumption for the three buildings of the case study after the energy retrofit. They are implemented using information from the retrofit portfolio and standard operation scenarios.

Simulations reproduce the monthly trends observed in the measurements but they underestimate the heating energy consumption, which results in a significant PG for all three buildings, with the largest gap for B1. Although the difference between models and measurements is significant, it is expected as well. Indeed, as detailed in Chapter 3, dissipated power, IAT, occupancy, window opening and DHW use are quite different from standard building operation scenarios.

Through the sensitivity analyses, we identified that IAT is, by far, the most energy-driving modeling input. Among other impacting parameters are dissipated energy scenarios, the occupancy and some features of building envelope. Because of the weight of IAT, a variation of 10% of the parameter values in operation schedules can affect heating energy consumption by more than three times that of a similar modification in the characteristics of the building envelope and the energy systems. Therefore, the former group of inputs is expected to hold the largest part of modeling uncertainties over building heating energy consumption and are selected as the targets of BEM calibration.

4.2.3 Performance gap and calibration

4.2.3.1 Manual calibration: tuning of modeling inputs

To close the initial PG, the selected strategy relies on the manual tuning of NC-BEM inputs. These inputs are the energy-driving parameters highlighted through sensitivity analyses and that were investigated in previous data analyses. Five input parameters are selected in the calibration process: IAT, dissipated power, occupancy, window opening and ventilation. An assumption is made, as IAT considered in standard scenarios and in sensitivity analyses is a temperature setpoint, while the IAT scenarios used to tune NC-BEM are measured IAT. There no parameters related to ventilation that are monitored in our sensor network. However, scenarios are adapted to match the peaks of occupancy in the instrumented housings. DHW use is not considered due to its negligible impact on

heating energy consumption.

The effect of window opening on heating energy consumption could not be assessed from sensitivity analyses. Window opening data processing provides typical daily patterns with the averaged opening duration over each hour of the day. The opening width of the windows is not measured. There is a significant scientific literature on the modeling and identification of patterns for window opening and closing [8, 9]. However, studies focus on the drivers of window opening and on the opening duration and not on the opening width of windows, while this input is not necessarily stationary and it depends on the intended action of the occupants and on the types of windows.

Consequently, the effect of window opening on heating energy consumption is tested separately, without other tuned parameters. The goal of the study is to calibrate building energy models, and there would be too much uncertainty to consider window opening in the calibration process. Several maximum opening widths are tested. The strategy is the following: initially a window is always opened at maximum width, and the opening duration profiles are used as adjustment coefficients. For instance, for a window that can be opened at an 80% maximum opening width and that is opened for half-an-hour, the maximum opening is set to 80% and the opening duration is 50% (half an hour). Since the time-step of the window-opening time series is one hour, it is then considered as equivalent to one hour of opening with 40% opening width. All windows are set with the same maximum opening width for a given simulation. This process is summarized in Figure 4.6.

A total of five tests are conducted for the manual calibration. At first, NC-BEM are tuned with one parameter at a time (IAT, dissipated electric power and occupancy), and only in instrumented housings for which there are available data. Indeed, data analyses have highlighted that typical operation profiles for dissipated power and occupancy are specific to a given housing. All non-instrumented apartments are modeled with standard scenarios. Then, all three inputs are tuned together (including the modification of the ventilation schedules). Finally, an extrapolation of IAT tuning in non-instrumented housings is implemented. Considering the measurements over the three studied buildings, it might be reasonable to assume that non-instrumented apartments may exhibit similar IAT patterns. Therefore, IAT scenarios are extrapolated to non-instrumented housings using the average of IAT measurements in instrumented housings withing the same buildings: 22.2°C, 22.7°C and 24.3°C for B1, B2 and B3, respectively. Three tests are run for window opening with a maximum opening width of 20%, 40% and 80% (which is the largest opening width for the type of windows installed in considered buildings) and for

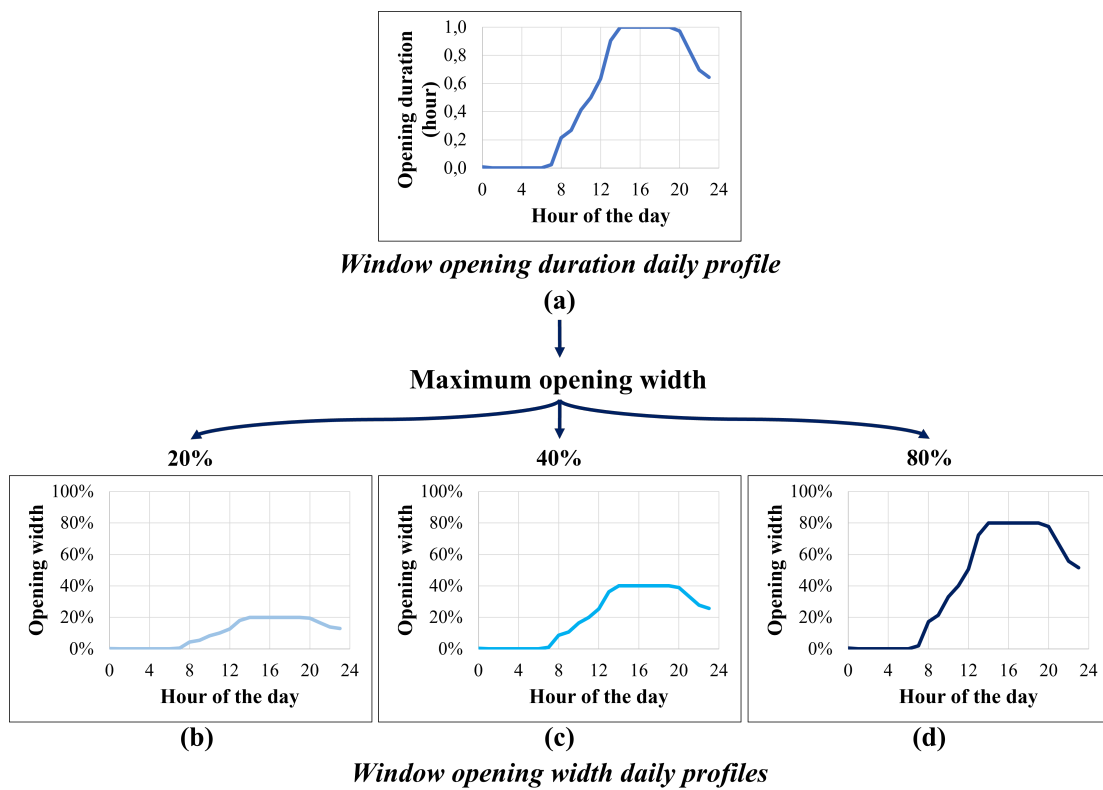


Figure 4.6: Process of data modification to test window opening configurations in building energy models – A maximum opening width (%) is used with a daily profile figuring window opening duration (a) to provide daily profiles with window opening width: here are examples with 20% (b), 40% (c) and 80% (d) maximum opening width.

4.2 Modeling results and discussions

each of the building case study. Figure 4.7 summarizes the different tests for window opening and the manual calibration of BEM.

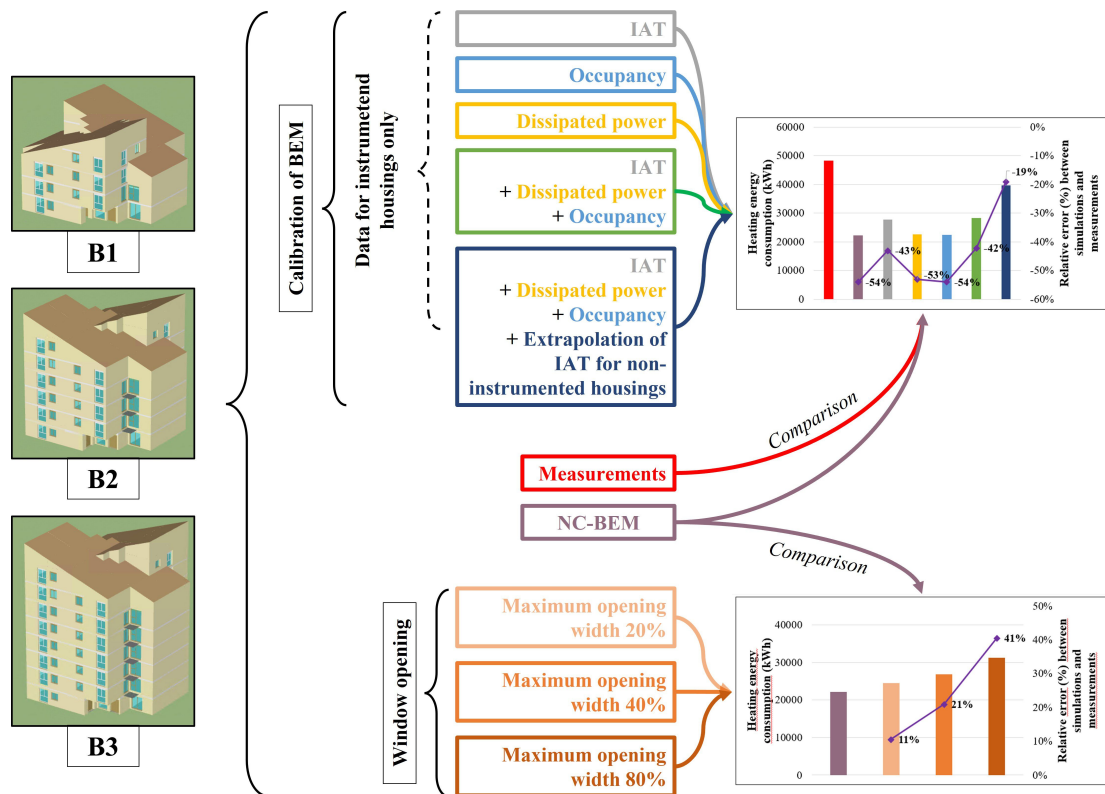


Figure 4.7: Manual calibration and window opening tests for B1, B2 and B3 – Five calibration tests are performed for each building with the tuning of the IAT, the occupancy, the dissipated electric power, the combination of the three parameters, and the additional extrapolation of IAT in non-instrumented housings – Three tests are conducted for each building on window opening with 20%, 40% and 80% maximum opening width – Results are compared with measurements and simulations from NC-BEM.

4.2.3.2 Comparison: simulations versus measurements

BEM of each building are manually calibrated step by step. IAT, dissipated electric power and occupancy are tuned separately. When one parameter is tuned, the other two are set with standard scenarios as for NC-BEM. Then all three parameters are tuned at the same time and, finally, the tuning is complemented with the extrapolation of IAT for non-instrumented housings. Results are summarized in Table 4.5 and depicted in Figure 4.8.

The parameter-by-parameter tuning provides several insights. IAT tuning up to measured temperature levels significantly increases the heating energy consumption compared to NC-BEM, which was expected, considering results of sensitivity analyses. There is 11%, 15% and 14% of heating energy consumption increase for B1, B2 and B3, respectively. However, the resulting heating energy consumption remains lower than the measurements as it only accounts for two apartments in each building for IAT.

Occupancy and dissipated power are the other two operation scenarios that are considered. The difference after calibration remains very low. With dissipated power, B1 and B2 show a 1% increase, while for B3 the increase is lower than a percent. For occupancy, only B2 exhibits around 1% of heating energy consumption increase. This low impact is supported by the results of the sensitivity analyses: with a 10% difference of the values of input parameters for all housings, the effect of dissipated electric power and occupancy over heating energy consumption remains under 10% and under 5% respectively in NC-BEM. As for IAT, these parameters are only monitored in a few apartments for each building. It is also interesting to note that for both occupancy and dissipated power, input tuning increases heating energy consumption. Both parameters translate to internal heat gains, which means that the cumulative heat gains from measurements is lower than that of standard scenarios.

When all tuned parameters in instrumented housings are combined – i.e. IAT, dissipated power, occupancy and ventilation fitted to occupancy patterns – the resulting increase in heating energy consumption is slightly higher than for IAT alone. As it accumulates the effect of all three tuned inputs, this result is expected as well.

The last test combines all tuned input parameters and extrapolates the IAT to non-instrumented housings (dark blue column of the bar graph in Figure 4.8 and in Figure 4.9). IAT levels differ from one housing to the other. Nevertheless, unlike for other measurements, daily IAT patterns are stable along the day. Furthermore, the measurements are much higher than the standard 19°C. Since IAT is the most energy-driving

4.2 Modeling results and discussions

Table 4.5: Results of step-by-step input calibration on heating energy consumption for B1, B2 and B3, over the heating season 2021/2022 – For each building, IAT, occupancy and dissipated power are tuned separately from each other, then they are tuned at the same time, and finally, the tuning of all three parameters is complemented with the extrapolation of IAT to non-instrumented housings.

		Heating energy consumption (kWh)	Measurements vs. simulations				
			MAE (kWh)	MAPE (%)	RMSE (kWh)	Relative error (%)	
B1	NC-BEM		22,182	3,277	73%	3,586	-54%
	Measures		48,400	/	/	/	/
	Tuning of inputs only for instrumented housings	IAT	27,799	2,575	54%	2,877	-43%
		Dissipated power	22,539	3,233	72%	3,537	-53%
		Occupancy	22,444	3,269	72%	3,577	-54%
		All parameters*	28,235	2,521	53%	2,819	-42%
All parameters* + extrapolation of IAT for non-instrumented housings		39,772	1,129	29%	1,281	-19%	
B2	NC-BEM		55,029	2,916	45%	3,340	-30%
	Measures		78,346	/	/	/	/
	Tuning of inputs only for instrumented housings	IAT	66,685	1,586	18%	1,946	-15%
		Dissipated power	55,722	2,828	42%	3,252	-29%
		Occupancy	55,266	2,885	43%	3,309	-29%
		All parameters*	67,822	1,486	17%	1,853	-13%
All parameters* + extrapolation of IAT for non-instrumented housings		97,027	2,335	42%	2,642	+24%	
B3	NC-BEM		73,307	4,587	47%	5,152	-33%
	Measures		110,005	/	/	/	/
	Tuning of inputs only for instrumented housings	IAT	89,460	2,815	22%	3,376	-19%
		Dissipated power	74,082	4,490	46%	5,052	-33%
		Occupancy	74,207	4,475	46%	5,038	-33%
		All parameters*	90,805	2,729	23%	3,256	-17%
All parameters* + extrapolation of IAT for non-instrumented housings		153,378	5,422	72%	5,888	+39%	

* includes ventilation scenarios adapted from peak activity in occupancy scenarios

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

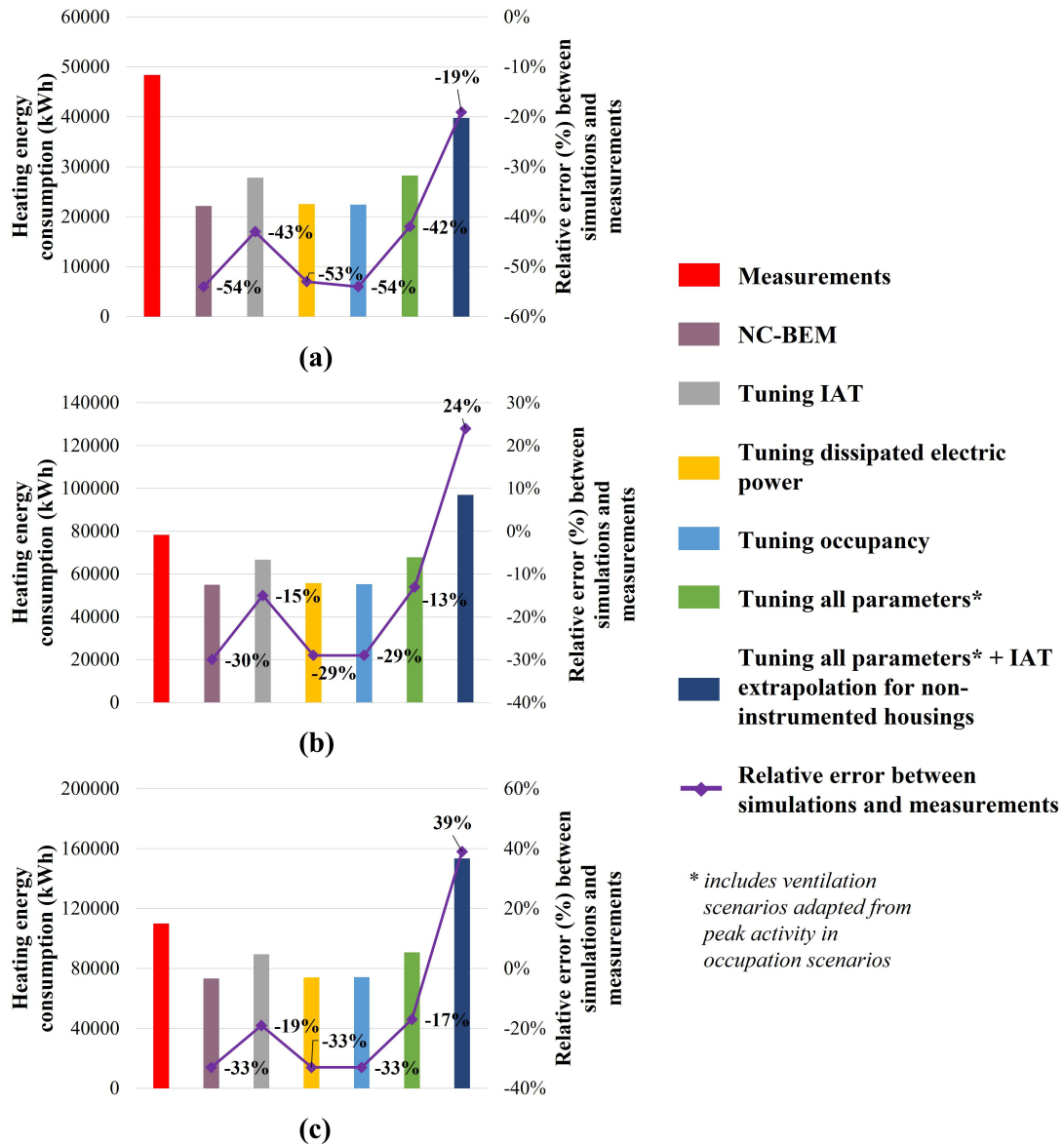


Figure 4.8: Comparison of heating energy consumption from measurements and calibration process for B1 (a), B2 (b) and B3 (c) over the heating season 2021/2022 – Measurements are figured in red, non-calibrated models are in brown and the tuning of inputs is depicted in grey, yellow, blue, green and dark blue for the tuning of IAT, dissipated power, occupancy, all parameters for instrumented housings, and the extrapolation of the IAT for non-instrumented housings, respectively – The purple line shows the relative error in % between simulations and measurements.

4.2 Modeling results and discussions

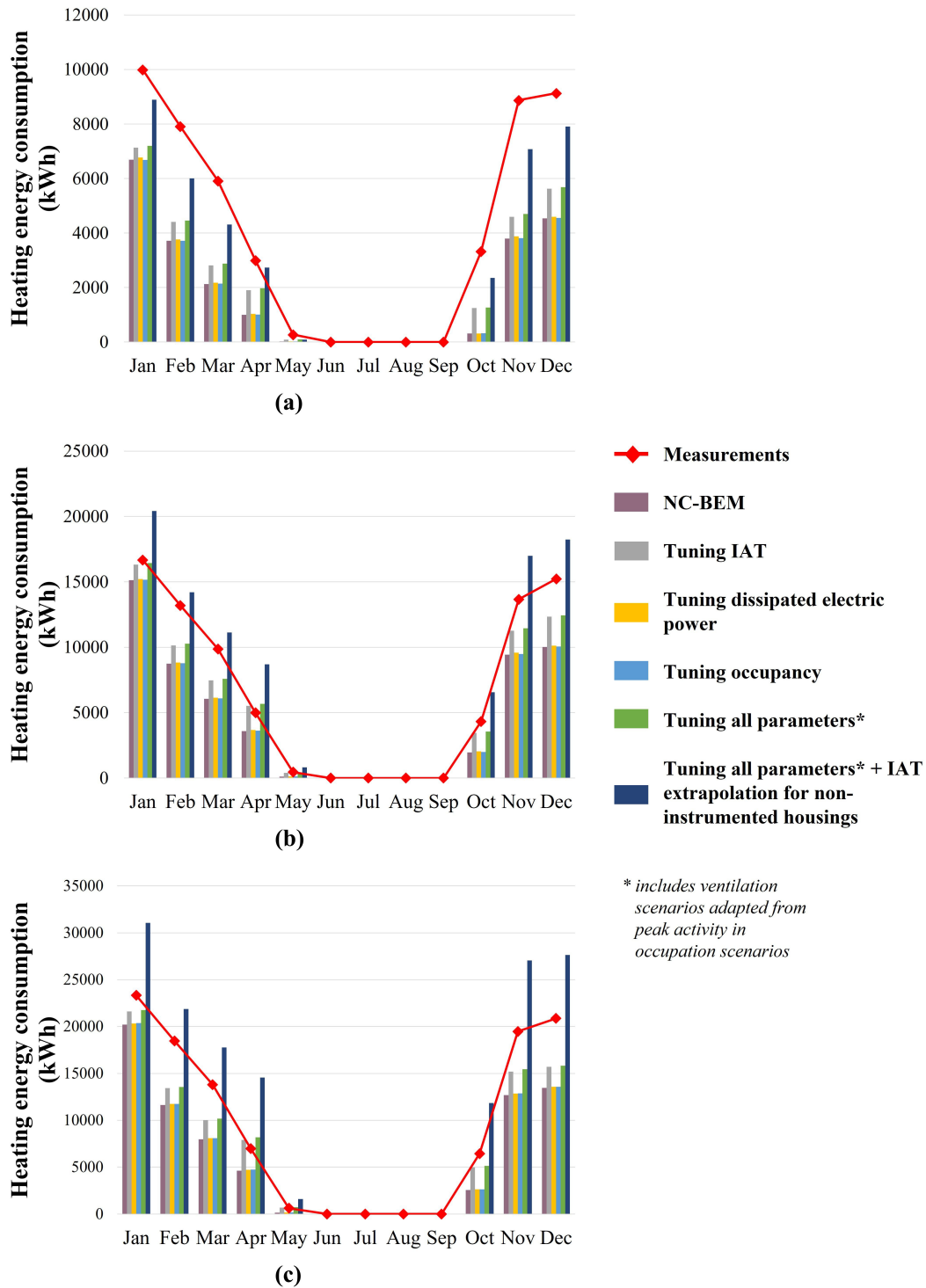


Figure 4.9: Monthly evolution and comparison of heating energy consumption from measurements and tuning of inputs for B1 (a), B2 (b) and B3 (c) – Non-calibrated models are in brown and the tuning of inputs is depicted in grey, yellow, blue, green and dark blue for the tuning of IAT, dissipated power, occupancy, all parameters for instrumented housings, and the extrapolation of IAT for non-instrumented housings, respectively – The red dotted line shows the measurements from our sensor network.

input parameter in the model, it would be expected that a more realistic representation of this input in non-instrumented housings could help bridge the energy PG. However, the results of this step show that the heating energy consumption is now much higher than the corresponding measurements, except for B1. The relative error is of -19%, +24% and +39% for B1, B2 and B3, respectively, and compared with the measurements. The smallest modeling error is achieved for B1, with a MAPE of 29%, while the MAPE for B2 and B3 is of 42% and 72%, respectively.

The extrapolation of IAT tuning in non-instrumented housings highlights the limits of the proposed strategy, specifically with the data collection focusing on a sample of the apartments in the considered buildings. The role of IAT set point in heating energy consumption is substantial. However, data could only be reliably collected for six apartments out of sixty-three, which represents less than 10% of the housings. Moreover, the range of temperatures in the different instrumented housings is significant and because of the size of the sample, no specific trend with respect to housing orientation, floor or any other parameter could be concluded. Therefore, using the average temperature of instrumented apartments to extrapolate to non-instrumented apartments probably is an assumption that is too strong with respect to the reality: it is quite unlikely that apartments may all reach the exact same IAT levels. This might be one of the reasons why the calibration process is not fully successful. Indeed, there are also other remaining uncertainties, regarding the impact of occupancy and dissipated power in other apartments. Finally, another potential source of errors might be related to the lack of energy metering data from our sensor network and the assumptions that were made to recover missing data, since out of the three considered buildings, only the heating consumption of B3 and of B1, B2 and B3 together were directly measured.

4.2.3.2.1 Window opening

Window opening is tested separately from other inputs. Indeed, window opening detection is addressed by our wireless sensor network but the opening width is not monitored. However, it is a significant parameter to consider as it relates to occupants' behavior and it is expected to impact on heat loss and heating energy consumption. Since there is too much uncertainty regarding this parameter, it is not included in the calibration process.

Three rounds of tests are performed for each building. In instrumented housings, profiles extracted from field data analyses and described in Appendix C.5 are used and modified according to the process described in Figure 4.6, with a maximum opening width of 20%, 40% and then 80% set for each window. In non-instrumented apartments,

4.2 Modeling results and discussions

windows are always closed. All windows of all housings are always closed as well in NC-BEM, as it is the standard configuration for this parameter. Results are summarized and compared with the simulations from NC-BEM in Table 4.6 and in Figure 4.10.

Table 4.6: Impact of window opening on the thermal energy consumption of B1, B2 and B3 over the heating season 2021/2022: 2, 4 and 1 instrumented housings with available window opening data are considered in these tests, out of total of 13, 21 and 29 apartments in B1, B2 and B3, respectively – Results show the heating energy consumption of the whole building – Window opening tests are conducted only for instrumented housings with available window opening data.

		Heating energy consumption (kWh)	Normalized heating energy consumption (kWh/m ²) *	Relative error (%)
B1	NC-BEM – No window opening	22,182	19.2	/
	Maximum opening width – 20%	24,559	21.2	+10.7%
	Maximum opening width – 40%	26,839	23.2	+21.0%
	Maximum opening width – 80%	31,195	27.0	+40.6%
B2	NC-BEM – No window opening	55,019	30.5	/
	Maximum opening width – 20%	73,415	40.6	+33.4%
	Maximum opening width – 40%	84,496	46.8	+53.6%
	Maximum opening width – 80%	92,496	51.2	+68.1%
B3	NC-BEM – No window opening	73,307	32.5	/
	Maximum opening width – 20%	77,190	34.2	+5.3%
	Maximum opening width – 40%	80,622	35.8	+10.0%
	Maximum opening width – 80%	85,877	38.1	+17.1%

* gross floor area

The first striking result is that the effect of window opening on the heating energy consumption is significant regardless of the number of instrumented housings or the number of sensors in a building, as illustrated in Figure 4.10. For instance, in B3, data are only collected in B3/2 and for three out of four installed sensors. Nevertheless, even with the lowest tested opening width (20%), there is still a 5.3% increase of the heating energy consumption for a building (B3). Then, B1 has two instrumented apartments with all eight operative sensors and B2 has four instrumented housings with two missing sensors

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

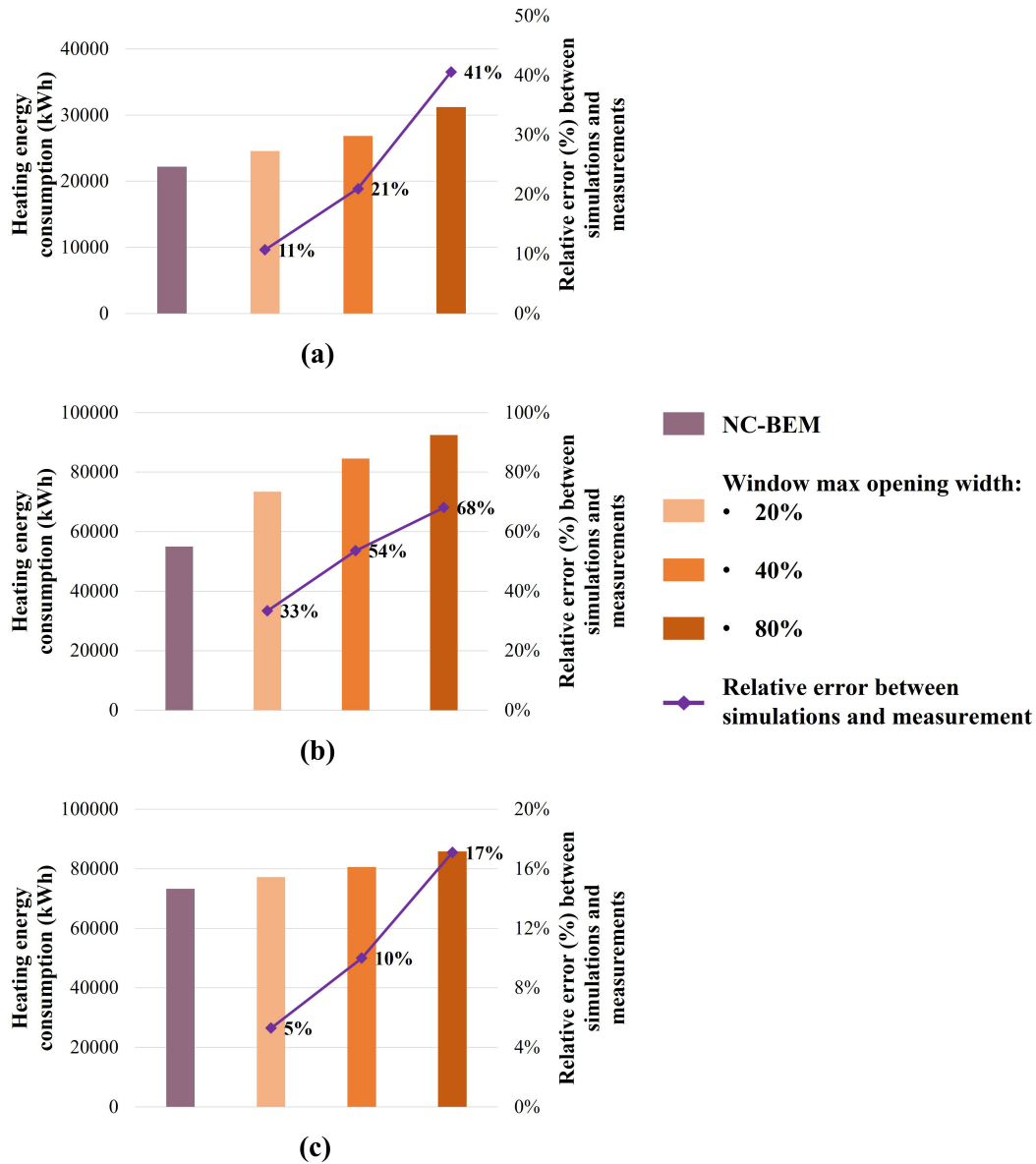


Figure 4.10: Tests of maximum opening width of windows with 20% (in beige), 40% (in orange) and 80% (in brown) maximum opening widths, compared with NC-BEM (in purple) for B1 (a), B2 (b) and B3 (c) – A total of eight, twelve and three sensors, for two, four and one instrumented housings are considered in B1, B2 and B3, respectively.

out of fourteen.

The substantial role of the window opening width is consistent. Comparing a small opening and a fully opened window, for the same duration, the heat loss in winter is significantly higher. Obviously, it is not realistic to account for all windows to be fully opened (80%) when the sensors detect a window opening. The opening width should vary, depending on a variety of parameters such as the purpose of the natural ventilation, the related activity from the occupant (ventilation after the night, cooking), climatic conditions (outdoor temperature, humidity or wind) or even social and cultural backgrounds [8]. Hence, it is of paramount importance to take this parameter into account and to find a way to either collect related data or to assess a representative averaged opening width.

However, it should be notified that these simulations have a main limitation. In BEM, IAT is set as the temperature setpoint in the different thermal zones. If a window is opened and there is a heat loss, the heating energy system compensates the heat loss to achieve the temperature setpoint. In the case study, there is no thermostat in housings. Therefore, if a window is opened, the system does not take this action into account and the heat demand is not impacted. However, this also means that measured temperatures in instrumented housings already account for window opening in the living rooms, which reinforces the questions regarding the actual temperature setpoints and the overall heating energy management in the studied buildings.

4.2.3.3 Conclusions on the performance gap and calibration

We addressed the manual calibration of BEM with the tuning of different input parameters. Energy-driving parameters identified through sensitivity analyses were selected. Those depicting the building operation and for which data were collected through the sensor network were modified to fit to modeling operation scenarios for the instrumented housings. The calibration was manually performed input by input in five steps.

The tuning of individual parameters confirmed the predicted impact from sensitivity analyses. Occupancy and dissipated electric power had a low impact while IAT, equated to temperature setpoints, resulted in a significant increase of heating energy consumption. We also observed that a change in occupancy and dissipated electric power to field-measured data increased heating energy consumption, meaning that in the instrumented sample, less heat gains than expected were generated.

We also proposed to extrapolate the tuning of IAT for non-instrumented housings using data collected in other housings, which finally resulted in a significant final energy

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

PG. For B2, the gap went from -30% to +24% after calibration and extrapolation. For B3, it was even larger, going from -30% to +39% before and after tuning. Only B1 really got closer to measurements, as the gap was reduced from -54% to -19% after calibration.

Several reasons may explain the large remaining PG. First of all, most parameter tuning was only performed for instrumented housings, which relates to two, four and two apartments in B1, B2 and B3, respectively. Occupancy had little observed impact but results showed that the tuning of dissipated power had slightly more effect. For considered housings, dissipated power tuning ended up increasing heating energy consumption due to lower internal heat gains. It may also end up decreasing the energy consumption, if some apartments have and use more electrical appliances. This parameter and its impact over the model remain quite uncertain.

A second significant aspect resides in the choice of IAT scenarios for non-instrumented housing. Measured IAT were assumed to be equal to temperature set points, that are the input data scenarios used in energy models. Using the operative temperature instead, to take into account the temperature of cold walls could be more realistic as a temperature set point to take into account the occupants' comfort. Then, for each non-instrumented household, we made a strong assumption using the average temperatures over the heating season in instrumented apartments from the same building, with 22.2°C, 22.7°C and 24.3°C for B1, B2 and B3 respectively. However, these values are questionable and for several reasons. There is significant range of temperature in apartments, from 21.4°C to 24.9°C in average in B1/2 and B3/0, while the instrumented sample with available data accounts for less than 10% of the whole housing stock of the case study. If the heating patterns are expected to be similar, the temperature level may significantly differ from one apartment to the other. Also, averaged temperatures for each building increase from the smaller to the larger building. Considering that an overall 2°C variation of IAT increased the heating energy consumption in the buildings by up to 55.2% to 68.3%, a small discrepancy in IAT for non-instrumented housings may significantly impact on the model output and the final energy PG.

A number of other modeling uncertainties remain. Most information regarding the building envelope and the energy systems are extracted from the retrofit portfolio. These documents are supposed to be the accurate summary of what was implemented onsite. Nevertheless, there were several missing information regarding the characteristics of some elements of the buildings such as the heating and DHW systems. There are also uncertainties in the measurements and the sensor network set up. Assumptions were made to complete the missing data from heating energy meters. These meters ended up being

difficult to successfully operate all along the instrumentation period. IAT could also be discussed on several critical points, including the location and number of the sensors in the housing, with only one sensor installed in the living room.

Furthermore, it should not be forgotten that non-negligible origins of the energy PG are often observed in the construction process, as discussed in Chapter 1. The quality of the retrofit process and the accuracy of the documentation was not verified nor compared, while the inputs related to the description of building energy systems and the envelope have a non-negligible cumulative impact on heating energy consumption. Among these parameters, the first-ranking is the dimension of walls. Although the BEM have been checked several times, small modeling errors cannot be excluded.

4.2.4 Practical conclusions on heating energy performances for the studied buildings

In the present chapter, we addressed the study of the energy PG through building energy modeling, the calibration of models and the comparison with measurements from our sensor network. This study raised several practical insights regarding the operation of the considered building case study, that are summarized in this section.

Sensitivity analyses highlighted the most energy-driving input parameters of the BEM with respect to the heating energy consumption. In other words, these parameters relate existing building features, whose improvement or optimization can significantly reduce heating energy consumption. Specifically, the thermal characteristics of the walls after retrofit still has a non-negligible effect. Considering that the envelope of buildings was not insulated prior to the retrofit, this action was definitely required.

In terms of energy consumption, it seems that the retrofit of the buildings achieved the targeted energy efficiency. The regulatory study that was conducted before the retrofit of the considered buildings predicted a heating energy consumption of 67,136 kWh, 96,244 kWh and 110,586 kWh for B1, B2 and B3, respectively, over a typical heating season. Measurements and calculations described in Section 4.2.1 showed a heating energy consumption of 48,405 kWh, 78,354 kWh and 110,468 kWh for B1, B2 and B3, respectively. Hence, B1 and B2 apparently exceeded heating energy consumption goals, while B3 met the objective, considering the assumptions that were used in the calculations of the heating energy consumption for B1 and B2.

However, there are still several lines of actions to further improve heating energy efficiency. Specifically, IAT measurements in instrumented apartments were much higher

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

than a standard 19°C, as they ranged from 21.4°C to 24.9°C in average over the heating season 2021/2022. This large 3.5°C-difference between two apartments, here B1/2 and B3/0, also highlighted the diversity of IAT levels in the instrumented sample. Hence, it would be reasonable to think that non-instrumented housings may also show a diversity of indoor temperature levels within a similar range. Furthermore, it was observed that there is currently no water law for heating management. IAT measurements are constant over day and night. These insights question the current heating energy management.

Several strategies could be considered. First, IAT should be reduced to a reasonable level. It would induce a significant decrease of the heating energy consumption in the three buildings. If the operative temperature – i.e. the average between the indoor temperature of cold walls and the air temperature – is considered as the heating temperature set point, it should still result in lower IAT in housings, since the building envelope was insulated during the retrofit. Temperature measurements in different location within the buildings could also be implemented, eventually with a thermostat to help optimizing heating energy management. Even after the renovation, there are no temperature monitoring points in the buildings to provide a feedback on the performances of the heating energy system. Water laws would complement these strategies to reduce heating over night periods and eventually fit to other specific heating needs in the three studied buildings. It is necessary to highlight that, in that case, the repetitive triggering of boilers to compensate the temperature decrease after the night could result in a higher energy consumption of the boilers and higher electricity consumption from the hot water pumps, compared to a constant temperature set point over the day. However, combined with the decrease of IAT, as suggested above, we would still expect to observe a significant decrease of the heating energy consumption.

Finally, the study of window opening depicts several trends as well, specifically over the summer period. We found out that in most apartments and rooms, windows are opened for a significantly longer time during the summer months (between May and September, depending on the housings) compared to other times of the year. There could be several reasons for this trend. Specifically, as many of the inhabitants, and specifically the occupants of instrumented housings, have been living in the buildings for many years prior to the energy retrofit, they may be used to open the windows during the summer because of the poor thermal insulation of the buildings before our research project. This is related to the rebound effect, as it may take up to several years for inhabitants of retrofitted buildings to adapt to their new environment and modify their habits [10]. The thermal insulation and ventilation of housings would need further checking as well. It is indeed very common to find summer overheating after the retrofit of residential buildings

[11] and that would drive on window opening. It may be related to unsuitable ventilation strategies or a choice of insulation materials that is adapted to meet the requirements in terms of winter heating energy savings but are unsuited for summer comfort perspectives. This also confirms that summer comfort in the buildings of the case study should be investigated in future works.

Conclusions

The energy performance gap is a major challenge in building energy modeling. To study and intend to close the performance gap after retrofit actions, a three-step strategy was proposed with the implementation of non-calibrated energy models as for a TRS, the identification of the main energy-driving inputs of the models, and the tuning of these inputs in NC-BEM using field data collected through our wireless sensor network.

As expected, NC-BEM resulted in a significant PG compared with measurements. An intuition was leading to the IAT as one of the main reasons for this gap, considering the heating centralized energy system and the very high IAT measurements observed in instrumented housings. This assumption was confirmed by sensitivity analyses that underlined the impact of IAT as the most energy-driving input. Other operation scenarios including occupancy and dissipated power were considered less impacting. Due to the weight of the IAT, it was observed that for a 10% variation of the values of input parameters, scenarios of building operation could result in up to more than three times the effect of a similar modification for parameters describing the building envelope and the energy systems.

The tuning of parameters was manually conducted input by input: IAT, occupancy and dissipated power alone, then all together, and complemented with an extrapolation towards more realistic IAT levels in non-instrumented housings. Although the individual tuning of inputs had a logical effect, the extrapolation of IAT was not fully successful. It raised several questions on the remaining uncertainties in the BEM. Part of these uncertainties relate to heating energy metering for the comparative study. Due to the defect of some heating energy meters, several assumptions were made, and considering the difficulties in maintaining the thermal energy metering solution over the time of our project, even the available measurements could be questioned. Then, on the other side, other uncertainties remain for the buildings' features, the modeling process and the retrofit work that was conducted. Retrospectively, these could be investigated with a different strategy for the instrumentation and the building energy modeling processes. A larger housing sample

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

but with less diversity in the measurements could be a relevant option to ensure an easier management of the data collection and to target specific parameters. A larger number of IAT measurement points in instrumented apartments should be considered. Also, IAT could be replaced by the operative temperature, taking into account data collected for the indoor temperature of cold walls.

These observations are also shared in the study of window opening. Collected data provided an opening duration. However, the opening width was not monitored and it was concluded as a substantial energy-driving parameter. A monitoring strategy would be required to collect related data.

It is necessary to underline as well a limitation of the sensitivity analysis performed in our study, with "one-at-a-time" method. As energy models distinctively consider operation scenarios, it is expected that most of these parameters, such as the occupancy, the dissipated power and window opening may be related and not independent from each other. This should be further explored using more adapted global sensitivity analyses techniques.

Nevertheless, for the same reasons than the calibration process was not entirely successful, it also offered relevant insights, specifically regarding the current heating energy management and on the improvements that would be necessary for an even better performing sensor network. Furthermore, the calibration strategy still remains relevant, and it is expected to provide better performances with slightly more available data and a few less uncertainties on key building features.

Finally, there is a large pool of opportunities to explore regarding the characterization of the energy consumption and the impact of the renovation on the considered retrofitted buildings. Although it is the main energy end-use, only the heating energy consumption was studied. Other significant end-uses should be further investigated, including domestic hot water consumption and usage of electric appliances. Furthermore, so far, the study has been conducted at building scale. Many housings did not participate in the instrumentation campaign and it resulted in the difficulties reported above regarding the calibration of building energy models. Hence, a study at the scale of apartments could be considered to explore a smaller spatial granularity, and try and characterize different behaviors and energy demand patterns based on the specificity of the occupants. Finally, there is also the question of the summer thermal comfort, a highly topical issue with respect to the requirements of the French Environmental Regulation, that should be taken into account, and on which window opening patterns, as an example of the occupants' behavior, would be expected to play an important role as well.

Bibliography

- [1] ADEME, *Climat, air et énergie — Chiffres clés*. 2018. 142
- [2] IZUBA, “Pleiades — Des outils logiciels pour l’optimisation énergétique et environnementale des bâtiments.” <https://www.izuba.fr/logiciels/outils-logiciels/>. 143
- [3] D. Coakley, P. Raftery, and M. Keane, “A review of methods to match building energy simulation models to measured data,” *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 123–141, sep 2014. 144
- [4] E. Fabrizio and V. Monetti, “Methodologies and Advancements in the Calibration of Building Energy Models,” *Energies 2015, Vol. 8, Pages 2548-2574*, vol. 8, pp. 2548–2574, mar 2015. 144
- [5] M. H. Kristensen and S. Petersen, “Choosing the appropriate sensitivity analysis method for building energy model-based investigations,” *Energy and Buildings*, vol. 130, pp. 166–176, oct 2016. 146
- [6] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, and G. Augenbroe, “A review of uncertainty analysis in building energy assessment,” *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 285–301, oct 2018. 146
- [7] V. Masatin, E. Latõšev, and A. Volkova, “Evaluation Factor for District Heating Network Heat Loss with Respect to Network Geometry,” *Energy Procedia*, vol. 95, pp. 279–285, 2016. 154
- [8] 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*, vol. 58, pp. 188–198, dec 2012. 161, 171
- [9] X. Dai, J. Liu, and X. Zhang, “A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings,” *Energy and Buildings*, vol. 223, p. 110159, sep 2020. 161
- [10] R. Galvin, “Making the ‘rebound effect’ more useful for performance evaluation of thermal retrofits of existing homes: Defining the ‘energy savings deficit’ and the ‘energy performance gap’,” *Energy and Buildings*, vol. 69, pp. 515–524, feb 2014. 174
- [11] K. J. Lomas and S. M. Porritt, “Overheating in buildings: lessons from research,” *Building Research and Information*, vol. 45, no. 1-2, pp. 1–18, 2017. 175

CHAPTER 4 – ENERGY MODELING, CALIBRATION AND PERFORMANCE GAP

General conclusions

The present thesis originates from the ANDRE research project funded by CAMEO SAS, a company specializing in energy saving certification, the landlord *Marne-et-Chantereine Habitat*, and researchers from Université Gustave Eiffel, with laboratories ESYCOM and COSYS LISIS focusing on building energy modeling and calibration, Lab'URBA for social sciences and GRETTIA for data sciences. The ANDRE project aimed to investigate a major topic, taking into account the current challenges in building energy efficiency. Energy retrofit is indeed one of the main line of actions to reduce building energy consumption. Considering the age and state of the existing housing stock in France, many buildings would need to improve their overall environmental footprint. Consequently, landlords are highly encouraged to implement ambitious energy retrofit actions over their building stock but the results of the renovation are at stake. It is common to observe a performance gap, which characterizes the difference between the expected results of the thermal renovation and the actual energy performance of buildings resulting from the retrofit. With this in mind, the landlord *Marne-et-Chantereine Habitat* provided three existing social residential buildings from its stock, as a case study to investigate building energy performance in a retrofit context, using a multi-field approach. This manuscript discussed the research work conducted on data collection and analyses from a wireless sensor network, to improve the calibration of building energy models.

In the introducing chapter, we provided an overview of building energy modeling as a topic. The most implemented methods for building energy modeling were presented to choose the fittest approach for our study, and to guide strategies to bridge the energy performance gap. Data-driven approaches are among the most popular techniques nowadays, due to the increasing amount of collected field data on many aspects of the building energy performance. However, the need for a highly interpretable and detailed energy model to investigate implemented retrofit actions led us to choose a more conventional physics-based approach. The review of the literature highlighted many potential origins to the energy performance gap, along with adapted strategies to close it. With respect to

GENERAL CONCLUSIONS

the case study, we decided to investigate building post-occupancy as a source of modeling uncertainties that could result in an energy performance gap. From this point of view, it was highlighted that a better understanding of building operation to calibrate energy models would be the adapted strategy.

Hence, the first step would be to collect field data to characterize post-occupancy building operation. The case study supporting this research is a three-building complex of sixty-three social housings, built in 1974. Deep energy retrofit actions were conducted from the summer 2020 to the summer 2021 in the three buildings. The data collection was performed using a wireless sensor that was deployed onsite. It comprised 170 sensors distributed in an eight-apartment sample, in common areas and on energy systems, and complemented by a weather station installed on the nearby university campus. At building scale, data were collected over three years. In housings, data were collected over one year and half. Aside from the large number of sensors, a specificity of the monitoring solution was the targets of the instrumentation: energy consumption, indoor environment quality, usages and local weather. The review of the literature on similar instrumentation solutions for building energy monitoring showed that equivalent or more performing set ups would mostly be industrial – with consecutive human and financial means – or relying on building energy management systems. As a matter of fact, long-term management of monitoring solutions is a complex task and such a sensor network inevitably faces technical challenges. As a conclusion to the management of the sensor network over the full term of the project, it was clear that there is room for improvements regarding the measurements, the choice of monitoring technologies, and the automation of the network supervision. Nevertheless, our instrumentation solution offered a relevant feedback on wireless data acquisition, that is a hot topic for building efficiency applications. It also plainly filled its first and main objective: to provide a large amount of diverse field data to explore building energy operation and help calibrating building energy models.

For field data analyses, we focused on five parameters related to the energy consumption, the indoor environment quality, and the occupants' behavior in instrumented housings: indoor air temperature, occupancy, window opening, domestic hot water use and electric dissipated power. Data pre-processing highlighted that raw measurements were not necessarily optimal in terms of analyses. They required modifications, such as for occupancy detection and domestic hot water use. Nevertheless, passed these constraints, the analyses of selected parameters provided relevant insights. For each parameter, different patterns were observed. These patterns were apartment-specific. A major finding related to indoor air temperature and heating patterns. As newly retrofitted buildings, it would have been expected to find reasonable temperature levels and a close-to-optimised heat-

GENERAL CONCLUSIONS

ing energy management. On the contrary, measured air temperature was largely above standards and there was no apparent heating management over day, night, weekdays and weekends in the buildings, which would save a significant amount of energy. Furthermore, the diversity of operation profiles observed from collected data questioned the reliability of standard scenarios that proved to be very different from the reality. Although, it may be quite difficult to use such an instrumentation campaign for every retrofit project, this could lead a middle ground with different perspectives: small instrumentation batches prior to the retrofit to collect dedicated field data, case-study-specific standard datasets with more diversity in building operations and/or customizable profiles based on easy-to-implement measurements such as the occupancy, that drives most of indoor activity.

We used the findings from field data analyses over the instrumented sample to replace energy modeling assumptions. We focused on heating energy consumption after the retrofit, as the main energy end-use in apartment buildings. The modeling and calibration strategy was divided into three steps. At first, non-calibrated energy models of the retrofitted buildings were implemented based on the retrofit portfolio and standard data for the investigation of the initial performance gap, and the identification of energy-driving inputs, using sensitivity analyses. Once the energy drivers were highlighted, corresponding data analysis results were integrated to building energy models. They were then used to tune energy-driving inputs and to calibrate the building energy models. As expected, the non-calibrated building energy models exhibited a large gap compared to measurements. Sensitivity analyses clearly depicted the impact of one specific input parameter: the indoor air temperature. Along with other operation schedules, including the occupancy and the dissipated electric power, these inputs were then tuned one-by-one to observe their respective effect, and then together. The extrapolation of indoor air temperature tuning to non-instrumented housings was also performed. However, it resulted in an almost equivalent energy performance gap with over-heating energy consumption for two buildings.

Consequently, limitations were underlined in the overall data collection, the building energy modeling and the calibration strategy. Coming back to the beginning of the instrumentation process, due to budget and enrollment restrictions, it was decided to limit the number of participating housings and to increase the number of sensors per apartment. Also, the data collection started much earlier than the modeling process for obvious reasons, such as the need of a significantly large amount of data and project management constraints – both the instrumentation and the energy modeling were conducted by our research team. From a data analysis point-of-view, and considering that collected data would also be used by other members of the ANDRE project for their respective research

GENERAL CONCLUSIONS

goals, this strategy was meaningful.

For building energy modeling applications, the strategy could have been implemented differently. A few potential ideas would be:

- To target a larger housing sample with less sensors but specifically selected measurements, with reinforced monitoring to avoid extensive data loss and long-term management issues;
- To start the energy modeling process at the same time as the instrumentation: a simplified energy model with sensitivity analyses could highlight energy-driving input parameters to focus on and modifications could be implemented in the sensor network to comply with the needs of the energy models;
- To include the construction party in the research project to ensure there would be less missing information regarding the building features.

Despite many possible improvements, the research project and its results presented in this manuscript pave the ground for further research on building energy modeling calibration complemented with the use of wireless sensor networks. Our study focused on heating energy consumption and highlighted the related most energy-driving parameters to monitor: indoor air temperature, occupancy, dissipated energy and window opening. We contemplate the idea that the study of other building energy end-uses based on collected field data would help designing customized end-use-related instrumentation packages for practical field studies and the efficient characterization of building energy consumption. Consequently, there is a significant amount of data that could not be analyzed over the time of this thesis. We expect that cross-analyses would lead to a deeper understanding of occupants' behavior and its impact on energy consumption. Future works should focus on the investigation of the detailed electrical energy consumption from connected plugs and sensors with clamp ampere meters, as well as the characterization of the summer thermal comfort and the indoor air quality with cold wall temperatures, relative humidity and CO₂ concentration. All these different parameters could provide relevant insights to contribute to the improvement of an efficient energy retrofit with the integration of occupants' behaviors. Finally, this work remains to be integrated to the research conducted by the other members of the ANDRE project. The combination of instrumentation, building physics, social sciences and data sciences is a novel approach. Advanced data analytic and data-driven energy models could support data analysis goals listed above. Moreover, our findings stressed how each different housing exhibited unique patterns of occupancy, dissipated power, domestic hot water use and window opening. Some patterns could be related to characteristics of the occupants of instrumented housings. However, a dedicated

GENERAL CONCLUSIONS

socioeconomic approach could bring other perspectives on the link between socioeconomic aspects of housings and energy-driving behaviors.

GENERAL CONCLUSIONS

Appendix A

Tools for building energy modeling

Building energy models are implemented through specific tools and software. This appendix offers a summary of these tools for physics-based and data-driven energy modeling. Hybrid models rely on the combination of tools from both categories.

A.1 Physics-based energy modeling

Physics-based modeling mostly depends on commercial and academic research software. It exists a significant variety of software from different countries, companies and research institutes with various embedded functions. Several reviews have been edited on the topic [1, 2]. The BEST (Building Energy Software Tools) directory also offers among the most complete overview of available tools [3]. From an initial list of seventy-two reported modeling tools, we focused on the fourteen most commonly used software in the literature and for industrial applications. Table A.1 focuses on the technical features of the modeling software, Table A.2 provides an overview of the types of buildings that can be modeled with the software, and Table A.3 summarizes the types of modeling outputs that can be expected.

APPENDIX A – TOOLS FOR BUILDING ENERGY MODELING

Table A.1: Summary of software for physics-based energy modeling – Technical characteristics.

Name	Free / Licence	Company – Country	Calculation	Programming compatibility	Standards compliance	Geometry input
ClimaWin	Licence	BBS Logiciels – France	Climawin	N.C.	RE2020, RT Ex	Information list / Floor plans / 3D-model
Pleiades	Licence	IZUBA – France	Pleiades	N.C.	RE2020, RT Ex, ACV C-	Information list and floor plans
Design Builder	Licence	DesignBuilder Software Ltd – UK	EnergyPlus, Radiance	N.C.	ASHRAE 90.1, LEED EAp2 and EAc1	Information list and 3D-model
e-Design HAP	Licence	Carrier – USA	Carrier HAP	N.C.	LEED, ASHRAE	Information list
EnergyPlus	Free	U.S. Department of Energy's (DOE) Building Technologies Office (BTO) – USA	EnergyPlus	C++, Python	LEED, ASHRAE	Information list
eQuest / DOE-2	Free	Lawrence Berkeley National Laboratory	DOE-2	None	ASHRAE	Information list and floor plans
ESP-r	Free	University of Strathclyde – UK	ESP-r	FORTRAN, C	BREEAM	Information list and floor plans
IES Virtual Environment	Licence	Integrated Environmental Solutions Ltd – USA	Apache dynamic thermal simulation engine	Python	LEED, BREEAM, GreenStar	Information list and 3D-model
Matlab Simulink	Licence	Matlab	Matlab Simulink	Matlab	None	Information list
Modelica Buildings Library	Free	Lawrence Berkeley National Laboratory / University of California – USA	Modelica	Modelica	None	Information list
Open Studio	Licence	Alliance for Sustainable Energy, LLC – USA	EnergyPlus, ESP-r, Radiance	Python, C++, Ruby, C#, JavaScript	ASHRAE	Information list and 3D-model
Revit Insight360	Licence	Autodesk – USA	Revit	Python, C#	N.C.	Information list and 3D-model
TAS	Licence	Environmental Design Solutions Ltd – USA	Energyplus	N.C.	LEED, ASHRAE 90.1, BREEAM	Information list and floor plans
TRNSYS	Licence	Thermal Energy System Specialists, LLC / CSTB – USA	TRNSYS	FORTRAN, C, C++, Excel/VBA, Matlab, EES	N.C.	Information list

A.1 Physics-based energy modeling

Table A.2: Summary of software for physics-based energy modeling – Applications: case studies.

Name	Energy systems	Office	Commercial	Industrial	Other tertiary sector	Residential	District-scale
ClimaWin		X	X			X	
Pleiades		X	X		X	X	
Design Builder		X	X	X	X		
e-Design HAP	X	X	X			X	
EnergyPlus		X	X		X	X	
eQuest / DOE-2		X	X		X	X	
ESP-r		X	X		X	X	
IES Virtual Environment		X	X		X	X	
Matlab Simulink	X	X	X	X	X	X	X
Modelica Buildings Library	X	X	X	X	X	X	X
Open Studio		X	X		X	X	
Revit Insight360		X	X		X	X	
TAS		X	X		X	X	
TRNSYS	X	X	X	X	X	X	X

APPENDIX A – TOOLS FOR BUILDING ENERGY MODELING

Table A.3: Summary of software for physics-based energy modeling – Applications: modeling outputs.

Name	Detailed dynamic thermal simulation / dynamic energy simulation	HVAC	Day-lighting	Indoor comfort	Computational fluid dynamics	Life cycle analysis	Parametric analysis Optimization	Code compliance	Audits and reports	BIM
ClimaWin	X							X	X	X
Pleiades	X		X	X		X	X	X	X	X
Design Builder	X	X	X		X		X	X	X	X
e-Design HAP		X								
EnergyPlus	X	X						X		
eQuest / DOE-2	X		X					X		
ESP-r	X		X				X			
IES Virtual Environment	X	X	X	X				X	X	X
Matlab Simulink	X	X					X			
Modelica Buildings Library	X	X					X			
Open Studio	X	X	X				X	X		
Revit Insight360	X	X	X			X	X			X
TAS	X	X	X							X
TRNSYS	X	X					X			

A.2 Data-driven energy modeling

The recent research interest and popularity of data-driven BEM is significantly supported by the accessibility of many opens-source tools. Most tools are coding platforms that benefit from several pre-coded packages. Non-expert users can easily implement many algorithms, although the tuning of these algorithms still requires extensive knowledge on the topic. The following Table A.4 summarizes the most encountered tools for data-driven modeling, along with popular packages and implemented data-driven techniques.

Table A.4: Summary of coding tools, packages and applications for data-driven energy modeling.

Tool	Available packages	Applications
IBM SPSS Statistics [4]	/	Autoregressive models [5]
IBM SPSS Statistics [4]	/	Autoregressive models [5]
Python	Statsmodel [6]	Statistical regressions [7]
	Scikit-Learn [8]	SVR, statistical regressions [7, 9], ensemble models [10], DT [11]
	NeuroLab [12]	ANN [10]
	Tensorflow [13]	ANN, DNN [14]
	Keras [15]	Deep learning [16]
Weka Software [17]	/	k-NN [18], DT [19], statistical regressions, ANN, SVR [20]
Matlab [21]	/	ANN [22], DNN [23], ensemble models [24], unsupervised and reinforcement learning [25]
	LibSVM [26], FarutoUltimate [27]	SVR [14, 28, 29]
	Neural Network Toolbox	ANN [30, 31]
mySVM [32]	/	SVR [22]
R [33]	/	SVR [34]

Bibliography

- [1] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, “Physical energy and data-driven models in building energy prediction: A review,” *Energy Reports*, vol. 8, pp. 2656–2671, nov 2022. 185
- [2] ASSOCIATION APOGEE, “Revue pratique des logiciels de simulation énergétique dynamique (SED),” 2015. 185
- [3] Best Directory, “Best Directory — Building Energy Software Tools.” <https://www.buildingenergysoftwaretools.com/>, 2022. 185
- [4] IBM, “IBM SPSS Statistics.” <https://www.ibm.com/products/spss-statistics>. 189
- [5] G. R. Newsham, B. J. Birt, C. Arsenault, L. Thompson, J. Veitch, S. Mancini, A. Galasiu, I. Macdonald, and G. Burns, “Do green buildings outperform conventional buildings? Indoor environment and energy performance in North American offices (Report RR-329),” tech. rep., 2012. 189
- [6] J. Perktold, S. Seabold, and J. Taylor, “StatsModels: Statistics in Python — statsmodels 0.9.0 documentation.” <https://www.statsmodels.org/stable/index.html>. 189
- [7] P. Dagnely, T. Ruelle, T. Tourwé, and E. Tsiporkova, “Predicting hourly energy consumption. Can you beat an autoregressive model?,” *Proceeding of the 24th Annual Machine Learning Conference of Belgium and the Netherlands, Benelearn, Delft, The Netherlands*, 2015. 189
- [8] Scikit-Learn, “Machine learning in Python.” <https://scikit-learn.org/stable/>. 189
- [9] C. Zhang, Y. Zhao, X. Zhang, C. Fan, and T. Li, “An Improved Cooling Load Prediction Method for Buildings with the Estimation of Prediction Intervals,” *Procedia Engineering*, vol. 205, pp. 2422–2428, jan 2017. 189
- [10] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy and Buildings*, vol. 147, pp. 77–89, jul 2017. 189
- [11] S. Papadopoulos, E. Azar, W.-L. Woon, and C. E. Kontokosta, “Evaluation of tree-based ensemble learning algorithms for building energy performance estimation,” *Journal of Building Performance Simulation*, pp. 1–11, jul 2017. 189
- [12] Université du Québec à Montréal, “NeuroLab 0.3.5, Neural Network Library for Python.” <https://pythonhosted.org/neurolab/>. 189

BIBLIOGRAPHY

- [13] TensorFlow, “TensorFlow.” <https://www.tensorflow.org/>. 189
- [14] K. P. Amber, R. Ahmad, M. W. Aslam, A. Kousar, M. Usman, and M. S. Khan, “Intelligent techniques for forecasting electricity consumption of buildings,” *Energy*, vol. 157, pp. 886–893, 2018. 189
- [15] Keras, “Keras Documentation.” <https://keras.io/>. 189
- [16] C. Fan, J. Wang, W. Gang, and S. Li, “Assessment of deep recurrent neural network-based strategies for short-term building energy predictions,” *Applied Energy*, vol. 236, no. July 2018, pp. 700–710, 2019. 189
- [17] University of Waikato, “Weka 3 - Data Mining with Open Source Machine Learning Software in Java.” <https://www.cs.waikato.ac.nz/ml/weka/index.html>. 189
- [18] O. Valgaev and F. Kupzog, “Building power demand forecasting using K-nearest neighbors model - initial approach,” in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1055–1060, IEEE, oct 2016. 189
- [19] Z. Yu, F. Haghghat, B. C. Fung, and H. Yoshino, “A decision tree method for building energy demand modeling,” *Energy and Buildings*, vol. 42, pp. 1637–1646, oct 2010. 189
- [20] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, “Short-term load forecasting in a non-residential building contrasting models and attributes,” *Energy and Buildings*, vol. 92, pp. 322–330, apr 2015. 189
- [21] Mathworks, “MATLAB - MathWorks - MATLAB Simulink.” <https://www.mathworks.com/products/matlab.html>, 2022. 189
- [22] X. Li, J.-h. Lu, L. Ding, G. Xu, and J. Li, “Building Cooling Load Forecasting Model Based on LS-SVM,” in *2009 Asia-Pacific Conference on Information Processing*, pp. 55–58, IEEE, jul 2009. 189
- [23] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, “Deep learning for estimating building energy consumption,” *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016. 189
- [24] Z. Wang, Y. Wang, and R. S. Srinivasan, “A novel ensemble learning approach to support building energy use prediction,” *Energy and Buildings*, vol. 159, pp. 109–122, jan 2018. 189
- [25] E. Mocanu, P. H. Nguyen, W. L. Kling, and M. Gibescu, “Unsupervised energy prediction in a Smart Grid context using reinforcement cross building transfer learning,” *Energy and Buildings*, vol. 116, pp. 646–655, mar 2016. 189

BIBLIOGRAPHY

- [26] C. chung Chang and C. jen Lin, “LIBSVM - A Library for Support Vector Machines,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, pp. 1–39, 2013. 189
- [27] Y. Li, “LIBSVM-FarutoUltimate: a toolbox with implements for support vector machines based on libsvm,” 2011. 189
- [28] S. Paudel, M. Elmitri, S. Couturier, P. H. Nguyen, R. Kamphuis, B. Lacarrière, and O. Le Corre, “A relevant data selection method for energy consumption prediction of low energy building based on support vector machine,” *Energy and Buildings*, vol. 138, pp. 240–256, mar 2017. 189
- [29] D. Liu, Q. Chen, and K. Mori, “Time series forecasting method of building energy consumption using support vector regression,” in *2015 IEEE International Conference on Information and Automation*, pp. 1628–1632, IEEE, aug 2015. 189
- [30] A. Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, and A. Vinci, “Electrical consumption forecasting in hospital facilities: An application case,” *Energy and Buildings*, vol. 103, pp. 261–270, sep 2015. 189
- [31] M. A. Biswas, M. D. Robinson, and N. Fumo, “Prediction of residential building energy consumption - A neural network approach,” *Energy*, vol. 117, pp. 84–92, 2016. 189
- [32] Technical University of Dortmund, “mySVM - TU Dortmund.” <https://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/index.html>, 2004. 189
- [33] The R Foundation, “The R Project for Statistical Computing.” <https://www.r-project.org/>, 2022. 189
- [34] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, “Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique,” *Energy and Buildings*, vol. 126, pp. 94–103, aug 2016. 189

Appendix B

Sensor network for building energy monitoring

B.1 Connected objects and communication protocols for building energy monitoring

This appendix provides additional context to the Chapter 2, specifically regarding wireless communication in our sensor network. It defines and discusses several concepts, including the Internet of Things, and LoRaWan and GPRS communication protocols.

B.1.1 The Internet of Things

National instrumentation plans, such as in Europe or the United States [1, 2, 3, 4] pushed for the large installation of smart energy meters. These meters have boosted the recent development of a complementing technology with a large range of small-size sensors, with low energy consumption and cost: the Internet of Things (IoT).

The Internet of Things is a part of the information and communication technologies (ICT). The first occurrence of this concept is recent and goes back to 2005, with the research of the Massachusetts Institute of Technology on networked radio-frequency identification [5, 6]. Nowadays, most IoT technologies are supported by radio frequencies.

A definition of IoT could refer the definition given by the International Telecommunication Union: "a global infrastructure for the Information Society, enabling advanced services by interconnecting (physical and virtual) things based on, existing and evolving,

interoperable information and communication technologies” [7]. However, this definition is very broad and applies to a diversity of fields including health, security, energy, agriculture, transportation or industries. To relate to our study, we define IoT as the network solution including the sensors and meters, the communication protocols and supporting equipment, and the data collection and storage solutions.

B.1.2 Communication protocols

B.1.2.1 Overview

The IoT market is supported by a variety of communication protocols. The most common for the instrumentation of buildings and resilient cities are:

- LoRaWan, a ”Low Power, Wide Area (LPWA) networking protocol designed to wirelessly connect battery operated ‘things’ to the internet in regional, national or global networks. It targets key Internet of Things (IoT) requirements such as bi-directional communication, end-to-end security, mobility and localization services” [8]. It is the most commonly implemented communication protocol for applications related to connected cities, industrial monitoring or agriculture [9, 10];
- Sigfox [11] is the competing network to LoRaWan and was designed in France. Unlike LoRaWan, it is a private network, which has several national operated opened networks. Sigfox is also common for smart cities and buildings instrumentation. Hence, many sensors can be operated by both technologies;
- ZigBee [12], another competing network with LoRaWan and Sigfox;
- MODBUS [13] differs from the previous communication protocols. It focuses on automated machine communication but is also frequently found on energy meters;
- Power line carriers traditionally rely on the electricity grid: data are simultaneously communicated along with electricity power. Although it is relevant for applications in remote places with low network connection, many issues have been reported regarding wireless data transmission disturbed by the nearby electric grid [14];
- M-Bus, standing for Meter-Bus, is a communication protocol usually implemented by utility providers to collect data from smart gas, heat or oil meters [15];
- BACnet, standing for building automation and control network, is, as the name suggests, a communication protocol for building automation applications, developed by ASHRAE [14];
- Wi-Fi is a common communication protocol for data communication with wide

B.1 Connected objects and communication protocols for building energy monitoring

application to access the Internet using computers and smart-phones. It can be used for wireless building instrumentation but has several drawbacks, including serious security issues and a significant cost because of a higher energy consumption compared with other protocols. It is usually preferred when the amount of data to transfer is too significant for other communication protocols because of live-transmission or high quality data requirements [14];

- GPRS (general packet radio service) is a famous communication protocol used in building metering systems when the instrumentation requirements can not comply with constraints of other networks, such as using a small data acquisition time-step.

In our sensor network, LoRaWan and GPRS protocols are implemented due to their availability on the IoT market and ease of implementation.

B.1.2.2 LoRaWan

B.1.2.2.1 LoRa, LoRaWAN and LoRa Alliance

LoRaWan means long rang wide area network. This communication protocol targets data communication between connected objects [16]. LoRa is the supporting peer-to-peer communication technology. Historically, LoRa technology was developed from 2012 by a French start-up company, Cycleo, and is now owned by Semtech company [17].

LoRaWan is developed by a non-profit association, the LoRa Alliance [8], to provide a standardized and inter-operable network. Among the deciding members, there are manufacturing companies such as Cisco, IBM or SagemCom, and network operators including Bouygues Telecom with its LoRa subsidiary, Objenious.

B.1.2.2.2 Technical characteristics

LoRaWan protocol targets long-distance data communication with low energy consumption. It depends on free radio frequencies – 868 MHz in France – and the Internet for a small data throughput (0.3 to 50 Kbps – smaller than for 2G communication). This results in the low energy consumption of communicating devices, which is a significant advantage considering they may be located in remote places and powered with batteries.

Another aspect of LoRaWan protocol is the two-way communication with connected objects: data can be collected and sent to a sensor. This is a useful feature to check on communicating systems and to implement remote activation of devices [9].

LoRaWan network architecture is shaped in a star-of-star configuration. A central server communicates with gateways that are used as data aggregators and communication

APPENDIX B – A SENSOR NETWORK FOR BUILDING ENERGY MONITORING

relays for sensors, and also to assess the LoRaWan network quality. Gateways communicate with sensors through the LoRaWan network. Between gateways and storage servers, the Internet is used.

Finally, the LoRaWan, unlike competing networks such as Sigfox, is an opened network. Any user can implement an instrumentation solution using LoRaWan, either with operated or private networks. Operated networks are managed by national operators. This option is usually preferred for the ease of use, because the operator manages the network implementation and maintenance, subject to a connectivity fee. Private networks have advantages over operated networks regarding security issues, network usage and for large amounts of sensors when connectivity cost is not optimal.

B.1.2.2.3 Benefits and drawbacks

LoRaWan is a widely implemented and easily accessible communication protocol for building energy monitoring. There is a significant theoretical range of applications with long range communication: 15 km for free field, 7 to 10 km for suburban areas, several hundred meters in dense urban areas and a dozen vertical floors in buildings. LoRa technology also provides other benefits. As described above, the energy consumption for data communication is also low, the communication protocol implements a two-way communication between servers and sensors and LoRaWan is an opened network.

In practice, a major drawback is the limited use of the operated LoRaWan network with a restriction of 1% of bandwidth usage at a time. It limits applications, specifically when a large number of sensors are set up with small data acquisition time-step. The obvious replacement solution is the private LoRaWan network. However, the initial cost is higher, it not a cost-effective solution for a small number of sensors, and it requires a day-to-day automated management since the network is managed by the user.

B.1.2.3 GPRS

GPRS (general packet radio service) is the second communication protocol used in our sensor network. It evolved from GSM network (Global System for Mobile communication). Unlike LoRa, GPRS is fully supported by the Internet and directly communicates with servers without gateways. Data throughput is also more significant than for LoRaWan. A significant advantage of GPRS is the absence of limitations regarding network and bandwidth use, which is why it is selected for electricity metering consumption that required one-minute data acquisition time-step.

B.1 Connected objects and communication protocols for building energy monitoring

B.1.2.4 Discussion: the recent evolution and the impact of IoT communication protocols

IoT connected objects for building energy monitoring are mostly supported by communication networks including LoRaWan, Sigfox and Zigbee, due to their many advantages over other solutions available on the market. Recent news highlighted upcoming changes over IoT communication protocols and also question their environmental impact.

National operator Objenious has announced in 2022 the final discontinuation of its LoRaWan operated network by 2024 [18]. It will be replaced with emerging networks such as LTE-M (long term evolution - machine) and NB-IoT (narrowband internet of things), two communication protocols using cellular network and adapted to low power wide area applications. Sigfox also faced changes with a judicial settlement that forced the Sigfox company to target inter-operability of its products with other communication networks [19].

These changes raise questions regarding the environmental impact of the IoT. Over our research project, we observed a major discrepancy between the goal behind IoT applications for energy efficiency and its practical use. In buildings, IoT is implemented to collect operation data, resulting in a better understanding and optimization of building energy consumption. The global objective is to reduce the environmental footprint of the building sector. However, the environmental footprint of the IoT solution should be questioned as well. We noticed that in case of a malfunction, a sensor was automatically replaced by our contractor. Eventually, the manufacturer would recover the sensor to try and understand the issue, but none of our malfunctioning sensors have ever been repaired to our knowledge, and the design of these sensors shows very little room for repairs anyway. A perfect example is the batteries that power the sensors. If a sensor is designed to be water and dust resistant, as many of the sensors we used, there is no mean for the user to change batteries when they are worn out. Sensors are simply replaced and hopefully recycled, but there is no available information on this topic. Furthermore, the needs of an easy and quick installation and access to data supports this design strategy. Considering the changes announced in IoT communication protocols, it is quite likely that most operating sensors will have to be replaced, at least for the oldest ones because they might not be compatible with the newest implemented communication protocols.

B.2 Calibration checking of IEQ sensors for hous- ings

The majority of installed sensors could not be calibrated by our research team before their installation. Calibration is performed upon manufacturing. However, a compensating solution is to check upon measurements in a known environment prior to their installation. Because of the tight installation schedule most sensors were directly installed on site by contractors. For the part of the instrumentation managed by our research team, similar reasons pushed for a quick installation as well. However, IEQ ELSYS CO₂ sensors could undergo a checking process.

ELSYS CO₂ sensors combined five measurements, among which four were dedicated to IEQ characterization – including IAT, relative humidity, luminosity and CO₂ concentration – and one to occupants’ behavior monitoring – presence detection [20]. The accuracy and operating range of the sensors can be found in Table 2.6. The resolution of the measurements is 0.1°C, 0.1%, 1 lux and 1ppm for temperature, humidity, luminosity and CO₂, respectively. The checking process was the following. Ten similar sensors with the same characteristics were placed together above a shelf in an office room in our university building. Measurements were performed over five days, including three working days and two days of weekend. Considering the location of the sensors, indoor conditions and measurements were expected to be very similar between the ten sensors. However, as it was an office space and conditions were not fully controlled, especially regarding the presence and number of occupants, occupancy detection was not considered for the calibration checking. Measurements are depicted in Figure B.1.

The results of the calibration checking highlight three types of insights. Temperature and humidity measurements are very similar among the ten sensors. The graphic analysis highlights the same variations over time. A reference sensor is randomly selected among the ten sensors – considering that the visual inspections show measurements are very close to each other. The difference of measurement for each time-step is calculated between the reference sensor and the remaining nine sensors. For temperature measurements, the average difference is of 0.06°C, with a standard deviation of 0.07°C and a maximum difference of 0.9°C over the five days of monitoring. For relative humidity the average difference is of 0.25% with a standard deviation and maximum difference of 0.45% and 2%, respectively. This confirms observations from temperature and humidity graphs.

Luminosity curves show that the sensors seem to detect similar events. However, the amplitude of luminosity peaks is very different. The sensor IEQ5 is selected as the

B.2 Calibration checking of IEQ sensors for housings

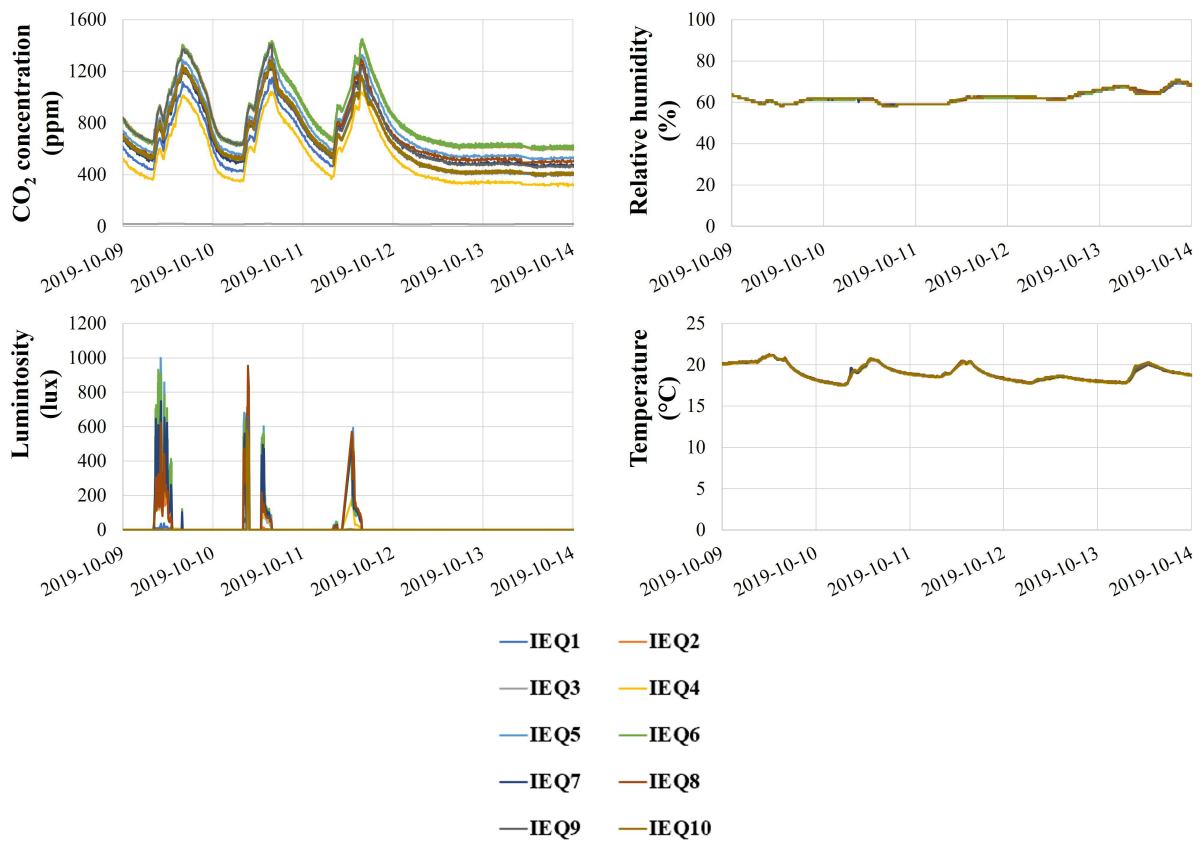


Figure B.1: Measurements from IEQ sensors over five days for calibration checking.

BIBLIOGRAPHY

reference sensor since it shows the highest luminosity peaks. The differences range from 6.6 ± 27.6 lux between IEQ5 and IEQ6 to 33.0 ± 140.2 lux with IEQ10. Such a large range of differences can find several explanations. Although the office space provides stable conditions, as described above, they are not fully controlled over time. Furthermore, due to a slightly different position on the shelf, sensors may not receive the same amount of light. Over the weekend there is no luminosity detected (curtains are closed). Hence, differences may originate from people opening the curtains, the door to the corridor (with lighting), or switching on and off desk lamps. The sensors might not be exposed in the exact same way to these light sources. Nevertheless, it is not possible to exclude a defect of luminosity sensing probes on specific sensors.

CO₂ concentration measurements highlight the necessity to check upon sensor calibration before installation. Variations over time are similar. However, there is a gap between the measurements of the different sensors. Considering the difference between two extreme sensors (IEQ4 and IEQ6, for instance), the averaged difference is of 312.3 ± 33.6 ppm, going up to 396 ppm. This gap is very significant, even with the 50-ppm accuracy of the sensors taken into account. Moreover, IEQ4 shows a CO₂ concentration under 400 ppm which is impossible, considering that the reference CO₂ concentration is around 415 ppm (outdoor CO₂ concentration). Unfortunately, at the time of the installation, there were no calibrated CO₂ meters available for a reference comparison. It was not possible either to send the sensors back to the manufacturer because of the tight installation schedule. However, in the sensor datasheet, the manufacturer specifies that these sensors have an internal automatic calibration routine: the routine "calibrates the sensor to set 400 ppm to the lowest value that has been read in the last period of approximately 8 days. This means that in an eight-day period, the sensor must be exposed to fresh (well ventilated) air at least once for the calibration to work". Then if discrepancies would still be observed in measurements – low CO₂ concentration values, for instance – they would be cross-analyzed with window opening. When a window would be opened for a significant amount of time, the CO₂ concentration in the apartment would be considered equal to the CO₂ concentration of the air outside. Values of measured CO₂ concentration would then be corrected accordingly.

Bibliography

- [1] X. Liu, C. Marnay, W. Feng, N. Zhou, and N. Karali, "A Review of the American Recovery and Reinvestment Act Smart Grid Projects and Their Implications for China," Tech. Rep. January, Lawrence Berkely National Laboratory, China Energy Group, 2017. 193
- [2] D. R. Obey, "Text - H.R.1 - 111th Congress (2009-2010): American Recovery and Rein-

BIBLIOGRAPHY

- vestment Act of 2009,” 2009. 193
- [3] European Commission — Joint Centre — Smart Electricity Systems and Interoperability, “Smart Metering deployment in the European Union.” <https://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union>, 2014. 193
- [4] U.S. Energy Information Administration (EIA), “How many smart meters are installed in the United States, and who has them?,” 2018. 193
- [5] F. Wortmann and K. Flüchter, “Internet of Things - Technology and Value Added,” *Business and Information Systems Engineering*, vol. 57, no. 3, pp. 221–224, 2015. 193
- [6] L. Atzori, A. Iera, and G. Morabito, “Understanding the Internet of Things - Definition, potentials, and societal role of a fast evolving paradigm,” *Ad Hoc Networks*, vol. 56, no. December, pp. 122–140, 2017. 193
- [7] TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU, “An overview of internet of things,” tech. rep., 2012. 194
- [8] LoRa Alliance, “Home page — LoRa Alliance™.” <https://lora-alliance.org/>, 2019. 194, 195
- [9] A. Augustin, J. Yi, T. Clausen, W. M. Townsley, A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, “A Study of LoRa: Long Range & Low Power Networks for the Internet of Things,” *Sensors*, vol. 16, p. 1466, sep 2016. 194, 195
- [10] G. Pasolini, C. Buratti, L. Feltrin, F. Zabini, C. De Castro, R. Verdone, and O. Andrisano, “Smart City Pilot Projects Using LoRa and IEEE802.15.4 Technologies,” *Sensors*, vol. 18, p. 1118, apr 2018. 194
- [11] Sigfox, “Sigfox - The Global Communications Service Provider for the Internet of Things (IoT),” 2019. 194
- [12] “Zigbee Alliance.” <https://zigbee.org/>. 194
- [13] The Modbus Organization, “The Modbus Organization.” <https://www.modbus.org/>, 2019. 194
- [14] M. W. Ahmad, M. Mourshed, D. Mundow, M. Sisinni, and Y. Rezgui, “Building energy metering and environmental monitoring – A state-of-the-art review and directions for future research,” *Energy and Buildings*, vol. 120, pp. 85–102, may 2016. 194, 195
- [15] M-Bus, “The Standard for Remote Reading of Smart Meters,” 2022. 194

BIBLIOGRAPHY

- [16] EBDS Wireless & Antennas, “La technologie LoRa,” tech. rep., 2016. 195
- [17] Semtech, “Semtech - Homepage.” <https://www.semtech.com/>, 2019. 195
- [18] Journal du Net, “Objenious arrête son réseau LoRaWAN,” 2022. 197
- [19] Journal du Net, “Sigfox : couverture et avenir du réseau,” 2022. 197
- [20] ELSYS, “ERS LoRaWAN room sensor for measuring indoor environment.” <https://www.elsys.se/en/ers/>, 2022. 198

Appendix C

Data processing and analyses

C.1 Data quality assessment

This appendix provides an overview of the performances of the sensor network in terms of data collection over the instrumentation period. Table C.1 summarizes data acquisition rates for each type of sensors in instrumented housings, in common areas, for energy systems and for the weather station. Acquisition rates are calculated for data collection between the installation date of each sensor and the May the 31st, 2022, selected as the final date for data analyses. Acquisition rates result from the comparison between the number of collected data points (after the processing of duplicates, but before other pre-processing steps) and the number of expected data points.

It is necessary to specify that a high data acquisition rate does not necessarily translate into a high data availability rate for data analyses. Some sensors may collect and transfer data but these data may not be exploitable. This relates specifically to contact temperature sensors for DHW pipes in housings, and for heating and DHW energy meters. For the former, temperature measurements can be acquired but if the measurements are under the temperature threshold, data cannot be considered for further analysis. A suspected explanation to this issue is that some occupants may have temporarily removed and wrongly re-located the temperature probe of the sensor, such as in B2/1 and B2/2. Heating and DHW energy metering is performed with an energy meter and a pulse counting sensor for data communication. It may happen that the meter malfunctions but that the pulse sensor keeps on sending data. Then collected data cannot be exploited in the analyses.

APPENDIX C – DATA PROCESSING AND ANALYSES

Table C.1: Summary of data collection rates over the instrumented buildings and housings.

HOUSINGS									
Buildings	B1		B2				B3		TOTAL
Housings	B1/2	B1/3	B2/0	B2/1	B2/2	B2/5	B3/0	B3/2	TOTAL
Sensor with clamp ampere meters	N.A.	95.2%	95.2%	N.A.	N.A.	N.A.	98.0%	N.A.	95.8%
Pulse sensor for electric smart meter	52.3%	53.4%	54.4%	52.0%	53.7%	0%	63.6%	70.3%	49.0%
Connected plug	72.5%	0%	46.3%	0%	0%	77.9%	72.9%	69.9%	53.8%
Pulse sensors for gas meter	N.A.	0%	99.3%	N.A.	99.6%	N.A.	99.6%	N.A.	99.5%
Contact temperature of heaters	57.7%	98.0%	76.1%	77.9%	79.0%	78.1%	58.0%	62.8%	72.3%
Contact temperature of DHW pipes	60.6%	97.1%	77.1%	81.7%	79.1%	0.0%	89.7%	0.0%	61.8%
Indoor temperature of cold walls	63.0%	97.2%	77.1%	N.A.	78.4%	76.1%	40.6%	77.1%	72.5%
Temperature, humidity, luminosity, CO ₂ , presence	59.5%	67.1%	59.4%	N.A.	N.A.	58.0%	56.6%	59.2%	59.8%
Window opening detection	/	/	/	/	/	/	/	/	/
TOTAL	67.5%	23.3%	41.3%	26.2%	72.9%	59.5%	73.2%	65.0%	55.4%
COMMON AREAS AND ENERGY SYSTEMS									
Buildings	B1		B2				B3		TOTAL
Sensor with clamp ampere meters	N.A.		N.A.				17.7%		17.7%
Pulse sensor for electric smart meter	44.7%		27.0%				47.0%		37.8%
Thermal energy meter for heating	99.6%		99.6%				99.3%		99.5%
Thermal energy meter for DHW	55.4%		99.5%				99.6%		57.7%
Indoor temperature and humidity	99.2%		99.7%				98.0%		99.1%
Presence detection	100%		100%				100%		100%
TOTAL	52.2%		29.7%				34.7%		37.8%
WEATHER STATION									
TOTAL	92.4%								

C.2 Details of clusters for IAT in apartments

This appendix presents clustering results for B1/2 and B1/3 (Figure C.1 (a) and (b)), B2/0 and B2/5 (Figure C.2 (a) and (b)), and B3/0 and B3/2 (Figure C.3 (a) and (b)), corresponding to data analyses in Section 3.3.2.1.

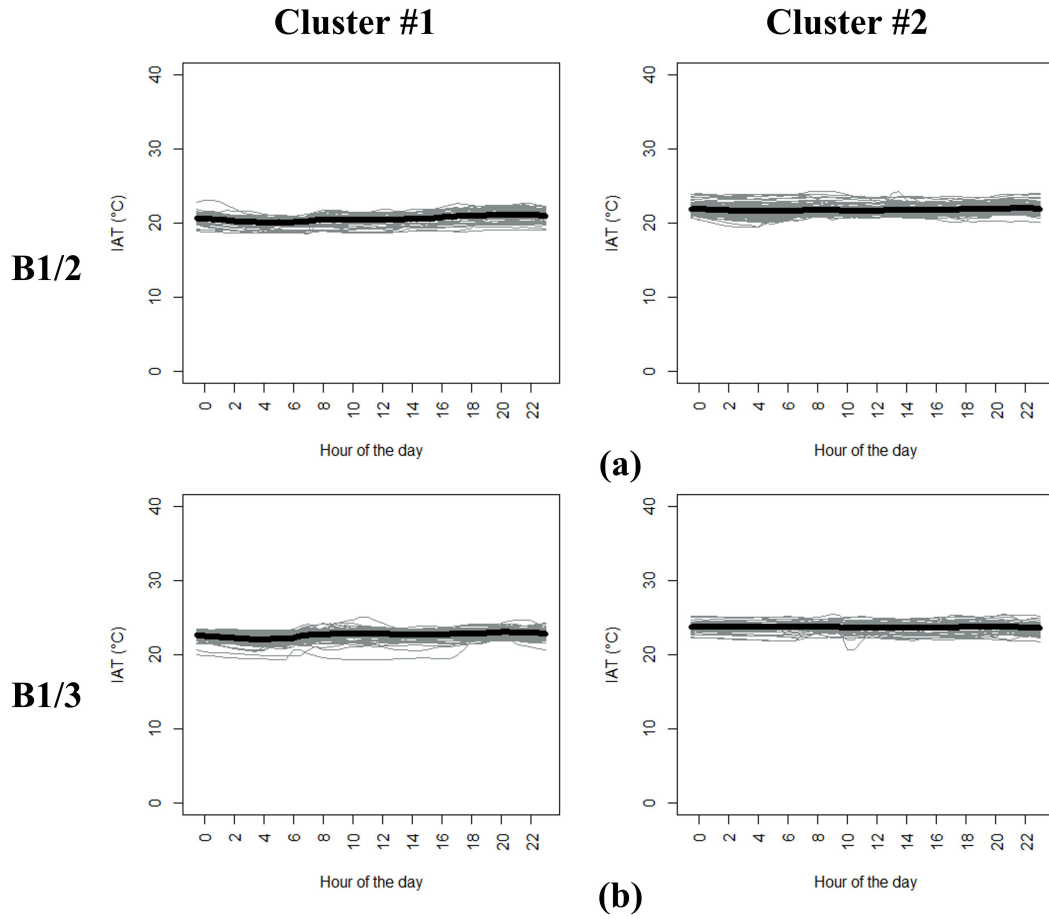


Figure C.1: IAT clusters for B1/2 (a) and B1/3 (b).

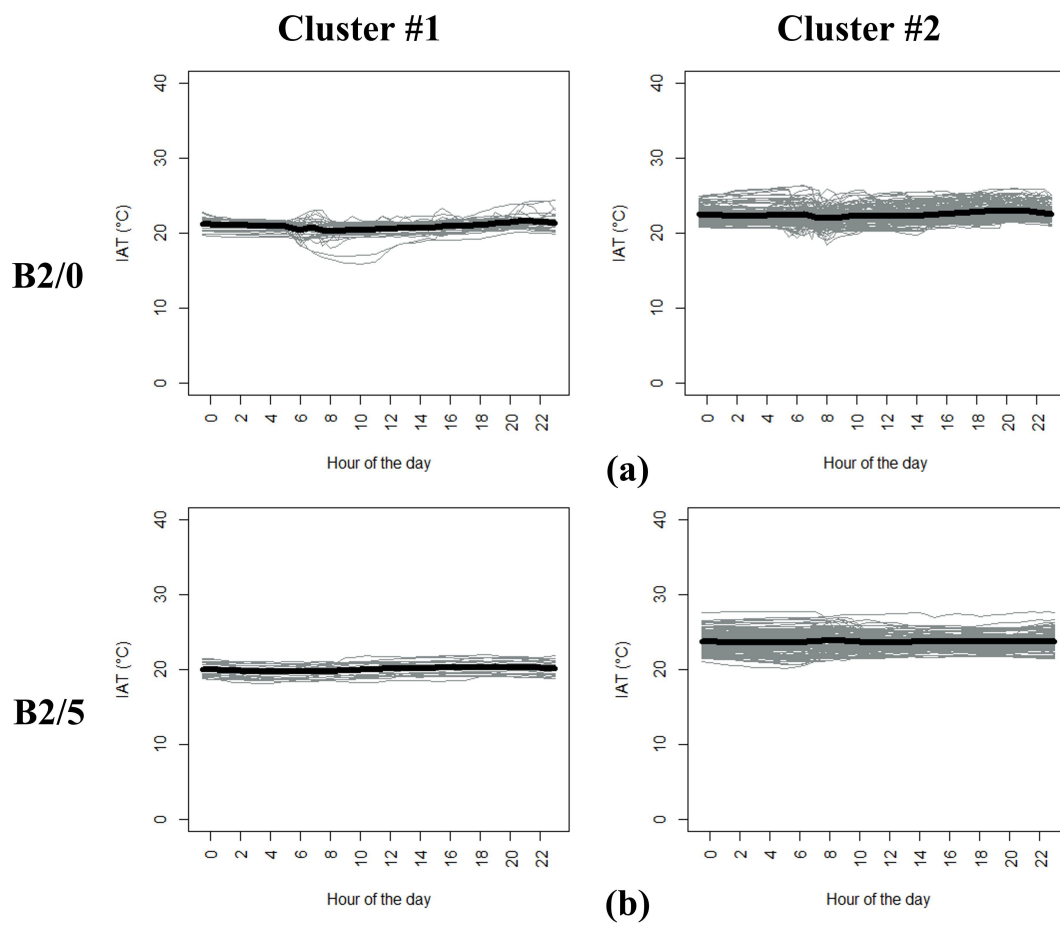


Figure C.2: IAT clusters for B2/0 (a) and B2/5 (b).

C.2 Details of clusters for IAT in apartments

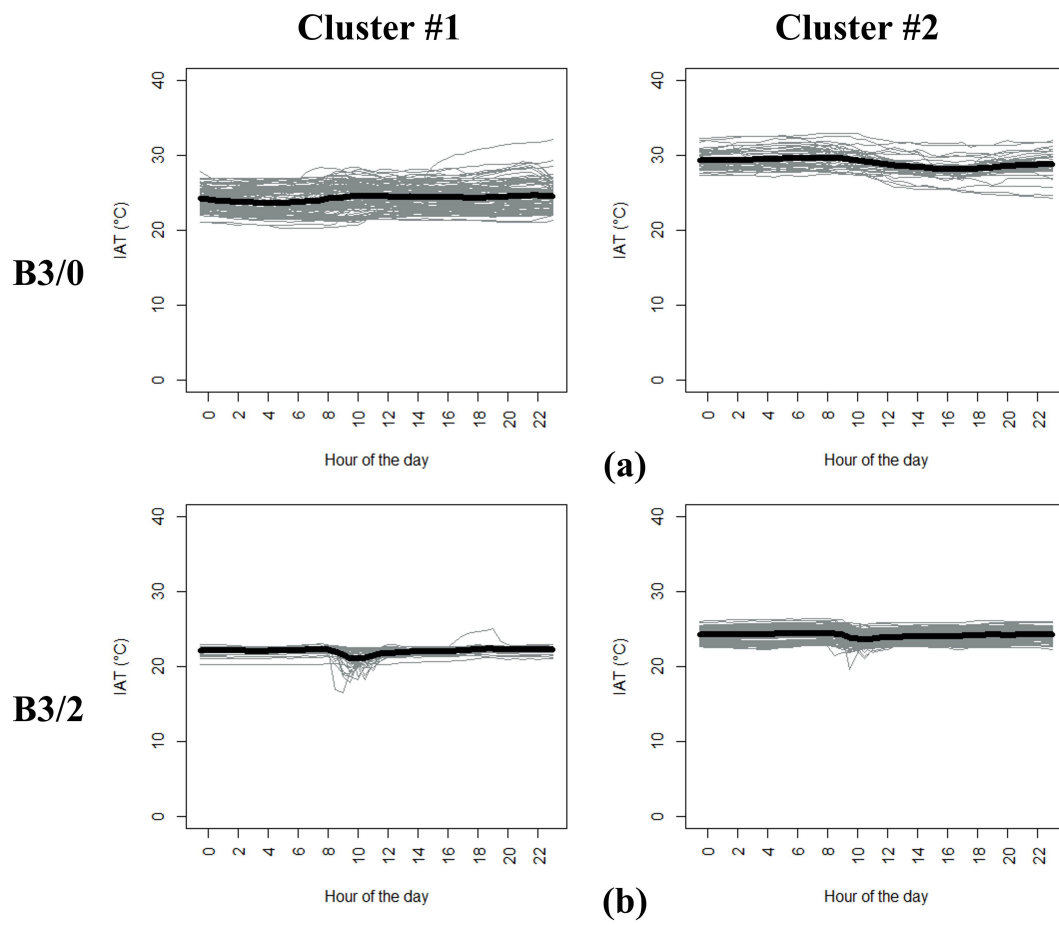


Figure C.3: IAT clusters for B3/0 (a) and B3/2 (b).

C.3 Density curves for IAT in apartments

The present appendix provides complementary density curves for instrumented apartments B1/2 and B1/3 (Figure C.4 (a) and (b)), B2/0 and B2/5 (Figure C.6 (a) and (b)), B3/0 and B3/2 (Figure C.6 (a) and (b)). Figures on top-left and bottom-left confront data distributions between daytime and nighttime. Figures on the top-right and bottom-right oppose data distributions in weekdays and weekends.

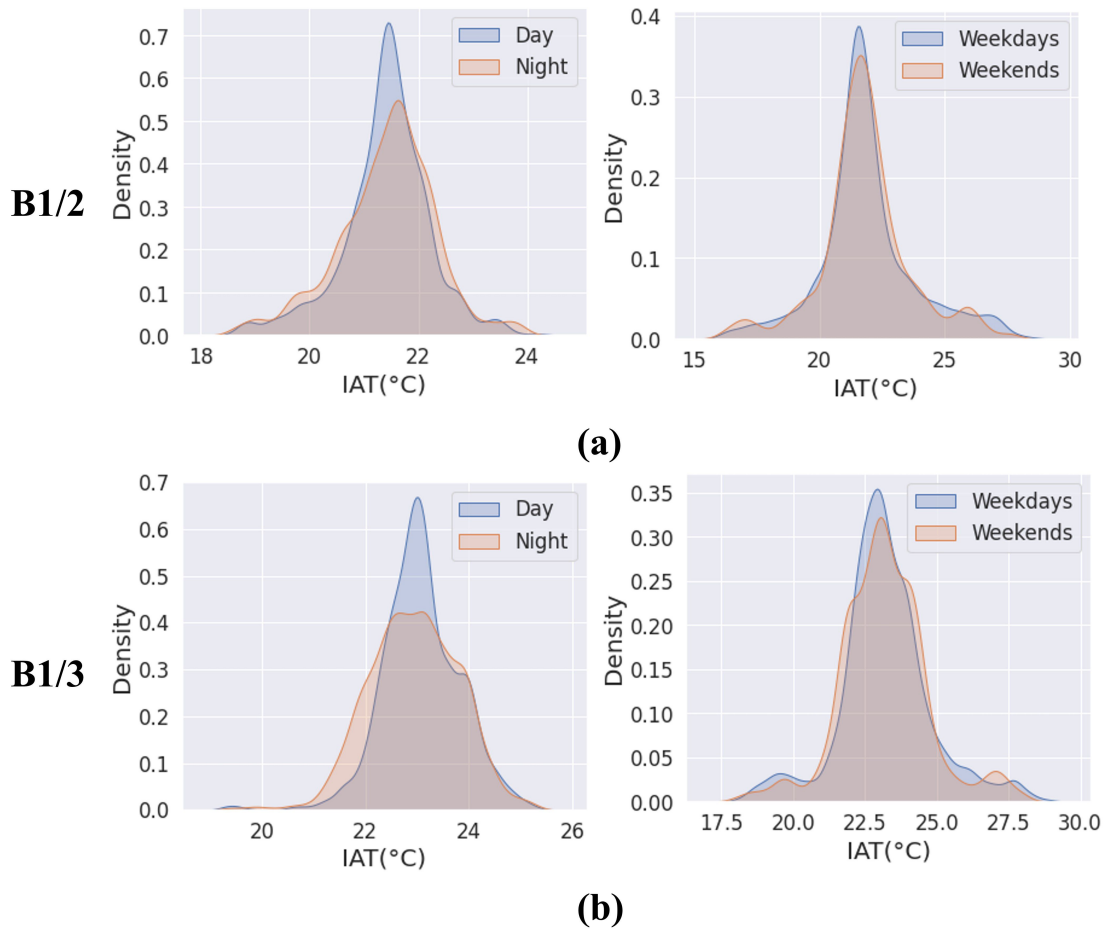


Figure C.4: Density curves for IAT in B1/2 (a) and B1/3 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.

C.3 Density curves for IAT in apartments

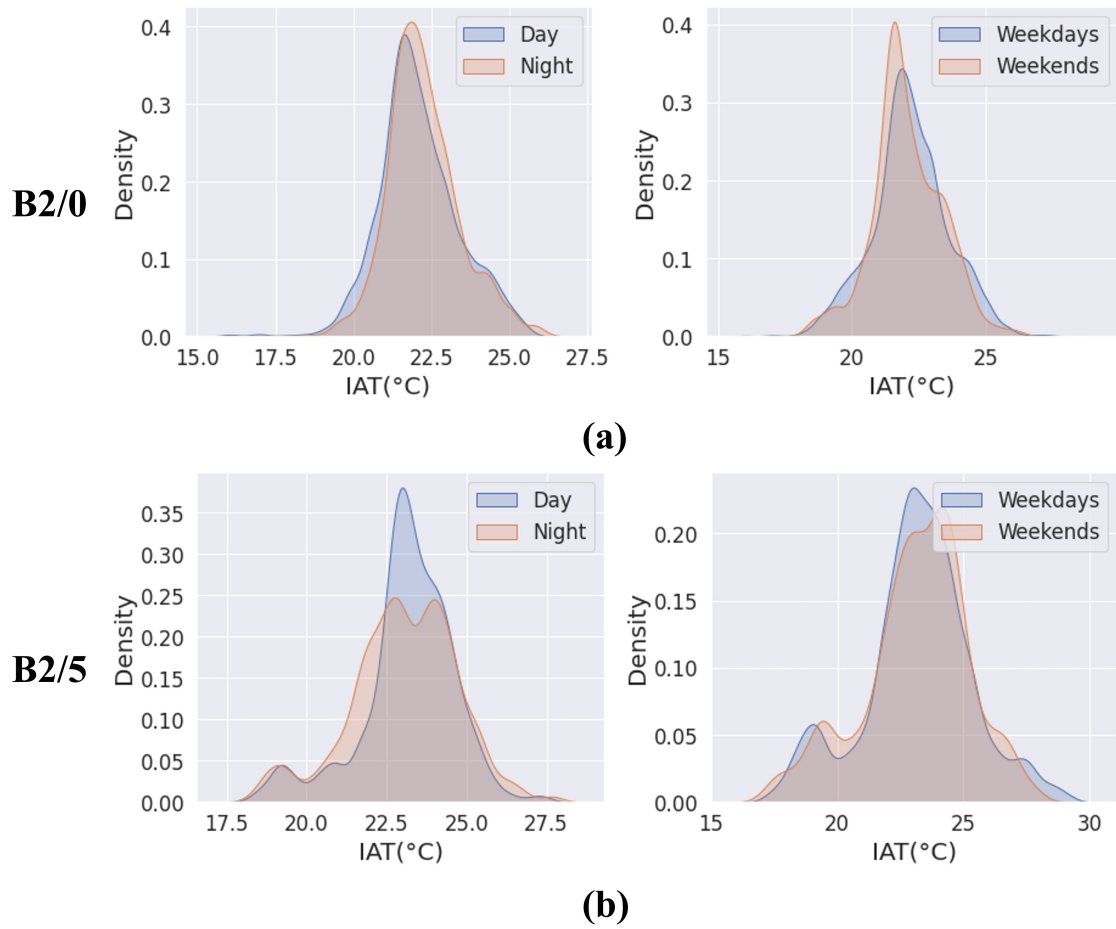


Figure C.5: Density curves for IAT in B2/0 (a) and B2/5 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.

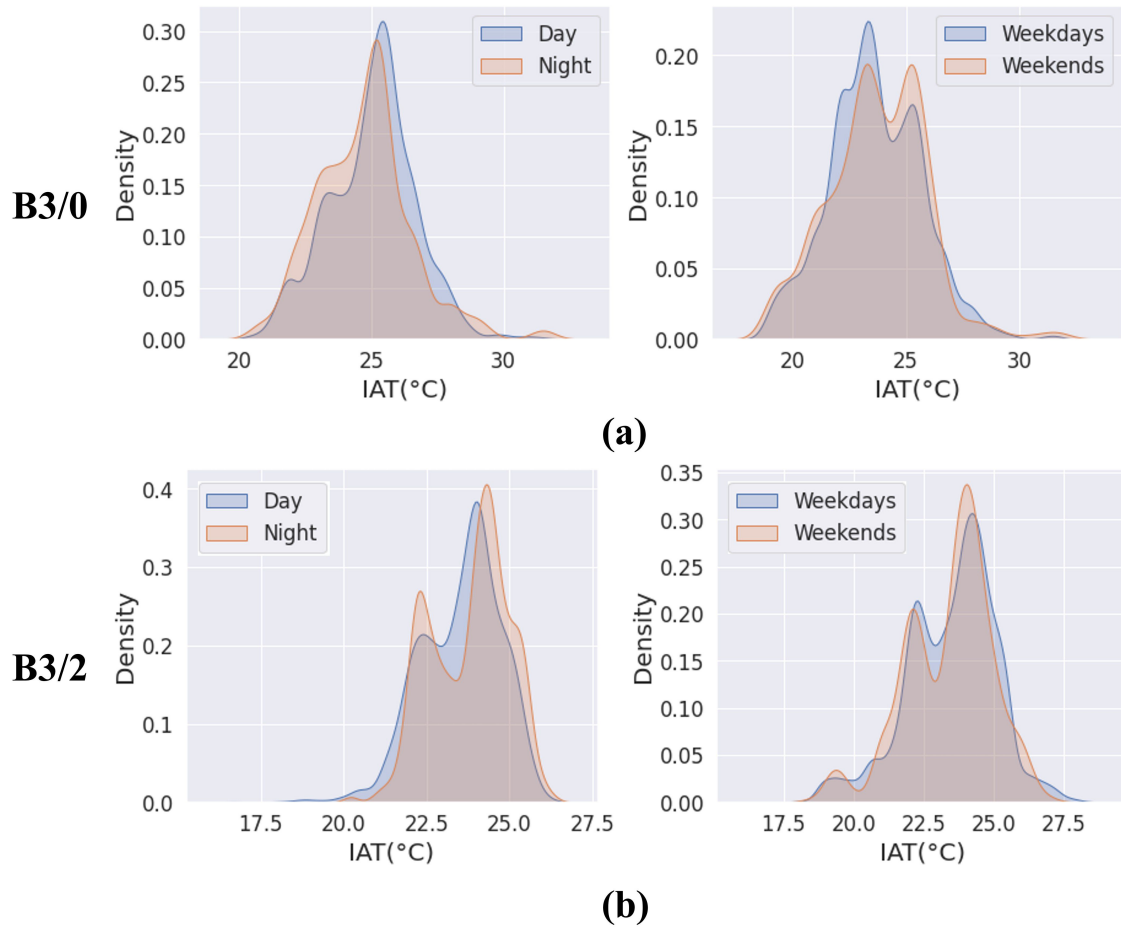


Figure C.6: Density curves for IAT in B3/0 (a) and B3/2 (b) – day and night density curves are on the top and bottom left, weekdays and weekends density curves are on the top and bottom right.

C.4 Details of clusters for dissipated electric power in apartments

This appendix complements analyses in Section 3.3.2.3 regarding electric dissipated power clustering for B1/2 and B1/3 (Figure C.7 (a) and (b)), B2/0, B2/1 and B2/2 (Figure C.8 (a), (b) and (c)), and B3/0 and B3/2 (Figure C.9 (a) and (b)).

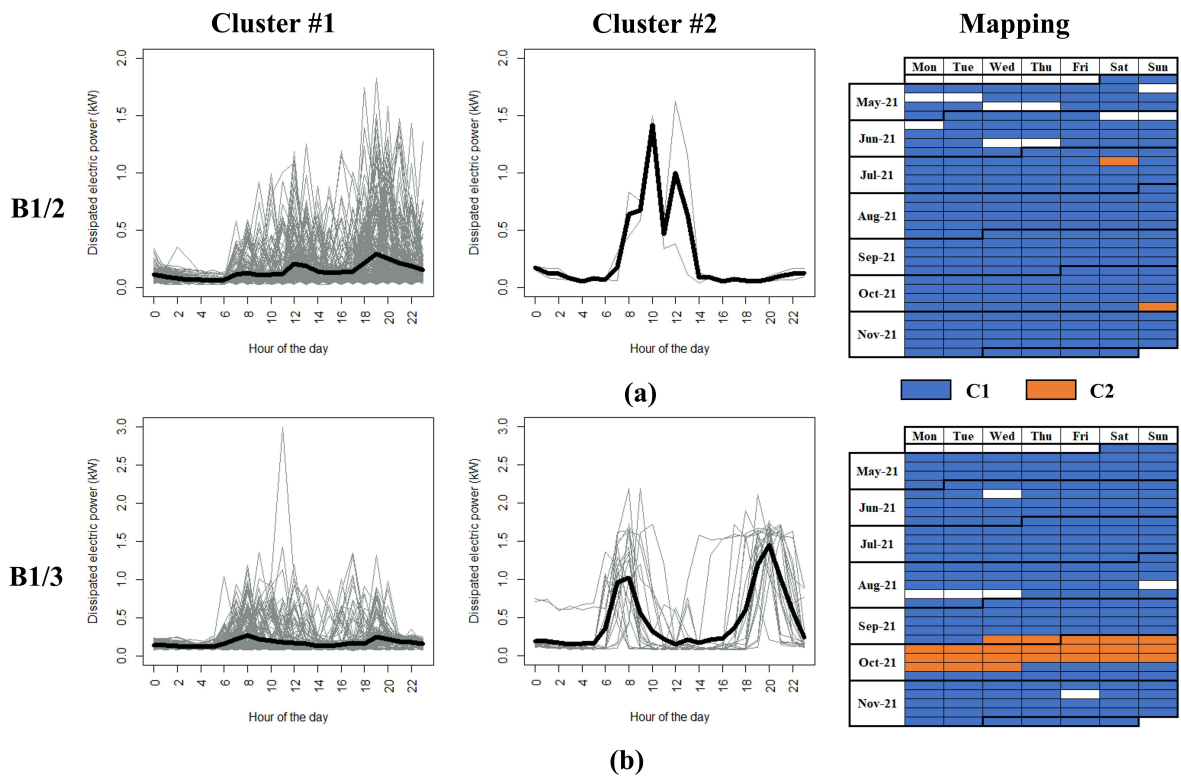


Figure C.7: Clusters and cluster mapping for dissipated electric power in B1/2 (a) and B1/3 (b).

A second round of clustering is also performed after removing outliers identified during the first clustering tests. The second test runs are implemented for B1/2, B2/0, B2/2, B3/0 and B3/2, for which outlying profiles were highlighted. Detailed results from the second round of clustering are provided in Figure C.10.

APPENDIX C – DATA PROCESSING AND ANALYSES

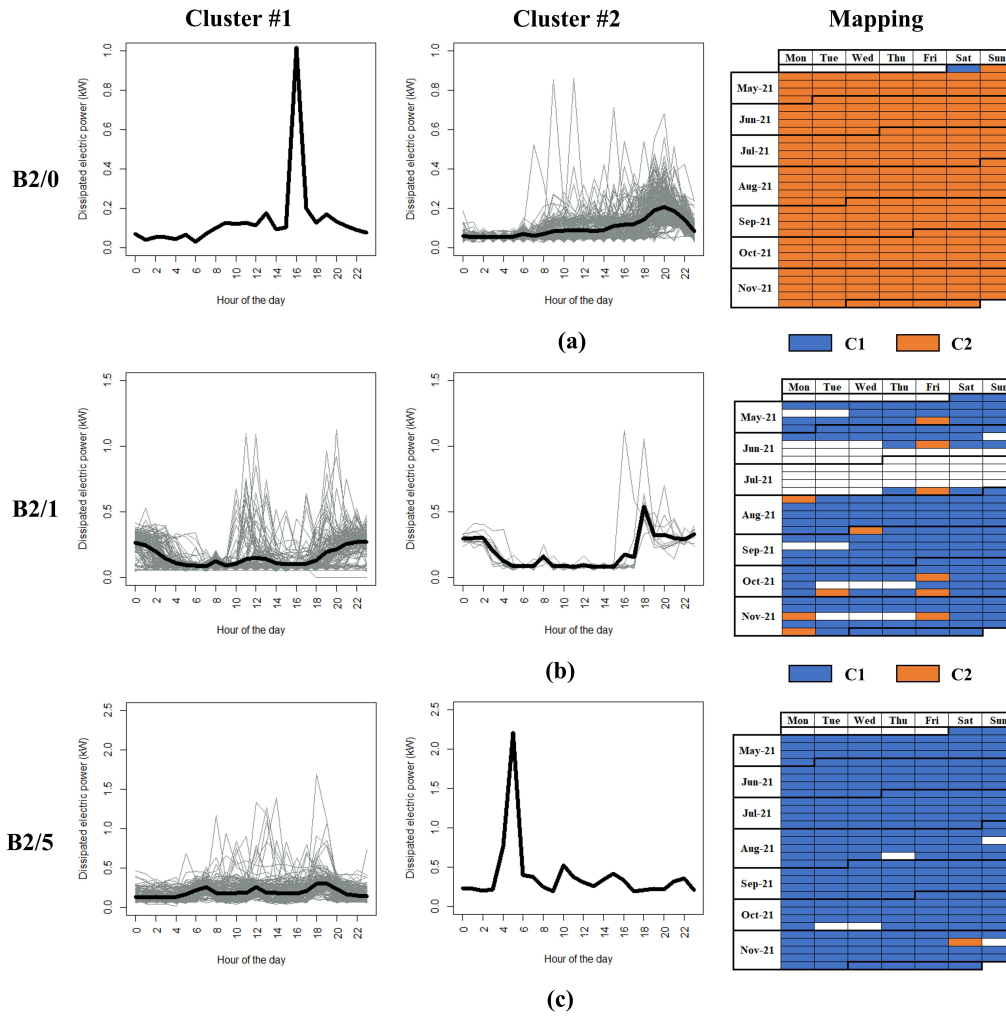


Figure C.8: Clusters and cluster mapping for dissipated electric power in B2/0 (a), B2/1 (b) and B2/2 (c).

C.4 Details of clusters for dissipated electric power in apartments

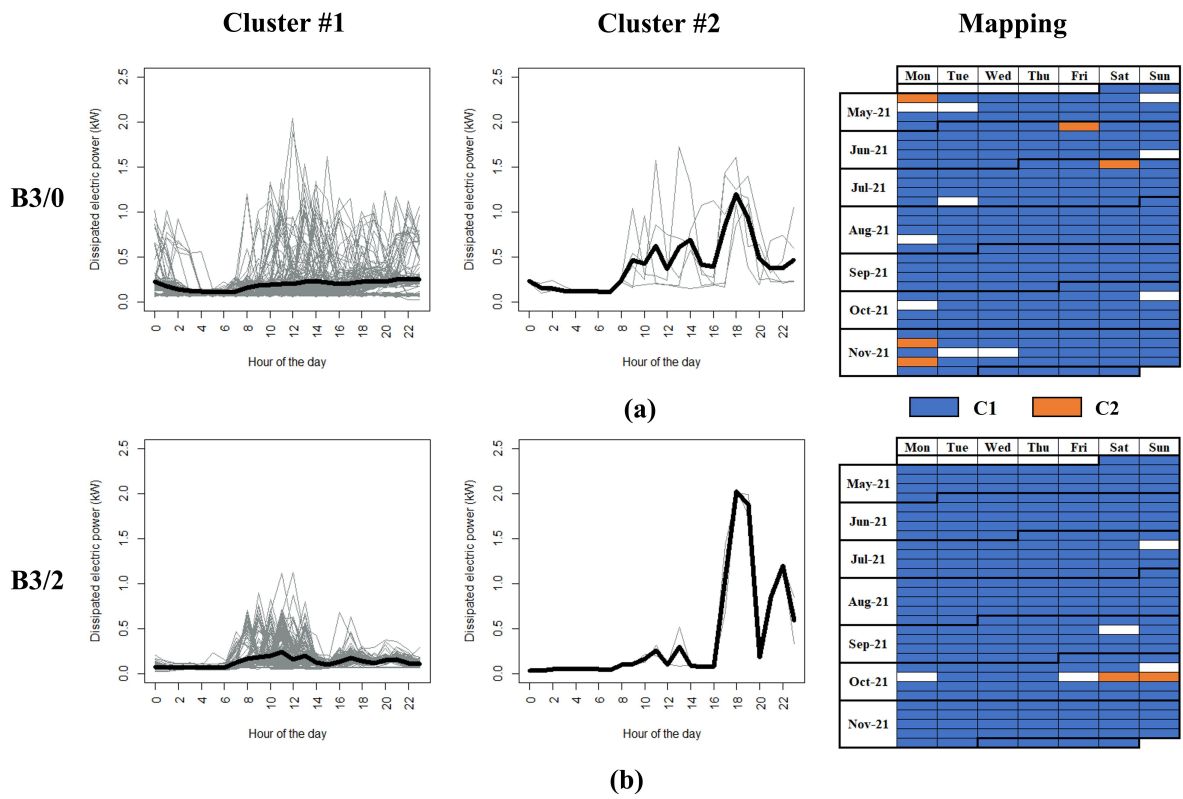


Figure C.9: Clusters and cluster mapping for dissipated electric power in B3/0 (a) and B3/2 (b).

APPENDIX C – DATA PROCESSING AND ANALYSES

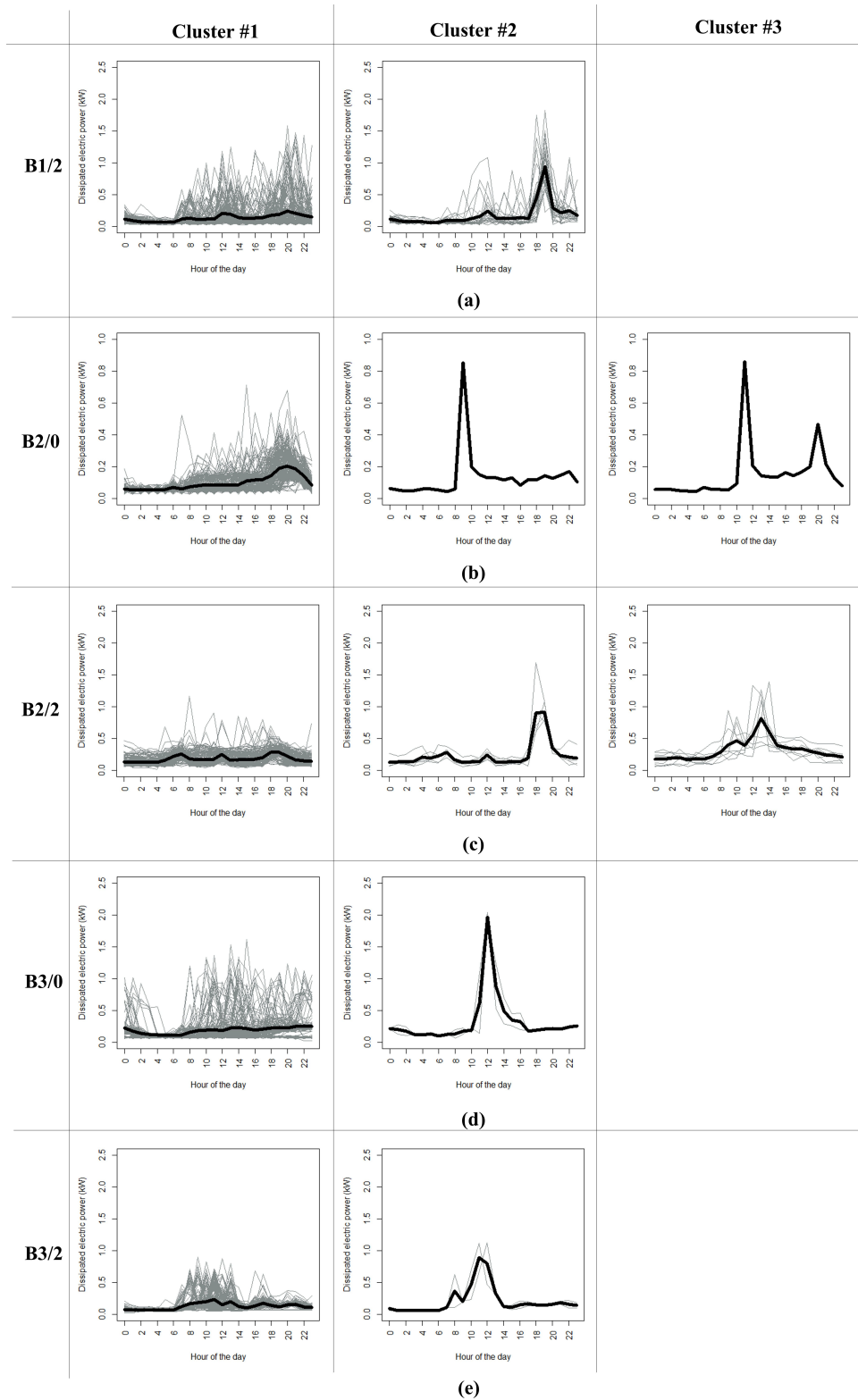


Figure C.10: Clusters for dissipated electric power from the second round of clustering, in B1/2 (a), B2/0 (b), B2/2 (c), B3/0 (d) and B3/2 (e).

C.5 Details of clusters and mapping for window opening

This appendix shows clusters and calendar mappings for window opening in apartments B1/2 and B1/3 (Figure C.11), B2/0, B2/1, B2/2 and B2/5 (Figure C.12), and B3/2 (Figure C.13), related to discussions in Section 3.3.2.4. It is complemented with the summary of cluster distribution for B1 (Table C.2), B2 (Table C.3) and B3 (Table C.4).

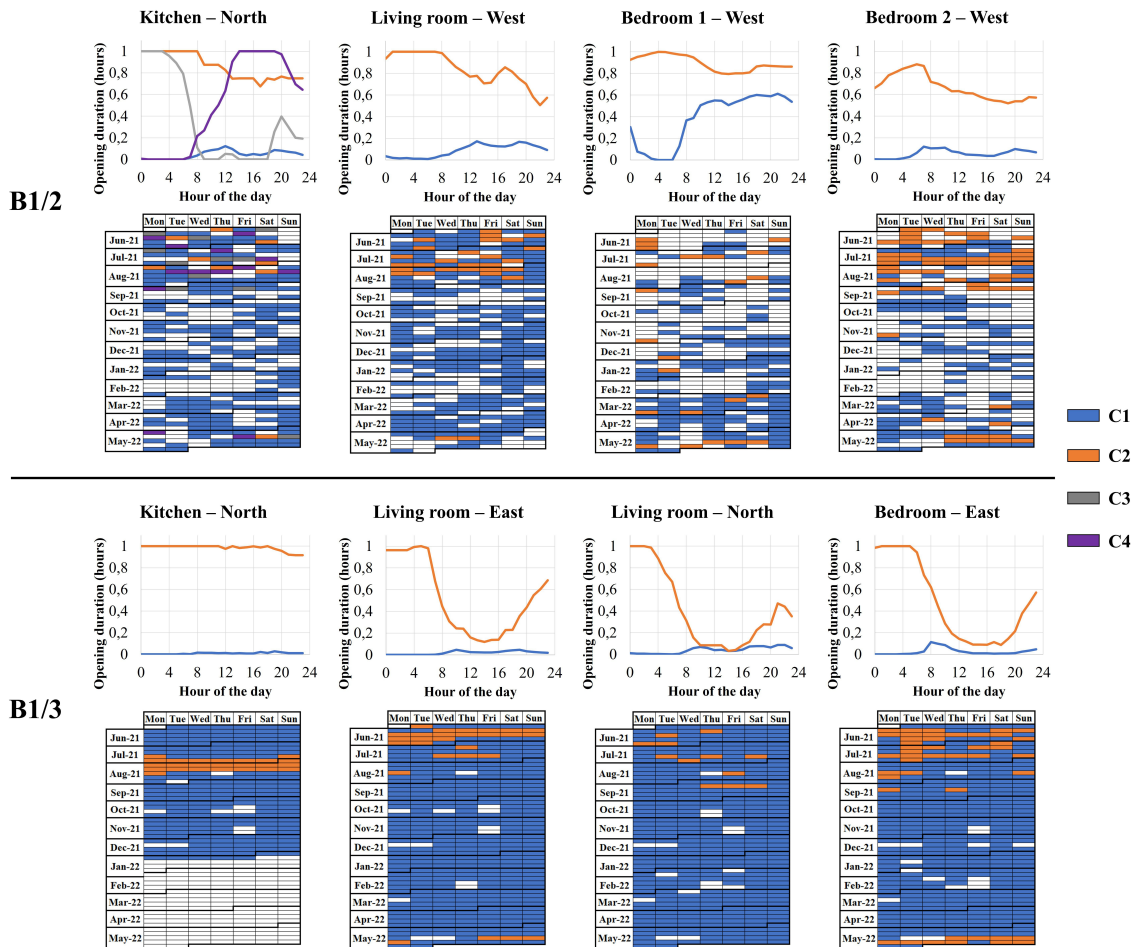


Figure C.11: Mean profile from clusters and cluster mapping for window opening in B1.

APPENDIX C – DATA PROCESSING AND ANALYSES

Table C.2: Clustering summary for window opening in B1.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
B1/2	Kitchen – North	Year-round	One random day per week – June to August		
	Living room – West	September to June	July to August	/	/
	Bedroom 1 – West	September to June	July to August	/	/
	Bedroom 2 – West	September to April	May to August	/	/
B1/3	Kitchen – North	Year-round	Last week of July to middle of August	/	/
	Living room – East	July to May	June	/	/
	Living room – North	Year-round	Not considered	/	/
	Bedroom – East	August to April	May to July	/	/

C.5 Details of clusters and mapping for window opening

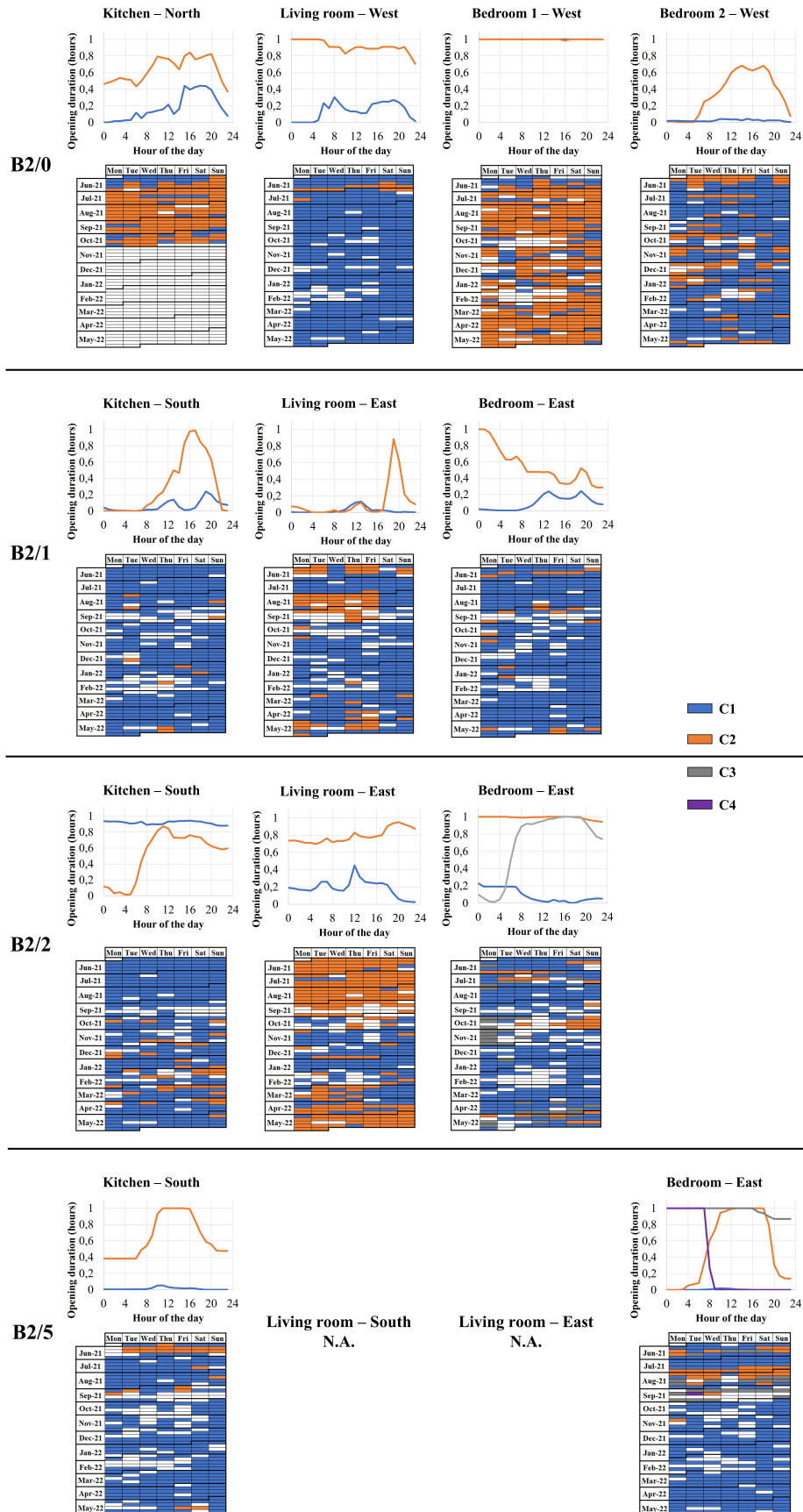


Figure C.12: Mean profile from clusters and cluster mapping for window opening in B2.

APPENDIX C – DATA PROCESSING AND ANALYSES

Table C.3: Clustering summary for window opening in B2.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
B2/0	Kitchen – North	First two weeks of June	Mid-June to October	Rest of year closed	
	Living room – West	Year-round except for...	... the last week of July	/	/
	Bedroom 1 – West	Not considered	Year-round	/	/
	Bedroom 2 – West	Year-round except for...	...one random day per week	/	/
B2/1	Kitchen – South	Year-round	Not considered	/	/
	Living room – East	September to April, July	May, June, August	/	/
	Bedroom – East	Year-round	Not considered	/	/
B2/2	Kitchen – South	Year-round except for...	...one random day per week from October to March	/	/
	Living room – East	September to March	April to August	/	/
	Bedroom – East	Year-round	Not considered	Not considered	/
B2/5	Kitchen – South	July to May	June	/	/
	Living room – South	No available data			
	Living room – East	No available data			
	Bedroom – East	September to July	August	Not considered	Not considered

C.5 Details of clusters and mapping for window opening

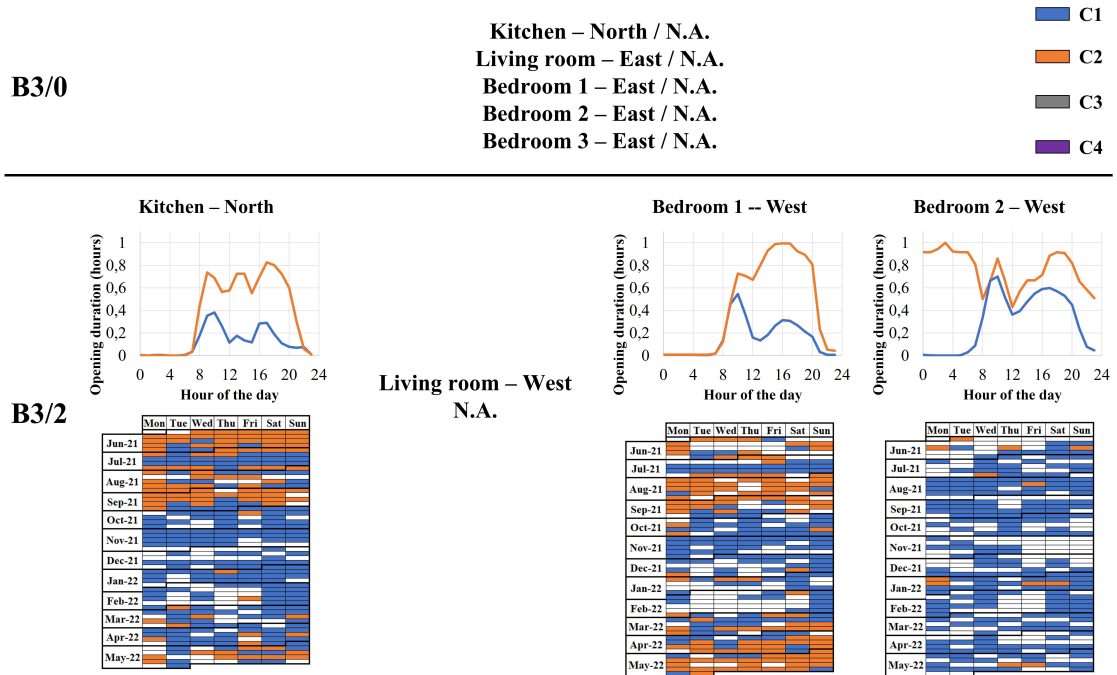


Figure C.13: Mean profile from clusters and cluster mapping for window opening in B3.

Table C.4: Clustering summary for window opening in B3.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
B3/0	Kitchen – North	No available data			
	Living room – East	No available data			
	Bedroom 1 – East	No available data			
	Bedroom 2 – East	No available data			
	Bedroom 3 – East	No available data			
B3/2	Kitchen – North	October to April, July	May, June, August, September	/	/
	Living room – West	No available data			
	Bedroom 1 – West	October to March, July	April to June, August, September	/	/
	Bedroom 2 – West	Year-round	Not considered	/	/

C.6 Details of cluster mapping for DHW in apartments

This appendix provides calendar mapping for DHW use in apartments B1/2, B1/3, B2/0 and B3/0 (Figure C.14 (a), (b), (c), and (d), respectively) corresponding to results and analyses in Section 3.3.2.5.

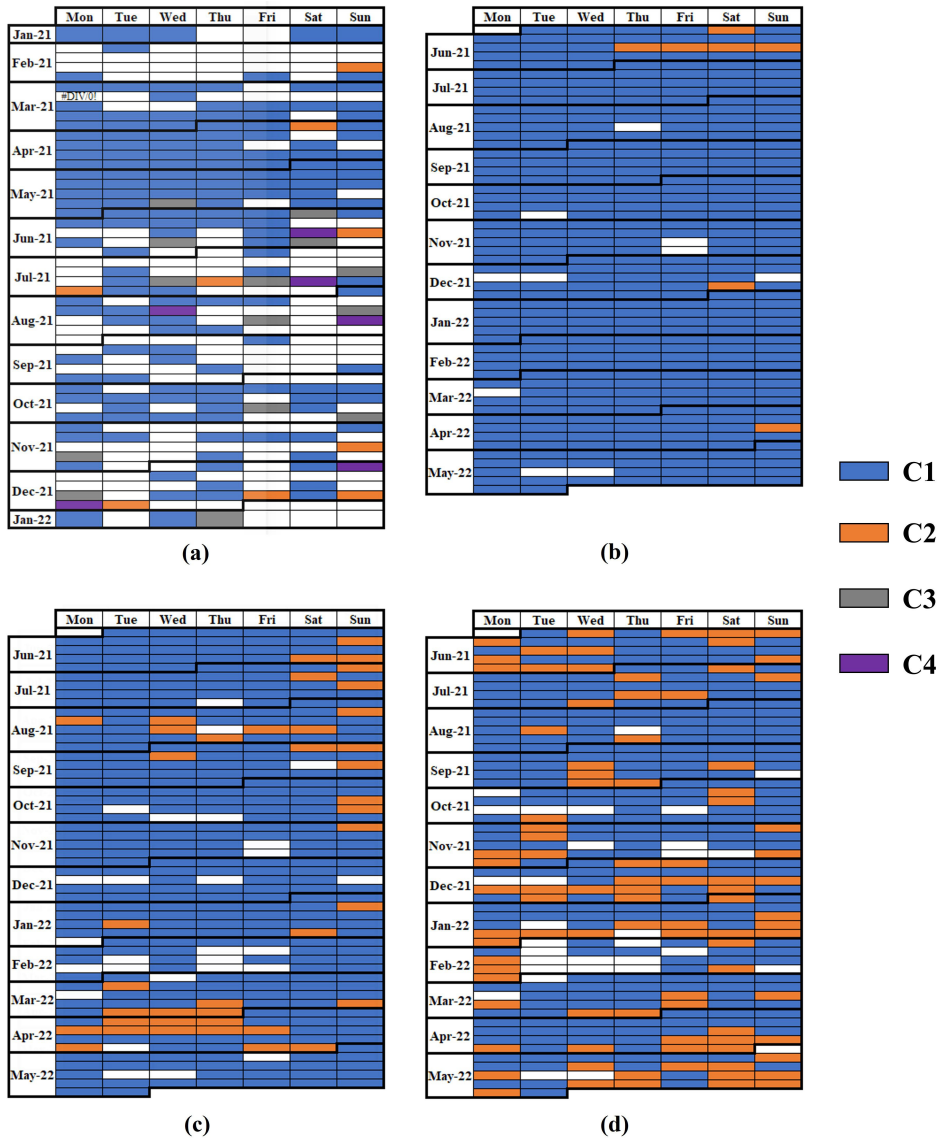


Figure C.14: Cluster mapping for DHW use in B1/2 (a), B1/3 (b), B2/0 (c) and B3/0 (d).

C.6 Details of cluster mapping for DHW in apartments
