



HAL
open science

Modelling household activities in instrumented dwellings

Huynh Phan

► **To cite this version:**

Huynh Phan. Modelling household activities in instrumented dwellings. Mechanics [physics]. Université de Bordeaux, 2022. English. NNT : 2022BORD0009 . tel-03796541

HAL Id: tel-03796541

<https://theses.hal.science/tel-03796541>

Submitted on 4 Oct 2022

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE SCIENCES PHYSIQUES ET DE L'INGENIEUR

Institut de Mécanique et d'Ingénierie (I2M) de Bordeaux

Spécialité: MÉCANIQUE

Présentée par

Ly Huynh PHAN

Modélisation de l'activité des ménages dans le bâtiment résidentiel instrumenté

Soutenue le: 10 Janvier 2022

MEMBRES DU JURY :

Mme. Monika WOLOSZYN, Professeure, Université Savoie Mont Blanc, France, Président

M. Nizar BOUGUILA, Professeur, Concordia University, Canada, Rapporteur

M. Christian INARD, Professeur, Université de La Rochelle, France, Rapporteur

Mme. Valérie CAMPS, Maître de conférences, Université Paul Sabatier, France, Examinatrice

M. Cuong PHAM, Maître de conférences, Post and Telecommunications Institute of Technology, Vietnam, Examineur

M. Laurent MORA, Professeur, Université de Bordeaux, France, Co-directeur

M. Stéphane PLOIX, Professeur, Grenoble INP, France, Co-directeur

M. Thomas RECHT, Maître de conférences, Université de Bordeaux, France, Co-encadrant

INVITÉ :

M. Éric VORGER, Docteur, Kocliko, France

Acknowledgements

The journey is long enough for me to learn many things. During the journey, I received many supports from many people. Without them, it is incredibly difficult for me to complete my Ph.D. Firstly, I would like to say thanks to my supervisors Prof. Laurent, Prof. Stéphane and Dr. Thomas. They have been very patient, supportive and enthusiastic. The beginning period was very difficult but they always believed me and helped me a lot. Secondly, I would like to thank my family in Vietnam, my parents and my brother, who always support me in any case. Then, I would like to thank my teacher (Dr. Quoc Dung) and friends, both in Vietnam (Dan, Hiep, Linh, Huong, Duc, Long, Tien, Luyen, ...) and in Bordeaux (Thanh, Nhan, Truc, Thy, Q.Anh, Trang, Tu, ...). They help me a lot and we have had beautiful periods during these years. I would give special thanks to my girlfriend Phuong Thao. Without you, I can not finish my road. You always help me, support me and make me more confident in the most difficult times. Thank Google, Reddit, StackOverflow, Sciencedirect, ... for all of your content and knowledge.

Last but not least, this work was funded by the Regional Council of Nouvelle Aquitaine within the joint research lab GP2E between I2M laboratory and technical center Nobatek/INEF4.

Abstract

Simulation tools support engineers to size Heating, Ventilation, and Air Conditioning (HVAC) systems and optimize the building design. It is proved that occupant behaviors are important to explain the discrepancy between simulated and actual energy consumption. Better information on the occupant activities might reduce this gap. Recently, many scientists have focused on modeling occupants' activities to improve the performance of the simulation. Though, an activity is affected by the context in houses, which gives the information to fully understand this activity. To address this question, stochastic approaches were proposed with statistical data to model activity profiles based on household characteristics (building type, income, *etc.*). Nevertheless, their outputs only provide average information based on different contexts (locations, appliances, *etc.*). Thus, they are not specific and are unverifiable for particular houses. The agent-based approach is another way to model, in a contextualized manner, the occupants' behaviors in houses. It has been developed to satisfy occupant needs. However, the agent-based approach is challenging to implement in households with many members due to their numerous interactions. Data-based models with measurements have also been proposed to estimate occupant activities. They aim to determine the systematic relationships between variables using machine learning techniques. Due to the measurements, the context is considered. However, most of these models are not analytical and they do not provide occupant-understandable explanations. The problem tackled by this thesis is to use data from a given context, which comes from sensors and questionnaires, to estimate contextualized occupant activities (cooking, sleeping, taking a shower, *etc.*) and their related energy impacts. The necessary features for the estimation should be determined to reduce data noise and save the instrumentation budget. Activity estimations should consider the context to have more precise results. These approaches must be verifiable and provide

understandable information for inhabitants to help them enhance their behaviors in order to reduce energy consumption. In this thesis, a mobile application has been developed to collect labels of activities from the occupants. These labels have then been combined with the data from sensors and the context information of the house to detect and simulate occupants' activities and their related energy impacts. The proposed method focused on three steps:

- estimating occupant activities with sensors, the context, and the activity labels. It starts by determining the most relevant features for the activity estimation. Then, a consequences-based Bayesian Network is built using context factors and the knowledge coming from observations and questionnaires which are given to occupants to ask about their activities. Two houses have been used for testing. Cross-validation and F1-score are used to evaluate the model quality. The results show that it could estimate several activities (cooking, washing dishes, washing clothes, *etc.*) with the F1-score from 0.77 to 0.88.
- evaluating the energy impacts of occupant activities with a set of sensors, knowledge, and activity labels. The thesis focuses on the effects of activities on electricity consumption. A Bayesian Network is built for each activity to evaluate its impacts on electricity consumption. The model has been tested with several activities in the studied houses. To validate, the model compares statistically the electricity consumption between activities and between the weekdays.
- simulating occupant activities and their energy impacts, which is particularly useful in the building renovation. Results from previous models are re-used and combined with Monte-Carlo simulation to simulate sequences of occupants' activities and their associated electricity consumption. Comparisons between simulated and actual electricity consumption of appliances are made to verify the model quality.

Keywords : occupant activity modelling, building energy simulation, Bayesian Network, machine learning, contextualized occupant activity, instrumented dwellings.

Résumé en français

Depuis la révolution industrielle au milieu du XVIII^e siècle, parallèlement au développement des applications de haute technologie, l'influence des activités humaines sur l'environnement, et en particulier sur le climat, a considérablement augmenté. La production de gaz à effet de serre (GES) a augmenté de 260 % depuis 1750. L'émission accrue de GES tels que le CO₂ par la combustion de combustibles fossiles est à l'origine du dérèglement climatique de la planète, l'un des problèmes les plus graves auxquels le monde est confronté aujourd'hui. Alors que la consommation d'énergie représente plus de 80 % des GES mondiaux, en Europe, les bâtiments résidentiels en représentent plus de 25 % en 2018, d'où la nécessité de réduire la consommation d'énergie dans les bâtiments résidentiels. Les outils de simulation aident les ingénieurs à dimensionner les systèmes Chauffage Ventilation Climatisation (CVC), à optimiser la conception d'un bâtiment et à améliorer son efficacité énergétique. Cependant, la performance énergétique d'un bâtiment n'est pas seulement influencée par ses caractéristiques physiques mais aussi par les impacts des occupants à travers leur présence et leurs activités. Ces impacts sont considérés à chaque phase d'un projet de construction ou de rénovation, depuis les premières esquisses jusqu'à l'exploitation. En phase de conception, dans la plupart des cas (y compris pour l'outil réglementaire), des scénarios horaires de température de consigne et d'occupation des locaux sont pris en compte pour décliner les profils types de puisage d'eau chaude sanitaire et d'utilisation des appareils électroménagers, d'éclairage artificiel, d'ouverture des volets, etc. Cependant, le retour d'expérience montre souvent une différence considérable entre les valeurs prédites et la consommation réelle. Il s'avère que l'impact des occupants est l'un des principaux facteurs expliquant cet écart. Une meilleure information sur les activités des occupants pourrait améliorer les performances des outils de simulation énergétique et réduire cet écart. Récemment, en plus d'améliorer l'enveloppe du bâtiment, des

efforts ont été faits pour améliorer les systèmes de gestion et de contrôle de l'énergie dans les bâtiments résidentiels en considérant les activités des occupants comme faisant partie du système. L'objectif n'est pas seulement de réduire la consommation d'énergie mais aussi de satisfaire le confort des occupants. Il est donc crucial de modéliser les comportements des occupants et leurs impacts sur la performance énergétique du bâtiment.

Dans cette thèse, les activités des occupants sont définies comme ce qu'une personne ou un groupe fait ou a fait dans un certain endroit pendant un certain temps. Actuellement, de nombreux scientifiques s'attachent à modéliser ces activités afin d'améliorer les performances des outils de simulation énergétique. Initialement, les scénarios des activités sont programmés et définis pour chaque zone thermique. Ces scénarios sont déterminés sur la base de conditions standard ou d'une analyse statistique des observations. Cependant, ces scénarios sont reproductibles pour de nombreux ménages alors que leurs caractéristiques (membres, type de bâtiment, professionnel, etc.) sont différentes. Par conséquent, ils ne sont pas assez adaptés pour modéliser l'activité des occupants avec sa diversité. L'activité d'un occupant peut en effet être affectée par le contexte dans les maisons, qui fournit l'ensemble des circonstances environnantes qui peuvent être utilisées pour caractériser la situation de cette activité (Henricksen et Indulska, 2005; Xuan Hoa Binh et al., 2010). Par exemple, une famille de quatre personnes cuisine différemment d'une famille d'un seul membre. Une personne peut dormir plus longtemps le week-end qu'en semaine.

Pour prendre en compte ces informations, de nombreuses études utilisent des modèles stochastiques basés sur des données statistiques telles que les enquêtes sur l'emploi du temps. Dans ces modèles, les profils d'activité sont modélisés en fonction des caractéristiques générales du ménage (membres, type de bâtiment, occupation, etc.). Ces profils sont ensuite utilisés pour reproduire et simuler des scénarios d'activité. Cette approche, facile à mettre en œuvre grâce à la disponibilité de données statistiques, a permis de résoudre le problème de la prise en compte de la diversité des ménages. En conséquence, les modèles stochastiques sont devenus l'approche la plus populaire de modélisation des activités ces dernières années. Cependant, les résultats actuels de ces modèles fournissent des informations moyennes basées sur diverses maisons dans des contextes différents (pays, emplacements, habitudes des occupants, appareils, etc.)

Ainsi, les informations contenues dans les modèles stochastiques sont un mélange de données, qui sont invérifiables pour un bâtiment spécifique dans un contexte particulier.

L'approche basée sur les agents est une autre façon d'étudier les interactions occupant-bâtiment et occupant-occupant par le biais d'agents auto-organisés, qui agissent pour atteindre leur confort prédéfini. Ces modèles sont développés pour satisfaire les besoins des occupants dans des maisons spécifiques afin que le contexte soit intégré dans le processus de décision pour réaliser des activités. Cependant, l'approche basée sur les agents est très complexe et difficile à mettre en œuvre dans les ménages composés de nombreux membres en raison de leurs interactions rapidement complexes. En outre, la diversité des caractéristiques, de la personnalité et des habitudes entraîne la difficulté de définir le confort des agents, qui est un facteur important dans les systèmes à base d'agents. En outre, des modèles basés sur des données mesurées ont été proposés pour reconnaître les activités des occupants. Cette approche vise à déterminer les modèles et les relations systématiques entre les variables connexes en utilisant des techniques d'exploration de données et d'apprentissage automatique. Grâce aux données mesurées, elle permet de surmonter les limites de l'approche stochastique dans la vérification et l'évaluation des modèles. Cependant, la plupart de ces modèles ne sont pas analytiques et ne fournissent pas d'explications compréhensibles aux occupants. En outre, les résultats de ces modèles sont déterministes, de sorte qu'ils ne tiennent pas compte de la diversité des activités des occupants. En outre, la dépendance totale aux données rend cette approche difficile à intégrer avec les connaissances d'experts, qui sont utiles pour de nombreuses estimations d'activités avec des données limitées.

Le problème abordé par cette thèse est de déterminer et de traiter les données d'un contexte donné, provenant de capteurs et de questionnaires, qui pourraient être utilisées pour estimer les activités des occupants (cuisiner, dormir, se doucher, etc.) et leurs impacts énergétiques dans une maison spécifique, qui sont utiles dans la simulation énergétique et les bilans énergétiques. Les variables requises et leurs caractéristiques utiles pour l'estimation doivent être déterminées pour réduire les capteurs redondants et économiser le budget d'installation. Les estimations d'activité doivent prendre en compte le contexte spécifique pour rendre les résultats plus précis et spécifiques. Ces approches d'estimation sont vérifiables et fournissent des informations compréhensibles, ce qui est utile aux habitants pour comprendre leurs activités et

ensuite améliorer certaines activités inefficaces afin de réduire la consommation d'énergie et d'améliorer leur confort dans la maison.

Pour répondre aux objectifs ci-dessus, outre les mesures, il faut collecter les informations des activités des occupants. Cependant, cette tâche est difficile dans les logements en raison de la politique privée. Ainsi, dans cette thèse, une application mobile a été développée pour collecter les étiquettes d'activité des occupants. Ces étiquettes sont ensuite combinées avec les données des capteurs et les informations contextuelles de la maison pour détecter et simuler les activités des occupants (cuisiner, dormir, se doucher, etc.) et leurs impacts énergétiques, notamment en termes de consommation électrique. La méthode proposée se concentre sur trois étapes:

1. Estimer l'activité des occupants à l'aide de plusieurs capteurs, du contexte et des étiquettes d'activité recueillies auprès des occupants. Cette étape commence par déterminer les caractéristiques les plus liées à l'activité. Ensuite, un modèle basé sur la connaissance est construit à partir des données des capteurs, de la connaissance experte de l'activité et du contexte à partir de l'observation et du questionnaire, qui est remis aux occupants pour les interroger sur leurs activités. Il s'appuie sur le réseau bayésien pour modéliser l'activité d'un occupant avec des relations probabilistes de cause à effet basées sur un réseau défini basé sur les conséquences. Le réseau bayésien est choisi car ses paramètres sont faciles à comprendre et ses résultats sont probabilistes, ce qui est nécessaire pour prendre en compte la diversité des activités des occupants. Deux bâtiments résidentiels en France avec des activités différentes ont été utilisés pour le test. Pour vérifier la qualité du modèle, une validation croisée avec la métrique du score F1 est utilisée pour évaluer la précision du modèle. Les résultats montrent que le modèle est capable d'estimer certaines activités (cuisine, vaisselle, lessive, soins personnels, etc.) avec un score F1 de 0,77 à 0,88. Pour analyser les profils d'activité estimés, une comparaison statistique entre les profils d'activité estimés et observés. Les résultats montrent que le modèle peut être utile pour déterminer les profils d'activité dans des maisons spécifiques, ce qui est utile dans le processus de vérification de performance énergétique des bâtiments.
2. Lorsque l'on considère les activités des occupants, outre les heures et les durées, leurs influences sur les consommations énergétiques du bâtiment doivent être

prises en compte. La deuxième tâche de cette thèse est d'évaluer les impacts énergétiques des activités des occupants sur le bâtiment à l'aide d'un ensemble de capteurs, de contextes et d'activités collectées. La thèse se concentre sur la consommation d'électricité induite par les activités des occupants. Les appareils dépendants d'une activité sont déterminés sur la base des connaissances recueillies auprès du ménage. Ensuite, un modèle d'utilisation est construit pour chaque activité des occupants afin de déterminer et d'évaluer comment les occupants utilisent les appareils électriques pour mener cette activité dans la maison. Le modèle a été testé avec plusieurs activités dans les maisons étudiées. Plus précisément, le modèle compare et évalue la consommation d'électricité entre les activités et entre les jours de la semaine. Pour valider la méthodologie, l'électricité utilisée par les appareils a été reproduite et simulée avec les étiquettes d'activités observées. La comparaison entre cette simulation et la mesure montre que la méthodologie est utile pour déterminer l'électricité consommée par les appareils dans les activités des occupants. Cependant, la qualité du modèle n'est pas constante lorsque les données d'entraînement sont limitées ou que les occupants effectuent des changements de comportement dans l'utilisation des appareils.

3. Pour rénover des bâtiments, des simulations énergétiques sont généralement utilisées pour estimer leur performance énergétique actuelle et pour étudier des stratégies de rénovation. Afin d'améliorer les performances de ces simulations dans des logements spécifiques, il est nécessaire d'utiliser des activités des occupants contextualisées plutôt que des activités représentatives issues de modèles statistiques. Ainsi, cette tâche vise à simuler les activités des occupants et leurs impacts énergétiques, notamment sur la consommation d'électricité domestique dans des ménages spécifiques. Dans ce travail, les résultats des modèles précédents sont réutilisés et combinés avec une simulation de type Monte Carlo pour simuler les séquences d'activités des occupants et leur consommation d'électricité sur une longue période. En particulier, les profils d'activité sont déterminés à partir du modèle d'estimation des activités. Ces profils sont utilisés pour prédire l'information des activités des occupants. Une fois ces activités déterminées, le modèle d'utilisation des appareils est utilisé pour simuler l'électricité utilisée par chaque appareil dans cette activité. Différentes activités

ont été utilisées pour la validation dans les études de cas. L'électricité consommée par les appareils impliqués dans les activités est simulée sur une longue période. Ensuite, ces résultats sont comparés aux mesures pour vérifier la performance du modèle. Il est démontré que la méthodologie proposée est utile pour simuler la consommation électrique des appareils dans des maisons spécifiques. Cependant, cette méthode est limitée dans le traitement des changements d'habitudes des occupants dans l'utilisation des appareils, changements pouvant intervenir dans le temps ou dépendre de la saison.

En résumé, la méthodologie proposée comprend trois modèles pour aborder différents sujets liés à la modification d'activités contextualisées des occupants dans les bâtiments résidentiels. Premièrement, un modèle d'estimation des activités basé sur un réseau bayésien est proposé pour estimer les activités des occupants à partir des données des capteurs et du contexte de la maison. Ensuite, un modèle basé sur les conséquences est proposé pour chaque activité afin de prédire la consommation électrique des appareils liés à cette activité. Enfin, les deux modèles ci-dessus sont combinés pour simuler les activités des occupants et leurs impacts énergétiques, ce qui pourrait être utile dans la sélection des stratégies de rénovation des bâtiments. Cette méthodologie a été mise en œuvre dans des études de cas spécifiques, et les résultats montrent que la méthodologie peut estimer plusieurs activités des occupants dans les logements et prédire l'électricité consommée par les appareils associés. En outre, en analysant les données antérieures sur les activités des occupants et leurs effets sur l'environnement, les scientifiques pourraient fournir un retour d'information et des recommandations aux occupants pour qu'ils améliorent certaines de leurs pratiques. De cette façon, ils pourraient à la fois réduire leur consommation d'énergie et améliorer leur confort dans les bâtiments.

Keywords : modélisation de l'activité des ménages, simulation énergétique des bâtiments, réseau bayésien, l'apprentissage artificiel, l'activité contextualisée des occupants, bâtiment résidentiel instrumenté.

Nomenclature

AR	Activity Related
BEMS	Building Energy Management System
BN	Bayesian Network
BPS	Building Performance Simulation
CBBN	Consequences-Based Bayesian Network
CO₂	Carbon Dioxide
DT	Decision Tree
DTC	Dynamic Time-series Clustering
EMMS	Energy Management and Monitoring System
EPBD	Energy Performance of Building Directive
EPC	Energy Performance Certification
EU	European Union
GHG	Greenhouse Gases
HVAC	Heating, Ventilation and Air Conditioning
IG	Information Gain
kWh	Kilo-Watt-Hours
ppm	Parts Per Million
RE	Relative Error
SVM	Support Vector Machine
TUS	Time Use Survey

Table of contents

Nomenclature	xi
List of Figures	xiv
List of Tables	xx
Introduction	1
1 Increasing impacts of occupant activities in residential buildings	5
1.1 Energy consumption in residential buildings	6
1.2 Occupant activities and energy consumption in residential buildings . . .	10
1.3 Conclusion	20
2 Existing works dealing with occupant activity modelling	21
2.1 Modeling occupant activity in residential buildings	21
2.2 Existing approaches for occupant activity modeling	22
2.3 Conclusion	42
3 Problematic: Contextualized activities models in residential buildings	43
3.1 Introduction	44
3.2 General definitions and concepts	44
3.3 Contextualized activity models: opportunities and challenges	49
3.4 Problematic: contextualized activities modeling in dwellings	52
3.5 Conclusion	57
4 Data monitoring and case studies descriptions	59
4.1 Case study descriptions	60
4.2 Sensors for occupant activity estimation	64
4.3 User interface application to collect activities labels	70
4.4 Conclusion	75

5	Estimation of contextualized activities	77
5.1	Introduction	78
5.2	Methodology to estimate contextualized activities	78
5.3	Results: the cooking activity	96
5.4	Conclusions	115
6	Estimation of the electricity consumption in occupants' activities	117
6.1	Introduction	118
6.2	Problematic: modelling the electricity consumption of appliances in houses	118
6.3	Methodology to model the electricity consumption of domestic appliances for a related activity	122
6.4	Implementation and results	130
6.5	Conclusion	145
7	Application to the design phase in building renovation strategy	147
7.1	Introduction	148
7.2	Contextualized occupants' activities in evaluations of building renovation strategies	148
7.3	Simulation of occupants' activities and their impacts on related appliances	151
7.4	Implementation and results	157
7.5	Conclusion	175
	Conclusion and perspectives	179
	Bibliography	185

List of Figures

1.1	Final energy consumption by sector in EU, 2018 (Eurostat, 2018)	6
1.2	Energy consumption by end-use in residential buildings in Europe (Eurostat, 2018)	8
1.3	Two categories of occupant behaviors (Hong et al., 2017)	11
1.4	General parameters for energy simulation (Maile et al., 2007)	14
1.5	Profiles of the merged activity types whilst people at home (Wilke et al., 2013)	16
1.6	An example of differential explanations for occupants to understand the impacts of their behaviors on doors/windows (Pal et al., 2019)	19
2.1	The general model for occupant activity estimation	22
2.2	An example of occupancy schedules (number of occupants) input in the software Pleiades	23
2.3	Stochastic model of occupant activity of Wilke et al. (2013)	30
2.4	Probabilistic model of occupant behavior of Aerts et al. (2014)	30
2.5	Two groups of activity in dwellings according to Aerts et al. (2014)	31
2.6	Structure of an agent-based model according to Macal et North (2010)	33
2.7	The context factors of an agent’s activity from Ha et al. (2006)	34
2.8	(a) Dynamic BN (b) Elements of Bayesian Network (Tijani et al., 2015)	38
2.9	An example of static Bayesian Network (Amayri et al., 2016)	38
2.10	Dynamic Bayesian Network for the movements of the door in office (Tijani et al., 2015)	39
3.1	Illustration of an activity	47
3.2	Example of statistical activity profile in a residential building studied by Wilke et al. (2013)	48
3.3	Differences between representative activities and contextualized activities approaches	50
3.4	Applications of contextualized activities in building renovation	52

3.5	General methodology to model contextualized occupant activities and their related electricity consumptions	56
4.1	Building plan of the case study in Cestas, France	61
4.2	Building plan of the case study in Bordeaux, France	63
4.3	The positioning of PIR sensors, power sensors, contact sensors and some of their images in Cestas site	68
4.4	The position of PIR sensors, power sensors, contact sensors and some of their images in the Bordeaux site	69
4.5	Protocol and sensors data management system in this work	70
4.6	The user interface of the developed mobile application for collecting labels	71
4.7	The calendar of labeled/unlabelled days (left) and an example of an unlabelled day (right) in the developed application	72
4.8	An example of a labeled day (left) and an example of the function of labeling a multi-room activity (right) in the developed application	73
4.9	The features of saving typical patterns of activities (right) and auto-fill similar days (left) in the developed application	73
5.1	The relationship between occupant activity and context	79
5.2	An example of cooking estimation problem	80
5.3	The general methodology to estimate contextualized occupants activities	81
5.4	Preprocessing data step	83
5.5	Examples of the five levels (very low - yellow, pink - low, violet - medium, green - high, gray -very high) of a microwave determined by k-means in the Cestas site	90
5.6	General elements of a Bayesian Network: structure and parameters (Amayri, 2017)	92
5.7	Conceptual model of occupants' activities and their consequences	94
5.8	The structure of the proposed Consequences-based Bayesian Network (CBBN)	95
5.9	Sorted Information Gain between features and three activities: cooking breakfast, cooking lunch and cooking dinner in the Cestas household . .	99
5.10	Consequences-Based Bayesian Networks for the cooking activities in Cestas site	101
5.11	Recursive feature elimination with cross-validation with cooking activity in Cestas household (A: breakfast, B: lunch and C: dinner)	105
5.12	The learning curves with the cooking activities in the Cestas household (A: breakfast, B: lunch and C: dinner)	106

5.13	Some examples of measurements (toaster, microwave, coffee machine and motion) from 10/09/2019 06:00:00 to 12/09/2019 23:30:00 in Cestas	108
5.14	Distributions of the cooking activities deduced from 1000 simulations of the cooking activities from 10/09/2019 06:00:00 to 12/09/2019 23:30:00 in Cestas. It gives the number of simulation that the cooking activity happens (red color) and does not happen (in blue color) for each timestep.	109
5.15	Comparison between the observed and the estimated cooking profile in the Cestas household	110
5.16	Sorted Information Gain between features and three activities: cooking breakfast, cooking lunch and cooking dinner in Bordeaux house	111
5.17	Consequences-Based Bayesian Networks for the cooking activities in Bordeaux site	112
5.18	Some measurements (robot, microwave, kettle, motion) from 17/04/2021 06:00:00 to 18/04/2021 23:30:00 in Bordeaux site	113
5.19	Distributions of the cooking activities deduced from 1000 simulations of the cooking activities from 17/04/2021 06:00:00 to 18/04/2021 23:30:00 in Bordeaux site. It gives the number of simulation that the cooking activity happens (red color) and does not happen (blue color) for each timestep.	114
5.20	Comparison between the observed and the estimated cooking profile in the Bordeaux site	115
6.1	Illustration for the appliances consuming electricity during the activity or during a part of the activity.	124
6.2	The proposed Bayesian Network to represent the relationship between an activity and the electricity consumed by an appliance	125
6.3	Illustration for the appliances consuming electricity during the activity or during a part of the activity.	126
6.4	An example of detecting a starting timestep and a stopping timestep of an appliance consuming electricity after an activity ends	127
6.5	The structure of Bayesian Network for predicting AR consumption of appliances for a given activity	128
6.6	Algorithm to simulate the AR consumption level of an appliance for a related activity	129
6.7	Examples of the electricity used by the appliances (robot, yogurt maker, foods processor) that are rarely used in the Cestas household	131

6.8	The estimated CPDs, distributions of values and visualisation of the of AR consumption of the coffee machine for cooking breakfast in the Cestas household	132
6.9	The estimated CPDs, distributions of values and visualisation of the AR consumption of the toaster for cooking breakfast in the Cestas household	133
6.10	The estimated CPDs, distributions of values and visualisation of the AR consumption of the microwave for cooking breakfast in the Cestas household	134
6.11	The estimated CPDs, distributions of values and visualisation of the AR consumption of the microwave for cooking lunch in the Cestas household	134
6.12	The estimated CPDs, distributions of values and visualisation of the AR consumption of the microwave for cooking dinner in the Cestas household	135
6.13	The estimated CPDs, distributions of values and visualisation of the AR consumption of the coffee machine for cooking lunch in the Cestas household	136
6.14	The estimated CPDs, distributions of values and visualisation of the AR consumption of the coffee machine for cooking dinner in the Cestas household	136
6.15	The Bayesian Network structure for predicting the electricity consumption related to the cooking activity in the Cestas household . . .	137
6.16	Comparisons between simulated and observed daily AR consumption of appliances for cooking breakfast in the Cestas household (A: coffee machine, B: toaster, C: microwave)	139
6.17	Comparisons between simulated and observed daily AR consumption of appliances for cooking lunch in the Cestas household (A: coffee machine, B: toaster, C: microwave)	140
6.18	Comparisons between simulated and observed daily AR consumption of appliances for on the cooking dinner in the Cestas household (A: toaster, B: microwave)	141
6.19	Comparisons of conditional probability distributions of AR consumption levels of the coffee machine for cooking lunch between months in the Cestas household.	142
6.20	Comparisons of conditional probability distributions of AR consumption levels of the microwave for cooking lunch between months in the Cestas household.	142

6.21	Comparisons between simulated and observed daily AR consumption of the washing machine for washing clothes in Cestas household	144
6.22	Comparisons between simulated and observed daily AR consumption of the tumble dryer for drying clothes in Cestas household	144
6.23	Comparisons between simulated and observed daily AR consumption of the dishwasher for washing dishes in Cestas household	145
6.24	Comparisons between simulated and observed daily AR consumption of the audio and video for entertaining in Cestas household	145
7.1	Simulation of contextualized activities and its application in building renovation	151
7.2	Proposed methodology to simulate the electricity consumed by appliances	153
7.3	A conceptual model of the relationship between an activity, a timestep and a type of day	154
7.4	Flowchart of the algorithm to simulate occupants activities and the AR electricity consumption of involved appliances	156
7.5	Different cooking activity profiles between days in the Cestas household from 2019-09-01 to 2019-12-30	159
7.6	Different cooking activity profiles between weekdays and weekend in the Cestas household from 2019-09-01 to 2019-12-30	159
7.7	Comparison between measured and simulated weekly electricity consumption of the microwave in the Cestas household from 2020-01-01 to 2020-02-01	161
7.8	Comparison between measured and simulated weekly electricity consumption of the toaster in the Cestas household from 2020-01-01 to 2020-02-01	161
7.9	Comparison between measured and simulated weekly electricity consumption of the coffee machine in the Cestas household from 2020-01-01 to 2020-02-01	162
7.10	Weekday and weekend average electricity consumption patterns for the coffee machine with cooking profiles in the Cestas household from 2020-01-01 to 2020-02-01	164
7.11	Weekday and weekend average electricity consumption patterns for the toaster with cooking profiles in the Cestas household from 2020-01-01 to 2020-02-01	165

7.12	Weekday and weekend average electricity consumption patterns for the microwave with cooking profiles in the Cestas household from 2020-01-01 to 2020-02-01	166
7.13	Weekday and weekend average electricity consumption patterns for the tumble dryer with drying clothes profiles in the Cestas household from 2020-01-01 to 2020-02-01	167
7.14	Weekday and weekend average electricity consumption patterns for the washing machine with washing clothes profiles in the Cestas household from 2020-01-01 to 2020-02-01	168
7.15	Weekday and weekend average electricity consumption patterns for the dishwasher with washing dishes profiles in the Cestas household from 2020-01-01 to 2020-02-01	169
7.16	Weekday and weekend average electricity consumption patterns for the audio video devices with entertaining profiles in the Cestas household from 2020-01-01 to 2020-02-01	170
7.17	Different cooking activity profiles between days in the Bordeaux household from 2021-04-05 to 2021-05-16	171
7.18	Different cooking activity profiles between weekdays and weekend in the Bordeaux household from 2021-04-05 to 2021-05-16	172
7.19	Comparison between measurement and simulated weekly electricity consumption of the microwave in Bordeaux site from 2021-05-26 to 2021-06-24	173
7.20	Comparison between measured and simulated weekly electricity consumption of the kettle in the Bordeaux household from 2021-05-26 to 2021-06-24	173
7.21	Comparison between measured and simulated weekly electricity consumption of the oven in the Bordeaux household from 2021-05-26 to 2021-06-24	174
7.22	Comparison between measured and simulated weekly electricity consumption of the toaster in the Bordeaux household from 2021-05-26 to 2021-06-24	174

List of Tables

2.1	Summary of existing approaches in occupant activity model	41
3.1	Proposed context elements at home completed from Kashif (2014) . . .	45
4.1	Household members characteristics in the case study in Cestas, France .	61
4.2	Household members characteristics in the case study Bordeaux, France .	63
5.1	Summary of the generated features each activity in each location	87
5.2	Manual thresholds of air temperature, relative humidity, luminosity and CO ₂ concentration	89
5.3	Manual thresholds of features related to occupants' movements and doors/windows openings	89
5.4	An example of the activity-location information collected from the Cestas household through questionnaires	97
5.5	Groups of selected features for estimating cooking breakfast, cooking lunch and cooking dinner in the Cestas household	100
5.6	Examples of estimated conditional probability distributions of the microwave in cooking breakfast and cooking lunch in the Cestas household	102
5.7	Comparison of F1-score among different approaches for the cooking estimation in the Cestas household	103
5.8	Results of the estimation of some occupants' activities in the Cestas household	107
5.9	Comparison of F1-scores among different approaches in cooking estimation in Bordeaux site	113
6.1	Possible appliances used in occupants' activities in the Cestas household	130
6.2	The type of relation between activities and the AR consumption of appliances in the Cestas household.	132
6.3	Testing RE of the simulated AR consumption of appliances in the Cestas household	138

6.4	Testing RE of the simulated electricity consumption of related appliances in several activities in the Cestas household	143
7.1	Comparison between measured and simulated weekly energy consumption of studied appliances in the Cestas household from 2020-01-01 to 2020-02-01	163
7.2	Comparison between measured and simulated weekly energy consumption of studied appliances related to the cooking activity in the Bordeaux household from 2021-05-26 to 2021-06-24	175

Introduction

Context

Since the Industrial Revolution in the mid-18th century, along with the development of high-technology applications, human activities have had dramatically increase in influence on the environment, especially climate. Productions of greenhouse gases (GHGs) has risen by 260% since 1750 (Julia, 2012). The increased emission of GHGs such as CO₂ concentration *via* the combustion of fossil energy, results in global warming, one of the most serious issues that the world is facing today. To tackle these issues, the European Union (EU) has formulated targets for 2020 since 2012. By 2020, the GHGs emission in the EU should be reduced by 20% from 1990 levels and 20% of primary energy consumption should be saved (Directive Parliament, 2012). Recently, in 2016, the target of the reduction of GHGs emissions was set to 45%, and the reduction of primary energy consumption was targeted to be at least 30% by 2030 (European Commision, 2016). While energy consumption accounts for over 80% of global GHGs, in Europe, residential buildings contribute more than 25% of them in 2018 (Eurostat, 2018). It leads to the necessity of reducing energy consumption in residential buildings.

To deal with it, European Commision (2015) stated that the goal of reducing energy consumption in buildings could be achieved by three main approaches:

- Concerning building's envelopes: building envelopes, or shell, innovative building materials, components, and designs of buildings.
- Dealing with building installations: Smart energy monitoring and management systems (EMMS) for heating, cooling, ventilation, and other systems can be used in new constructions or renovated buildings.
- Taking into account renewable energy technologies: Nearly-zero-energy buildings

take advantage of renewable energy technologies.

There is a wide range of technologies that can be used in each of these approaches. For example, using a combination of optimal design techniques to minimize heat gain in summer and heat loss in winter, using passive heating and cooling techniques or using daylight to reduce lighting needs. However, there is still potential for enhancing these approaches because building performance is not only influenced by its physical characteristics but also governed by occupant activities (Jia et al., 2017). In recent years, efforts have been made to improve the EMMS in residential buildings by considering occupant activities as a part of systems. The goal is to not only reduce energy consumption but also satisfy human comfort in buildings. For example, occupants' locations help HVAC systems determine which zone needs heating/cooling. The lighting systems could work efficiently with occupants' presence and their activities such as sleeping, cooking.

Building energy simulation is a useful tool to optimize building's envelopes. During the building design phase, energy simulations can be repeatedly performed to support architecture in sizing HVAC systems and selecting building elements. To renovate an existing building, energy simulation tools can be used to verify its current energy performance and to evaluate impacts of potential retrofitting actions.

However, despite the development of building energy performance simulation, there is still a significant difference between simulated and actual energy consumption in residential buildings. Many scientists believe that occupant activities are major factors in explaining discrepancies between simulated and actual energy consumption in residential buildings. Better information on occupant activities might improve the performance of energy simulation tools and reduce this gap (Guerra-Santin et Itard, 2012; Julia, 2012; Hong et al., 2017). On the other hand, by analyzing energy impacts of occupants' activities, we could give feedback and recommendations to occupants. This information is useful for occupants to reduce energy consumption and improve their comfort.

In summary, modeling occupant activities is useful in many applications such as:

- **Energy Performance Verification:** it is an essential task in renovating existing buildings. To achieve it, energy simulations are performed to assess current

building conditions. Occupant activity models can contribute useful information to these simulations.

- **Simulation during design phase:** energy simulation tools are helpful to predict building energy performance in the design phase. Occupant activity models can give useful information to improve the accuracy of these simulations.
- **Energy monitoring and management systems (EMMS):** objectives of EMMS are to satisfy occupants' comforts and reduce energy consumption. To work properly, the EMMS can not neglect occupants' activities. For example, it is wasted energy for turning on light when nobody is at home. Therefore, modeling occupants' activities is useful to these systems.
- **Energy feedback, recommendation:** investigating occupants' activities and their energy impacts are helpful to enhance occupants' awareness and recommend better practices to them.

Recently, many scientists have been focusing on modeling occupants' activities to improve energy simulations (Page et al., 2008; Tanimoto et al., 2008; ?; Vorger, 2014; Aerts et al., 2014). In the beginning, fixed schedules of activities were used popularly but these scenarios were applied for households with different characteristics (members, type of building, professional, *etc.*). To consider characteristics, stochastic models with statistical data have been proposed to model occupant activities and their energy impacts. This approach is convenient to implement and consider the diversity of households due to the availability of statistical data. However, this approach aims to model representative activities of similar households and their results are not specific and are not verifiable.

To deal with these issues, data-based models with measurement have been proposed to study occupants' activities. This approach aims to determine patterns and systematic relationships between related variables using data mining and machine learning techniques. These models are specific to particular households so that their results are verifiable. However, most of these models are black-box, they do not provide human-understandable explanations. In addition, they are deterministic models so that they do not consider the diversity of occupants' activities. Multi-agents is another popular approach to consider interactions between occupants and buildings. This approach is developed to satisfy human needs. This approach can work well with

specific households and its outputs are easy to understand. But, this approach is too complex and difficult to implement in the case of households with many members.

Problematic

The primary objective of this research is to establish a methodology to generate estimators and simulators of human practices in specific dwellings. It aims to contribute in particular to improvements of not only energy simulations but also energy feedbacks in residential buildings. These estimators and simulators must be verifiable and understandable, which are limited in existing approaches.

In this thesis, following works are focused on:

- A) estimating contextualized occupants' activities with a set of basic sensors, context information and activity labels collected from occupants. Necessary sensors should be detected for the reduction of redundant sensors. Finally, estimated results should be compared to observed labels to validate model accuracy. This model is presented in chapter 5.
- B) estimating energy impacts of contextualized occupants' activities with a set of basic sensors, context information, and activity labels collected from occupants. Results should be compared to measurements for model verification. This model is presented in chapter 6.
- C) simulating contextualized occupants' activities and their energy impacts, which are useful in building renovation. Simulated results should be compared to measurements for validation. This model is presented in chapter 7.

Chapter 1

Increasing impacts of occupant activities in residential buildings

Contents

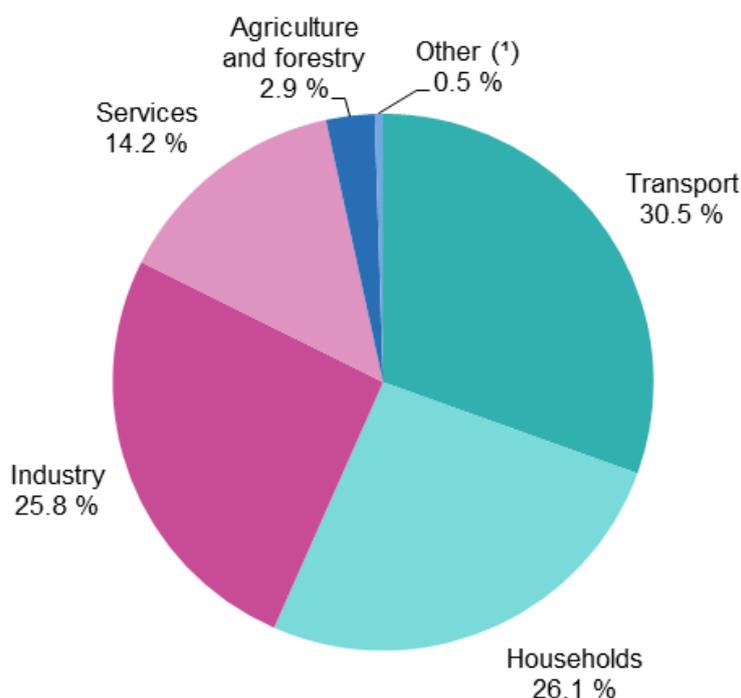
1.1 Energy consumption in residential buildings	6
1.1.1 The important role of residential buildings in energy consumption	6
1.1.2 Energy and human practices in residential buildings	7
1.1.3 Energy efficiency in residential buildings	8
1.2 Occupant activities and energy consumption in residential buildings	10
1.2.1 Impacts of occupant activities on energy consumption	10
1.2.2 Occupant activity models in the simulation of energy performance	12
1.2.2.1 Energy simulations in residential buildings	14
1.2.2.2 Energy Monitoring and Management Systems (EMMS)	17
1.2.2.3 Occupant feedbacks and recommendations	18
1.3 Conclusion	20

1.1 Energy consumption in residential buildings

1.1.1 The important role of residential buildings in energy consumption

The residential buildings sector plays a crucial role in energy consumption. Households consume energy to achieve various purposes and satisfy occupants comfort in buildings. In Europe, a survey examined by Eurostat (2018) showed that the residential buildings contribute more than 26% of the total final energy consumption in 2018 (see figure 1.1). Therefore, the sector of residential buildings is a major lever to reduce energy consumption.

Final energy consumption by sector, EU-27, 2018
(% of total, based on tonnes of oil equivalent)



(*) Data on "international aviation" are not included in category Transport and hence are included in the category "Other".

Source: Eurostat (online data code: nrg_bal_s)



Figure 1.1 – Final energy consumption by sector in EU, 2018 (Eurostat, 2018)

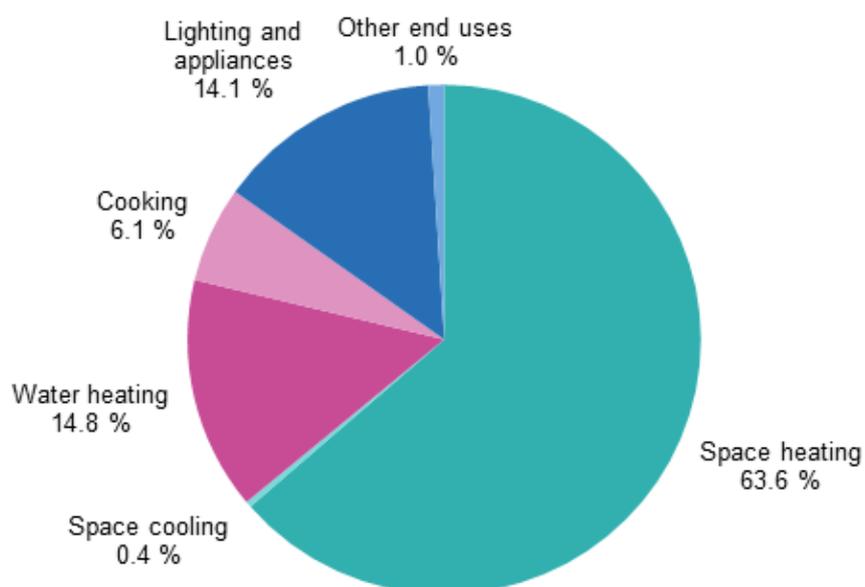
1.1.2 Energy and human practices in residential buildings

Recently, many studies have examined the factors that have significant influences on energy consumption in residential buildings (Page et al., 2008; Janda, 2011; Eurostat, 2020). They are interested in understanding more precisely the energy consumption phenomena to address the major problems resulting in wasted energy in residential buildings. In literature, Eurostat (2020) mentioned five major categories of energy use in residential buildings:

- **Space heating and space cooling:** that is the energy required to balance the thermal gains (HVAC systems, solar gains, internal gains, *etc.*) and losses (ventilation, transmission, infiltration, *etc.*). The purpose is to maintain occupants in comfortable indoor conditions (temperature, humidity, *etc.*).
- **Domestic hot water:** that is the energy required to heat water to a comfortable temperature for the occupant.
- **Cooking:** that is the energy required to support occupants during cooking (gas, electricity consumption of fridge, robot, microwave hot plates, *etc.*).
- **Lighting and electrical appliances:** that is the energy required to operate the electrical appliances, which could be used for different purposes such as entertainment, study (computer, TV, *etc.*), and artificial lighting.
- **Other end uses:** that is the energy required for another end uses.

Specifically, a field study made by Eurostat (2020) shows that in 2018, in residential buildings in Europe, the final energy for space heating and cooling takes up more than 64% of the total energy consumption. The remaining energy is mostly for domestic hot water heating (14.8%), specific electricity (14.1%) of plug load, lighting and cooking (6.1%). This distribution is shown in figure 1.2.

Final energy consumption in the residential sector by use, EU-27, 2018



Source: Eurostat (online data code: nrg_bal_c)

eurostat 

Figure 1.2 – Energy consumption by end-use in residential buildings in Europe (Eurostat, 2018)

The demand for saving energy encourages scientists to improve energy efficiency in residential buildings and reduce energy consumption in these end-uses.

1.1.3 Energy efficiency in residential buildings

Energy efficiency aims reducing the amount of energy consumed while maintaining or improving the quality of services provided in buildings. European Parliament (2010) defined nearly zero-energy building (nZEB) as a standard for reaching energy efficiency in buildings: *"a building that has a very high energy performance. The nearly zero or very low amount of energy required should be covered to a very significant extent by energy from renewable sources, including energy from renewable sources produced on-site or nearby"*. Sharmin et al. (2014) stated some factors that can improve both energy consumption and occupants comforts:

- space heating/cooling, water heating and electric appliances;
- household water usage;
- thermal performance of building envelope;
- indoor air quality.

In Europe, ambitious action plans have been implemented at the national level focusing on the residential buildings sector, which includes many objectives: to increase the mobilization of all building actors, to define more restrictive regulatory rules, to provide financial incentives for households, to offer significant financial support to building owners, and to support research and development projects. The **European Energy Performance of Building Directive (EPBD)** sets minimum requirements of energy performance for new buildings and undergoing retrofitting buildings. In 2002, the **EPBD** was adopted to encourage the development of energy efficiency of buildings (European Parliament, 2002). EPBD was continued to be recast in 2010 and amended in 2012 (European Parliament, 2010, 2012). Besides, for the evaluation of building energy performance, **EPBD** provides **Energy Performance Certificate (EPC)** as a framework for the assessment of the building energy performance. Importantly, the calculation method considers differences between the retrofitted buildings, the buildings undergoing renovations and the existing buildings. To take account of existing buildings, these **EPCs** are compulsory for residential buildings that are rented or sold. However, the **EPCs** do not require these buildings to meet some energy efficiency levels but assess their current energy performance. The purpose is to compare the energy performance between the existing buildings and renovated buildings.

In France, the authorities also established several standards and regulations concerning new buildings. The latest French thermal regulation which defines energy performance standards of buildings is the "**Réglementation Environnementale 2020**" (RE 2020). This regulation extends the focus to six main energy-uses: heating, cooling, lighting, domestic hot water, auxiliary equipment (ventilators and heat pumps), and electrical appliances.

In recent years, efforts have been made to develop innovative and smart building systems to improve the energy efficiency in buildings. For example, Cheung et al. (2005) proposed an integrated passive design approach to reduce the cooling demand in the context of Hong Kong. The results showed that more than 30% of the energy

could be saved by adding extruded polystyrene thermal insulation in walls. In another article, Zhou et al. (2014) investigated a real-time energy control approach for a home management system for household economic benefits in UK. However, there is still potential for enhancing these systems. Building energy performance is influenced not only by its physical characteristics but also by many other factors such as operations, maintenance, occupant activities, indoor and outdoor environmental conditions (Jia et al., 2017).

1.2 Occupant activities and energy consumption in residential buildings

1.2.1 Impacts of occupant activities on energy consumption

In dwellings, occupants' actions are what a person or a group does or has done, which can be represented by controlling events on the indoor equipment. Activity is a consistent set of actions with the same intention. An activity covers a time slot whereas actions are events. For example, occupants can do actions such as opening doors/windows or turning on the light. They can also perform activities like cooking, studying and watching television, which include sets of actions with different possible sequences. As mentioned above, activities are considered as an important factor regarding energy performance in residential buildings. Page et al. (2008) explained that the impacts of occupants on building energy performance could be assessed by their presences (internal heat gains, emissions of CO₂, water vapor, *etc.*), the activities they perform (daily activities such as cooking, lighting or actions of door/windows) and their interactions with the controls of building systems. Janda (2011) also stated that "*buildings don't use energy: people do*". People use energy to satisfy specific daily life activities such as preparing food, supplying heat and light, and maintaining health and comfort (Kashif et al., 2011).

Hong et al. (2017) grouped occupant activities into two categories: adaptive behaviors and non-adaptive behaviors, which are described in figure 1.3.

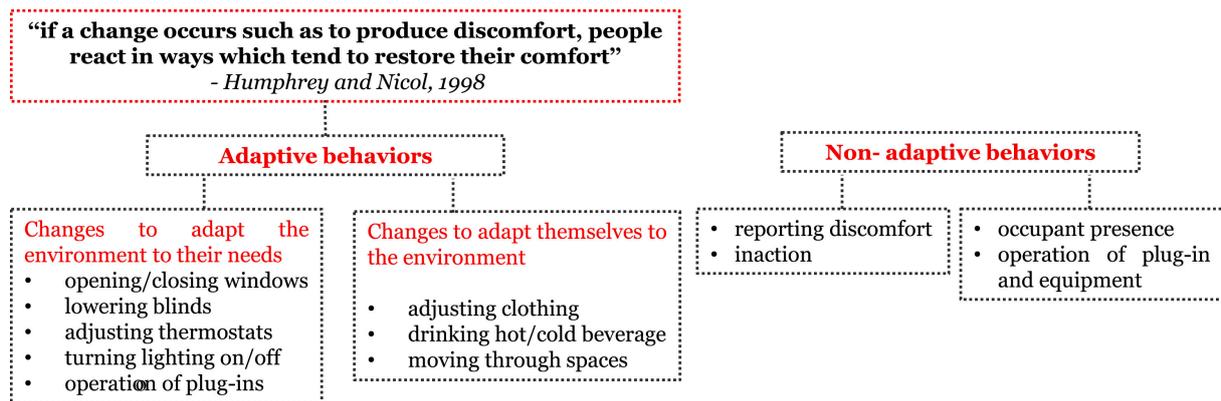


Figure 1.3 – Two categories of occupant behaviors (Hong et al., 2017)

Humphreys et Nicol (1998) stated that *"if a change occurs such as to produce discomfort, people react in ways which tend to restore their comfort"*. Thus, adaptive behaviors are performed by occupants to make changes to adapt the environment to their needs/comfort. To do it, Hong et al. (2017) explained that occupants either adapt:

- the indoor environment to their needs or preferences: opening or closing windows/doors, turning lighting on/off, *etc.*
- their behavior to their environment: adjusting clothing or moving through the building, drinking, *etc.*

Non-adaptive behaviors such as occupant presence, operations of electrical equipment, as well as reporting complaints regarding discomfort influencing the final building performance. These actions are more related to occupant habits such as daily activities (cooking, cleaning, relaxing, *etc.*) or occupant preferences than adapting the environment for their comforts. However, uses of plugs also refer to adaptive operations of the personal heater, fans, and electrical appliances.

But, how do occupant activities affect building energy performance? Hong et al. (2017) mentioned that the relationship between occupant activities and building energy performance is two-fold. On the one hand, environmental conditions trigger occupants to interact with the building control systems, causing changes to energy loads. On the other hand, adaptive activities performed by occupants make changes to indoor environmental conditions. For example, occupant activities influence the energy demand for space heating in two ways: the presence of people with its impacts (CO₂, heat gain, *etc.*) leads to changes in indoor temperature and heating/cooling demand, occupants adjust the heating system, windows or other appliances to maintain their

comfort. In another example, occupants might open the window to achieve better comfort conditions regarding the CO₂ concentration. This action leads to changes in the building's environmental conditions and energy flows: an increase in heating energy consumption due to an increase in the ventilation airflow rate.

Many studies examined that energy consumption at household level is largely due to occupant activities, and this is an international phenomenon (Fathi et Kavooosi, 2021; Chen et al., 2021; Pang et al., 2020). Particularly, Richardson et al. (2008) analyzed the survey data of energy consumption in the UK to build a model for the energy demand of domestic buildings. They showed that the energy consumption of buildings was highly related to the profiles of occupants and their activities while they are at home. Andersen et al. (2009) also conducted a study in Denmark, the result showed that the use of heating was strongly related to the action of windows and the average age of inhabitants, the respondents' thermal sensation and gender also influenced the use of lighting. In another article, Lin et Hong (2013) used statistical analysis to study occupancy and activities patterns. Based on the results, they stated that "*occupants with a wasteful work style consumed up to double the energy of the standard occupants, while austere work-style occupants used half of the energy of the standard occupants*". Besides, Yousefi et al. (2017) investigated the influence of occupant activities in residential buildings in Iran and they showed that occupant activities could change the heating and cooling loads up to 90% in a warm climate.

1.2.2 Occupant activity models in the simulation of energy performance

As mentioned in section 1.1.3, improving energy efficiency is a crucial objective not only to reduce energy consumption and enhance human comfort in residential buildings. To achieve this goal, many approaches have been proposed, such as using building energy simulation for the design optimization, installing smart devices for energy monitoring and HVAC systems management, or considering renewable energy technologies (Jia et al., 2017; Nagy et al., 2015). Due to the impacts of occupant activities on building energy performance, many efforts have been made to integrate these activities into proposed approaches. The goal is to help the approaches to work more realistically and

more precisely (Hong et al., 2017; Leroy et Yannou, 2018).

Occupants' activities in these approaches can be divided into two modelling approaches:

- *Representative occupant activities*: those are the typical patterns, which represent the activities in similar households. They are useful in implementing the model on a population scale when the diversity of dwellings needs to be concerned.
- *Contextualized occupant activities*: those are the activities of occupants in a specific type of house or apartment. They are useful in implementing applications and assessing the impacts of occupants on particular households.

This information of occupant activities is useful in many applications regarding building energy. Jia et al. (2017) presented three major benefits of the capability of detecting and modeling occupant activities earlier in the design:

- Integrating occupant activity information in simulation should lead to more realistic energy consumption results
- Occupant activity models provide fundamental support to building control systems
- Actual occupant information would improve building performance and services

Hong et al. (2017) also stated that occupant activity models could support building energy designers, modelers, operators, and managers to develop specific energy efficiency measures and to evaluate technology adoption levels by taking into account the occupant activity impacts. Furthermore, these studies can help to evaluate retrofit measures at the building level and to assess the potential for renovation penetration.

Amayri (2017) also presented six services that could gain benefits from the estimation of occupancy or occupant activities, which comprise: *simulation for design, contributing to physical models (heat gains, heat loss, etc.) of the buildings, explanations for the energy consumption in the building, behavioral data model for recommending occupant regarding energy, handling discrepancies in building managements, and mirror and key performance indicators.*

In short, modeling occupant activity is useful and important in:

- energy simulation in residential buildings to improve building design and support Energy Performance Verification process;

- energy monitoring and management systems to support HVAC control systems, electricity load management and reduce wasted energy;
- occupant feedbacks to encourage better practices regarding energy impacts by enhancing human awareness.

1.2.2.1 Energy simulations in residential buildings

Energy simulation tools have proved to be effective and powerful in studying the energy performance and thermal comfort during the building's life-cycle. They combine dynamic simulation with computational modeling and draw upon the disciplines of heat and mass transfer, thermodynamics, lighting, building technology, thermal and visual comfort, numerical methods, and human activity. The purpose of these tools is to predict or estimate the energy performance of buildings with given parameters (see in figure 1.4). Balvedi et al. (2018) stated that building energy simulations are highly efficient and low-cost tools for analyzing and optimizing building design and system. During the design phase, they help to enable comparisons of different design alternatives and analysis of uncertain factors influencing the energy performance in buildings. Moreover, with existing buildings, energy simulations could be used to assess the current conditions and then evaluate refurbishment scenarios.

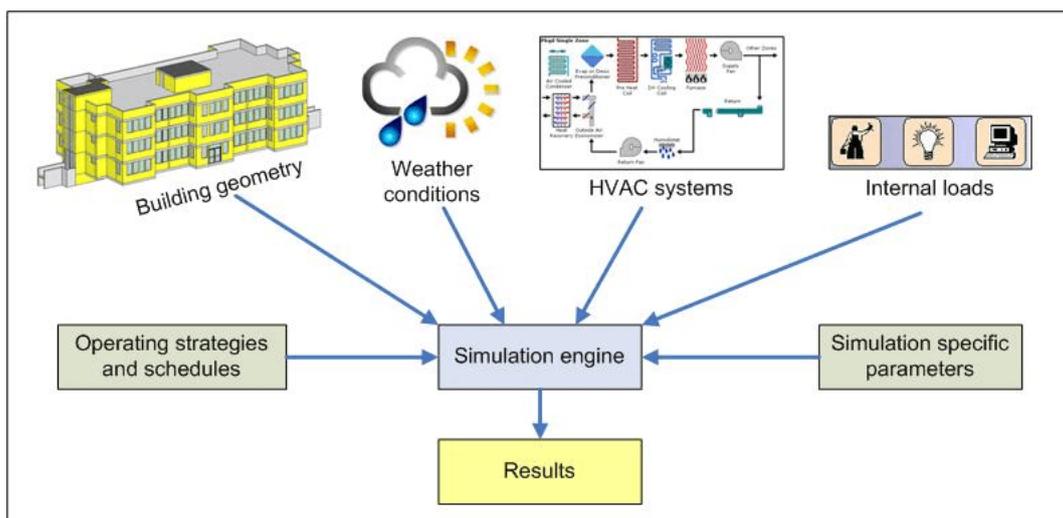


Figure 1.4 – General parameters for energy simulation (Maile et al., 2007)

Today, numerous energy simulation tools (EnergyPlus, TRNSYS, eQUEST, PLEIADES, etc.) are available to support the design phase of high-performance buildings. They can be divided into two types: *design tools*, which focus not only

on sizing HVAC equipment but also on the building's envelope, and *simulation tools*, which simulate the comfort analysis. In France, the Pleiades software is one of the most popular dynamic thermal simulation tools. These tools have become essential for the design stage or renovation of energy efficient buildings (Maile et al., 2007).

Despite the development of energy simulation tools, there are still issues, which need to be solved. Many energy simulation tools pay more attention to the building envelope than the interaction between occupants and buildings. Most of the world standards for buildings consider occupant impacts by typical patterns and deterministic scenarios. Regulations state that one building should meet all occupants' expectations and requirements. Thus, they consider predefined schedules (occupancy, activities) and responses to an exceeded threshold. These standards and scenarios are used as inputs for building energy simulation. Nevertheless, in low-energy buildings, the impacts of occupants become an important consumption factor. Occupants consume energy and affect indoor conditions through their interactions with the building. They cause a significant discrepancy between the simulated and actual energy use. Many studies stated that occupants' activities are the major elements in explaining discrepancies between simulated and actual energy consumption (Delghust et al., 2012; Gram-Hanssen, 2013; Laurent et al., 2013).

Yang et al. (2014) showed that the fixed operation schedule based on certain codes or the surveys could lead to energy wastes and occupants' discomfort since the different preferences (hobbies, habits, needs, *etc.*) of humans. Hoes et al. (2009) also studied that in a standard type of buildings, the metabolic heat gain was found to be an essential and sensitive input parameter for energy simulation tools to assess the building performance. The heat gain has a direct relation with occupants' activities. Therefore, they are important input parameters influencing the results of building performance simulations. In another study, Aerts et al. (2014) mentioned that a systematic error could exist between predicted and actual energy consumption. The consumption of high-energy buildings is usually underestimated, whereas the consumption of low-energy buildings is often overestimated. D'Oca et al. (2014) also examined that the impacts of occupants' activities have been attributed to discrepancies between predicted and actual building performance: the gap may exceed 300% in extreme cases. In this study, the energy consumption simulated could increase by 61%

when integrating the occupant activity model. By only modifying configurations of two parameters (windows opening and thermostat set-point schedules) related to occupants' activities, the simulated consumption could increase by 50%.

To deal with it, many scientists are interested in modelling representative occupants' activities in residential buildings instead of using scenarios. The most popular approach is the stochastic model, which simulates representative profiles of occupant activities based on statistical data. An example of these profiles is presented in figure 1.5, which was determined by Wilke et al. (2013) with a statistical model. This model is based on three types of time-dependent quantities: *the probability to be at home; the conditional probability to start an activity whilst being at home; the probability distribution function for the duration of that activity.*

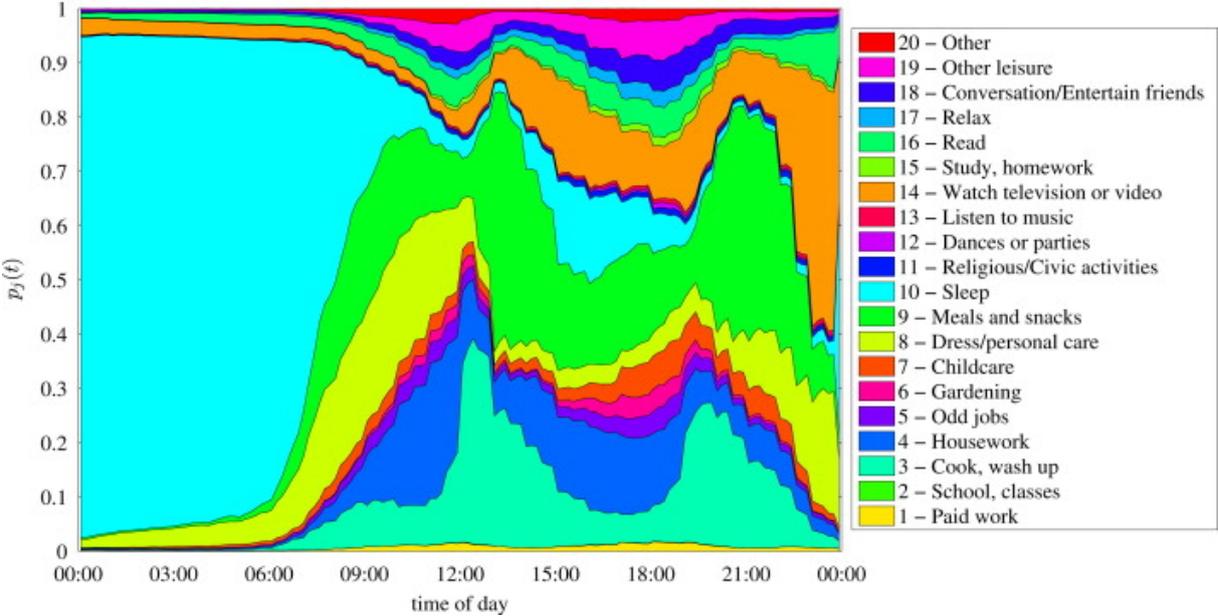


Figure 1.5 – Profiles of the merged activity types whilst people at home (Wilke et al., 2013)

Although the objective of the Energy Performance Certificate (EPC) is to compare buildings for policy reasons rather than accurately predict the actual energy consumption, it is essential to reduce the gap between predicted and actual consumption. Guerra-Santin et Itard (2012) stated that "*more accurate information on the actual user behavior and the identification on behavior patterns to build energy-user profiles might improve the energy predictions in the energy performance regulations. Although accurate energy prediction is not the aim of EPC, a better estimation of the actual energy performance and the actual energy savings expected from the introduction of*

tightening of building regulations, could be achieved".

In the design phase, representative activities are commonly used in energy simulation due to the flexibility and availability of the survey data. However, contextualized activities should be used in energy simulation to help the results to be more precise when renovating particular houses. It is useful in evaluating the building design proposals.

Besides, energy simulation is also essential in the energy performance verification process. To achieve this objective, contextualized activities could be used in simulation to assess the current conditions in existing buildings, then suggest the renovation plans to improve their energy performances.

1.2.2.2 Energy Monitoring and Management Systems (EMMS)

Modeling occupant activities is not only useful in the design phase but also in the operation phase. Integration of occupant activity can help smart systems (HVAC, lighting, *etc.*) to operate more effectively and to reduce wasted energy. For example, if people are not at home, HVAC and lighting systems should switch to standby mode to reduce energy use. Besides, considering occupant activity could help these systems operate more efficiently in case of unexpected situations, such as leaving work without turning off light.

In a study, Masoso et Grobler (2010) analyzed the energy consumption in commercial buildings in Botswana and South Africa. The research was examined in the hot and dry climates and the results show that more energy is used during non-working hours (56%) than during working hours (44%) and the golden rule is "*if you don't need it, don't use it*".

Recently, many scientists have proven that the estimation of occupant activity could significantly enhance the efficiency of the EMMS in buildings. For example, Agarwal et al. (2010) showed that detecting occupancy and using its information for the HVAC control module could reduce by 10% to 15% its energy consumption. In another research, Nagy et al. (2015) presented that artificial lighting systems could also gain benefits from occupant activities estimation. Particularly, by detecting occupant status and integrating it into the lighting systems, they achieved 38% energy savings but still maintained occupant visual comfort.

On the other hand, Dong et Lam (2014) also proposed a methodology for integrated building heating and cooling control based on the prediction of occupant activity patterns and local weather conditions. The results show that the energy consumption could be reduced by 30% in the heating season and by 18% in the cooling season, compared to a conventional scheduled temperature set-points control rules. In these systems, contextualized activities are popularly focused because they aim to optimize the energy performance and human comfort in specific houses.

1.2.2.3 Occupant feedbacks and recommendations

Besides the techniques interacting directly with the buildings, many scientists focus on occupants to improve energy efficiency in houses. Their objective is to enhance occupants' knowledge of their activities and their impacts on consumed energy. Then they could adjust their practices for reducing wasted energy and improving their comfort. Consequently, many energy feedbacks and recommendation approaches were proposed to occupants (Zhang et al., 2021; Wang et al., 2022; Trinh et al., 2021). The feedback tools aim to provide insights of the impacts of their activities on building energy, especially in energy consumption (Lee et al., 2020). This information helps occupants to understand and handle the finances related to energy usages, e.g. electricity bills, gas bills, *etc.* On the other hand, the recommendation systems could support occupants evaluate their activities regarding energy consumption and humans' comfort. Their past activities could be analyzed to suggest better practices (Ben-Haim, 2021; Varlamis et al., 2022).

For example, Hong et al. (2017) mentioned that using direct feedback and recommendations could improve occupants' activities and save up to 20% of energy consumption in houses. Based on the energy feedback and the recommendations, occupants tended to change their activities positively to save energy consumption and enhance their comforts.

Pal et al. (2019) proposed a framework to unmask the relationship between occupants' behaviors (actions of doors/ windows) and indoor conditions in buildings. As an example shown in figure 1.6, this approach helps occupants to better understand the consequences of their actions and activities. For example, opening doors and windows could increase the airflow between indoor and outdoor and improve the air

quality of a room. From these explanations, occupants could enhance their activities to achieve better thermal comfort and air quality.

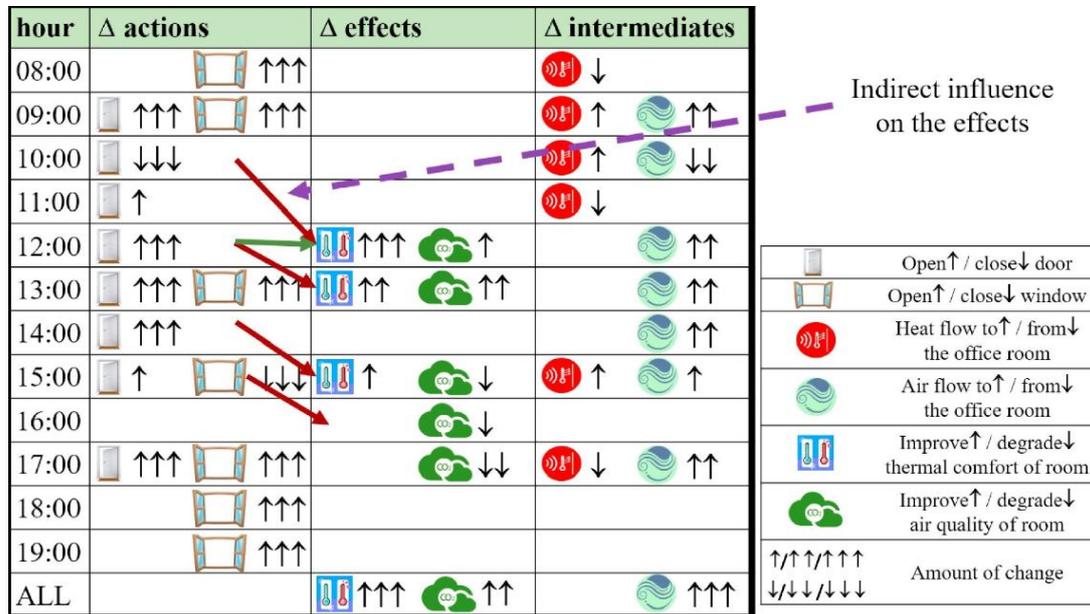


Figure 1.6 – An example of differential explanations for occupants to understand the impacts of their behaviors on doors/windows (Pal et al., 2019)

In another study, He et al. (2021) proposed a framework to estimate the energy-saving potentials of occupant activity improvement. They divided these potentials into four groups: baseline, wasteful, moderate and austere, which represent different levels of their energy-saving consciousness. The framework was implemented in a case study and the results showed that the building energy consumption could be reduced by up to 9.5% with the moderate activity improvement, and up to 21% with the austere activity improvement.

Azar et Al Ansari (2017) stated that "Significant energy savings can be achieved in buildings by altering how occupants use and operate various building systems. The first step to successfully induce such change requires a thorough assessment and understanding of the actual drivers and motivators of existing behaviors". Vassileva et Campillo (2014) also mentioned that providing occupants feedbacks on their energy usages is an effective way to save energy in the domestic sector. This information could lead to behavioral changes and help reduce energy consumption between 15% and 25% in some cases. In another article, Yu et al. (2011) used clustering and classification analysis to identify energy-inefficient/energy-efficient occupant behaviors. The results obtained could help occupants to modify their behavior to reduce about 2092 MJ per capita per year.

1.3 Conclusion

To deal with the issue of climate change, it is essential to reduce greenhouse gases emissions, which are mostly produced through energy consumption. The sector of residential buildings plays a crucial role in total final energy consumption in Europe, so saving energy consumption in houses is essential. Among many factors, occupants with their activities have significant impacts on the energy performance of residential buildings. These activities are diverse depending on households. Investigating them could help reduce the energy wasted and enhance the energy performance in dwellings.

Specifically, modeling of occupants' activities is useful for many services to improve energy efficiency in residential buildings:

- reducing the uncertainty related to occupant behavior and decreasing the discrepancies between simulated and actual energy consumption. These contributions are important for the design phase of the building and the energy performance verification process.
- helping energy monitoring and management systems (HVAC, artificial lighting, *etc.*) to work more efficiently *i.e.* reducing energy waste and improving occupant comfort.
- providing feedbacks and recommendations on occupants' activities. This information could enhance the occupants' knowledge, which could encourage them to change some of their activities to reduce energy consumption and improve their comforts.

Thus, it is definitely interesting to model occupant activities in residential buildings.

Chapter 2

Existing works dealing with occupant activity modelling

Contents

2.1 Modeling occupant activity in residential buildings	21
2.2 Existing approaches for occupant activity modeling	22
2.2.1 Fixed-schedules approach	22
2.2.2 Data-based approach or deterministic approach	24
2.2.3 Probabilistic (statistical) approach	25
2.2.4 Agent-based (multi-agent) approach	32
2.2.5 Knowledge-based approach	36
2.2.6 Summary	40
2.3 Conclusion	42

2.1 Modeling occupant activity in residential buildings

In general, occupant activity modeling aims to determine information related to occupants' behavior in residential buildings. To do it, scientists have been interested in exploring the relationships between occupant activity and other variables such as indoor, outdoor conditions, the house size, occupants' professional status, or the time of day. These relationships could help to deduce the impacts of occupant activities

on energy consumption. A causal model structure for modeling occupant activity is presented in figure 2.1.

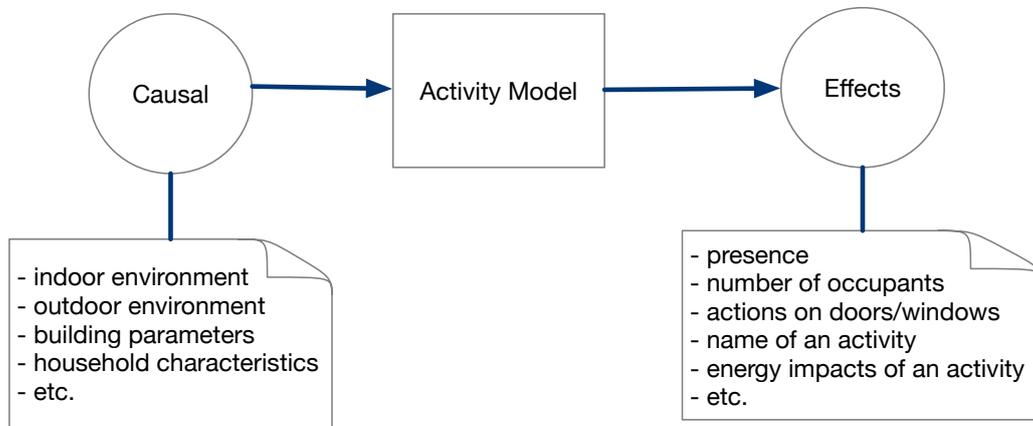


Figure 2.1 – The general model for occupant activity estimation

Depending on expected purposes, required input variables and expected output information are different among existing models. Balvedi et al. (2018) divided occupant activity models into two groups: those concerning occupancy and those about actions on components of objects of a building (windows, doors, blinds, *etc.*). To support building energy simulation (BES) softwares, many scientists were interested in the modeling of more information such as numbers of occupants, occupant activities, and activity profiles, *etc.*

2.2 Existing approaches for occupant activity modeling

In the literature, Gaetani et al. (2016) sorted occupant activity models into four main categories: fixed-schedules, data-based (deterministic), probabilistic (statistic, stochastic), and agent-based (multi-agents). In addition to these categories, many recent studies have proposed knowledge-based approaches with Bayesian networks to model occupant activity in dwellings in recent years (Tijani et al., 2015; Amayri et al., 2017).

2.2.1 Fixed-schedules approach

The fixed schedules approach is the simplest form of occupants' activity model. It focuses on conventional values of occupants' behavior. Occupants' activities are

considered as fixed actions/reactions to the building depending on the convention values. This approach usually studies the occupancy (presence, number of occupants) or occupants' actions on doors/windows. The scenarios/schedules of this information are fixed and then could be set hourly for each thermal zone. Occupants could adjust them to be more comfortable in their environment (Vorger, 2014; Gaetani et al., 2016). For instance, the action of windows opening/closing can be set based on the indoor temperature (if the indoor temperature is greater than 27°C, then the windows will be opened, or there is a probability for it to happen).

This approach is commonly used in BES softwares because of its simplicity. Scenarios/schedules are generally extracted from occupants' calendars, standard conditions or statistical analysis of actual observations. To do it, clustering methods could be applied to determine the typical patterns, which are then used to construct the scenarios of occupant activities. After that, these scenarios are implemented as inputs for BES softwares (Balvedi et al., 2018; Guerra Santin, 2011; Aerts et al., 2014). A basic example of occupancy inputs (number of occupants) for BES is presented in figure 2.2. This scenario of the occupancy is generally set based on the calendar of occupants.

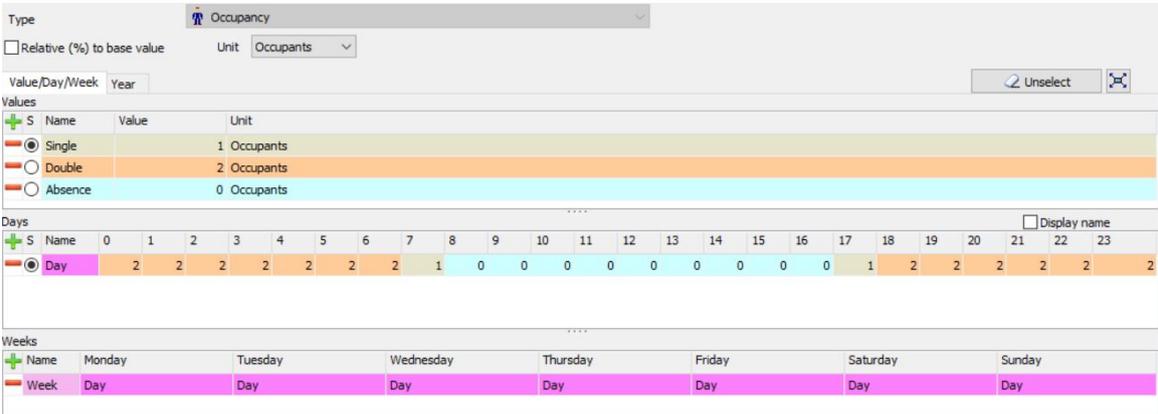


Figure 2.2 – An example of occupancy schedules (number of occupants) input in the software Pleiades

While simple, predefined rules and scenarios have some limitations. Firstly, scenarios and schedules are defined for a long period and rarely change over time. Hence, they lead to fully repeatable and predictable activities and the variations of occupant activity are not considered (Borgeson et Brager, 2008; Fabi et al., 2013). These scenarios and schedules are applied to all buildings and all types of occupants. It is not realistic because configurations and characteristics of each building and occupant

significantly influence activities. Gram-Hanssen (2010) stated that personal preferences or habits play an essential role in operating activities. Thus, it is not efficient for operating occupant activity based on predefined rules and scenarios.

2.2.2 Data-based approach or deterministic approach

The second approach is data-based (deterministic) approach. In this approach, an activity is defined as a meaningful label representing what occupants do in an interval of time. The objective is to determine these activities by exploring consistent patterns and systematic relationships between them and environment variables (Jia et al., 2017). This approach generally works as a supervised learning model, which classifies labels of activities depending on other information. Data mining and machine learning are popular techniques in that context. To validate these models, a dataset is divided into two parts: training part and testing part. The models are built with the training dataset and then validate estimated labels with the testing dataset.

Recently, many researchers have used data-based models to estimate both occupants' presence, their actions on components in the building (doors, windows, appliances, *etc.*) and their daily activities (cooking, eating, *etc.*). For example, Amayri (2017) applied Decision Tree (DT) and Support Vector Machine (SVM) to estimate levels of occupancy in an office, which is based on measurements captured by power sensors for three years. The estimated occupancy level, the power consumption for each laptop and the electricity consumption for the visitor's desk were then applied DT to determine activities (presentation, working computer, absence) in the office. In another study, Ren et al. (2015) combined data mining technique, clustering technique, and DT to gain knowledge about the behavior of occupants adjusting their thermostat settings and heating system operations in houses. Yu et al. (2011) also analyzed the information of end-use loads (electricity, gas, kerosene), indoor temperature, and occupants' behaviors (actions on electrical appliances, lighting, heating water, *etc.*) in a residential building, then applied the K-Means algorithm for clustering and the C4.5 algorithm for classifying. The goal was to identify energy-efficient and energy-inefficient activities (using appliances, heating water, lighting, *etc.*) in order to motivate occupants to improve their practices at home. From datasets containing power and environmental variables measurements captured by sensors and volunteers survey, Alhamoud et al.

(2015) applied Random Forest classifier to recognize some occupant activities (eating, watching, etc.) and to extract activity patterns that were then used for energy consumption estimation and energy-saving recommendation systems. In another study, Li et Dong (2017) installed sensors to capture indoor environmental variables in four residential buildings. Then, they applied Artificial Neural Network and Support Vector Regression to predict the presence of occupants in both room-level and house-level.

The data-based approach is the most popular method to recognize occupant activity in buildings in real-time energy monitoring. The advantage of this approach is the ease of implementation due to the development of machine learning and data mining tools. Moreover, these models of this approach are flexible to estimate many kinds of activities in buildings. However, there are some limitations in this approach:

- This approach aims to estimate the deterministic labels. These labels can not represent the diversity of occupants' activities. Jia et al. (2017) also stated that "*occupant behavior is complex, which is not in a mode of **if-then**. Driving factor is influential but not determinant, and models solely based on driving factors that do not have sufficient prediction performance*". Therefore, these models are limited to integrate into BES.
- Data-based models are generally black-box models with complex mathematical functions. As a result, it isn't easy to understand the causal-effect relationship between the activities and their related variables.

2.2.3 Probabilistic (statistical) approach

To deal with the diversity of activities, many researchers have proposed probabilistic models taking into account the randomness in occupant activities modeling in buildings. Nicol et Humphreys (2004) suggested that "*there is no precise temperature at which everyone will open a window, but the higher the temperature, the higher probability of opening the window*". In these models, occupants' activities are represented by the probability distribution based on predetermined driving factors in each interval of time. For example, to model an a based on a set of variables x , the probabilistic model aims to determine the probability distributions $p(a|x)$ that the activity happens in each interval of time.

This approach is generally the most common way of modeling occupant activity in the research context. In BES, probabilistic models are usually used to analyze the energy impacts of occupants' activities on energy consumption and on air quality conditions in buildings. **Statistical analysis** and **stochastic models** are two popular kinds of models in this approach. They are commonly used to model representative activities of occupants from statistical data.

Statistical analysis

In statistical analysis, the numerical relationship between an occupant activity and available information (indoor environment, electricity use, time series, *etc.*) is estimated. Then, the probability of occurrence of the studied activity is deduced from related parameter inputs using regression models. In this approach, logistic regression (logit) is a popular method to estimate the probability p_a that the activity a happened is estimated as in equation 2.1.

$$\begin{aligned} \text{logit}(p_a) &= \log\left(\frac{p_a}{1-p_a}\right) = w_0 + \sum_{i=1}^n w_i \times x_i \\ \Leftrightarrow p_a &= \frac{\exp(w_0 + \sum_{i=1}^n w_i \times x_i)}{1 + \exp(w_0 + \sum_{i=1}^n w_i \times x_i)} \end{aligned} \quad (2.1)$$

with w_i is a coefficient corresponding to x_i - variable affecting the probability that activity a happens.

The work of Haldi et Robinson (2008) is an early study with statistical analysis. Authors collected the data of activities, thermal comfort, adaptive actions and environmental conditions in eight Swiss offices. Logistic regression was then applied to predict occupants' actions on windows, doors, fans, and blinds to achieve their predefined comforts. The study also considered personal activity as clothing or drinking in the model. In another article, Andersen et al. (2013) studied actions on windows/doors in Danish dwellings. The dataset including environmental parameters, windows/door states, moments of the day, seasons, and presences, are measured by sensors in 15 dwellings. Logistic regression was applied to infer the probability of the actions (opening or closing) on the windows. Vorger (2014) also applied logistic regression to estimate the transition probabilities among the actions on windows/doors in dwellings, based on the indoor temperature, the time of day and the day of the week in survey data. Then, the model was used to simulate the actions on windows/doors in

a BES software.

Statistic analysis is a popular and flexible approach in modeling actions on doors/windows (opening/closing) or the occupancy (occupied/unoccupied status) in buildings. This approach could use both survey data and sensor data, which makes it possible to be applied in both population scale and in particular buildings. However, there are still some limitations in statistic analysis as follows:

- Logistic regression model was widely used in this approach. The mathematical formula in this model is quite difficult to explain the relationships between occupants' behaviors and related variables.
- This approach is limited to estimate more complex activities in houses such as cooking, studying or sleeping due to the complex relationships between these activities and other variables in buildings.

Stochastic models

Stochastic models aim to determine representative activities in households with similar given characteristics (building type, owner, etc.). They assumed that in different buildings, occupants with different attributes (age, gender, etc.) could have different habits, and they lead to different activities. Thus, building characteristics and individual attributes should be taken into account. In these models, an activity is generally represented by a starting probability distribution and a duration distribution or a transition probability distribution.

Multinomial logit model, a variation of logit model (equation 2.1), is a popular method to identify the starting probabilities of activities in stochastic models. Specifically, with M activities ($a_1, \dots, a_{M \in N^*}$) and n variables ($x_1, \dots, x_{n \in N^*}$), the logit function for each activity is described in equation 2.2.

$$\begin{aligned}
 L_1 = \text{logit}(p_{a_1}) &= w_{1,0} + \sum_{i=1}^N w_{1,i} \times x_i \\
 &\dots \\
 L_M = \text{logit}(p_{a_M}) &= w_{M,0} + \sum_{i=1}^N w_{M,i} \times x_i
 \end{aligned}
 \tag{2.2}$$

With $w_{i,j}$ is a coefficient corresponding to an activity a_i and a variable x_j . A starting

probability $p(a_i)$ of an activity a_i has the form as in equation 2.3.

$$prob(a_i) = \frac{L_i}{\sum_{j=1}^M L_j} \quad (2.3)$$

Markov Chain is a popular algorithm to determine transition probabilities among activities. A transition probability T_{ij} is a probability that an activity j starts shortly after an activity i stops. In Markov Chain, it is assumed that the probability of the current state depends only on the latest state. In the case of N activities, the transition probability T_{ij} has the form as in equation 2.4.

$$T_{ij}(t) = P(a_{t+1} = j | a_t = i), i, j \in [1, 2, \dots, N] \quad (2.4)$$

with $P(a_t = i)$ is the probability that the activity i happens at the time t .

Markov Chain and multinomial logit model are usually combined together to simulate occupant activities in BES.

An early study in this approach was proposed by Page et al. (2008). Authors built a generalized stochastic model to simulate the occupancy in buildings. In this study, occupancy data was collected by motion sensors in twenty zones of a building. Based on this data, authors estimated probabilities of presence and transition probabilities (absent \rightarrow present or present \rightarrow absent), which were related to the time of the day. Then, inhomogeneous Markov Chain and Hidden Markov Model were applied to simulate the occupancy in each building zone. In another research, Haldi et Robinson (2009) applied a discrete Markov process to study the actions on windows in offices. The study measured indoor temperature and outdoor conditions (temperature, humidity, wind speed, etc.) in an office for seven years. Then, a continuous-time model was inferred that could be efficiently used for a fast calculation of windows opening and closing durations.

Recently, stochastic models have been applied to study occupants' activities in residential buildings. Researchers paid attention to occupant daily activities (cooking, watching, cleaning, etc.). The objective of these models is to determine representative activities so that they usually use time-use surveys (TUS) instead of sensors data. TUS are national surveys conducted to identify how much time humans spend on various activities (paid work, household chores, family care, personal care, etc.). In these surveys, both information of activities, household characteristics, and individuals' attributes are recorded.

For example, Yamaguchi et al. (2011) proposed a stochastic occupant activity model based on survey data to simulate a high-temporal resolution electricity demand profile of a household. In this paper, 27 activities were considered to build profiles of activities in dwellings. Moreover, 24 variables related to types of day and family attributes were used to classify the data. With each activity, three parameters were focused: an Average Ongoing Minutes of the activity that occurs in a day (AOM), a standard deviation of the AOM, a Percentage of respondents who adopt the behavior (PB) at a specific time of a day. Similarly, Muratori et al. (2013) developed a stochastic model to simulate usages of electricity. An inhomogeneous Markov chain with nine activities was calibrated with a 10-minutes resolution for five types of individuals: employed woman, employed man, unemployed woman, unemployed woman, and child. The type of the day was also separated into week-day and weekend. The model of occupant behavior had $2 \times 5 \times 144$ profiles of individuals, each of them had a matrix with dimension 9×9 of activities.

In another study, Wilke et al. (2013) also proposed a stochastic model based on French TUS to simulate occupants' time-dependent activities in residential buildings. In this approach, an activity was represented by three types of time-dependent quantities: a probability to be at home, a conditional probability to start an activity whilst being at home and a probability distribution function for the duration of the activity. The model of activities was carried out once the occupancy model had been established. When an occupant is present at home, activity is determined to start through their starting probabilities. To estimate these probabilities, authors used a multinomial logit model with 41 characteristics of the individual (carer, day, healthy, gender, etc.). Then, once an activity started, its duration was estimated by using the probability duration function, which was based on the Weibull probability distribution. During the activity, if occupants leave the house (unoccupied), the activity would be interrupted immediately. The outline of this model is presented in figure 2.3. Vorger (2014) implemented this model to simulate 21 occupant activities in *Pleiades*, a widely BES software in France (Peuportier et Blanc, 1990).

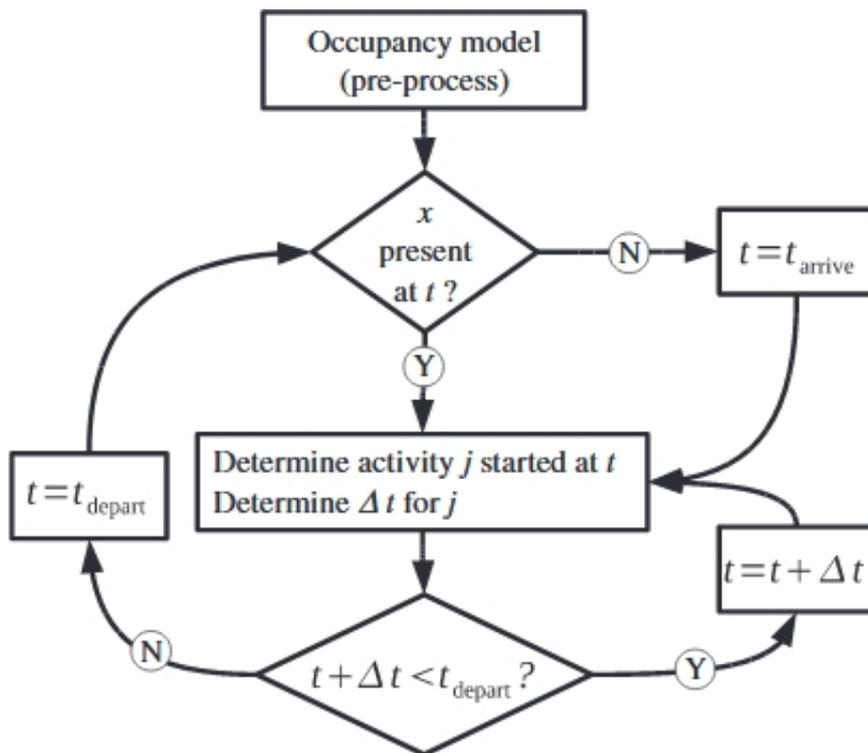


Figure 2.3 – Stochastic model of occupant activity of Wilke et al. (2013)

However, the model did not consider different types of activities and possible interactions between them. To address these issues, Aerts et al. (2014) derived realistic occupancy and activity data in residential buildings from the Belgian Time-Use Survey and Household Budget Survey. The author proposed a stochastic activity model, which is presented in figure 2.4.

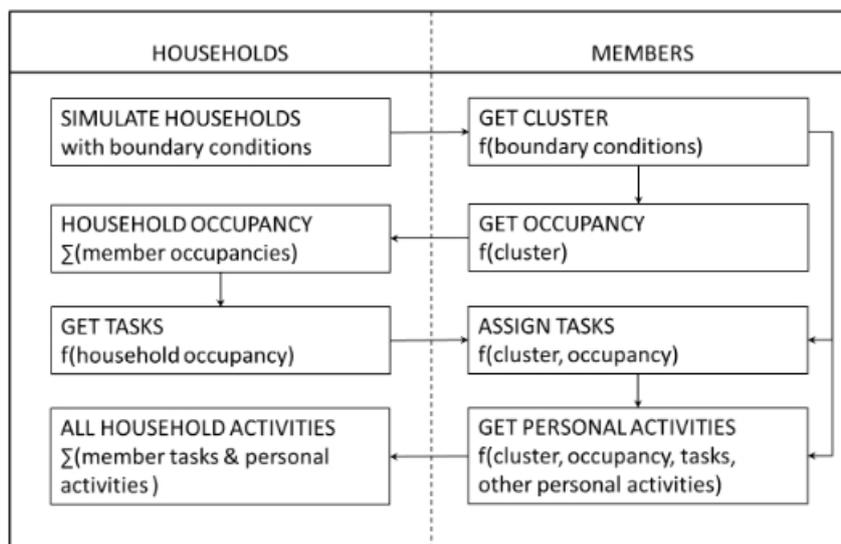


Figure 2.4 – Probabilistic model of occupant behavior of Aerts et al. (2014)

In this model, three occupancy patterns (at home or awake, sleeping, absent) were identified based on six socio-economic features (age, type of day, type of dwelling, employment, gender, household size). Then, among 272 activities, a stochastic activity model was developed to deal with nine activities in dwellings. Aerts et al. (2014) defined an activity as a label describing what occupants do during a period of time with three parameters: a start distribution, a duration probability distribution, and an assignment probability distribution. Activities were divided into two groups: tasks and personal activities (figure 2.5). Tasks were usually performed by only one of the household members. Personal activities can be performed independently from other household members. Tasks were first modeled at the household level, and then they were assigned to the available household member. After the tasks were assigned, personal activities were supplemented. When an activity started, a duration was assigned based on the duration probability distribution. The assignment probability distribution was used to assign this activity to a household member. Once the activities were simulated, they were used linked to the use of appliances for simulating the electricity demand in the dwelling.

	Activities	Type
1	preparing food	task
2	vacuum cleaning	task
3	ironing	task
4	doing the dishes	task
5	doing laundry	task
6	using a computer	personal activity
7	watching television	personal activity
8	listening to music	personal activity
9	taking a bath or shower	personal activity

Figure 2.5 – Two groups of activity in dwellings according to Aerts et al. (2014)

Stochastic models with TUS data are popular to deal with more human activities in houses, which are limited in statistic analysis. The development and the availability of time-use survey data help these models to be more convenient and more flexible to apply in population scale. Moreover, these models are useful in BES in the building design phase, in which the dataset of the household cannot be available. However, there are still some limitations as follows:

- Aerts et al. (2014) and Vorger (2014) showed that even with dwellings having similar building characteristics and individuals attributes, there are still significant discrepancies of the heating needs and the thermal comfort, which were mainly linked to the diversity and the average of occupant activities in the survey data. Therefore, stochastic models with representative activities are not specific in terms of particular households with their contexts.
- Because the representative activities of stochastic models are average, they could not be specific to actual households. In this case, it is difficult to verify these results.

2.2.4 Agent-based (multi-agent) approach

In efforts to reduce energy consumption in residential buildings, many researchers have been interested in investigating interactions between occupants and building and among occupants in specific buildings. To address these issues, the agent-based approach was proposed to model complex systems composed of interacting agents in an environment. Agents can be occupants or building components. They have behaviors, often described by simple rules and interactions with other agents, which in turn influence their behaviors. A special feature of these agents is the *self-organization*, which was defined by Di Marzo Serugendo et al. (2005) as "*the mechanism or the process enabling a system to change its organisation without explicit external command during its execution time*". Macal et North (2010) mentioned three elements of an agent-based model (see figure 2.6): *a set of agents with their attributes and behaviors, the relationships and methods of interaction between agents and an environment that agents interact with each other*. Each agent plays an important role in the whole system. The state of each agent influences other agents *via* relationships and interactions. In contrast, the state of the whole system also affects the state of each agent. According to Lee et Malkawi (2014), "*the purpose of the agent-based modeling is to mimic a real-world occupant: an autonomous agent that interacts with both its environment and other agents, and makes behavior decisions based on the level of its thermal comfort*".

There have been many studies applying the agent-based approach to model occupants' activities. In 1997, Sierhuis et Clancey (1997) proposed BRAHMS (Business Re-design Agent-based Holistic Modeling System), an activity-based multi-agent

modeling environment, allows to model knowledge in situated actions and learning in human activities. Each activity is considered as an agent with pre-conditions, actions, and consequences. BRAHMS focuses on collaboration, communication, and the situation in daily activities.

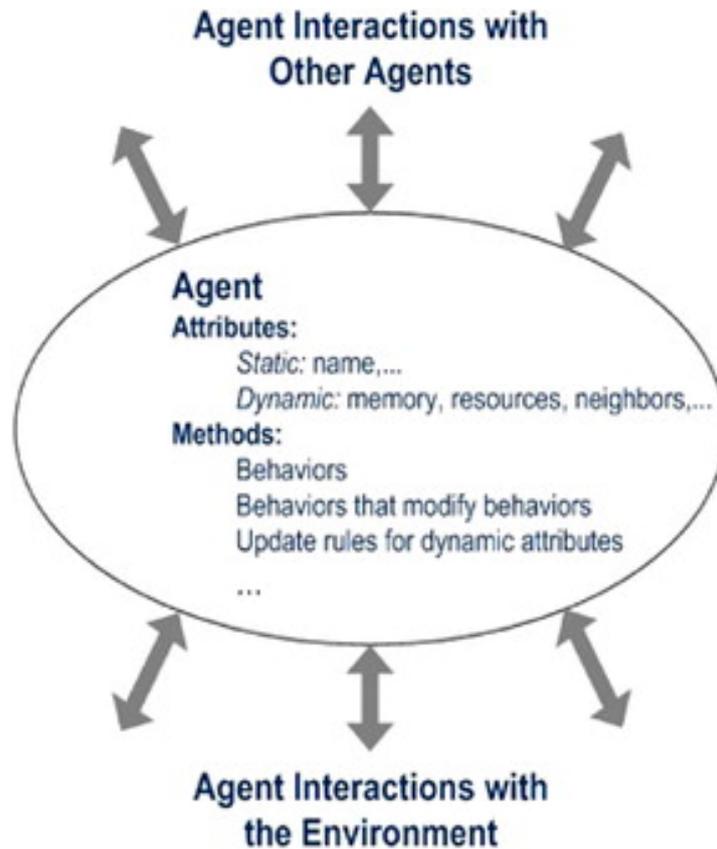


Figure 2.6 – Structure of an agent-based model according to Macal et North (2010)

In addition, to represent the agent’s activity, Ha et al. (2006) proposed a model of activity which includes a context 5W1H (What, When, Where, Why, Who and How), and mapped it in a home context (object, time, space, occupant and environment) (figure 2.7).

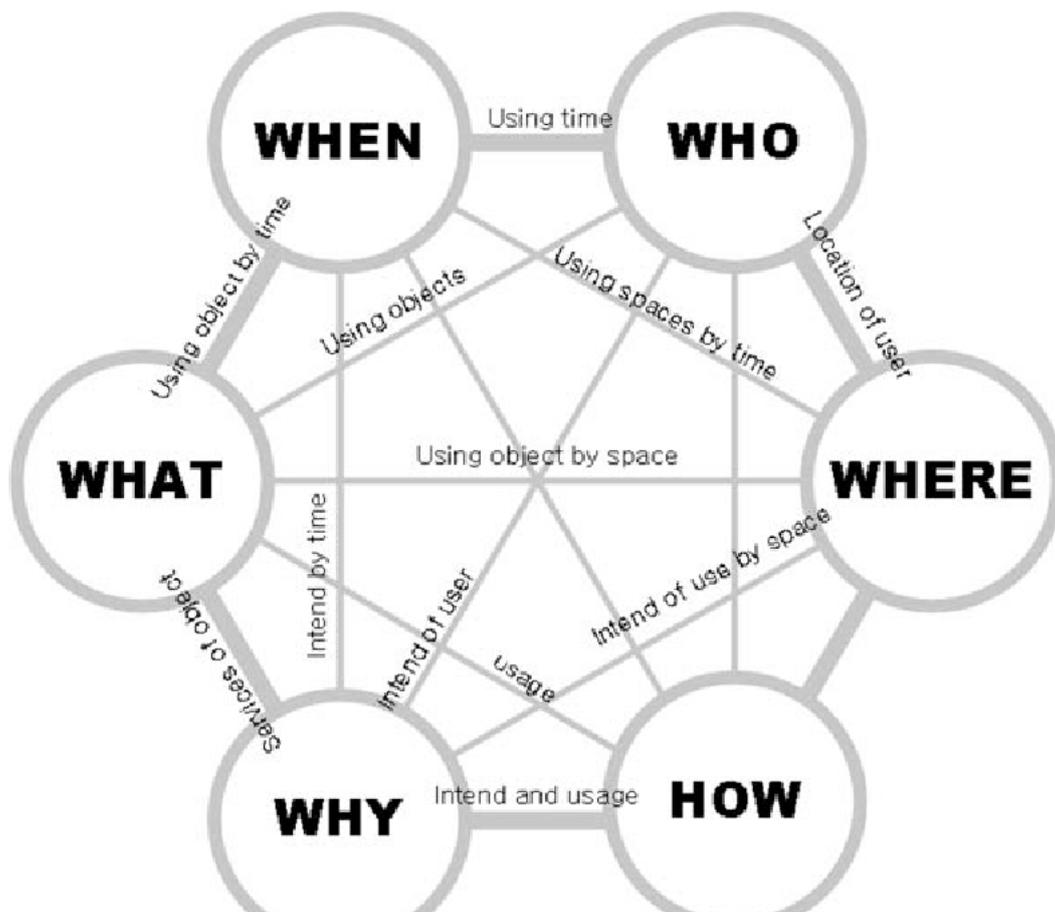


Figure 2.7 – The context factors of an agent’s activity from Ha et al. (2006)

Kashif et al. (2013) proposed an agent-based model with the BRAHMS language and the 5W1H context to study occupant activity patterns in energy management strategies. In this model, authors defined activity as "*the actions or reactions of occupants, in relation to its environment, to satisfy occupants’ needs*". Specifically, the model is connected to a dynamic thermal simulation software that interacts with agents and adapts the environmental conditions in each time slot. Agents do not react immediately in response to environmental conditions. They have their own perceptions, which are influenced by their characteristics (psychological, physiological) and their knowledge (experience, comfort). Agents use these perceptions to deliberate and make decisions on their activities. For example, an agent feels discomfort due to a high indoor temperature and wishes to open the window because from experience, it knows that opening the window could help to be more comfortable. Then, it asks other agents to open the window, if they accept, it is ready to open the window. The agent deliberates its comfort with the opening of windows based on other conditions (outdoor environment), if the comfort is

acceptable, the agent decides to open the window. Consequently, after this action, the environmental conditions are updated, and due to the comparison between perceived comfort and expected comfort, the knowledge of the agent is modified. This model is attractive because it seems to reproduce the reality of occupants' behavior, taking into account the phenomena of perception, memory, and deliberation.

In another study, Guivarch et al. (2012) proposed Amadeus, an Adaptive Multi-Agent System that could assist the occupants in an ambient system. It assumes that if Amadeus performs similar to what occupants did in similar situations, these activities can enhance occupants' satisfaction. Amadeus can learn from its observation, how and in what situations occupants perform their activities. Thus, it can adapt its behavior to the occupants' activities even if they change their preferences.

Lee et Malkawi (2014) also proposed an agent-based model for occupants' activity in commercial buildings. The activity model concerned three essential processes: defining activities, identifying activity triggers and measuring/quantifying activities. This study considered activities that affect the occupant thermal comfort such as window use, blind use, or fan use. To do an activity, an agent uses a process of three steps: perceive, think and act. Firstly, an agent observes and perceives environmental information and the thermal conditions of the space. Then, based on the beliefs (user-defined, random distributed), the expected comfort is estimated for each activity. Then, the agent thinks and ranks the activities based on their comforts. After that, the activity is decided to be executed by the agent to maintain or improve its comfort level.

Recently, Albouys et al. (2019) proposed SMACH, a Multi-agent Simulation of Human Activity in the Household platform, to simulate human activity and its electricity consumption at the household scale. There are two main elements in this system: a multi-agent system simulating human activity and its impacts on the electricity consumption of electrical appliances; a multi-zone thermal model simulating indoor temperature and the energy consumption for room heating. These elements are combined to generate load curves linked to human activity, which is useful in investigating the electrical energy consumption in existing households.

Multi-agent models are advanced in modeling activities at both the individual level and household level. They outperform if-then rules of occupant behavior because it considers both the interactions among agents and the relationships between agents and

environment (Jia et al., 2017). It is efficient for energy monitoring systems such as HVAC with the interactions among occupant activities in the same zone. For example, the actions of the fan affect the actions of the windows in the same room. In addition, the multi-agents approach can be combined with other approaches for modeling the activities of agents. However, some limitations still exist in this approach as follows:

- The goal of this approach is to determine activities that help occupants to achieve expected comfort. Thus, these activities represent expected activities rather than actual activities. In addition, agents usually choose "*the optimal activities to achieve their comforts*" so that the diversity of occupants activities could be not considered (Lee et Malkawi, 2014; Vorger, 2014).
- It is difficult to define the conditions that the agent feels comfortable, the function that links the environment to the psychological state of the agent is not explained because of the diversity, the complexity of occupants' characteristics, sociology, psychology, personality, and their habits (Kashif et al., 2011). These definitions are usually defined by the belief of experts, but they are not precise for all occupants in the building. For example, the comfortable temperature for a child is different from the comfortable temperature of an old man. Therefore, it is not precise to define which temperature is *comfortable* for occupants, which conditions that occupants need to react to improve their comforts.
- Due to the interactions between agents, the increase of the number of agents could significantly increase the complexity of the model. Each agent does not act independently, this approach needs to define the interactions and the impacts among these agents. This task is very difficult when there are many agents. For example, it is necessary to define at least 2^n interactions with n individuals in a house. Moreover, each interaction can be related to various variables such as time, location, environment, personal characteristics, *etc.*

2.2.5 Knowledge-based approach

Many other research works related to occupants' behavior are outside the scope of the mentioned categories. In fact, some experts have experience and their knowledge can be useful in modeling occupant activity. The expert knowledge could help to

define the starting probability of the variables related to the activities. Therefore, knowledge-based models were proposed to use expert knowledge in the occupant activity model. This kind of model represents human activity with probabilistic cause-effect relations, which are defined by expert knowledge or are deduced from observations (Tijani et al., 2015).

Bayesian Network (BN), a probabilistic graphical model, is the popular method in this approach. BN comprises two main elements: *a directed acyclic graph (DAG)*, which represents the cause-effect relationship between variables *via* nodes, edges (X to Y - cause node to effect node), and *a set of parameters in all conditional probability distributions (CPDs)*, which represents the conditional probability of relationships linking causes to effects between variables in the above DAG. These conditional probability distributions could be deduced from observations or expert knowledge. The causal graph DAG also has to be determined and refined. To construct this causal graph, the first approach consists of learning the causality directly from data, the second approach is based on expert knowledge about occupant cognition in different contexts and activities.

There are two categories of Bayesian networks: static and dynamic BN. Static BNs only concern a single slice of time and are not appropriate for analyzing a system that changes over time. Dynamic BNs describe how variables influence each other over time based on a model derived from historical data. The main elements in a BN and an example of dynamic BN are shown in figure 2.8. In addition, figure 2.9 shows an example of a static Bayesian Network, which was proposed by (Amayri et al., 2016). In this study, the author used the level (L-low, M-medium, H-high) of the power consumption and of the microphone (acoustic pressure) to estimate the level of occupancy in an office.

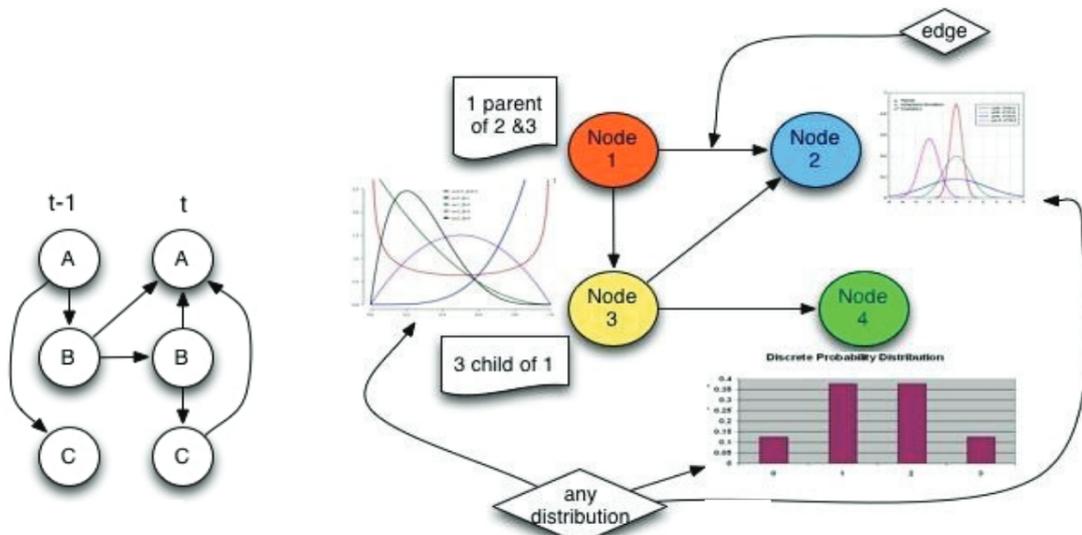


Figure 2.8 – (a) Dynamic BN (b) Elements of Bayesian Network (Tijani et al., 2015)

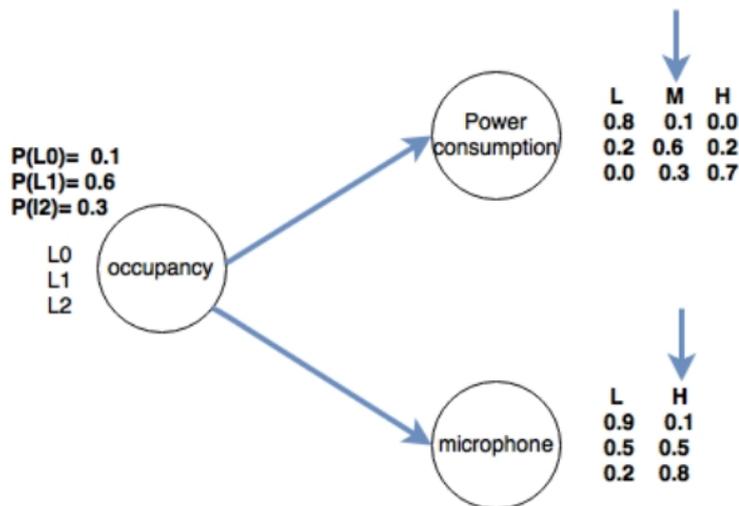


Figure 2.9 – An example of static Bayesian Network (Amayri et al., 2016)

Hawarah et al. (2010) proposed a Bayesian Network for predicting the use of ovens in houses. In this model, the structure of the model is defined by experts, which includes three causal nodes (hour, type of day, month), which influence the starting probability of cooking-oven activity. The duration and the energy consumption are two effect nodes that are caused by cooking-oven activity and the above causal nodes. The conditional probability distribution at each node is calculated using databases, which consists of the energy consumed by the services in 27 houses in France.

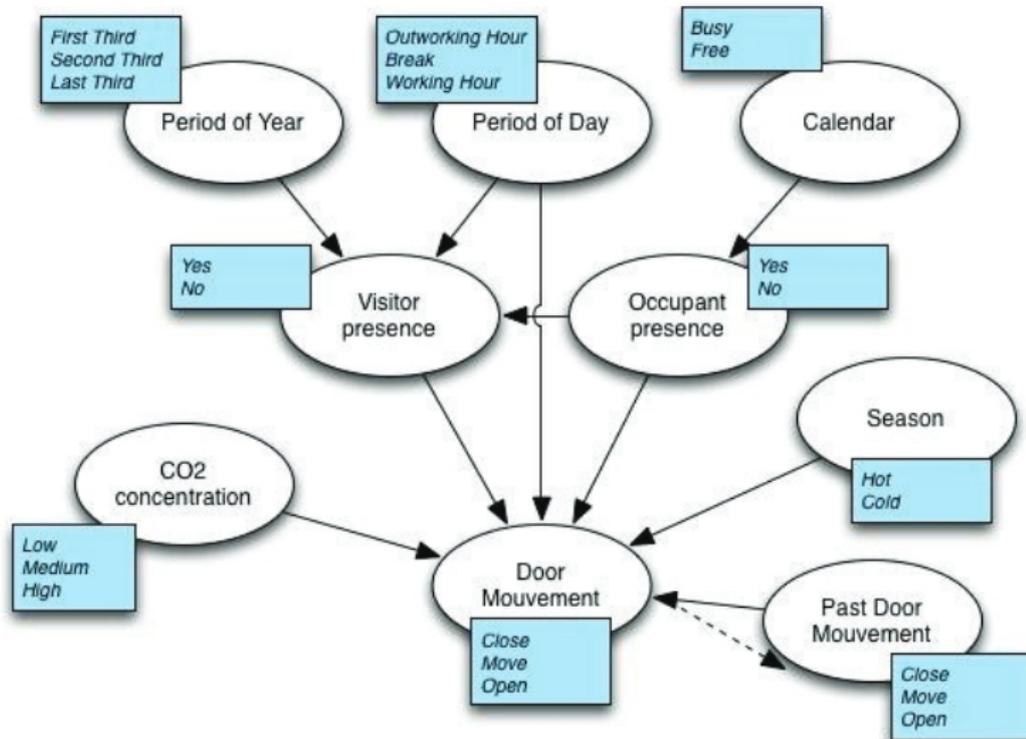


Figure 2.10 – Dynamic Bayesian Network for the movements of the door in office (Tijani et al., 2015)

Tijani et al. (2015) also proposed a dynamic Bayesian Network for predicting the actions of the door (open, close, move) in an office. In this model, the authors considered periods of the year (first, second, third), periods of the day (outworking, break, working), the calendar (busy, free), the visitor presences (yes, no), the occupant presences (yes, no), the seasons (hot, cold), the levels of CO₂ concentration (low, medium, high) and the past door movement as variables influencing the actions of the door. The structure of this Bayesian Network (figure 2.10) was designed by expert knowledge and the conditional probability distribution is deduced from the data.

The knowledge-based approach with the Bayesian Network is a flexible and friendly approach for humans to understand and verify the results of the model. It is useful for different objectives in activity modeling. It can be used to estimate both deterministic states as data-based models and probability distribution as probabilistic models. In addition, two main elements in this approach make it adaptive to changes in the buildings: the conditional probabilities distribution is adaptive to learn new data and the network is adaptive to modify the structures of data, which are limited in data-based models. In addition, this approach allows experts to use their knowledge and experience

to contribute to the models: define the network, assume the conditional probabilities between variables, *etc.* These expert contributions are very useful in terms of lacking information, or to adjust the model to be more specific to case studies. However, there are some limitations in this approach as follows:

- To design the network structure and estimate the conditional probability distributions, this approach requires much data to ensure the consistency and the accuracy of both the network and parameters. An alternative approach is to use expert knowledge, it is more efficient and realistic but it also requires the knowledge of experts in the concerned domain.
- It is quite similar to the multi-agents approach that the network in the model could be complex when there are many variables. However, it can be helpful to use expert knowledge to simplify it, and the relationships between variables in this approach require fewer efforts to be defined than the interactions between agents.
- Existing models in this approach mostly focused on occupancy or occupants' actions. There are limited studies in modeling occupants' activities in residential buildings and their energy impacts.

2.2.6 Summary

The summary of the above approaches is shown in table 2.1. In general, fixed schedules and data-based approach with deterministic models can not consider the diversity of activities, which is important in energy simulation. To address this issue, statistical approach is used popular because it is easy to implement and it is adaptable to households characteristics. This approach aims at representative activities so that it is not specific to particular houses. Only multi-agents approach and knowledge-based approach are easy for human to explain in modeling contextualized activities in specific houses. However, multi-agents models are very complex to implement in households with many members while existing knowledge-based models only focused on the occupancy or occupants' interactions on building components. Thus, to improve the energy simulation in specific households, it is interesting to propose an approach in modeling contextualized activities that can deal with above limitations.

Table 2.1 – Summary of existing approaches in occupant activity model

	Fixed schedules	Data-based	Statistical	Multi-agents	Knowledge-based
How an activity is represented ?	fixed conventional values, fixed schedules	deterministic label describing what occupants do	probability distributions for an activity to happen	actions or reactions to satisfy occupants' needs in a specific context	causal-effect model between activity and related variables
Kind of activity	occupancy, actions on building elements (doors, windows, thermostat, etc.)	occupancy, actions on building components, daily activities	occupancy, actions on building components, daily activities	occupancy, actions on building components, daily activities	occupancy, actions on building components
How time is managed?	discrete	discrete	continuous or discrete	continuous or discrete	discrete
Can be used for simulation?	yes	no	yes	yes	yes
Is generative or deterministic?	deterministic	deterministic	generative	generative/deterministic	generative
Is verifiable?	no	yes	no	yes	yes
Is easy to transform to another context?	no	yes	yes	no	yes
Is easy to implement?	easy to implement	easy to implement	easy to implement	very complex to define agents and its context, especially when there are many agents	complex if there are many nodes but experts knowledge can be used to define
How people preferences are modelled?	calendar, rules	learn from data	learn from data	defined comforts, actions, interactions of agents	causal-effect relationship
Is easy to explain?	easy	difficult	difficult	easy	easy
Is specific for particular households ?	no	yes	no	yes	yes

2.3 Conclusion

In the building sector, to take account of impacts of occupants on many useful applications such as energy simulation, energy management, monitoring systems and energy feedback, many different approaches have been proposed to estimate occupant activity. Based on the objective and the complexity of these approaches, we divided them into five main groups: fixed-schedules, data-based, statistical-based, knowledge-based, and agent-based.

The objective of this chapter was to present an overview of the advantages and the limitations of the existing approaches to address the contribution of this thesis. Among many applications, this research aims to improve the estimation of occupant activity in energy simulation in residential buildings. There is a significant discrepancy between estimated occupant activity and actual occupant activity. Most of the existing activity models in energy simulation are constructed based on survey data. However, the hybrid of survey data makes these models difficult to be verified and not proper for a specific house. It leads to the necessity of an approach in estimating contextualized occupant activity in residential buildings.

Dealing with the above issue, our research aims to consider the context in the activity models. To do it, our works use the knowledge-based approach to:

- estimate occupant activity in houses with specific contexts, which depends on both information of contexts and sensors data.
- estimate the energy impacts of contextualized occupant activity in the house in order to explore the consequences of activities when they are performed.

The methodology and the implementation of these models will be presented in next chapters (chapter 5, 6).

Chapter 3

Problematic: Contextualized activities models in residential buildings

Contents

3.1 Introduction	44
3.2 General definitions and concepts	44
3.2.1 Time discretization	44
3.2.2 The home context	44
3.2.3 Occupant activity	46
3.2.4 Consequences of an activity	48
3.3 Contextualized activity models: opportunities and challenges	49
3.3.1 Difference between representative and contextualized activities	49
3.3.2 The reminder of limitations in existing occupant activity models	50
3.3.3 Contributions of contextualized occupant activities modeling	51
3.4 Problematic: contextualized activities modeling in dwellings	52
3.4.1 Objectives and research questions	53
3.4.1.1 Concerned activities	53
3.4.1.2 Research questions	53
3.4.2 Proposed methodology to model contextualized activities	54
3.5 Conclusion	57

3.1 Introduction

In chapter 2, we analyzed and discussed the existing approaches dealing with occupant activity modeling, their capabilities and their limitations. To better match existing households, we are going to propose a contextualized activity model. This chapter includes three main sections with different objectives. Section 3.2 presents the general definitions and the principles of modeling contextualized occupants' activities. Section 3.3 memorizes the capabilities and the limitations of existing occupant activity models and presents the benefits of contextualized activity model. Finally, section 3.4 states the problem with the objectives, research questions and a proposed methodology for the design of contextualized occupants' activity model.

3.2 General definitions and concepts

3.2.1 Time discretization

The time is continuous with a finite number of seconds. Occupants do not change their activities every second. Considering these activities at every second could be impractical. In addition, occupants better manage their activities with qualitative and discrete times. Thus, we define a timestep as a discrete interval of time in which contextualized activities are investigated.

Timesteps with small resolution (1-minute, 5-minutes, *etc.*) yield a better precision about the activities, but it leads to computation time and extra labelling efforts from occupants. Larger resolution (1-hour, 2-hours, *etc.*) reduces the labelling effort but the accuracy decreases (Albatayneh et al., 2015; Haigh et al., 2011).

3.2.2 The home context

There are significant discrepancies in human activities between houses. Context is an important factor affecting human behavior and explaining this gap (Kashif et al., 2011). Henricksen et Indulska (2005) defined the context of an activity as "*the set of circumstances surrounding it that are potentially relevant to its completion*". Similarly, Xuan Hoa Binh et al. (2010) defined context as "*any information that can be used to*

characterize the situation of an occupant at home". In particular, they identified five main elements of the home context: the occupants, the time, the location, the environment and the objects. In a household, individuals could do activities in different locations. For example, a family member studies in the bedroom, whereas another member does it in the living room and sometimes in the bedroom. Thus, to specify this information, this work supplements the above definition with an additional term *located activity*, which describes the location where an activity is assumed to be located. The considered context elements are presented in table 3.1.

Table 3.1 – Proposed context elements at home completed from Kashif (2014)

Context element	Definition	Example
Occupant	It is a person who is at home. It can be a family member or a visitor.	Family members (parents, children, <i>etc.</i>), visitors (friends, neighbors, <i>etc.</i>)
Time	It is discretized timestep and composed of time slot during one. It covers day of the week, day-month-year and hour:minute:second.	Sunday 03-10-2021 18:00:00 to Sunday 03-10-2021 18:30:00
Location	It can be a room or an outdoor location	Living room, kitchen, bathroom, bedroom, garden
Environment	indoor environment	Air temperature, air relative humidity, CO ₂ concentration, lighting
Objects	They are domestic electric appliances and others (systems, furniture, <i>etc.</i>)	Microwave, toaster, coffee machine, HVAC, windows, doors
Located activity	It is the relation between an activity and a location where it happens	Reading in the bedroom and reading in the living room are different activities

Each context element can have some characteristics. For example, an indoor environment includes an air temperature, a relative humidity, a luminosity and a CO₂

concentration. In the same manner, an electric appliance has a name, a location, its energy consumption, *etc.*

The context elements can affect occupant activities and reciprocally, occupant activities could influence some elements of the context (Hong et al., 2017). In order to model occupant activity in dwellings, the context in which occupant activities happened must be recorded. A method to collect the information about occupant activities and the context is presented in chapter 4.

3.2.3 Occupant activity

Milliken et Honeycutt (2004) terms an activity as ways occupants strive to satisfy their physical needs, to survive or to be comfortable. Similarly, (Kashif et al., 2013; Xuan Hoa Binh et al., 2010) defined an activity as "*the actions or reactions, usually in relation to its environment, which purpose to satisfy occupant needs according to the current context*". This definition states that an activity could include some actions and interactions with objects. However, the definition requires to define the needs of inhabitants, which are complex to identify for households with many activities.

In a statistical approach, Wilke et al. (2013) defines an activity with two parameters: starting probabilities and probability distributions for durations. Aerts et al. (2014) divided activities into two groups: tasks and personal activities. Besides the starting probabilities and the distribution for possible duration, an occupant assignment probability and an activity compatibility are considered to assign activities to the most suitable individual. In addition to the starting time and the duration, detailed information about the impacts of activities on the environment and building energy consumption is also necessary to consider and evaluate possible energy management solutions.

In this work, a located activity is defined as what a person or a group does or has done during a time step in a given place. It includes:

- a label; environment, doors, windows,
- a timestep; domestic appliances, *etc.*);
- possible actions on some of context
- a location;
- controllable variables (indoor
- impacts on the context (changing

temperature, consuming electricity, etc.); • participants.

During a located activity, occupants can have some actions on some context controllable variables (turning on the domestic appliances, opening windows, etc.). These actions could affect some context elements and cause energy impacts as increasing the air temperature or consuming electricity. An example of this definition is presented in figure 3.1.

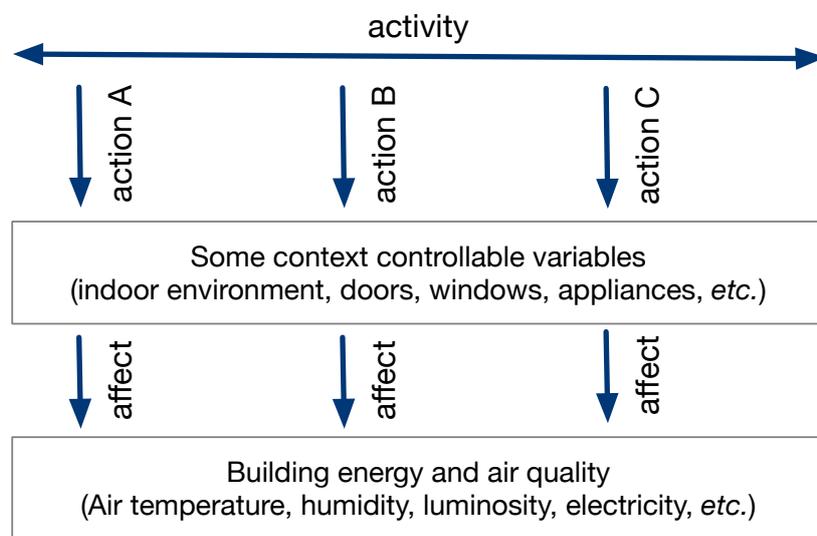


Figure 3.1 – Illustration of an activity

In the next, an activity will implicitly denote a located activity. In a simulation, an activity is usually simulated *via* representative patterns during the time of the day. These patterns usually include the probabilities that occupants' activities happen in each timestep. To consider it, statistical approaches define the non-located activity profiles as distributions, which present the conditional probability that non-located activities are performed given the time of day (Wilke et al., 2013; Aerts et al., 2014; Tanimoto et al., 2008). Figure 3.2 presents an example of a statistical activity profile in residential buildings proposed by Wilke et al. (2013).

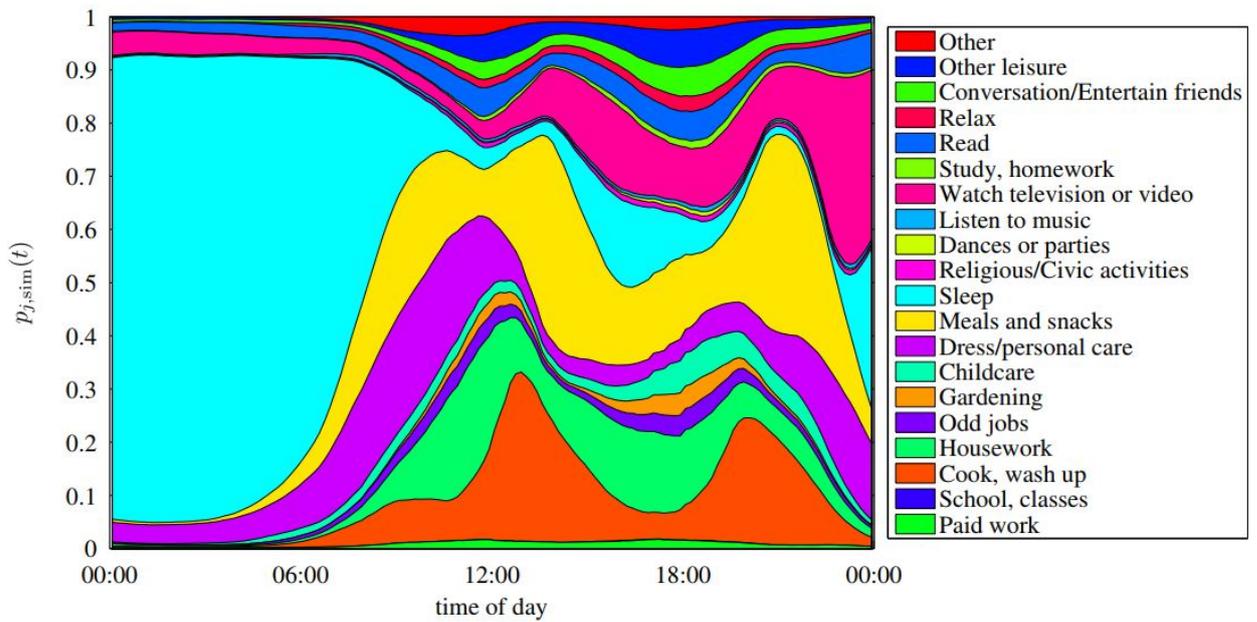


Figure 3.2 – Example of statistical activity profile in a residential building studied by Wilke et al. (2013)

3.2.4 Consequences of an activity

As mentioned above, an activity could comprise related actions that affect the indoor environment and consume energy in the house (Ouyang et Hokao, 2009; Pal et al., 2019; Chen et al., 2021). In this thesis, we define the consequences of an activity as its potential impacts on some controllable variables of the context (environment, actions on domestic appliances, actions on doors/windows, etc.) regarding building energy performance and human comfort when the activity happens.

For example, when someone studies in a living room, he can open windows and doors: these actions lead to a reduction of the CO₂ concentration in the room. For another instance, it is assumed that many people are cooking in the kitchen. It increases the CO₂ concentration and the air temperature. They often use a microwave or a toaster that consumes electricity. Investigating these consequences could provide occupants better understanding of their energy consumption, which could suggest improvements or changes in their practices.

3.3 Contextualized activity models: opportunities and challenges

3.3.1 Difference between representative and contextualized activities

As mentioned in chapter 2, many proposed statistical approaches did efforts to analyze the households' characteristics (number of people, building type, income, *etc.*) to estimate occupant activities in residential buildings from time-use surveys data (Aerts et al., 2014; Vorger, 2014; Wilke et al., 2013). However, representative activities contain average activities corresponding to aggregated data related to different contexts (national statistics combined with different types of actions, type of populations, type of places). Their outputs are not specific to particular households. Even in similar houses, the patterns of occupant activities are still diverse. For example, in a study carried out by Zaraket (2014), the author showed that for similar single households, the simulated electricity consumption of washing machine could fluctuate from 0 kWh/month to 19 kWh/month and the range of watching TV is from 0 kWh/week to 10 kWh/week. In another study, Aerts et al. (2014) analyzed that with the same building characteristics, the annual electricity consumption of a Belgian household with two people aged 25-39 working full-time can range from 200 kWh to 12000 kWh.

The objective of the statistical approaches is to model average activities for households with similar characteristics. Even though a household with average characteristics could exist and match by chance. Average households are virtual in the way that an average household behaving like the average of households is difficult to be found and therefore it leads to the validation issue in these households. More generally, an approach aggregating different contexts can not aim to address contextualized activities because it does not correspond to any particular context. Figure 3.3 presents an example of the above limitations in the statistical approach. To model representative activities for a household, the statistical data with different households is used to build an average model to determine representative activities in similar households. Because this model contains the information of many households, its results are not specific to any household and they are difficult to be verified. Besides, to estimate contextualized

activities in a household, a contextualized activity model is built based on the dataset of this household. It helps the models' results to be specific to the household and thus, to be verifiable.

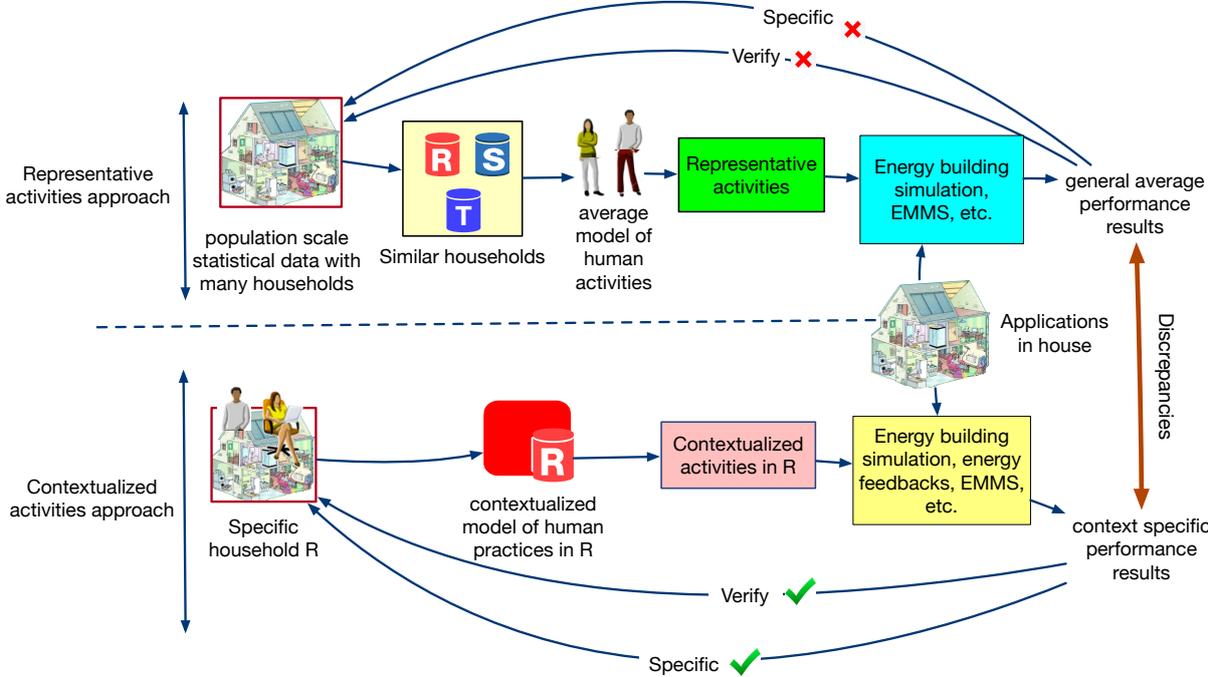


Figure 3.3 – Differences between representative activities and contextualized activities approaches

3.3.2 The reminder of limitations in existing occupant activity models

In chapter 2, some limitations in existing activity models have been presented as follow:

- Firstly, most of existing models focuses only on actions on windows/doors/blinds or ventilation. Only a few models concern occupants' daily activities such as cooking, watching television, and their energy impacts.
- Secondly, some data-based models with measurements were proposed to model occupant activities. However, the task of collecting labels is challenging. Cameras could be used but they are rarely permitted except in laboratories since the privacy concerns.
- Thirdly, most of the models are black-box models, data goes in, labels come out. These models contain complex mathematical functions to determine the relationship between the inputs and the outputs so that they are difficult for

many people to understand. Particularly in artificial neural networks, inputs are transformed through multiple layers with different functions to estimate the activity. Thus, in these models, relationships between activities and related variables are quite difficult to explain. However, these explanations are of great importance to give feedback and recommendations and enhance occupants' awareness of their activities compared to other households. Indeed, it could be useful for them to change some of their practices to save energy and improve their comfort in the house.

- Finally, statistical models are popular to simulate daily occupants' activities with survey data. However, statistical models' outputs are average and it leads to the question of the verification of such models in a specific context.

3.3.3 Contributions of contextualized occupant activities modeling

When modeling occupant activity, validation is an important task to evaluate the model performance. However, this task is challenging with representative activities because the average households are virtual and they do not have the actual dataset of occupants' activities. Thus, in existing households where the information of occupants' activities can be collected, contextualized activities are more interesting than representative activities because they are verifiable and more specific.

Modeling contextualized activities and their electricity consumption can contribute to many applications. To renovate an existing building, verifying its current energy performance is an important task. BPS software is usually used with the current building envelope and settings to support this step. In this task, contextualized activities can contribute some useful inputs (occupancy, actions on doors/windows, electricity consumption, *etc.*) to improve the performance of the simulations. Selecting retrofitting actions is another important task in building renovation. This task usually uses BPS software to evaluate the energy impacts of potential renovation strategies. Though occupants' activities after retrofitting actions can not be known, contextualized activities can be useful to construct activity profiles in energy simulations. These potential applications of contextualized activities in building renovation is presented in figure 3.4.

On the other hand, contextualized activity model can provide feedbacks and recommendations to occupants on their activities and energy impacts. They are

useful to enhance occupants' awareness of their practices and their energy impacts. Besides, contextualized activity model can extend the databases of occupants' activities in dwellings. This database is more specific than the statistical dataset. Thus, when considering a specific household, this database can be used to select the most representative household of this household.

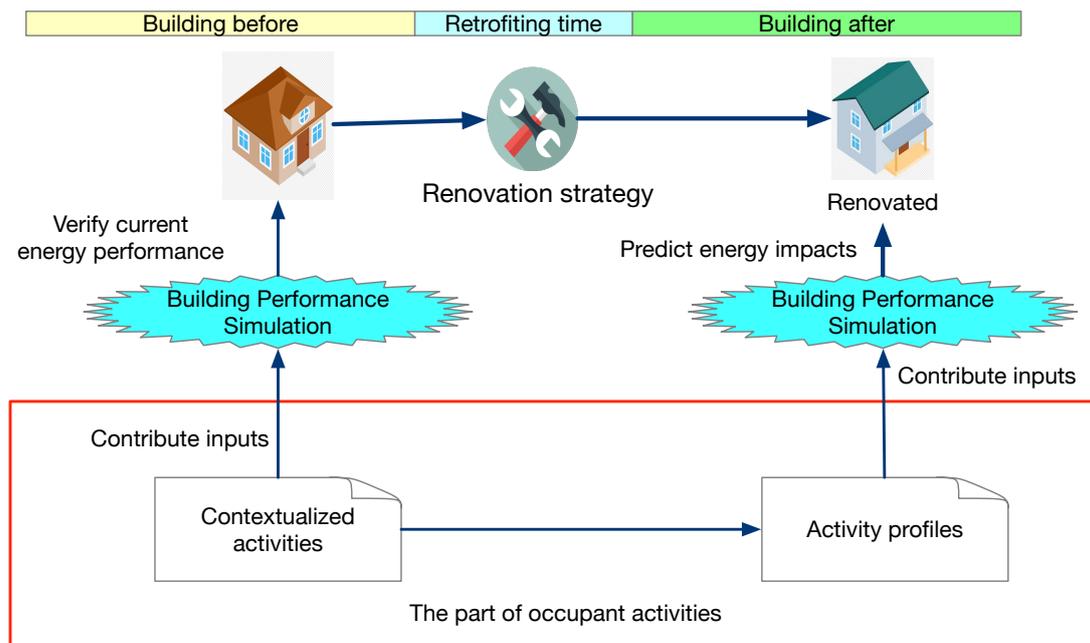


Figure 3.4 – Applications of contextualized activities in building renovation

With the above benefits, it is motivating to model contextualized occupant activity in houses. To deal with the limitations of existing approaches, contextualized activity models must satisfy these requirements:

- to be human-understandable;
- to consider the diversity of occupant activities.

3.4 Problematic: contextualized activities modeling in dwellings

This study aims to propose a methodology to overcome limitations of the existing models of activities in dwellings. In this section, we state the problem of modeling contextualized activities and their related energy impacts.

3.4.1 Objectives and research questions

3.4.1.1 Concerned activities

At home, many activities can be investigated. Wilke et al. (2013) modeled twenty activities such as relaxing, studying, cooking, *etc.* (see figure 3.2). They include both in-home activities and outside activities. In addition, many activities are difficult to distinguish, such as relaxing, reading and watching television. In another study, Aerts et al. (2014) focused on a set of eight activities related to electricity consumption: preparing food, vacuum cleaning, ironing, doing the dishes, doing laundry, using a computer, watching television and listening to music. However, some activities such as listening to music, using a computer or watching television are challenging to classify independently except using surveys. Here, the work focuses on modeling several activities that could have significant impacts on energy, particularly electricity consumption in houses.

The following activities are concerned:

- Doing personal care;
- Drying clothes;
- Washing dishes;
- Washing clothes;
- Cooking (breakfast, lunch, dinner);
- Entertaining (using a computer, watching television, listening to audio, *etc.*).

3.4.1.2 Research questions

The objective of the research is to address the following questions:

Question 1: *How to estimate contextualized occupant activities in specific residential buildings with a set of sensors?*

In this question, the objective is to propose a model to use sensors data to estimate concerned occupants' activities. The relationship between them needs to be investigated. Besides, the model must satisfy the requirements mentioned in the previous section.

Question 2: *How to evaluate the electricity consumption of activities in a given context?*

Occupants' activities can cause different types of energy impacts such as electricity consumption, impacts on HVAC systems or impacts on indoor air quality. However,

some energy impacts are difficult to measure directly and thus are difficult to validate. Therefore, this question aims to analyze relationships between activities and electricity consumed by appliances, which can be measured directly from sensors. For example, we would like to study when occupants used to cook, which appliances they used, and how they consumed electricity.

Question 3: *How to simulate occupant activities and their consumed electricity in specific houses?*

To support the evaluation of the building design in a renovation strategy, the objectives of this question is to :

- Investigate activity profiles and simulate occupant activities.
- Simulate the electricity consumption dependent on occupant activities.

3.4.2 Proposed methodology to model contextualized activities

This section proposes a general methodology with three contextualized models corresponding to three research questions:

1. Activity estimation model

- *Objective:* estimate occupant activities and construct activity profiles in specific households with a set of sensors. Specifically, at every time step of the day, the model aims to detect the labels of activities that are corresponding to in-situ measurements. This model should satisfy the requirements mentioned in previous section.
- *Inputs:* measurements of indoor conditions (temperature, humidity, etc.), movements, actions on doors/windows, and electricity consumption of appliances.
- *Outputs:* labels of activities in each time step of the day and constructed activity profile.
- *Validation:*
 - comparisons between estimated activities and observed activities.
 - statistically comparisons between estimated activity profiles and actual activity profiles.

2. Electricity model

- *Objective:* determine and evaluate the energy impacts, particularly in electricity consumption of contextualized occupant activities. Specifically, this model aims to analyze the used appliances and how they consumed electricity involved with concerned activities.
- *Inputs:* activities labels and actions on domestic electric appliances.
- *Outputs:* distributions of electricity consumption of appliances dependent on occupants' activities.
- *Validation:* the model is used to re-simulate electricity consumed by appliances. These simulations are statistically compared with measured electricity.

3. Activity simulation model

- *Objective:* simulate occupant activities and their consumed electricity. This simulation is used to evaluate the retrofitting actions when renovating existing buildings.
- *Inputs:* activity profile determined from estimation model and electricity model.
- *Outputs:* simulated activities and electricity consumed by related appliances in a long-term period of time.
- *Validation:* comparison of consumed electricity of appliances between contextualized approach and measured consumptions.

These models are combined together to represent the occupant activities and their related electricity consumption in houses (see figure 3.5).

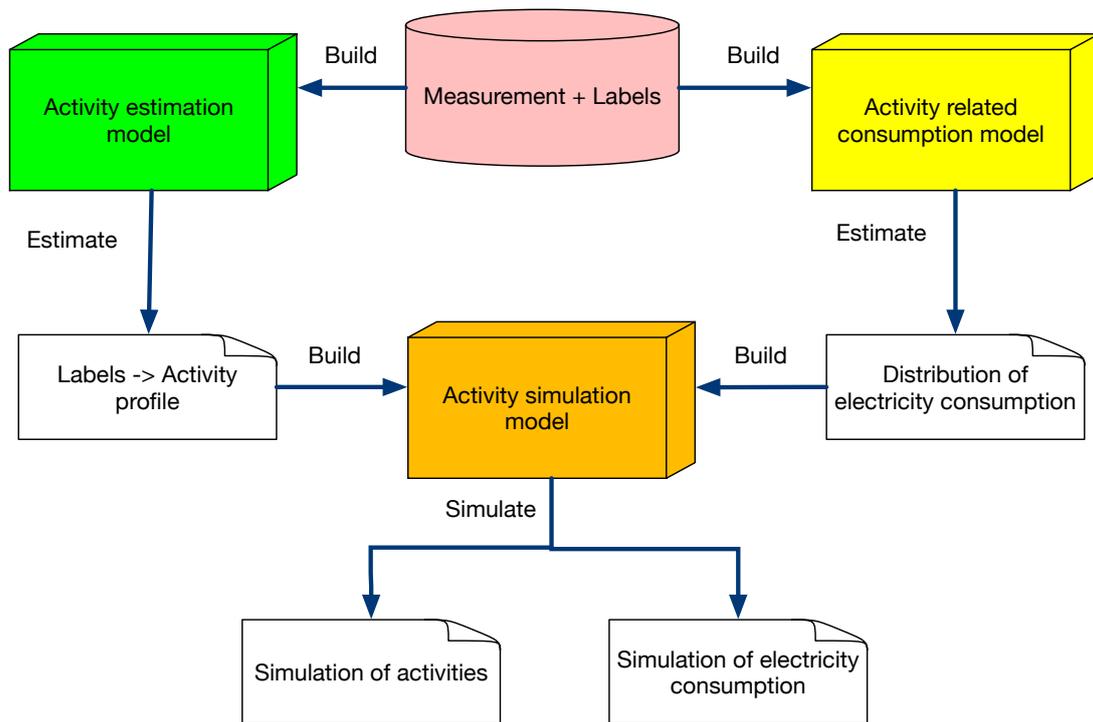


Figure 3.5 – General methodology to model contextualized occupant activities and their related electricity consumptions

The steps to model a contextualized activity in the house are presented as follows:

1. The necessary dataset is collected from installed sensors and from occupants in situ. To consider the context, the methodology uses both measurement and labels of activities.
2. From the collected dataset, an **activity estimation model** is built to estimate occupant activities. This model is then applied to extrapolate the labels of activities that are missing in the dataset. Estimated labels from the **activity estimation model** are used to build the activity profiles.
3. An **activity related electricity consumption model** is built based on the measurements and activity labels. This model determines the relationships between activities and related electricity consumption of appliances.
4. The estimated activity profile and the electricity consumption distribution determined by the **electricity model** are used to build an **activity simulation model**. This model simulates occupants' activities and electricity consumed by related appliances in long-term periods of time.

3.5 Conclusion

Existing activity models have limitations (see chapter 2). There are not verifiable (statistical ones), hardly understandable (data-based ones) and highly complex (agent-based ones). To deal with these issues, this study aims to propose a methodology to model contextualized occupant activities that would:

- be verifiable and human-understandable;
- enhance awareness of occupant on activities regarding energy in situ;
- support energy simulation in verifying building energy performance and selecting retrofitting actions in building renovation.
- give explanations on occupant activities and their energy consumption by investigating the relationship between them;
- provide outputs taking into account the diversity of occupant activities. Thus, it should be a generative model.

To reach these contributions, three contextualized activity models are proposed with the objectives as follows:

1. **activity estimation model** - to estimate labels of occupant activities, based on the measurements and the contexts in houses.
2. **electricity consumption model** - to estimate the consequences and their electricity impacts when activity happens in houses.
3. **activity simulation model** - to simulate sequences of activity and their electricity consumption in specific houses.

These models are aggregated together to represent the impacts of occupant activities in houses. The descriptions and the implementations of these models will be presented in the next chapters. Specifically, chapter 5 presents the methodology of the activity estimation model. Chapter 6 focuses on the activity consequences model and the activity simulation model is described in chapter 7.

Chapter 4

Data monitoring and case studies descriptions

Contents

4.1 Case study descriptions	60
4.1.1 Cestas	60
4.1.1.1 Building plan	60
4.1.1.2 Household descriptions	60
4.1.2 Bordeaux	62
4.1.2.1 Building plan	62
4.1.2.2 Household descriptions	63
4.2 Sensors for occupant activity estimation	64
4.2.1 Sensors in existing activity models	64
4.2.1.1 PIR Motion Sensor	64
4.2.1.2 Power sensor	65
4.2.1.3 Door/windows contact sensor	65
4.2.1.4 CO ₂ Sensor	66
4.2.1.5 Acoustic sensors	66
4.2.1.6 Air temperature, luminosity and relative humidity sensors	66
4.2.2 Sensors location in the case studies	67

4.2.2.1	Cestas	67
4.2.2.2	Bordeaux	68
4.2.3	Protocol to monitor, process and store sensors data	69
4.3	User interface application to collect activities labels	70
4.3.1	Collecting activities labels from occupants	70
4.3.2	Proposed approach: a mobile application	71
4.3.3	Reconciliation of collected labels	74
4.4	Conclusion	75

4.1 Case study descriptions

For this thesis work, two French households were studied: a house in Cestas and an apartment in Bordeaux, both in the South West of France. The description of the buildings and the characteristics of each household is given in this section.

4.1.1 Cestas

4.1.1.1 Building plan

The building plan of the case study in Cestas is shown in figure 4.1 with the following features:

- *Description*: It is a renovated rural house with an inhabited timber frame extension inhabited. The house is on one level. It includes a kitchen-living room-entrance unit as well as three bedrooms, one of which has a private bathroom, the kitchen and the meal room are in the same heating zone (see figure 4.1).
- *Dimensions*: The house has an approximate floor area of 150 m² and a height of 2.50m. Thus, the volume of the house is approximately 375 m³.
- *Location*: Cestas (near Bordeaux), France.

4.1.1.2 Household descriptions

- *Members*: Many studies focus on the number of members, the range of the age youngest child, the number of vehicles and the total monthly income as main

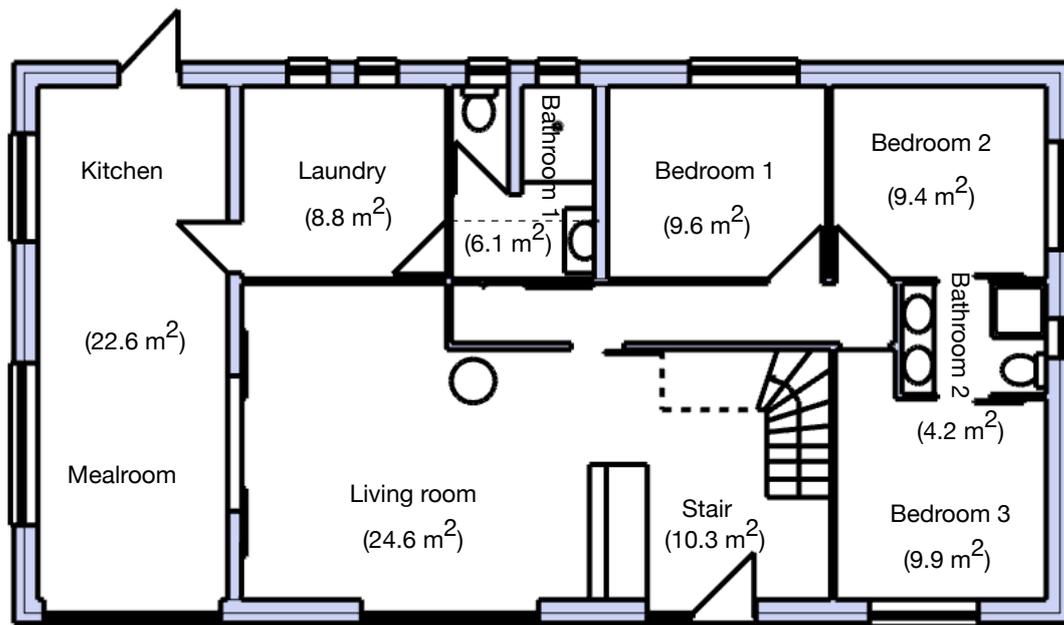


Figure 4.1 – Building plan of the case study in Cestas, France

household features (Wilke et al., 2013; Vorger, 2014; Aerts et al., 2014). The Cestas household is composed of five members with the youngest child in the range of 5-12 years old. There are more than two vehicles and the monthly income is greater than 4000 €/month. The characteristics of the members are described in table 4.1.

Table 4.1 – Household members characteristics in the case study in Cestas, France

Number	Gender	Age	Job	Study level
1	Man	+ 40 years old	Full-time	Master or higher
2	Woman	+ 40 years old	Full-time	Master or higher
3	Woman	18-40 years old	No	Bac-Bac+4
4	Woman	-18 years old	No	Secondary
5	Man	-18 years old	No	Lower than secondary

• *Appliances*: The following electrical appliances are available in this household:

- Microwave;
- Coffee machine;
- Washing machine;
- Oven;
- Toaster;
- Dishwasher;

- | | | |
|------------------|-----------------|-----------------------|
| - Cooking robot; | - Hair dryer; | - 2 desktop computers |
| - Tumble dryer; | - Hi-fi system; | and 3 laptops; |
| - Iron; | - Radio; | - Television; |
| - Vacuum; | - Internet box; | - Fridge. |

4.1.2 Bordeaux

4.1.2.1 Building plan

The building plan of the case study in Bordeaux are shown in figure 4.2 with the following features:

- *Description:* It is an urban apartment of type "Echoppe" on street in northeast facade, on court in southwest facade. The house has three levels and the rooms in each level are shown in figure 4.2.
- *Dimensions:* The house has three levels with the following surface area: cellar 40 m², 1st floor 75 m², 2nd floor 35 m².
- *Location:* Bordeaux (in the sector of the Pessac barrier) in France.

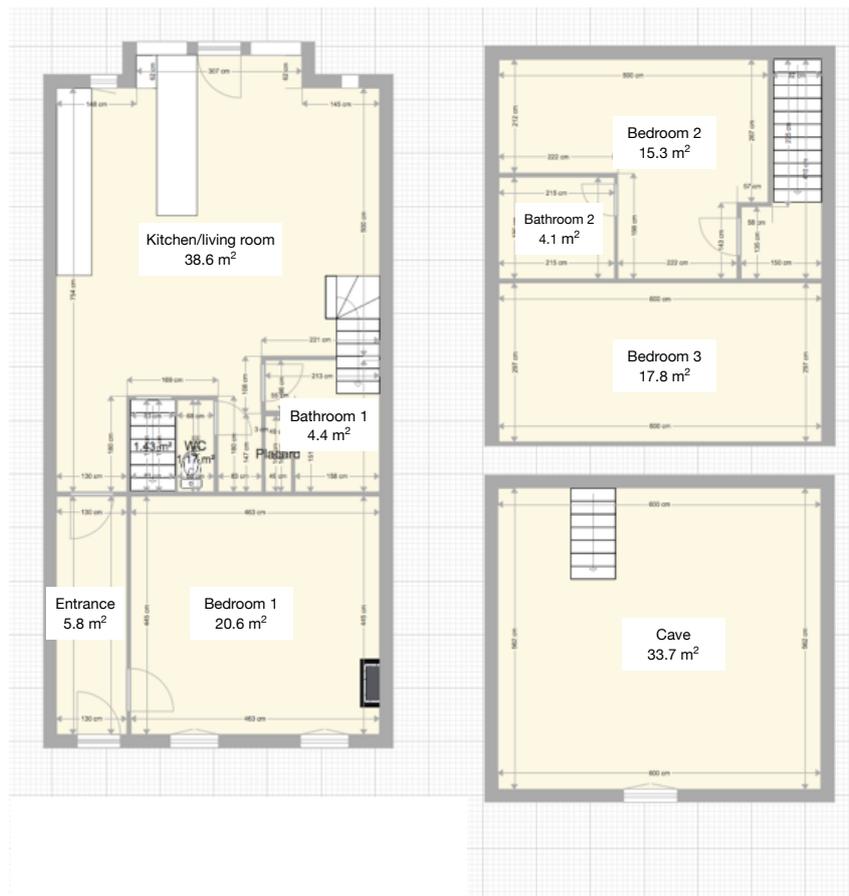


Figure 4.2 – Building plan of the case study in Bordeaux, France

4.1.2.2 Household descriptions

- *Members:* The household has four members with the youngest child in the range 5-12 years old. There are more than two vehicles and the monthly income is from 1900€/month to 4000€/month. The characteristics of the members are described in table 4.2.

Table 4.2 – Household members characteristics in the case study Bordeaux, France

Number	Gender	Age	Job	Study level
1	Man	+ 40 years old	Full-time	Master or higher
2	Woman	+ 40 years old	Full-time	Master or higher
3	Woman	-18 years old	No	Lower than secondary
4	Man	-18 years old	No	Bac-Bac+4

- *Appliances:* The following electrical appliances are available in this household:

- Oven;
- Hotplate;
- Microwave;
- Coffee machine;
- Toaster;
- Boiler;
- Cooking robot;
- Washing machine;
- Dishwasher;
- Tumble dryer;
- Vacuum;
- Hi-fi system;
- Internet box;
- Laptops;
- Television;
- Fridge.

4.2 Sensors for occupant activity estimation

The estimation of occupant activities highly depends on collected features from installed sensors and discovers relationships between these features with human practices. The use of sensors comes from the assumption that when occupants perform some activities, they interact with the surroundings, which could affect environmental conditions (CO₂ concentration, air temperature, air humidity, *etc.*) and consumes electricity when use appliances are used (microwave, computer, TV, *etc.*). It raises the question of determining sensors that need to be installed. This section analyzes common sensors used in existing works dealing with occupant activities and presents a protocol to manage sensors data in the case studies.

4.2.1 Sensors in existing activity models

4.2.1.1 PIR Motion Sensor

A passive infrared sensor (PIR sensor) is generally an electronic sensor that measures infrared (IR) light radiation emitted from objects which generate heat in its fields of view. PIR sensor detects changes in the amount of infrared radiation impinging upon it that varies depending on the temperature and surface characteristics of the objects in front of the sensor. When an object, such as a person, passes in front of a PIR sensor, the sensor detects this movement based on the difference between the room temperature and the body temperature. PIR sensors are usually used to detect the inhabitant presence in buildings. Elbayoudi et al. (2019) used PIR sensors in estimating and identifying trends in activities of daily living or activities of daily working. From these trends, the authors could predict the evolution of human behaviors. In another

study conducted by Amayri (2017), PIR sensors were used to detect the occupancy and to support the estimation of activities in residential buildings. Akbari et Haghghat (2021) used information from PIR motion sensors to extract the household's daily occupancy patterns and to link them to load-shifting of the energy consumption.

4.2.1.2 Power sensor

A power consumption sensor is generally an electronic sensor used with an electric socket to measure the electricity use in this socket. The sensor captures the instantaneous energy consumed by various plugged-in appliances. It provides information on how humans use electrical appliances and their related energy consumption. Amayri (2017) used power consumption sensors with Bayesian Network to estimate occupancy in the office and human practices in residential buildings. Sonta et al. (2018) used the data of electricity consumption and unsupervised learning technique to classify occupant activities.

4.2.1.3 Door/windows contact sensor

A contact sensor is generally a magnetic sensor used to detect the opening/closing of doors or windows in buildings. It usually comprises two parts - a sensor and a magnet. The sensor is usually placed on the stationary frame to avoid damage, and the magnet is usually placed on the moving portion of the door/windows. Whenever the door is opened, the magnet moves and becomes separated from the sensor. This event triggers the sensor to detect the opening/closing of doors/windows. Lee et Malkawi (2014) used the actions on doors/windows in modeling occupant activities with agent-based systems. Agents could change these actions to optimize their defined comforts. Alyafi (2019) analyzed the relationships between the actions on doors/windows and the effects on indoor environmental conditions such as air temperature and CO₂ concentration. These identified relationships were used to recommend better actions on doors/windows.

4.2.1.4 CO₂ Sensor

A CO₂ sensor is used to measure the concentration of carbon dioxide present in the air, in particles per million (ppm) of CO₂, which could represent a feature of the air quality in the building. This sensor usually provides useful information related to human presence. Indeed, through respiration or breathing, the human body uses oxygen and produces CO₂. Wolf et al. (2019) proposed a Markov-switching model with CO₂ level for occupant activity estimation. Amayri (2017) used CO₂ concentration with Decision Tree algorithm to estimate the number of occupants in the office. Alyafi (2019) also used a physical model and an optimization method to analyze the impacts of occupants' actions on doors/windows on indoor conditions such as CO₂ concentration, air temperature, *etc.*

4.2.1.5 Acoustic sensors

A single omnidirectional microphone can be used to recognize various activities in a room. Omnidirectional microphones can pick up sound from virtually any direction. They are considerably cheaper than having multiple unidirectional microphones, and prove to be much advantageous in places where it is required to track/listen to multiple sources like in meetings, and discussions, *etc.* Amayri (2017) and Padmanabh et al. (2009) used acoustic sensors in detecting occupancy and estimating occupant activities in conference rooms. In another study conducted by Kim et al. (2020), acoustic sensors were used with deep learning technology to detect and monitor occupant behavior in single-person households.

4.2.1.6 Air temperature, luminosity and relative humidity sensors

An air temperature, luminosity and air relative humidity sensor is generally an electronic sensor used to measure the temperature and the relative humidity of an air, which represents the environmental conditions. This sensor provides potentially useful information related to the consequences of occupant activities and their effects on the building. Chen et al. (2021) and Alyafi (2019) mentioned that the windows behaviors could affect environmental conditions (temperature, relative humidity). Fabi et al. (2013) proposed the models of occupant's behavior patterns to investigate how different

behavior patterns influence indoor air temperature and relative humidity. Results showed that the indoor environmental quality is related to the occupants' behavior.

4.2.2 Sensors location in the case studies

In this work, the following sensors in modeling occupant activity were used:

- PIR motion sensors;
- Power sensors;
- Door/windows contact sensors;
- CO₂ sensors;
- Air temperature, luminosity and relative humidity sensors.

The installation and the management of the sensors in the case studies are presented in the next.

4.2.2.1 Cestas

In this case study, the following sensors were installed:

- 7 PIR motion sensors;
- 10 contact sensors;
- 16 power sensors;
- 5 CO₂ sensors and 4 air temperature and relative humidity sensors.

There are many available domestic appliances in this household but just some of them were measured:

- Each power sensor was used to measure the electricity consumption of a coffee machine, a toaster, a cooking robot, a washing machine, a dishwasher, a tumble dryer and a fridge.
- A microwave and an oven were measured by a same power sensor. In the next, the microwave will implicitly denote the consumption of both the microwave and the oven.
- Audio and video devices in the living room were measured by a same power sensor. In the next, the audio and video devices will denote to the consumption of these appliances.

Figure 4.3 presents the location of PIR motion sensors, power sensors and door/windows contact sensors in the rooms in the Cestas case study. Some pictures of installed sensors are also shown in this figure.

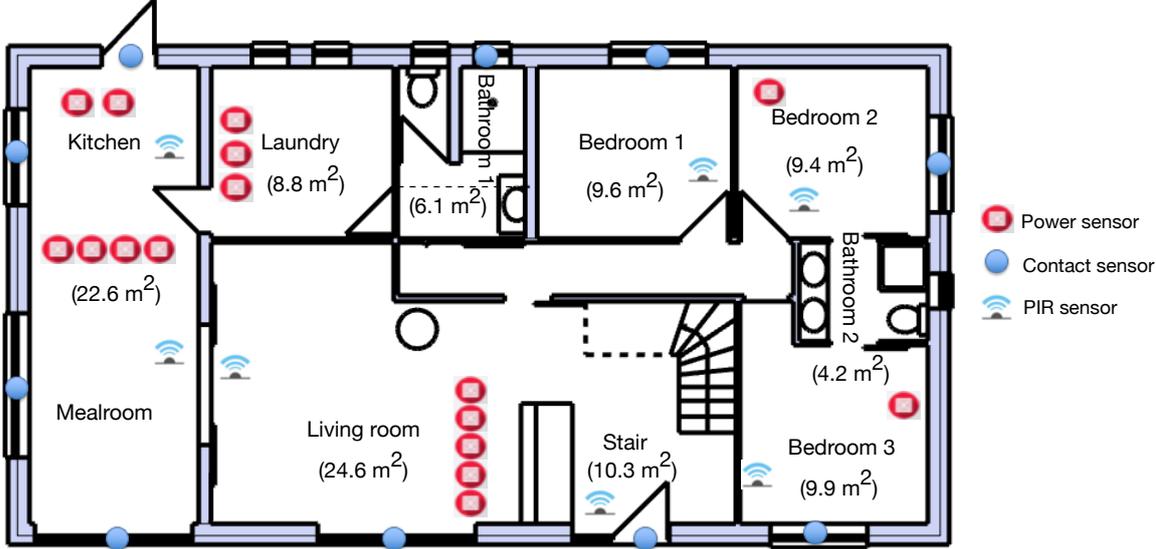


Figure 4.3 – The positioning of PIR sensors, power sensors, contact sensors and some of their images in Cestas site

Sensor data in this case study were recorded for a consecutive period of about one and a half years (December 2018 to September 2020). The frequencies of data are different depending on the type of sensors.

4.2.2.2 Bordeaux

In this case study, the following sensors were installed:

- 5 PIR motion sensors;
- 9 contact sensors ;
- 10 power sensors;
- 3 CO₂ sensors and 3 air temperature and relative humidity sensors.

Figure 4.4 presents the locations of PIR motion sensors, power sensors and door/windows contact sensors in the rooms in the Bordeaux case study. Some pictures of installed sensors are also given in this figure.

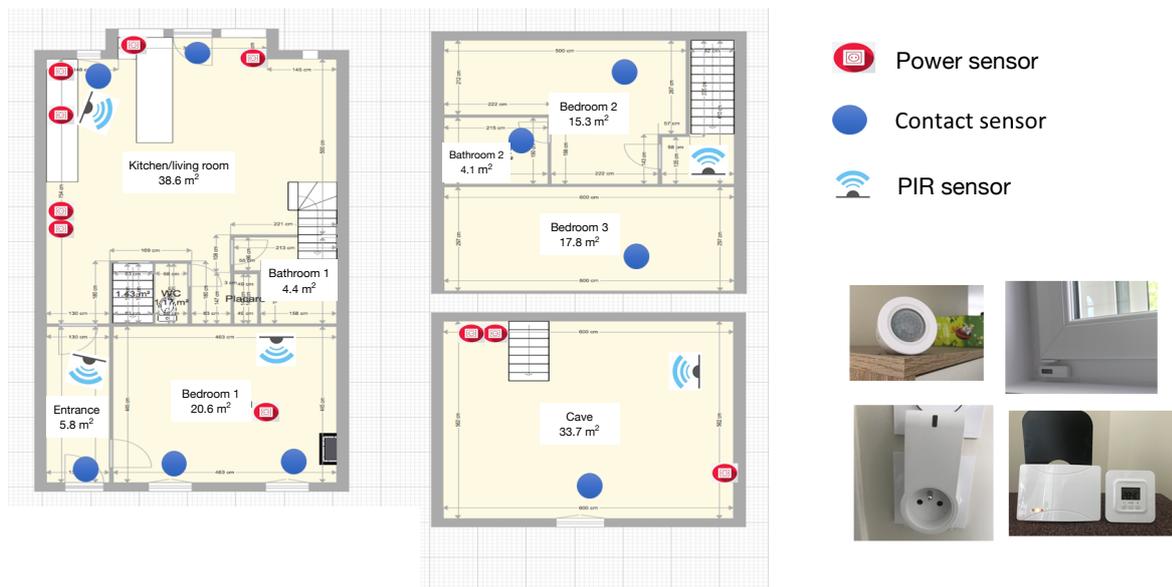


Figure 4.4 – The position of PIR sensors, power sensors, contact sensors and some of their images in the Bordeaux site

Sensor data in this case study were recorded for a consecutive period of about nine months (October 2020 to July 2021). The frequencies of data are different depending on the type of sensors.

A large number of installed sensors in the case studies leads to costly implementation. It raises the question: Is it possible to reduce some of these sensors but ensure the quality of the occupant activity models? Answering this question can help us to save the budget for future implementation campaigns.

4.2.3 Protocol to monitor, process and store sensors data

In each case study, the task of data processing can be summed up to six steps presented in figure 4.5. These steps are:

1. Once sensors are set up on sites, they measure specific data with different frequencies depending on the type of sensors.
2. Via specific communication channels (LORA, EnOcean, etc.) of the installed sensors, the data are sent to a local computer.

3. An acquisition program has been developed in the LabView environment. It gathers all the data in CSV files.
4. The generated CSV files are pushed to a remote server thanks to a secure communication channel (SSH).
5. A service is running on the server that reads the CSV files and sends the data to a storage database. The database is post-processed to add metadata and generate alerts in the events of missing data or outliers detection.
6. The raw and clean data can be viewed in a graphical interface from which the data can be extracted and used elsewhere to create user models.

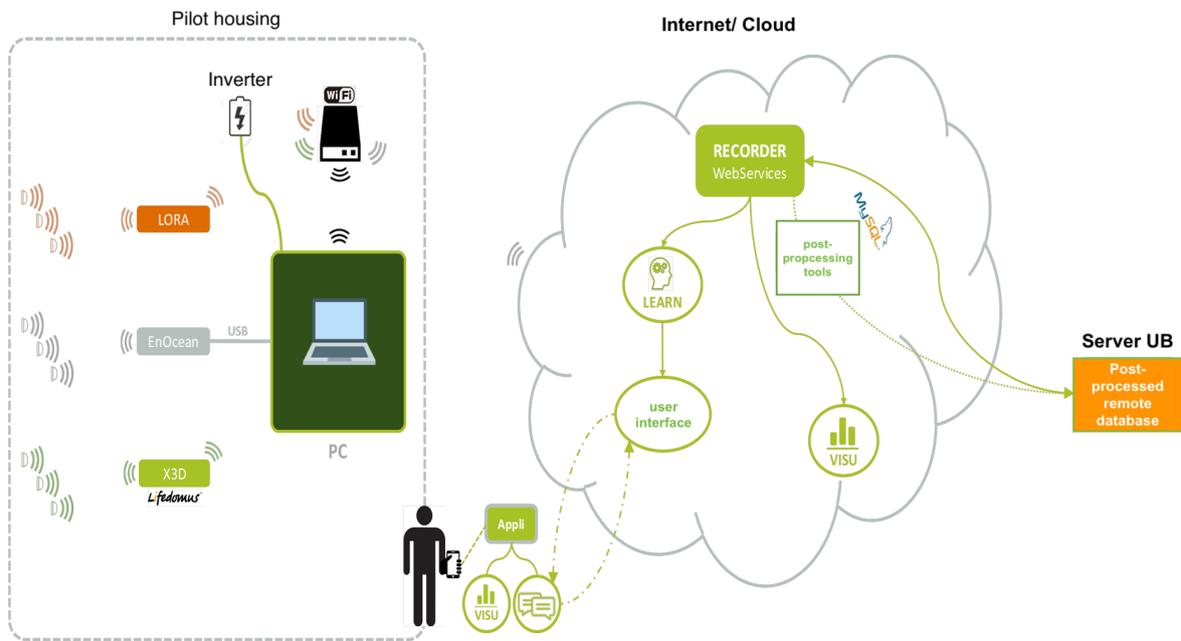


Figure 4.5 – Protocol and sensors data management system in this work

To ensure the storage, provision, and graphic representation of the data, we opted for a solution combining a time series oriented database (*InfluxDB*) and a web interface visualisation tool (*Grafana*).

4.3 User interface application to collect activities labels

4.3.1 Collecting activities labels from occupants

As discussed in chapter 3, in spite of being very suitable for the estimation and modeling of specific contexts, contextualized activity models require a phase of observation of

occupants’ practices which is difficult to envisage outside of a laboratory context due to the privacy policy (Amayri, 2017). Another approach is to use several surveys to collect the labels of human activities in the house. However, these statistical labels are average and are not specific to estimate human activities in particular houses. Besides, houses are complex case studies, where inhabitants perform their activities in different rooms. So, it requires to have an approach that allows inhabitants to provide their activities labels independently. In addition, occupants could feel uncomfortable if they need to spend too much time providing activities labels. Thus, their conveniences and their comforts should be considered in the way of collecting labels.

4.3.2 Proposed approach: a mobile application

To collect the activity labels, a mobile application has been developed during the Ph.D. It has the following features:

- The application is distributed to every member of the household. It allows occupants to provide their labels independently and privately. Each member is identified through his/her name. The collected labels are then saved into a private remote database. These steps are described in figure 4.6.

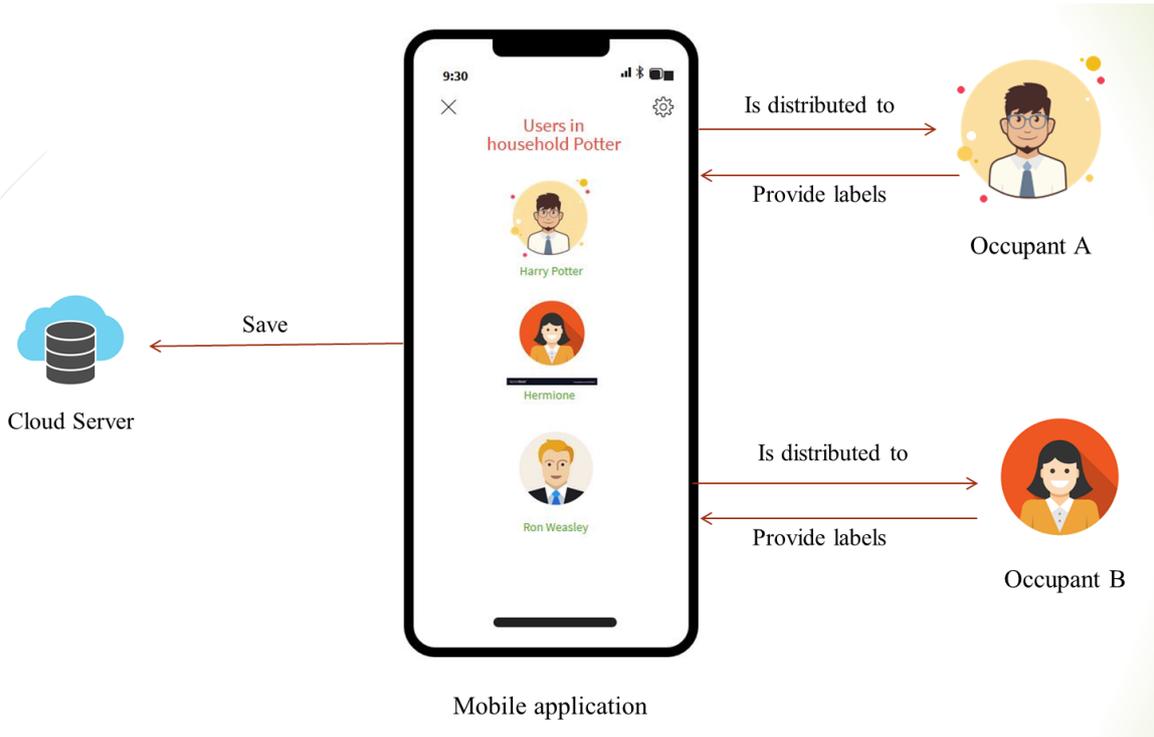


Figure 4.6 – The user interface of the developed mobile application for collecting labels

- To help occupants to review their activities, the application can show calendar and historical labels. Figure 4.7 shows an example with an unlabeled day. The days where labels have been collected, are marked with green circles. The red circles belong to the days where activities were not labeled.



Figure 4.7 – The calendar of labeled/unlabelled days (left) and an example of an unlabelled day (right) in the developed application

- The application helps occupants to provide labels of activities in the current day or previous days. Figure 4.8 presents application during the labeling of activities for a specific day. Considering occupants' comfort in labelling, 30-minutes timesteps are used in this application. A day is divided into 30-minutes timesteps, the occupant labels the activities for each timestep of the day. Some activities could be performed in different rooms and they are considered as different activities. For example, an individual can study in the bedroom or in the living room. Studying in the bedroom is different from studying in the living room. The application allows occupants to specify their locations.

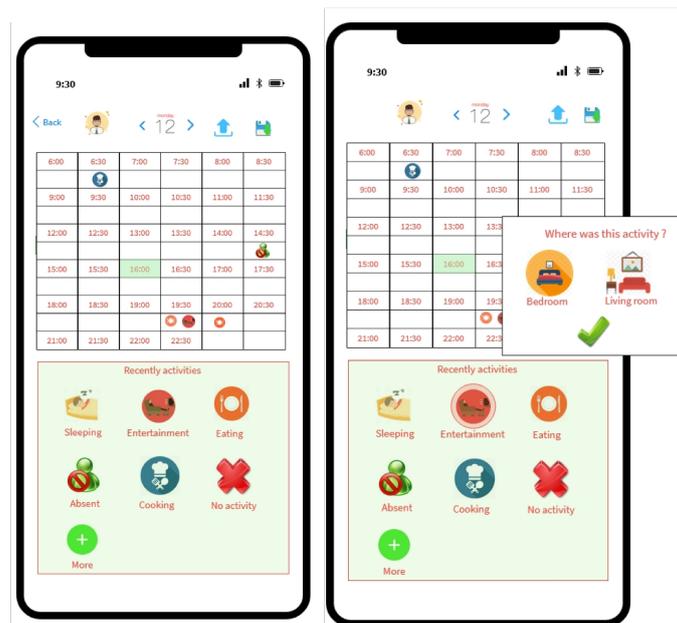


Figure 4.8 – An example of a labeled day (left) and an example of the function of labeling a multi-room activity (right) in the developed application

- We assume that humans have habits and they can have similar patterns of activities for many days. To help occupants to save time and review their habits of activities in the past, the application allows the occupant to save their typical patterns of activities. Users can use these typical daily patterns to fill similar days regarding their habits. Figure 4.9 shows an example of a typical pattern (on Monday) and the way to save a typical pattern of activities in a day.

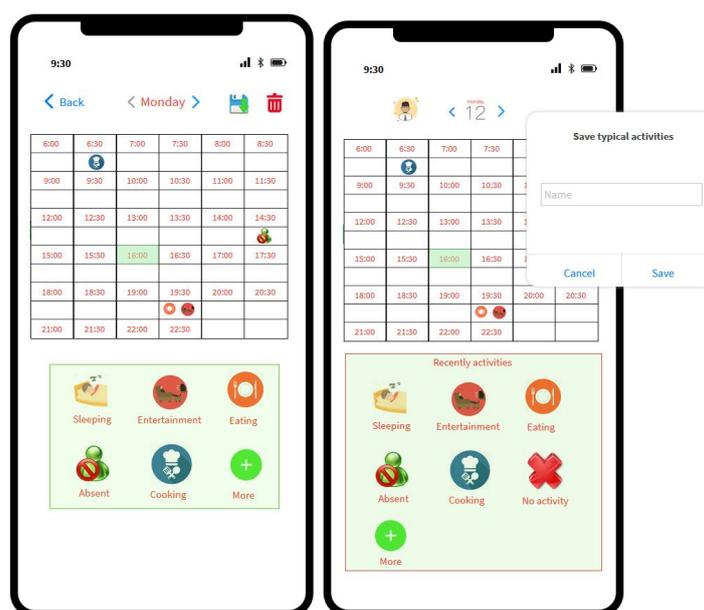


Figure 4.9 – The features of saving typical patterns of activities (right) and auto-fill similar days (left) in the developed application

Using the mobile application can help occupants to feel more comfortable in providing labels because of it is flexible to be used in many places and in many times. This application could be applied to different households with different contexts. To do it, some configurations need to be collected to modify some settings (number of members, list of activities, locations of activities, *etc.*) of the application. The calendar and the typical patterns are useful for inhabitants to remember their missing activities based on their past habits.

However, there are limitations in the application:

- It requires a lot of information from an household like the list of inhabitants, the list of activities and located activity settings of each inhabitant (mentioned in chapter 3). It is a challenge to implement at population-scale with numerous houses.
- Occupants could loose interest in the application over time. For example, they could feel bored or forget to use the application.

To deal with these issues, we proposed a notification feature. It allows sending notifications to occupants to ask them to use the application if they didn't do it for a long time (several days). Though similar to many existing applications in smartphones, notifications could annoy the occupants. Other approach such as interactive learning (Amayri, 2017) should be considered to deal with this problem.

4.3.3 Reconciliation of collected labels

Because the application is distributed to each inhabitant, the labels collected from all inhabitants can be inconsistent. For example, an inhabitant can declare that he studied in the living room at 8 p.m while another one declares that she did not study at that time because she was outside. Therefore, the collected labels need to be reconciled in order to build a consistent model for the whole household. To do it, we consider each activity independently and the collected labels of an activity are reconciled using the logical **OR** operator. For example, in a timestep t , the labels of an activity collected by n inhabitants are l_t^1, \dots, l_t^n with $l_t^i \in \{0, 1\} \forall i, t \geq 0$. The reconciled label l_t in this timestep is described in equation 4.1.

$$l_t = \mathbf{OR}(l_t^1, l_t^2, \dots, l_t^n) = l_t^1 \vee l_t^2 \vee \dots \vee l_t^n \quad (4.1)$$

4.4 Conclusion

In this chapter, two households in Cestas and Bordeaux have been presented to implement and validate the proposed contextualized activity models. Occupants can perform their activities, interact with the surroundings, affect environmental conditions and use electrical appliances, *etc.* It leads to the necessity of using sensors to capture these effects. Some existing sensors in activity models are analyzed and the following sensors were selected and installed in the case studies:

- PIR motion sensors;
- Power consumption;
- Door/Windows contact sensors;
- CO₂ sensors;
- Air temperature and relative humidity sensors.

The case studies are described with their building plan, household structures, and the installation and the management of the sensors.

To model contextualized activities in dwellings, most of the existing approaches require a phase of observation of human practices or using cameras. These requirements are challenging to practice outside of a laboratory context. To deal with it, we have proposed a mobile application to support occupants during the labeling process. They could use the application to provide their labels of activities at convenient moments.

The next chapters 5, 6, 7 present the methodology of contextualized activity models and their implementations in the case studies.

Chapter 5

Estimation of contextualized activities

Contents

5.1 Introduction	78
5.2 Methodology to estimate contextualized activities	78
5.2.1 Problem statement	78
5.2.2 Preprocessing sensors data	82
5.2.3 Generating features	84
5.2.4 Determining meaningful features	87
5.2.5 Building activity estimation model	91
5.2.5.1 Bayesian Network	91
5.2.5.2 Consequences-based Bayesian Network for activity estimation	93
5.2.6 Validating the model	95
5.3 Results: the cooking activity	96
5.3.1 Case study 1: Cestas	97
5.3.2 Case study 2: Bordeaux	111
5.4 Conclusions	115

5.1 Introduction

To consider impacts of humans on energy consumption, many approaches have been proposed. They estimate occupant activities and their related energy impacts. There are still limitations in these models, especially in modeling occupant activities in specific households (mentioned in chapter 2). To deal with these limitations, a general methodology with activity models to tackle the research questions were presented in chapter 3. The objective of this chapter is to answer the question: "*How to estimate occupant activities in specific residential buildings with a set of sensors?*".

Specifically, a general approach is proposed to determine sensors that could be useful to estimate occupant activities in the house. In our study, different sensors were installed to estimate activities, including motion detection, power measurement, CO₂ concentration, indoor air temperature, air relative humidity, luminosity, and door/window opening. The proposed approach is inspired by a knowledge-based approach with a graphical model. It starts by determining the most meaningful features. Then, estimation algorithms are proposed which rely on Bayesian Network. They aim to explore the relationships between human activities and energy-related context factors in the house. Further, the model must be human-understandable and be verifiable.

In this chapter, one example of the cooking activity is used to illustrate all methodological aspects. Then, the model is implemented in the two case studies (Cestas and Bordeaux) to test the methodology and its generalization. The estimated results are then compared to actual results to assess the model accuracy.

5.2 Methodology to estimate contextualized activities

5.2.1 Problem statement

As mentioned in the previous chapters, occupants' activities play an essential role in the building energy consumption. Determinations of occupants' activities are crucial to assess the energy impacts of occupants on building performance and evaluate building energy performance. It is challenging to detect occupants' activities directly using

cameras due to the privacy policy. The context factors could provide information about these activities. During an activity, occupants can have many actions such as opening doors or operating appliances. These actions affect and cause changes to some context factors (see figure 5.1). This relationship is helpful to deduce occupant activities from the context. For example, to study, an occupant goes to the bedroom, closes the door and turns on the light in the room. Its actions affect the indoor environment in the bedroom. Its presence increases the CO₂ concentration, its actions on the lighting system increase the luminosity in the room. In addition, many context factors can be collected through some basic sensors (temperature, doors contact, power consumption, etc.), which are more acceptable than cameras or wearable sensors.

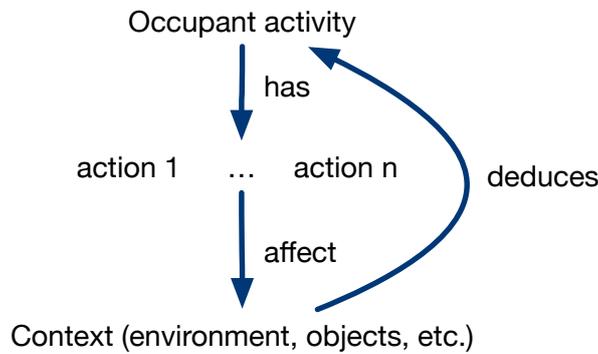


Figure 5.1 – The relationship between occupant activity and context

Here, the work proposes a model M to estimate the states of occupants' activities from sensors, which are installed to measure the context information. In formula, at timestep t , the context is represented by the vector X_t with n factors x_{kt} , $k \in [1; n]$, state a_t of an activity a is determined as in equation 5.1.

$$a_t = M(X_t) = M(x_{1t}; ..x_{kt}..; x_{nt}) \quad (5.1)$$

with:

$$a_t = \begin{cases} 0 & \text{activity } a \text{ does not happen at } t \\ 1 & \text{activity } a \text{ happens at } t \end{cases}$$

For example, in September 2nd, 2019, from 18:00 to 18:30, when the data of air relative humidity, luminosity and temperature in the kitchen are available, we would like to estimate whether occupants cook during this period (see figure 5.2).

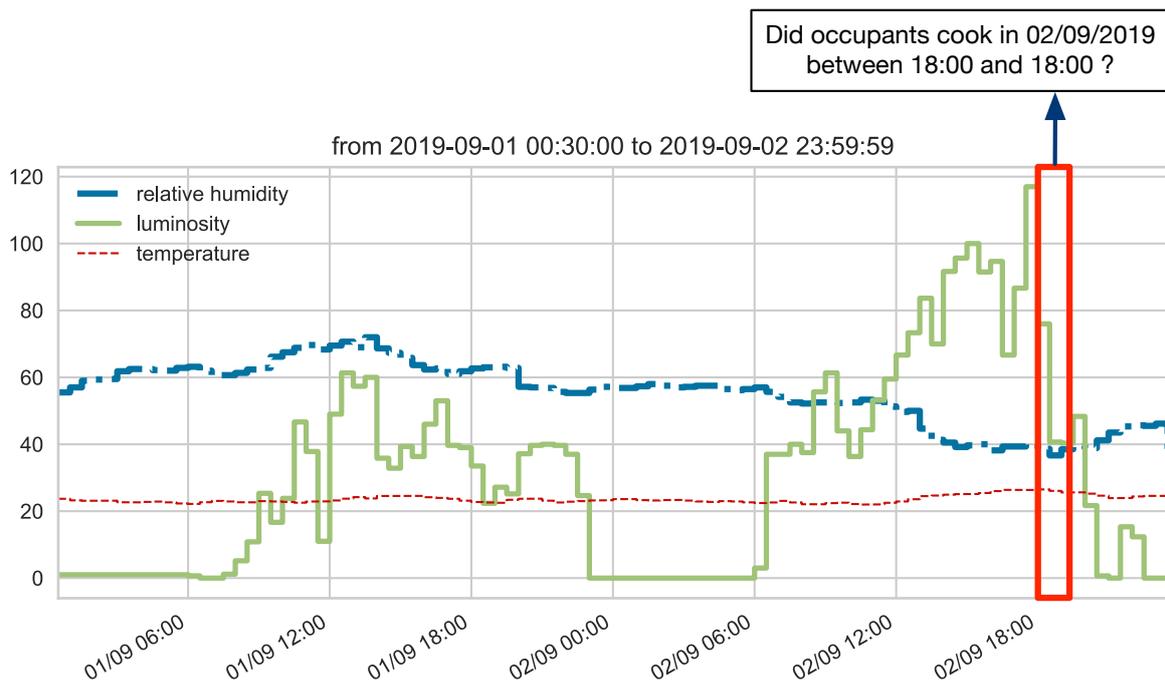


Figure 5.2 – An example of cooking estimation problem

Objectives and contributions

To estimate occupant activity, this study deals with the two following problems:

1. Determining a set of meaningful information to estimate occupant activity. There are many factors in a home context. Some information is useful to estimate occupant activity while others are not. Thus, detecting information that is related to the activity is important.
2. Building a model to deduce the states of activity from selected information. To tackle the limitations of the existing approaches (not verifiable, deterministic output or not understandable) mentioned in chapter 2, the model must meet the following requirements:
 - be friendly for humans to understand and explain the parameters and results;
 - have stochastic outputs to integrate into energy simulations;
 - be verifiable.

General methodology

To address the above problems, a methodology is proposed in figure 5.3 with five main steps as follows:

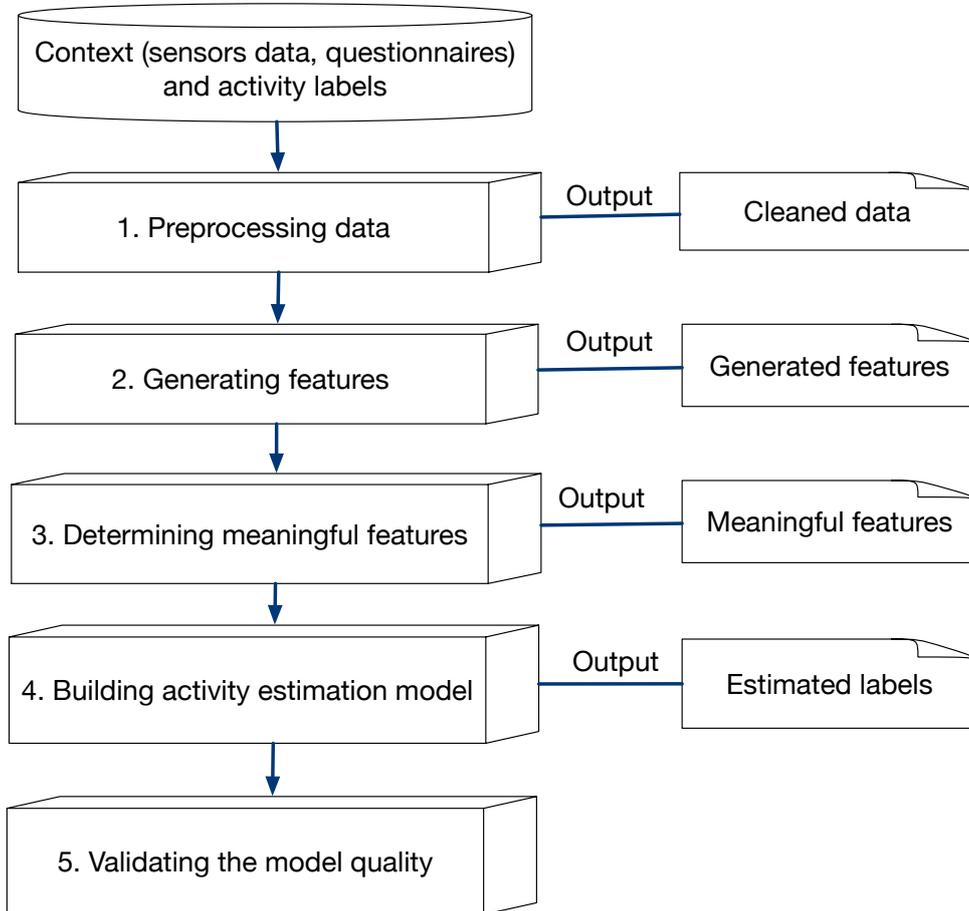


Figure 5.3 – The general methodology to estimate contextualized occupants activities

1. **Preprocessing sensors data:** the raw data of sensors could be not clean, not in the exact time resolution because of the different sensor's settings and could face signal and network issues. Thus, it is essential to clean and normalize the data.
2. **Generating features:** many features can be hidden from sensors data. Constructing them is useful to have more information on occupant activities.
3. **Determining meaningful features:** among many constructed features, some of them are not relevant depending on the considered estimation. Selecting meaningful features and removing others is essential. Information gain is a criteria for evaluating the contribution of one variable to another. In this work, this criteria is used to determine which features are most relevant to occupant activities.
4. **Building stochastic activity model:** a model needs to be proposed to

deduce occupants' activities from selected features. As discussed in chapter 2, Bayesian Network is a probabilistic graphical model that provides understandable parameters and outputs. It satisfies the requirements of the expected model. Thus, BN is used to estimate occupant activity in this work.

5. **Validating the model:** The validation is a contribution of the model. This step is to evaluate the model quality by comparing estimated activities to the observed labels.

These steps are sequentially presented in the next sections.

5.2.2 Preprocessing sensors data

Removing abnormal data

Despite using reliable sensors, some abnormal data points, which are significantly different from others, could be observed. Removing these points is essential to ensure the reliability and quality of data. Many reasons can cause abnormal values. In literature, Ayadi et al. (2017) presented three sources leading to the anomalies: noises or errors, unexpected events and malicious attacks. Noises or errors come from sensors' faults. They often last for a short period and they may occur frequently (Bettencourt et al., 2007). Unexpected events are often caused by changes in the real world (fire detection, earthquake, *etc.*) and they often last for a long period (Rajasegarar et al., 2006). Malicious attacks are related to network security and abnormal data points can be injected or modified by external factors (Baig, 2011). To detect abnormal values of sensors, many approaches have been proposed such as supervised learning, unsupervised learning or statistical analysis (Hill et Minsker, 2010; Ayadi et al., 2017).

Detecting anomalies of sensors data is a challenging task and it needs a lot of efforts. The thesis does not focus on this problem and we just consider noises in sensors data. To avoid the other types of anomaly, we remove the period of data that contains anomalies for a long period. This task is performed manually with supports of visualisation tools. To detect abnormal data points, we identify whether they exceed valid physical ranges, which are different from sensors. For example, energy consumption and the CO₂ concentration must not be negative. In addition, the statistical method proposed by Amayri (2017) is applied to determine whether a data point is significantly different

from neighborhoods. Although these methods are limited to many types of anomalies, they are selected because they are easy to implement to both discrete and continuous data and have low computation costs.

Filling missing values

Missing data points frequently exist in event-triggered sensors (contact sensors, PIR sensors, *etc.*), in which no data points are recorded if no change is detected. In addition, due to the removing anomalies step, several data points would be missing. Filling these missing points is essential to complete the dataset. Many methods can be applied to tackle this task. In this work, the filling methods are selected depending on types of sensors:

- **with contact sensors**, which detect discrete states (open/close) of objects (doors, windows), missing points represent that the objects do not change their states. Therefore, these points are filled by the latest recorded state of the objects;
- **with PIR sensors**, data points are recorded in each period of time or if a movement is detected. Thus, missing data points of these sensors are set by zero (0) values. This method is also applied to power sensors, which work similarly to PIR sensors;
- **with ambiance sensors** (air temperature, air relative humidity, luminosity, CO₂ concentration), which capture continuous values, missing values should be interpolated from other values. As mentioned above, we only consider short periods of missing data. Linear interpolation is a popular method to deduce time series missing values depending on neighbor values. This method is used in this work because it can work efficiently in short periods (less than 1 hour) of missing values (Lepot et al., 2017; Picornell et al., 2021; Demirhan et Renwick, 2018).

In summary, to clean the raw dataset, two tasks are applied: filtering (removing) invalid data points and filling missing ones (see in figure 5.4).



Figure 5.4 – Preprocessing data step

5.2.3 Generating features

To estimate occupants' activities, the BN needs to cross information from the context factors with labelled activities in the learning phase. Although sensors measurements contain some context information directly, other features are not directly available from measurements data and should be derived from it. Generating these additional features is a crucial step to estimate occupant activities. For example, a power sensor is installed to deduce the energy consumption of a television every 10 minutes. To detect the activity of watching television activity every 30 minutes, the energy consumed by the television of the three corresponding 10 minutes timesteps should be extracted then summed up. Another example, indoor temperature and outdoor temperature can be used to extract the gap between indoor and outdoor temperature.

To determine which features are to be generated, the knowledge of the local context needs to be considered. As mentioned in chapter 3, we consider six main elements for a context: the occupant, the time, the location, the environment, the objects and the activity-location. Below, we discuss the generation of features based on this decomposition.

Time

Occupants' activities can be affected by the timestep of the day or the type of day (weekday, weekend). The influence often comes from occupants' calendars. For example, occupants can cook breakfast on weekdays earlier than on weekends because they need to get up earlier on weekdays. Many scientists have been studied the time when the activity happens (Page et al., 2008; Widén et Wäckelgård, 2010; Wilke et al., 2013; Aerts, 2015). They stated that occupants' activities could be affected by the timestep of the day, the day of the week or the season of the year (Wilke et al., 2013; Vorger, 2014; Aerts, 2015; Amayri et al., 2017). Depending on these studies, we use the timestep and the type of day (weekend or weekday) as features representing the time element for a context. The season of the year is not used because the considered dataset in this work does not cover many seasons.

Location and located activity

There are many locations in a house. As discussed in chapter 3, an activity is associated with a specific location. When occupants perform an activity, they can affect surrounding context factors. These factors can be in many locations but considering all of them is a challenging task, especially when there are many activities and many locations. To simplify this problem, when studying an activity, we only consider features in the activity's location, which can be given by the located activity element.

Environment

Occupants' activities can affect indoor environment conditions through actions (as discussed in chapter 3). Gathering the environment features can contribute useful information to deduce the actions or the activities. For example, CO₂ concentration and luminosity can provide information related to human presences (Widén et al., 2012; Amayri et al., 2017; Wolf et al., 2019). Indoor air temperature and relative humidity are useful to investigate actions on doors/windows (Fabi et al., 2013; Chen et Pollino, 2012). Therefore, features related to environmental conditions should be generated. As discussed in chapter 4, to consider environmental conditions in a room, we study the CO₂ concentration, the air temperature, the air relative humidity and the luminosity in the room. For each variable in a timestep, many features can be generated such as average value, maximum value, minimum value, standard deviation, difference, *etc.* Each of these features can contribute different information. Because this work aims to study activities in each timestep, we focus on some features that cover the information of a variable over a timestep. Therefore, in each timestep, averaged temperature, averaged CO₂ concentration, averaged relative humidity and averaged luminosity are taken into account. In addition, it is better to investigate differences of variables between timesteps, which can be caused by the impacts of activities. Thus, in each timestep $t \in N^*$, additional compound features are considered: temperature trend defined as $\Delta T_k = T_t - T_{t-1}$, CO₂ concentration trend defined as $\Delta CO_{2,t} = CO_{2,t} - CO_{2,t-1}$, luminosity trend defined as $\Delta L_t = L_t - L_{t-1}$, relative humidity trend defined as $\Delta RH_t = RH_t - RH_{t-1}$.

Occupants

It is difficult to detect occupancy (presence, absence) directly. Cameras could be used but they are limited in many situations. To deal with it, many scientists used motion sensors to get information about occupants through movements detection (Amayri et al., 2017; Elbayoudi et al., 2019; Akbari et Haghghat, 2021). As discussed in chapter 4, PIR sensors were installed in this work to detect whether there is any movement in a room. This information can be used to deduce the number of motions or the duration of motions detected in each timestep. In this work, we extract the duration of motions detected because it is easier to explain than another.

Objects

Features of objects can describe the actions that occupants do during their activities. For example, states of doors/windows present whether these doors/windows are opened, the electricity consumption of an appliance can show if occupants use this appliance. Therefore, the energy consumption of electric appliances and the states of doors/windows for each timestep should be generated. This information is useful to consider actions of occupants on objects and then detect occupants' activities. Here, the settings of objects are collected through questionnaires. They include information on available domestic appliances, doors, and windows in each location. This knowledge is useful to determine which objects should be investigated because we can not investigate an object that does not exist. For each object, many features can be extracted. For example, power sensors can deduce the electricity consumption of appliances every time they are used. This information can be used to estimate the maximum value, the minimum value or the standard deviation of the electricity consumption of the appliance during a timestep. In this work, for each appliance, its total energy consumption is extracted for each timestep. This feature is selected because uses of the appliance are considered over the timestep. It is important because an appliance can be used many times. For doors and windows, features such as the number of doors/windows opening or the total duration (minutes), in which doors/windows are opened, can be generated. Similar to the features related to occupants' movements, the total duration is selected in this work because it is easier to explain.

Summary

In brief, from the cleaned sensors dataset, features are generated in each timestep to estimate an activity in its location as summarized in table 5.1.

Table 5.1 – Summary of the generated features each activity in each location

Generated feature(s)	Sensor(s)	Context element
Averaged temperature and trend temperature ΔT	Ambience	Environment
Averaged CO ₂ and trend CO ₂ concentration ΔCO_2	Ambience	Environment
Averaged luminosity and trend luminosity ΔL	Ambience	Environment
Averaged relative humidity and trend relative humidity ΔRH	Ambience	Environment
Timestep, type of day (weekday, weekend)	All sensors	Time
Total minutes that movements are detected	PIR motion	Occupant
Total minutes that doors/windows open	Contact	Objects
Total energy consumption of domestic appliances	Power sensors	Objects

The considered context elements in a house at each time step t is represented as a set of generated features $X_t = \{x_{1t}, x_{2t}, \dots, x_{nt}\}$. The choice of duration of the time step is highly context-dependent. In this work, because 30-minutes timesteps were selected to collect activity labels (discussed in chapter 4), they are used to extract features and build models.

5.2.4 Determining meaningful features

From the previous step, a large set of features is generated. However, some features can be meaningful to estimate an activity but they do not provide meaningful information for other activities. For example, the air relative humidity is not useful in estimating the studying activity, the air temperature is not useful in detecting the watching television activity. Removing them from a model can help reduce noises in the model and save the budget of sensors installation. This step addresses the problem "Determining the set of meaningful features in occupant activity estimation".

In literature, there are three approaches of features selection: filtered-based wrapper-based and embedded (Khaire et Dhanalakshmi, 2019; Szidónia et László, 2017). The wrapped-based methods build multiple models with many subsets of features and select the set of features that results in the best performance model. Filtered-based methods use statistical techniques to evaluate the relationship between each feature and the target variable. These evaluations are then compared to determine which features are selected. In the embedded approach, optimal features subsets are determined during the building of the model. The feature search process is embedded in the model so that the learning process and the feature selection can not be separated (Kashef et al., 2018). The embedded approach is only suitable for some specific models such as Decision Tree, Random Forest and it is not adapted to BN. Here, the objective is not only to improve the model performance but also to assess the relationship between features and occupants' activities. Thus, the filtered-based approach is selected for the meaningful features determination. *Information Gain (IG)* is a popular criteria in the filtered-based approach evaluating how much information a feature provides regarding a variable (Szidónia et László, 2017). This criteria is useful to determine meaningful features in building an estimation model (Amayri, 2017; Odhiambo Omuya et al., 2021).

To compute the information gain, it is necessary to discretize features which contain continuous values. Additionally, the inputs of a Bayesian Network must be categorical values. Thus, a discretization step needs to be applied to discrete features comprising continuous values. In brief, we need two steps to select meaningful features: *discretization of features* and *information gain estimation*.

Discretization of features

The objective of discretization is to determine cutting points to partition the range of a feature into a small number of intervals. Some automated algorithms could be used to estimate the optimal splitting points: unsupervised ones such as k-means clustering (Wang et Song, 2011) or supervised ones such as the Class-Attribute Interdependence Maximization (CAIM) (Kurgan et Cios, 2004) or the Minimum Description Length Principle (MDLP) (Fayyad et Irani, 1993). However, these approaches need a predefined number of ranges and most of them are based on predefined criteria and they do not

concern the context and the nature of features. To tackle it, some scientists use their knowledge on features to define more proper numbers of ranges and cutting points manually (Amayri, 2017; Tijani et al., 2015).

While the best numbers of ranges are not known, we discrete environment-related features manually in this work. Their splitting points are manually defined by the knowledge of their nature. Table 5.2 shows some examples of cutting points for these features, obtained by manual discretization. Similarly, features on doors/windows opening and motions are also discretized manually. Table 5.3 shows examples of manual discretization for these features.

Table 5.2 – Manual thresholds of air temperature, relative humidity, luminosity and CO₂ concentration

Feature	Manual ranges
trend temperature (°C)	$(-\infty, -2)$, $[-2, -1)$, $[-1, -0.5)$, $[-0.5, 0.5)$, $[0.5, 1)$, $[1, 2)$, $[2, +\infty)$
trend humidity (%)	$(-\infty, -11)$, $[-10, -6)$, $[-5, -3)$, $[-2, 2)$, $[3, 5)$, $[6, 10)$, $[11, +\infty)$
luminosity (lux)	$[0, 300]$, $[301, 500]$, $[501, 750]$, $[751, 1000]$, $[1000, +\infty)$
CO ₂ concentration (ppm)	$[0, 400]$, $[401, 700]$, $[701, 1000]$, $[1001, 2000]$, $[2001, +\infty)$

Table 5.3 – Manual thresholds of features related to occupants' movements and doors/windows openings

Feature	Manual ranges
Number minutes that motions have been detected (minutes)	$[0, 5]$, $[6, 15]$, $[16, 30]$
Number minutes that doors/windows have been opened (minutes)	$[0, 5]$, $[6, 15]$, $[16, 30]$

Contrary to the above features, appliances have different power ranges, so it is difficult to define their cutting points based on knowledge. As discussed above, unsupervised algorithms can be used to discretize features automatically. Many unsupervised algorithms can be applied such as k-means, Equal-Width or Equal-Frequency. In this work, k-means clustering is used because it is easy to

implement and is flexible with different ranges of data. To apply k-means, the number of clusters needs to be provided. However, it is difficult to define the optimal number of clusters for all appliances. In this work, for the sake of simplicity, we assume that an appliance can have five levels (intuitively selected) of energy consumption during each timestep into five levels: very low, low, medium, high and very high electricity consumption. Figure 5.5 presents an example of the five levels for a microwave energy consumption with a 30-minute step determined by k-means.

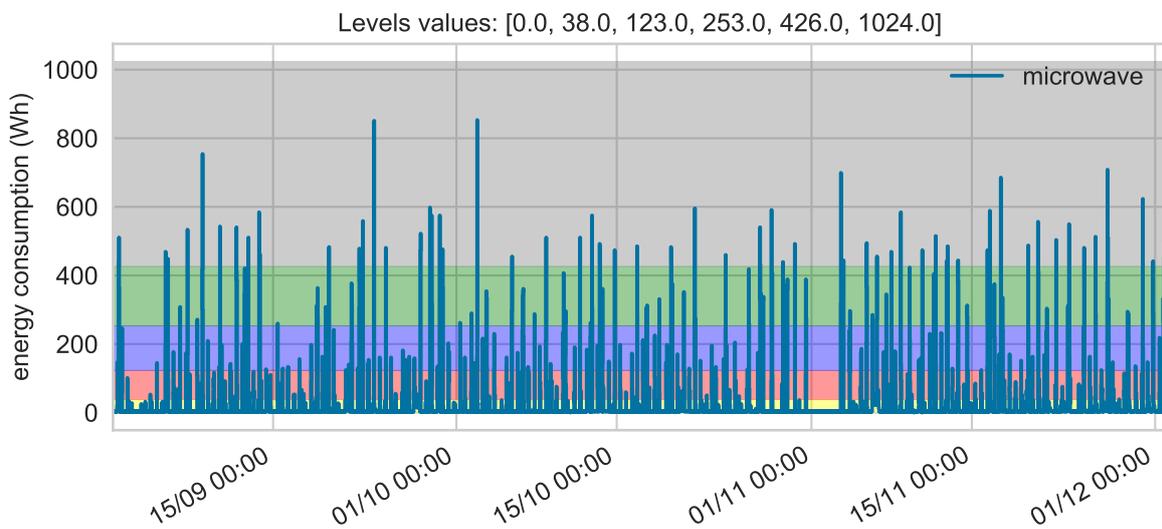


Figure 5.5 – Examples of the five levels (very low - yellow, pink - low, violet - medium, green - high, gray -very high) of a microwave determined by k-means in the Cestas site

Information Gain estimation for meaningful features selection

Information Gain is based on the notion of *entropy* (ENT) in the information theory. Entropy measures the disorder of the data categorized by a target (see equation 5.2). The higher the entropy is, the higher the disorder associated with this target is.

$$ENT(y) = - \sum_{i=0}^{n-1} -p(y = Y_i) \times \log p(y = Y_i) \quad (5.2)$$

where:

- $n \in N^*$: the number of category of the target y ;
- y : a discrete target with a value domain defined as $\text{dom}(y) = \{Y_0, \dots, Y_{n-1}\}$;

- $ENT(y)$: the entropy of the target y ;
- $p(y = Y_i)$: the probability for y to be equal to the value Y_i .

Information Gain (IG) estimates the reduction of labels' impurity for each feature, the higher information gain is, the more meaningful feature is. The IG between a feature x and an activity y is defined in equation 5.3:

$$IG(x, y) = ENT(y) - ENT(y|x) \quad (5.3)$$

with $ENT(y|x)$: the conditional entropy of y given x .

When the IGs of features are determined, they are used to sort the features. Then the most meaningful features with the highest IGs are selected.

An issue that needs to be addressed here is the number of selected features. There is no optimal number for all activities and it is context-related. In fact, the number of features could be detected based on the predefined criteria (threshold of information gain) or manually selected by using the knowledge. In addition, the Silhouette method (Rousseeuw, 1987; Subbalakshmi et al., 2015) and the Elbow method (Clayman et al., 2020; Congming et al., 2021) could be used to analyze the relationship between the number of features and the model performance. These analyses are useful to determine the optimal number.

5.2.5 Building activity estimation model

A model must be built to deduce occupant activities from selected meaningful features. As mentioned above, Bayesian Network is chosen to estimate occupants' activities from selected meaningful features.

5.2.5.1 Bayesian Network

Bayesian Network (BN) is a graphical model to present relationships among variables. The knowledge of BN is interpreted through two main elements: a network structure and parameters, which contain probability distributions (see figure 5.6). The network structure includes nodes representing features and edges, which express the cause-effect relationships between these features. The probability distributions comprise the prior probabilities $p(X_i)$ of nodes X_i and conditional probability distributions $p(X_j|X_i)$ between node X_i and node X_j .

In BN, the data is fixed and the probability distributions are used to describe the uncertainty in both model parameters and outputs. In formula, the Bayesian rule is described in equation 5.4.

$$P(\theta|y) = \frac{p(y|\theta) \times p(\theta)}{p(y)} \quad (5.4)$$

where:

- $P(\theta|y)$: the conditional probability of θ given y ;
- $p(y|\theta)$: the conditional probability of y given θ ;
- $p(\theta)$: the prior probability.

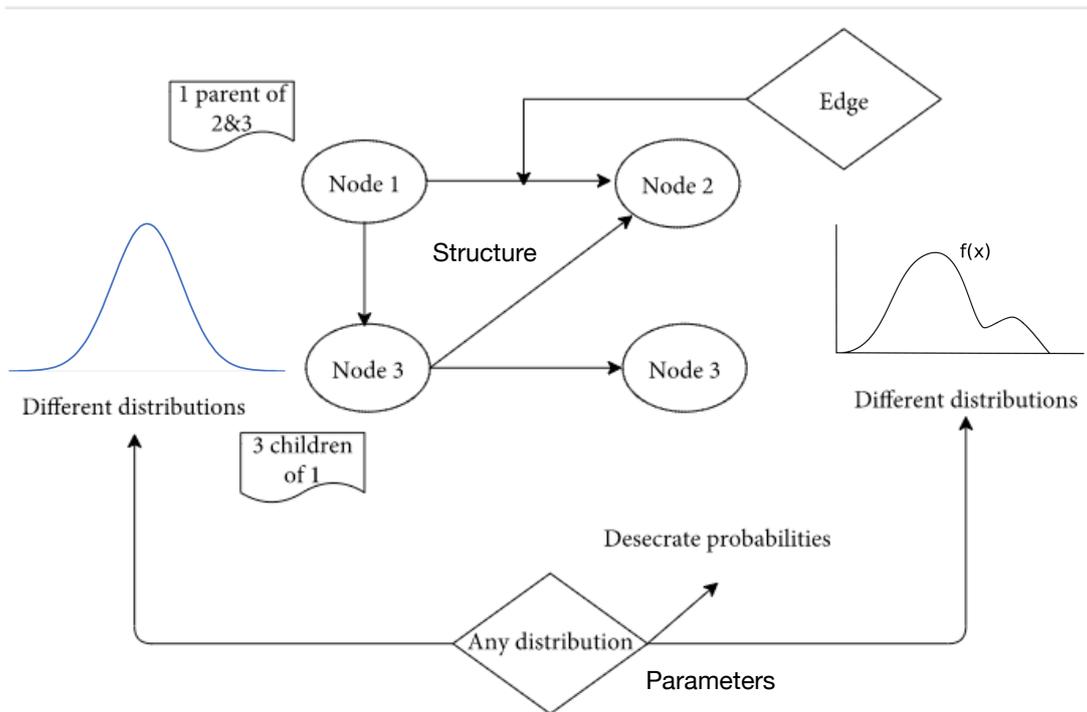


Figure 5.6 – General elements of a Bayesian Network: structure and parameters (Amayri, 2017)

Bayesian Network is a stochastic model that can estimate the probability of an activity with given features. Additionally, it could be used as a supervised classifier (Bayesian Network classifier), which determines the activity by optimizing the joint probabilistic presented in equation 5.5.

$$p(x_0, x_1, \dots, x_n) = \prod_{i=0}^n p(x_i | Pa(x_i)) \quad (5.5)$$

where

- $n \in N^*$: the number of nodes;

- $p(x_0, x_1, \dots, x_n)$: joint probability of nodes in the network;
- x_i : node i in the network;
- $Pa(x_i)$: parents of the node x_i .

The learning of the Bayesian Network can be divided into two parts: structure learning and parameters learning. In the literature, there are three main structure learning approaches: constraint-based, score-based and hybrid ones (Scutari et al., 2019; Sun et al., 2021). The constraint-based approach uses conditional independence tests to learn structure. The score-based approach learns the structure by optimizing a predefined criterion for each candidate (Asvatourian et al., 2020). The hybrid approach combines both constraint-based and score-based approaches (Caravagna et Ramazzotti, 2021). To determine the parameters, there are many approaches such as MAP-based (Guo et al., 2017), maximum likelihood (Redner et Walker, 1984), or minimax algorithm (Gao et al., 2019). However, both structure learning approaches and parameters learning approaches require a large amount of data, which is a challenge in many case studies. To deal with this issue, some scientists used expert knowledge to define the structure, especially when the number of features is small (Tian et al., 2018; Hawarah et al., 2010; Amayri et al., 2017). In this work, due to the limited quantity of activities labels, the structure of BN is defined by expert knowledge and the parameters are deduced directly from the dataset.

5.2.5.2 Consequences-based Bayesian Network for activity estimation

In houses, inhabitants tend to repeat activities to satisfy their needs (food, water, entertainment, *etc.*). These activities become habits and form daily activities of occupants with a timetable reasonably regular (Kashif, 2014). For example, an occupant goes to school from 9:00 to 17:00. In the morning, he often wakes up at around 7:00 to do personal care and prepares breakfast between 7:30 and 8:30. In the afternoon, he comes home at around 17:30. He wants to sleep before 23:00 so that he does other activities (taking a shower, washing clothes, cooking dinner, studying, entertaining, *etc.*) between 17:30 and 23:00. In many studies, occupants' activities are also stated to be associated with their habits, which are represented by the time in the day, the day in the week, *etc.* (Tanimoto et al., 2008; Hong et al., 2015).

To do activities, occupants can do actions and interact with household objects

(appliances, doors, windows, *etc.*). These behaviors cause effects on environmental factors (air temperature, relative humidity, *etc.*) and energy consumption (see figure 5.1). In particular, occupants' presence can increase CO₂ concentration in a room, their operations of domestic appliances consume electricity and their actions on doors/windows could cause variations of temperature or luminosity levels, *etc.* (Kashif et al., 2013; Pal et al., 2019). For example, in the evening, occupants want to cook dinner so that they go to the kitchen and turn on the light. CO₂ concentration and luminosity in the kitchen increase as consequences. Then, they use the microwave to heat foods and the kettle to boil water. These actions cause electricity consumption. Other example: to study, an occupant goes to his bedroom. To be quiet and have privacy, he closes doors and windows. Then, he turns on the lamp and the computer to do his homework. Obviously, his actions lead to increased luminosity, CO₂ concentration, electricity consumption and a change of the door's state (open→close).

Based on the above lots of examples, occupants' activities and consequences can be described as follows: Occupants have some needs in usual times in the house. When the regular times come, occupants do one or many activities to satisfy their needs. To do an activity, they can do actions on household objects, which can affect some environmental factors and consume energy. This causal relationship is represented by a conceptual model in figure 5.7.

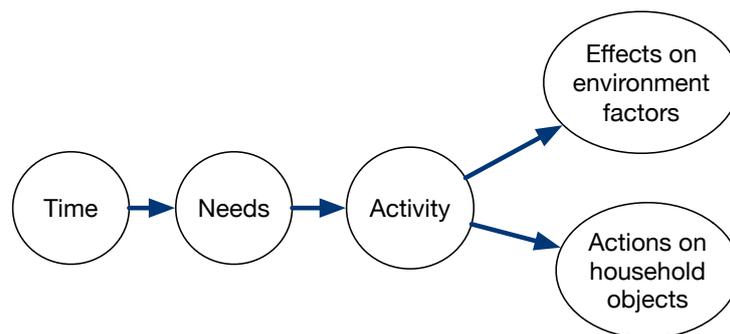


Figure 5.7 – Conceptual model of occupants' activities and their consequences

In this work, the objective is to deduce occupants' activities from the context factors represented by features. As mentioned in table 5.1, features are divided into four groups: environment (E), time (T), occupant (O) and objects (C). Linking these groups to the above conceptual model, a Consequences-Based Bayesian Network is proposed in figure 5.8.

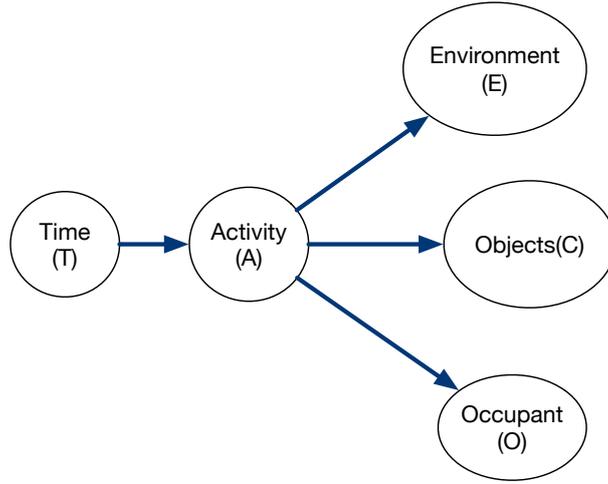


Figure 5.8 – The structure of the proposed Consequences-based Bayesian Network (CBBN)

The joint probability of the network is described in equation 5.6:

$$p(A, E, C, T, O) = p(A|T) \times p(E|A) \times p(C|A) \times p(O|A) \quad (5.6)$$

The conditional probability that an activity A is performed given selected features is described in equation 5.7:

$$p(A|E, C, T, O) = \frac{p(A|T) \times p(E|A) \times p(C|A) \times p(O|A)}{p(T) \times p(E) \times p(C) \times p(O)} \quad (5.7)$$

In each time step of a day, the activity state a^* can be determined by the Equation 5.8.

$$a^* = \underset{a \in \{0,1\}}{\operatorname{argmax}} p(a|T) \times p(E|a) \times p(C|a) \times p(O|a) \quad (5.8)$$

In summary, to build a Bayesian Network for activity estimation, we:

- Divided selected features into four groups: time, environment, occupant and objects.
- Built a consequences-based Bayesian Network with the above groups of features and the defined structure.

5.2.6 Validating the model

When a model is built, validation is an essential part of evaluating the model performance. Here, to validate the proposed model, we verify whether the model can accurately estimate occupant activities.

To do it, the dataset was divided into two parts: training and testing. The model was built with the training data and it is tested with the testing data. In a house, for an

activity, occupants can spend less time doing the activity than not doing it. For example, in a day, occupants only spend few hours cooking and they do not do it in the rest of the day. It can lead to an issue of imbalanced data: distributions between labels of activity are uneven. To tackle this issue, we use F1-score to verify the accuracy of the model. It uses both precision and recall, which are based on the confusion matrix. The value of the F1-score is from 0 to 1. The higher F1-score is, the better model is. The formula of the F1-score is described in equation 5.9.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (5.9)$$

where:

- precision: the number of predicted positive labels that actually belong to the positive class;
- recall: the percentage of positive labels that are predicted correctly.

Besides accurate estimation, it is important to verify if the model can predict occupant activities, which is useful in uncertainty analysis and to construct activity profiles. Thus, we compared statistically the predicted activity profile with the observed activity profile. The model has good performance if the trends of these activity profiles are similar.

To assess the model generalization, the methodology was implemented and validated in two different case studies in France: Cestas and Bordeaux.

5.3 Results: the cooking activity

In houses, occupants cook regularly to satisfy their physical needs of food, which is a basic need to survive in life. During the cooking activity, occupants do many actions and they highly affect both indoor environments (Hager et Morawicki, 2013; Zenissa et al., 2020) and energy consumption (Kashif, 2014; Alves et al., 2021). In 2018, about 6.1% of final energy consumption in Europe was consumed by cooking activity in residential buildings (Eurostat, 2018). Thus, estimating cooking activity is useful in assessing energy consumption in houses. In this section, the estimation of cooking activity is presented to illustrate the proposed methodology.

Occupants can cook in different ways because of their various needs in periods of

the day. For example, households can have individuals studying at school from 09:00 to 17:00. In the morning, they do not spend much time cooking because they need to leave home early. In the afternoon, they could eat at school. In the evening, they often spend more time cooking. These differences could cause different effects on building energy (electricity consumption, indoor environment, *etc.*). Considering this issue, we considered three kinds of the cooking activity: cooking breakfast, cooking lunch and cooking dinner are considered corresponding to three periods of the day:

1. cooking breakfast: in the morning, from 6:00 a.m to 10:59 a.m;
2. cooking lunch: in the afternoon, from 11:00 a.m to 17:59 p.m;
3. cooking dinner: in the evening, from 18:00 p.m to 23:00 p.m.

5.3.1 Case study 1: Cestas

A dataset of the Cestas household from 01/09/2019 to 30/12/2019 is used to implement and validate the methodology. To estimate cooking activity, it is essential to determine the relationship with the context. Firstly, the location of this activity needs to be known to generate useful features. As mentioned in the methodology, this information is extracted from activity-location, a context element. Figure 5.4 shows an example of activity-location information collected from the Cestas household through questionnaires. It can be seen that in this case study, cooking activity happens in the kitchen.

Table 5.4 – An example of the activity-location information collected from the Cestas household through questionnaires

Activity	Location (s)
Entertaining	Bedroom, Living room
Cooking	Kitchen
Doing personal care	Bathroom
Washing dishes	Kitchen

Thus, the contextual factors in the kitchen must be used to extract relevant features. In general, there are four groups of features corresponding to four context elements: environment, time, occupants and objects (see table 5.1). However, the objects such

as domestic appliances, windows, doors in houses are different. Many scientists investigated that households could have different appliances ownership, which is affected by many characteristics (building type, income, professional, *etc.*) (Wilke et al., 2013; Vorger, 2014; Aerts et al., 2014; Zaraket, 2014). Therefore, in this case study, knowledge is used to determine the list of objects in the kitchen:

- domestic appliances: microwave, toaster, coffee machine, fridge, dishwasher, food processor, robot for cooking and yogurt maker;
- one main door and one window.

Power sensors have been installed to capture the electricity consumption of these appliances (see chapter 4). The information related to CO₂ was not measured because of a technical issue. After cleaning sensors data, features are generated every 30-minutes:

- | | |
|---|---|
| <ul style="list-style-type: none"> • air temperature and air temperature trend; • air relative humidity and air relative humidity trend; • luminosity and luminosity trend; • number motions have been detected; • number minutes the main door has been opened; | <ul style="list-style-type: none"> • number minutes windows have been opened; • timestep, type of day; • energy consumption of microwave, toaster, coffee machine, fridge, dishwasher, food processor, robot for cooking and yogurt maker. |
|---|---|

To be simpler, the features related to doors, windows and domestic appliances are denoted by their names. For example, "*microwave*" denotes the consumption of the microwave, "*door*" denotes the duration that it has been opened. Within these generated features, many features do not provide meaningful information to the estimation of cooking activity. As mentioned in the methodology, Information Gain is used to determine the importance of features based on information theory. To do it, features need to be discretized. As discussed in section 5.2, manually cutting points are defined for the features related to the environment, occupant, doors and windows. An unsupervised learning algorithm (k-means) is used to determine five electricity consumption levels of domestic appliances.

Figure 5.9 presents the estimated information gain between the features and three concerned activities (cooking breakfast, cooking lunch and cooking dinner). It is shown

that the cooking activity is generally related to the time of day, the energy consumption of linked appliances and the movement information (motions), which represents the presence of occupants. The environment features (temperature, luminosity, *etc.*) and the actions on doors/windows do not provide much information for the cooking activity estimation. In addition, consumption of some appliances (food processor, cooking robot, yogurt maker) are not the subject of many contributions. It can be explained by the fact that these appliances are not used regularly. Therefore, using them to estimate the cooking activity is not helpful in this case study.

Cooking breakfast		Cooking lunch		Cooking dinner	
Feature	IG	Feature	IG	Feature	IG
Timestep	0.23	Microwave	0.37	Microwave	0.37
Microwave	0.22	Timestep	0.23	Timestep	0.27
Toaster	0.20	Fridge	0.15	Motion	0.25
Coffee machine	0.15	Motion	0.10	Toaster	0.04
Motion	0.14	Temperature trend	0.04	Fridge	0.04
Luminosity	0.02	Luminosity	0.03	Luminosity	0.03
Humidity trend	0.02	Toaster	0.03	Temperature trend, humidity trend, dish washer, door, windows, food processor, robot, yogurt maker, temperature, humidity, luminosity trend, weekday, coffee machine	0.00
Fridge	0.01	Humidity trend	0.03		
Temperature trend	0.01	Coffee machine	0.02		
Dish washer, door, windows, food processor, robot, yogurt maker, temperature, humidity, luminosity trend, weekday	0.00	Dish washer, door, windows, food processor, robot, yogurt maker, temperature, humidity, luminosity trend, weekday	0.00		

Figure 5.9 – Sorted Information Gain between features and three activities: cooking breakfast, cooking lunch and cooking dinner in the Cestas household

Additionally, the results show differences in cooking activities between periods of the day. In particular, occupants usually use microwave, toaster and coffee machine for cooking breakfast. To prepare the lunch, the microwave and the fridge are used frequently. In the evening, the microwave is generally used alone for cooking. These signatures provide insights into how occupants cook in the house.

When the Information Gain of features was determined, the most meaningful features, which have the highest IG, were used to build a Bayesian Network to estimate the cooking activity. As discussed in the methodology section, it is challenging to determine the number of features to select in building activity estimation models. In this study, for the sake of simplicity, the number of meaningful features is chosen intuitively to build the estimation model:

- Cooking breakfast: Timestep, microwave, toaster, coffee machine and motion;
- Cooking lunch: Timestep, microwave, fridge and number of motions;
- Cooking dinner: Microwave, timestep and number of motions.

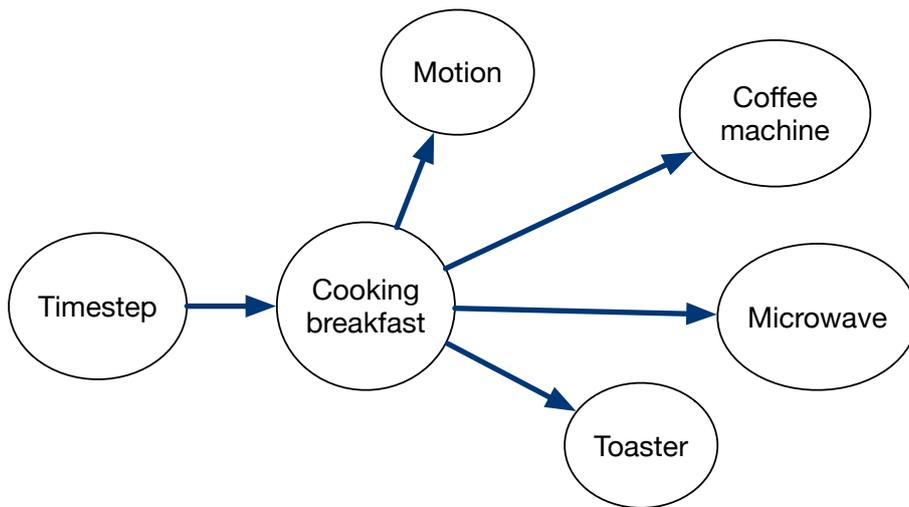
Selected features are then divided into four groups (environment, time, occupant, objects) described in table 5.5.

Table 5.5 – Groups of selected features for estimating cooking breakfast, cooking lunch and cooking dinner in the Cestas household

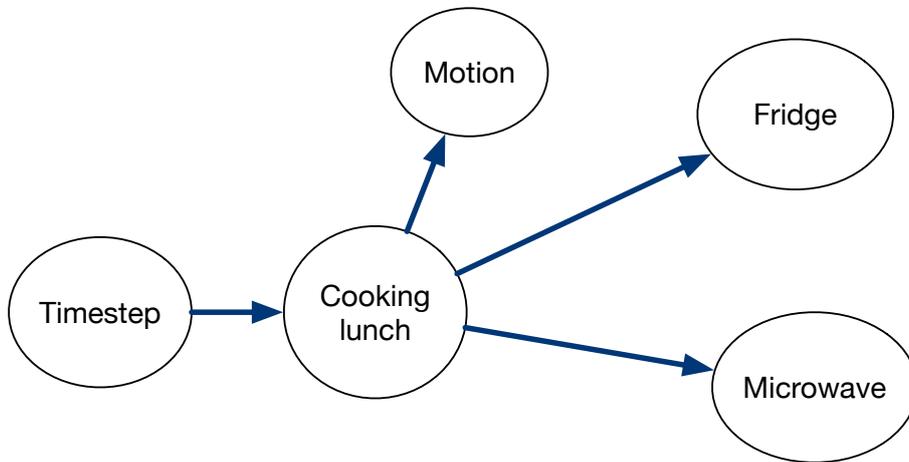
	Cooking breakfast	Cooking lunch	Cooking dinner
Environment	None	None	None
Time	Timestep	Timestep	Timestep
Occupant	Motion	Motion	Motion
Objects	Microwave, toaster, coffee machine	Microwave, fridge	Microwave

Corresponding to these groups of features, the consequences-based Bayesian Networks for the studied three kinds of cooking (cooking breakfast, cooking lunch and cooking dinner) are respectively shown in the part A, B and C of figure 5.10 (A:breakfast, B:lunch and C:dinner). It is shown that their structures are different. A feature can be meaningful for the estimation of a kind of cooking but not meaningful in estimating others. For example, in this case, the coffee machine feature is essential to estimate the cooking breakfast but it is not relevant to estimate the cooking lunch and the cooking dinner.

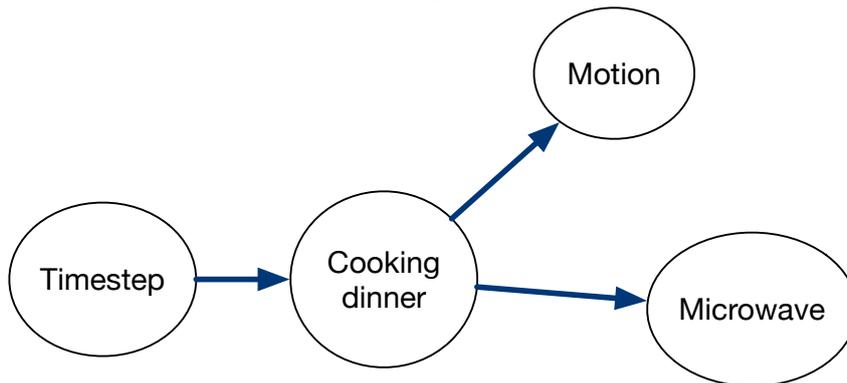
Parameters of the BN, which include prior probability distributions and conditional probability distributions, are deduced directly from the observed data. Table 5.6 presents examples of estimated conditional probability tables for two activities: cooking breakfast and cooking lunch. The results show that the microwave consumes more electricity in cooking breakfast than in cooking lunch. In the morning, 24% of the time the microwave consumes very low electricity during the cooking breakfast activity, 62% of the time it consumes from low to medium electricity. To cook lunch, 50% of the time the microwave consumes very low electricity, 33% of the time it consumes from medium to very high electricity.



(A) The cooking breakfast activity



(B) The cooking lunch activity



(C) The cooking dinner activity

Figure 5.10 – Consequences-Based Bayesian Networks for the cooking activities in Cestas site

Table 5.6 – Examples of estimated conditional probability distributions of the microwave in cooking breakfast and cooking lunch in the Cestas household

	Cooking breakfast (No)	Cooking breakfast (Yes)	Cooking lunch (No)	Cooking lunch (Yes)
Microwave (very low)	0.99	0.24	0.99	0.50
Microwave (low)	0.00	0.38	0.01	0.19
Microwave (medium)	0.00	0.24	0.00	0.18
Microwave (high)	0.00	0.11	0.00	0.09
Microwave (very high)	0.01	0.03	0.00	0.06

To validate the built Bayesian Network, F1-score is computed with 3-folds cross-validation, in which the dataset is divided into three parts, two parts are for training and one part is for testing. Additionally, we compare the performance of the Consequences-Based Bayesian Network (CBBN) to:

- Basic supervised learning models: Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression and Linear Regression.
- Bayesian Network with structure learned directly from the data with basic algorithms: Chow-Liu, Hill-Climbing, Greedy and Exact.

Table 5.7 presents the comparisons of F1-score among the approaches. The result shows that the proposed approach and the Random Forest are generally better than other approaches in this case study. In contrast, Random Forest does not meet the mentioned requirements: being human-understandable and having stochastic outputs. Consequences-Based Bayesian Network with a high F1-score (0.87 for cooking breakfast, 0.86 for cooking lunch, 0.86 for cooking dinner) verifies that the proposed model accurately estimates this activity in this case study.

Table 5.7 – Comparison of F1-score among different approaches for the cooking estimation in the Cestas household

	Cooking breakfast	Cooking lunch	Cooking dinner
CBBN	0.87	0.86	0.86
Decision Tree	0.80	0.81	0.81
Random Forest	0.81	0.85	0.88
Logistic regression	0.70	0.60	0.60
SVM	0.70	0.60	0.63
BN with Chow-Liu	0.73	0.60	0.70
BN with Greedy	0.80	0.75	0.73
BN with Hill-Climbing	0.75	0.70	0.70
BN with Exact	0.80	0.74	0.75
BN with Exact	0.80	0.74	0.75

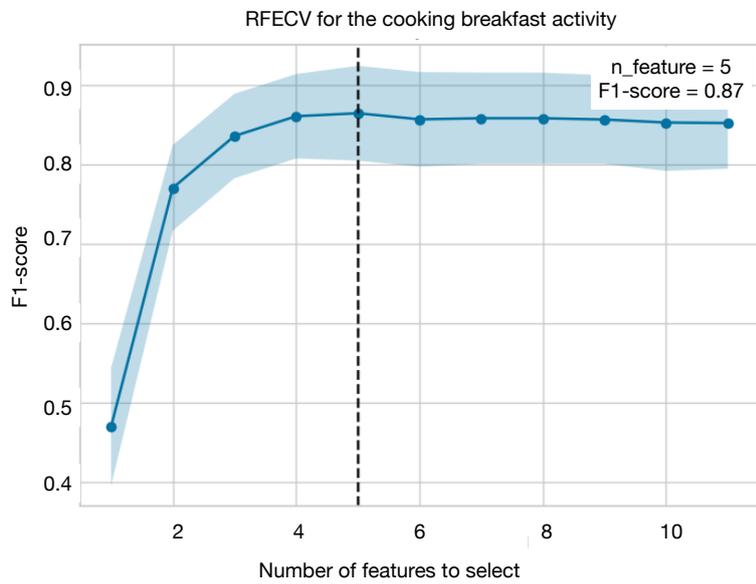
As mentioned in the previous section, it is difficult to determine the number of features in estimation models. To tackle this issue, recursive feature elimination with cross-validation (RFECV) could be used to investigate the relationships between the number of features and the model quality (Misra et Singh, 2020). In this algorithm, the weakest features are eliminated step-by-step. Cross-validation is applied to determine F1-score with every set of features. The parts A, B and C of figure 5.11 present respectively the implementation of this algorithm on three activities: cooking breakfast, cooking lunch and cooking dinner. It is shown that many features do not contribute to the activity estimation and removing them does not affect the model quality. In particular, six features with the smallest IG do not help to improve the model quality to estimate cooking breakfast activity. For the cooking dinner estimation, adding more features could even decrease the model accuracy because it could lead to an overfitting issue. RFECV is a useful tool to support the determination of the number of features that we should keep in the model.

Furthermore, the set of data is another factor influencing the quality of the model. Due to the challenges in collecting activity labels (mention in chapter 4), it is interesting to study the question "How much data is needed to build the model?". In fact, it is difficult to find the correct answer because the required data is different from activities. The part A, B and C of figure 5.12 show the learning curves, which present respectively the relationships between the dataset and the performance of the estimations model for cooking breakfast, cooking lunch and cooking dinner respectively. It is shown that the

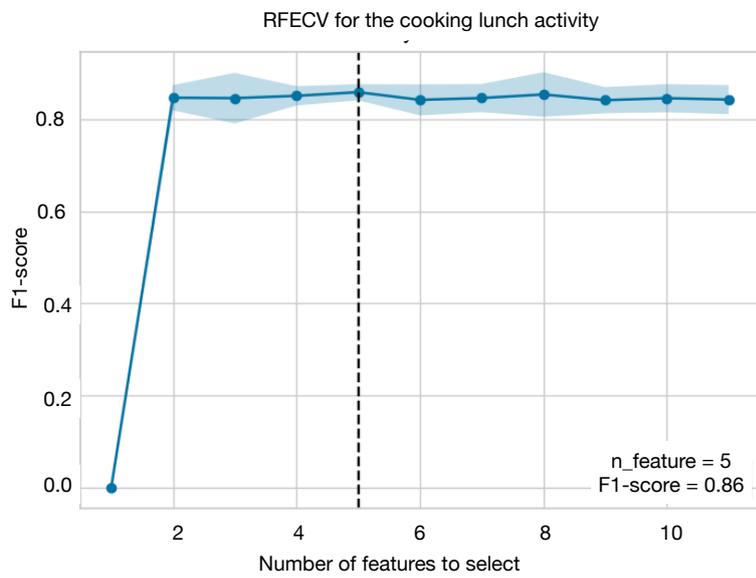
results are more consistent with more data. Particularly, the model which is trained by more training instances has a better testing score. It also avoids overfitting in learning, which is presented by the less discrepancy between the training and testing scores.

Though, remaining to collect activity labels in the long term is not practical. Over time, occupants can be confused by the questionnaires if they do not receive any results or feedbacks from their works. They can lose motivation and wonder about their benefits.

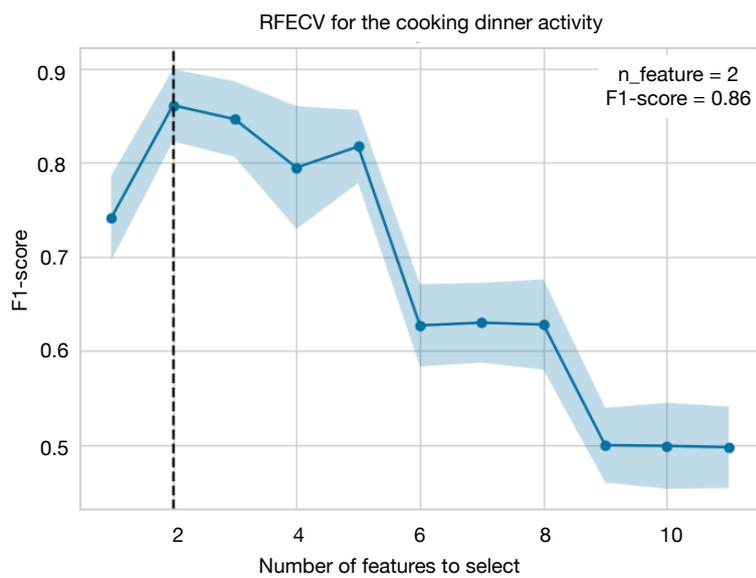
Table 5.8 summaries the estimation results for some activities in this site. It is shown that the electricity consumption of domestic appliances provides significant information to activity estimation models. It is presented by the fact that the features of domestic appliances are the most important features in these models. In addition, the timestep and the movements of occupants are also important. Lacking these features could lead to a poor performance of the model. For example, with the doing personal care activity, due to the privacy policy and technical issues, motions and electricity consumption of appliances could not be recorded in the bathroom. As consequence, the estimation model of this activity has a low performance (F1-score: 0.61).



(A) Recursive feature elimination with cross-validation with cooking breakfast

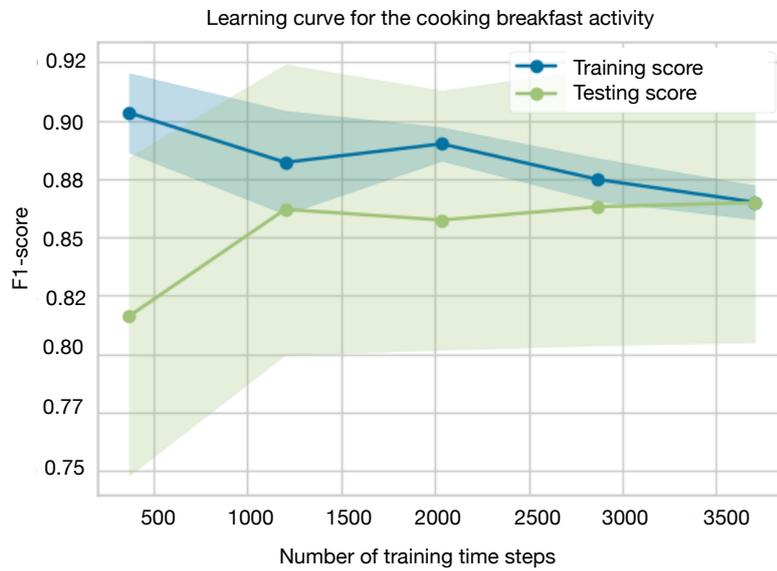


(B) Recursive feature elimination with cross-validation with cooking lunch

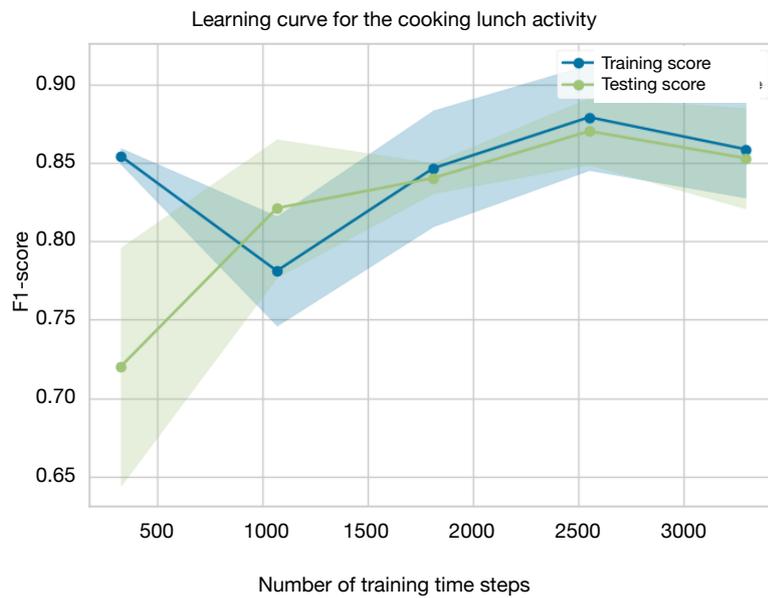


(C) Recursive feature elimination with cross-validation with cooking dinner

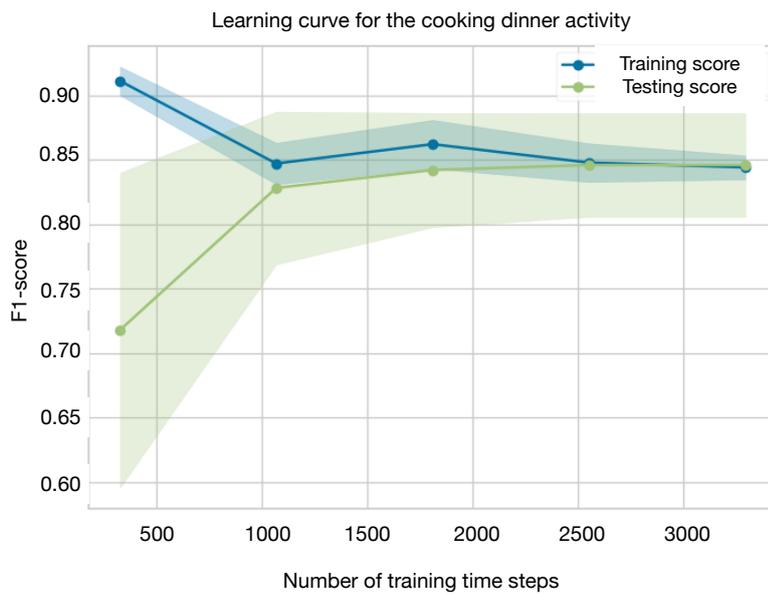
Figure 5.11 – Recursive feature elimination with cross-validation with cooking activity in Cestas household (A: breakfast, B: lunch and C: dinner)



(A) The learning curve with cooking breakfast



(B) The learning curve with cooking lunch



(C) The learning curve with cooking dinner

Figure 5.12 – The learning curves with the cooking activities in the Cestas household (A: breakfast, B: lunch and C: dinner)

Table 5.8 – Results of the estimation of some occupants' activities in the Cestas household

Summary of the estimation of contextualized activities in Cestas				
Activity	Description	Location	Selected features	F1-score
Cooking breakfast	Preparing foods for the breakfast	kitchen	microwave, coffee machine, toaster, motions, timestep	0.87
Cooking lunch	Preparing foods for the lunch	kitchen	microwave, fridge, motions, timestep	0.86
Cooking dinner	Preparing foods for the dinner	kitchen	microwave, motions, timestep	0.86
Entertaining	Using appliances to entertain (watching TV, audio-visual, etc.)	living room	audio-video, computer, CO ₂ , motions	0.77
Washing dishes	Using washing dishes machine to wash dishes	kitchen	washing dishes machine	0.82
Washing clothes	Using the washing machine to wash clothes	laundry room	washing clothes machine	0.86
Drying clothes	Using laundry machine to dry clothes	laundry room	dryer clothes machine	0.88
Doing personal care	Do-it-yourself in the bathroom	bathroom	humidity trend, luminosity	0.61

Beside accurately estimating labels, from observed features which are in four groups (E-environment, C-objects, T-time, O-occupant), CBBN could simulate occupant activity (A) with equation 5.10 as follow:

$$p(A|E, C, T, O) = \frac{p(A|T) \times p(E|A) \times p(C|A) \times p(O|A)}{p(T) \times p(E) \times p(C) \times p(O)} \quad (5.10)$$

In this work, the Monte-Carlo method is combined with CBBN in activity simulation. Figure 5.13 presents measurements from 10/09/2021 to 12/09/2021. Corresponding to these values, 1000 simulations are examined and the results are shown in figure 5.14. It is shown that the proposed model with the stochastic outputs can be used to reproduce states (red-yes, blue-no) of the cooking activity from in-house measurements. This kind of results could be helpful in energy simulation, particularly in building energy verification, in which it is necessary to assess the current energy conditions in the houses. Furthermore, the simulations could be used to construct the profile of occupants' activities, which are useful in improving the performance of energy simulation in the design phase of the building renovation strategy.

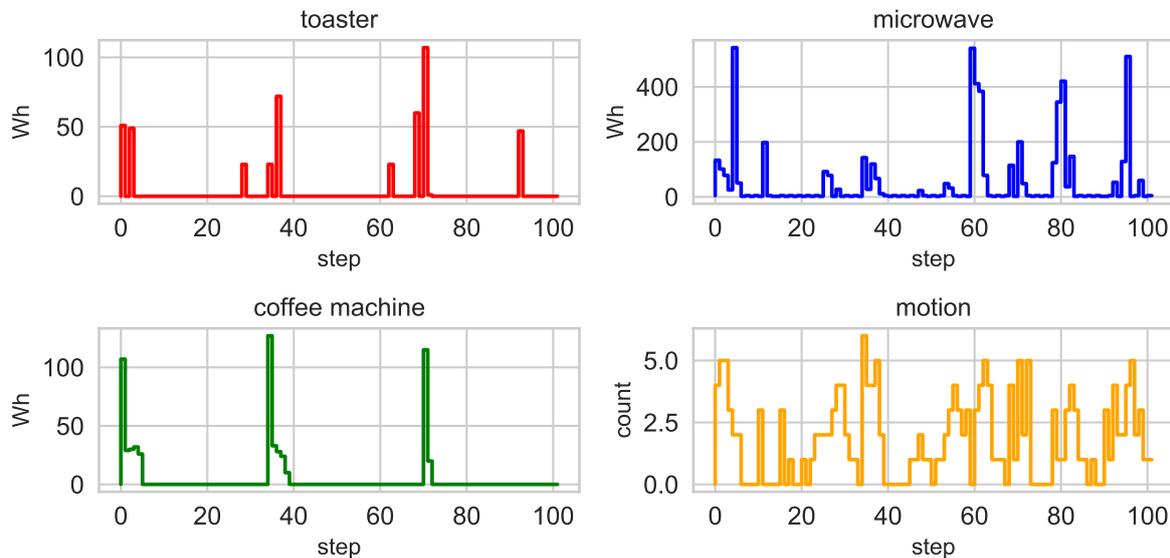
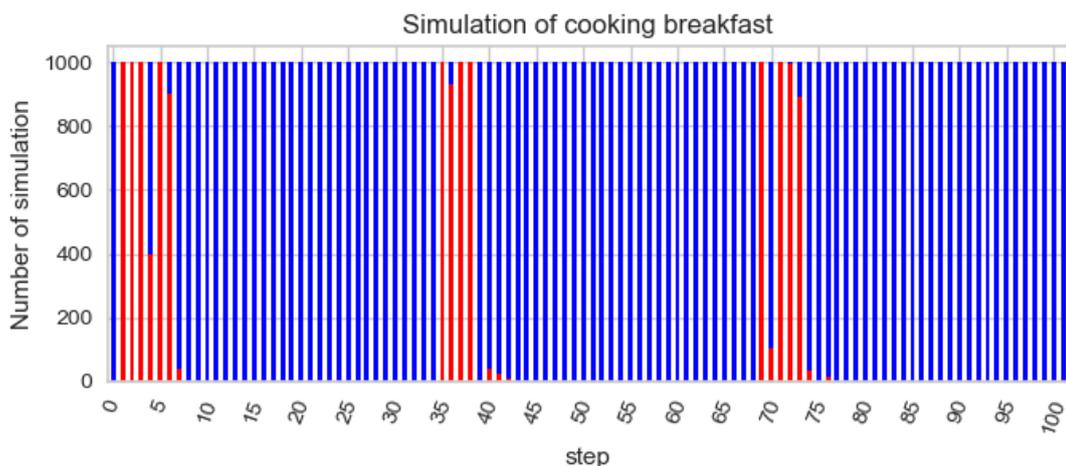
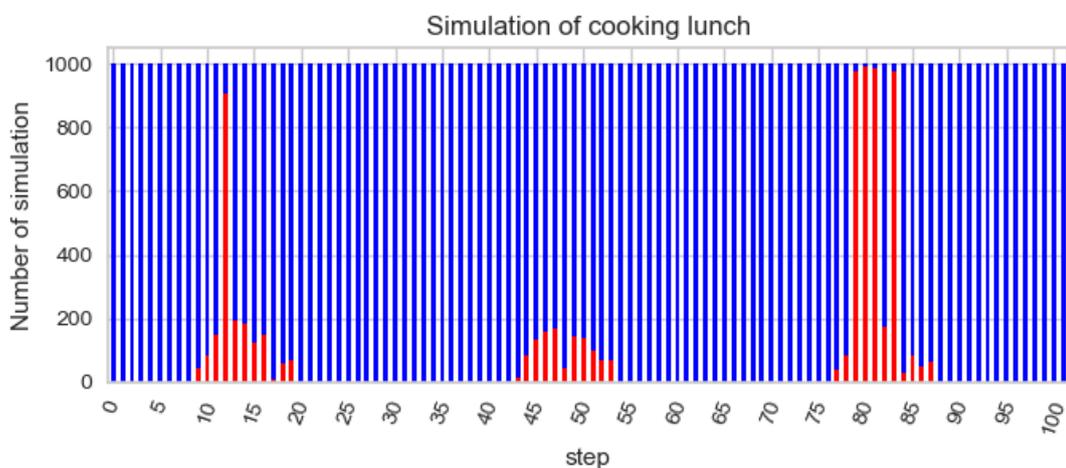


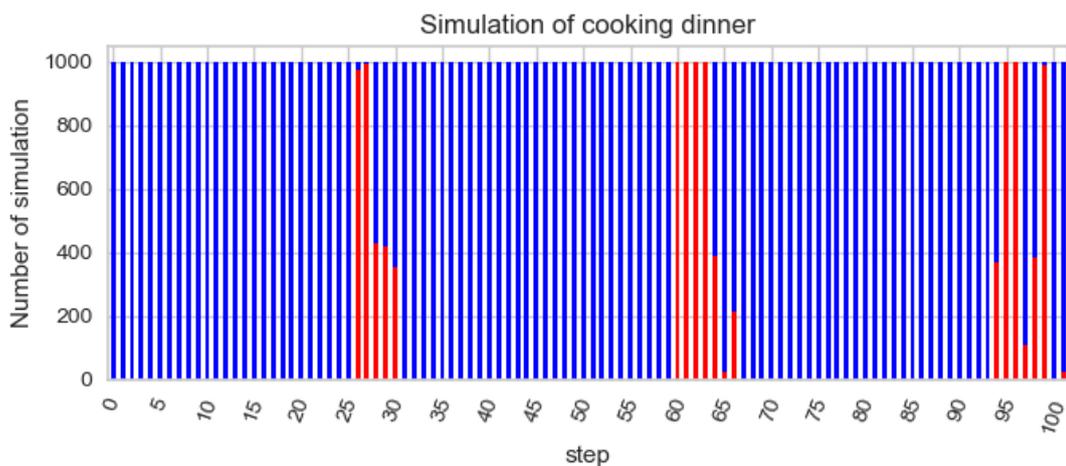
Figure 5.13 – Some examples of measurements (toaster, microwave, coffee machine and motion) from 10/09/2019 06:00:00 to 12/09/2019 23:30:00 in Cestas



(A) The cooking breakfast activity



(B) The cooking lunch activity



(C) The cooking dinner activity

Figure 5.14 – Distributions of the cooking activities deduced from 1000 simulations of the cooking activities from 10/09/2019 06:00:00 to 12/09/2019 23:30:00 in Cestas. It gives the number of simulation that the cooking activity happens (red color) and does not happen (in blue color) for each timestep.

Although F1-score can provide a global view of the estimation model quality, it is interesting to evaluate its performance on building activity profiles. To do it, we compare statistically the estimated activity profile determined by the contextualized approach and the observed activity profile, which is determined directly from the recorded data. The comparison of cooking profile in December 2019 between the contextualized approach and the observed data is shown in figure 5.15.

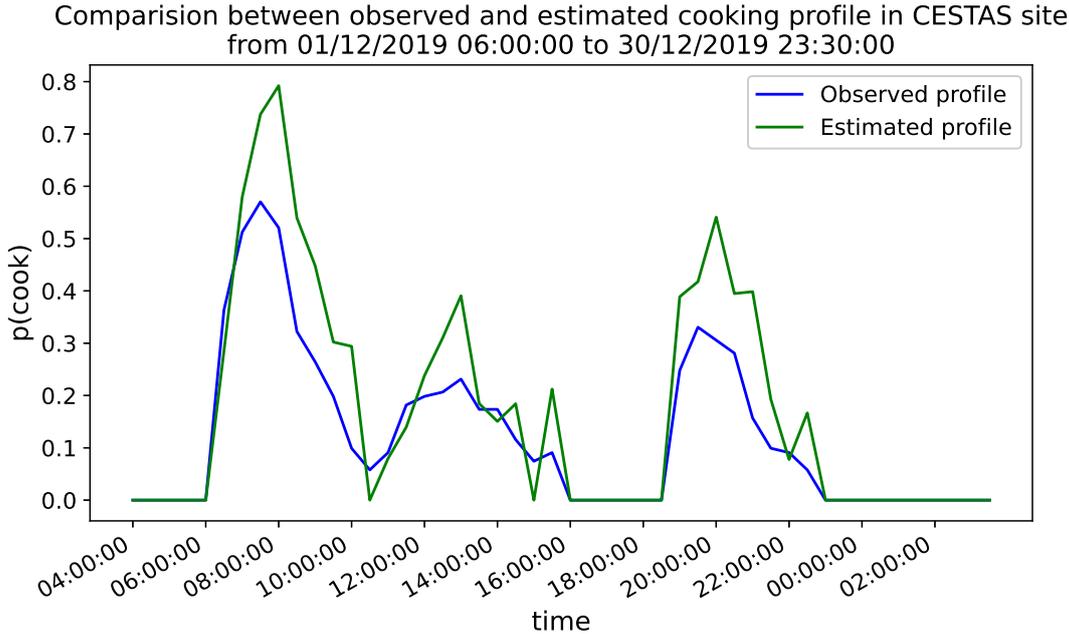


Figure 5.15 – Comparison between the observed and the estimated cooking profile in the Cestas household

The result demonstrates that the estimated cooking profile has a similar trend of cooking profile as the observed data. In both profiles, occupants tend to spend more time cooking breakfast and dinner than cooking lunch. It can be explained by the fact that most of individuals in the Cestas household work outside from morning to evening. Most of the differences in the probability distributions between these profiles are less than 10 %. Thus, the estimation model could be useful in constructing a cooking activity profile. However, most of time, the model underestimates the probability distribution of cooking activity. For example, in the morning, the estimated probability of cooking at 08:00 is 55 % while the actual one is 70 %. This issue can be caused by errors of the model or behavior changes of occupants.

5.3.2 Case study 2: Bordeaux

On this site, occupants cook in the kitchen, which has the following objects:

- a door;
- domestic appliances: kettle, toaster, robot, oven, microwave.

Due to technical issues, the contact sensor and the CO₂ concentration sensor were not installed. Thus, generated features only include air temperature, air temperature trend, luminosity, luminosity trend, air relative humidity, air relative humidity trend, motion, timestep, type of day and electricity consumption of domestic appliances (kettle, toaster, robot, oven, microwave).

Similar to the implementation in the site of Cestas, IG was applied to determine meaningful features. Figure 5.16 shows the estimated IG between the same three studied activities (cooking breakfast, cooking lunch, cooking dinner) and generated features. It is shown that the motions, the timestep and the features of appliances are the most meaningful features to estimate these activities. However, occupants in the site of Bordeaux cook differently from occupants in Cestas. In this household, occupants usually use the kettle for cooking breakfast and the microwave for both cooking lunch and cooking dinner. In addition, the robot is used in preparing food for lunch.

Cooking breakfast		Cooking lunch		Cooking dinner	
Feature	IG	Feature	IG	Feature	IG
Kettle	0.40	Microwave	0.44	Microwave	0.41
Timestep	0.37	Timestep	0.23	Timestep	0.35
Motion	0.20	Motion	0.15	Motion	0.18
Luminosity	0.03	Robot	0.11	Temperature trend	0.03
Temperature, temperature trend, humidity, humidity trend, luminosity trend, robot, toaster, oven, microwave, weekday	0.00	Luminosity	0.04	Luminosity	0.03
		Temperature trend	0.03	Humidity trend, temperature, humidity, luminosity trend, weekday, robot, kettle, oven, toaster	0.00
Temperature, humidity, humidity trend, luminosity trend, kettle, toaster, oven, weekday	0.00	Temperature, humidity, humidity trend, luminosity trend, kettle, toaster, oven, weekday	0.00		

Figure 5.16 – Sorted Information Gain between features and three activities: cooking breakfast, cooking lunch and cooking dinner in Bordeaux house

To build the model, features are manually selected to estimate cooking activities as follows:

- Cooking breakfast: time, kettle, motion;
- Cooking lunch: microwave, time, robot, motion;
- Cooking dinner: microwave, time, motion.

Corresponding to these features, the Consequences-Based Bayesian Networks for the three activities (cooking breakfast, cooking lunch and cooking dinner) are respectively shown in the part A, B and C of figure 5.17 (A: breakfast, B:lunch and C:dinner).

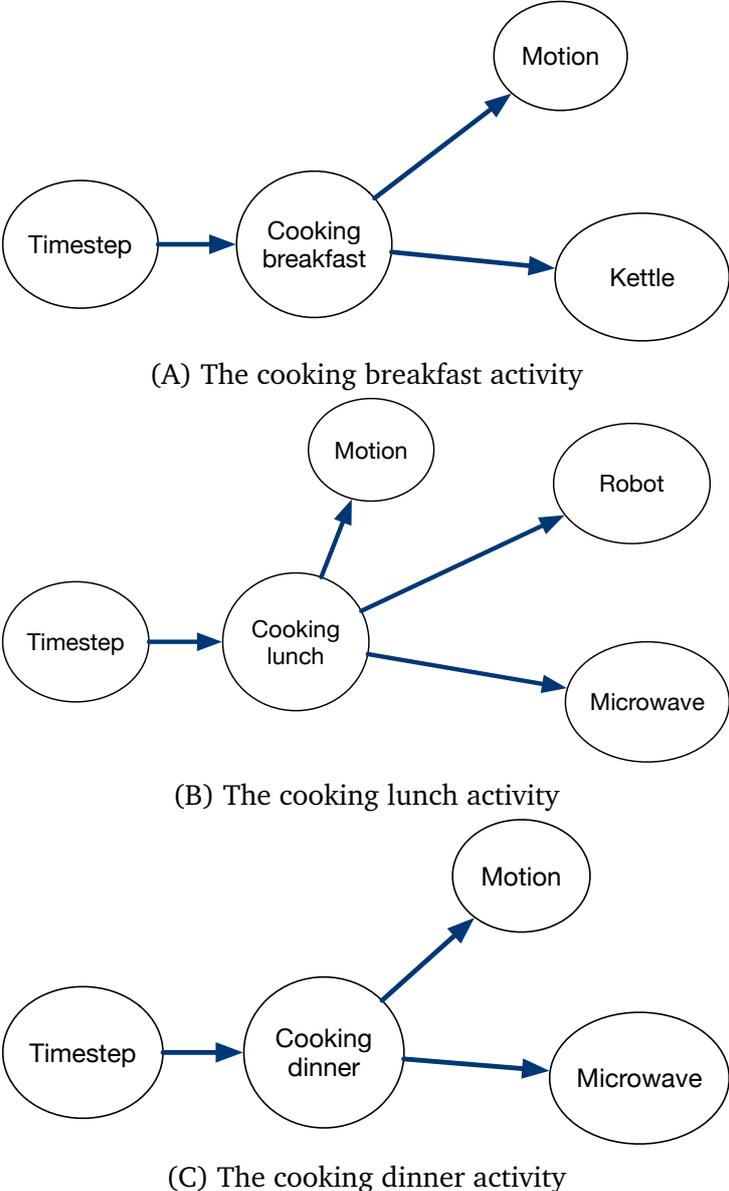


Figure 5.17 – Consequences-Based Bayesian Networks for the cooking activities in Bordeaux site

Table 5.9 presents the comparisons of F1-score between different models. It is shown that the proposed approach, the Decision Tree and the Random Forest algorithms are generally better than other approaches. The F1-score of the CBBN for cooking breakfast, cooking lunch and cooking dinner are 0.74, 0.81 and 0.80 respectively.

Table 5.9 – Comparison of F1-scores among different approaches in cooking estimation in Bordeaux site

	Cooking breakfast	Cooking lunch	Cooking dinner
CBBN	0.74	0.81	0.8
Decision Tree	0.76	0.83	0.81
Random Forest	0.73	0.82	0.82
Logistic regression	0.6	0.6	0.6
SVM	0.61	0.62	0.56
BN with Chow-Liu	0.62	0.65	0.6
BN with Greedy	0.65	0.7	0.51
BN with Hill-Climbing	0.6	0.65	0.68
BN with Exact	0.65	0.68	0.52

Figure 5.18 presents some measurements from 17/04/2021 to 18/04/2021. 1000 simulations are examined to deduce the cooking activities from these variables. The simulated results are shown in figure 5.19.

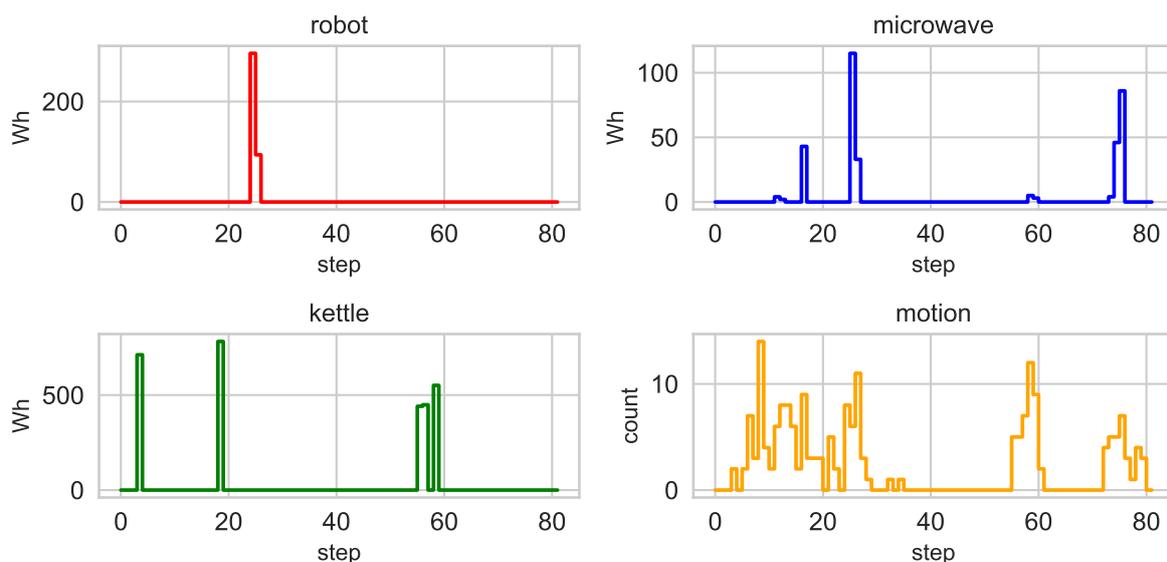
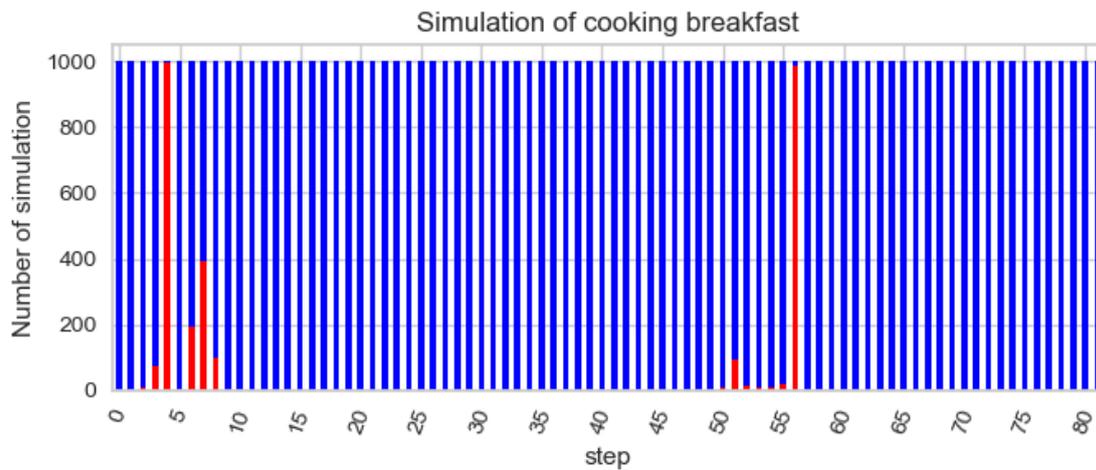
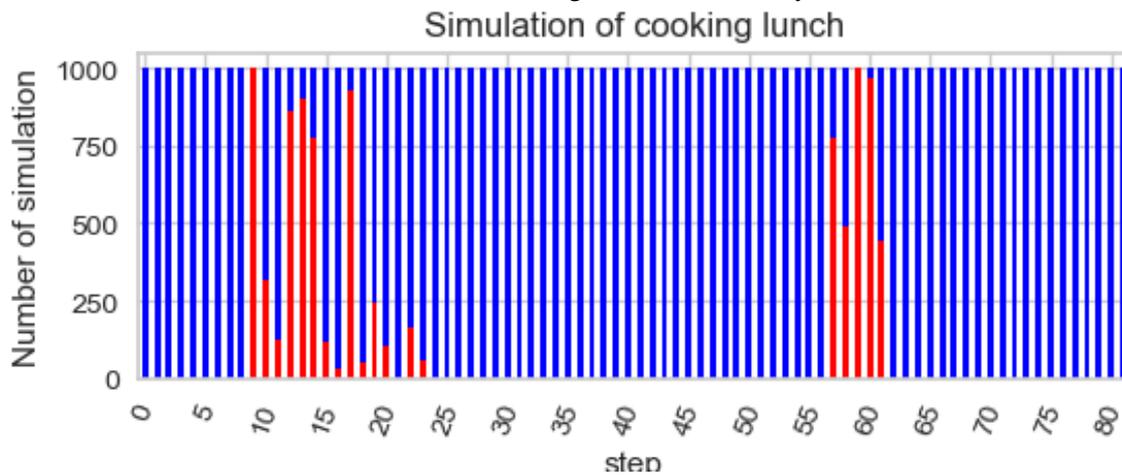


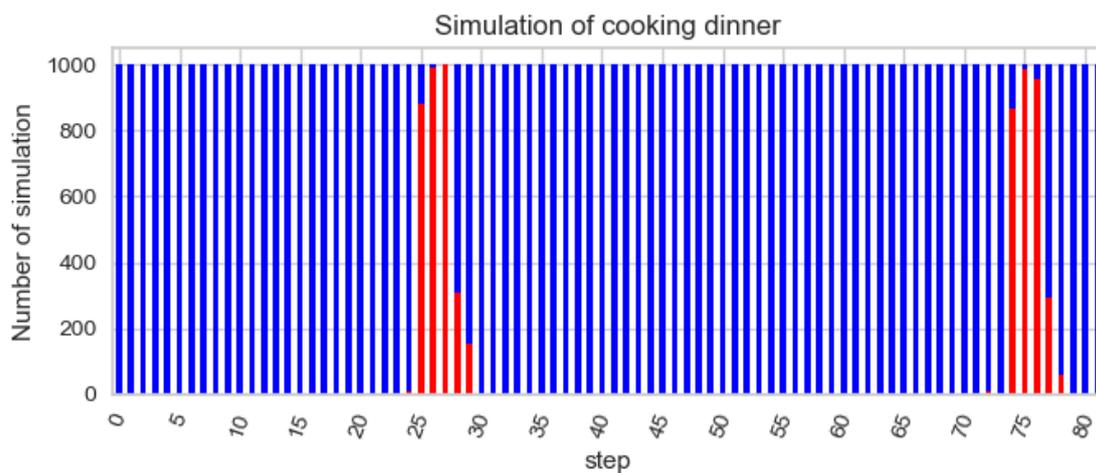
Figure 5.18 – Some measurements (robot, microwave, kettle, motion) from 17/04/2021 06:00:00 to 18/04/2021 23:30:00 in Bordeaux site



(A) The cooking breakfast activity



(B) The cooking lunch activity



(C) The cooking dinner activity

Figure 5.19 – Distributions of the cooking activities deduced from 1000 simulations of the cooking activities from 17/04/2021 06:00:00 to 18/04/2021 23:30:00 in Bordeaux site. It gives the number of simulation that the cooking activity happens (red color) and does not happen (blue color) for each timestep.

The comparison of the cooking profile between the contextualized approach and the actual data in April 2021 is shown in figure 5.20. The results highlight that the estimated cooking profile has a similar trend to the actual profile. These results tend to show that the proposed methodology can be satisfactorily applied to other case studies with new contexts.

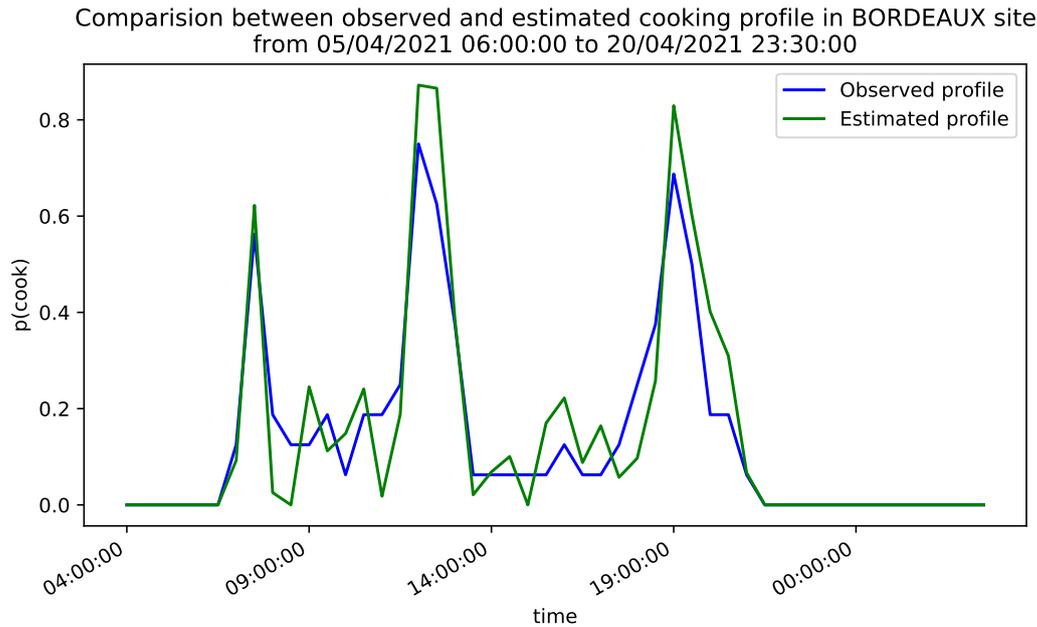


Figure 5.20 – Comparison between the observed and the estimated cooking profile in the Bordeaux site

5.4 Conclusions

This chapter proposes a methodology to estimate occupants' activities from *in-situ* measurements. The knowledge is used to generate useful features from sensors data. However, some features could not provide useful information to the estimation of all occupants' activities. Thus, Information Gain is applied to select the most meaningful features. Then, Consequences-Based Bayesian Networks are built to investigate the relationships between selected features and the considered occupants' activities. The network is human-understandable and has stochastic outputs.

A crucial contribution in this methodology is the verifiability, which is limited in stochastic activity models. Expressly, F1-score with 3-folds cross-validation is used to evaluate the quality of the estimation model. In addition, we statistically compared

the estimated activity profile with the observed activity profile to evaluate the model potential in constructing activity profiles. Cooking activity with three instances (cooking breakfast, cooking lunch, cooking dinner) was modeled in two sites in France (Cestas and Bordeaux) to illustrate and test the methodology.

The results show that domestic appliances' features and the occupant's motions are the most meaningful features in modeling occupants' activities in these case studies. Lacking this information can decrease the model quality. Moreover, the number of selected features and the size of the dataset also affect activity models. An example of the cooking activity has been presented to show that the model could simulate this activity in different case studies. In addition, it presents that occupants have different patterns of activities in different households. More investigations into the ways that occupants do activities, particularly in using appliances, are conducted in next chapter 6.

Chapter 6

Estimation of the electricity consumption in occupants' activities

Contents

6.1 Introduction	118
6.2 Problematic: modelling the electricity consumption of appliances in houses	118
6.2.1 Occupants' activities and the consumption of electrical appliances	118
6.2.2 Existing works	119
6.2.3 Objectives and research questions	120
6.3 Methodology to model the electricity consumption of domestic appliances for a related activity	122
6.3.1 Related domestic appliances used for an activity	122
6.3.2 Determining the activity related consumption of appliances	122
6.3.2.1 The appliance consumes electricity mostly depending on the occupancy	123
6.3.2.2 The appliance consumes electricity during the activity	123
6.3.2.3 The appliance consumes electricity after the activity	125
6.3.3 Modelling the activity related electricity (AR) consumption of appliances	127
6.3.4 Evaluating the model performance	129

6.4 Implementation and results	130
6.5 Conclusion	145

6.1 Introduction

Occupant activities play important roles in explaining discrepancies in electricity consumption among houses. Occupants can have distinct habits (activity profiles), which can be performed in various ways. In some activities, appliances are used to support occupants' purposes and their operations consume a certain amount of electricity (e.g., use of a microwave to cook, use of a washing machine to wash clothes, *etc.*). Thus, determining the consumed electricity of these appliances is an essential step to analyze the global energy consumption in a house. Many studies proposed stochastic electricity models to simulate appliances and their related electricity uses, based on statistical data (Wilke et al., 2013; Aerts et al., 2014; Zaraket, 2014). Similar to representative activities, stochastic electricity models present average results, not specific to particular houses.

Modeling the electricity consumption of appliances is useful in explaining and comparing the energy consumption among households. In this chapter, we focus on this task in specific instrumented households.

6.2 Problematic: modelling the electricity consumption of appliances in houses

6.2.1 Occupants' activities and the consumption of electrical appliances

In European residential buildings, appliances are of great importance and should be investigated to improve the energy performance, in addition to heating and cooling systems. They consume more than 20% of the final energy consumption (Françoise, 2018).

The electricity consumption of appliances can be different among dwellings (Firth

et al., 2008; Le et Pitts, 2019). Huebner et al. (2016) investigated that English households used electrical appliances in various ways and consumed different electricity levels for their needs. To explain this diversity, many studies focused on building types and household characteristics, which affect ownerships and qualities of appliances (Jones et Lomas, 2016; Zou et Mishra, 2020). However, even if houses have similar building characteristics in the same location, there are discrepancies among appliance usages (Firth et al., 2008). Many studies conducted that the diversity of occupant activities causes this gap (Vorger, 2014; Zaraket, 2014; Aerts et al., 2014).

In houses, while some appliances work mainly based on occupancy (e.g., lighting systems or heating systems), many electrical appliances work for specific activities (microwave, toaster, television, *etc.*). Studying relations between these activities and related appliances could give more explanations on differences in electricity uses between households. Ahmadi-Karvigh et al. (2016) also stated that determining the consumption of electrical appliances affected by activities could detect potential energy savings and give feedback and recommendations to occupants. Recently, many scientists linked appliances' uses to occupant activities to simulate electricity demand and load profiles in households (Vorger, 2014; Aerts et al., 2014; Tanimoto et al., 2008).

6.2.2 Existing works

Electricity uses of appliances could be modeled directly without occupants' activities. Hawarah et al. (2010) used Bayesian Network and statistic data to model the uses of some appliances, then predicted the electricity consumption in houses. In another study, Murray et al. (2018) uses the Signature Detection Algorithm (SDA) to recognize and model over time the electricity consumption of appliances in houses. Though, modeling only appliances is difficult to unmask impacts of occupant behavior on the electricity consumption in houses because their relationships are not investigated.

To tackle this issue, many scientists proposed statistical approaches to model relations between occupants' activities and appliances in dwellings (Wilke et al., 2013; Aerts et al., 2014; Vorger, 2014). In these approaches, depending on household characteristics (income, household size, *etc.*), a logistic regression algorithm was used to determine probabilities that appliances exist in households. Typical power demands were then determined depending on survey data using uniform distributions. To predict

the consumption of many electrical appliances for residential activities, the authors used the logistic regression algorithm to infer conditional probabilities that an appliance is in use while an activity is taking place. Then, the electricity consumption of appliances was estimated depending on the typical power and the activities' durations.

In another study, Kashif (2014) proposed a multi-agents model to investigate specifically the electricity consumption of a fridge in a kitchen. The fridge was focused because it contributed a certain amount of electricity in a house and its consumption can be affected by not only occupants' activities but also by other factors (indoor temperature, random actions, *etc.*). To model its consumption, the author considered both its consumption affected by the cooking activity (by analyzing the consumption of the cooker in the kitchen) and its consumption caused by other factors (normal compressor cycle, other activities, *etc.*).

However, some limitations exist in these above models:

- Firstly, the models are based on statistical data, which comprises households with various characteristics (building type, income, individuals, *etc.*). Therefore, their results are average, not specific and challenging to be verified.
- Secondly, typical power demands cannot present the diversity of how occupants use appliances during their activities. For example, in the morning, occupants can use a microwave to heat the milk in a short time. In the evening, they need to defrost foods (chicken, meat, *etc.*) so that they can use the microwave for a longer time. It is obvious that the power demands of the microwave in these cases are different. Therefore, using the typical power demand of the microwave is not efficient in this case.

Thus, it is interesting to analyze and model relationships between occupants' activities and electrical appliances' use in specific houses.

6.2.3 Objectives and research questions

In this work, an activity is a label provided by an occupant bringing an information about a continuous period. A set of electrical appliances can be associated with an activity. During an activity, different actions can be done by occupants. The actions induce modification of the consumption of appliances but just some of the actions can

be known. To deal with this issue, activities can give a context of the actions carried out. Given an activity and a related appliance, a consumption profile can be associated.

Because activities are independently provided by occupants, more than one activity can be recorded in a timestep. An appliance can be affected by several activities at the same timestep and it is challenging to distinguish impacts of each activity on its electricity consumption. To simplify the problem, we assume that an appliance cannot be impacted by more than one activity in a timestep. When there is more than one activity related to an appliance in a timestep, the electricity consumption of the appliance should be associated with the most relevant activity (defined manually or intuitively).

In a home, an appliance can consume electricity during many timesteps in a day and a certain amount of them are impacted by occupants' activities. An activity with actions can affect the consumption of different types of electrical appliances. Firstly, they can trigger some automation appliances depending on the occupancy like ventilation, lighting or heating systems. Secondly, during an activity, occupants' actions can affect the electricity consumption of domestic appliances (microwave, toaster, television, *etc.*) In this work, we do not study automation appliances whose consumptions are mostly based on the occupancy because occupants' activities are not needed. Many approaches have been proposed to particularly model occupancy and its energy impacts on heating systems (Wang et al., 2020; Pang et al., 2020; D'Oca et al., 2014; Fabi et al., 2013) and on lighting systems (Page et al., 2008; Vorger, 2014; Nagy et al., 2015).

The objective of this chapter is to study the part of the consumption of an appliance that is related to a given activity in specific households, which is denoted by Activity Related consumption (AR consumption) of the appliance. This task is useful to predict the consumption of electrical appliances when occupants' activities happen. To achieve this goal, two issues need to be studied:

- Firstly, the appliances whose consumption is dependent on activities must be identified.
- Secondly, for a given activity, the AR consumption of domestic appliances need to be estimated.

As mentioned in chapter 3, the following activities with related appliances are considered:

- entertaining;
- drying clothes;
- washing dishes;
- washing clothes;
- cooking.

The principles and methods used are described in the next sections.

6.3 Methodology to model the electricity consumption of domestic appliances for a related activity

6.3.1 Related domestic appliances used for an activity

To analyze the impacts of occupants' activities on the electricity consumption, it is essential to determine their related impacted domestic appliances. We should not model the electricity used by a not impacted appliance. To address this issue, when modeling an appliance, based on statistical methods, many scientists used the household characteristics to predict the ownership of the appliance, which represents the probability for a household to have the appliance and then applied a logistic regression model to assign it to occupants' activities (Wilke et al., 2013; Vorger, 2014; Zaraket, 2014; Aerts, 2015). This approach is popular in predicting the diversity of the appliances' settings in a household where this information is not available. Though, the estimated ownership is an average output from many similar households and is not specific to particular households. To tackle this issue, we collect this information directly from occupants through questionnaires. These questions comprise the available appliances in the house and the possible appliances used in occupants' activities.

6.3.2 Determining the activity related consumption of appliances

During an activity, occupants can do many actions and these actions can induce modifications of the electricity consumption on appliances. That will cause different impacts on their electricity consumption. To determine the AR consumption of an appliance for a given activity, actions on the appliance should be identified. However, most of the actions are difficult to detect. It needs many efforts from occupants to record their actions and the corresponding times. To deal with this issue, labels of the activity,

which are information provided by occupants, can give a context of the actions carried out. Thus, this work simplifies the problem by assessing relations between the recorded activity labels and the AR consumption of the appliance.

For an activity, appliances can consume electricity in different ways. For instance, the consumption of an appliance can be delayed when an activity happens such as the washing machine (in washing clothes) or the dishwasher (in washing dishes). In other activities, appliances can be used during the activity such as the television or the computer (in entertaining). Dealing with this diversity, based on approaches proposed by Widén et Wäckelgård (2010) and Aerts et al. (2014), there are three types of relations between an activity and the AR consumption of an appliance:

- the appliance consumes electricity mostly depending on the occupancy;
- the appliance consumes electricity during the activity;
- the appliance consumes electricity after the activity.

6.3.2.1 The appliance consumes electricity mostly depending on the occupancy

For automation appliances such as lighting or heating appliances, they are operating during presences. The presence of occupants during activities can affect operations of these appliances. As discussed in section 6.2.3, labels of activities are generally not needed to consider these appliances and they are not studied in this work.

6.3.2.2 The appliance consumes electricity during the activity

Some appliances can be used during an activity or during a part of an activity. Occupants usually stop these appliances before or when they finish the activity. Therefore, the AR consumption of the appliances is only considered within the activity (see in figure 6.1). This type of relation concerns the activities using appliances such as the television, the computer, the audio-visual devices, the iron or the aspirator, which are started and stopped only by occupants.

This type of relation can also be applied to some appliances used in cooking such as the food processor and the cooking robot. On the other hand, many appliances used in cooking are set-points appliances such as the toaster, the microwave, the coffee machine, the yogurt maker, the kettle, *etc.*, which are started by occupants and stop

automatically based on predefined conditions. However, occupants usually do other actions for cooking and do not usually change to other activities while these appliances are performing. Thus, for simplicity, we can assume these appliances that consume electricity during the cooking activity.

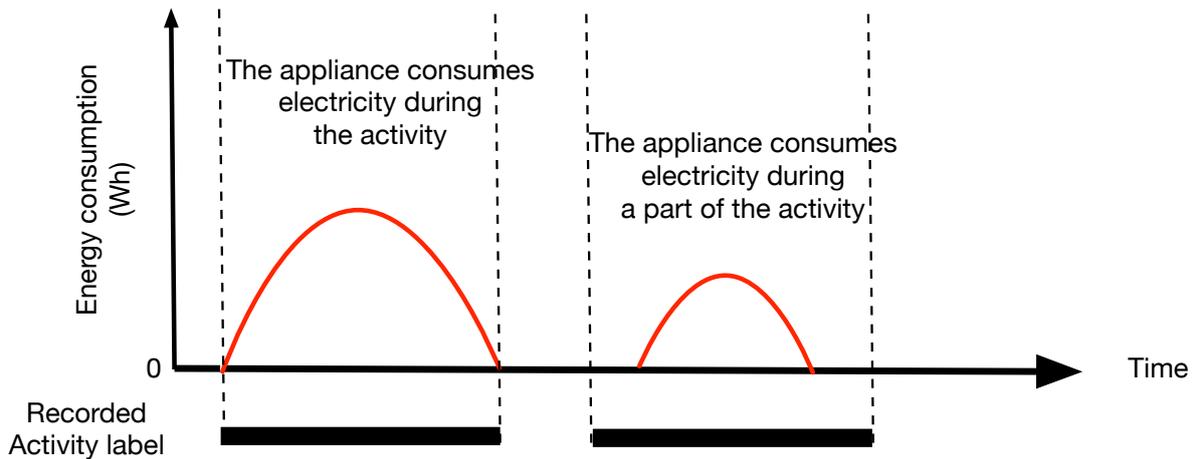


Figure 6.1 – Illustration for the appliances consuming electricity during the activity or during a part of the activity.

The consumption of the above appliances is only impacted by occupants during their activities. Thus, the electricity consumption of an appliance can be affected by both occupants' activities and other factors. For instance, a fridge is usually associated with the cooking activity. During the activity, occupants can open/close the fridge to put in or take out foods and these actions modify its consumption. The consumption of the fridge can also be impacted when the air temperature increases or when occupants do specific non-cooking actions such as putting foods in after going to the market, taking out juices to drink, *etc.* In this work, we do not consider the fridge because using only activities is not enough to study its electricity consumption.

Because an appliance with this type of relation consumes electricity during the activity, when studying the AR consumption of the appliance, we study the relation between its consumption and its related activity in the same timestep. As discussed above, in an activity, occupants can do various actions on the appliance and they induce different AR consumptions of the appliance. Therefore, it is not efficient to use a typical AR consumption of the appliance for its related activity. Considering

all possible AR consumption of the appliance is also challenging. In this work, for simplicity, as mentioned in chapter 5, we assumed that an appliance could have five levels of AR consumption: very low (VL), low(L), medium(M), high(H), very high(VH). These levels are detected using k-means clustering. This algorithm is used because it's easy to implement and analyze. There is a typical consumption value for each AR consumption level, which is the average of all values at this level.

To represent relationships between an activity and the AR consumption levels of an appliance, we use a Bayesian Network shown in figure 6.2. In a timestep t with a given activity label a_t , to evaluate the AR consumption of an related appliance, probability distributions $p(x_t|a_t)$ need to be estimated with x_t is the AR consumption level of the appliance. These distributions provide insights into how appliances are used in occupants' activities.

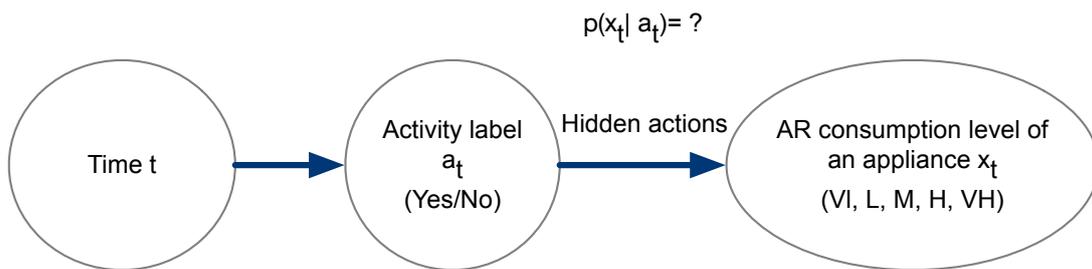


Figure 6.2 – The proposed Bayesian Network to represent the relationship between an activity and the electricity consumed by an appliance

6.3.2.3 The appliance consumes electricity after the activity

An activity can be related to only one appliance such as washing clothes (related to washing machine), drying clothes (related to tumble dryer) or washing dishes (related to the dishwasher). To do the activity, the appliance is started by occupants and they then stopped automatically based on predefined settings (mode, duration, etc.) Because the activity uses only one appliance, occupants usually declare the activities only at the timestep when they start the appliance. It can be inferred that the activity ends after this action. However, the appliance starts to consume electricity or continues to consume the electricity after this timestep (see in figure 6.3). Its AR consumption is not limited to the timestep when the activity is declared. It should be the total electricity that the appliance consumes from starting to stopping.

Therefore, to determine the AR consumption of an appliance with this type of

relation, it is necessary to identify starting timestep and stopping timestep of the appliance. Starting timesteps can be determined directly using the labels provided by occupants, which declare when an appliance is started. To detect stopping timesteps, the way that the appliance consumes electricity should be concerned. For appliances such as washing machine, dishwasher and tumble dryer, from starting to stopping, they often consume a little electricity in the beginning, their consumptions increase significantly in the middle and they decrease in the end. These signatures are used to identify stopping timesteps of the appliance, which should have the following conditions:

- the electricity consumption is very low and less than a threshold at this timestep, this threshold can be defined for each specific appliance using data analysis;
- there is a decreasing trend of the energy consumption in previous timesteps;
- there is no increasing trend of energy consumption after this timestep if there is no starting point in the next timesteps.

Figure 6.4 illustrates this example and the way to detect a starting timestep and a stopping timestep of an activity and a related appliance.

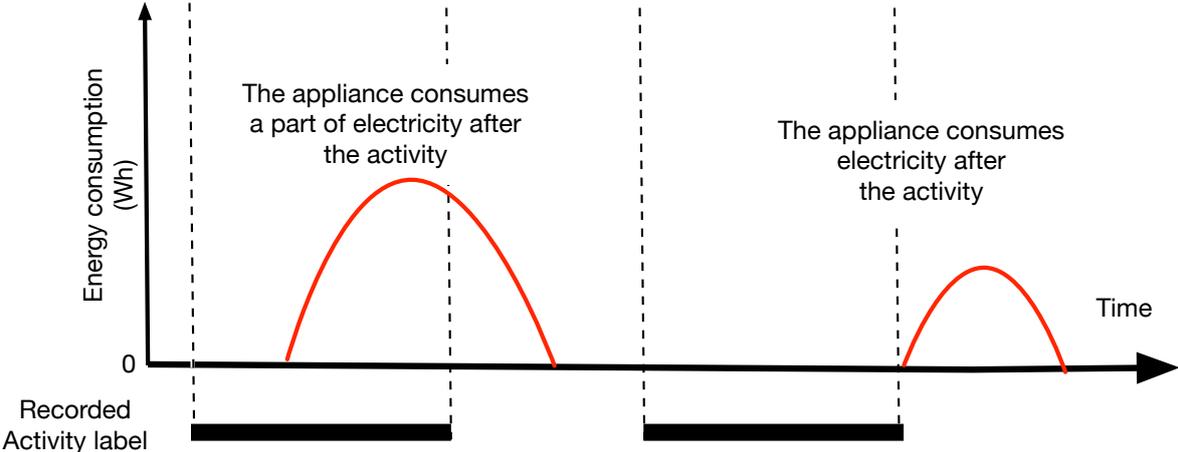


Figure 6.3 – Illustration for the appliances consuming electricity during the activity or during a part of the activity.

Then, to investigate the AR consumption of an appliance for a related activity, at each timestep t , distributions $p(x_t|a_t)$ are determined with x_t is the AR consumption

level of the appliance and a_t is a given label of the activity. The representation of this relationship is similar to the Bayesian Network presented in figure 6.2.

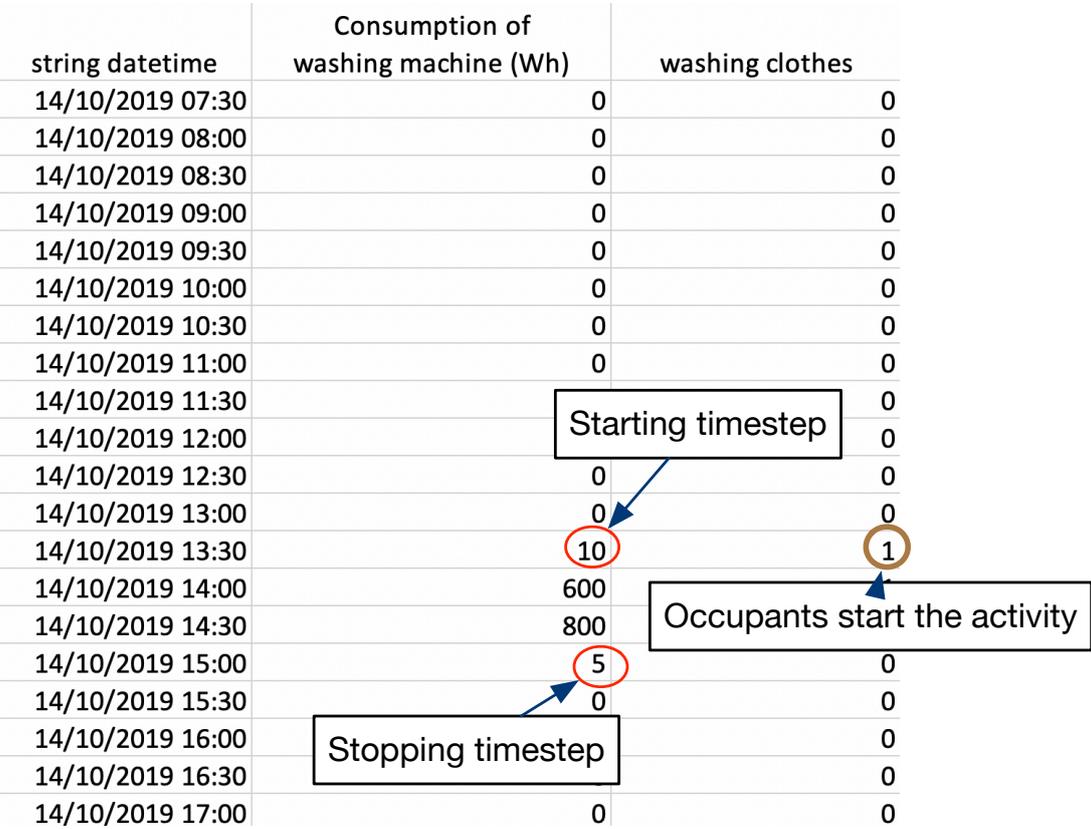


Figure 6.4 – An example of detecting a starting timestep and a stopping timestep of an appliance consuming electricity after an activity ends

6.3.3 Modelling the activity related electricity (AR) consumption of appliances

This section aims to propose an AR consumption model to predict the AR consumption of an appliance for a given activity. Modeling occupants’ activities is not the objective of this model so that they should be considered by an independent process. Thus, the collected labels are used directly to build the model and assess the performance of the predicted electricity consumption.

The two above steps are combined to determine the AR consumption of an appliance when a related activity happens. Firstly, activity dependent appliances are determined using knowledge about the household. The AR consumption model must represent relationships between the activity and these related appliances. These relationships should be understandable by humans. In addition, the diversity of AR consumption

should be considered. Bayesian Network is a probabilistic graphical model that can meet these requirements because it can present causal relationships between the activity and the appliances.

Thus, to represent impacts of an activity and dependent appliances, a Bayesian Network is built as in figure 6.5. At each timestep t , conditional probability distributions between an activity with a given label a_t and the AR consumption levels of its related appliances are determined using the methods described in section 6.3.2, which are based on the type of the relationship between an activity and an appliance.

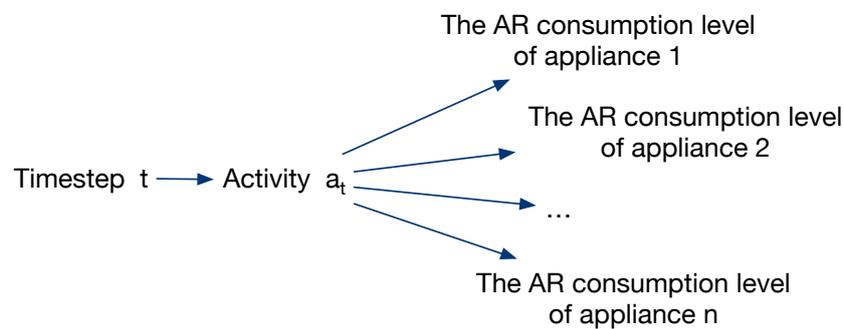


Figure 6.5 – The structure of Bayesian Network for predicting AR consumption of appliances for a given activity

Figure 6.6 illustrates schematically the algorithm which is used to predict the AR consumption of an appliance associated with a given activity in each simulation as follows:

1. In each timestep t , two situations are considered:
 - If the activity happens ($a_t = 1$): the AR consumption level of the appliance is predicted through the Monte-Carlo method and estimated CPDs, then move to the next timestep if it is not the final timestep t_{end} .
 - If the activity does not happen ($a_t = 0$): move to the next timestep if it is not the final timestep t_{end} .
2. The simulation model is terminated when it reaches the final timestep t_{end} .

When the AR consumption levels of the appliance were predicted, they are used to deduce the AR consumption through typical AR consumption values corresponding to each AR consumption level.

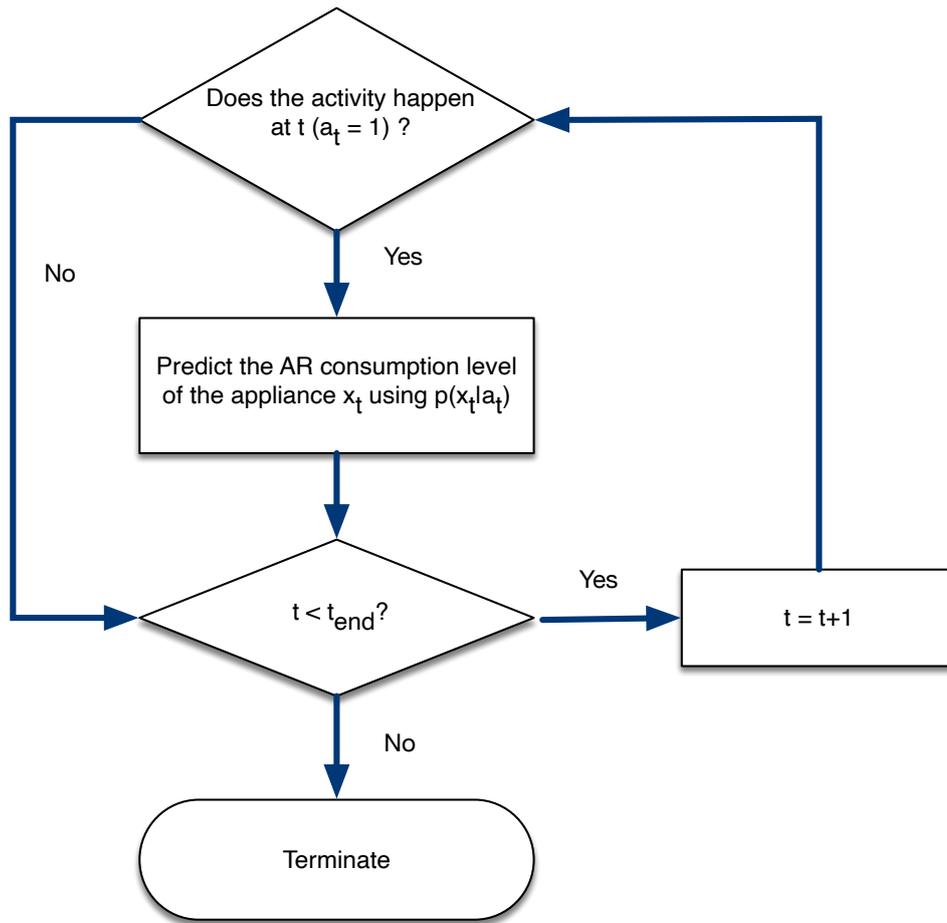


Figure 6.6 – Algorithm to simulate the AR consumption level of an appliance for a related activity

6.3.4 Evaluating the model performance

To evaluate the proposed AR consumption model, we verified whether it accurately simulates the AR consumption of appliances for several studied activities. To assess the performance of the simulated AR consumption of an appliance for a given activity a , we use a simple relative error (RE) as follows:

$$RE = \frac{AR_{simulated} - AR_{actual}}{AR_{actual}} \times 100\%$$

where:

- AR_{actual} is the average measured daily AR consumption of the appliance for a ;
- $AR_{simulated}$ is the average simulated daily AR consumption of the appliance for a .

To avoid the overfitting issue, the dataset is divided into a training and a testing part. First, the model is built with the training data and then the testing data are used for

the performance evaluation. In addition to the ER, boxplots are used to evaluate the diversity of the simulated electricity consumption.

6.4 Implementation and results

To illustrate the methodology, we tested three kinds of cooking activity (cooking breakfast, cooking lunch and cooking dinner) with the Cestas household, which correspond to three periods of the day:

1. cooking breakfast: in the morning, from 6:00 a.m to 10:59 a.m;
2. cooking lunch: in the afternoon, from 11:00 a.m to 17:59 p.m;
3. cooking dinner: in the evening, from 18:00 p.m to 23:00 p.m.

The dataset includes sensors data and collected labels from 01/09/2019 to 30/12/2019. This dataset is divided into two parts:

- Training part: 01/09/2019 to 30/11/2019;
- Testing part: 01/12/2019 to 30/12/2019.

Determination of appliances related to an activity

Table 6.1 shows appliances which are measured and related to concerned occupants' activities in the Cestas household.

Table 6.1 – Possible appliances used in occupants' activities in the Cestas household

Activity	Possible used electrical appliances
Cooking	microwave, toaster, coffee machine, food processor, robot, yogurt maker
Entertaining	audio-visual devices
Washing clothes	washing machine
Drying clothes	tumble dryer machine
Washing dishes	dishwasher

It is shown that appliances related to three kinds of cooking activity are: the microwave, the toaster, the coffee machine, the food processor, the robot and the yogurt

maker. However, from 09/2019 to 12/2019, the food processor, the robot and the yogurt maker were rarely used in cooking, which is presented in figure 6.7. Thus, the cooking activity is studied with three electrical appliances: the microwave, the toaster, and the coffee machine in this case study.

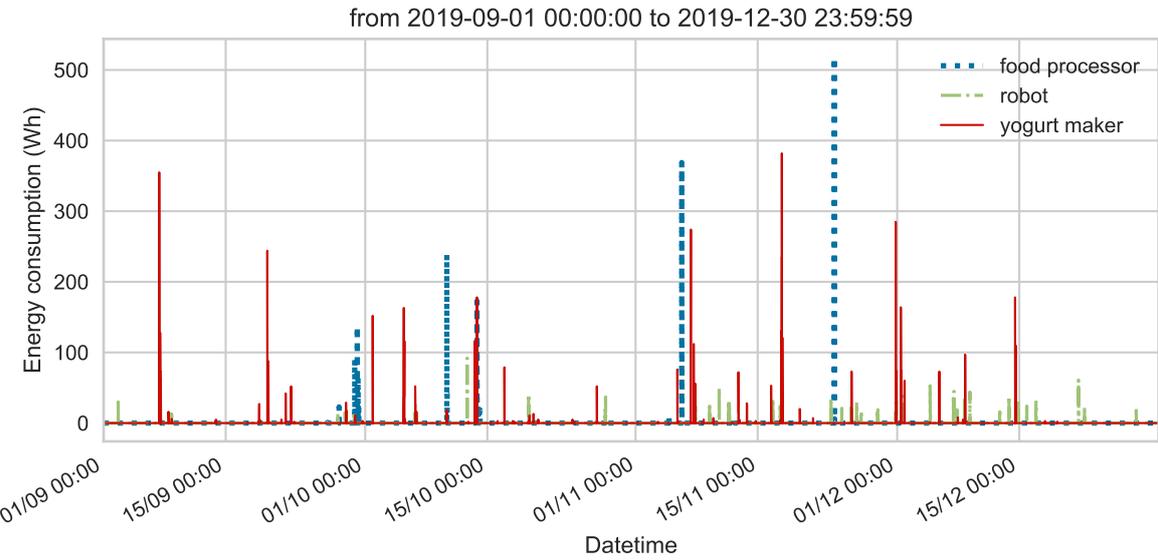


Figure 6.7 – Examples of the electricity used by the appliances (robot, yogurt maker, foods processor) that are rarely used in the Cestas household

Determination of the activity related consumption of appliances

Once the dependent appliances are determined, the next step in the methodology is to estimate the distributions of the AR consumption of the appliances for related activities. These distributions represent the diversity of ways that occupants use the appliances for their needs. To do that, the type of relation between appliances and activities needs to be identified. Based on the discussion in section 6.3.2, the types of relation between concerned activities and the AR consumption of related appliances in the Cestas household is shown in table 6.2. It is shown that the relation between kinds of the cooking activity (cooking breakfast, cooking lunch and cooking dinner) and related appliances is "the appliance consumes electricity during the activity". Thus, as mentioned in the methodology, to determine conditional probability distributions (CPD) between the AR consumption levels of an appliance and a related activity, we study the relation between the consumption of the appliance and collected labels of the activity in the same 30-minutes timestep.

Table 6.2 – The type of relation between activities and the AR consumption of appliances in the Cestas household.

Activity	Appliances	Type of relation
Cooking	Microwave, toaster, coffee machine, food processor, robot, yogurt maker	The appliance consumes electricity during the activity
Entertaining	Audio-visual devices	The appliance consumes electricity during the activity
Washing clothes	Washing machine	The appliance consumes electricity after the activity
Drying clothes	Tumble dryer machine	The appliance consumes electricity after the activity
Washing dishes	Dishwasher	The appliance consumes electricity after the activity

The estimated CPDs of the coffee machine, the toaster and the microwave for the cooking breakfast are respectively presented in figures 6.8, 6.9 and 6.10.

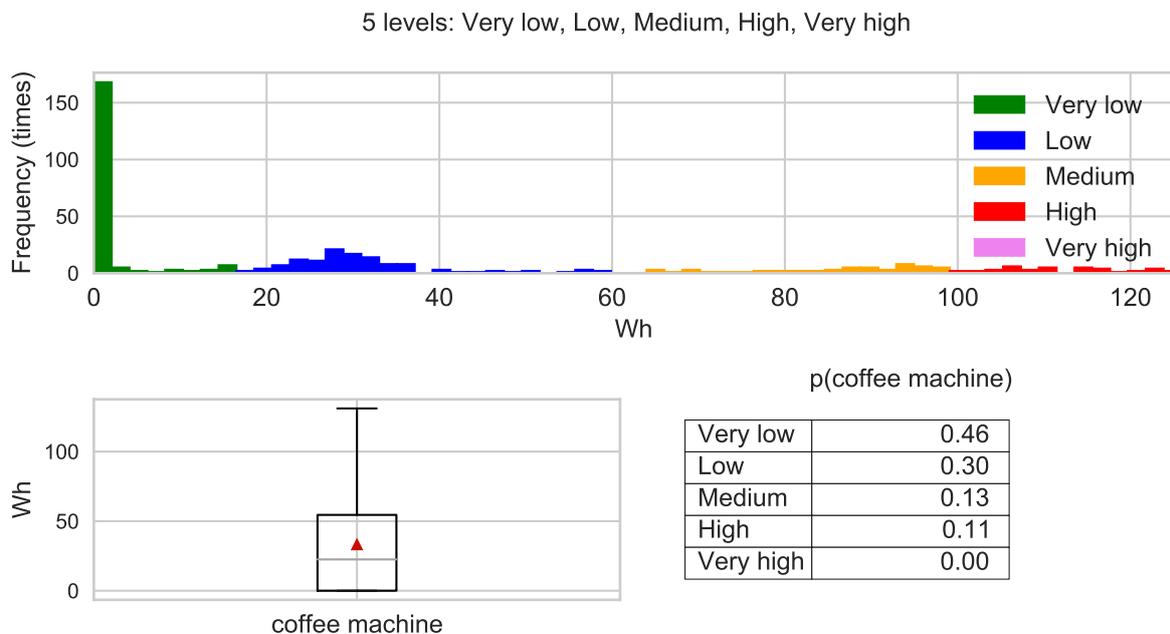


Figure 6.8 – The estimated CPDs, distributions of values and visualisation of the of AR consumption of **the coffee machine** for **cooking breakfast** in the Cestas household

The figures show that an appliance can be used differently for cooking breakfast. The coffee machine consumes average of 40 Wh when occupants cook breakfast, it consumes a very low AR consumption level 46% of the time, a low level 30% of the time and 24% of time for others. The toaster consumes average of 25 Wh for cooking breakfast, it consumes a very low AR consumption level 40% of the time, a low level 33% of the time and 27% for others. On the other hand, the microwave consumes more AR electricity than above appliances with an average of 100 Wh, it consumes a very low AR consumption level 27% of the time and only 12% of the time for high or very high levels.

An appliance can be used differently depending on activities. For example, figure 6.11 and figure 6.12 respectively present estimated CPDs of the microwave for cooking lunch and cooking dinner. They show that only the microwave is used frequently in both these activities. However, its AR consumption for cooking dinner is more than for cooking lunch. In particular, the microwave consumes an average of 180 Wh for cooking lunch while the average AR consumption for cooking dinner is 210 Wh. It can be explained that occupants use this appliance for a longer duration in the evening, when they need to prepare more food than in the lunch.

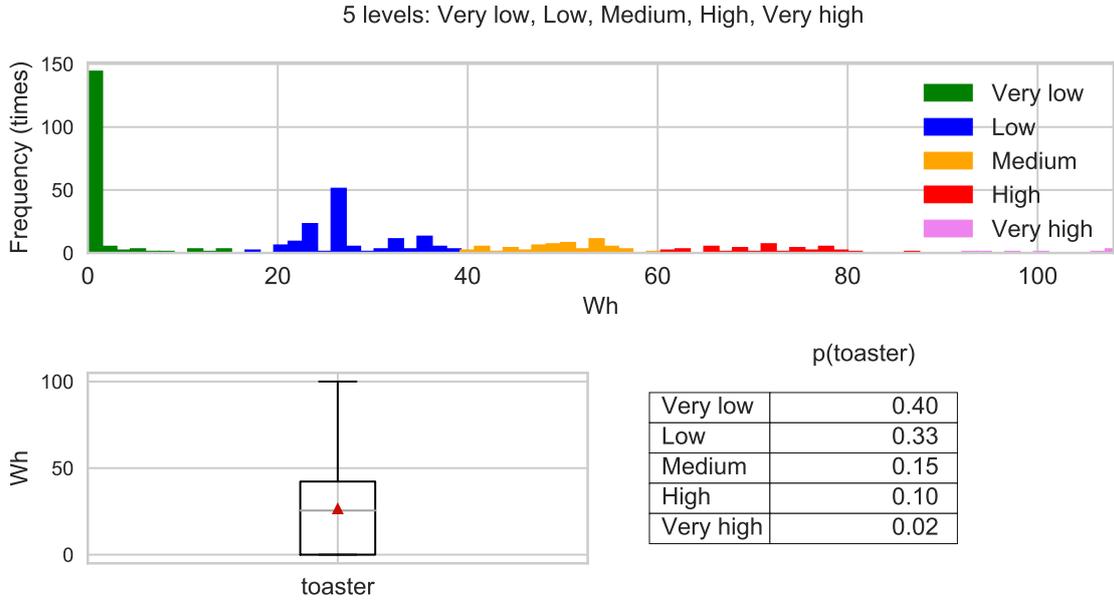


Figure 6.9 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the toaster** for **cooking breakfast** in the Cestas household

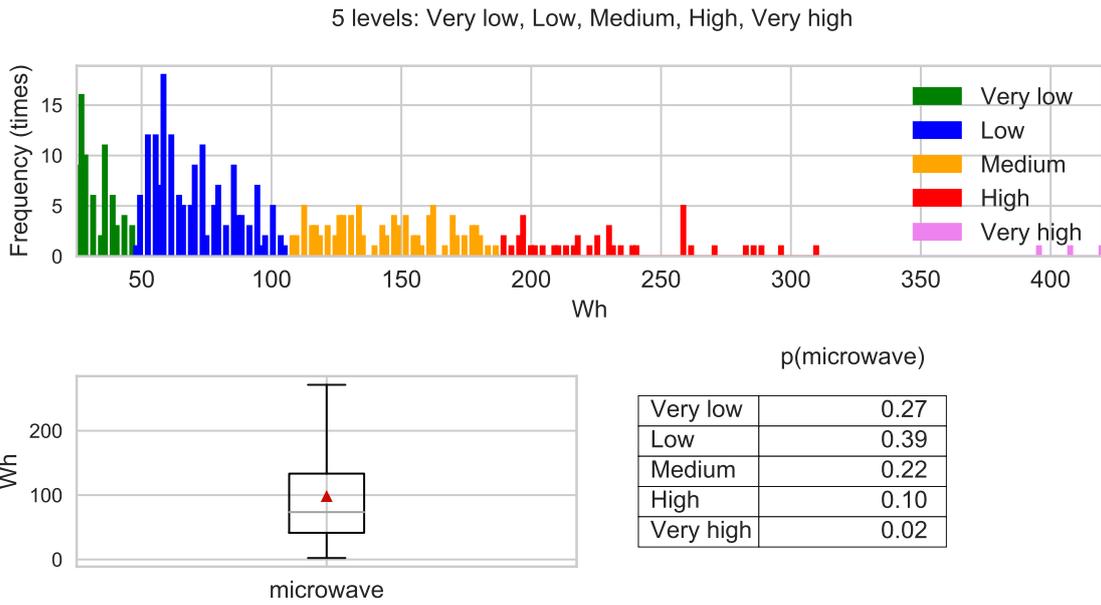


Figure 6.10 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the microwave for cooking breakfast** in the Cestas household

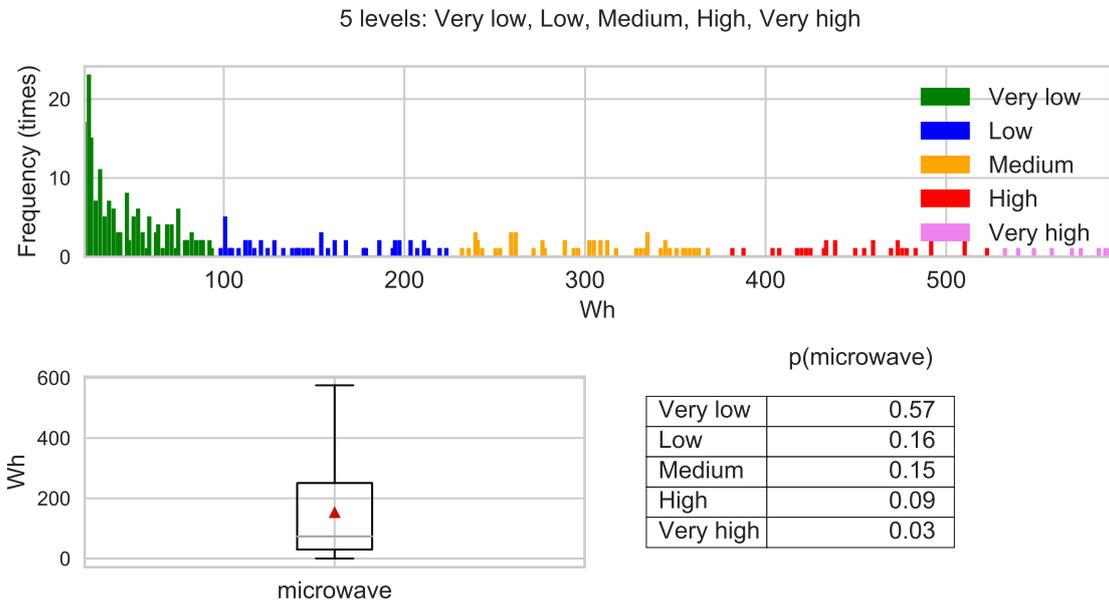


Figure 6.11 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the microwave for cooking lunch** in the Cestas household

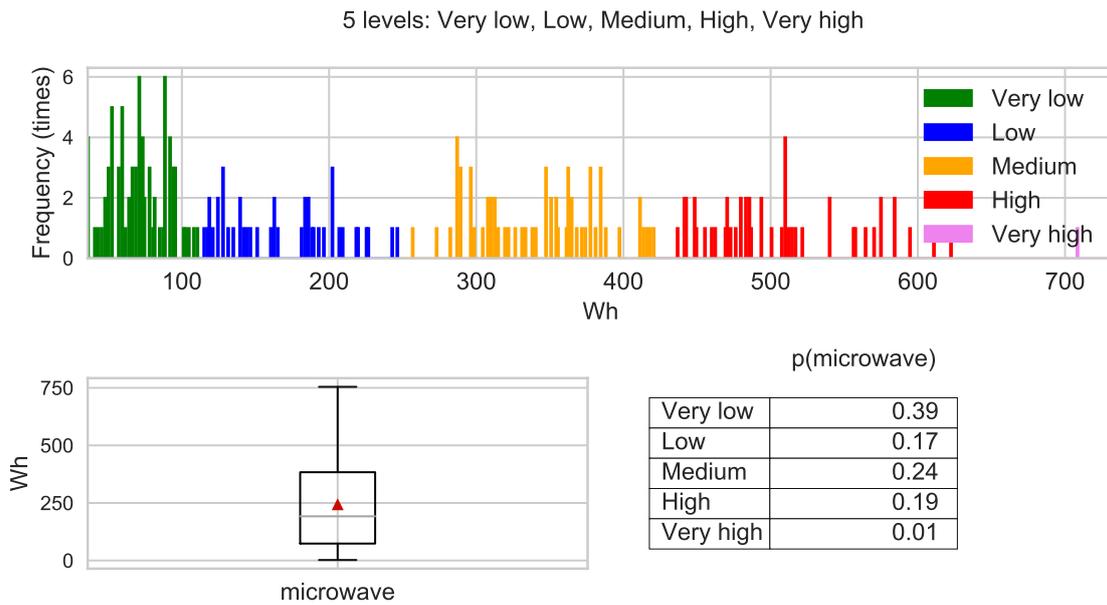


Figure 6.12 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the microwave** for **cooking dinner** in the Cestas household

An appliance can be used frequently in an activity but it is rarely used in other activities. For example, figure 6.13 and figure 6.14 present estimated CPDs of the coffee machine for cooking lunch and for cooking dinner. It is shown that the coffee machine is rarely used for these kinds of cooking while it is more frequently used for cooking breakfast (see figure 6.8).

During an activity, particularly in the three kinds of cooking activity (cooking breakfast, cooking lunch and cooking dinner), occupants spend different durations using appliances and it leads to various electricity consumptions.

This section implemented the method to determine the conditional probability distributions between the AR consumption level of an appliance and a given activity. In the next section, these distributions are used to simulate the AR consumption of the appliances.

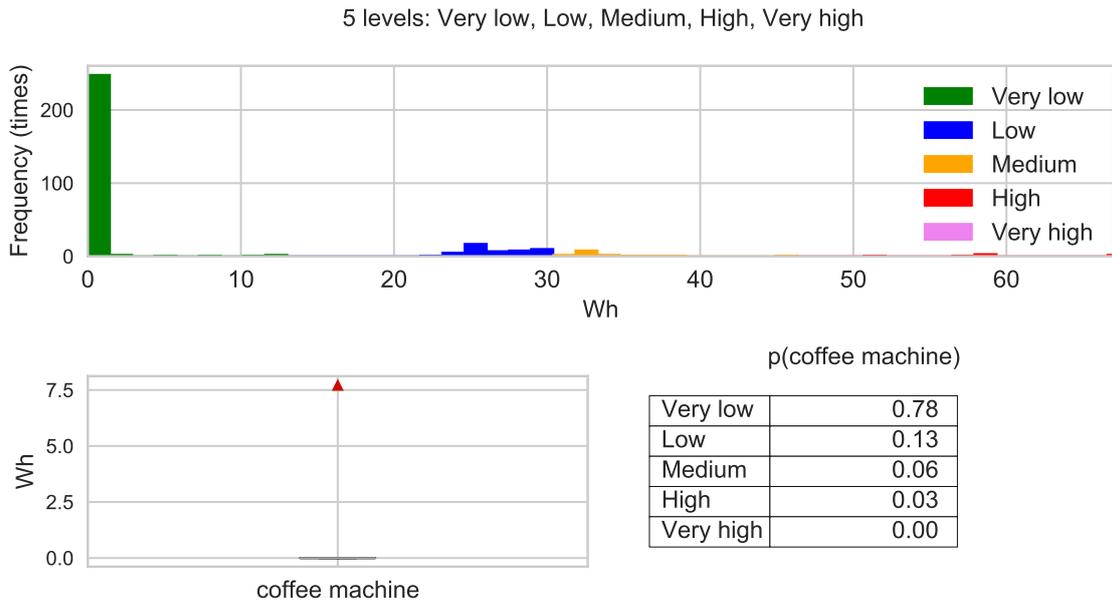


Figure 6.13 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the coffee machine** for **cooking lunch** in the Cestas household

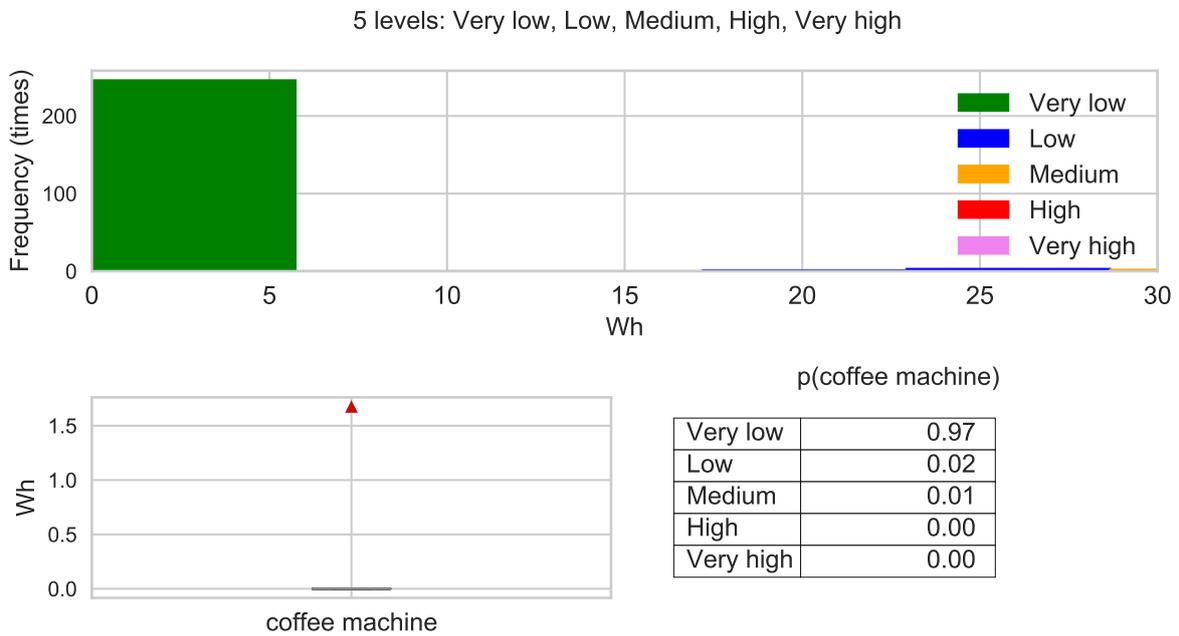


Figure 6.14 – The estimated CPDs, distributions of values and visualisation of the AR consumption of **the coffee machine** for **cooking dinner** in the Cestas household

Modelling the activity related electricity (AR) consumption of appliances

This section focuses on predicting the AR consumption of appliances for occupants' activities. As mentioned in the methodology, the collected labels of occupant activities are directly used to assess the performance of the AR consumption simulation itself, rather than its combination with an activity simulation model.

In previous steps, the microwave, the toaster and the coffee machine were determined as the available appliances related to the three kinds of cooking, the Bayesian Network structure to predict the AR consumption of appliances that is dependent on the different kinds of cooking activity is presented in figure 6.15. The structure is simple because the available related appliances are limited. In fact, to cook in the Cestas household, occupants usually use many appliances such as microwave, oven, hotplate, kettle, etc. It is interesting to study the AR consumption of these appliances for cooking. However, just some appliances were available with measurements so that we only studied the AR consumption of these appliances.

The structure is used for both the AR consumption model for cooking breakfast, the AR model for cooking lunch and the AR model for cooking dinner. However, their parameters (CPDs) are estimated separately and they are different. In each timestep, once the cooking breakfast, cooking lunch or cooking dinner is detected, the network is used to predict the AR consumption levels of each appliance using estimated CPDs.

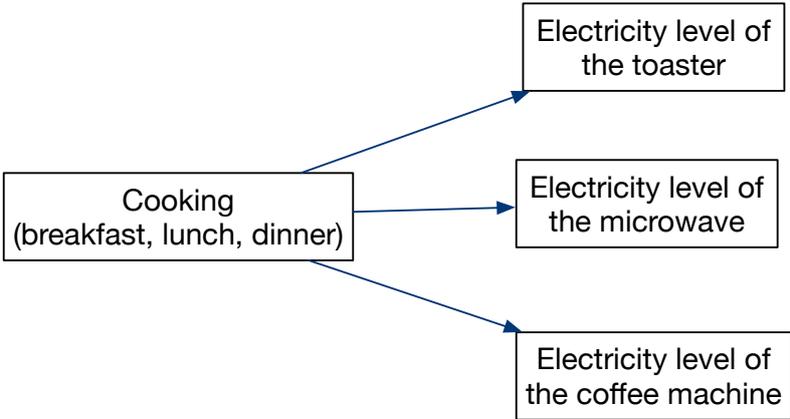


Figure 6.15 – The Bayesian Network structure for predicting the electricity consumption related to the cooking activity in the Cestas household

Evaluation of the model

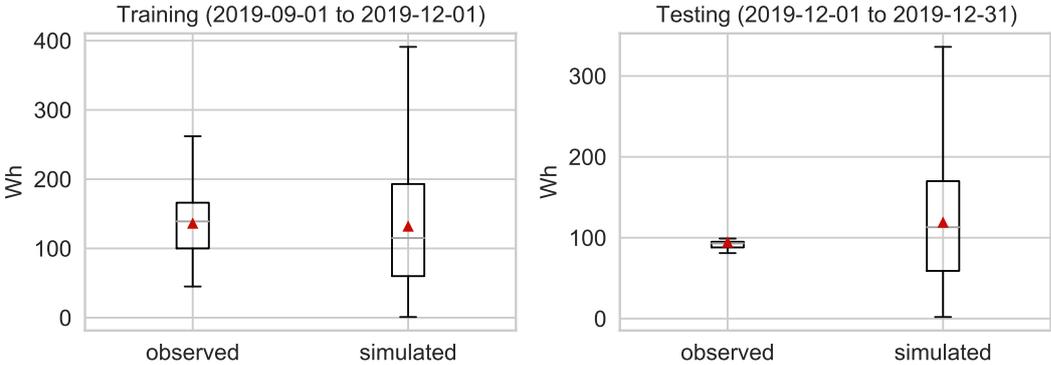
As seen in figure 6.14, the coffee machine was rarely used for cooking dinner. Therefore, the uses of the coffee machines for this kind of cooking were not considered. Table 6.3 presents the estimated RE of simulated AR consumption of appliances used in this case study. The result shows that the AR consumption of many appliances can be predicted accurately with low REs (-3% for the microwave in cooking lunch, 2% for the microwave in cooking dinner, -6% for the toaster in cooking breakfast, *etc.*

Table 6.3 – Testing RE of the simulated AR consumption of appliances in the Cestas household

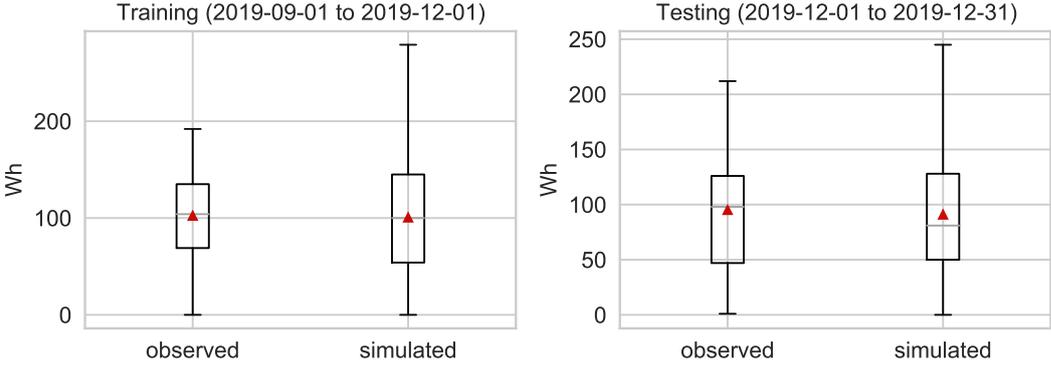
Activity	Microwave	Toaster	Coffee machine
Cooking Breakfast	-13%	-6%	20%
Cooking Lunch	-3%	-14%	1000%
Cooking Dinner	2%	-12%	-

In addition to the RE, the variance of the AR consumption of appliances needs to be considered. To do it, boxplots are used to illustrate comparisons between observed and simulated daily AR consumption of the appliances for the different kinds of cooking activity. The result of the cooking breakfast is presented in figure 6.16. It can be shown that the model is useful in both reproducing but also simulating the appliance uses and their variety of appliances' uses, which are represented by the similar boxplots in both training data and testing data. In particular, for cooking breakfast, the microwave actually consumes an average of 400 Wh/day with 75% of consumption values lying above 250 Wh/day. The average simulated AR consumption of this appliance is 380 Wh/day. More than 75% of consumption values are greater than 200 Wh/day. The results of the toaster are very good when both its average AR consumption and its variance are very similar between simulated and observed. The toaster actually consumes an average of 100 Wh/day while the simulated one is 97Wh/day. Most of the simulated values and observed values of this appliance are between 50 Wh/day and 130 Wh/day. The simulated and observed average AR consumption for the coffee machine are very similar, but the simulated values are much more spread out than the observed values. Similarly, comparisons of the cooking lunch activity and the cooking dinner activity are respectively presented in figure 6.17 and figure 6.18. It can be

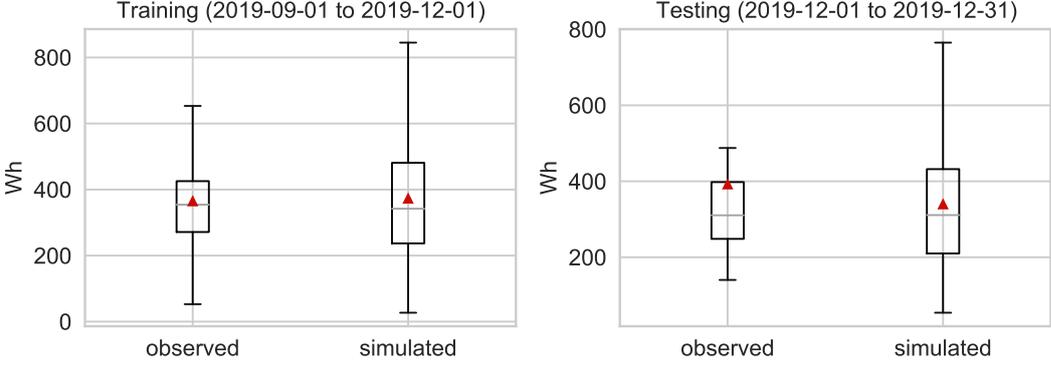
seen that the microwave was used more in cooking dinner than other kinds of cooking (cooking breakfast and cooking lunch). Simulated values of the AR consumption of the microwave are similar to the observed values in both cooking lunch and cooking dinner. Most of the AR consumption of the toaster is used for cooking breakfast.



(A) coffee machine

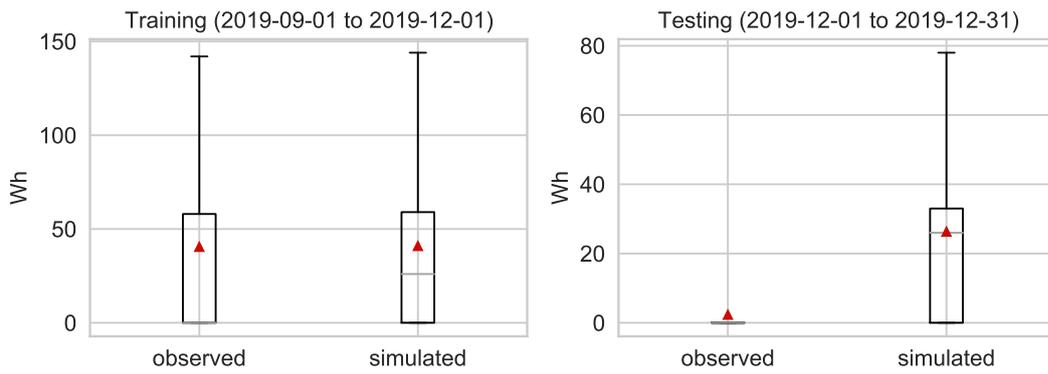


(B) toaster

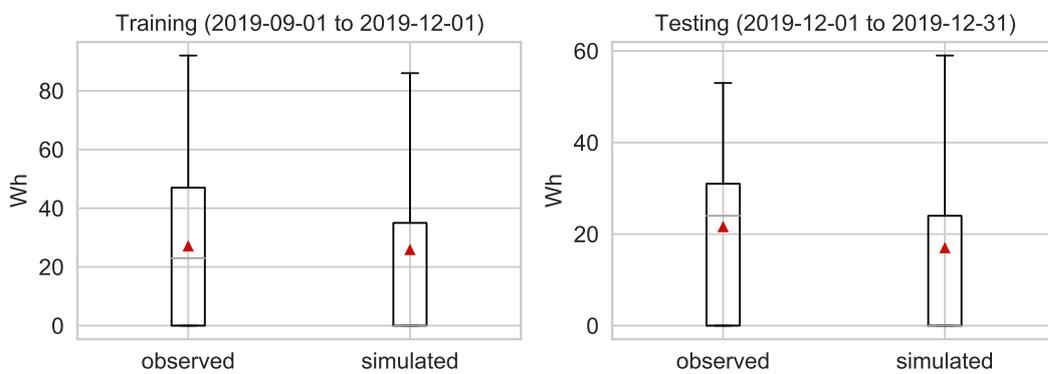


(C) microwave

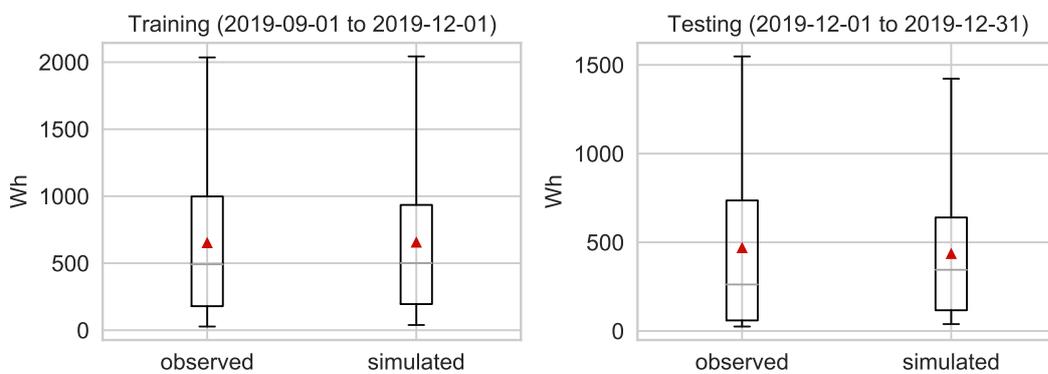
Figure 6.16 – Comparisons between simulated and observed daily AR consumption of appliances for **cooking breakfast** in the Cestas household (A: coffee machine, B: toaster, C: microwave)



(A) coffee machine

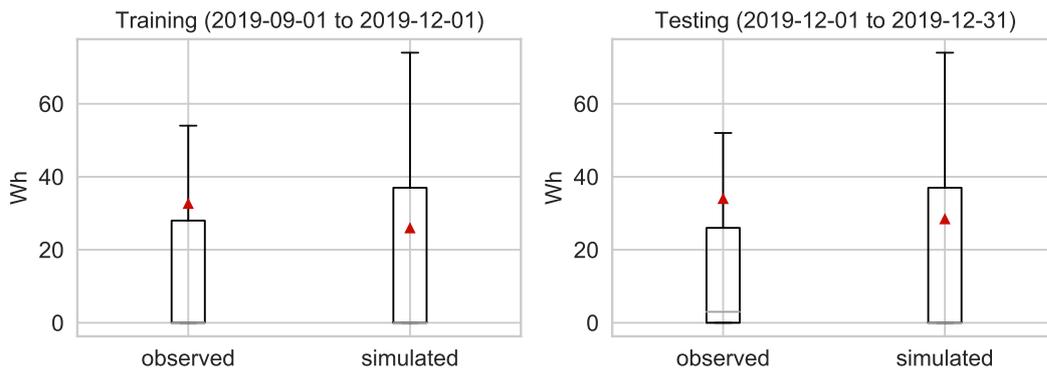


(B) toaster

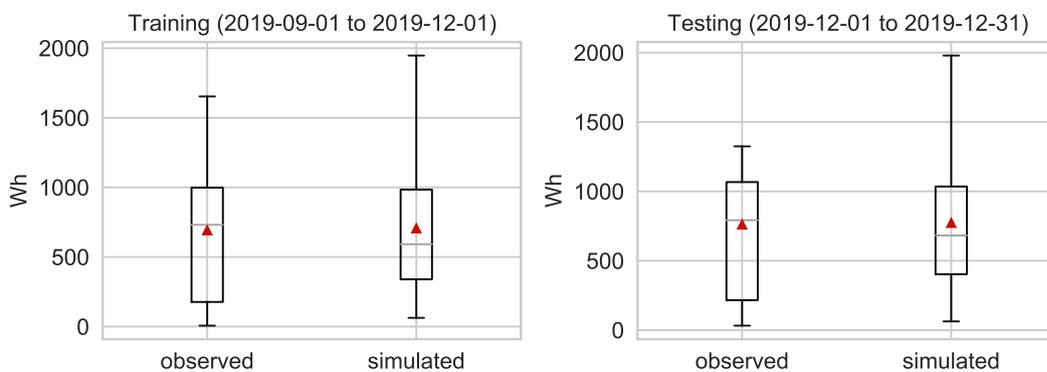


(C) microwave

Figure 6.17 – Comparisons between simulated and observed daily AR consumption of appliances for **cooking lunch** in the Cestas household (A: coffee machine, B: toaster, C: microwave)



(A) toaster



(B) microwave

Figure 6.18 – Comparisons between simulated and observed daily AR consumption of appliances for on the **cooking dinner** in the Cestas household (A: toaster, B: microwave)

However, the model did not work with the coffee machine used for cooking dinner. The simulated daily electricity consumption was eleven times higher than the observed one (ER was 1000%). It is necessary to investigate factors causing the gap. As described in the methodology, the model works based on combining activities labels and estimated CPDs. Because the observed labels were used so that the gap between simulated and observed electricity consumption was caused by differences between estimated CPDs and actual CPDs, which represent ways of occupants using appliances for their activities. Figure 6.19 presents CPDs of the coffee machine in the cooking lunch during the training and testing periods. It shows that occupants changed their ways of using the coffee machine for cooking lunch. During training periods (September, October and November 2019), the coffee machine was used in preparing the lunch. In the testing period (December 2019), occupants did not use or rarely used the coffee machine for this activity. Because the model was built from the training data, it is difficult for the

model to handle this change, leading to incorrect predictions.

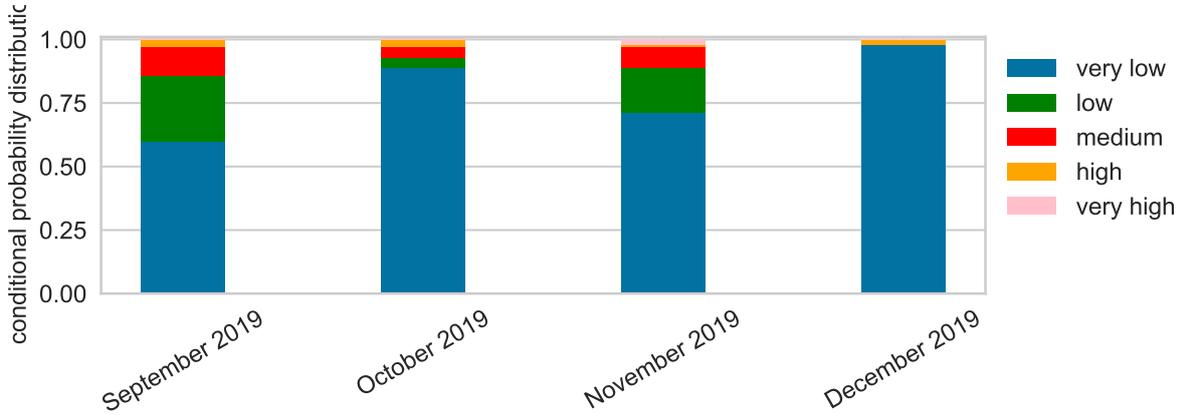


Figure 6.19 – Comparisons of conditional probability distributions of AR consumption levels of **the coffee machine** for **cooking lunch** between months in the Cestas household.

Contrary, the model can work well if the change is not significant. For example, the model accurately predicted the electricity consumption of the microwave for cooking lunch, in which the CPDs in the training data were similar to the ones in the testing data (see in figure 6.20).

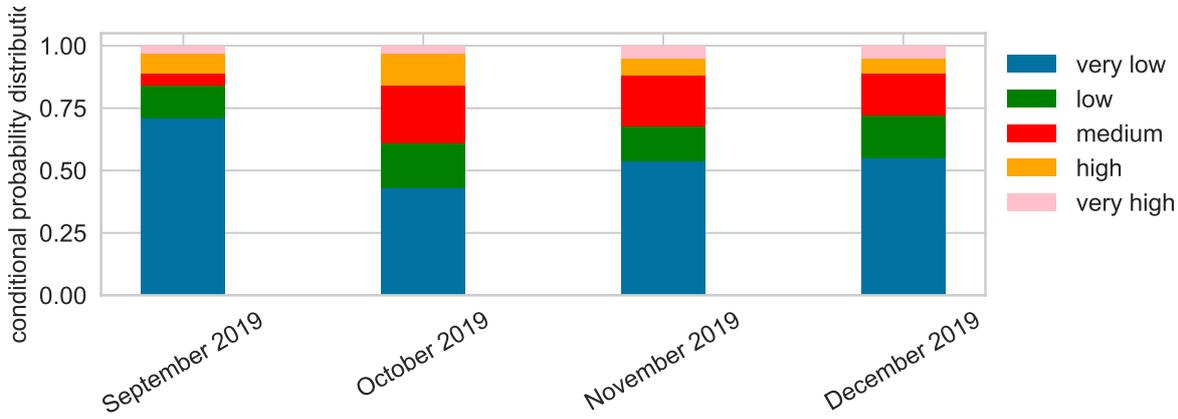


Figure 6.20 – Comparisons of conditional probability distributions of AR consumption levels of **the microwave** for **cooking lunch** between months in the Cestas household.

Table 6.4 summarizes testing REs corresponding to several other activities (washing clothes, drying clothes, washing dishes and entertaining) and their related appliances in the Cestas household.

Table 6.4 – Testing RE of the simulated electricity consumption of related appliances in several activities in the Cestas household

Appliance	Activity	Location	RE
Washing machine	Washing clothes	Laundry room	-25%
Tumble dryer	Drying clothes	Laundry room	-21%
Dishwasher	Washing dishes	Kitchen	-5%
Audio+video	Entertaining	Living room	-70%

Statistical comparisons between simulated and observed daily AR consumption of appliances related to these activities are presented in figures 6.21, 6.22, 6.23 and 6.24. It is shown that the model can work well with most of the considered activities and their related appliances. In particular, the washing machine actually consumes an average of 700 Wh/day for washing clothes while the simulated one is 550 Wh/day. The actual values of this appliance are more spread out than the simulated values. The tumble dryer consumes a lot of energy for drying clothes in this case study. It consumes an average of 3000 Wh/day and 75% of the time that it consumes more than 1500 Wh/day. The results of the dishwasher for washing dishes are similar to the observed values.

However, results of the audio and video devices for entertaining are poor. It overestimates the AR consumption of these appliances both in training data and testing data. When checking the data for these appliances (audio and video devices), we found that a lot of electricity consumed by them was not for the entertaining activity. Using only the entertaining activity can not be enough to simulate their electricity consumption.

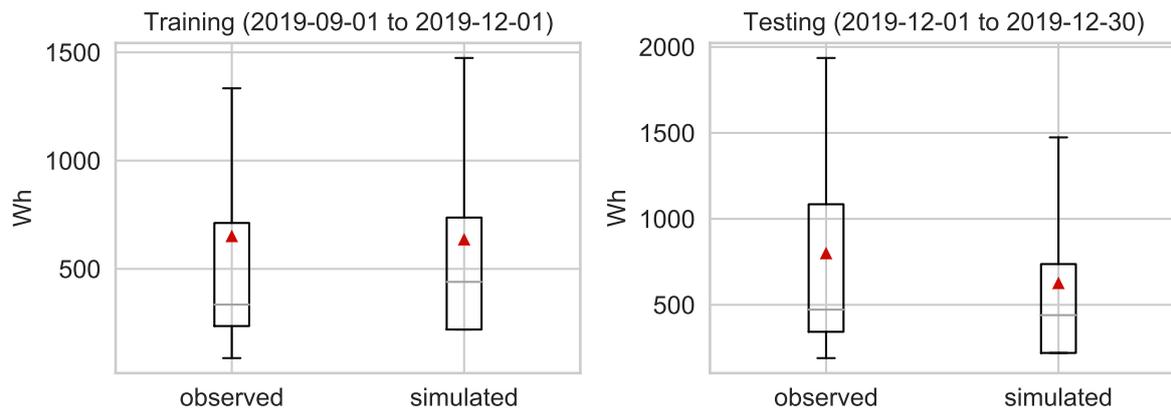


Figure 6.21 – Comparisons between simulated and observed daily AR consumption of **the washing machine for washing clothes** in Cestas household

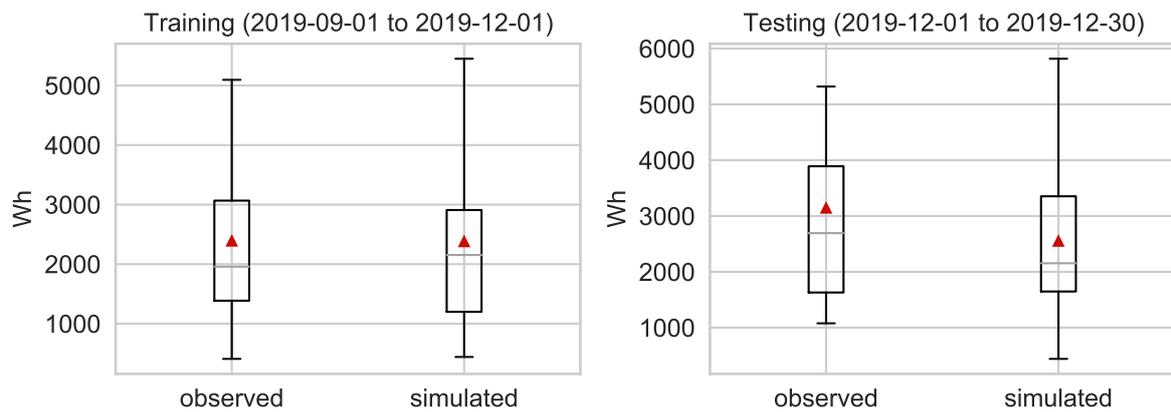


Figure 6.22 – Comparisons between simulated and observed daily AR consumption of **the tumble dryer for drying clothes** in Cestas household

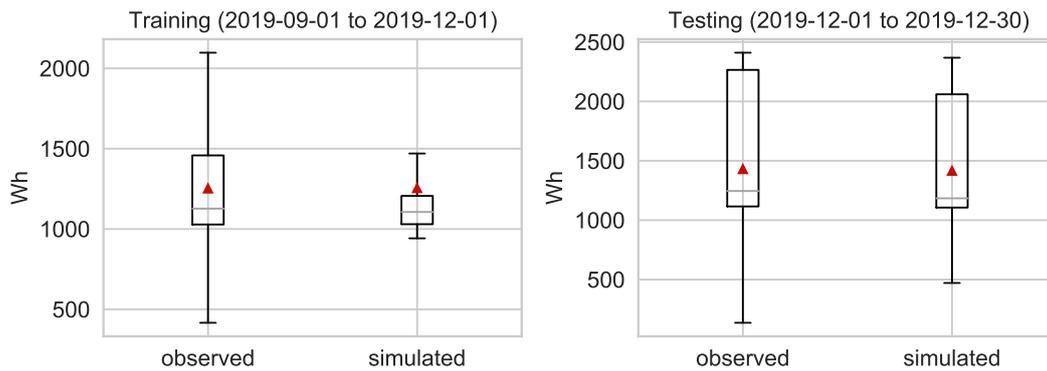


Figure 6.23 – Comparisons between simulated and observed daily AR consumption of **the dishwasher for washing dishes** in Cestas household

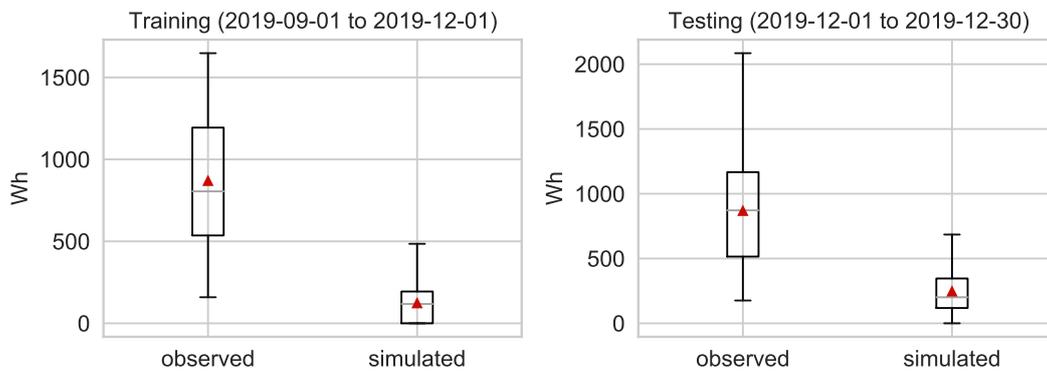


Figure 6.24 – Comparisons between simulated and observed daily AR consumption of **the audio and video for entertaining** in Cestas household

6.5 Conclusion

Occupants' activities play an essential role in explaining the discrepancies in the electricity consumption between similar houses. In a house, occupants can do many actions and they can modify the electricity consumption of appliances. Determining these actions is challenging but activities provided by occupants can give contexts of the actions carried out. While an appliance can consume electricity at many timesteps in a day, a certain amount of its consumption is impacted by occupants' activities through actions. Thus, determining the activity related (AR) consumption of appliances is a necessary task to assess the energy impacts of occupants' activities. Many statistical models have been proposed to deal with this task. The models are based on statistical data including many households with different characteristics, locations, *etc.* Therefore,

their outputs are the average of many households and not specific for particular dwellings.

To tackle this issue, this chapter proposed a methodology to predict the AR consumption of appliances when their related activities were identified in specific households. Questionnaires were used to identify domestic appliances related to occupants' activities. Relations between an activity and the AR consumption of an appliance were also discussed. For a given activity, the AR consumption of an appliance can be induced during the activity or after the activity. Then, an AR consumption model based on a Bayesian Network was proposed. This model aims to predict the AR consumption of appliances when activities are known. To evaluate the model, the model was tested with several activities and their associated appliances in the Cestas household. The results show that occupants can use an appliance in different ways for their activities in this house. For instance, a coffee machine was usually used for cooking breakfast but it was rarely used for cooking lunch or cooking dinner. Furthermore, the AR consumption of an appliance can be different between days. For example, the daily AR consumption of a washing machine in the house can fluctuate between 250 Wh and 750 Wh while the one for a dishwasher is between 1 kWh to 1.5 kWh. In addition, the results showed that when occupants' activities were identified, the model could reproduce and predict the AR consumption of appliances and its diversity in this house with REs between -25% and 20% in most cases and statistical comparisons. However, there are some limitations in the model:

- Some analyses show that appliances can change their AR consumption for a given activity over time and the model is limited in handling it. But, the model can be useful to investigate these changes.
- The AR consumption model aims to predict the total AR consumption of appliances in each timestep. It does not consider the continuity between timesteps.

In the chapter 7, the AR consumption models will be combined with the activity profiles (determined from the occupant activity models) to simulate the electricity consumed by appliances in periods when occupants' activities could not be known, which is useful in evaluating retrofitting actions in a building renovation.

Chapter 7

Application to the design phase in building renovation strategy

Contents

7.1 Introduction	148
7.2 Contextualized occupants' activities in evaluations of building renovation strategies	148
7.2.1 Motivations	148
7.2.2 Objective	150
7.3 Simulation of occupants' activities and their impacts on related appliances	151
7.3.1 Problem statement	151
7.3.2 General methodology	152
7.3.3 Generating occupants activity profiles using estimation model	153
7.3.4 Determining impacts of an activity on related electrical appliances	154
7.3.5 Simulating electricity consumed by related electrical appliances	154
7.3.6 Validation	157
7.4 Implementation and results	157
7.4.1 Case study: Cestas	158
7.4.2 Case study: Bordeaux	170
7.5 Conclusion	175

7.1 Introduction

Energy simulations can be conducted to analyze and evaluate building designs based on predefined criteria. Using optimization algorithms, these simulations support finding optimal solutions among a large number of candidate designs and settings (Tian et al., 2018). Thus, accurately simulating energy simulation has benefits in facilitating simulated changes to many facets of building and providing results showing the effects of such changes on energy usage (Shiel et al., 2018). It is also fundamental to design low-energy buildings, renovate existing buildings and thus promote sustainability (Ascione et al., 2021).

Besides energy consumed for heating, cooling or domestic hot water needs, many scientists also consider the electricity consumed by activity-dependent domestic appliances (microwave, oven, washing machine, television, *etc.*) (Wilke, 2013; Aerts, 2015; Vorger, 2014; Zaraket, 2014). Though, most of the models are stochastic models, which has a limitation in verification (mentioned in chapter 2 and chapter 6). To tackle it, this chapter proposes a verifiable activity-dependent model to simulate electricity consumed by domestic appliances in a specific house.

7.2 Contextualized occupants' activities in evaluations of building renovation strategies

7.2.1 Motivations

Renovating building is a process to retrofit, restore, rehabilitate, and renovate aspects of existing buildings (after their initial construction and exploitation) to improve their looks and energy performance (Santangelo et Tondelli, 2017; Regnier et al., 2018). The building renovation can have significant impacts on the future heating strategies, cooling strategies and patterns of electricity demands (Sharif et Hammad, 2019).

According to Regnier et al. (2018), building retrofits can generally have many

benefits: saving building's electricity consumption, reducing global greenhouse gas emissions, reducing energy costs and improving indoor environmental quality (thermal comfort, indoor air quality, protection against external noise, *etc.*) Lu et al. (2021) also mentioned that building renovation is an effective approach to reduce the energy consumption and identify energy-saving opportunities based on building conditions, types, and functionalities. In addition, it can have a relatively lower environmental impact and life cycle cost compared to new constructions (Lee et al., 2019). There are many potential strategies for a building renovation. The simplest and popular approach is to upgrade existing equipments to improve their energy efficiency (Regnier et al., 2018). Additionally, many actions can be performed on building envelopes, HVAC systems or lighting systems in order to retrofit the building energy efficiency (Sharif et Hammad, 2019; Pannier et al., 2021).

A building renovation generally has six main steps: building assessment, recommissioning, goal setting, determining the level of retrofit, selection of suitable retrofitting interventions, and implementation (Lu et al., 2021). Within these steps, the selection of a renovation strategy is a complex task. There could be many possible renovation solutions that can be performed on a building envelope and systems. Their performance can be affected by building actors (building designer, investor, owner, tenant, *etc.*), by constraints or by people preferences (Taillandier et al., 2016; Pannier et al., 2021; Lu et al., 2021).

To support the decision-making in selecting retrofitting options, BES softwares are popularly used to predict the impacts of these retrofitting actions on the building energy performance (Regnier et al., 2018; Lu et al., 2021; Yan et al., 2015). These softwares are used to simulate the energy performance of buildings with proposed retrofitting solutions. Comparisons between different simulated variants are considered to select one or several optimal solutions. Thus, the accuracy of BES software could affect the quality of selected renovation strategies. As discussed in chapter 1, occupants' activities are useful to reduce gaps between simulated energy consumption and actual consumption. Thus, these activities are useful in evaluating possible renovations strategies. While many statistical models have been proposed to consider occupants' activities in BES, their results are average and not specific to particular households (as discussed in chapter 2 and chapter 3). Thus, when applying BES to evaluate renovation

strategies in an existing house, it could be interesting to use contextualized activities and their energy corresponding impacts than using average results from statistical models.

7.2.2 Objective

As discussed above, contextualized activities could be more beneficial than representative activities in selecting renovation strategies. However, when evaluating a retrofit action, we cannot know future occupants' activities after it. In a house, occupants tend to repeat their successful activities and they could become patterns of activities with a timetable fairly regular (Xuan Hoa Binh et al., 2010). Historical data of occupants' activities are useful in investigating these patterns and their impacts on energy consumption.

As mentioned in previous chapters, occupants' activities can affect energy consumption in different ways. For example, occupants can affect operations of HVAC systems due to their presence. They can use electrical appliances to support their activities or have influences on indoor air quality through their actions on doors/windows, *etc.* However, their energy impacts on HVAC systems or indoor air quality were not measured in our case studies. Thus, this study focuses only on how occupants' activities affect the consumption of electrical appliances, which can be measured directly by power sensors.

In this chapter, we propose a methodology, which is based on measurements and historical activities labels, to simulate occupants' activities and the electricity consumption of related appliances in a specific house. To do it, the estimation model proposed in chapter 5 and the activity related (AR) electricity consumption model proposed in chapter 6 are combined. This methodology could provide useful occupant-related inputs to BPS softwares in order to support the selection of renovation strategies (see in figure 7.1).

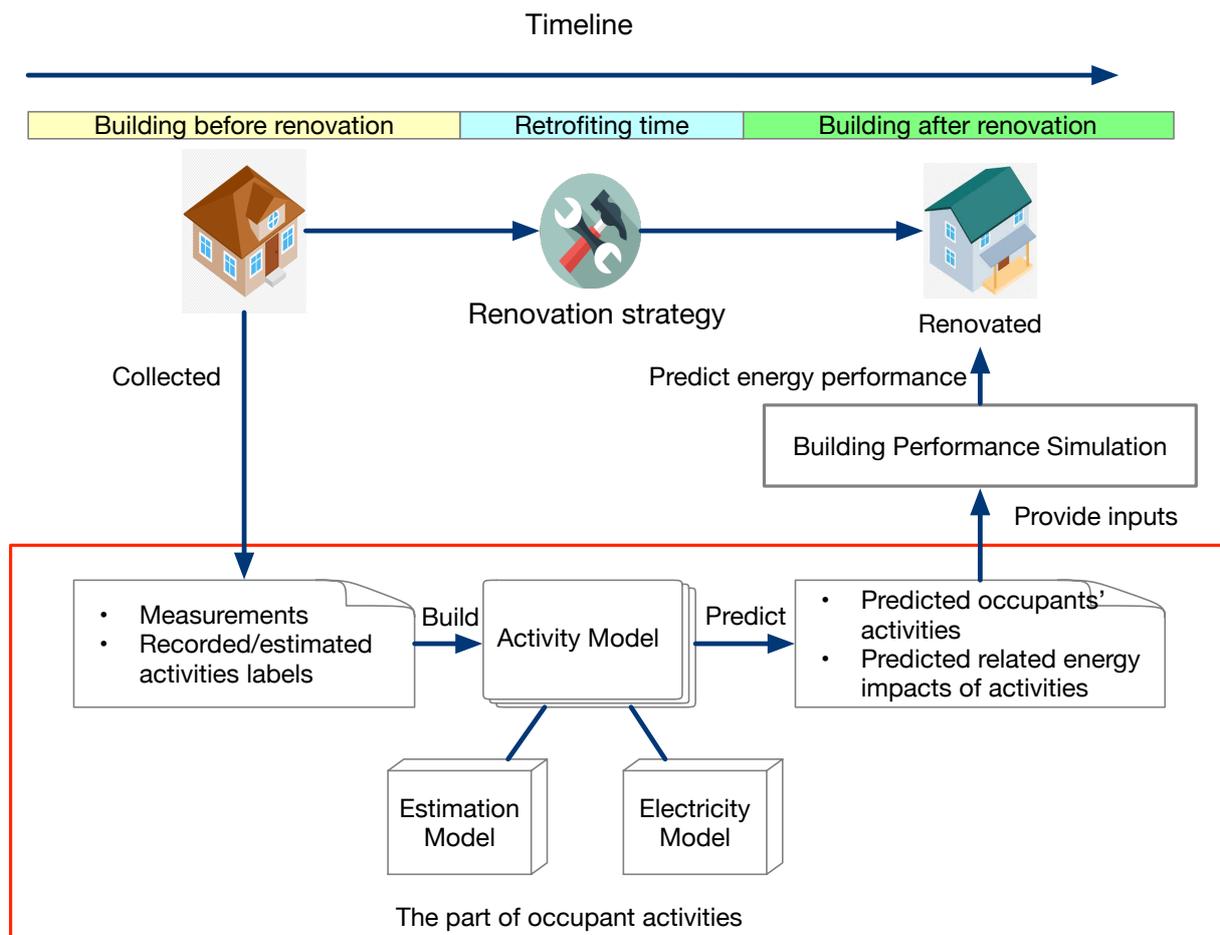


Figure 7.1 – Simulation of contextualized activities and its application in building renovation

7.3 Simulation of occupants' activities and their impacts on related appliances

7.3.1 Problem statement

When building models supporting renovating an existing house, there are generally three periods:

- before renovation time (historical period);
- renovating time (model design period);
- after renovation time (simulation period).

During renovating time, occupants' activities and their related electricity consumption in renovated time are not known. To tackle this issue, this work aims to:

- Build a model from data collected in historical period, which includes two following main components:
 - activity profiles: which contain probabilities for activities to happen at each timestep. These probabilities are used for the simulation of occupant activities.
 - impacts of occupants' activities on related electrical appliances: which are used to link the simulated activities to AR electricity consumption of appliances.
- Use built model to simulate occupants activities and AR electricity consumption of appliances in simulation period.

7.3.2 General methodology

Figure 7.2 presents the methodology to build models from a historical data and combine them to simulate activity related electricity consumption of appliances in a simulation period. The main tasks in the methodology can be described as follows:

1. **Generating activity profiles from the historical data:** from historical measurements and collected activity labels, an activity estimation model is built with the methodology proposed in chapter 5. As discussed in that chapter, collecting labels in a long period is challenging. Using only collected labels in a short period can be insufficient to determine activity profiles. The activity estimation model can be used to determine labels in periods in which measurements were available but labels were missing. The trained activity profiles are then determined from the estimated labels;
2. **Determining impacts of the activity on related electrical appliances from the historical data:** from historical measurements and collected activity labels, an AR electricity consumption model is built with the methodology proposed in chapter 6. This model contains the relationship between occupant activities and related domestic appliances;
3. **Simulating AR electricity consumption of related electrical appliances in the simulation period:** the determined trained activity profiles are used to simulate occupant activity labels in the simulation period. These simulated labels are

combined with the determined impacts of the considered activity on related appliances to simulate their electricity consumption.

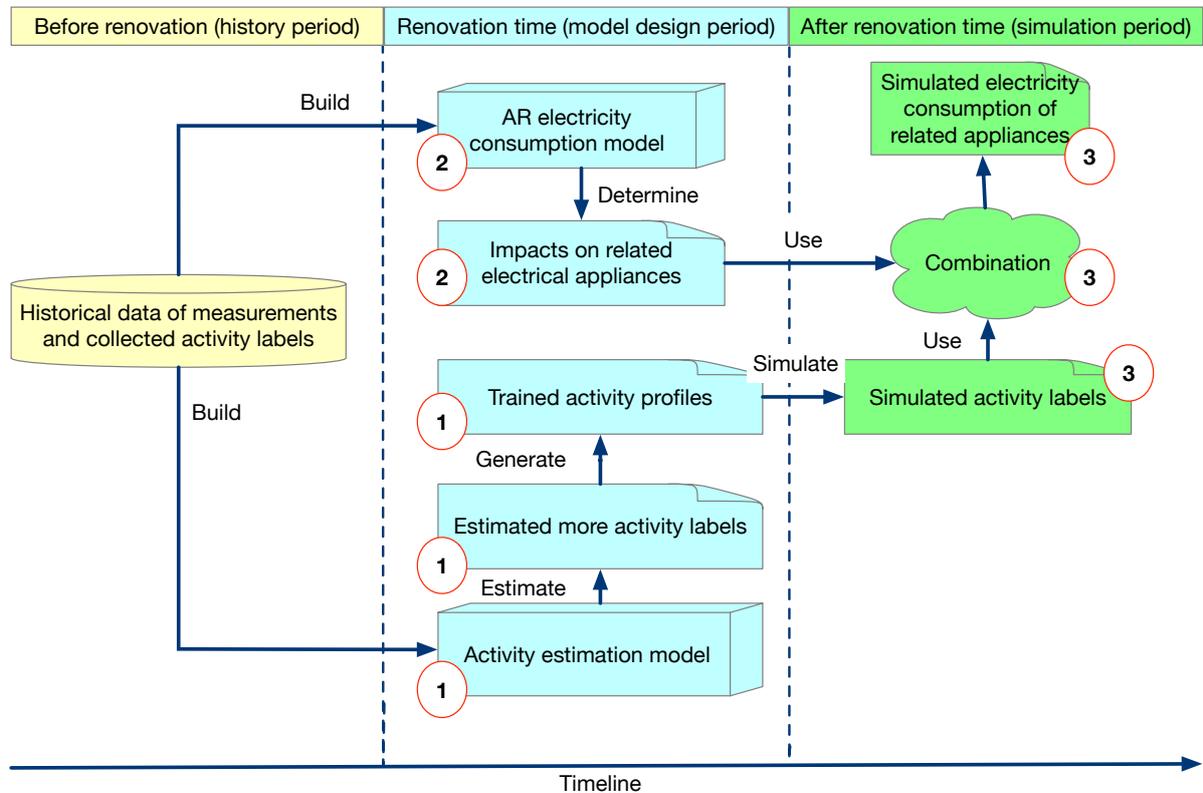


Figure 7.2 – Proposed methodology to simulate the electricity consumed by appliances

7.3.3 Generating occupants activity profiles using estimation model

This first task aims to determine trained activity profiles from activity labels estimated from the activity estimation model. These profiles can be used in simulating occupant activity labels in energy simulation softwares. Many studies stated that the time of the day and the day of the week could affect occupant activities in houses (Tanimoto et al., 2008; Page et al., 2008; Wilke et al., 2013; Vorger, 2014). In another study, Aerts (2015) applied hierarchical clustering to investigate the profiles of occupant activity in Belgium households. She also pointed out that their activities are different between weekdays and weekends. Their jobs and their habits can contribute to this diversity.

In this work, the timestep of the day and the type of day (weekday, weekend) are used in modelling occupant activity profile. When collecting labels, several activities can be in the same timestep and they are collected from many occupants. It is challenging

to identify sequences of these activities. For simplicity, previous activities are not considered when we model an activity independently. Figure 7.3 presents a conceptual model representing the relationship between the considered activity, the timestep and the type of day. Based on this conceptual model, in each timestep t of a day with the corresponding type of day d , we estimate conditional probability distributions $p(a|t, d)$ for an activity a to be performed.

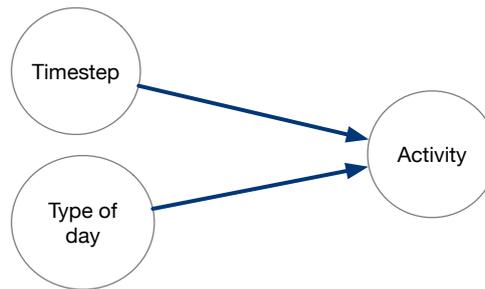


Figure 7.3 – A conceptual model of the relationship between an activity, a timestep and a type of day

7.3.4 Determining impacts of an activity on related electrical appliances

In this second main task, the AR electricity consumption model (proposed in chapter 6) is built from historical measurements and estimated activity labels. The objective is to determine the relationship between occupant activities and related electricity consumption of appliances. In particular, to consider an activity a and a related appliance with five AR consumption levels x (very low, low, medium, high, very high), conditional probability distributions $p(x|a)$ are determined. These distributions represent how the appliance consumes electricity to support occupants to perform the considered activity.

7.3.5 Simulating electricity consumed by related electrical appliances

Results of the two above tasks are combined in this third main task. The trained activity profiles with the distributions are used to simulate occupants' activities and the outputs of the electricity model are used to interpolate their related electricity consumption.

Figure 7.4 presents the main steps to do a simulation of an activity a and the AR electricity consumption level x of an appliance in a simulation period with a starting timestep t_{start} and a stopping timestep t_{end} :

1. In the beginning, the necessary distributions $p(a|t, d)$ and trained profiles $p(x|a)$ are determined based on historical dataset (measurements and estimated labels) using models described in the first task and the second task;
2. From the starting timestep t_{start} , in each index i with the timestep t_i and the type of day d_i , label of the activity $a(t_i)$ is simulated through the determined probability distribution $p(a(t_i)|t_i, d_i)$ and the Monte-Carlo method;
3. Then, the activity label is checked whether it happens at t_i :
 - (a) If a happens at t_i ($a(t_i) = 1$), for each relevant appliance x , the electricity level $x(t_i)$ of this appliance is simulated using the probability $p(x(t_i)|a(t_i))$ and the Monte-Carlo method;
 - (b) Then, the methodology moves to the next timestep $t_i + 1$;
4. The algorithm repeats the three previous steps until it reaches the stopping timestep ($t_i = t_{end}$).

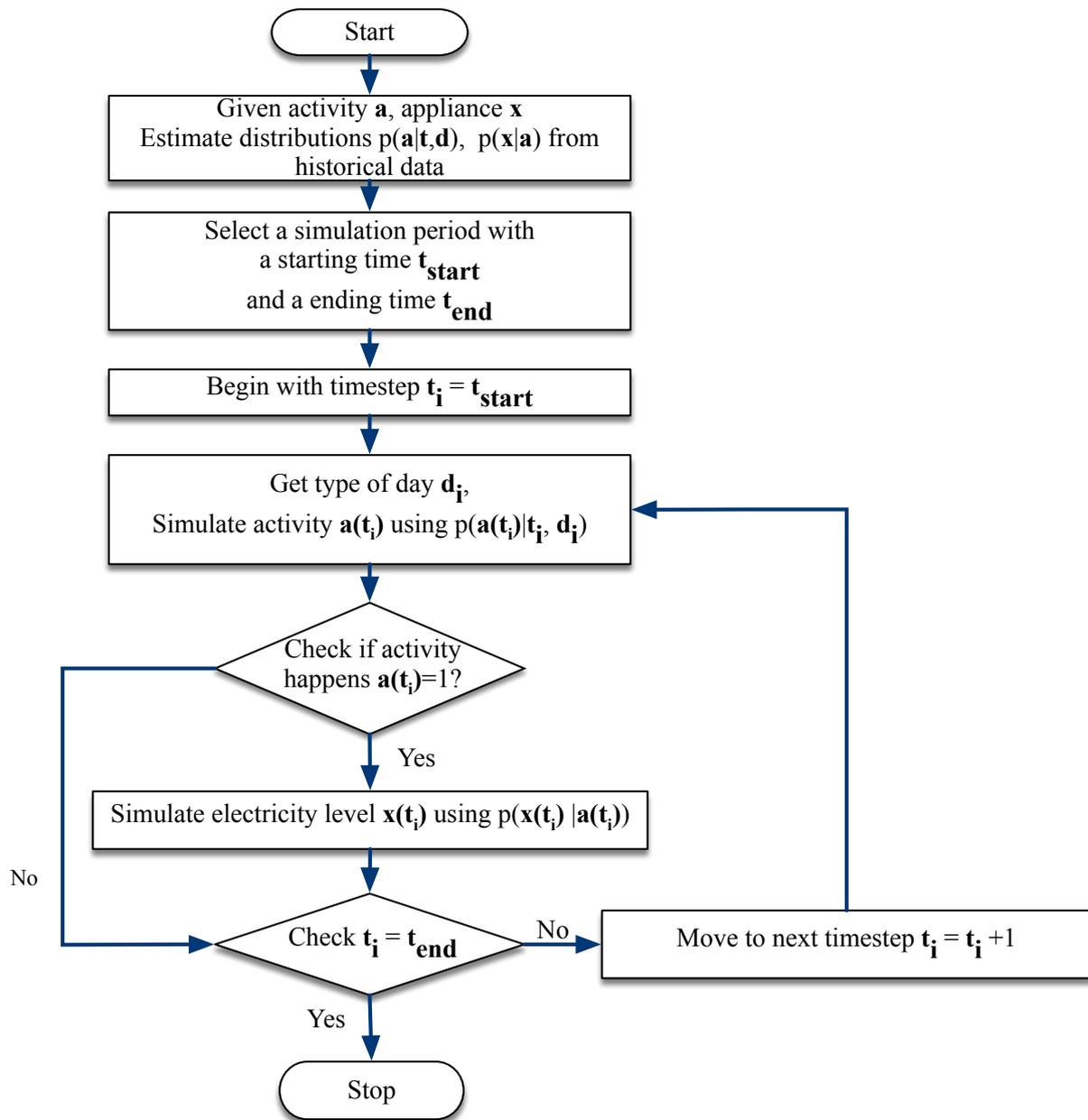


Figure 7.4 – Flowchart of the algorithm to simulate occupants activities and the AR electricity consumption of involved appliances

7.3.6 Validation

Firstly, we divide dataset into two parts: historical part and simulation part. The historical part is used to build models. Then, the built models are used to simulate occupants' activities and their AR consumption in the simulation part. To evaluate the methodology, we compare the simulated electricity consumption of each appliance to the measured one in the simulation part. Because the type of day (weekday, weekend) is considered in activity profiles, weekly electricity consumption is taken into account instead of daily electricity consumption. In this section, we use a relative error indicator (RE) to evaluate the proposed methodology as in equation 7.1.

$$RE = \frac{wx_{\text{simulated}} - wx_{\text{measured}}}{wx_{\text{measured}}} \times 100(\%) \quad (7.1)$$

with:

- *RE*: relative error is a value representing the difference between two values in percentage (%);
- $wx_{\text{simulated}}$: average simulated weekly electricity consumption;
- wx_{measured} : average measured weekly electricity consumption.

In addition to the ER, the diversity of the electricity consumption of appliances is important. To take it into account, statistical comparisons between them are performed and represented by boxplots.

7.4 Implementation and results

To test the methodology and its generalization, it was implemented in two case studies: Cestas and Bordeaux. In this section, to be consistent with previous chapters, the cooking activity and its related appliances are used to illustrate the methodology. In the Cestas household, a 3-months dataset from September 2019 to December 2019 is used to train the simulation model, and a 1-month dataset from January 2020 to February 2020 is used for testing. In the Bordeaux household, a dataset from April 05th 2021 to May 16th 2021 is used for training and a dataset from May 24th 2021 to June 28th is used for testing.

7.4.1 Case study: Cestas

Generating occupants activity profiles using estimation model

Generating trained activity profiles from the activity estimation model is the first step in the methodology (see figure 7.2). For the cooking activity in the Cestas household, three months dataset of occupant activities (from September 2019 to December 2019) is used to investigate its profiles.

Figure 7.5 illustrates these cooking profiles of every day of the week. The habits of cooking have similar patterns during the weekdays. The frequency of cooking in the morning is more than in the afternoon and the evening. That is represented by the higher cooking probability in the morning (from 06:00 to 11:00) compared to other periods of the day. Occupants tend not to spend much time cooking lunch from Monday to Friday, which is represented by the low probability of cooking between 11:00 and 17:00. It can be explained that household members work mostly at their workplace and do not come back home until the evening, except on Thursday. On Thursday, occupants spend more time cooking lunch than on the other days. It can be explained because lessons in schools and universities are finished earlier, so children come back home around noon. Due to a longer presence in the house on weekends, occupants have more time to cook lunch and dinner.

As mentioned in the methodology, this work considers two different occupant activity profiles: weekday and weekend profiles. Regarding the cooking activity, figure 7.6 shows these patterns in this case study. There are clear discrepancies between them, especially in cooking in lunch. These patterns include the distributions $p(a|t, d)$, which represent the probability for the cooking activity a to happen at the timestep t and the type of day d . For example, on a weekday, the probability that occupants cook at 8 a.m is 70% while this probability on the weekend is 40%.

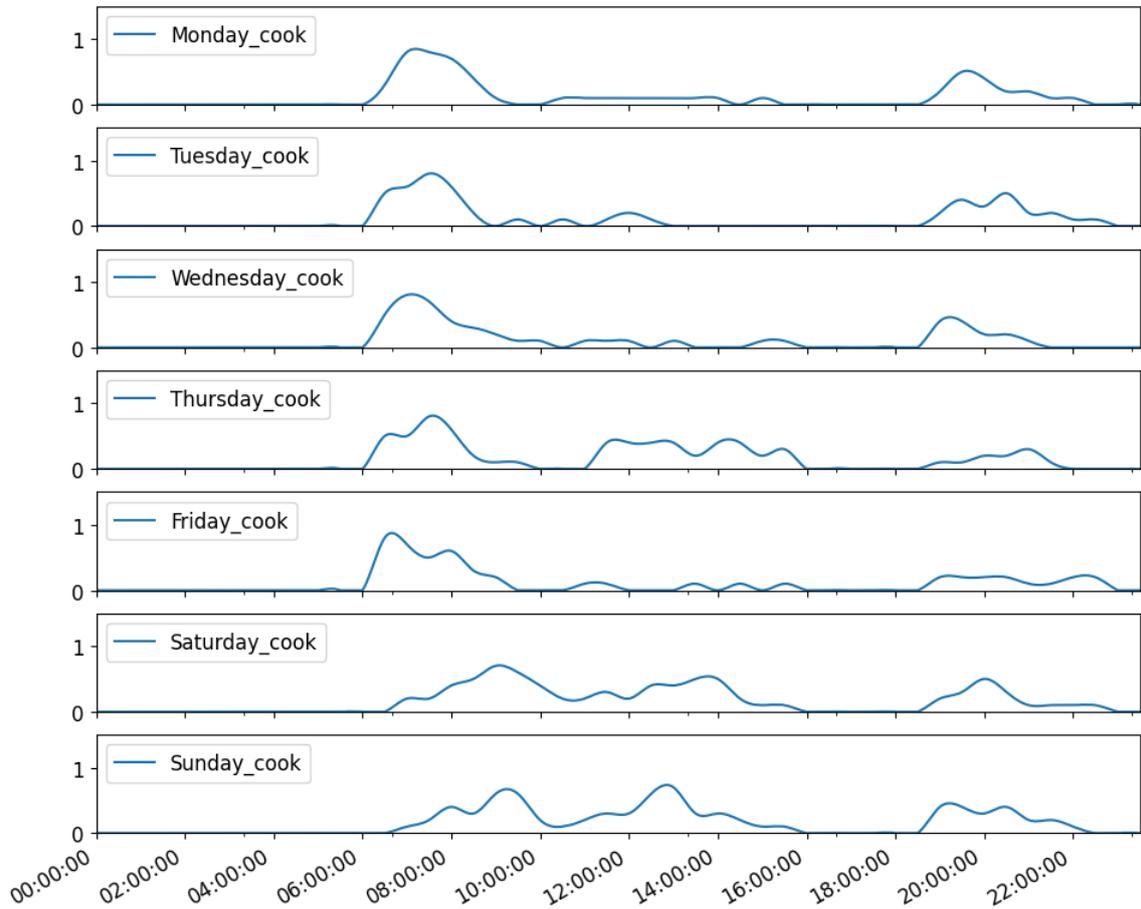


Figure 7.5 – Different cooking activity profiles between days in the Cestas household from 2019-09-01 to 2019-12-30

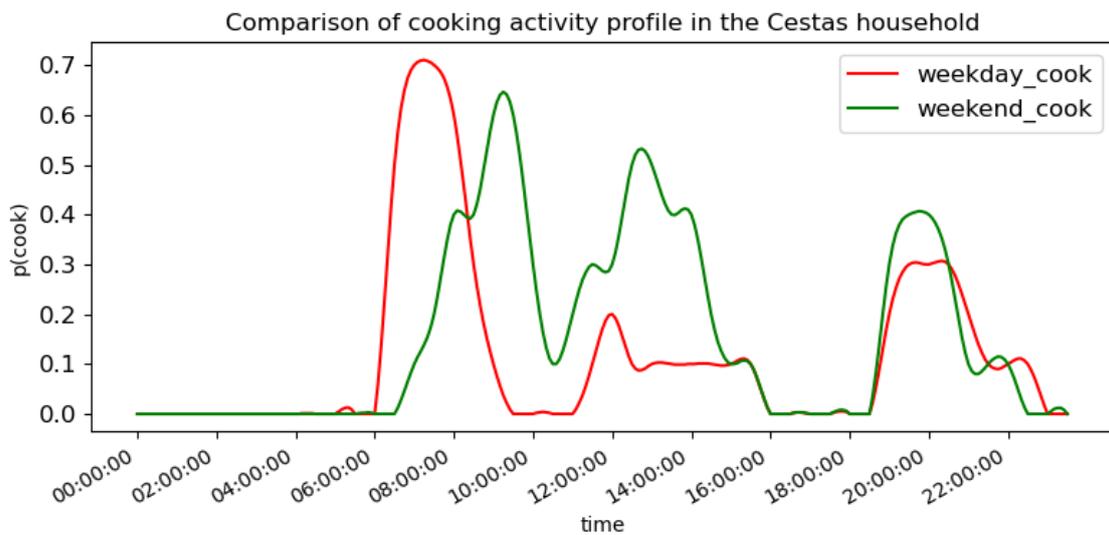


Figure 7.6 – Different cooking activity profiles between weekdays and weekend in the Cestas household from 2019-09-01 to 2019-12-30

Determining impacts of the activity on related electrical appliances

In the Cestas household, many appliances exist to support occupants cooking such as microwave, oven, toaster, hotplate, food processor, *etc.* However, as discussed in chapter 6, some of them were rarely used (food processor, yogurt maker, robot) or not measured (hotplate). For the microwave and the oven, because they were measured by the same power sensor, the microwave is used to represent both of them. So, we considered only three appliances related to three kinds of cooking activity (cooking breakfast, cooking lunch and cooking dinner) that were measured in this case study: the microwave, the toaster and the coffee machine. The impacts of each of them on the relevant appliances are studied independently. The results of this task were presented in chapter 6.

Simulating activity related electricity consumption and validation

Because the cooking activity is divided into three kinds (cooking breakfast, cooking lunch, cooking dinner), the energy consumed by an appliance is the total electricity consumed by it for each of them as in equation 7.2.

$$E(a)_{cook} = E(a)_{breakfast} + E(a)_{lunch} + E(a)_{dinner} \quad (7.2)$$

with $E(a)_{cook}$, $E(a)_{breakfast}$, $E(a)_{lunch}$, $E(a)_{dinner}$: the electricity respectively consumed by the appliance a for all kinds of cooking, for cooking breakfast, for cooking lunch and for cooking dinner.

Figure 7.7 shows comparison between measured and simulated weekly electricity consumption of the microwave. It is shown that the microwave actually consumes an average of 10 kWh/week with 75% of consumption values lying above 8.8 kWh/week. The average simulated result from the methodology for this appliance is 8.3 kWh/week so that the relative error is -17%. More than 75% of consumption values are greater than 6.8 kWh/week.



Figure 7.7 – Comparison between measured and simulated weekly electricity consumption of **the microwave** in the Cestas household from 2020-01-01 to 2020-02-01

The result of the toaster is presented in figure 7.8. The household consumes an average of 0.87 kWh/week for using the toaster for cooking. The consumption values of this appliance are quite constant and only fluctuate between 0.8 kWh/week and 0.9 kWh/week. The simulated values show a much higher dispersion than the measured values. The simulated values fluctuate between 0.22 kWh/week and 1.28 kWh/week. More than 75% of simulated values are below 0.9 kWh/week and the average value is 0.76 kWh/week, the corresponding relative error is -13%.

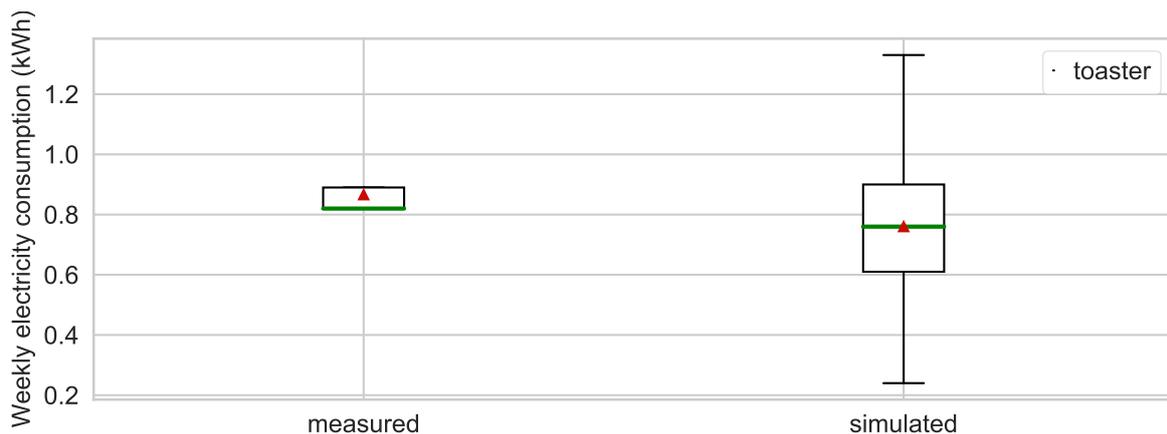


Figure 7.8 – Comparison between measured and simulated weekly electricity consumption of **the toaster** in the Cestas household from 2020-01-01 to 2020-02-01

The energy uses of the coffee machine in this household are quite similar to the toaster regarding weekly electricity consumption (see figure 7.9). Occupants consume

from 0.5 kWh/week to 0.83 kWh/week for the toaster and the average consumption is 0.75 kWh/Week. The simulated values have a higher dispersion than the measured values. The simulated values can vary between 0.1 kWh/week to 1.6 kWh/week. More than 75% of simulated values are above 0.7 kWh/week and the average value is 0.9 kWh/week with the relative error is 20%.

The summary of the results of the studied appliances related to the cooking activity in the Cestas household is presented in table 7.1 with their average value and standard deviation of the weekly consumption. The table shows that when the proposed methodology can simulate the weekly electricity consumption of most of the above appliances with a relative error between -17% and 20%. However, the simulations always have higher dispersion the measure values. The consumption of audio and video devices was underestimated with a relative error of -68%. As discussed in chapter 6, a certain amount of electricity consumed by audio and video appliances was not linked to occupants' activities. That caused the significant difference between simulated and measured electricity consumption of these appliances.

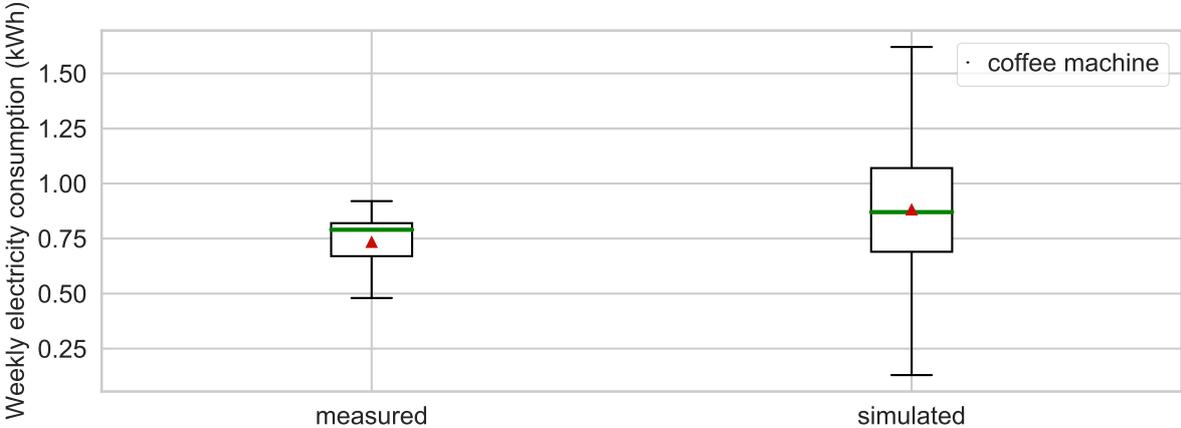


Figure 7.9 – Comparison between measured and simulated weekly electricity consumption of **the coffee machine** in the Cestas household from 2020-01-01 to 2020-02-01

Table 7.1 – Comparison between measured and simulated weekly energy consumption of studied appliances in the Cestas household from 2020-01-01 to 2020-02-01

Appliance	Measured	Simulated	Relative error
Microwave	10 ± 1.8 kWh	8.3 ± 1.9 kWh	-17%
Toaster	0.87 ± 0.18 kWh	0.76 ± 0.21 kWh	-13%
Coffee machine	0.75 ± 0.1 kWh	0.9 ± 0.26 kWh	20%
Tumble dryer	12.3 ± 3.5 kWh	9.8 ± 4.4 kWh	-20%
Dishwasher	7.9 ± 2.3 kWh	6.5 ± 2.7 kWh	-18%
Washing machine	3.0 ± 2.1 kWh	2.8 ± 1.6 kWh	-7%
Audio and Video devices	6.6 ± 1.2 kWh	2.1 ± 1.2 kWh	-68%

In addition to modeling weekly energy consumption, the methodology can generate average electricity patterns of appliances. Figure 7.10 depicts weekday and weekend average electricity consumption patterns of the coffee machine in the Cestas household. The figure shows that the simulated pattern has the same trend as the measured pattern on both weekdays and weekends. The patterns were strongly linked to the cooking activity in the morning. It was logical because we discussed that the coffee machine was mainly used in the morning. The comparison is similar to the toaster (see figure 7.11), which was also mainly used in the morning.

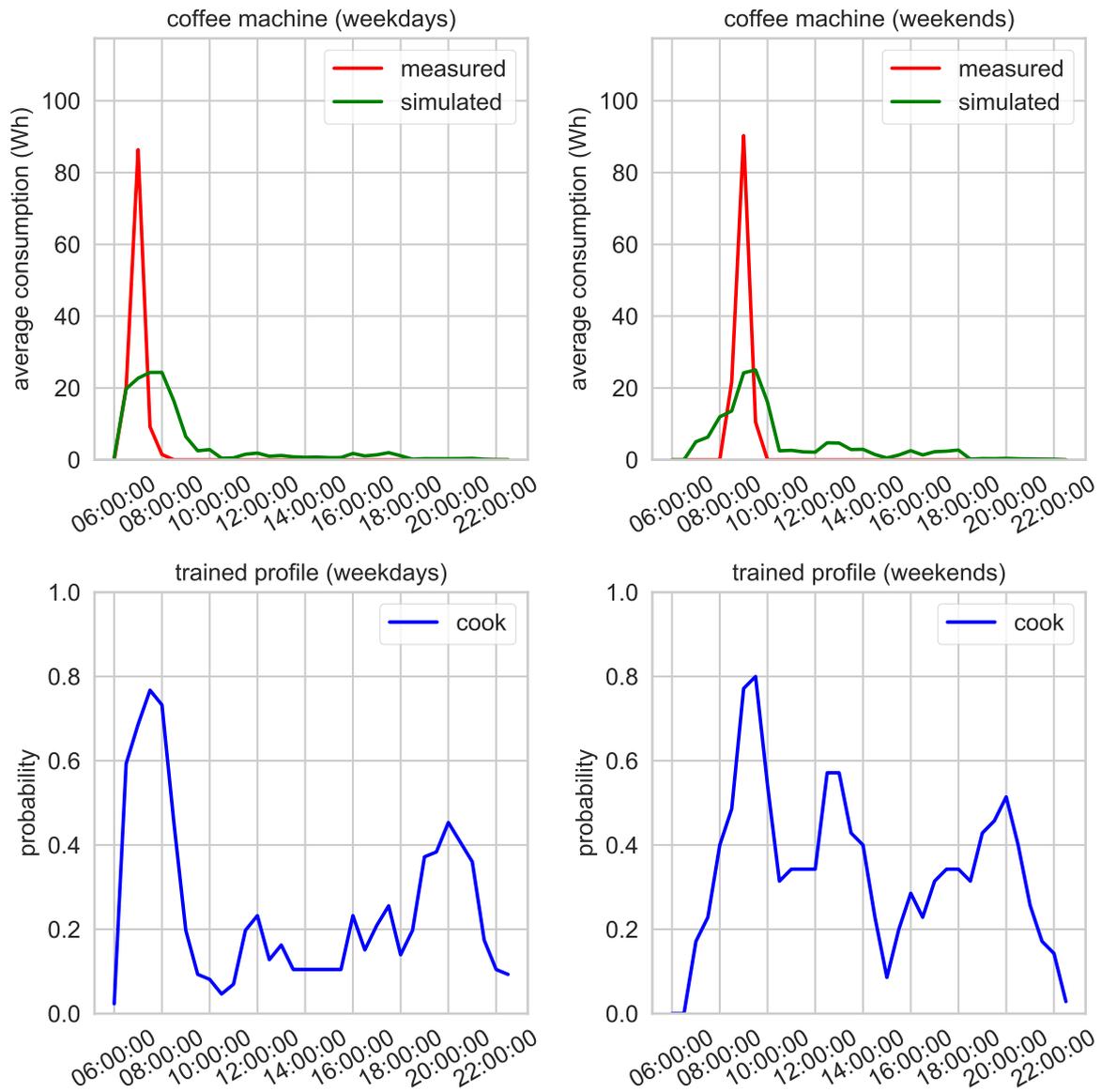


Figure 7.10 – Weekday and weekend average electricity consumption patterns for **the coffee machine** with **cooking** profiles in the Cestas household from 2020-01-01 to 2020-02-01

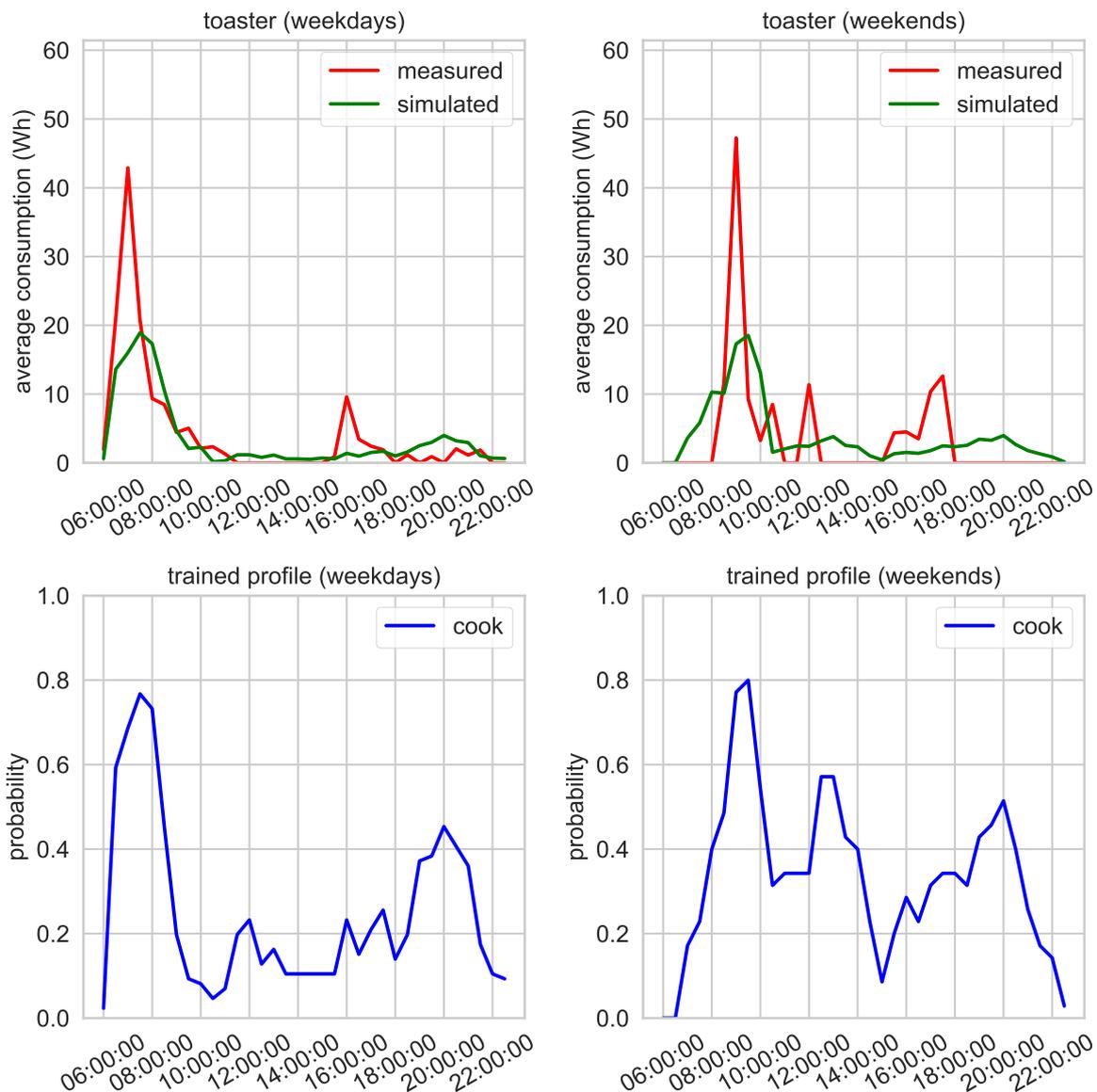


Figure 7.11 – Weekday and weekend average electricity consumption patterns for **the toaster** with **cooking** profiles in the Cestas household from 2020-01-01 to 2020-02-01

Figure 7.12 shows weekday and weekend average electricity consumption of the microwave. With this appliance, there are fewer discrepancies between simulated and measured on weekdays than weekends. The simulated weekday pattern is similar to measure one, which describes that the microwave was used more frequently for breakfast and dinner than for lunch. On weekends, there are more differences between the patterns, especially in lunch and dinner.

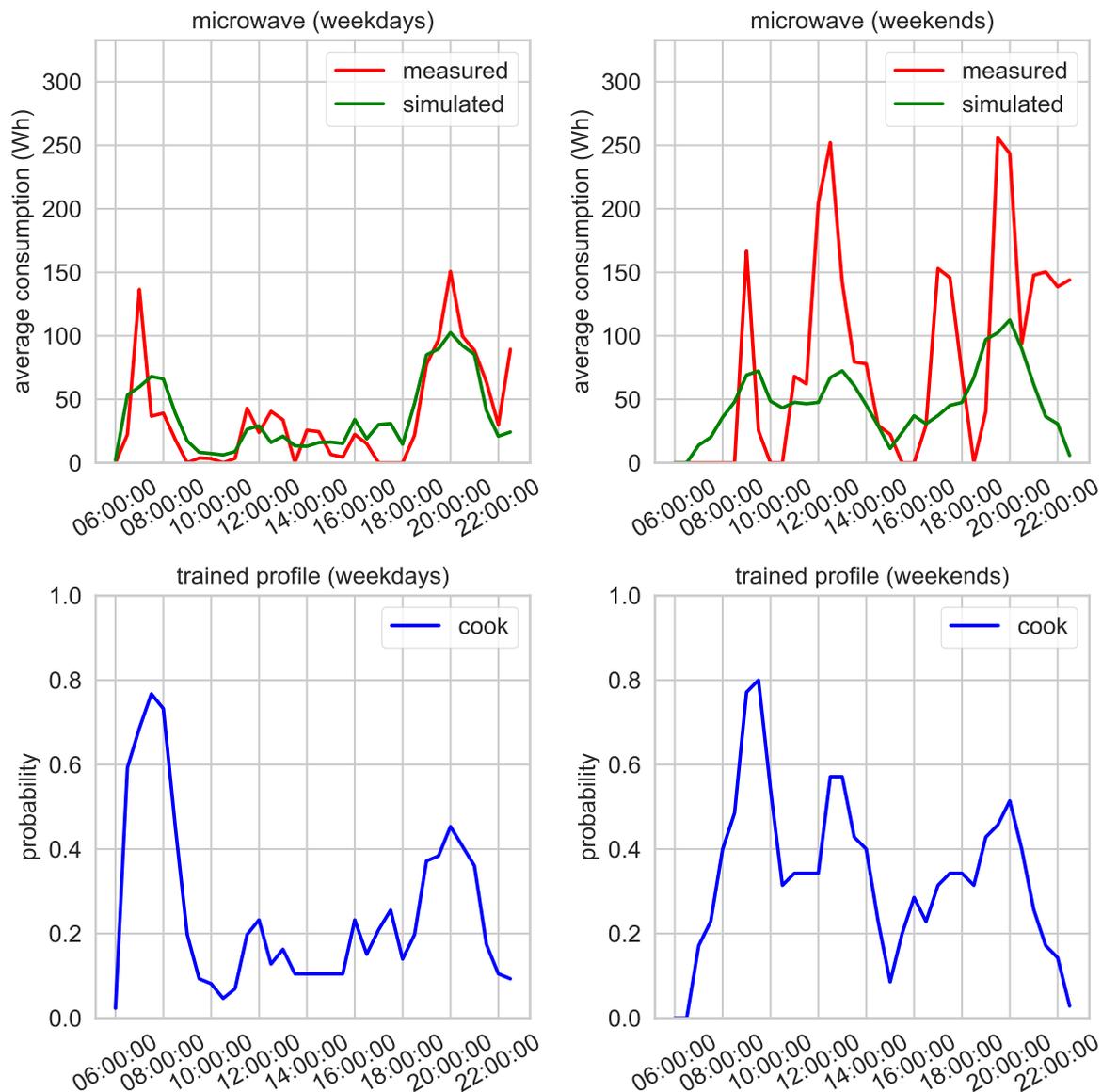


Figure 7.12 – Weekday and weekend average electricity consumption patterns for **the microwave** with **cooking** profiles in the Cestas household from 2020-01-01 to 2020-02-01

Figures 7.13, 7.14 and 7.15 respectively presents average electricity patterns of the tumble dryer, the washing machine and the dishwasher. In general, simulated patterns have the same trends as measured patterns on both weekends and weekdays. However, the simulated patterns did not capture peaks of measured values. Despite detecting one peak value, the simulations contain several lower values close to that value. It can be explained that occupants do not usually use the appliances at a fixed timestep but among several close timesteps associated with occupants' activities. There are more discrepancies between them on weekends. Two reasons can cause the discrepancies.

Firstly, occupants can change their activity profiles and it leads to modifications of the electricity profiles. Secondly, just a few weekend days (8 days) are used to generate the testing weekly electricity consumption patterns. Therefore, the patterns can have underfitting issues.

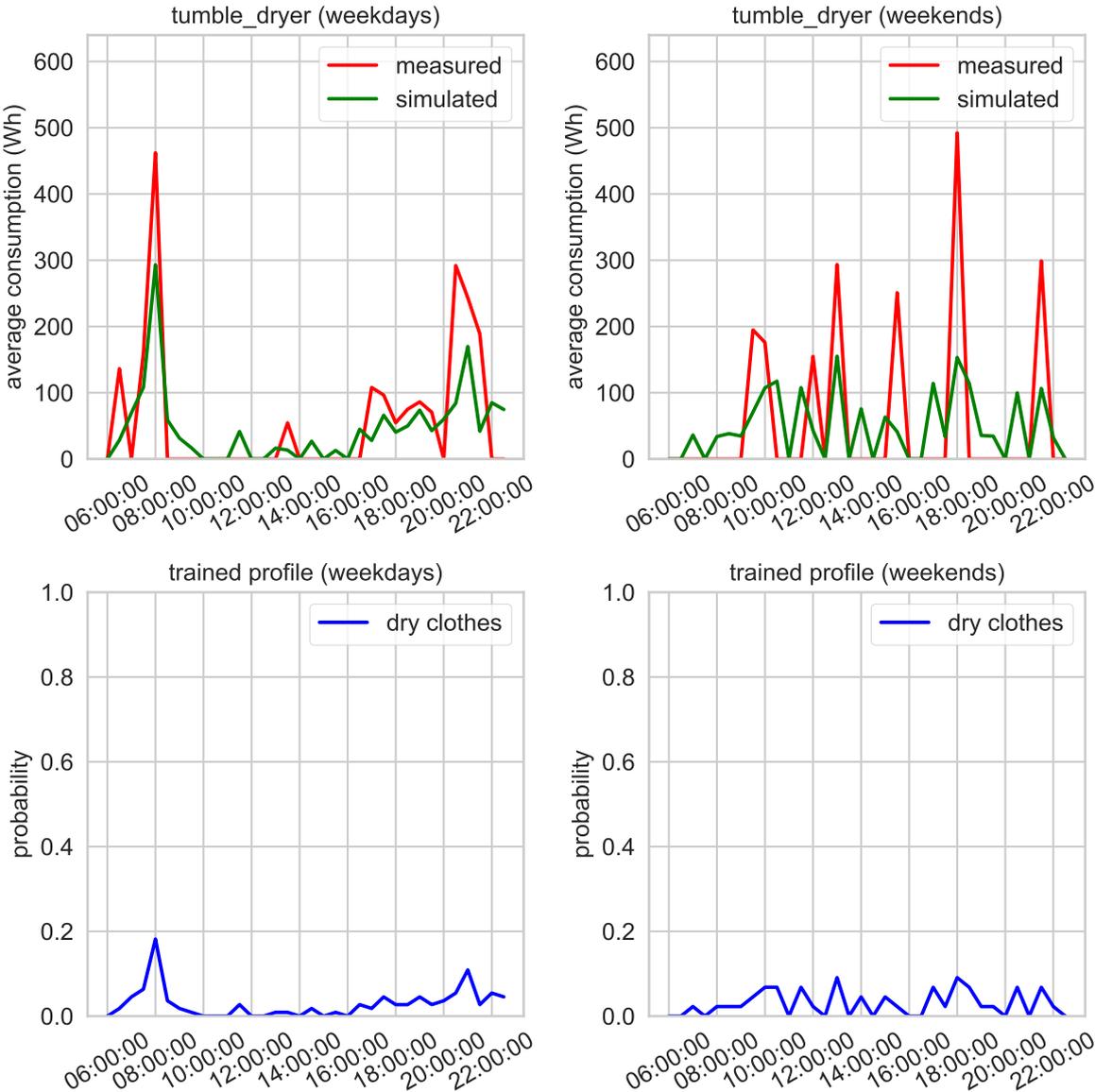


Figure 7.13 – Weekday and weekend average electricity consumption patterns for **the tumble dryer** with **drying clothes** profiles in the Cestas household from 2020-01-01 to 2020-02-01

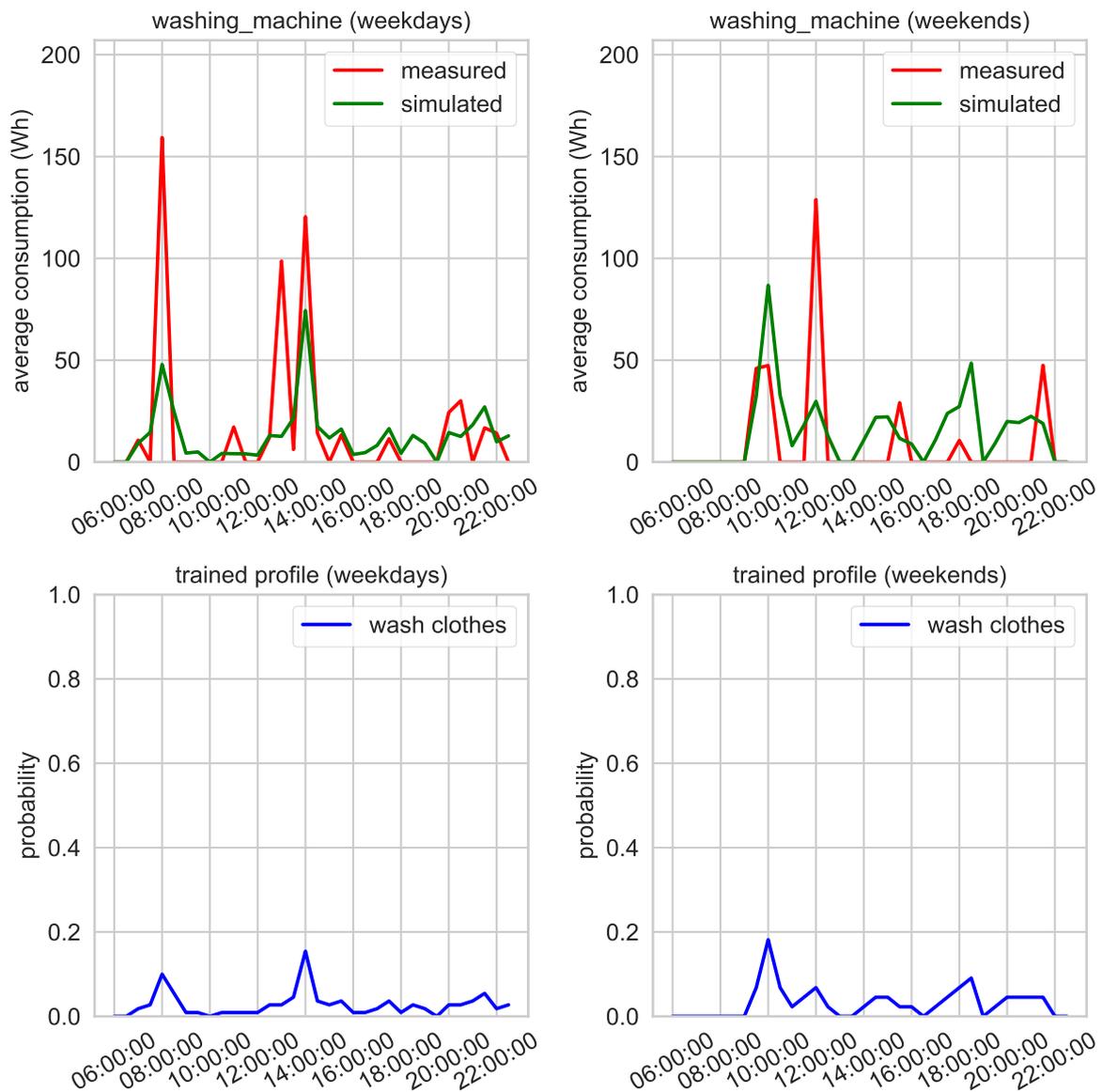


Figure 7.14 – Weekday and weekend average electricity consumption patterns for **the washing machine with washing clothes** profiles in the Cestas household from 2020-01-01 to 2020-02-01

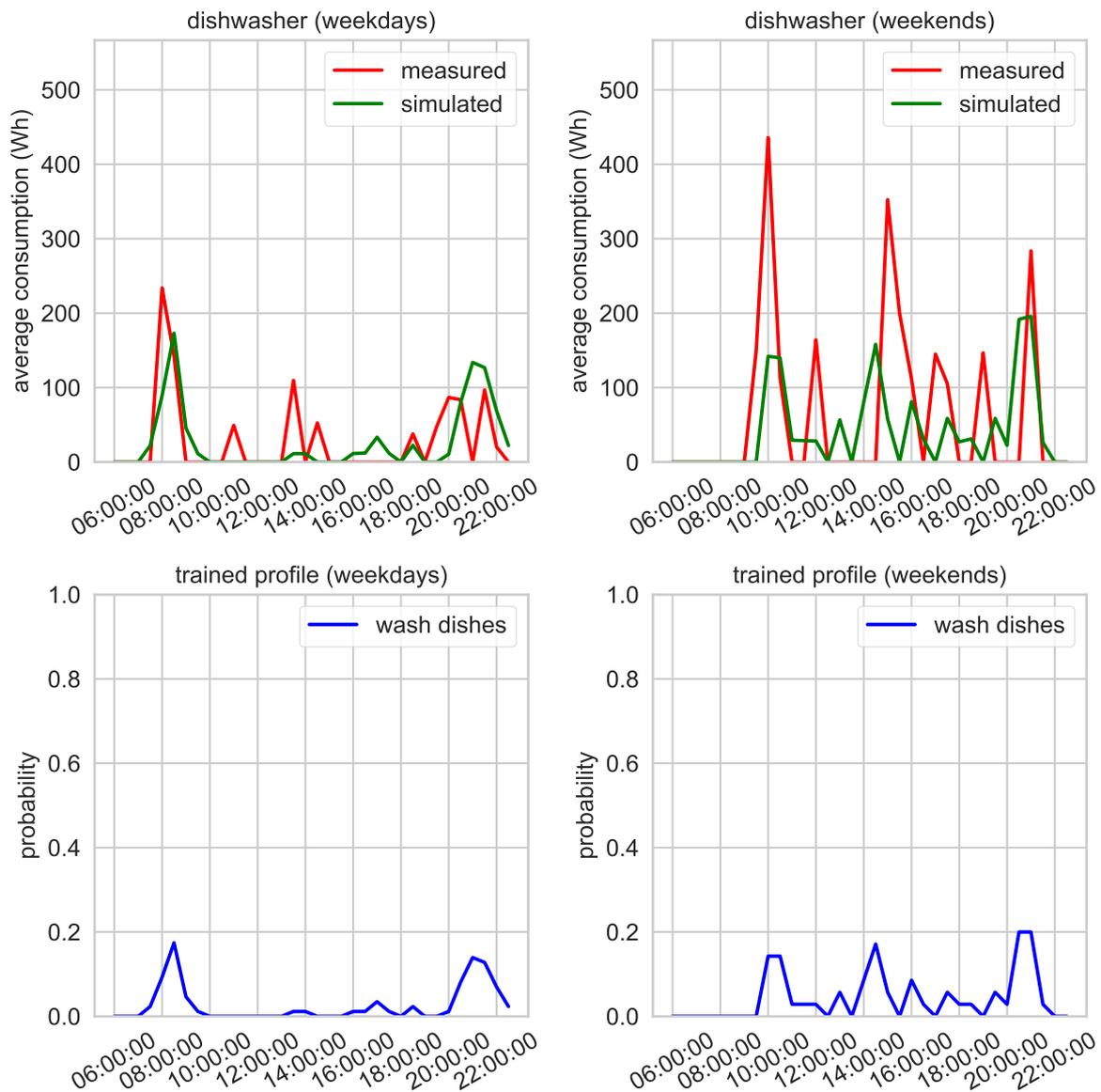


Figure 7.15 – Weekday and weekend average electricity consumption patterns for **the dishwasher** with **washing dishes** profiles in the Cestas household from 2020-01-01 to 2020-02-01

Regarding audio and video devices, their average electricity consumption patterns are shown in figure 7.16. It is obvious that simulated values are much less than measured ones. There are links between measured consumption and trained entertaining profiles but they are not completed. These devices consume a certain amount of electricity even when the entertaining activity tends not to happen.

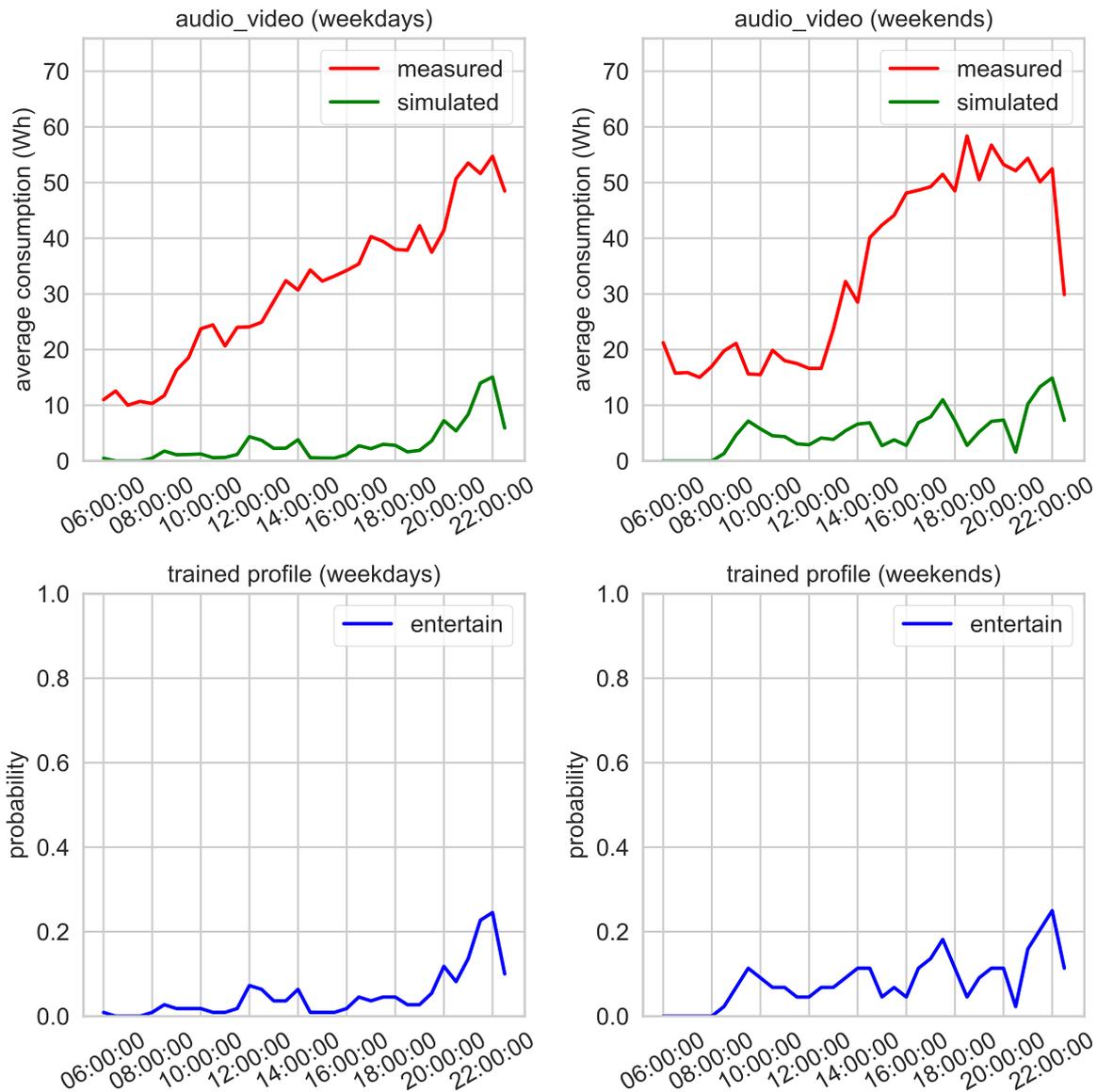


Figure 7.16 – Weekday and weekend average electricity consumption patterns for **the audio video devices** with **entertaining** profiles in the Cestas household from 2020-01-01 to 2020-02-01

7.4.2 Case study: Bordeaux

To test whether the methodology works in different contexts, we reproduce the same kind of analysis and steps, which were applied in the Cestas household for the Bordeaux household.

Firstly, the estimated labels are used to generate the activity profile. Particularly, the cooking activity profile in every day of the week is presented in figure 7.17.

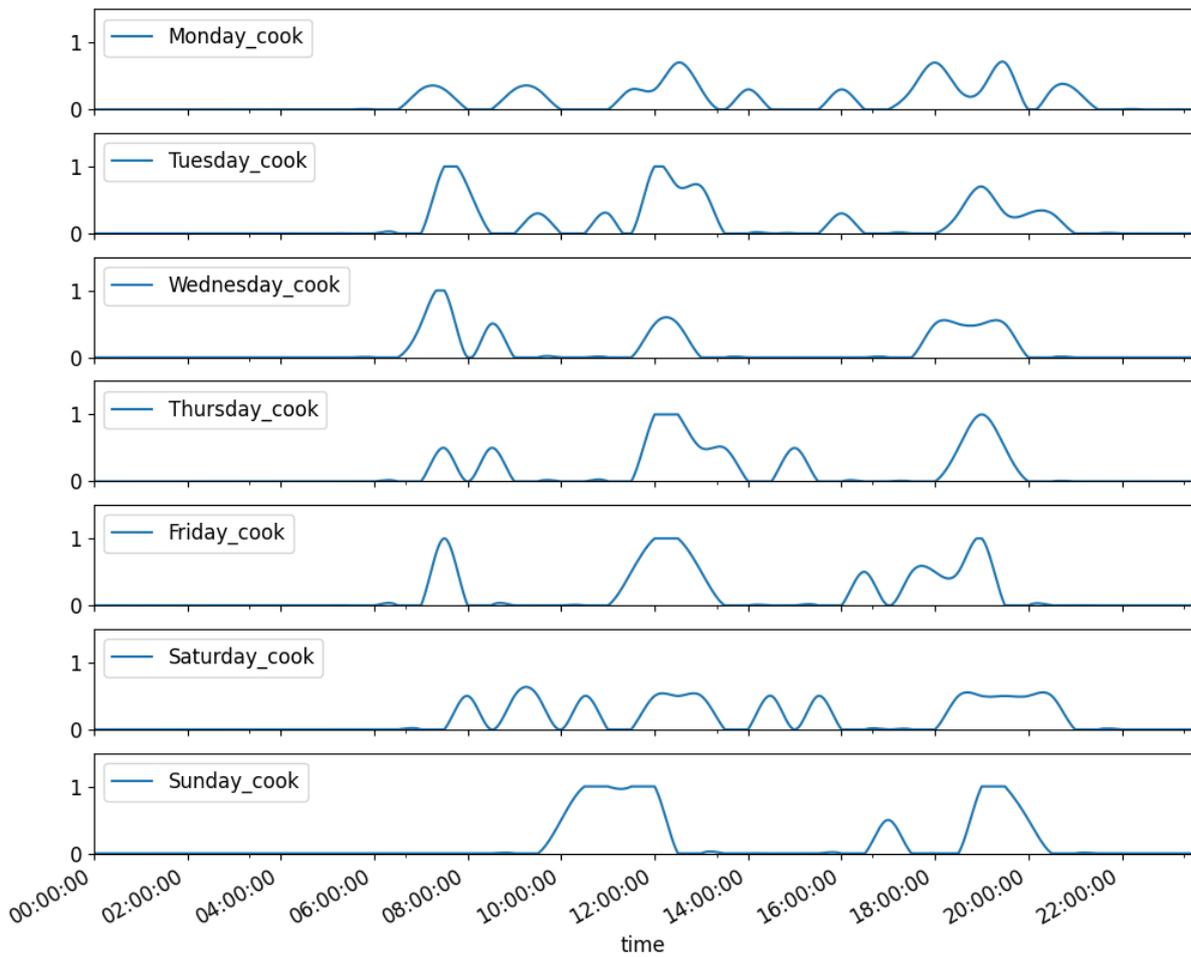


Figure 7.17 – Different cooking activity profiles between days in the Bordeaux household from 2021-04-05 to 2021-05-16

The profiles of cooking activity are summarized to weekday profile and weekend profile, which are illustrated in figure 7.18. The main gap between these profiles locates in cooking breakfast. It is shown that occupants tend to cook breakfast more or earlier on weekdays than on weekends. The probabilities of cooking lunch on weekends are generally higher than during the weekdays. Compared to the Cestas household, the Bordeaux household has a higher probability of cooking lunch and dinner during the weekdays.

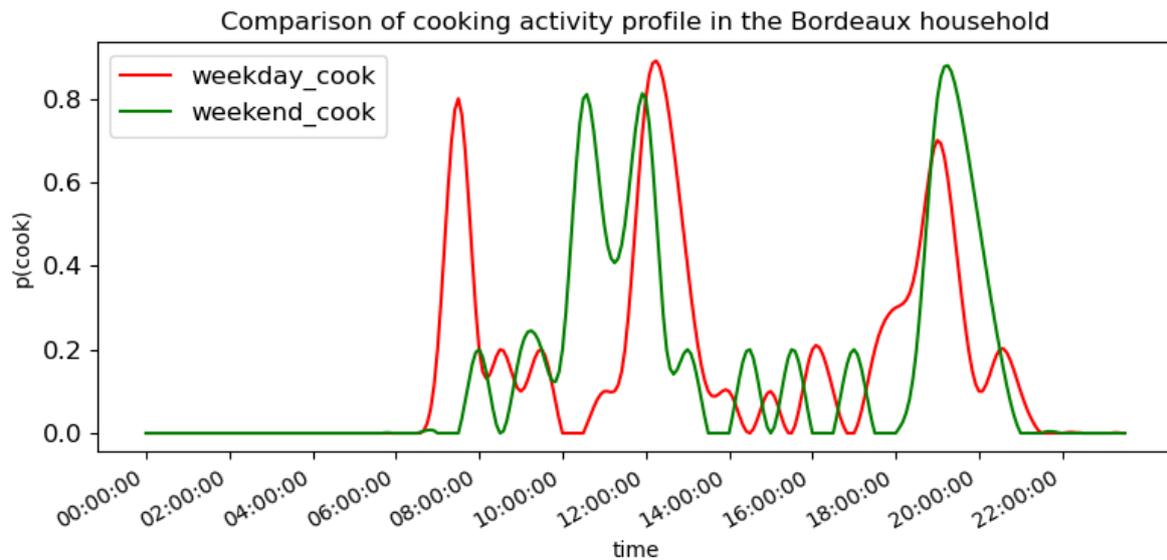


Figure 7.18 – Different cooking activity profiles between weekdays and weekend in the Bordeaux household from 2021-04-05 to 2021-05-16

In this household, four appliances associated with three kinds of cooking activity (cooking breakfast, cooking lunch and cooking dinner), which were measured, are considered: microwave, oven, toaster and kettle. Results of the microwave use simulation are shown through boxplots in figure 7.19. Occupants use an average of 2.2 kWh/week for the microwave but there are variations in their usages. More than half of the time, microwave consumes from 1.9 kWh/week to 2.9 kWh/week. In some weeks, the weekly used electricity can be less than 1.5 kWh or more than 3 kWh. In the proposed model, the average value is 2.1 kWh/week with a relative error of 5%. More than 50% of weekly consumption values in this approach are between 1.9 kWh to 3 kWh/week.

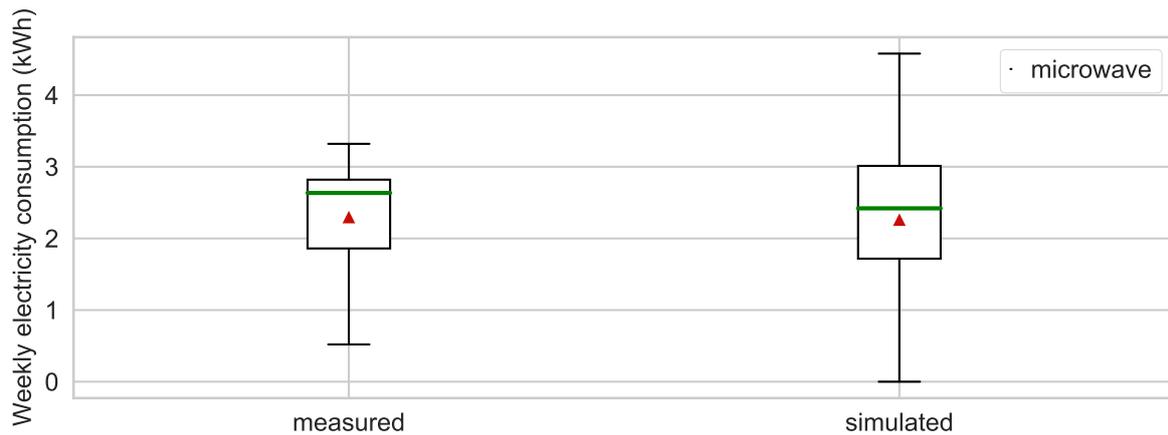


Figure 7.19 – Comparison between measurement and simulated weekly electricity consumption of **the microwave** in Bordeaux site from 2021-05-26 to 2021-06-24

For the kettle, it consumes a certain of electricity with an average of 4.6 kWh/week. Most of weeks consumes from 1.5 kWh/week to 7.5 kWh/week for this appliance. In the proposed model, more than 50% of consumption values are between 4 kWh/week and 6.3 kWh/week while the average value is 5.1 kWh/week with a relative error of 10%. These results and comparisons are presented in figure 7.20.



Figure 7.20 – Comparison between measured and simulated weekly electricity consumption of **the kettle** in the Bordeaux household from 2021-05-26 to 2021-06-24

Similarly, the results of the oven are presented in figure 7.21. This appliance consumes a certain amount of energy with an average of 3.8 kWh per week. During the testing period, the electricity used by the oven is always more than 3 kWh/week. For this appliance, the simulated average value is 4 kWh/week and its dispersion is

similar to the measured one. The relative error in this case is 5%.

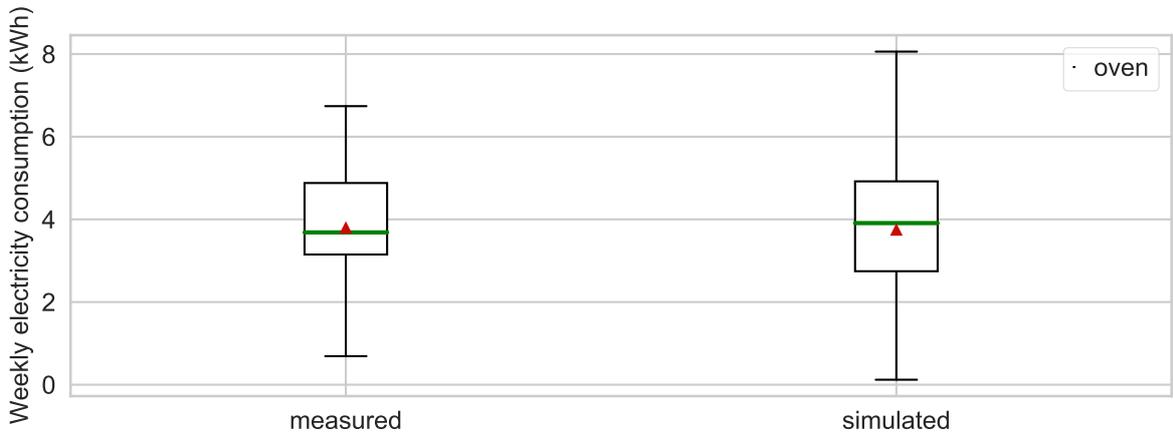


Figure 7.21 – Comparison between measured and simulated weekly electricity consumption of **the oven** in the Bordeaux household from 2021-05-26 to 2021-06-24

However, regarding the relative error, the results are not good with the toaster, which is presented in figure 7.22. The uses of the toaster of the household during the testing period are very small and stable with an average of 0.18 kWh/week. The methodology overestimated the consumption of this appliance. The simulated average consumption value is 0.42 kWh/week, which is two times (233%) more than the measured average value, the corresponding relative error is 133%. It can be explained that there can be a change in occupant behavior in using the toaster in the testing period. In that time, occupants spent less time using the toaster than in the training period.



Figure 7.22 – Comparison between measured and simulated weekly electricity consumption of **the toaster** in the Bordeaux household from 2021-05-26 to 2021-06-24

The summary of the results of the above appliances is presented in table 7.2 with their average values and standard deviations of the weekly consumption. The table shows that the methodology can simulate the electricity consumption of most of the above appliances with a relative error between 5% and 10%. It proves that the methodology can be adapted to this case study.

Table 7.2 – Comparison between measured and simulated weekly energy consumption of studied appliances related to the cooking activity in the Bordeaux household from 2021-05-26 to 2021-06-24

Appliance	Measured	Simulated	Relative error
Microwave	2.2 ± 0.8 kWh	2.1 ± 1 kWh	5%
Toaster	0.18 ± 0.1 kWh	0.42 ± 0.3 kWh	133%
Kettle	4.6 ± 2.1 kWh	5.1 ± 2.4 kWh	10%
Oven	3.8 ± 1.8 kWh	4 ± 1.8 kWh	5%

7.5 Conclusion

This chapter proposes a methodology to simulate occupants' activities and the electricity consumption of the involved appliances in specific households. This model combines two proposed models: the activity estimation model described in chapter 5 and the activity related electricity consumption model proposed in chapter 6. The activity profiles are determined by the activity estimation model. Profiles are used to predict the labels of activities. When an activity happens, the activity related consumption model is used to simulate the energy consumption of the associated domestic appliances. The methodology was tested in two households in France with different characteristics and settings. The results show that the methodology can simulate the electricity consumption and its diversity of most of the studied appliances in our case studies. It is represented by a tested relative error between -17% and 20%. On the other hand, the methodology can generate weekday and weekly average electricity consumption of appliances. It is presented that simulated patterns can represent the same trends as measured patterns.

In building renovations, BPS softwares are popularly used to evaluate the potential energy gains of retrofitting actions. To improve the performance of BPS softwares,

many approaches have been proposed to take account of occupants' activities and their energy impacts. However, most of them are based on statistical approaches and not specific for particular households. When renovating an existing house, it is interesting to consider occupants' activities and their energy impacts that are specific to this house. This methodology is useful to study contextualized activities and their energy impacts, particularly in electricity consumption, in a specific household. Firstly, activity profiles generated by the estimation model, which is built based on historical activities labels directly collected from occupants, are more specific than representative profiles determined by statistical approaches. Secondly, by using knowledge, the methodology can identify some information such as locations of activities, available appliances, list of appliances involved to an activity, which are uncertain in statistical approaches. Thirdly, the AR electricity consumption model is based on measurements so that it contains specific information about how appliances consume electricity for occupants' activities. In addition, the methodology models activities with their specific locations. That is useful to consider the consumption related to activities in each thermal zone. Finally, the methodology is easy to understand and verifiable.

However, the methodology needs the historical dataset including measurements, activity labels and context information. Thus, it requires much effort in sensors installation and occupants interaction to collect the necessary data. Moreover, occupants can change their behaviors over time, particularly in the use of appliances. This transition is difficult to be handled in the methodology. To deal with it, the idea is to build sub-models to consider different periods of the year. Additionally, interactive learning, which is proposed by Amayri (2017), can be integrated to adapt the energy behavior of occupants to new contexts over time. In addition, the methodology simplified relationships between activities which are useful to consider the consistency and the continuity of energy impacts. On the other hand, occupants rather perform activities and use appliances in close timesteps than a fixed timestep. Thus, regarding the average electricity consumption patterns, considering them in each 30-minutes can be too specific.

Conclusion and perspectives

General conclusion

Simulation tools support engineers to size HVAC systems and optimize the design of a building. However, building energy performance is not only influenced by its physical characteristics but also governed by occupant impacts through the presence and activities. It is investigated that occupants' impacts are essential to explain discrepancies between predicted and actual energy consumption in the design phase. Better information on the occupant activities might improve the performance of energy simulation tools and reduce this gap. In recent years, besides enhancing the building envelope, efforts have been made to enhance Energy Management and Monitoring Systems (EMMS) in residential buildings by concerning occupant activities as a part of the system. The goal is not only to reduce energy consumption but also to satisfy occupants' comfort. The determination of occupants' activities cannot be generally directly obtained by sensors or cameras because of the privacy concerns, modelling occupants' activities is thus an important key in monitoring systems and energy simulation.

The problem tackled by this thesis is to determine and process data from a given context, coming from sensors and questionnaires, that could be used to estimate occupants' activities (cooking, washing clothes, *etc.*) and their related energy impacts in a specific home, which are useful in both energy verification processes and evaluations of retrofitting actions in building renovations. The meaningful features for the estimation should be determined to reduce redundant sensors and save installation budgets. The approach should be verifiable and provide understandable information to inhabitants for suggesting their behavior changes.

The task of gathering labels is challenging in family households due to the diversity

of their members and activities. Using camera is rarely allowed since the private policy. To tackle this issue, a mobile application was developed to collect labels from individuals. These labels were combined with measurements to estimate and simulate occupants' activities (cooking, sleeping, washing clothes, *etc.*) and their related energy impacts. The scope of work includes contextualized models with three tasks:

1. **estimating occupant activities** with many sensors, the context, and activity labels collected from occupants. It starts by determining the features of sensors most related to the activity. Then, a knowledge-based model is built using meaningful features, expert knowledge of the activity, and the context coming from observation and questionnaires, which were given to occupants to ask labels of their activities. It relied on a Bayesian Network to model occupants activities with probabilistic cause-effect relations based on a defined consequences-based network. The estimation model was then applied to extrapolate labels to another dataset. The estimated labels were used to construct activity profiles in energy simulation.

Two instrumented houses with several activities were used to test the model. The time of the day, the uses of appliances and the occupancy, which is represented by the motions and CO₂ concentration, are determined as meaningful features in the model. Cross-validation with the F1-score metric was used to evaluate the accuracy of the estimation. The results were better with high energy activities and were worse with low energy activities. Besides, when the prerequisites of measurements and labels are met, the model was proven to have good performance in constructing activity profiles in the case studies.

2. **evaluating the energy impacts associated with occupants' activities** with a set of sensors, the context, and labels. The thesis focused on the activity related electricity consumption of domestic appliances. The uses of appliances are different between activities in houses, which leads to different patterns of the electricity consumption related to activities. To consider it, an activity related electricity consumption model, which was based on Bayesian Network, was built to investigate electricity levels (very low, low, medium, high, very high) consumed by domestic appliances related to occupants' activities (*i.e.* microwave, toaster, washing machine, *etc.*).

The model was tested with several activities in the studied house. There were discrepancies of the electricity levels of appliances within an activity and between activities. To evaluate the methodology, we compared the simulated electricity (by the proposed electricity model) and observed electricity consumption (by sensors) of several domestic appliances. When the labels of activities were determined, the model is useful in investigating the electricity levels used for occupants' activities. In the case study, the model were efficient in simulating the related electricity consumption of appliances associated to several activities. Results showed that occupants can use appliances in different ways for their activities. Their behaviors of using the appliances can change over time and the model was limited to these changes.

3. **simulating occupants' activities and their related energy impacts**, which can be useful to support selecting of retrofitting actions for a existing dwelling. In this work, activity profiles (determined by the activity estimation models) were combined with the activity related electricity model to simulate activities and their related electricity consumption of domestic appliances. This contextualized approach aims to improve the energy simulation in building renovations when the sensors data and labels of activities are available.

Several activities were tested in the case studies. To evaluate the methodology, simulated electricity consumption of appliances was statistically compared to measured ones. The results showed that the approach was useful in simulating occupants activities and the related domestic electricity consumption in these cases.

When the measurements, labels and the context of the house are available, the contextualized models are suitable in modelling occupants activities and their related domestic electricity consumption in specific households. It helps to contribute some useful information to BPS softwares to reduce the gap between simulated and actual energy simulation. This benefit can be helpful in both verifying current energy performance and evaluating refurbishments of existing dwellings. Contextualized models can give occupants insights of energy impacts of their activities. They can use this information to improve their behavior in houses regarding the energy consumption. In addition, selections of meaningful features related to occupants activities can be

useful to save budget for future installation campaigns. However, some limitations exist in this approach:

- The approach only considered several activities related to domestic appliances. It did not have good performance with activities that are not related to electrical appliances.
- The appliances, which are not only associated to occupants' activities, are not concerned (*i.e.* fridge, artificial lighting, *etc.*).
- The approach is limited in considering the changes of occupants' activity profiles and their behaviors of using appliances over time, which usually happens when the season or the calendar of individuals changes.
- Occupants' activities can affect many factors in houses such as HVAC systems, lighting systems, the electricity consumption, *etc.* However, the approach only focused on the impacts on electricity consumption of domestic appliances.
- Even if the mobile application is convenient to used in many cases, remaining it in the long-term is a challenge for individuals, especially when they find no interactions between them and the results.

Future work

The work done in this thesis is focusing on the impacts of occupant activities on domestic appliances in houses. This work can be extended to other effects as the actions on doors/windows, artificial lighting, or the indoor environment (temperature, humidity, *etc.*). Though these variables are complex and not only affected by occupant activities, studying them is interesting to investigate the occupant habits and preferences. More knowledge as individuals' calendars, seasons, vacations can be added into the model to take into account the possible behavior changes.

In the short-term future, interactive learning can be integrated to the model to improve the process of collecting labels and enhance the model performance. In particular, with interactive learning approach, when some conditions are satisfied (the estimation error of the model, the density of the neighborhood, or the minimum weight of the concerned activity), the questions are sent to occupants to ask about their activities. It is a gradually process with the following benefits:

- Interaction with occupants to collect labels so that their motivations are increased. This process is useful in building the model when the initial dataset is not large enough. Using interactive learning helps the model to be trained in parallel to the task of collecting labels. Thus, the model quality is improved over time when the dataset increases. Moreover, considering the time to task with the defined criteria helps to reduce the unnecessary requests sent to occupants.
- By interacting with occupants, model with interactive learning approach can update its parameters to adapt new situations. This feature is useful in tackling the issue of the behavior changes of occupants.

In the mid-term future, contextualized models maybe give useful information to study occupants activities in statistical datasets. Even though contextualized models focus on a specific house, it is interesting to investigate whether they works with similar houses. There are many statistical datasets having measurements but occupants' activities are missing. Completing them is useful to study energy impacts of occupants activities on dwellings in a population scale.

In the long-term future, contextualized activities and their energy impacts can provide useful recommendations to people. They can use these recommendations to evaluate ways to modify their activities.

Publications

- Phan, H., Recht, T., Mora, L., et Ploix, S. (2021). Consequences-based graphical model for contextualized occupant activities estimation in connected buildings. In *17th International Conference of the International Building Performance Simulation Association, BS2021*, Brugge, Belgium.
- Phan, H., Recht, T., Mora, L., et Ploix, S. (2020). Contextualized model for occupants'activities estimation in connected buildings. In *Conférence IBPSA France 2020*, Reims, France

Bibliography

- Aerts, D. (2015). *Occupancy and Activity Modelling for Building Energy Demand Simulations, Comparative Feedback and Residential Electricity Demand Characterisation*. PhDThesis, Vrije Universiteit Brussel, Belgium.
- Aerts, D., Minnen, J., Glorieux, I., Wouters, I., et Descamps, F. (2014). A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and Environment*, 75:67–78.
- Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., et Weng, T. (2010). Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 1–6, New York, NY, USA. Association for Computing Machinery.
- Ahmadi-Karvigh, S., Becerik-Gerber, B., et Soibelman, L. (2016). A framework for allocating personalized appliance-level disaggregated electricity consumption to daily activities. *Energy and Buildings*, 111:337–350.
- Akbari, S. et Haghghat, F. (2021). Occupancy and occupant activity drivers of energy consumption in residential buildings. *Energy and Buildings*, 250:111303.
- Albatayneh, A., Alterman, D., Page, A., et Moghtaderi, B. (2015). The Significance of Time Step Size in Simulating the Thermal Performance of Buildings. *Advances in Research*, 5:1–12.
- Albouys, J., Sabouret, N., Haradji, Y., Schumann, M., et Inard, C. (2019). SMACH: Multi-agent Simulation of Human Activity in the Household. In Demazeau, Y.,

- Matson, E., Corchado, J. M., et De la Prieta, F., editors, *Advances in Practical Applications of Survivable Agents and Multi-Agent Systems: The PAAMS Collection*, Lecture Notes in Computer Science, pages 227–231, Cham. Springer International Publishing.
- Alhamoud, A., Xu, P., Englert, F., Reinhardt, A., Scholl, P., Boehnstedt, D., et Steinmetz, R. (2015). Extracting Human Behavior Patterns from Appliance-level Power Consumption Data. In Abdelzaher, T., Pereira, N., et Tovar, E., editors, *Wireless Sensor Networks*, Lecture Notes in Computer Science, pages 52–67. Springer International Publishing.
- Alves, C. A., Vicente, E. D., Evtyugina, M., Vicente, A. M. P., Sainnokhoi, T.-A., et Kováts, N. (2021). Cooking activities in a domestic kitchen: Chemical and toxicological profiling of emissions. *Science of The Total Environment*, 772:145412.
- Alyafi, A. A. (2019). *Generation of explanations for energy management in buildings*. PhDThesis, Université Grenoble Alpes, France.
- Amayri, M. (2017). *Estimation de l'occupation dans le bâtiment*. PhDThesis, Université Grenoble Alpes, France.
- Amayri, M., Ngo, Q.-D., et Ploix, S. (2016). Estimating Occupancy from Measurement and Knowledge with Bayesian Networks. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 508–513, United States of American. IEEE. event-place: Las Vegas, NV, USA.
- Amayri, M., Ngo, Q.-D., Safadi, E. A. E., et Ploix, S. (2017). Bayesian network and Hidden Markov Model for estimating occupancy from measurements and knowledge. In *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, pages 690–695, Romania. IEEE. event-place: Bucharest.
- Andersen, R., Fabi, V., Toftum, J., Corgnati, S. P., et Olesen, B. W. (2013). Window opening behaviour modelled from measurements in Danish dwellings. *Building and Environment*, 69:101–113.

- Andersen, R. V., Toftum, J., Andersen, K. K., et Olesen, B. W. (2009). Survey of occupant behaviour and control of indoor environment in Danish dwellings. *Energy and Buildings*, 41(1):11–16.
- Ascione, F., Bianco, N., Iovane, T., Mastellone, M., et Mauro, G. M. (2021). Conceptualization, development and validation of EMAR: A user-friendly tool for accurate energy simulations of residential buildings via few numerical inputs. *Journal of Building Engineering*, 44:102647.
- Asvatourian, V., Leray, P., Michiels, S., et Lanoy, E. (2020). Integrating expert's knowledge constraint of time dependent exposures in structure learning for Bayesian networks. *Artificial Intelligence in Medicine*, 107:101874.
- Ayadi, A., Ghorbel, O., Obeid, A. M., et Abid, M. (2017). Outlier detection approaches for wireless sensor networks: A survey. *Computer Networks*, 129:319–333.
- Azar, E. et Al Ansari, H. (2017). Framework to investigate energy conservation motivation and actions of building occupants: The case of a green campus in Abu Dhabi, UAE. *Applied Energy*, 190:563–573.
- Baig, Z. (2011). Pattern recognition for detecting distributed node exhaustion attacks in wireless sensor networks. *Computer Communications*, 34(3):468–484.
- Balvedi, B. F., Ghisi, E., et Lamberts, R. (2018). A review of occupant behaviour in residential buildings. *Energy and Buildings*, 174:495–505.
- Ben-Haim, Y. (2021). Feedback for energy conservation: An info-gap approach. *Energy*, 223:119957.
- Bettencourt, L. M. A., Hagberg, A. A., et Larkey, L. B. (2007). Separating the Wheat from the Chaff: Practical Anomaly Detection Schemes in Ecological Applications of Distributed Sensor Networks. In Aspnes, J., Scheideler, C., Arora, A., et Madden, S., editors, *Distributed Computing in Sensor Systems*, Lecture Notes in Computer Science, pages 223–239, Berlin, Heidelberg. Springer.
- Borgeson, S. et Brager, G. (2008). Occupant Control of Windows: Accounting for Human Behavior in Building Simulation. Technical report, UC Berkeley: Center for the Built Environment.

- Caravagna, G. et Ramazzotti, D. (2021). Learning the structure of Bayesian Networks via the bootstrap. *Neurocomputing*, 448:48–59.
- Chen, S., Zhang, G., Xia, X., Chen, Y., Setunge, S., et Shi, L. (2021). The impacts of occupant behavior on building energy consumption: A review. *Sustainable Energy Technologies and Assessments*, 45:101212.
- Chen, S. H. et Pollino, C. A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software*, 37:134–145.
- Cheung, C. K., Fuller, R. J., et Luther, M. B. (2005). Energy-efficient envelope design for high-rise apartments. *Energy and Buildings*, 37(1):37–48.
- Clayman, C. L., Srinivasan, S. M., et Sangwan, R. S. (2020). K-means Clustering and Principal Components Analysis of Microarray Data of L1000 Landmark Genes. *Procedia Computer Science*, 168:97–104.
- Congming, S., Wei, B., Wei, S., Wang, W., Liu, H., et Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2021.
- Delghust, M., Laverge, J., Janssens, A., Erck, C. V., et Taelman, C. (2012). The influence of user behaviour on energy use in old dwellings: case-study analysis of a social housing neighbourhood. page 8.
- Demirhan, H. et Renwick, Z. (2018). Missing value imputation for short to mid-term horizontal solar irradiance data. *Applied Energy*, 225:998–1012.
- Di Marzo Serugendo, G., Gleizes, M.-P., et Anthony, K. (2005). Self-Organization in Multi-Agent Systems. *Knowledge Eng. Review*, 20:165–189.
- Directive Parliament (2012). EUR-Lex - 32012L0027 - EN - EUR-Lex.
- Dong, B. et Lam, K. P. (2014). A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation*, 7(1):89–106.

- D'Oca, S., Fabi, V., Corgnati, S. P., et Andersen, R. K. (2014). Effect of thermostat and window opening occupant behavior models on energy use in homes. *Building Simulation*, 7(6):683–694.
- Elbayoudi, A., Lotfi, A., et Langensiepen, C. (2019). The human behaviour indicator: A measure of behavioural evolution. *Expert Systems with Applications*, 118:493–505.
- European Commision (2015). EUROPA - Energy Performance of Buildings: Technology Information Sheet | SETIS - European Commission.
- European Commision (2016). Clean energy for all Europeans - Energy - European Commission.
- European Parliament (2002). Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings.
- European Parliament (2010). Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings.
- European Parliament (2012). Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC with EEA relevance.
- Eurostat (2018). Energy statistics - an overview.
- Eurostat (2020). Energy consumption in households - Statistics Explained.
- Fabi, V., Andersen, R. V., et Corgnati, S. P. (2013). Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings. *HVAC&R Research*, 19(5):635–645.
- Fathi, S. et Kavooosi, A. (2021). Effect of electrochromic windows on energy consumption of high-rise office buildings in different climate regions of Iran. *Solar Energy*, 223:132–149.
- Fayyad, U. et Irani, K. (1993). Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, France.

- Firth, S., Lomas, K., Wright, A., et Wall, R. (2008). Identifying trends in the use of domestic appliances from household electricity consumption measurements. *Energy and Buildings*, 40(5):926–936.
- Françoise, D. (2018). L'énergie et le bâtiment : les données chiffrées pour la France depuis 1950 | RESPONSABILITÉ & ENVIRONNEMENT. Technical Report 90, Annales des Mines, France.
- Gaetani, I., Hoes, P.-J., et Hensen, J. L. M. (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings*, 121:188–204.
- Gao, X.-g., Guo, Z.-g., Ren, H., Yang, Y., Chen, D.-q., et He, C.-c. (2019). Learning Bayesian network parameters via minimax algorithm. *International Journal of Approximate Reasoning*, 108:62–75.
- Gram-Hanssen, K. (2013). Efficient technologies or user behaviour, which is the more important when reducing households' energy consumption? *Energy Efficiency*, 6(3):447–457.
- Gram-Hanssen, K. (2010). Standby Consumption in Households Analyzed With a Practice Theory Approach. *Journal of Industrial Ecology*, 14(1):150–165.
- Guerra Santin, O. (2011). Behavioural Patterns and User Profiles related to energy consumption for heating. *Energy and Buildings*, 43(10):2662–2672.
- Guerra-Santin, O. et Itard, L. (2012). The effect of energy performance regulations on energy consumption. *Energy Efficiency*, 5(3):269–282.
- Guivarch, V., Camps, V., et Péninou, A. (2012). AMADEUS: an adaptive multi-agent system to learn a user's recurring actions in ambient systems. *ADCAIJ : Advances in Distributed Computing and Artificial Intelligence Journal*, 1.
- Guo, Z.-g., Gao, X.-g., Ren, H., Yang, Y., Di, R.-h., et Chen, D.-q. (2017). Learning Bayesian network parameters from small data sets: A further constrained qualitatively maximum a posteriori method. *International Journal of Approximate Reasoning*, 91:22–35.

- Ha, D. L., Ploix, S., Zamai, E., et Jacomino, M. (2006). A home automation system to improve household energy control. *IFAC Proceedings Volumes*, 39(3):15–20.
- Hager, T. J. et Morawicki, R. (2013). Energy consumption during cooking in the residential sector of developed nations: A review. *Food Policy*, 40:54–63.
- Haigh, S., Ghosh, B., et Madabhushi, G. (2011). Importance of time step discretisation for nonlinear dynamic finite element analysis. *Canadian Geotechnical Journal*, 42:957–963.
- Haldi, F. et Robinson, D. (2008). On the behaviour and adaptation of office occupants. *Building and Environment*, 43(12):2163–2177.
- Haldi, F. et Robinson, D. (2009). Interactions with window openings by office occupants. *Building and Environment*, 44(12):2378–2395.
- Hawarah, L., Ploix, S., et Jacomino, M. (2010). User Behavior Prediction in Energy Consumption in Housing Using Bayesian Networks. In Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., et Zurada, J. M., editors, *International Conference on Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, pages 372–379, Poland. Springer Berlin Heidelberg.
- He, Z., Hong, T., et Chou, S. K. (2021). A framework for estimating the energy-saving potential of occupant behaviour improvement. *Applied Energy*, 287:116591.
- Henricksen, K. et Indulska, J. (2005). Developing context-aware pervasive computing applications: Models and approach. *Pervasive and Mobile Computing*, 2:37–64.
- Hill, D. J. et Minsker, B. S. (2010). Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling & Software*, 25(9):1014–1022.
- Hoes, P., Hensen, J. L. M., Loomans, M. G. L. C., de Vries, B., et Bourgeois, D. (2009). User behavior in whole building simulation. *Energy and Buildings*, 41(3):295–302.
- Hong, T., D'Oca, S., Turner, W. J. N., et Taylor-Lange, S. C. (2015). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. *Building and Environment*, 92:764–777.

- Hong, T., Yan, D., D'Oca, S., et Chen, C.-f. (2017). Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, 114:518–530.
- Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z., et Oreszczyn, T. (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy*, 177:692–702.
- Humphreys, M. A. et Nicol, J. F. (1998). Understanding the adaptive approach to thermal comfort. In *ASHRAE Transactions*, volume 104, pages 991–1004, United State.
- Janda, K. B. (2011). Buildings don't use energy: people do. *Architectural Science Review*, 54(1):15–22.
- Jia, M., Srinivasan, R. S., et Raheem, A. A. (2017). From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, 68:525–540.
- Jones, R. V. et Lomas, K. J. (2016). Determinants of high electrical energy demand in UK homes: Appliance ownership and use. *Energy and Buildings*, 117:71–82.
- Julia, H. (2012). Greenhouse gases rise by 260% since Industrial Revolution.
- Kashef, S., Nezamabadi-pour, H., et Nikpour, B. (2018). Multilabel feature selection: A comprehensive review and guiding experiments. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 8.
- Kashif, A. (2014). *Modélisation du comportement humain réactif et délibératif avec une approche multi-agent pour la gestion énergétique dans le bâtiment*. PhDThesis, Université Grenoble Alpes, France.
- Kashif, A., Le, X. H. B., Dugdale, J., et Ploix, S. (2011). Agent based Framework to Simulate Inhabitants' Behaviour in Domestic Settings for Energy Management. In *International Conference on Agents and Artificial Intelligence*, volume 3, pages 190–199, Italy.

- Kashif, A., Ploix, S., Dugdale, J., et Le, X. H. B. (2013). Simulating the dynamics of occupant behaviour for power management in residential buildings. *Energy and Buildings*, 56:85–93.
- Khaire, U. M. et Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*.
- Kim, J., Min, K., Jung, M., et Chi, S. (2020). Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition. *Building and Environment*, 181:107092.
- Kurgan, L. et Cios, K. (2004). CAIM discretization algorithm. *IEEE Transactions on Knowledge and Data Engineering*, 16(2):145–153. Conference Name: IEEE Transactions on Knowledge and Data Engineering.
- Laurent, M.-H., Allibe, B., Oreszczyn, T., Hamilton, I., et Tigchelaar, C. (2013). Back to reality: How domestic energy efficiency policies in four European countries can be improved by using empirical data instead of normative calculation. In *ECEEE Summer Study proceedings*, pages 2057–2070, France.
- Le, V. T. et Pitts, A. (2019). A survey on electrical appliance use and energy consumption in Vietnamese households: Case study of Tuy Hoa city. *Energy and Buildings*, 197:229–241.
- Lee, E., Kang, M., Song, J., et Kang, M. (2020). From intention to action: Habits, feedback and optimizing energy consumption in South Korea. *Energy Research & Social Science*, 64:101430.
- Lee, J., Shepley, M., et Choi, J. (2019). Exploring the effects of a building retrofit to improve energy performance and sustainability: A case study of Korean public buildings. *Journal of Building Engineering*, 25:100822.
- Lee, Y. S. et Malkawi, A. M. (2014). Simulating multiple occupant behaviors in buildings: An agent-based modeling approach. *Energy and Buildings*, 69:407–416.
- Lepot, M., Aubin, J.-B., et Clemens, F. (2017). Interpolation in Time Series: An Introductory Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water*, 9:796.

- Leroy, Y. et Yannou, B. (2018). An activity-based modelling framework for quantifying occupants' energy consumption in residential buildings. *Computers in Industry*, 103:1–13.
- Li, Z. et Dong, B. (2017). A new modeling approach for short-term prediction of occupancy in residential buildings. *Building and Environment*, 121:277–290.
- Lin, H.-W. et Hong, T. (2013). On variations of space-heating energy use in office buildings. *Applied Energy*, 111:515–528.
- Lu, Y., Li, P., Lee, Y. P., et Song, X. (2021). An integrated decision-making framework for existing building retrofits based on energy simulation and cost-benefit analysis. *Journal of Building Engineering*, 43:103200.
- Macal, C. M. et North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162.
- Maile, T., Fischer, M., et Bazjanac, V. (2007). Building Energy Performance Simulation Tools - a Life-Cycle and Interoperable Perspective. Working Paper 107, Stanford Digital Repository.
- Masoso, O. T. et Grobler, L. J. (2010). The dark side of occupants' behaviour on building energy use. *Energy and Buildings*, 42(2):173–177.
- Milliken, M. E. et Honeycutt, A. (2004). *Understanding human behavior: a guide for health care providers*. Thomson Delmar Learning, Clifton Park, NY, 7 edition. OCLC: 54408066.
- Misra, P. et Singh, A. (2020). Improving the Classification Accuracy using Recursive Feature Elimination with Cross-Validation. *International Journal on Emerging Technologies*, 11:659–665.
- Muratori, M., Roberts, M. C., Sioshansi, R., Marano, V., et Rizzoni, G. (2013). A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, 107(C):465–473.

- Murray, D. M., Stankovic, L., Stankovic, V., et Espinoza-Orias, N. D. (2018). Appliance electrical consumption modelling at scale using smart meter data. *Journal of Cleaner Production*, 187:237–249.
- Nagy, Z., Yong, F. Y., Frei, M., et Schlueter, A. (2015). Occupant centered lighting control for comfort and energy efficient building operation. *Energy and Buildings*, 94:100–108.
- Nicol, F. et Humphreys, M. (2004). A stochastic approach to thermal comfort-Occupant behavior and energy use in buildings. *ASHRAE transactions*, 110:554–568.
- Odhiambo Omuya, E., Onyango Okeyo, G., et Waema Kimwele, M. (2021). Feature Selection for Classification using Principal Component Analysis and Information Gain. *Expert Systems with Applications*, 174:114765.
- Ouyang, J. et Hokao, K. (2009). Energy-saving potential by improving occupants' behavior in urban residential sector in Hangzhou City, China. *Energy and Buildings*, 41(7):711–720.
- Padmanabh, K., Vanteddu, A. M., Sen, S., Katru, S., Kumar, A., C, S., Vuppala, S., et Paul, S. (2009). iSense: a wireless sensor network based conference room management system. *SenSys09: The 7th ACM Conference on Embedded Network Sensor Systems*, pages 37–42.
- Page, J., Robinson, D., Morel, N., et Scartezzini, J. L. (2008). A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40(2):83–98.
- Pal, M., Alyafi, A. A., Ploix, S., Reignier, P., et Bandyopadhyay, S. (2019). Unmasking the causal relationships latent in the interplay between occupant's actions and indoor ambience: A building energy management outlook. *Applied Energy*, 238:1452–1470.
- Pang, Z., Chen, Y., Zhang, J., O'Neill, Z., Cheng, H., et Dong, B. (2020). How much HVAC energy could be saved from the occupant-centric smart home thermostat: A nationwide simulation study. *Applied Energy*, page 116251.
- Pannier, M.-L., Recht, T., Robillart, M., Schalbart, P., Peuportier, B., et Mora, L. (2021). Identifying optimal renovation schedules for building portfolios: Application in a

- social housing context under multi-year funding constraints. *Energy and Buildings*, 250:111290.
- Peuportier, B. et Blanc, I. (1990). Simulation tool with its expert interface for the thermal design of multizone buildings. *International Journal of Solar Energy*, 8(2):pages 109 – 120. Publisher: Taylor & Francis.
- Picornell, A., Oteros, J., Ruiz-Mata, R., Recio, M., Trigo, M. M., Martínez-Bracero, M., Lara, B., Serrano-García, A., Galán, C., García-Mozo, H., Alcázar, P., Pérez-Badia, R., Cabezudo, B., Romero-Morte, J., et Rojo, J. (2021). Methods for interpolating missing data in aerobiological databases. *Environmental Research*, 200:111391.
- Rajasegarar, S., Leckie, C., Palaniswami, M., et Bezdek, J. C. (2006). Distributed Anomaly Detection in Wireless Sensor Networks. In *10th IEEE Singapore International Conference on Communication Systems*, pages 1–5, Singapore.
- Redner, R. A. et Walker, H. F. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review*, 26(2):195–239. Publisher: Society for Industrial and Applied Mathematics.
- Regnier, C., Sun, K., Hong, T., et Piette, M. A. (2018). Quantifying the benefits of a building retrofit using an integrated system approach: A case study. *Energy and Buildings*, 159:332–345.
- Ren, X., Yan, D., et Hong, T. (2015). Data mining of space heating system performance in affordable housing. *Building and Environment*, 89:1–13.
- Richardson, I., Thomson, M., et Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40(8):1560–1566.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
- Santangelo, A. et Tondelli, S. (2017). Occupant behaviour and building renovation of the social housing stock: Current and future challenges. *Energy and Buildings*, 145:276–283.

- Scutari, M., Graafland, C. E., et Gutiérrez, J. M. (2019). Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. *International Journal of Approximate Reasoning*, 115:235–253.
- Sharif, S. A. et Hammad, A. (2019). Simulation-Based Multi-Objective Optimization of institutional building renovation considering energy consumption, Life-Cycle Cost and Life-Cycle Assessment. *Journal of Building Engineering*, 21:429–445.
- Sharmin, T., Gül, M., Li, X., Ganev, V., Nikolaidis, I., et Al-Hussein, M. (2014). Monitoring building energy consumption, thermal performance, and indoor air quality in a cold climate region. *Sustainable Cities and Society*, 13:57–68.
- Shiel, P., Tarantino, S., et Fischer, M. (2018). Parametric analysis of design stage building energy performance simulation models. *Energy and Buildings*, 172:78–93.
- Sierhuis, M. et Clancey, W. J. (1997). Knowledge, Practice, Activities and People. *AAAI Technical Report SS-97-01*, 142-48:7.
- Sonta, A. J., Simmons, P. E., et Jain, R. K. (2018). Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach. *Advanced Engineering Informatics*, 37:1–13.
- Subbalakshmi, C., Krishna, G. R., Rao, S. K. M., et Rao, P. V. (2015). A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *Procedia Computer Science*, 46:346–353.
- Sun, B., Zhou, Y., Wang, J., et Zhang, W. (2021). A new PC-PSO algorithm for Bayesian network structure learning with structure priors. *Expert Systems with Applications*, 184:115237.
- Szidónia, L. et László, L. (2017). Gabor Feature Selection Based on Information Gain. *Procedia Engineering*, 181:892–898.
- Taillandier, F., Mora, L., et Denys, B. (2016). Decision support to choose renovation actions in order to reduce house energy consumption – An applied approach. *Building and Environment*, 109.

- Tanimoto, J., Hagishima, A., et Sagara, H. (2008). A methodology for peak energy requirement considering actual variation of occupants' behavior schedules. *Building and Environment*, 43(4):610–619.
- Tian, Z., Zhang, X., Jin, X., Zhou, X., Si, B., et Shi, X. (2018). Towards adoption of building energy simulation and optimization for passive building design: A survey and a review. *Energy and Buildings*, 158:1306–1316.
- Tijani, K., Ngo, Q. D., Ploix, S., Haas, B., et Dugdale, J. (2015). Towards a General Framework for an Observation and Knowledge based Model of Occupant Behaviour in Office Buildings. *Energy Procedia*, 78:609–614.
- Trinh, K., Fung, A. S., et Straka, V. (2021). Effects of Real-Time Energy Feedback and Normative Comparisons: Results from a Multi-Year Field Study in a Multi-Unit Residential Building. *Energy and Buildings*, 250:111288.
- Varlamis, I., Sardianos, C., Chronis, C., Dimitrakopoulos, G., Himeur, Y., Alsalemi, A., Bensaali, F., et Amira, A. (2022). Smart fusion of sensor data and human feedback for personalized energy-saving recommendations. *Applied Energy*, 305:117775.
- Vassileva, I. et Campillo, J. (2014). Increasing energy efficiency in low-income households through targeting awareness and behavioral change. *Renewable Energy*, 67:59–63.
- Vorger, E. (2014). *Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment*. PhDThesis, École nationale supérieure des Mines de Paris, France.
- Wang, C., Pattawi, K., et Lee, H. (2020). Energy saving impact of occupancy-driven thermostat for residential buildings. *Energy and Buildings*, 211:109791.
- Wang, H. et Song, M. (2011). Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming. *The R Journal*, 3(2):29.
- Wang, W., Du, W., Cheng, C., Lu, X., et Zou, W. (2022). Output feedback control for energy-saving asymmetric hydraulic servo system based on desired compensation approach. *Applied Mathematical Modelling*, 101:360–379.

- Widén, J., Molin, A., et Ellegård, K. (2012). Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations. *Journal of Building Performance Simulation*, 5(1):27–44.
- Widén, J. et Wäckelgård, E. (2010). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, 87(6):1880–1892.
- Wilke, U. (2013). *Probabilistic Bottom-up Modelling of Occupancy and Activities to Predict Electricity Demand in Residential Buildings*. PhD Thesis, École Polytechnique Fédérale de Lausanne, Switzerland.
- Wilke, U., Haldi, F., Scartezzini, J.-L., et Robinson, D. (2013). A bottom-up stochastic model to predict building occupants' time-dependent activities. *Building and Environment*, 60:254–264.
- Wolf, S., Møller, J. K., Bitsch, M. A., Krogstie, J., et Madsen, H. (2019). A Markov-Switching model for building occupant activity estimation. *Energy and Buildings*, 183:672–683.
- Xuan Hoa Binh, L., Kashif, A., Ploix, S., et Dugdale, J. (2010). Simulating inhabitant behaviour to manage energy at home. In *Conférence IBPSA France 2010*, pages 28–36, Moret-Sur-Loing, France.
- Yamaguchi, Y., Fujimoto, T., et Shimoda, Y. (2011). Occupant Behavior Model for Households to Estimate High- Temporal Resolution Residential Electricity Demand Profile. In *12th International Building Performance Simulation Association, BS2011*, volume 12, Sydney.
- Yan, D., O'Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., et Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107:264–278.
- Yang, Z., Li, N., Becerik-Gerber, B., et Orosz, M. (2014). A systematic approach to occupancy modeling in ambient sensor-rich buildings. *Simulation*, 90(8):960–977.

- Yousefi, F., Gholipour, Y., et Yan, W. (2017). A study of the impact of occupant behaviors on energy performance of building envelopes using occupants' data. *Energy and Buildings*, 148:182–198.
- Yu, Z. J., Haghghat, F., Fung, B. C. M., Morofsky, E., et Yoshino, H. (2011). A methodology for identifying and improving occupant behavior in residential buildings. *Energy*, 36(11):6596–6608.
- Zaraket, T. (2014). *Stochastic activity-based approach of occupant-related energy consumption in residential buildings*. PhDThesis, Ecole Centrale Paris, France.
- Zenissa, R., Syafei, A., Surahman, U., Sembiring, A., Pradana, A., Ciptaningayu, T., Ahmad, I., Assomadi, A., Boedisantoso, R., et Hermana, J. (2020). The Effect of Ventilation and Cooking Activities Towards Indoor Fine Particulates in Apartments. *Civil and Environmental Engineering*, 16.
- Zhang, Y., Jin, Y., et Li, Y. (2021). Enhanced energy harvesting using time-delayed feedback control from random rotational environment. *Physica D: Nonlinear Phenomena*, 422:132908.
- Zhou, S., Wu, Z., Li, J., et Zhang, X.-p. (2014). Real-time Energy Control Approach for Smart Home Energy Management System. *Electric Power Components and Systems*, 42(3-4):315–326.
- Zou, B. et Mishra, A. K. (2020). Appliance usage and choice of energy-efficient appliances: Evidence from rural Chinese households. *Energy Policy*, 146:111800.

