

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

MODÉLISATION DU COMPORTEMENT DES OCCUPANTS POUR
DES APPLICATIONS DE GESTION DE LA DEMANDE
D'ÉLECTRICITÉ DANS LE SECTEUR RÉSIDENTIEL

THÈSE PRÉSENTÉE
COMME EXIGENCE PARTIELLE DU
DOCTORAT EN GÉNIE ÉLECTRIQUE

PAR
LUIS FERNANDO RUEDA VASQUEZ

SEPTEMBRE 2021

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN GÉNIE ÉLECTRIQUE (PH.D.)

Direction de recherche :

Prof. Kodjo AGBOSSOU Directeur de recherche

Prof. Sousso KELOUWANI Codirecteur de recherche

Jury d'évaluation :

Prof. Kodjo AGBOSSOU Directeur de recherche

Prof. Sousso KELOUWANI Codirecteur de recherche

Prof. Loïc BOULON Président du jury

Prof. Javier SOLANO Évaluateur externe

Prof. Mohamed OUF Évaluateur externe

Thèse soutenue le 03 septembre 2021

Remerciements

Je tiens à remercier sincèrement mon directeur de recherche, le Professeur Kodjo Agbossou, et mon codirecteur, le Professeur Souso Kelouwani, pour la qualité de leur encadrement. Je voudrais également les remercier pour leurs conseils et commentaires constructifs, ainsi que pour leur orientation à toutes les étapes de ma recherche. Je saisis également l'occasion de remercier le Professeur Alben Cardenas et M. Nilson Henao qui ont suivi et supervisé de près le développement de mon travail. Je leur exprime ma gratitude pour leurs précieuses contributions et leur disponibilité tout au long de cette thèse.

J'aimerais également remercier les chercheurs du Laboratoire des Technologies de l'Énergie (LTE) d'Hydro-Québec pour leur contribution scientifique, ainsi que pour le temps qu'ils ont consacré à nos échanges, réunions et révisions de documents relatifs à mes travaux de recherche. D'autre part, je voudrais remercier les organismes subventionnaires qui ont fourni l'appui financier et les ressources nécessaires pour mener à bien ce travail de recherche et mon programme d'études. Ces organismes sont le LTE d'Hydro-Québec, le Conseil de Recherche en Sciences Naturelles et en Génie du Canada (CRSNG), la Fondation UQTR et Mitacs.

Finalement, je souhaiterais remercier toutes les personnes de l'Institut de Recherche sur l'Hydrogène, en particulier mes collègues du Laboratoire d'Innovation et de Recherche en Énergie Intelligente (LIREI) avec qui j'ai eu le plaisir de travailler et de partager de très bons moments.

Je remercie Dieu d'avoir été mon guide et ma force dans toutes les étapes de ma vie.

Je veux dédier ce travail de recherche à :

Mes grands-parents,

parce qu'ils ont été et seront toujours mon inspiration et la lumière qui éclaire mon chemin.

Mes parents,

pour m'avoir appris qu'avec du travail et du dévouement, les rêves peuvent devenir réalité.

Vous êtes mon modèle et exemple dans la vie.

Mon frère et ma belle-sœur,

pour leur amour, leurs conseils et leur soutien constant dans la réalisation de mes rêves.

Mes beaux-parents, mon beau-frère et sa famille,

merci pour votre soutien et amour inconditionnels tout au long de cette aventure.

Mon épouse,

parce que cela n'aurait pas été possible sans ton amour, ta compréhension et ton soutien.

Merci mon amour de m'avoir donné la force de mener à bien ce travail.

Tu donnes du sens à ma vie, je t'aime.

Résumé

Au Canada et plus particulièrement au Québec, pendant la période hivernale, le réseau électrique peut être fortement sollicité lors des pics de consommation. Les hivers rigoureux du Québec et la généralisation du chauffage électrique des bâtiments, en raison du faible coût des systèmes résistifs et de l'énergie renouvelable hydro-électrique, expliquent en partie cette utilisation intensive du réseau électrique. En plus de ces spécificités, le profil de consommation d'électricité résidentielle québécois est fortement influencé par les habitudes de vie des résidents comme, par exemple, les horaires d'occupation et les préférences de réglage de température du chauffage des logements. Néanmoins, ceci est également un levier important pour la gestion efficace du réseau électrique. Par conséquent, afin de répondre à la demande électrique, tout en maintenant le confort des clients, la gestion intelligente des charges résidentielles et une bonne connaissance des habitudes et préférences de consommation des ménages jouent un rôle fondamental pour l'utilité électrique.

Conformément aux tendances en matière de réseaux électriques intelligents, les avancées dans les méthodes d'intelligence artificielle (IA) et la popularité croissante de l'internet des objets (IoT), les systèmes de gestion de la demande résidentielle (HEMS) se développent très rapidement. Cependant, bien que plusieurs projets pilotes aient démontré les avantages de la gestion locale de la demande, la participation des utilisateurs est primordiale. L'acceptation et l'engagement de la part des clients avec la mise en œuvre des actions visant l'économie ou le déplacement de la consommation d'électricité influencent la performance de la gestion de la demande. De ce fait, la modélisation du comportement des individus et son intégration dans les outils de simulation des bâtiments et les HEMS sont des défis révélés à maintes reprises dans la littérature scientifique.

À travers cette thèse nous proposons des stratégies pour la modélisation du comportement des occupants, en particulier, l'occupation dans les bâtiments résidentiels. De cette manière, en utilisant des méthodes d'apprentissage axées sur les données, nous cherchons à contribuer avec une représentation réaliste de la présence des individus, tant dans des environnements simulés que dans des applications réelles axées sur la gestion énergétique des bâtiments. Pour atteindre cet objectif, la première partie du travail de recherche est consacrée à la modélisation probabiliste des profils individuels d'occupation dans des bâtiments résidentiels. Cette partie du travail considère la durée des états d'occupation et l'hétérogénéité du comportement des individus dans le calcul des probabilités de transition afin de générer des profils d'occupation quotidiens plus réalistes. Ainsi, les profils générés peuvent être utilisés dans des applications telles que la simulation de la performance des bâtiments, la prévision de la demande et la génération de profils de consommation en utilisant des approches ascendantes (bottom-up).

Par ailleurs, dans la deuxième partie de cette étude, nous cherchons à fournir aux systèmes de gestion locale de la demande des informations sur la probabilité de présence actuelle et future des individus. Ainsi, l'objectif est de contribuer au développement des stratégies de gestion intelligente du système de chauffage électrique permettant de maintenir un équilibre entre la consommation et le confort des clients. Pour atteindre cet objectif, nous avons exploré l'utilisation de modèles de Markov cachés à durée explicite et de modèles basés sur le hasard en tant que mécanismes pour la détection et la prévision non supervisée et en ligne de la présence des occupants. De plus, ces modèles sont intégrés à une architecture de gestion locale basée sur des méthodes d'apprentissage automatique, pour réaliser le contrôle prédictif par zone du système de chauffage par plinthe électrique d'une résidence. L'architecture proposée intègre un tarif dynamique, permettant à la fois de répondre aux besoins du client en termes de confort thermique des occupants et à ceux du distributeur en termes de réduction de la puissance en heures de pointe. De même, le caractère non supervisé et automatisé de la méthode proposée réduit l'intervention humaine lors de l'apprentissage, de l'exécution et de l'adaptation du système. Enfin, une infrastructure expérimentale a été mise en place pour valider les techniques proposées dans un système avec un environnement réaliste.

Table des matières

Remerciements	i
Résumé	iii
Table des matières	v
Table des figures	viii
Liste des acronymes	x
Chapitre 1 - Introduction	1
1.1 Contexte général de la recherche	1
1.2 Le comportement humain et la demande énergétique des bâtiments	5
1.3 Problématique de recherche.....	7
1.4 Objectifs et contributions.....	9
1.4.1 Objectifs	9
1.4.2 Contributions	10
1.5 Méthodologie adoptée	11
1.5.1 Étapes de la méthodologie	12
1.6 Infrastructure de recherche	14
1.7 Structure de la thèse.....	14
Chapitre 2 - État de l'art	16
2.1 Modélisation du comportement	16
2.1.1 Modélisation du comportement à partir des enquêtes d'emploi du temps	18
2.1.2 Modélisation du comportement à partir de mesures sur place	23
2.2 Gestion locale de la demande	28
2.2.1 Stratégies de commande	32

2.2.1.1	Méthodes de commande classique	33
2.2.1.2	Méthodes basées sur des règles ou des conditions	33
2.2.1.3	Méthodes basées sur le contrôle prédictif et optimal	34
2.2.2	Contrôle des bâtiments centré sur l'occupant	35
2.2.2.1	Prevision de l'occupation	36
2.3	Synthèse du chapitre	37
Chapitre 3 - Approche pour la modélisation de l'occupation dans les logements		39
3.1	Introduction	39
3.2	Article 1 : A Probabilistic Model of Household Occupant Behavior for Home Energy Management Applications	40
3.2.1	Contexte	40
3.2.2	Méthodologie	41
3.2.2.1	Hypothèses et considérations de la proposition	41
3.2.2.2	Étapes de l'approche proposée	41
3.2.3	Résultats	43
3.3	Synthèse du chapitre	59
Chapitre 4 - Approche pour la gestion de la demande résidentielle basée sur la détection et la prévision de l'occupation		60
4.1	Introduction	60
4.2	Article 2 : A comprehensive review of approaches to building occupancy detection	61
4.2.1	Contexte	61
4.2.2	Méthodologie	62
4.2.3	Résultats	62
4.3	Article 3 : Online unsupervised occupancy anticipation system applied to residential heat load management	64
4.3.1	Contexte	64
4.3.2	Méthodologie	65
4.3.2.1	Hypothèses et considérations de la proposition	65
4.3.2.2	Étapes de l'approche proposée	66

4.3.3	Résultats	67
4.4	Synthèse du chapitre.....	99
Chapitre 5 - Conclusions et recommandations		101
5.1	Recommandations.....	104
5.2	Annexe.....	106
Bibliographie		107
Annexe A - Système de mesure et de contrôle déployé		124
A.1	Système de mesure.....	124
A.1.1	L'impact des portes et fenêtres sur les variables mesurées	126
A.2	Modélisation de la dynamique thermique du bâtiment.....	128
A.3	Formulaire pour collecter la valeur d'occupation réelle du logement.....	130
Annexe B - Caractérisation du signal de puissance		132
B.1	Caractérisation du signal de puissance totale.....	132
B.1.1	Prétraitement du signal	133
B.1.2	Analyse d'autocorrélation	134
B.1.3	Détection des périodes de pointe locale	135
B.1.3.1	Analyse multi-échelle	135
B.1.3.2	Analyse du début et de la fin des périodes de pointe locale .	136
B.2	Analyse de charges domestiques.....	138
B.2.1	Chauffage des espace et éclairage	138
B.2.2	Chauffe-eau	139
B.2.2.1	Détection des événements	140
B.3	Analyse de corrélation croisée entre les habitudes de consommation.....	143
B.4	Détection des évènements de charge électrique.....	144

Table des figures

1-1	Variations de la température extérieure au Québec et exemple de son impact dans la demande d'électricité d'un résidence.	2
1-2	Puissance installée par Hydro-Québec et besoins de puissance en période de pointe [11]-[19].	3
1-3	Schéma bloc décrivant l'interaction entre les occupants et les bâtiments. Adapté de [30].	5
1-4	Exemple de la diversité de l'utilisation des appareils électroménagers basé sur le set de données IRISE [37].	6
1-5	Méthodologie de recherche adoptée.	12
2-1	Probabilité moyenne d'occupation par sous-populations.	19
2-2	Réseaux bayésiens des processus markovien de premier order, d'order supérieur et semi-markovien d'une séquence discrète à deux états.	20
2-3	Nombre d'articles par type de capteur [56].	24
2-4	Pourcentage d'articles par type d'algorithme [56].	27
2-5	Stratégies de gestion de la demande. Adapté de [5].	30
2-6	Architecture typique d'un HEMS [111].	32
2-7	Schéma conceptuel du contrôle des bâtiments centré sur l'occupant [131]. . .	35
3-1	Schéma bloc de l'approche proposé pour la modélisation de l'occupation dans les bâtiments résidentiels.	42
4-1	Structure de la stratégie de gestion centrée sur l'occupant.	66

A-1	Schéma de l'architecture proposée.	124
A-2	Photos du montage expérimental.	125
A-3	Exemple de l'impact des portes intérieures sur les variables mesurées.	127
A-4	Exemple de l'impact des portes intérieures sur le système de chauffage.	128
A-5	Circuit équivalent RC de la dynamique thermique des pièces.	129
A-6	Circuit équivalent RC de la dynamique thermique des pièces.	130
B-1	Schéma-bloc de l'approche proposée pour la caractérisation du signal de puissance.	133
B-2	Puissance totale d'une maison réelle pendant la saison d'hiver 2017-2018.	133
B-3	Exemple du résultat obtenu avec l'étape de lissage.	134
B-4	Autocorrélation du signal de puissance totale présenté dans la FIGURE B-2.	135
B-5	Noyaux multi-échelles utilisés.	136
B-6	Exemple des résultats obtenus avec l'analyse multi-échelle.	137
B-7	Distribution de probabilité pour le début et la fin des périodes de pointe obtenues avec l'échelle de 30 minutes et 120 minutes.	138
B-8	Exemple des résultats obtenus avec l'analyse de convolution.	139
B-9	Distribution de probabilité pour le début et la fin des périodes de plus forte demande d'énergie pour le chauffage des espace et l'éclairage.	140
B-10	Approche proposée pour la détection des événements du chauffe-eau.	140
B-11	Détection des événements du chauffe-eau.	141
B-12	Analyse de la durée des états de fonctionnement du chauffe-eau.	141
B-13	Analyse de la durée des états de fonctionnement du chauffe-eau pendant la journée.	142
B-14	Distribution de probabilité pour le début et la fin des périodes de plus forte demande d'énergie pour le chauffe-eau.	142
B-15	Analyse de corrélation croisée normalisée.	143

Liste des acronymes

ATC	Accumulateurs Thermiques Centraux
ATL	Accumulateurs Thermiques Locaux
CPP	Critical Peak Pricing
CVC	Chauffage, Ventilation et Climatisation
DR	Demand Response
DSM	Demand Side Management
ED	Enquêtes sur les Déplacements
EDHMM	Explicit Duration Hiden Markov Model
EET	Enquêtes d'Emploi du Temps
FOMC	First-Order Markov Chains
FSM	Finit State Machine
GDP	Gestion de la Demande de Puissance
GDR	Gestion de la Demande Résidentielle
GES	Gaz à Effet de Serre
HEMS	Home Energy Management Systems
HMM	Hiden Markov Model
HOMC	Higher-Order Markov Chains
IA	Intelligence Artificielle
IMA	Infrastructure de Mesurage Avancée
IoT	Internet des Objets
ML	Machine Learning

MPC	Model Predictive Control
NILM	Non-Intrusive Load Monitoring
RTP	Real-Time Pricing
RÉI	Réseau Électrique Intelligent
SPL	Systèmes de Production Locale
SST	Systèmes de stockage thermique
TOU	Time-of-use

Chapitre 1 - Introduction

Cette thèse a été réalisée dans le contexte de la gestion de la demande d'électricité chez les clients résidentiels que ce soit pour les bénéficiaires des fournisseurs d'électricité, tel qu'Hydro-Québec¹, mais également pour ceux de leurs utilisateurs. Ainsi, elle vise entre autres l'analyse et la proposition de nouvelles méthodes pour la modélisation du comportement des occupants, afin de proposer des stratégies de gestion de la consommation résidentielle permettant de favoriser les économies d'énergie et la diminution de l'appel de puissance, tout en respectant le confort des usagers. Nous présentons dans ce chapitre une description du contexte de la recherche, du problème abordé, des objectifs et de la méthodologie adoptée dans la thèse. Par la suite, nous décrivons les principales contributions.

1.1 Contexte général de la recherche

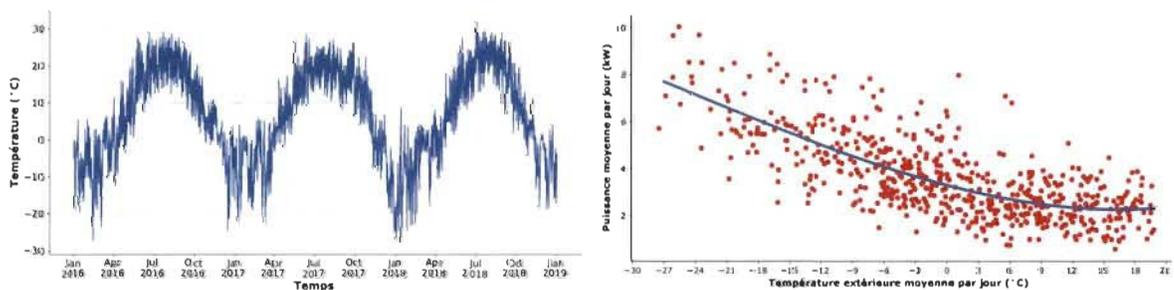
Le secteur énergétique est dans une phase de transformation fortement influencée par le réchauffement climatique, l'imminent épuisement des ressources d'énergie fossile et la transition énergétique qui entraîne une croissance de la demande d'électricité produite à partir d'énergies renouvelables. Dans le secteur énergétique, le service d'électricité est un facteur fondamental de l'activité économique moderne [1]. En conséquence, le réseau électrique évolue radicalement et se transforme en un Réseau Électrique Intelligent (RÉI), qui se caractérise par une optimisation pour l'utilité et pour le client des ressources disponibles qui ont tendance à se décentraliser, une participation active des clients à la production et à la transaction de l'électricité et une numérisation et sécurisation du réseau électrique [2]. Le RÉI est un réseau électrique moderne qui permet une communication bidirectionnelle entre les fournisseurs d'énergie et les consommateurs, l'intégration des énergies renouvelables et des

1. Hydro-Québec est responsable de la production, du transport et de la distribution de l'électricité au Québec.

véhicules électriques, la gestion de la demande en temps réel et la modélisation des habitudes de consommation des clients [2], [3]. Par ailleurs, dans le scénario des RÉI, les clients, en plus d'être des consommateurs, peuvent jouer le rôle de producteurs d'électricité et d'une certaine manière contribuer à la gestion du réseau.

Parmi les acteurs du réseau électrique, le secteur résidentiel fait l'objet d'une attention particulière en raison de la croissance de la consommation d'électricité, de son potentiel d'économie d'énergie, et de la gestion de la puissance durant les périodes de pointe [5], [6]. De ce fait, l'efficacité énergétique, la gestion de la demande de puissance (GDP) et la fiabilité du réseau sont devenues des priorités pour les gestionnaires de réseau de distribution.

Dans les pays nordiques comme le Canada², et en particulier dans la province de Québec³, où cette étude est menée, la consommation annuelle d'énergie dans le secteur résidentiel est fortement influencée par les saisons et les conditions climatiques. La FIGURE 1-1a présente les variations de la température extérieure à Trois-Rivières, QC pour une période de trois années consécutives (2016-2018) [4]. On observe dans cette période une température moyenne autour de 5°C, et des valeurs minimales proches de -30°C. De plus, la FIGURE 1-1b montre un exemple de la corrélation entre la demande d'électricité et la température extérieure pour un résidence québécois avec chauffage électrique des espaces dans laquelle nous avons effectué des mesures sur place. Cet exemple révèle comment la consommation d'énergie augmente



(a) Température extérieure - Trois-Rivières, QC [4]

(b) Demande d'électricité vs température extérieure

FIGURE 1-1 Variations de la température extérieure au Québec et exemple de son impact dans la demande d'électricité d'un résidence.

2. Dans le contexte international, en 2017, le Canada était le sixième plus grand producteur et consommateur d'énergie électrique au monde [7].

3. En 2015, le Québec était la province ayant la plus forte consommation d'électricité au Canada avec 35%, suivi de l'Ontario et de l'Alberta avec 28% et 11% respectivement [8].

avec la baisse des températures, ce qui entraîne un écart important entre la consommation d'été et d'hiver. En fait, des études ont révélé que la courbe de puissance est étroitement liée au chauffage des espaces et de l'eau. Ces deux charges représentent ensemble environ 80% des besoins d'énergie des ménages québécois [9]. Néanmoins, le chauffage des espaces suscite un intérêt particulier, car il peut entraîner une augmentation significative de la demande d'électricité pendant les heures de pointe en raison de l'utilisation du chauffage par un nombre élevé de clients en même temps [10].

La FIGURE 1-2 montre les moments lorsque la demande de puissance a atteint ou a dépassé la capacité installée du réseau d'Hydro-Québec au cours des dernières années. Ces pointes de consommation se sont produites en hiver généralement au cours de la semaine, entre 6h et 9h ou entre 16h et 20h. Cela coïncide avec les périodes de la journée où les personnes sont les plus actives dans la maison, produisant un effet sur la demande d'électricité connue comme les périodes de pointe quotidiennes [20]. L'effet des habitudes ou routines de gens dans les différents secteurs de l'économie explique cette forte consommation. Par exemple, dans une résidence, les occupants prennent une douche et préparent un repas le matin, avant de partir au travail ; puis, le soir, au retour du travail, ils montent la température, préparent un repas et regardent la télévision. Ces comportements vont affecter largement la demande d'énergie dans des périodes spécifiques. Par ailleurs, la FIGURE 1-2 montre que durant les années 2010, 2012

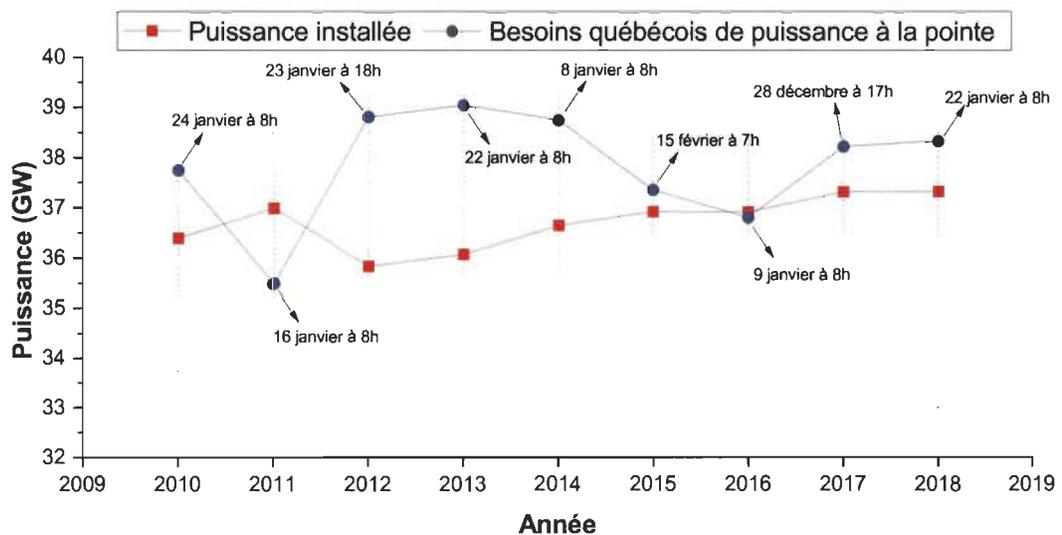


FIGURE 1-2 Puissance installée par Hydro-Québec et besoins de puissance en période de pointe [11]-[19].

à 2015, 2017 et 2018, la capacité installée a été dépassée par la demande pendant la période de pointe, ce qui oblige Hydro-Québec à faire des achats à court terme sur des marchés voisins, à prix très élevés [21]. De plus, la présence de ces périodes de pointe exige que Hydro-Québec accorde des investissements additionnels, car le réseau doit être sur-dimensionné afin de pouvoir répondre à tout moment aux dépassements momentanés de puissance. Également, vu que l'efficacité du réseau de distribution est affectée (p. ex. augmentation des pertes de transport), Hydro-Québec doit faire des efforts supplémentaires pour pouvoir maintenir les niveaux de qualité exigés par la loi et les autorités, qui surveillent le fonctionnement des systèmes de distribution et de transport électrique.

Donc, tenant compte des problématiques liées à l'utilisation de l'énergie qui sont décrites précédemment, la gestion de la demande d'électricité au niveau résidentiel (GDR) est une des avenues prometteuses pour réduire les pointes de consommation et diminuer l'empreinte de carbone associée à la consommation d'énergie. Ainsi, l'utilisation des systèmes de gestion de la demande résidentielle aussi connus comme HEMS (Home Energy Management System, en anglais) permettrait d'améliorer les fonctions du réseau et de mieux moduler la courbe de charge, par exemple, en déplaçant la consommation à des périodes plus convenables. Toutefois, pour aborder le problème de la gestion énergétique des bâtiments, il est nécessaire d'adopter une approche holistique incorporant tous les facteurs qui influencent la consommation d'énergie des bâtiments, y compris les besoins particuliers des clients. En fait, les habitudes quotidiennes des clients et leurs préférences en matière de confort peut introduire de grandes fluctuations dans la capacité de flexibilité de la charge⁴, par conséquent, sa modélisation est parmi les principaux défis pour appliquer des stratégies de gestion de la demande [24]. Par exemple, Clewanger et al. [25] ont montré à travers des simulations paramétriques et la modélisation énergétique que des changements dans le comportement des occupants peuvent modifier la consommation annuelle d'énergie jusqu'à 75% dans les bâtiments résidentiels et 150% dans les bâtiments commerciaux. Par conséquent, au cours des dernières années, la modélisation du comportement est devenue un domaine de recherche important conformément

4. La flexibilité énergétique d'un bâtiment est définie dans l'annexe 67 du programme IEA-EBC (International Energy Agency-Energy in Buildings and Communities Programme, en anglais) comme « la capacité de gérer la demande et la production en fonction des conditions climatiques locales, des besoins des utilisateurs et des exigences du réseau » [22], [23].

aux tendances en matière de gestion de la demande et du réseau électrique intelligent.

1.2 Le comportement humain et la demande énergétique des bâtiments

Dans le contexte de l'analyse énergétique des bâtiments, le comportement humain fait référence à l'interaction entre les occupants et les immeubles afin d'atteindre le niveau de confort et de sécurité qu'ils souhaitent, et pour préserver un environnement intérieur sain [26], [27]. Il peut être décrit par la présence humaine et son interaction avec les appareils, les systèmes de contrôle (p. ex. le chauffage, ventilation et climatisation (CVC) et les éléments du bâtiment (p. ex. les portes et fenêtres) [28]. Par ailleurs, le comportement humain peut être modélisé à différents niveaux de résolution, en commençant par l'état d'occupation d'un espace intérieur (présence/absence) et en augmentant le degré de détail avec le nombre de personnes, leurs attributs et même les activités qu'elles réalisent [29].

La FIGURE 1-3 montre que la présence et le comportement des occupants ont un impact indéniable sur la performance énergétique des bâtiments. Ainsi, lorsque des personnes se trouvent dans une pièce, même sans interagir avec l'environnement, elles modifient le micro-climat intérieur [31]. La chaleur provenant du corps humain et le processus de respiration affectent particulièrement la température intérieure, l'humidité et la concentration de CO₂. De plus, lorsque les personnes interagissent activement avec le bâtiment (p. ex. la préparation des repas, en ouvrant et en fermant les fenêtres ou en allumant les lumières), leur influence est encore plus considérable [32], ce qui se reflète également sur la consommation et le coût

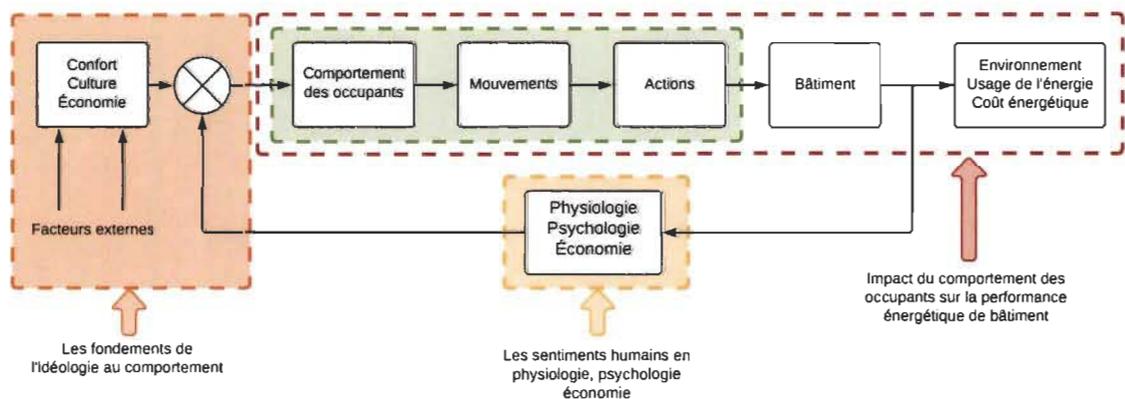


FIGURE 1-3 Schéma bloc décrivant l'interaction entre les occupants et les bâtiments.
Adapté de [30].

de l'énergie des bâtiments. Un facteur à considérer est la popularité croissante de l'internet des objets (IoT), et des maisons intelligentes. Ces technologies apprennent des habitudes des utilisateurs, en automatisant l'utilisation, par exemple, de l'éclairage ou des systèmes CVC, permettant ainsi d'optimiser le confort des utilisateurs et la consommation d'énergie des bâtiments. Par conséquent, la causalité entre la présence, le comportement et la performance énergétique des bâtiments peut être modifiée, entraînant de nouvelles dynamiques entre les utilisateurs et les immeubles.

Le schéma bloc présenté à la FIGURE 1-3 montre également que le comportement humain est affecté par différents facteurs externes, comprenant la dimension économique, sociologique, psychologique, physiologique et ergonomique. L'intégration de ces facteurs externes dans la modélisation du comportement est un aspect important qui a été mis en évidence dans la littérature [33]. Cela permet une description plus robuste des motivations qui conduisent les occupants à interagir avec l'enveloppe et les systèmes du bâtiment pour obtenir les conditions de confort désirées [34]. De ce fait, Zaraket [35] et Robinson [36] relèvent la présence des occupants dans un bâtiment et les activités qu'ils y entreprennent comme des processus stochastiques complexes à comprendre, à mesurer et à anticiper.

La littérature scientifique a montré que le climat, les caractéristiques sociodémographiques et économiques, et la nature stochastique du comportement humain entraînent des variations considérables de la consommation d'énergie entre les ménages [35]. Par exemple, Seryak et

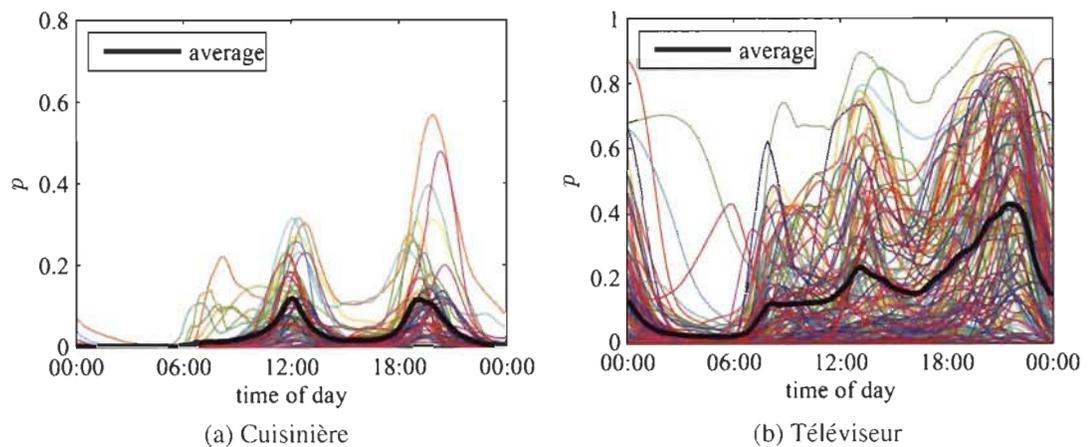


FIGURE 1-4 Exemple de la diversité de l'utilisation des appareils électroménagers basé sur le set de données IRISE [37].

Kissock [38], ainsi que Emery et Kippenhan [39] ont analysé la consommation d'énergie d'un groupe de maisons sur une longue période, et ont constaté que différents occupants peuvent avoir des consommations d'énergie variées en raison de leurs habitudes comportementales particulières. En fait, tel que présenté dans la FIGURE 1-4, l'hétérogénéité du comportement humain et les facteurs externes l'affectant sont reflétés dans la disponibilité et l'utilisation des appareils électroménagers tels que les cuisinières ou dans les activités comme regarder la télévision.

Par conséquent, au cours des dernières années, un intérêt particulier a été accordé à la modélisation du comportement humain. Notamment, les IEA-EBC (International Energy Agency – Energy in Buildings and Communities) Annexes 66 [40] et 79 [41] ont contribué à la définition et à l'étude du comportement humain dans les bâtiments. Ils visent ainsi à contribuer à l'intégration de l'occupation et le comportement des occupants dans le processus de conception et l'exploitation des bâtiments afin d'améliorer à la fois la performance énergétique et le confort des occupants. Dans cette perspective, [42], [43], et [44] ont montré qu'une conception robuste et optimale des bâtiments est possible grâce à une meilleure compréhension du comportement humain. Ceci permet de réduire, dans les bâtiments, la consommation d'énergie ainsi que sa variation due à l'influence des occupants. Par ailleurs, des travaux tels que [45], [46] et [47] ont montré qu'en identifiant les habitudes de consommation des occupants et la relation entre les différentes charges, il est possible d'identifier les appareils qui ont un impact plus important sur les périodes de pointe, et de modéliser les variations du comportement dans le temps. Ainsi, il a été démontré que des économies d'énergie allant jusqu'à 30% peuvent être réalisées en appliquant des stratégies de contrôle basées sur l'occupant [48], [49].

1.3 Problématique de recherche

Au cours des dernières années, une attention importante a été accordée à l'amélioration de l'efficacité des systèmes et des technologies des bâtiments. De ce fait, son enveloppe et les équipements mécaniques et électriques améliorent leur performance. Cependant, le comportement des occupants a été sous-estimé lors de la conception, la construction, l'exploitation et la rénovation des bâtiments [41]. Cela s'explique en partie du fait que le comportement humain est complexe, stochastique, multidisciplinaire et difficile à mesurer [50], [51]. Cette

compréhension restreinte donne lieu à des hypothèses inappropriées. Cela se traduit par des résultats inexacts sur la performance énergétique des bâtiments, ainsi que des écarts importants dans leur optimisation, conception, diagnostic et simulation énergétique [30].

Le besoin d'une meilleure compréhension du comportement humain et de son impact sur la consommation énergétique des bâtiments devient plus important. En fait, des études ont révélé que l'un des plus grands obstacles à la mise en œuvre efficace des techniques de réponse à la demande dans le secteur résidentiel est l'absence d'une bonne connaissance des habitudes et des préférences de consommation des ménages [52]. Par conséquent, au cœur de cette thèse se trouve le défi révélé à maintes reprises dans la littérature, qui est de modéliser efficacement le comportement des individus. Notamment, nous ciblons la clientèle résidentielle en raison de son impact sur le réseau électrique québécois [9], [53]. Par ailleurs, nous adressons particulièrement la problématique de la modélisation de la présence/absence des individus, en tant qu'élément fondamental pour la compréhension du comportement des humains dans leurs résidences [29].

À son tour, cette problématique entraîne des défis spécifiques selon le type d'application ou d'analyse à réaliser. D'un côté, dans les environnements simulés, notamment dans les outils de simulation des bâtiments, ainsi que les applications de prévision et gestion de la demande, le comportement humain est généralement représenté de façon simplifiée (p. ex. des profils d'occupation prédéfinis), ou dans certains cas, il est tout simplement négligé. Cependant, des études ont montré l'importance d'utiliser des profils d'occupation et d'activités réalistes pour améliorer la performance des analyses effectuées [54]. En fait, Vorger [55] et Jia et al. [46] ont souligné que dans les outils de simulation énergétique, l'incertitude causée par le comportement des occupants est à l'origine d'un écart important entre les résultats de simulations et les mesures *in situ*. Pour adresser cette problématique, différents types de modèles peuvent être trouvés dans la littérature. Cependant, l'analyse des modèles comportementaux existants nous a permis d'identifier des limites sur des aspects tels que l'intégration de la dépendance temporelle du comportement et la représentation de sa variabilité en fonction des caractéristiques sociodémographiques des individus. De plus, les approches existantes négligent d'autres aspects comme l'interaction entre les utilisateurs d'un même ménage, ainsi que l'intégration des périodes d'absence prolongée et de maladie. Une discussion

plus détaillée sur ce sujet est présentée au Chapitre 2.

D'un autre côté, dans les cas pratiques, de nombreux efforts utilisant des mesures sur place ont été entrepris pour la modélisation de la présence/absence des individus. Comme résultats, un nombre importante des méthodes et technologies ont déjà été explorés dans ce sujet de recherche. De cet fait, il s'avère nécessaire un étude permettant d'identifier les points forts, les limites, les leçons apprises et les opportunités d'amélioration des approches déjà étudiées dans la littérature. Par ailleurs, la collecte de données sur l'occupation et le comportement humain a été mise en évidence comme l'un des principaux défis pour le développement de méthodes permettant sa modélisation. Le coût, le respect de la vie privée et d'autres facteurs socioéconomiques figurent parmi les principaux obstacles [40]. En raison de cette difficulté à observer l'occupation de façon explicite, les approches d'apprentissage supervisé couramment utilisées pour la détection de l'occupation sont difficilement applicables aux situations pratiques [56]. Une discussion plus détaillée à ce sujet est présentée au Chapitre 2. De plus, cela rend le défi encore plus complexe lors de la gestion des systèmes de chauffage par plinthe électrique. À cause de l'inertie thermique des bâtiments, les systèmes de chauffage ont besoin d'une période de temps considérable pour influencer la température d'une pièce. Ainsi, la détection et l'anticipation en ligne et non supervisée de l'occupation deviennent des défis importants à tenir en compte lors de la mise en œuvre des stratégies de gestion de la demande. Ceci est dans le but de maintenir un équilibre entre la consommation et le confort des clients simultanément.

1.4 Objectifs et contributions

1.4.1 Objectifs

L'objectif principal de cette thèse consiste à proposer des stratégies pour modéliser le comportement des occupants, en particulier la présence/absence des individus dans leurs résidences. Ce projet de recherche vise à favoriser une représentation réaliste de l'occupation des bâtiments résidentiels dans les outils de simulation énergétique et les systèmes de gestion locale de la demande. Les objectifs spécifiques de ce travail sont définis comme suit :

- Conception d'une approche stochastique permettant la modélisation des profils individuels d'occupation des résidences en tenant compte de manière explicite de leur dynamique temporelle et des caractéristiques sociodémographiques des individus.
- Réalisation d'une étude approfondie de la littérature scientifique en lien avec les techniques de la détection et la modélisation de la présence des occupants dans les bâtiments à partir des mesures *in situ*.
- Développement et validation expérimentale d'une approche non supervisée pour la détection et la prévision en ligne de l'occupation des résidences en utilisant des données mesurées sur place.
- Proposition et validation expérimentale d'une stratégie de contrôle prédictif et adaptatif des systèmes de chauffage par plinthe électrique en tenant compte de la présence d'occupants et d'un prix variable de l'énergie.

1.4.2 Contributions

Les principales contributions du projet sont :

- Conception d'un modèle stochastique à deux états basés sur l'analyse statistique et probabiliste des profils individuels d'occupation des résidences. Ce modèle permet la génération des profils d'occupation quotidiens réalistes, en se basant sur l'adaptation des réseaux bayésiens utilisés dans la modélisation de processus markoviens à durée explicite. De plus, le modèle proposé permet la prise en compte de la dynamique temporelle et la variabilité du comportement humain en utilisant une loi de transition non seulement conditionnée explicitement par la durée des états d'occupation, mais aussi par les caractéristiques sociodémographiques des individus.
- Développement d'un modèle stochastique à deux états basé sur les données mesurées sur place pour la détection et l'anticipation des périodes de présence/absence des occupants dans les logements. L'approche proposée minimise l'intervention humaine pendant le processus d'apprentissage du modèle et fournit une détection et une prévision en ligne de l'occupation du logement. À cette fin, l'approche proposée repose sur l'utilisation des modèles de Markov cachés à durée explicite et des modèles basés sur le hasard.

- Proposition d'une approche de gestion locale basée sur les méthodes d'intelligence artificielle, pour réaliser un contrôle prédictif du système de chauffage par plinthe électrique des logements. Cette approche comporte l'utilisation d'une fonction de coût pondérée par les informations sur l'occupation, incluant un prix variable de l'énergie pour contrôler le système de chauffage visant à maintenir l'équilibre entre la demande d'énergie et le confort des clients.
- Conception d'une architecture expérimentale pour la validation des systèmes de détection de l'occupation et de la gestion de la demande résidentielle. Le système permet la mesure des variables environnementales, le mouvement des occupants et de la consommation d'énergie, en favorisant également l'intégration de thermostats communicants. Ce système se sert de l'infrastructure de communication WiFi existante dans la plupart des domiciles, et utilise des équipements commerciaux et des logiciels open source, tandis que son déploiement s'effectue à partir d'un système embarqué à faible coût.

1.5 Méthodologie adoptée

La méthodologie adoptée dans cette thèse est décrite conformément à la FIGURE 1-5. Tout d'abord, une première phase de recherche bibliographique a permis d'identifier et de définir le problème à adresser. Par la suite, l'exploration détaillée de la littérature scientifique a été effectuée afin d'acquérir une maîtrise approfondie des notions relatives au domaine d'étude. Cette démarche a permis d'identifier les méthodes existantes dans l'état de l'art, ainsi que leurs limitations, contraintes et difficultés. Puis, des améliorations, des ajustements et de nouvelles méthodes sont proposés afin de répondre aux critères, aux exigences et aux particularités du problème abordé. Dans une dernière étape, l'efficacité des propositions est évaluée par simulation et expérimentalement, en effectuant des études comparatives avec les techniques existantes dans la littérature. Pour cela, nous utilisons des données réelles et des enquêtes d'emploi du temps. Quant aux données réelles, elles ont été obtenues à partir de campagnes de mesures effectuées dans des résidences situées dans la province du Québec.

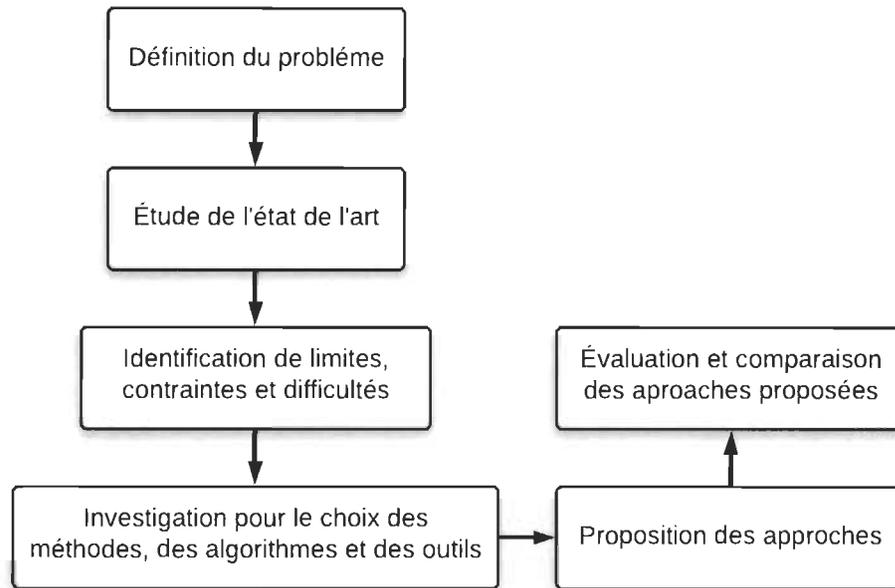


FIGURE 1-5 Méthodologie de recherche adoptée.

1.5.1 Étapes de la méthodologie

Dans ce qui suit, les étapes de la méthodologie sont détaillées conformément aux propositions décrites dans la section précédente.

- **Modélisation des profils individuels d'occupation** : L'étude exploratoire nous a permis d'identifier différentes approches et méthodologies proposées actuellement pour la modélisation de la présence. Notamment, nous avons retenu les modèles génératifs, en particulier les modèles Markoviens et certaines variantes connexes. Comme leur nom l'indique, les modèles génératifs sont capables de générer des données synthétiques à partir de l'échantillonnage des distributions. De plus, ces méthodes issues de l'analyse Bayésienne comportent des avantages tels que la possibilité de modéliser la structure temporelle des données et de gérer les incertitudes, ainsi que d'intégrer des connaissances *a priori*. Par ailleurs, cette thèse cible des améliorations dans le calcul de probabilités de transition des modèles stochastiques discrets en considérant des modèles sur la durée des états d'occupation plus adaptés au comportement réel des individus. Pour ce faire, des analyses de survie en utilisant des approches semi-paramétriques sont utilisées. Ces méthodes, contrairement aux modèles paramétriques, ne restreignent pas la

forme de la fonction de survie à une distribution particulière, favorisant la représentation des distributions de durée avec une forme multimodale. Enfin, différentes mesures de performance sont analysées et certaines méthodes de référence sont sélectionnées pour valider la méthode proposée.

- Détection et anticipation de l'occupation : Initialement, une analyse approfondie de l'état de l'art est réalisée pour identifier les technologies et méthodes existantes. Suite à cette revue de la littérature et conformément aux travaux présentés dans la première partie de la thèse pour la modélisation des profils d'occupation, les modèles de Markov cachés sont choisis comme approche pour effectuer une détection non supervisée de l'occupation. Par ailleurs, afin de fournir une meilleure représentation de la dynamique temporelle des profils de présence, des analyses de survie non paramétriques sont utilisées pour modéliser la durée des états d'occupation. Ainsi, un modèle de Markov caché à durée explicite est mis à profit pour la détection en ligne de la présence des individus. De plus, les modèles basés sur le hasard sont retenus pour prédire la présence des occupants. Ces modèles présentent une approche qui permet de saisir les probabilités de transition dépendant de la durée et d'incorporer des variables exogènes. Finalement, la performance des méthodes proposées est analysée de façon expérimentale.
- Gestion prédictive des systèmes de chauffage par plinthe électrique : En raison des conditions climatiques particulières du Québec, de l'impact du comportement humain sur la consommation électrique des ménages et de la grande influence du chauffage électrique sur les périodes de pointe, nous proposons une stratégie de gestion basée sur l'occupation pour le contrôle des systèmes de chauffage par plinthes électriques. De plus, compte tenu de l'inertie thermique des bâtiments, une stratégie de contrôle prédictif à horizon fini est proposée, laquelle intègre la prédiction de la présence des occupants et un tarification dynamique de l'énergie. Par ailleurs, pour valider la performance du système, une validation expérimentale est effectuée dans une résidence réelle pendant trois mois d'hiver.

1.6 Infrastructure de recherche

Toutes les ressources informatiques (ordinateurs, licences logicielles, etc.) et électroniques nécessaires pour l'implantation du système expérimentale et pour les études de modélisation ont été fournies par le Laboratoire d'Innovation et de Recherche en Énergie Intelligente (LIREI) de l'Institut de Recherche sur l'Hydrogène (IRH) et par le Département de Génie Électrique et Génie Informatique de l'Université du Québec à Trois-Rivières (UQTR). Des campagnes de mesure ont été réalisées dans des bâtiments résidentiels québécois afin d'obtenir certaines informations utiles à la validation des propositions. De plus, le groupe de recherche dispose d'un appartement expérimental doté d'un système permettant le sous-mesurage de la puissance active et réactive et de l'instrumentation nécessaire pour effectuer des essais et valider les modèles et les systèmes de gestion développés.

1.7 Structure de la thèse

Cette thèse est présentée en 5 chapitres, la suite du document est organisée ainsi :

- **Chapitre 2 :** Dans ce chapitre l'état de l'art général du problème de la modélisation de l'occupation et du comportement des occupants, les stratégies pour l'acquisition des données sur le comportement humain, ainsi que les stratégies et les méthodes utilisées pour le contrôle et la gestion locale de la demande sont présentées.
- **Chapitre 3 :** Ce chapitre présente sous la forme d'un article scientifique le travail effectué pour atteindre le premier objectif de la thèse. Dans cet article, nous présentons la méthode proposée pour la modélisation des profils d'occupation dans les logements. De plus, il comprend une étude approfondie de l'état de la technique correspondant au sujet traité, ainsi il décrit en détail la formulation mathématique de la méthode proposée, son évaluation et sa comparaison avec des approches précédentes de la littérature.
- **Chapitre 4 :** Nous présentons dans ce chapitre l'approche proposée pour la gestion de la demande résidentielle basée sur la détection et la prévision de l'occupation. Le travail réalisé dans cette partie de la thèse est présenté sous la forme de deux articles scientifiques à travers lesquels nous avons atteint les trois derniers objectifs de la thèse. Le premier de ces articles est consacré à la réalisation d'une revue de l'état de l'art

sur les technologies et les méthodes utilisées pour la détection et la modélisation de l'occupation. Par ailleurs, le deuxième des articles présentés dans ce chapitre est axé sur la formulation mathématique et la validation des méthodes proposées pour la détection et prévision en ligne de la présence des individus. Ainsi, il est présenté la conception et mise en œuvre d'une stratégie de gestion intelligent pour le système de chauffage par plinthe électrique basée sur l'occupation et un prix variable de l'énergie. De plus, l'architecture expérimentale développée pour la validation du système proposé est également décrite.

- **Chapitre 5 :** Finalement, nous présentons la conclusion, les recommandations et des travaux futurs possibles pour la suite de cette étude.

Chapitre 2 - État de l'art

Afin d'aborder chacune des thématiques liées à cette recherche, ce chapitre présente une description abrégée de l'état de l'art sur les différentes approches utilisées pour la modélisation de l'occupation et du comportement des occupants. Une analyse des approches pour la détection et prévision de l'occupation et des méthodes proposées dans la littérature pour la gestion de la demande résidentielle sont également présentées.

2.1 Modélisation du comportement

En raison de l'impact des habitudes humaines sur la consommation d'énergie des bâtiments, le développement de modèles du comportement des occupants est crucial afin de favoriser les économies d'énergie tout en respectant le confort des utilisateurs. Au cours des dernières années, la modélisation du comportement des occupants, la détection de l'occupation et la reconnaissance des activités ont attiré l'attention des chercheurs qui s'intéressent à la conception et à l'exploitation des bâtiments [28]. Ces modèles peuvent être utilisés pour des applications telles que la simulation de la performance des bâtiments et la prévision de la demande. Ces derniers peuvent aider le fournisseur électrique à anticiper les pics de demande d'énergie et à prendre des décisions sur la gestion de l'énergie, la commutation de la charge, la reconfiguration du réseau et le développement des infrastructures [57], [58]. Les modèles comportementaux peuvent être catégorisés en fonction de leur objectif et de leur approche de modélisation. Fabi [59] a distingué deux objectifs principaux, à savoir, la compréhension du comportement lui-même ou la détection des relations entre le comportement et la demande d'énergie, ainsi que les facteurs de variation. Cette thèse se concentre sur ce dernier objectif. En ce qui concerne les approches de modélisation, ils sont discutés ci-après :

- Approche déterministe : Il s'agit de la méthode la plus utilisée. Avec cette approche, le comportement humain est modélisé à partir des scénarios prédéfinis par l'utilisateur. Par exemple, des profils d'occupation définis selon l'heure de la journée [60].
- Approche stochastique : Contrairement aux méthodes déterministes, qui produisent une sortie unique, les modèles stochastiques utilisent des analyses probabilistes pour estimer la probabilité qu'une action donnée se produise à partir d'une entrée observée [60]. Les modèles probabilistes de comportement sont généralement établis à partir de campagnes de mesures au cours desquelles les variables étudiées (p. ex. la présence, le nombre de personnes ou l'état ouvert ou fermé d'une fenêtre) sont enregistrées en même temps que des données permettant de décrire l'environnement ou la population.
- Autres approches : D'autres approches ont été proposées dans la littérature, à savoir, l'approche orientée agent, l'approche sociologique et les modèles basés sur des actions. L'approche orientée agent permet de modéliser des systèmes aux dynamiques complexes composés d'agents autonomes qui interagissent entre eux et avec leur environnement. Dans le cas du comportement, cette méthode est basée sur trois aspects, à savoir, une population d'agents définis par des attributs et des comportements, un ensemble de règles définissant les relations et les interactions entre les agents et, finalement, un environnement avec lequel les agents interagissent. Cependant, l'objectif principal des agents est généralement défini pour atteindre le confort thermique, ce qui constitue une hypothèse forte souvent insuffisamment pour représenter un comportement réaliste [55]. L'approche sociologique vise à mieux comprendre le comportement lui-même, plutôt qu'à quantifier l'impact sur la consommation d'énergie [60]. Guy et al. [61] affirment que les actions visant à économiser l'énergie ne sont pas le résultat d'une décision rationnelle et informée prise par un individu. Au lieu de cela, ils affirment qu'il est nécessaire d'étudier les structures sociales dans lesquelles les décisions sont prises. Les modèles basés sur des actions définissent les comportements des occupants comme des actions (mouvements et actions de contrôle) [59]. Par rapport aux approches déterministes, ces modèles prennent en compte le caractère aléatoire qui résulte du changement irrégulier et non synchrone de l'occupation dans l'espace et le temps. De plus, comparé aux méthodes stochastiques, ces modèles conservent la relation

temporelle et spatiale de l'occupation et se révèlent plus pratique en raison du nombre réduit d'entrées.

La littérature scientifique a révélé que le comportement humain correspond à un processus stochastique et non déterministe, illustré par Nicol et Humphreys [50] comme suit : « il n'y a pas une température précise à laquelle tout le monde ouvre sa fenêtre, mais plus la température est élevée, plus la probabilité que la fenêtre soit ouverte est forte ». Par ailleurs, il a été mis en évidence que des facteurs externes, tels que la météo, peuvent affecter le comportement humain [62]. Cependant, la collecte de données sur le comportement humain représente l'un des principaux défis pour le développement de modèles stochastiques du comportement, en raison de critères tels que le coût et la vie privée des personnes [40]. Toutefois, plusieurs efforts ont été déployés au cours des dernières années pour surmonter ces enjeux.

Parmi les techniques plus utilisées pour la collecte de données sur le comportement humain figurent les enquêtes et les mesures sur place [40]. *Les enquêtes* peuvent fournir des informations détaillées sur la routine quotidienne d'un grand nombre de répondants et saisir la nature stochastique du comportement des occupants et sa variabilité en fonction des caractéristiques sociodémographiques [48]. *La méthode sur place* consiste à collecter et à surveiller les occupants dans leur environnement naturel (p. ex. à la maison ou au bureau), en utilisant des capteurs existants ou supplémentaires. La section 2.1.1 présente une analyse des modèles comportementaux basés sur des données d'enquête. La section 2.1.2 est consacrée aux modèles pour la détection de la présence basés sur des données mesurées sur place.

2.1.1 Modélisation du comportement à partir des enquêtes d'emploi du temps

Une bonne compréhension du comportement des individus à l'intérieur et à l'extérieur de la maison est d'une grande importance pour des secteurs tels que l'énergie [63] et les transports [64]. La littérature révèle que les enquêtes d'emploi du temps (EET) et les enquêtes sur les déplacements (ED) sont des sources d'information importantes pour atteindre une telle compréhension [65], [66]. En particulier, les données du sondage EET ont été largement utilisées pour le développement de modèles probabilistes de comportement des occupants, portant principalement sur les individus dans le secteur résidentiel [67]. Les données du sondage EET fournissent des informations détaillées sur la routine quotidienne d'un grand

nombre de répondants et permettent de capter la variabilité et la nature stochastique du comportement humain [26]. Un exemple de la diversité du comportement humain est présenté dans la FIGURE 2-1. Cette analyse a été générée en utilisant l'EET canadienne, cycle 29, et réalisée entre avril 2015 à avril 2016, comprenant un échantillon de 15,390 répondants [68].

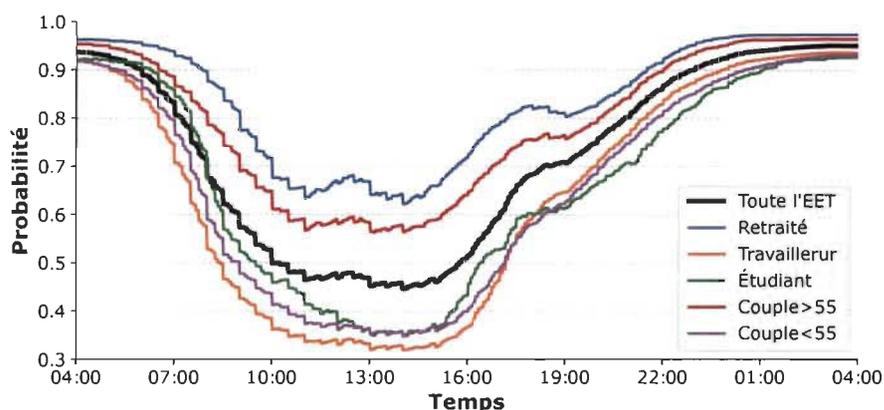


FIGURE 2-1 Probabilité moyenne d'occupation par sous-populations.

Cette figure montre la probabilité d'occupation des différentes sous-populations, telles que les retraités et les travailleurs, faisant ressortir l'importance de représenter de manière appropriée l'hétérogénéité du comportement des individus. Par ailleurs, la littérature révèle qu'indépendamment du pays dans lequel l'enquête est réalisée, ils fournissent un ensemble similaire d'informations sur les répondants, par exemple l'âge, le genre, l'état civil et l'activité principale. En conséquence, cela permet le développement de méthodes pouvant être répliquées, quel que soit l'enquête utilisée.

Parmi les techniques employées pour modéliser le comportement humain à l'aide des données du sondage EET, les chaînes de Markov de premier ordre (FOMC) et les chaînes de Markov d'ordre supérieur (HOMC) sont les approches les plus utilisées. La FIGURE 2-2 montre trois réseaux bayésiens décrivant le comportement d'une machine à deux états à travers le temps discret. Le premier réseau correspond au processus markovien classique dont l'état actuel est seulement conditionné par l'état précédent. Ceci est un processus sans mémoire, donc la durée de l'état n'est pas considérée dans le calcul des probabilités de transition. Le deuxième réseau illustre la représentation d'un processus markovien d'ordre N , dans lequel N a été défini égal à 3, de sorte que l'état actuel est conditionné par les trois états précédents. Le troisième réseau est un processus semi-markovien segmenté, également classé dans la

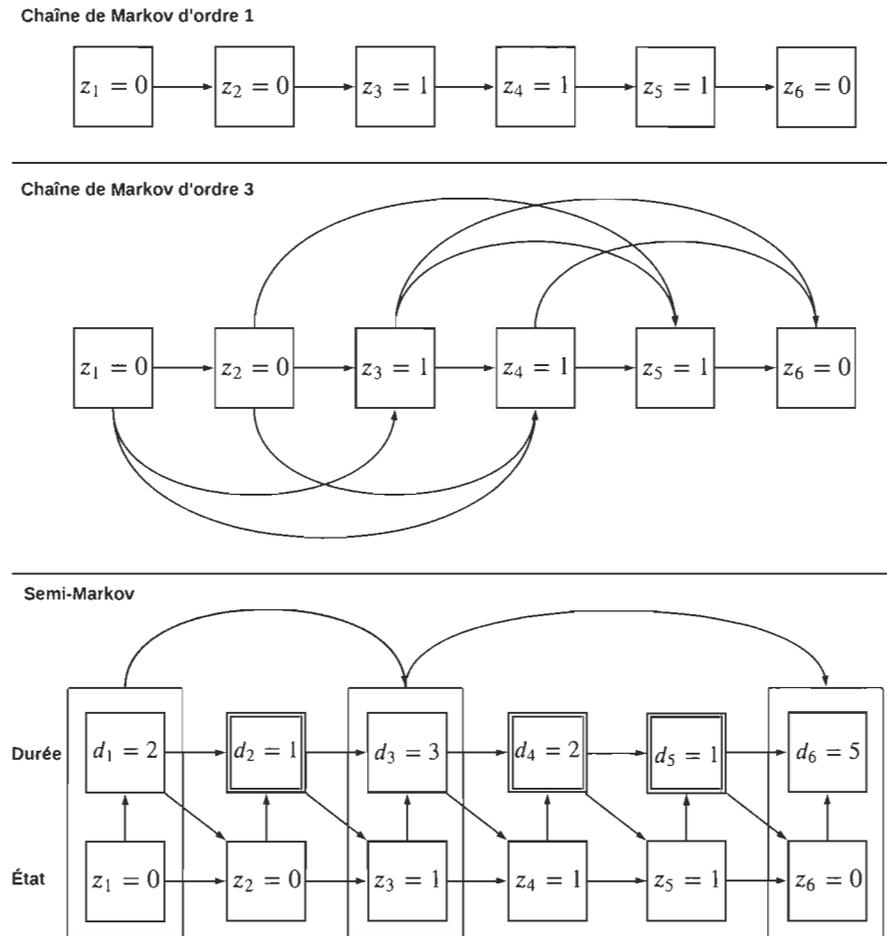


FIGURE 2-2 Réseaux bayésiens des processus markovien de premier order, d'ordre supérieur et semi-markovien d'une séquence discrète à deux états.

famille des HOMC. Dans ce type de réseau, la durée des états est considérée, de sorte que les états sont conditionnés par leur précédent à travers une transition markovienne, et la durée est conditionnée uniquement à l'état actuel. Ainsi, le processus est basé sur un compteur décremental, qui force une transition d'état à chaque fois qu'il atteint la valeur "1". Une fois la transition effectuée, une nouvelle durée aléatoire est générée pour le nouvel état, le reste du temps, la durée et l'état même évolue de façon déterministe.

Dans les études portant sur la modélisation du comportement, Richardson et al. [69] ont utilisé l'EET réalisée au Royaume-Uni en 2000 [70] et les chaînes de Markov de premier ordre pour générer des profils quotidiens d'occupation active¹ à une résolution de dix minutes. Avec cette méthode, les auteurs peuvent déterminer le nombre d'occupants actifs de la maison.

1. L'occupation active fait référence au moment où un utilisateur est dans la maison et effectue une activité autre que dormir.

Par ailleurs, le nombre de résidents du foyer et les jours de la semaine ont été utilisés comme paramètres pour caractériser la variabilité de l'occupation. Par la suite, Richardson a approfondi son modèle pour générer également des profils d'activités, afin de les intégrer dans une approche pour estimer la demande d'électricité résidentielle [71]. Widen et al. [72] ont proposé une technique de Monte-Carlo par chaîne de Markov de premier ordre pour générer des données sur la demande d'éclairage domestique à partir des modèles d'occupation et de la disponibilité de la lumière naturelle. Les auteurs ont présenté un modèle dans lequel toutes les personnes sont caractérisées par des paramètres tels que le type de domicile (maisons individuelles et appartements) et le type de jour (en semaine et en week-end). Par la suite, Widen et Wackelgard [73] ont retravaillé leur modèle pour générer des profils d'activités des occupants à partir desquels la demande d'électricité de la maison est déterminée.

De même, Muratori et al. [74] présentent une chaîne de Markov hétérogène pour générer des modèles d'activité. Les auteurs utilisent une méthode similaire à celle présentée par Widen et Wackelgard [73]. De plus, ils ont divisé les individus entre travailleurs, non-travailleurs, hommes, femmes et enfants pour évaluer les différences dans le comportement. Par ailleurs, Collin et al. [75] ont développé un modèle Markov du premier ordre à treize états, différenciant les personnes à partir d'attributs tels que la taille du ménage, les travailleurs et les non-travailleurs. Ensuite, les auteurs ont appliqué un modèle pour générer des profils de puissance en fonction des activités des utilisateurs. Ils ont employé une approche probabiliste pour identifier les appareils électriques partagés par les utilisateurs effectuant la même activité. Baptista et al. [76] ont proposé d'utiliser l'approche de la chaîne de Markov interactive présentée par Conlisk [77] pour intégrer l'interaction entre les individus d'un ménage. Pour cela, les matrices de probabilité de transition ont été conditionnées par l'activité actuelle d'un leader. De cette façon, les occupants étaient en mesure de coordonner explicitement leurs activités. Cependant, le travail ne présente que quelques résultats préliminaires, et l'approche n'a pas considéré une analyse exhaustive de la variabilité du comportement en fonction des caractéristiques des individus. À partir des articles présentés ci-dessus, nous pouvons souligner que la variabilité du comportement, due aux caractéristiques des individus et aux facteurs externes, n'a pas été suffisamment explorée. En outre, ces modèles présentent l'inconvénient du manque de mémoire inhérent aux chaînes de Markov du premier ordre. De ce fait, [78]

et [79] ont révélé que les modèles de Markov du premier ordre ont des limitations dans la représentation de la dynamique temporelle de la présence/absence des individus. Ce qui réduit sa capacité à prédire de manière réaliste les profils d'occupation et d'activités des occupants.

De plus, certains auteurs ont proposé l'utilisation de modèles semi-markovien et des modèles de Markov d'ordre supérieur. Tanimoto [80], Wilke et al. [37], [78], et Vorger [55] ont modélisé le comportement des occupants à partir d'un processus semi-markovien. Pour cela, Tanimoto utilise une distribution logarithmique pour générer la durée des activités, et Wilke et al. et Vorger utilisent une distribution de Weibull pour estimer la durée de la présence et des activités. Cependant, ces distributions ont des limites pour saisir le caractère multimodal qui se produit souvent dans les données de durée. D'autre part, Flett et al. [79], [81] ont présenté un modèle permettant de générer des profils d'occupation. Les auteurs ont développé une méthode de Markov d'ordre supérieur où de multiples matrices de transition de probabilité ont été générées en fonction de la durée de l'état existant. Aerts et al. [60], [82] ont présenté une méthodologie basée sur le regroupement hiérarchique (« hierarchical clustering », en anglais) pour identifier des patrons d'occupation typiques à partir d'un sondage EET belges. Ces patrons ont été utilisés pour calibrer un modèle probabiliste à trois états (absent, endormi, et à la maison mais éveillé) afin de générer des séquences individuelles d'occupation annuelle et quotidienne. Cependant, la littérature scientifique révèle que la plupart de ces modèles ne prennent en compte que certains groupes spécifiques de personnes en fonction de l'âge, du genre et du revenu [74]. Toutefois, en raison de la grande variabilité du comportement humain et de l'évolution du comportement des occupants (p. ex. tendance verte, vieillissement de la population, travail à domicile), il a été révélé la nécessité de développer des modèles utilisant un ensemble plus représentatif de caractéristiques sociodémographiques [48]. Par ailleurs, bien que certains modèles de Markov d'ordre supérieur aient été proposés, il existe encore des limites dans la modélisation de la durée des états, notamment dans la génération de distributions multimodales de la durée et la représentation des variations du comportement propres aux caractéristiques des occupants.

Les modèles de comportement des occupants existants présentent d'autres contraintes qui doivent être prises en compte pour rendre plus réaliste la prédiction de l'occupation et d'activités des individus [48]. Par exemple, la plupart des approches ont traité les personnes

comme des agents indépendants, ce qui ne reflète pas le comportement humain réel dans les logements à plusieurs occupants. En réalité, les occupants interagissent et peuvent donc effectuer des activités ensemble. La précision dans la modélisation de ces interactions peut avoir un impact significatif sur la simulation des courbes de charge. Les effets saisonniers et les longues périodes d'absence causées par les vacances ou les maladies sont d'autres aspects de la modélisation de l'occupation qui n'ont pas été suffisamment étudiés. Cependant, ces limites sont en partie le résultat des informations restreintes disponibles dans les sondages. Pour chaque ménage, en général, les enquêtes ne fournissent des informations que pour un ou deux journées typiques du répondant, fournissant peu d'informations sur les autres habitants du ménage et pour des périodes de temps plus prolongées. Quoiqu'il en soit, les enquêtes restent une source d'information très importante, dans la quête d'une meilleure compréhension du comportement humain.

2.1.2 Modélisation du comportement à partir de mesures sur place

Les variables environnementales à l'intérieur d'un espace fermé et la puissance sont fortement corrélés au comportement humain [27]. Pour cette raison, plusieurs types de capteurs sont utilisés pour surveiller ces variables, permettant ainsi une mesure indirecte de la présence des individus, le nombre de personnes, les activités, et l'interaction des personnes avec les éléments du bâtiment tels que les fenêtres, les rideaux et l'éclairage. Cette démarche a fait l'objet d'un grand intérêt ces dernières années en raison des innovations technologiques et la popularité croissante de l'internet des objets, permettant le déploiement de solutions plus économiques et faciles à installer [56]. Étant donné que les travaux de cette thèse portent sur la modélisation de l'occupation, l'état de l'art présenté ci-dessous est consacré à ce sujet. Toutefois, la plupart des approches et des technologies discutées peuvent être appliquées à des applications telles que la modélisation des activités [83] et de fenêtres [84].

De nombreuses technologies ont été utilisées pour détecter les informations relatives à l'occupation. Par exemple, les caméras optiques et thermiques ont été reconnues comme des dispositifs permettant une grande précision pour la détection de l'occupation [56]. Cependant, des préoccupations importantes liées à la confidentialité, aux coûts et aux ressources informatiques élevées de ces technologies ont été soulignées dans la littérature [85], [86]. Par

conséquent, les dispositifs de détection de l'occupation plus respectueux de la vie privée et plus abordables ont fait l'objet d'une attention significative [87]. Un aperçu des capteurs les plus utilisés dans les applications de détection d'occupation est présenté à la FIGURE 2-3. Par ailleurs, une description sommaire de ces capteurs est présentée ci-dessous :

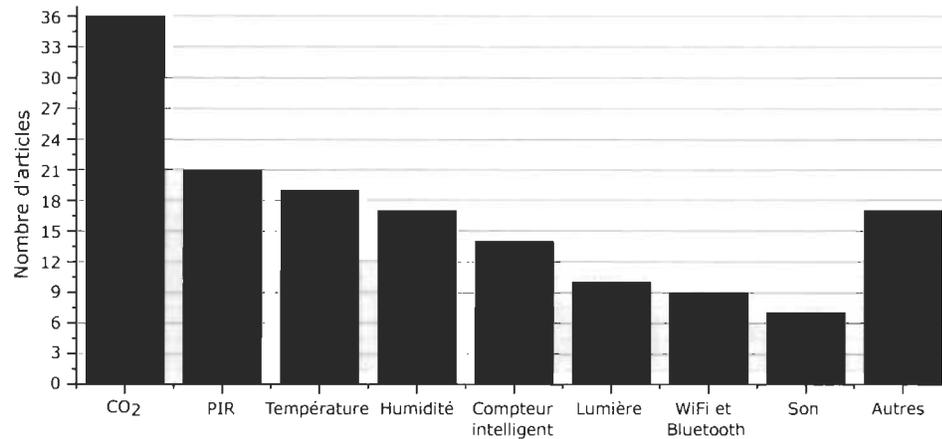


FIGURE 2-3 Nombre d'articles par type de capteur [56].

- **Capteurs environnementaux :** La surveillance des variables environnementales telles que la température, la luminosité, l'humidité, le dioxyde de carbone (CO₂) et la pression a été largement utilisée pour estimer l'occupation des maisons et des bâtiments. La surveillance de la concentration de CO₂ est l'une des méthodes les plus utilisées. Le dioxyde de carbone est exhalé par les humains à la suite d'une activité métabolique. Donc, la variation de sa concentration a prouvé l'efficacité à détecter la présence, et estimer le nombre de personnes ainsi que le niveau d'activité des utilisateurs dans les espaces fermés [88]. Cependant, la présence d'animaux domestiques, de plantes, d'échangeurs d'air et l'ouverture des portes/fenêtres peuvent biaiser la mesure de ces variables.
- **Capteur PIR :** Ce type de capteurs est utilisé pour détecter l'énergie thermique sous forme de radiation émise ou réfléchi par des personnes, des animaux ou d'autres objets. La mesure de cette radiation peut aider à détecter des changements dans l'environnement qui peuvent être associés à la présence et à l'activité humaine. Ainsi, les capteurs PIR constituent l'un des systèmes les plus courants utilisés pour la détection d'occupation [89] et pour le contrôle énergétiquement efficace des systèmes d'éclairage [90]. Par

contre, les capteurs PIR ne captent pas les occupants lorsqu'ils ne bougent pas pendant une longue période, par exemple lorsqu'une personne est assise devant un ordinateur.

- **WiFi et Bluetooth** : L'utilisation de nouveaux types de technologies, comme le WiFi et le Bluetooth, est apparue comme une solution possible à la nécessité d'une meilleure précision et fiabilité pour estimer la présence et le nombre de personnes. Par exemple pour améliorer la procédure d'évacuation en cas d'urgence ou pour accroître la satisfaction des clients dans les restaurants ou centres commerciaux [91]. Parmi les avantages de ces technologies, on peut souligner le fait que son infrastructure est largement disponible dans les espaces intérieurs, ce qui réduit les coûts de développement. En outre, une grande variété d'appareils tels que les smart phones ou les ordinateurs portables disposent d'une connectivité WiFi et Bluetooth intégrés. Toutefois, l'utilisation de ces technologies est sujette à des erreurs humaines, comme l'oubli du téléphone à la maison ou au travail, ou l'oubli de déclencher le Bluetooth.
- **Compteurs intelligents** : Au cours des dernières années, les informations sur la consommation d'énergie sont devenues une source d'information utilisée pour des applications telles que la détection d'occupation. Ceci est dû à l'intégration croissante des compteurs intelligents et de la relation étroite entre la consommation d'électricité et la présence et les activités des occupants à leur domicile. Par ailleurs, l'un des avantages de ces technologies est que la mesure peut être effectuée de manière non intrusive, réduisant ainsi la complexité des infrastructures nécessaires, tout en préservant la vie privée des occupants. Cependant, le coût des technologies pour le sous-mesurage de la puissance est un facteur à prendre en compte, ainsi que les techniques de surveillance non intrusive de charges (NILM), qui sont encore en cours de développement.
- **Combinaison de capteurs** : Comme présenté ci-dessus, la littérature a montré que chaque type de capteur présente des avantages et des limites uniques pour les applications de détection d'occupation [92], [93]. Par conséquent, la combinaison de capteurs a été présentée comme un moyen efficace de mettre à profit les avantages de différents capteurs pour améliorer les performances du système. Cependant, l'interopérabilité des capteurs, souvent issus de fabricants différents ou ne supportant pas les mêmes protocoles de communication, constitue un défi qui mérite encore des améliorations.

Cependant, bien qu'il existe un grand nombre de technologies pouvant être utilisées pour des applications telles que la détection de la présence et le comportement des occupants, le déploiement d'infrastructures de mesurage et de contrôle est un domaine qui n'a pas évolué de la même manière dans les secteurs commercial, institutionnel et résidentiel [94]. Dans les deux premiers, il existe généralement une infrastructure de mesurage et de contrôle, qui est déployée conjointement avec les systèmes de CVC, les systèmes de surveillance et les contrôles d'accès, afin de maintenir une bonne qualité de l'air intérieur et d'assurer un niveau de sécurité. Toutefois, bien que l'objectif principal de cette infrastructure ne soit pas orienté vers la détection et la modélisation du comportement, les informations fournies par ces systèmes peuvent être exploitées à de telles fins. En revanche, le secteur résidentiel manque d'infrastructures de mesure et de contrôle, ce qui limite l'application des méthodes de détection et de modélisation du comportement. De plus, le respect de la vie privée est un aspect très important à considérer, car il limite l'acceptation des utilisateurs à certains types de technologies. Par conséquent, la popularisation de la domotique, ainsi que des technologies telles que le WiFi et le NILM sont présentées comme des alternatives potentielles ayant une plus grande acceptation par les utilisateurs du secteur résidentiel.

D'autre part, en ce qui concerne les méthodes utilisées pour les systèmes de détection d'occupation, Rueda et al. [56] les ont classées en trois catégories, les méthodes analytiques, les méthodes basées sur les données et les méthodes basées sur les connaissances :

- Méthodes analytiques : Ces méthodes sont basées sur des modèles physiques qui nécessitent une connaissance détaillée des processus et des variables impliqués dans le système. En raison de la connaissance des équations dynamiques du système, ces types de modèles fournissent les résultats les plus proches du système réel. Cependant, ils sont l'un des types de techniques les plus complexes à mettre en œuvre en raison des connaissances approfondies qu'il faut avoir de toutes les variables impliquées dans le système. Pour la détection d'occupation, ces méthodes ont été principalement utilisées pour la modélisation des systèmes CVC afin d'identifier le nombre de personnes dans un espace intérieur sur la base des variations de la concentration de CO₂.
- Méthodes basées sur les données : Elles sont principalement basées sur des algorithmes d'apprentissage automatique et des approches statistiques et probabilistes, qui, à partir

de l'examen des données observées, peuvent déduire des connaissances pour faire des prédictions ou des classifications. Parmi les avantages de ce type de méthode, on peut citer la capacité d'identifier des relations non linéaires entre des variables indépendantes et dépendantes, un large champ d'applications, et la capacité d'apprendre, de prédire et d'améliorer les connaissances dans le temps. Cependant, l'un des principaux défis des méthodes orientées données est qu'elles nécessitent une quantité massive de données de qualité et non biaisées pour le processus d'apprentissage, ce qui a un impact significatif sur la précision des résultats.

- Méthodes basées sur la connaissance : Ces méthodes sont capables de résoudre des problèmes complexes en utilisant les connaissances fournies par les experts humains du domaine. Les systèmes experts offrent de bonnes performances, une maintenance facile et ne nécessitent pas de processus d'apprentissage ou de calibration du modèle. Cependant, le processus d'extraction, de structuration et d'organisation des connaissances du domaine provenant d'experts pour une application (acquisition de connaissances) est une tâche complexe et longue.

La FIGURE 2-4 montre qu'entre ces trois grandes familles d'algorithmes, les méthodes basées sur les données telles que les réseaux neuronaux artificiels (ANN) [86], les machines à vecteurs de support (SVM) [95], K-nearest neighbors (KNN) [92] et les modèles de Markov cachés (HMM) [96] sont parmi les plus utilisés pour la détection de l'occupation. Cependant,

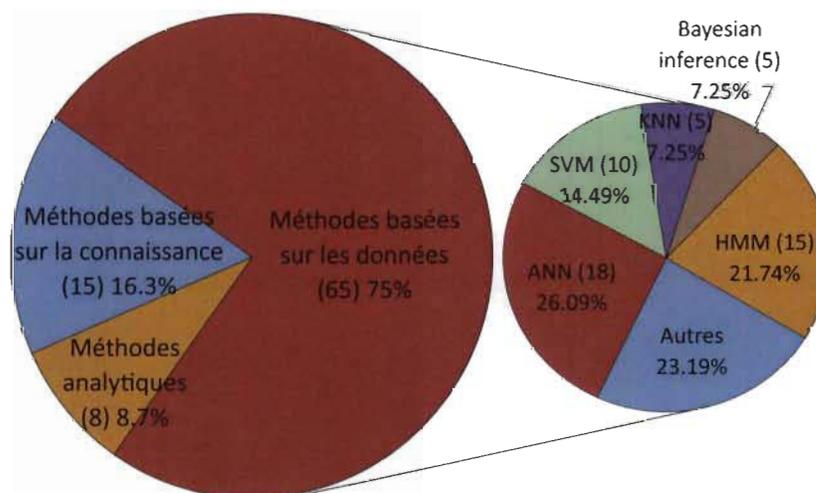


FIGURE 2-4 Pourcentage d'articles par type d'algorithme [56].

la littérature scientifique révèle que le développement de la plupart de ces modèles a été abordé comme un problème d'apprentissage supervisé. Pour cette raison, une connaissance complète de la variable à modéliser, dans ce cas, la présence, est requise afin d'effectuer la phase d'apprentissage [56]. De ce fait, cette approche est principalement adaptée aux applications hors ligne. Pour cela, la valeur réelle de l'occupation est généralement collectée manuellement par un utilisateur à l'aide des technologies très intrusives comme les caméras. Cependant, ces modèles sont difficilement applicables aux situations pratiques, où le contrôle en ligne est nécessaire alors que l'occupation ne peut être observée explicitement et que le modèle doit s'adapter au fil du temps [56]. C'est pourquoi des initiatives ont été prises pour développer des méthodes semi-supervisées et non supervisées telles que HMM [96], [97], le test de Page–Hinkley [98], ALOS [99], PresenceSense [100] et SHARK [101] comme alternatives pour surmonter ces limitations. Toutefois, la plupart de ces modèles ont négligé la dynamique temporelle du comportement, alors que la littérature montre que, par exemple, la saison, le jour de la semaine, l'heure et la durée sont des facteurs importants à prendre en compte pour obtenir une modélisation réaliste de l'occupation et du comportement [60], [102].

2.2 Gestion locale de la demande

La gestion locale de la demande (Demand Side Management (DSM), en anglais) fait référence aux technologies, actions et programmes visent à améliorer la gestion globale des systèmes électriques et l'efficacité énergétique en respectant les besoins et le confort des clients [2], [103]. Ces stratégies ont le potentiel de fournir des avantages à la fois pour les clients finaux que pour les services de réseau. Par exemple, les utilisateurs peuvent avoir des factures de services publics moins élevées, des tarifs plus bas ou des paiements négociés, tandis que l'utilité peut réduire les coûts et les pertes de transmission et de distribution (T&D) [5]. Dans la DSM, différents objectifs sont définis selon le niveau de gestion (transmission, distribution, demande) des opérateurs du réseau. Parmi ces objectifs, les plus souvent mentionnés dans la littérature sont :

- Améliorer la fiabilité et la résilience du système.
- Améliorer l'efficacité énergétique des appareils.

- Réduire les émissions de gaz à effet de serre (GES).
- Encourager l'augmentation de l'intégration des ressources énergétiques distribuées.
- Augmentation de la satisfaction et la flexibilité des clients.
- Minimiser le coût d'approvisionnement.
- Réduire des coûts d'exploitation et de maintenance.
- Minimiser la consommation électrique en puissance pendant les périodes de pointe.
- Favoriser l'implantation des approches transactionnelles pour l'échange de puissance entre le fournisseur et le consommateur.
- Encourager le client à faire une meilleure gestion de la consommation et de la production locale d'énergie.

L'intérêt principal des opérateurs du réseau consiste à contrôler la quantité d'énergie et de puissance transitant lors de périodes spécifiques. Cependant, les besoins du réseau électrique varient considérablement selon le lieu, l'heure de la journée, le jour de la semaine et la saison. Par conséquent, un bâtiment peut avoir besoin de gérer sa charge électrique de différentes manières pendant ces périodes. Par exemple, en réduisant la charge tout au long de l'année grâce aux stratégies d'efficacité énergétique, en déplaçant la charge à différents moments de la journée ou en augmentant la capacité de stockage d'énergie pour une utilisation ultérieure [5]. Parmi les stratégies de gestion de la demande qui peuvent être mises en œuvre dans les bâtiments pour gérer la charge, on peut mentionner l'efficacité, le délestage et le déplacement de charge. Celles-ci sont illustrées à la FIGURE 2-5.

- Efficacité : Il s'agit de la réduction continue de la consommation d'énergie des charges tout en assurant le même niveau ou un niveau amélioré de fonctionnalité. Cette stratégie permet de réduire la demande non seulement aux heures de pointe, mais également aux autres heures de la journée.
- Délestage : Ce type de stratégie permet de réduire la consommation d'électricité pendant une courte période de temps. Le délestage est généralement mis en place pendant les périodes de pointe et en cas d'urgence en utilisant par exemple un contrôle direct sur certains appareils, permettant de réduire les pics de charge.

- **Déplacement de charge** : Il fait référence à la capacité de modifier le moment où l'électricité est utilisée. Cela implique un déplacement de la charge des périodes de pointe aux heures creuses. Les applications courantes incluent l'utilisation du préchauffage de l'eau et des locaux avant la période de pointe, de dispositifs de stockage d'énergie et les déplacements de certaines charges des clients.

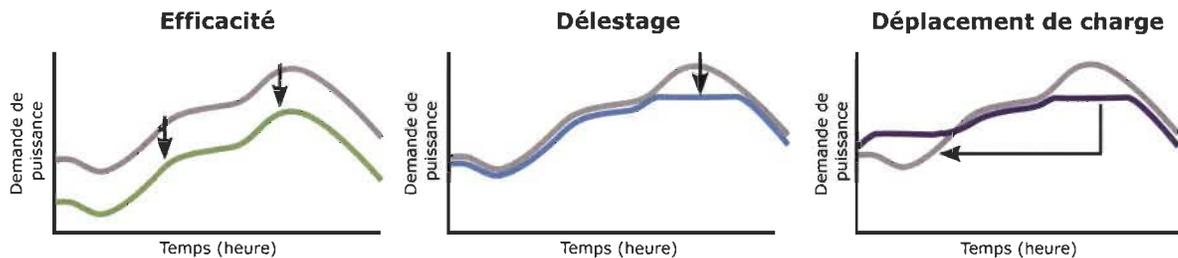


FIGURE 2-5 Stratégies de gestion de la demande. Adapté de [5].

Pour atteindre la réalisation de ces stratégies, les programmes de réponse à la demande (Demand Response (DR), en anglais) constituent une alternative qui vise à motiver les consommateurs finaux à réagir aux changements de prix à travers des programmes incitatifs au fil du temps en modifiant leurs habitudes normales d'utilisation de l'électricité [104]. En effet, ces stratégies cherchent à influencer de façon directe ou indirecte le profil de la demande et à couvrir les forts appels de puissance de façon efficace avec l'énergie disponible. Parmi les programmes de réponse à la demande fondés sur des changements de prix, nous pouvons mentionner la tarification en fonction de l'heure de la consommation (Time-of-use (TOU) pricing), le prix dynamique en temps réel (Real-Time Pricing (RTP)) et la tarification pour la période de pointe (Critical Peak Pricing (CPP)) [105]. D'autre part, concernant les programmes de réponse à la demande fondés sur des mesures incitatives, l'un des plus connus dans la littérature est le contrôle direct de la charge. Ce programme consiste à contrôler à distance les appareils ou les équipements avec un pré-engagement du client par l'entremise d'une compensation offerte par l'utilité : p. ex. un meilleur prix de l'énergie [106].

Néanmoins, afin de mettre en œuvre des stratégies pour déplacer la charge à des périodes plus convenables, de diminuer la pointe de consommation ou de moduler les charges qui présentent des pics transitoires importants, la modélisation des habitudes de consommation des clients et la flexibilité des charges électriques sont des facteurs importants à considérer. Les charges électriques peuvent être divisées en deux catégories : les charges non flexibles

et les charges flexibles. Les charges telles que l'éclairage et les appareils de cuisine sont des charges non flexibles dont les heures d'utilisation ne peuvent pas être déplacées. Les charges flexibles sont celles dont la durée et l'heure d'utilisation peuvent être contrôlées. Ces charges peuvent être divisées en deux catégories en fonction de leurs caractéristiques : charges reportables (Deferrable loads, en anglais) et charges à température variable (Temperature-shiftable loads, en anglais). Les charges comme la machine à laver et le lave-vaisselle sont des charges reportables, et les charges comme le chauffage et la climatisation sont charges à température variable [107].

L'avancement des technologies et la prolifération croissante de l'internet des objets ont poussé le développement des appareils intelligents communicants qui ont comme objectif de faciliter les tâches des occupants dans la résidence. Toutefois, de tels avancements permettent aussi l'intégration de ces appareils dans la mise en œuvre des stratégies de gestion de la demande, devenant ainsi la base du développement des systèmes de gestion de la demande résidentielle aussi connus comme HEMS (Home Energy Management System, en anglais) [108]. Les HEMS sont une approche qui permet gérer la consommation d'énergie en communiquant avec les appareils électroménagers et les fournisseurs d'énergie, en surveillant la consommation d'énergie et en recevant des informations (p. ex. les tarifs) afin de réduire la consommation en planifiant l'utilisation des appareils ménagers [109].

Par ailleurs, les systèmes de production locale (SPL), les accumulateurs thermiques locaux (ATL) et les accumulateurs thermiques centraux (ATC) peuvent faire partie des HEMSs [110]. Les SPL et le stockage électrique permettent l'optimisation des ressources et sources disponibles. Ils permettent le déplacement de charge vers une autre source d'énergie, qui est vu comme de délestage pour l'utilité électrique, mais pas pour le client. De plus, les systèmes de stockage thermique (ATL ou ATC) sont utilisés comme un moyen de gérer la demande de façon directe, et de déplacer les charges en dehors des périodes de pointe [10].

L'architecture typique d'un HEMS est présentée à la FIGURE 2-6. Le système se compose principalement d'un compteur intelligent en communication permanente avec l'utilité et d'un centre de contrôle équipé de logiciels et de protocoles de communication tels que WiFi, ZigBee, Bluetooth et MQTT. Le contrôleur sert à la coordination dans la réception/envoi des données de l'utilitaire à la maison intelligente et contrôle les signaux pour gérer

les appareils électriques. Des informations sur l'environnement intérieur et les conditions météorologiques sont également prises en compte dans HEMS pour déterminer les actions de contrôle appropriées et préserver le confort des individus [109]. De plus, le système peut compter sur la production et le stockage d'énergie locale (p. ex. panneaux photovoltaïques et batteries), ainsi que la possibilité d'intégrer des véhicules électriques.

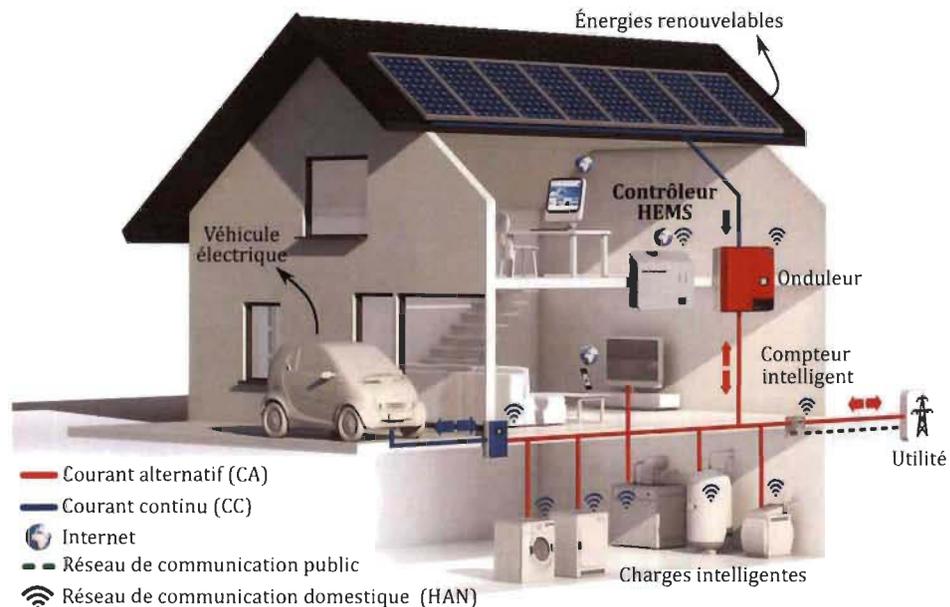


FIGURE 2-6 Architecture typique d'un HEMS [111].

2.2.1 Stratégies de commande

Les systèmes de chauffage décentralisés par plinthes électriques, qui sont majoritairement utilisés au Québec, ont un effet important sur la pointe de consommation. En raison de cette concomitance de la charge de chauffage avec la pointe, de la possibilité à moduler la puissance de ces charges résistives, de l'accessibilité aux thermostats communicant, ainsi que du potentiel d'économie d'énergie et de la réduction dans la facture des clients, cette thèse se concentre sur la commande de ces systèmes de chauffage.

La gestion de la demande et du microclimat intérieur des bâtiments est un problème d'optimisation qui vise généralement à atteindre un niveau élevé de confort, tant thermique que sur la qualité de l'air, tout en maximisant l'efficacité énergétique et en réduisant leurs coûts de fonctionnement. Cependant, en raison du nombre de variables utilisées et de la nature

stochastique de certaines d'entre elles, il s'agit d'un problème d'optimisation qui peut être complexe à résoudre. Parmi les méthodes proposées dans la littérature pour aborder ce défi, les méthodes de contrôle classique et les techniques de commande intelligente sont les plus utilisées.

2.2.1.1 Méthodes de commande classique

Les méthodes classiques de commande ont été largement adoptées dans les systèmes énergétiques des bâtiments en raison de sa simplicité de conception et de sa faible complexité de calcul lors de la détermination des signaux de commande [112]. À titre d'exemple, les commandes marche/arrêt ou les contrôleurs bang-bang sont très couramment utilisés dans les systèmes de bâtiments sans contrôle numérique. Cependant ces contrôleurs peuvent créer des problèmes d'inefficacité énergétique à cause des instabilités et des dépassements de la référence [113]. Par conséquent, l'utilisation des contrôleurs proportionnels intégraux (PI) et proportionnels intégraux et dérivés (PID) a été proposée. Ces techniques permettent un meilleur contrôle du système, et ils sont utilisés habituellement dans les bâtiments où les systèmes CVC sont équipés d'une commande numérique et de variateurs de fréquence (p. ex., les commandes à modulation de largeur d'impulsion). Toutefois, dans le cas des contrôleurs P, PI et PID, le choix des paramètres de contrôle est une tâche essentielle, car une mauvaise configuration pourrait rendre l'ensemble du système instable [112].

2.2.1.2 Méthodes basées sur des règles ou des conditions

Les contrôleurs à base de règles (RBC) sont fondés sur la logique inférentielle du type « if-then-else » et ils sont fréquemment utilisés dans plusieurs domaines, par exemple, pour la régulation de la température intérieure des bâtiments [114]. Parmi les méthodes les plus populaires, nous pouvons trouver des contrôleurs basés sur la logique floue, où un ensemble de règles est défini avec une hiérarchie de décisions [115]. En fait, ce type de méthode s'est révélé être une approche efficace, facile à comprendre et économique pour les applications d'automatisation industrielle [116]. De plus, c'est une méthode qui permet de résoudre des problèmes complexes de manière approximative, et qui peut être appliquée dans les cas où le système est difficile à modéliser [117]. Le principal inconvénient de la méthode RBC

est le caractère unique et empirique de chaque implantation et la forte dépendance sur la connaissance du système [118].

2.2.1.3 Méthodes basées sur le contrôle prédictif et optimal

La commande prédictive et optimale est une approche largement utilisée et acceptée par la communauté scientifique et l'industrie dans diverses disciplines [118]. Parmi les avantages des méthodes prédictives et optimales figure leur capacité à prendre en compte plusieurs objectifs d'optimisation et plusieurs contraintes sur les variables de contrôle. De plus, c'est une méthode qui est robuste aux perturbations et qui a une bonne réponse transitoire. La Commande Prédictive basée sur un Modèle (MPC) est une des stratégies les plus utilisées. Cette méthode utilise un modèle de représentation du comportement du système, permettant de faire des prédictions des variables de sortie par rapport aux entrées qui sont définies par un processus d'optimisation ou par estimation, comme les prédictions de température extérieure ou la prévision de l'occupation à l'intérieur du logement [10]. Le désavantage de cette dernière réside dans sa complexité vis-à-vis des modèles du système à contrôler. Cependant, cette méthodologie de commande est très robuste grâce au processus d'adaptation et très performante grâce aux optimisations récursives faites sur les variables du système [119].

La commande optimale est une autre approche très utilisée. Elle est basée sur l'optimisation d'un objectif en changeant les valeurs des variables de décision, tout en gardant les contraintes établies [120]. Parmi les méthodes d'optimisation nous pouvons trouver la programmation linéaire à nombres mixtes [121], [122], l'optimisation convexe [123], [124], la programmation dynamique [125], [126], la programmation stochastique [127], l'utilisation des processus de décision de Markov [128], l'application de théorie des jeux [129] et l'implémentation des méthodes metaheuristiques, tels que l'optimisation de l'essaim de particules (OEP) et les algorithmes évolutifs (AE) [113], [130]. Dans la grande variété des méthodes d'optimisation pour la commande optimale, le choix de la stratégie dépendra des conditions du problème, de la description du problème (p. ex., en nombres entiers ou continus, ou avec des équations linéaires ou non-linéaires) et finalement, du niveau d'incertitude des variables et de la condition du système.

2.2.2 Contrôle des bâtiments centré sur l'occupant

L'intégration d'informations sur le comportement humain dans les stratégies de contrôle des bâtiments commerciaux et résidentiels a été un sujet de grand intérêt ces dernières années. La FIGURE 2-7, présente le concept général des stratégies de contrôle des bâtiments centrées sur l'occupant (Occupant-centric building control, en anglais). Ainsi, le système de contrôle collecte des informations sur l'environnement intérieur (p. ex. la température, le CO₂ et l'humidité) et surveille l'interaction des personnes avec le bâtiment (p. ex. l'utilisation des fenêtres ou de l'éclairage) afin de déterminer en temps réel le confort des occupants. Avec ces informations, le contrôleur estime les actions de contrôle optimales, visant à préserver un environnement sain, tout en maintenant ou améliorant la satisfaction des occupants.

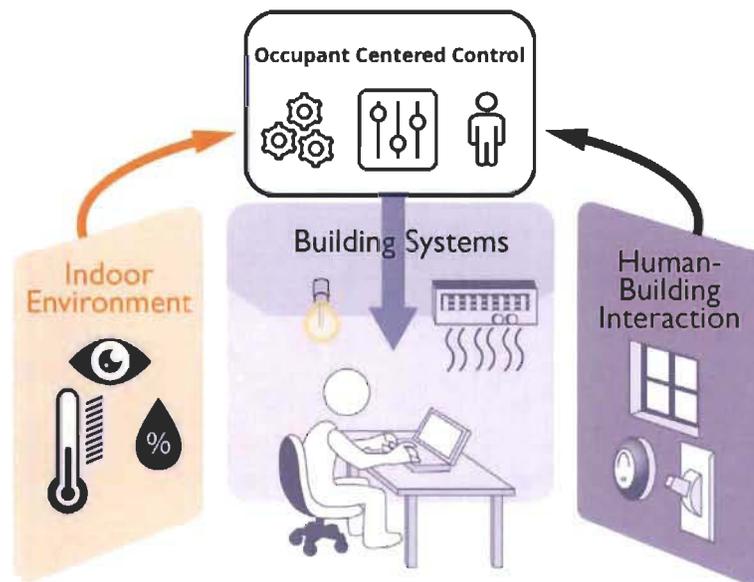


FIGURE 2-7 Schéma conceptuel du contrôle des bâtiments centré sur l'occupant [131].

Plusieurs travaux ont démontré qu'une quantité importante d'énergie peut être économisée en utilisant des stratégies de contrôle basées sur l'occupation [48]. Par exemple, Nguyen et al. [49] ont effectué une revue des efforts de recherche sur les bâtiments intelligents, axés sur les économies d'énergie et la reconnaissance de l'activité des utilisateurs. Les auteurs ont conclu que les stratégies de contrôle basées sur l'occupation peuvent réduire jusqu'à 40% la consommation d'énergie du système de chauffage, de ventilation et de climatisation (CVC). Milenkovic et al. [83] ont montré qu'en utilisant un contrôle des systèmes d'éclairage basé sur

les activités, il était possible de réaliser des économies d'énergie globales de 78,5% par an par rapport au contrôle manuel. De plus, en utilisant une stratégie de contrôle basée sur le comptage des personnes, il a été possible de réduire la consommation annuelle d'énergie de CVC de 39,9%. Par ailleurs, une stratégie de contrôle par zone pour les charges thermostatiques a été présentée dans [132]. Les auteurs ont démontré que leur stratégie permettait d'économiser plus de 15% de la consommation d'énergie tout en augmentant le confort thermique de plus de 25%.

Tel que présenté précédemment, différentes stratégies de contrôle ont été proposées dans la littérature, toutefois, parmi celles-ci, il a été démontré que la détection instantanée de l'occupation est plus adéquate pour des applications telles que le contrôle de l'éclairage. Cependant, d'autres systèmes comme les chauffages d'espace nécessitent un temps plus long pour avoir un effet sur la température d'une pièce. Pour cette raison, il est indispensable d'anticiper le comportement des individus afin de réaliser des économies d'énergie et maximiser le confort des occupants [99]. De ce fait, une analyse générale des approches les plus couramment utilisées pour la prédiction des périodes de présence/absence est présentée ci-après.

2.2.2.1 *Prevision de l'occupation*

Kleiminger et al. [133] classe les approches de prédiction de l'occupation en deux catégories : les algorithmes basés sur l'horaire et les approches contextuelles . D'une part, les *algorithmes basés sur les horaires* prédisent l'occupation à l'aide de données historiques sur l'état d'occupation du ménage. Par exemple, Scott et al. [134] ont utilisé les données des capteurs PIR et RFID pour estimer les modèles d'occupation de la maison. De plus, les auteurs ont utilisé la distance de Hamming et les algorithmes KNN pour prédire l'occupation de la maison afin de contrôler le chauffage. Nacer et al. [99] ont présenté une méthode basée sur les k-means et l'expectation-maximisation pour modéliser l'occupation des utilisateurs et prédire les événements d'arrivée et de départ. Mozer et al. [135] ont utilisé des capteurs de mouvement pour détecter l'occupation d'une maison et ont employé un réseau neuronal pour prédire l'occupation à l'aide d'observations récentes. De plus, les modèles de Markov ont également été explorés comme alternative pour prédire l'occupation à des horizons allant de 15 minutes à 24 heures [136], [137]. Une analyse comparative des performances des approches

de l'état de l'art a été présentée par Kleiminger et al. [133] et Li et Dong [136], soulignant que des prédictions d'une précision allant jusqu'à 90% peuvent être réalisées. D'autre part, *les approches contextuelles* estiment les états d'occupation futurs d'une maison en fonction du contexte actuel de ses occupants (l'emplacement ou l'activité). Par exemple, Krumm et Brush [138] ont combiné un algorithme de temps de conduite dérivé de données GPS avec un programme probabiliste (PS) pour estimer quand une personne sera à la maison ou absente. Gupta et al. [139] ont utilisé les données de déplacement GPS pour évaluer l'heure d'arrivée de l'utilisateur à son domicile ; par conséquent, le système contrôle de façon dynamique le thermostat pour préchauffer la pièce. Cependant, bien que les approches contextuelles soient présentées comme l'une des techniques les plus performantes pour la prédiction d'occupation, ces méthodes dépendent du fonctionnement permanent de dispositifs spécifiques par les individus (p. ex. le smartphone et le GPS). Par conséquent, ces méthodes sont plus intrusives dans le mode de vie des individus, et sont plus sensibles aux erreurs humaines telles que l'oubli du téléphone à la maison.

2.3 Synthèse du chapitre

À partir de la revue de littérature faite, il a été possible d'identifier les sources d'information disponibles, les technologies et les différentes méthodologies adoptées dans la littérature pour faire face au problème de la modélisation du comportement, en particulier la présence des occupants. Les points forts des méthodes existantes ont été présentés, ainsi que leurs limites, restrictions et difficultés. De plus, des concepts généraux portant sur les stratégies de gestion de la demande ont également été discutés.

Dans la littérature, les enquêtes d'emploi du temps et les mesures sur place figurent parmi les techniques les plus utilisées pour la collecte de données sur la présence et le comportement humain. En ce qui concerne les modèles basés sur des données des enquêtes, ils sont largement utilisés pour des applications de simulation et pour la génération de profils de consommation de groupes de bâtiments. Notamment, les FOMC et HOMC sont les approches les plus utilisées, cependant, la modélisation de la dépendance temporelle du comportement et la représentation de la variabilité des patrons de présence et d'activités parmi les individus sont mises en évidence en tant que limitation de ces modèles. Cela réduit sa capacité à prédire de

manière réaliste le comportement humain, ce qui révèle que des efforts supplémentaires sont encore nécessaires dans ce domaine. Par ailleurs, bien que certaines contraintes inhérentes aux informations fournies par les enquêtes soient identifiées, les EET sont considérées comme l'une des sources d'information les plus importantes pour arriver à une meilleure compréhension et modélisation du comportement humain.

Ce chapitre présente aussi une revue des méthodologies pour la détection et prévision de l'occupation et la gestion locale de la demande. Cette analyse soulève l'importance d'inclure des informations sur l'occupation et le comportement des individus dans les systèmes de gestion résidentielle, notamment dans le but de contribuer à l'amélioration de la performance énergétique des logements, tout en respectant le confort des usagers. Cependant, nous avons mis en évidence des contraintes éthiques et pratiques liées à la collecte de données d'occupation réelles, ainsi que l'importance d'anticiper l'occupation pour effectuer le contrôle du système de chauffage électrique. De ce fait, les stratégies de contrôle prédictif et les modèles de détection et d'anticipation de l'occupation basés sur l'apprentissage non supervisé sont reconnus comme des approches prometteuses pour répondre au problème décrit dans ce travail.

Chapitre 3 - Approche pour la modélisation de l'occupation dans les logements

3.1 Introduction

La littérature scientifique présente différents types de modèles qui ont été développés pour caractériser les services énergétiques des bâtiments. Ces modèles sont largement utilisés pour des applications telles que la prédiction de la charge du réseau, l'anticipation de l'impact des nouvelles technologies (p. ex. la domotique, la production et la gestion locale de l'électricité) et l'analyse de l'évolution des habitudes humaines (p. ex. le télétravail) [58]. Selon la nature des données utilisées et l'objectif visé, Swan et Ugursa [140] catégorise ces modèles en approches descendantes (« top-down ») et ascendantes (« bottom-up »). L'approche descendante utilise des données générales du secteur résidentiel, telles que les historiques de consommation, des données macro-économiques ou des données climatiques afin de calculer la consommation énergétique du secteur résidentiel. L'approche ascendante comprend, quant à elle, l'agrégation stochastique de l'ensemble des consommations électriques des équipements simulées à partir d'un schéma d'utilisation ou issu du comportement des occupants [141]. Parmi ces deux méthodes, l'approche ascendante est l'une des plus utilisées, car elle offre une flexibilité et des détails plus importants, permettant par exemple, l'intégration de technologies telles que les batteries, les voitures électriques et les panneaux photovoltaïques, ainsi que l'influence du comportement humain [55], [142]. Cependant, ce dernier est normalement considéré comme une variable déterministe, par exemple, en supposant un profil quotidien de soutirage de l'eau chaude ou de présence. Néanmoins, tel que Haple et al. [143] le précisent, le comportement humain correspond à un processus stochastique, qui a un impact significatif sur la consommation d'énergie des bâtiments. Par conséquent, pour représenter la variabilité sur la consommation

énergétique d'une ou plusieurs résidences, il est nécessaire d'intégrer le caractère stochastique du comportement des occupants. Ce qui met en évidence la nécessité de développer des modèles comportementaux réalistes. Ainsi, la section 3.2 présente sous la forme d'un article scientifique, le développement d'un modèle probabiliste à deux états permettant la génération des profils d'occupation quotidiens. Ce travail cherche donc à contribuer dans l'amélioration de l'état de l'art existant dans ce domaine de recherche.

3.2 Article 1 : A Probabilistic Model of Household Occupant Behavior for Home Energy Management Applications

Journal : IEEE Access (Volume : 9, Pages : 38187-38201)

Auteurs : Luis Rueda, Simon Sansregret, Brice Le Lostec, Kodjo Agbossou, Nilson Henao and Souso Kelouwani

Date de publication : 2 mars 2021

DOI : <https://doi.org/10.1109/ACCESS.2021.3063502>

3.2.1 Contexte

Les préférences et les habitudes de vie des humains ont un impact significatif sur la consommation d'énergie des bâtiments, ce qui a conduit au développement de modèles comportementaux permettant d'améliorer la conception et la simulation des performances des bâtiments, ainsi que d'améliorer les performances des systèmes de gestion de la demande afin de permettre des économies de consommation d'énergie tout en maintenant le confort des occupants. Parmi les méthodes de collecte de données sur les occupants, les enquêtes d'emploi du temps (EET) ont été largement utilisées, car elles fournissent un large échantillon incluant la variabilité du comportement humain et son lien avec des paramètres humains, sociaux, économiques et autres. Cependant, les modèles comportementaux existants présentent des limites sur des aspects tels que la représentation de la dynamique temporelle du comportement. Par exemple, certaines approches ne tiennent pas compte de la durée des états, et d'autres ont des limites pour bien modéliser les distributions de durée multimodale des périodes de présence/absence et des activités. De plus, les approches existantes utilisent un nombre restreint ou limité de caractéristiques sociodémographiques des individus, ce qui réduit leur

capacité à représenter la variabilité du comportement humain. Par ailleurs, d'autres limites liées à l'interaction entre les habitants d'un même logement, les effets des saisons et des longues absences sont mis en évidence dans la littérature.

3.2.2 *Méthodologie*

Cette première partie de la thèse a été réalisée en deux étapes. Initialement, une étude approfondie des approches pour la modélisation comportementale a été réalisée, permettant d'identifier les limites et les contraintes des modèles existants dans la littérature. Ensuite, en tenant compte de l'analyse effectuée, une stratégie pour la modélisation des profils individuels d'occupation dans les résidences a été proposée. Ce modèle est utilisé pour la génération de profils d'occupation synthétiques pour lesquels les caractéristiques sociodémographiques des utilisateurs, ainsi que la durée des périodes de présence/absence et l'heure de la journée sont explicitement prises en considération.

3.2.2.1 *Hypothèses et considérations de la proposition*

L'approche proposée pour la modélisation de la présence des occupants dans les logements considère les hypothèses suivantes :

- Le comportement de chaque individu est considéré comme indépendant, par conséquent la présence ou les activités des individus ne sont pas affectées par les actions ou les activités des autres membres du ménage.
- Les saisons n'influencent pas le comportement des individus.
- Les absences prolongées (p. ex. les vacances) et les variations exoradiques de comportement (par exemple, une maladie ou la perte d'emploi) ne sont pas pris en compte.

3.2.2.2 *Étapes de l'approche proposée*

L'approche proposée pour la modélisation de l'occupation dans les logements est illustrée par la FIGURE 3-1 et comporte les étapes suivantes :

- *Construction du modèle* : Dans cette étape du travail, un modèle basé sur les données est proposé. Ainsi, dans un premier temps, une étape de prétraitement et de mise en

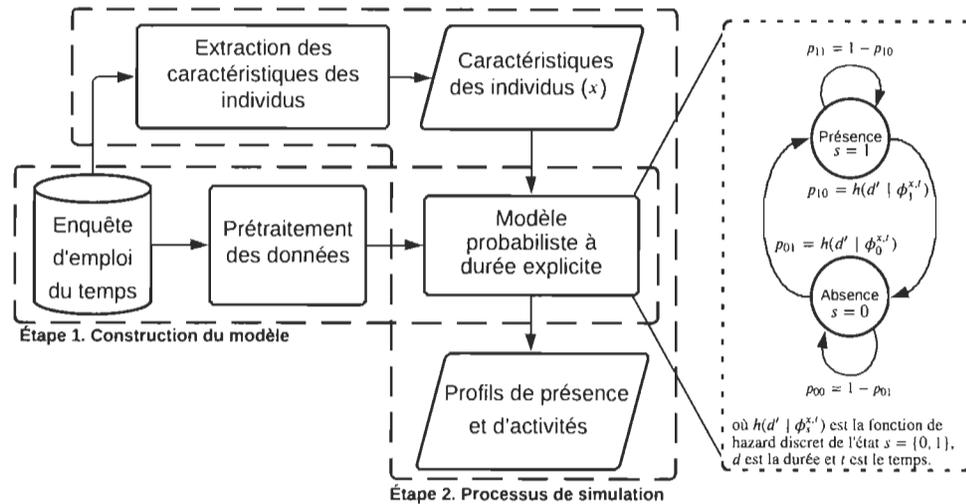


FIGURE 3-1 Schéma bloc de l'approche proposée pour la modélisation de l'occupation dans les bâtiments résidentiels.

forme des données brutes de l'enquête est effectuée. Cela permet d'identifier et corriger les informations manquantes ou d'éventuelles erreurs dans les données. Dans un second temps, nous présentons le développement d'un modèle probabiliste à deux états basé sur l'analyse statistique et probabiliste des profils individuels d'occupation, et les caractéristiques sociodémographiques des individus. Ceci comprend l'utilisation d'une loi de transition conditionnée de façon explicite à la durée des états d'occupation ainsi que la variabilité du comportement humain en fonction des caractéristiques des individus. Par ailleurs, dans le cadre de la construction du modèle, une étape de sélection des caractéristiques est menée. Avec cette analyse, l'objectif est d'identifier les paramètres d'entrée les plus significatifs du point de vue statistique pour le modèle.

- *Processus de simulation* : Une fois que le modèle est construit, les caractéristiques sociodémographiques des individus à simuler sont extraites de la même enquête. Ces caractéristiques sont nécessaires, parce qu'elles correspondent aux paramètres d'entrée du modèle. Ainsi, avec l'ensemble de ces caractéristiques, le processus de simulation est exécuté, ce qui permet d'obtenir les profils d'occupation pour chacun des individus.

Il faut noter que le processus de calibration du modèle proposé est effectué à partir de l'enquête sur l'emploi du temps du Canada d'avril 2015 à avril 2016, fournies par Statistique Canada [68]. Cependant, l'approche proposée peut être appliquée, quel que soit le pays dans

lequel l'enquête a été menée. Cela s'explique par le fait que, indépendamment de l'enquête, en général, les données du sondage EET fournissent un ensemble similaire d'informations sur les répondants, à savoir leur âge, leur genre, leur état civil, leur revenu et une description détaillée des activités réalisées au cours de la journée demandée.

3.2.3 Résultats

Les résultats principaux de cette première partie du travail de thèse peuvent être résumés comme suit :

- Une compréhension approfondie des principaux défis et possibilités pour la modélisation du comportement des occupants, en particulier l'occupation à partir de l'utilisation des données du sondage EET.
- Nous avons présenté les développements d'un modèle de Markov à durée explicite. Cette stratégie utilise la fonction de hasard (complément de la probabilité de survie conditionnelle) pour intégrer la durée des états dans le calcul de probabilités de transition. Les résultats mettent en évidence le potentiel de notre approche, en démontrant que pour le cas étudié, la performance de notre modèle est supérieure à celle des méthodes de la littérature utilisées comme base de comparaison.
- L'hétérogénéité du comportement des individus a également été prise en compte dans l'approche proposée. Pour cela, le modèle semi-paramétrique des risques proportionnels de Cox (régression de Cox) a été utilisé. Avec cette stratégie, il est possible d'estimer l'influence des variables exogènes sur la fonction de hasard. De ce fait, nous tenons compte de l'impact des caractéristiques sociodémographiques sur la durée des états et donc sur la probabilité de transition.
- Les caractéristiques ayant la plus grande signification statistique pour le modèle proposé ont été identifiées. Pour cela, une analyse paramétrique a été réalisée, permettant de réduire la complexité du modèle. Ainsi, nous considérons que les résultats de cette analyse fournissent une base pour les recherches futures, par exemple pour la formulation des questions les plus pertinentes pour une enquête ou pour sélectionner les variables les plus appropriées pour la calibration de modèles comportementaux.

Received February 17, 2021, accepted February 24, 2021, date of publication March 2, 2021, date of current version March 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063502

A Probabilistic Model to Predict Household Occupancy Profiles for Home Energy Management Applications

LUIS RUEDA¹, SIMON SANSREGRET², BRICE LE LOSTEC²,
KODJO AGBOSSOU¹, (Senior Member, IEEE), NILSON HENAO¹,
SOUSSO KELOUWANI³, (Senior Member, IEEE)

¹Department of Electrical and Computer Engineering, Hydrogen Research Institute, University of Quebec at Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada

²Laboratoire des Technologies de l'Énergie, Institut de Recherche Hydro-Québec, Shawinigan, QC G9N 7N5, Canada

³Department of Mechanical Engineering, Hydrogen Research Institute, University of Quebec at Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada

Corresponding author: Luis Rueda (luis.rueda@uqtr.ca)

ABSTRACT Due to the impact of human lifestyle on building energy consumption, the development of occupants' behavior models is crucial for energy-saving purposes. In this regard, occupancy modeling is an effective approach to intend such a purpose. However, the literature reveals that existing occupancy models have limitations related to the representation of occupancy state duration and the integration of occupancy variability among individuals. Accordingly, this paper proposes an explicit differentiated duration probabilistic model to generate realistic daily occupancy profiles in residential buildings. The discrete-time Markov chain theory and the semi-parametric Cox proportional hazards model (Cox regression) are used to predict household occupancy profiles. The proposed model is able to capture occupancy states duration and integrate human behavior variability according to individuals' characteristics. Moreover, a parametric analysis is employed to investigate these characteristics' impact on the model performance and consequently, select the most significant input variables. A validation process is conducted by comparing the model performance with that of previous methods, presented in the literature. For this purpose, the k cross-validation technique is utilized. Validation results demonstrate that the proposed approach is highly efficient in generating realistic household occupancy profiles.

INDEX TERMS Occupancy, behavior, survival analysis, hazard rate, markov-chain, Cox regression.

I. INTRODUCTION

The electric grid faces a transformation in terms of growth and development due to global warming, huge electricity demand, and the impending depletion of fossil energy resources. As a result, the smart grid is promoted as a conceptual framework for improving energy efficiency and advanced management of available resources [1]. The building sector accounts for more than 30% of total world energy consumption that is expected to increase by an average of 1.5% per year between 2012-2040 [2]. Consequently, this sector receives significant attention particularly due to its energy-saving potentials, which can be up to 30% [3].

Energy consumption in buildings is influenced by different factors, including climate, building envelope, building services and energy systems, indoor environment quality, building operation and maintenance, and occupants' behavior [4]. Among these factors, human behavior can be considered as a major issue towards enhancing the performance, design, and simulation of buildings [3]. Occupant behavior refers to the interaction between occupants and buildings in order to preserve a healthy indoor environment and acquire desired comfort, and security [5], [6]. It can be described by occupants presence, and their interaction with appliances, control systems (e.g., heating, ventilation, and air conditioning), and building elements (e.g., doors and windows) [7].

In the last decades, occupant behavior modeling, occupancy detection, and activity recognition have attracted significant attention among researchers whose focus is building

The associate editor coordinating the review of this manuscript and approving it for publication was Amjad Anvari-Moghaddam.

design and operation [7]. These models can be used for applications like building performance simulation and load forecasting, which can help utilities anticipate peak power demand and make decisions on power management, load switching, network reconfiguration, and infrastructure development [8], [9]. Generally, human behavior is dealt with as a deterministic variable, for example occupancy static schedules [5]. Nonetheless, occupant behavior has a stochastic nature and it is influenced by different factors such as weather conditions, physical characteristics of the house, and socio-demographic characteristics of individuals. For this reason, the development of probabilistic occupant behavior models has attracted researchers' attention to integrating variability of individuals' behavior patterns.

Moreover, a better understanding of human-building interactions can improve building energy performance and reduce its energy consumption while maintaining occupants' comfort [10]. Indeed, several works have demonstrated that a significant amount of energy can be saved by performing occupancy-based and activity-based control strategies. For instance, Nguyen *et al.* [11] surveyed intelligent buildings research efforts, focused on energy saving and user activity recognition. The authors concluded that occupancy-based control strategies can result in up to 40% energy savings in HVAC system. Georgievski *et al.* [12] presented an approach based on Hierarchical Task Network (HTN) planning combined with activity recognition. Their proposed method, applied to control lights and workstations in an office building and a restaurant, provided up to 80% energy savings. Layered hidden Markov models (LHMMs) and deterministic finite state machines (FSMs) were used in [13] to identify desk-related activities and count people in an office room. The authors showed that by using activity-based control over lighting systems, it was possible to achieve overall energy savings of 78.5% per year compared to manual control. Additionally, by using a people count-based control strategy, it was possible to reduce annual HVAC energy consumption by 39.9%. Scott *et al.* [14] used PIR and RFID sensors' information to predict home occupancy patterns for heating control. The authors analyzed three different control techniques (based on scheduled, always-on, and PreHeat algorithms) and consequently, obtained energy savings of up to 35%. Boait and Rylatt [15] presented a prototype to control the heating system in a house in the UK. The prototype utilized Bayesian inference to identify the occupancy state based on electricity consumption data and hot water temperature. The authors achieved up to 14.1% energy savings in the gas, consumed by the boiler (19.2 kWh/day). Likewise, heterogeneous occupancy patterns were used to evaluate a simulation-based optimization approach for the design of energy management systems (EMS), applied to microgrids [16], [17]. Moreover, a zonal control strategy of thermostatic loads was presented in [18]. The authors demonstrated that their strategy could save more than 15% energy consumption while increasing the thermal comfort by more than 25%.

A. BACKGROUND

A comprehensive understanding of individuals' behavior inside and outside of the house is of great importance for sectors such as energy [2] and transportation [19]. Literature reveals that Time-use survey (TUS) and Travel survey (TS) are important sources of information to achieve such comprehension [20], [21]. Particularly, TUS data has been widely used for the development of probabilistic occupant behavior models, which primarily focus on individuals in the residential sector [22]. It can provide detailed information about the daily routine of a large number of respondents and capture the stochastic nature of occupant behavior and its variability according to individuals' characteristics. Therefore, TUS is suitable for representing a population [5]. Moreover, the literature reveals that independent from the country in which the survey is conducted, TUS provides a similar set of information about respondents, for instance their age, gender, marital status, and income. Accordingly, it allows the development of highly replicable methods regardless of the survey location. Walker and Pokoski [23] proposed a model of residential demand based on customer behavior, captured from surveys. The authors were one of the first to use TUS data to model human behavior. They proposed an "availability" and a "proclivity" function to estimate the number of people in a household and their probability of doing a certain activity at specific time periods. However, one of the main drawbacks of this model, which limits its application scope, is its insufficient flexibility to analyze behavioral variability related to users' socio-demographic characteristics.

Among the techniques, used to model human behavior by using TUS data, first-order Markov chains (FOMC) and higher-order Markov chains (HOMC) are the most popular ones. A first-order MC is a memoryless stochastic process according to its Markovian property. Therefore, the transition probability between states only depends on the current state. Richardson *et al.* [24] used this technique to generate active occupancy¹ daily profiles at a ten-minute resolution based on the United Kingdom 2000 time-use survey [25]. The approach was able to indicate the number of active occupants in the house. However, only the number of household residents and the weekdays were used as parameters to characterize occupancy variability. Furthermore, Richardson *et al.* extended their occupancy model in another study [26] to generate activity profiles in order to develop a domestic electricity demand model. Widen *et al.* [27] proposed a first-order Markov Chain Monte Carlo technique to generate domestic lighting demand data from occupancy patterns and daylight availability. The authors presented a simplified model in which all persons were characterized by four parameters related to dwelling (detached houses and apartments) and day type (weekdays and weekends). Subsequently, in [28], they refined their initial model to generate occupant activity

¹Active occupancy refers to a moment when a user is in the house and performs an activity other than sleeping.

sequences and use them to estimate domestic electricity demand.

Likewise, Muratori *et al.* [29] presented a heterogeneous Markov chain to generate activity patterns. The authors used a method similar to the one presented by Widen and Wackelgard [28]. In addition, they included a small set of parameters to assess behavior differences among individuals, represented by working, non-working, male, female, and children classes. Collin *et al.* [30] developed a thirteen-state first-order Markov model, to create activity profiles, differentiated by household size as well as working and non-working individuals. Afterwards, the authors applied the model to generate demand profiles according to user activities. They employed a probabilistic function to identify electrical appliances, shared by users performing the same activity. Baptista *et al.* [31] proposed the utilization of the interactive Markov chain approach, presented by Conlisk, [32] to incorporate interaction between individuals in a household. For this purpose, the transition probability matrices were conditioned by the present activity of a leader. In this way, occupants were able to explicitly coordinate their activities. However, this approach did not consider a comprehensive analysis of behavior variability according to individuals' characteristics. The articles, presented above, emphasize that behavior variability has not been adequately explored. Moreover, they share the disadvantage of a memoryless property, yielded by the first-order Markov models. [33], [34] has revealed that first-order Markov models are not able to consistently predict state duration and thus, their generated occupancy or activity profiles often fail to reflect a realistic behavior.

Alternatively, some authors have proposed the use of higher-order Markov models. These probabilistic models do not possess the memoryless property of traditional Markov chain models. Therefore, the duration is not required to assume exponential or geometric distributions over continuous or discrete-time cases, respectively. Accordingly, Tanimoto [35], Wilke *et al.* [33], [36], and Vorger [37] captured occupants' behavior from a semi-Markovian process. In this process, the states are conditioned by their precedents through a Markovian transition, and the duration is conditioned by only the current state. In fact, the procedure is based on a decremental counter that forces a state transition every time it reaches "1". Once a new state is determined, the counter initial value is estimated by a duration distribution. For such an estimation, Tanimoto used a logarithmic distribution to generate the duration of activities. Wilke *et al.* and Vorger used a Weibull distribution to determine presence and activity duration. However, these distributions put limitations on capturing the multimodal character that often occurs in duration data.

Furthermore, Flett *et al.* [34], [38] presented a model to produce occupancy profiles. They evaluated the impact of relationships between cohabiting individuals on the overall active occupancy probability. The authors developed a higher-order Markov method where multiple probability transition matrices were generated according to the existing

state duration. Aerts *et al.* [39], [40] presented a methodology to identify 'typical occupancy patterns' from Belgian time-use data by using hierarchical clustering. These patterns were used to calibrate a three-state probabilistic model (absent, asleep, and at home or awake) to generate individual daily and yearly occupancy sequences. However, none of the above methods include a detailed set of parameters to assess behavior differences. This, in turn, has limited the analysis of individuals characteristics to age, gender, income and employment status. Consequently, it can be deduced that although some higher-order Markov-based methods have been proposed, there are still lacks in the representation of state duration, mainly related to occupant-specific behavior variations and multimodal duration distributions generation.

In addition, existing occupant behavior models have other limitations that must be addressed to enable a more reliable residential occupancy and activity prediction. For the simulation of behavior in multi-person households, most of the approaches have treated users as independent agents, which does not reflect actual human behavior in dwellings. In real life, occupants interact and thus, they can perform activities together. Accuracy in modeling such forms of interaction can significantly impact the simulation of load curves since energy consumption changes according to whether or not occupants share appliances. It should be noted that seasonal effects and long absence periods due to holidays and illness are additional concerns with occupancy modeling that have not been adequately studied. This can be partially associated with limited information available in the surveys. Generally, surveys supply daily data of respondents and skip information of all inhabitants and longer periods.

B. CONTRIBUTION

This paper aims to address the aforementioned issues related to designing a model capable of describing occupancy state duration while incorporating occupancy variability among individuals. Literature shows that the hazard-based model is a suitable approach for dealing with these concerns since it is able to capture duration-dependent transition probabilities and incorporate exogenous variables [41]. In fact, this method has proved its potential for applications such as windows opening and closing behavior modeling [42], activity-travel modeling, as well as in-home and out-of-home activity generation [19]. Accordingly, we present an explicit-duration probabilistic model that combines the semi-parametric Cox proportional hazards model (Cox regression) and the discrete-time Markov chain theory in order to effectively handle the challenges, faced in this study. Cox regression is a technique that at each time step estimates a state chances of making transition according to its duration up to that time [43]. This method is normally used for survival analysis. Moreover, unlike parametric models, the proportional hazards approach does not restrict the hazard function shape to a particular distribution. Furthermore, this method allows the hazard function to consider several explanatory variables and thus, enables the model to include behavior variability.

Additionally, regarding the diversity of individuals' characteristics, a parametric analysis is performed to assess their impact on the model capability to represent occupancy variability. Besides, a variance-based analysis and a backward elimination technique are used to identify the most statistically significant parameters of the proposed model. Moreover, an improvement of the approach, proposed by Wilke et al., [33] is presented. This improvement seeks to overcome the limitations of the Weibull distribution to fit the multimodal characteristic of duration data. In order to achieve this ambition, we utilize a Gaussian mixture model (GMM) along with a silhouette analysis. Hence, it is possible to automatically identify the number of clusters of each duration distribution in order to represent them as normally distributed sub-populations.

C. OUTLINE

In Section II, the time-use survey data, employed to calibrate the model, is described. Section III presents the probabilistic model, proposed to predict realistic occupancy profiles. Section IV explains the utilized parametric analysis. Section V describes the validation process and benchmarks the model performance with other approaches, proposed in the literature. In Section VI, the results and future prospects are discussed. Finally, the conclusions are presented in Section VII.

II. TIME-USE SURVEY DATA

This paper focuses on domestic occupancy modeling by using national time-use survey (TUS) data. In particular, the calibration process of the proposed model is performed based on the Canada TUS data from April 2015 to April 2016, supplied by Statistics Canada [44]. This survey comprises a sample size of 15,390 respondents of a target population that includes all persons, aged 15 years old or more in Canada. Each individual diary provides a detailed record of a wide variety of daily activities and their devoted time, location, and other participants (except for the respondent). In this survey, the diary starts at 4:00 AM with a list of 266 activities. The survey presents information of up to three simultaneous activities. In addition, it provides information about perceptions of time, unpaid work periods, well-being, paid work and education time, cultural and sports activities, transportation, as well as numerous socio-demographic characteristics.

A. OCCUPANCY PATTERNS DIFFERENTIATION

In order to represent variability and stochastic nature of the occupancy, a set of characteristics, defined by $x = \{x_1, x_2, \dots, x_M\}$ are used to describes each individual. As presented in TABLE I, these characteristics include personal information about the respondent (including age, occupation, and marital status), the family composition, and the household.

Individuals' characteristics enable the model to capture the occupancy variability of populations with different socio-demographic features. As illustrated in FIGURE 1, the

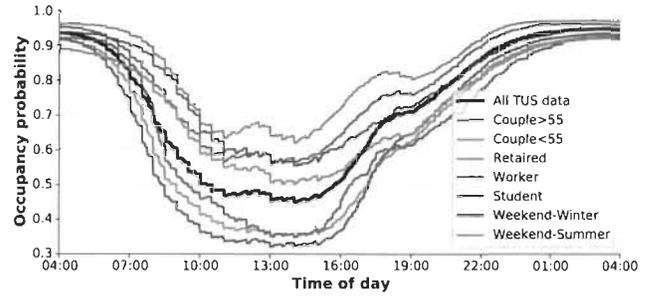


FIGURE 1. Occupancy patterns differentiation.

TABLE 1. Individuals characteristics.

Name	Description
x_1	Weekday
x_2	Type of day
x_3	Month
x_4	Sex of respondent
x_5	Age of respondent
x_6	Marital status of respondent
x_7	Respondent has a spouse/partner living in the household
x_8	Age difference between respondent and spouse/partner
x_9	Sex of respondent's spouse/partner living in the household
x_{10}	Type of partner respondent has living in the household
x_{11}	Child(ren) of the respondent living in the household
x_{12}	Age of respondent's youngest child in household
x_{13}	Age of youngest household member in respondent's household
x_{14}	Number of respondent's child(ren) in household
x_{15}	Number of respondent's child(ren) in household - 0 to 14 years
x_{16}	Age group of respondent's child(ren) in household
x_{17}	Number of respondent's parents in the household
x_{18}	Living arrangement of respondent's household (8 categories)
x_{19}	Living arrangement of respondent's household (11 categories)
x_{20}	Household size of respondent
x_{21}	Number of members in respondent's household (≥ 15 years)
x_{22}	Three generations or more in the respondent's household
x_{23}	Main activity - Last 12 months
x_{24}	Main activity of spouse/partner - 12 months
x_{25}	Dwelling type of the respondent's household
x_{26}	Educational attainment - Highest degree
x_{27}	Income - Personal income group (before tax)
x_{28}	Household income - Household income group (before tax)

probability of an individual's presence at home (notwithstanding the activity he/she performs) is different regarding the targeted sub-population. Furthermore, the time-varying nature of behavior is observed according to the hour of the day, the weekday, and the month of the year.

B. CENSORED DATA

Since the duration is taken from one-day diaries, information about the episodes that last until the end of the diary day (4 AM) is incomplete. This can be due to the fact that these episodes probably last for additional time and thus, their exact duration is unknown (right-censored observations). Accordingly, these events, considered as censored data, are identified and included in the estimation of the transition probabilities.

Different statistical methods such as complete-data analysis, imputation approach, and dichotomized data-bases

analysis are commonly used to deal with censored data [45]. However, likelihood-based strategies like Kaplan-Meier estimator and Cox-regression have been promoted as effective methods that can handle all available information, whether it presents censored data.

III. MODELING APPROACH

The proposed approach to modeling residential occupant behavior consists of a non-homogeneous semi-Markov process (NHSMP)² to predict individuals' occupancy profiles according to temporal information (e.g., hour, weekday, and month) and socio-demographic characteristics.

A. EXPLICIT-DURATION PROBABILISTIC MODEL

Individual occupancy sequences are generated through a two-state probabilistic model that produces daily profiles with a time-resolution of 10 min. Occupancy profile is represented by the vector $\vec{z} = [z_1, z_2, \dots, z_N]$ with a length of $N = 144$. This profile can be represented by a finite-state machine (FSM) with two possible states $s = \{0, 1\}$. In discrete-time $k = 1, 2, \dots, N$, the occupancy state, z_k is zero when the machine is at the "absence" state and one when it remains in the "presence" state. Therefore, by using the chain rule, the probability of the sequence $P(\vec{z})$ can be described in terms of conditional probabilities, expressed by:

$$P(\vec{z}) = P(z_1) P(z_2|z_1) P(z_3|z_2, z_1) \dots P(z_N|z_{N-1}, \dots, z_1) \quad (1)$$

Since equation 1 covers occupancy profile dynamic, its calculation becomes a complicated challenge if the current state of the machine has a long dependency on its previous state. Consequently, the dimension of the conditional laws becomes too large when the sequence lengthens. However, the development of approximations and the adoption of hypotheses that reduce the history of states sequence can bring the problem to an affordable dimension. Accordingly, a hypothesis of medium-term independence is introduced in order to truncate the calcul of $P(\vec{z})$. Such hypothesis implies that the probability of changing a state depends on the current state and its duration, explained by,

$$P(z_{k+1}|z_k, z_{k-1}, \dots, z_2, z_1) = P(z_{k+1}|z_k, d_k) \quad (2)$$

By applying this hypothesis, the sequence of states, \vec{z} can be represented as a Semi-Markov or explicit duration process $P(z_{k+1}|z_k, d_k)$, in which the duration, d_k is considered as an incremental counter, restarted ($d_k = 1$) when a transition is made ($z_{k+1} \neq z_k$). In this study, the system has only two states. Therefore, when a transition must be made, the new state, z_{k+1} is selected in a deterministic manner. By applying Bayes' theorem and chain rule to the expression on the right side of the equation 2, the transition probability can be

defined as,

$$P(z_{k+1} | z_k, d_k) = \frac{P(d_k | z_{k+1}, z_k) P(z_{k+1} | z_k)}{P(d_k | z_k)} \quad (3)$$

When estimating the transition probability, two situations can occur. These cases account for state change, expressed by equation (4), or state preserve, defined by equation (5), as below.

$$P(z_{k+1} \neq s | z_k = s, d_k = d') \quad (4)$$

$$P(z_{k+1} = s | z_k = s, d_k = d') \quad (5)$$

These two situations are mutually exclusive. As a result, by analyzing the circumstance in which a change in state takes place (according to the literature [46], [47]), the expressions on the right side of equation 3 can be reformulated as,

$$P(d_k = d' | z_{k+1} \neq s, z_k = s) = f(d' | \theta_s) \quad (6)$$

$$P(z_{k+1} \neq s | z_k = s) = \frac{1}{\mathbb{E}[d_k = d' | z_{k+1} \neq s, z_k = s]} \quad (7)$$

$$P(d_k = d' | z_k = s) = \frac{S(d' | \theta_s)}{\mathbb{E}[d_k = d' | z_{k+1} \neq s, z_k = s]} \quad (8)$$

where $f(d_k = d' | \theta_s)$ is the state duration probability, $S(d_k = d' | \theta_s)$ is the survival function, associated with state s , $\mathbb{E}[d_k = d' | z_{k+1} \neq s, z_k = s]$ is the mathematical expectation of state duration distribution, and θ_s is a set of parameters, related to state duration distribution.

Consequently, by combining equations (3) and (6)-(8), the probability of transition between states can be described by,

$$P(z_{k+1}|z_k = s, d_k = d') = \begin{cases} h(d' | \theta_s) & \text{if } z_{k+1} \neq s \\ 1 - h(d' | \theta_s) & \text{otherwise} \end{cases} \quad (9)$$

where $h(d' | \theta_s)$ is the discrete hazard function of state s , which can be expressed as,

$$h(d' | \theta_s) = \frac{f(d' | \theta_s)}{S(d' | \theta_s)} = 1 - \frac{S(d' + 1 | \theta_s)}{S(d' | \theta_s)} \quad (10)$$

Besides, time dependency is an important factor in human behavior modeling. For this reason, temporal information is included to the model by calculating the hazard function in hourly intervals $t \in [0, 1, \dots, 23]$. Accordingly, the model parameters, θ are affected not only by the state s , but also by the time of day t . This results in a hazard function with the form $h(d' | \theta_s^t)$.

In order to include occupancy variability in the model, we must evaluate the impact of individuals' characteristics (covariates) on the hazard function. Accordingly, we use the Cox proportional hazard model (Cox regression). This semi-parametric method allows to estimate the relationship between the hazard rate and the explanatory variables without any assumptions about the shape of the baseline hazard

²A non-homogeneous semi-Markov process (NHSMP) is a generalization of Markov chains and the renewal process, where transition probability between states are time-varying and depend on sojourn times.

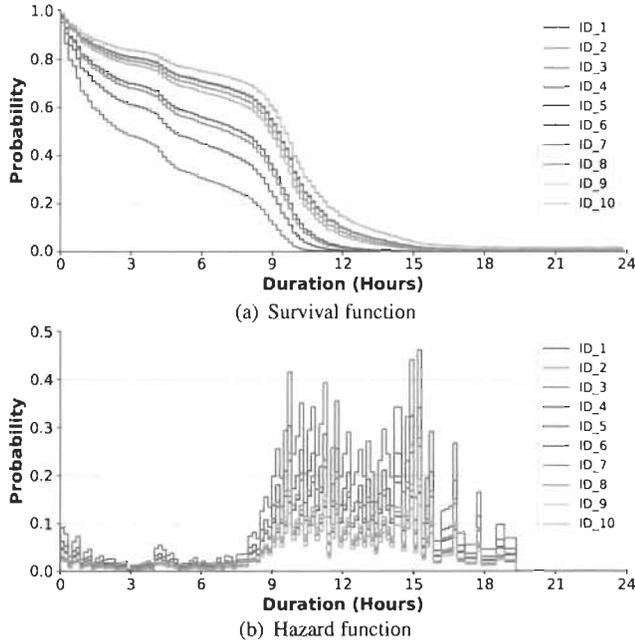


FIGURE 2. Survival and hazard function variability.

function [43], based on,

$$h(d' | x, \theta'_s) = h(d' | \phi_s^{x,t}) = h_0(d' | t, s) e^{\beta(x,t,s)x'} \quad (11)$$

where $h_0(d' | t, s)$ is a non-parametric baseline hazard function, which relates to the value of the hazard when all covariates are zero ($x = 0$), $e^{\beta(x,t,s)x'}$ is a duration-independent function to represent the effect of the individuals' characteristics, and ϕ represents the parameters of the hazard function $\phi = \{h_0, \beta\}$. FIGURE 2 shows ten randomly selected individuals' survival and hazard functions for absence periods between 4 and 5 AM. It can be observed that individuals 3 and 9 are the ones with the highest and lowest transition probability, respectively. Periods of absence that begin in this time range are likely to last for 9 to 10 hours. Thereafter, their related transition probability increases considerably. In fact, this figure is an example to depict the heterogeneity of survival functions that promotes the computation of the hazard function of each individual. This calculation can be handled by Cox-regression methods since they can integrate the influence of exogenous variables (individuals characteristics) on transition probabilities and consequently, lead to more realistic occupancy profiles.

From (11), the conditional transition probability for the two-state FSM can be explained by matrix M :

$$M = \begin{bmatrix} 1 - h(d' | \phi_0^{x,t}) & h(d' | \phi_0^{x,t}) \\ h(d' | \phi_1^{x,t}) & 1 - h(d' | \phi_1^{x,t}) \end{bmatrix} \quad (12)$$

where $h(d' | \phi_0^{x,t})$ and $h(d' | \phi_1^{x,t})$ are the hazard rates when $s = 0$ and 1, respectively.

B. SIMULATION PROCESS

In order to illustrate the dynamic of the two-state machine, used for the generation of the individual occupancy profiles,

Algorithm 1 Occupancy Profile Generation

Input: Individuals characteristics: x_1, x_2, \dots, x_M

Output: Discrete state sequence: z_1, z_2, \dots, z_N

```

1 begin
2   sampling  $z_1 \sim [\pi_0, \pi_1]$ 
3   initialize  $d_1 = 1$ 
4   for  $k=2,3,\dots,N$  do
5     sampling  $\tau_k \sim \text{Bern}(h(d = d_{k-1} | \phi_{z_{k-1}}^{x,t_k}))$ 
6     update  $z_k = z_{k-1} \oplus \tau_k$ 
7     update  $d_k = (1 - \tau_k)d_{k-1} + 1$ 
8   end
9 end

```

the pseudo-code of the simulation process is presented in Algorithm 1. In this algorithm, the transition signal, $\tau_k \in \{0, 1\}$ is sampled in every time-step from a Bernoulli distribution with $\alpha = h(d' | \phi_{z_{k-1}}^{x,t_k})$ (see line 5 of the Algorithm 1). According to this transition signal, the algorithm (i) updates the occupancy state, z_k by using an XOR function, and (ii) restarts the counter, d_k , which determines the elapsed time in each state.

Once the simulation process is completed, a general verification of the model performance is carried out.³ Hence, the model capacity to describe the occupancy of the whole TUS population is analyzed. FIGURE 3(a) presents a comparison between overall presence probabilities that have been obtained from the survey and the simulation data. This Figure shows the proposed approach effectiveness to represent the population's behavior.

Furthermore, the model capability to describe the duration of the system states has been analyzed. FIGURE 3(b) indicates a comparison between cumulative probability functions (CDF) of absence duration at 6 am. Although in this case, the distribution of duration follows a bimodal shape, the proposed approach is able to adequately reproduce its related pattern.

C. PERFORMANCE INDICATORS

In the following sections of the article, the performance of the model is evaluated and compared with other existing approaches in the literature by using the following metrics.

- 1) Mean absolute error (MAE): This metric is used to determine the model capability to represent the average occupancy state of different populations. The MAE can be expressed as,

$$MAE = \frac{1}{N} \sum_{k=1}^N \left| \bar{P}_s^{mod}(k) - \bar{P}_s^{tus}(k) \right| \quad (13)$$

³A detailed analysis of the results and their comparison with other approaches, presented in the literature, are discussed in Section V

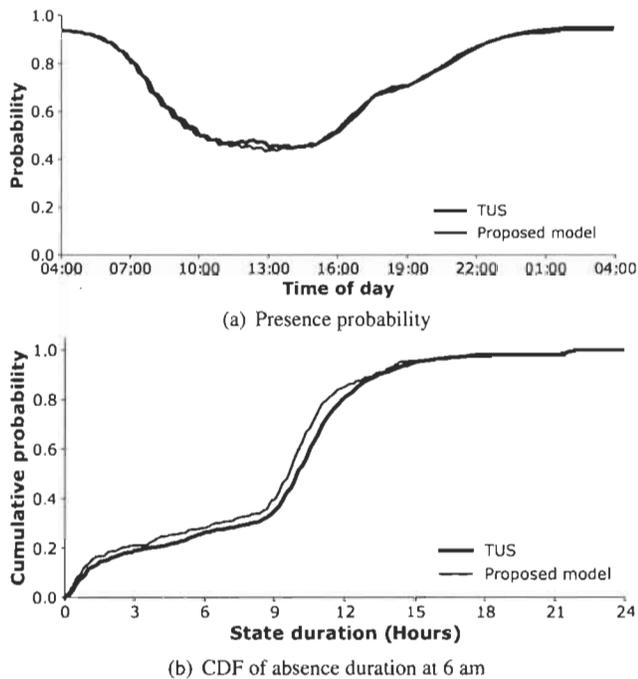


FIGURE 3. Simulation results for all the survey population.

where $\bar{P}_s^{mod}(k)$ and $\bar{P}_s^{tus}(k)$ are average probabilities of the state s at time-step k , derived from the model simulation and the TUS data, respectively.

- 2) First-Wasserstein distance: Also known as Earth mover's distance (EMD), this metric assesses the model ability to represent the duration of the occupancy states. This indicator is defined by,

$$W_1 = \sum_{d=1}^{144} \left| D_s^{mod}(d) - D_s^{tus}(d) \right| \quad (14)$$

where, $D_s^{mod}(d)$ and $D_s^{tus}(d)$ are CDFs of duration of the state s , derived from the the model simulation and the TUS data, respectively.

IV. PARAMETRIC ANALYSIS

Considering the diversity of individuals' characteristics, provided by the survey, a global sensitivity analysis (GSA) is performed. GSA makes it possible to quantify each characteristic contribution to the model output variability. As a result, the most important input variables of the model can be identified. The GSA is conducted by exploiting variance-based and regression methods.

A. VARIANCE-BASED ANALYSIS

Variance-based methods evaluate the impact of model input variability on the uncertainty of its output. Therefore, given a model with the form $y = f(x_1, x_2, \dots, x_M)$, the first-order Sobol' index is employed to quantify the relative contribution of x_i to the uncertainty of y , while excluding the interaction effect of other parameters. Additionally, the total effect index is used to evaluate the total effect of x_i considering its inter-

action with all others parameters. These two steps are carried out by (15) and (16), respectively,

$$S_i = \frac{V_{x_i}(E_{x_{\sim i}}(y | x_i))}{V(y)} \quad (15)$$

$$S_{Ti} = \frac{E_{x_{\sim i}}(V_{x_i}(y | x_{\sim i}))}{V(y)} = 1 - \frac{V_{x_{\sim i}}(E_{x_i}(y | x_{\sim i}))}{V(y)} \quad (16)$$

where V and E denote the variance and expected value operators, respectively, x_i stands for the i -th characteristic and the set $x_{\sim i}$ contains all input variables except for x_i .

FIGURE 4 presents the result of the sensitivity analysis, performed on 10 trials of 45000 samples. The results, shown by FIGURE 4(a), evidence that most characteristics have first-order sensitivity between 4 AM and 5 PM. This coincides with the period of the day during which the variation in occupancy is the highest due to activities such as working or studying. Besides, it is observed that the characteristics between x_{20} and x_{28} have a slightly higher sensitivity index, which allows us to draw inferences about their significance.

Moreover, FIGURE 4(b) depicts that the total-index is greater than the first-order index for all the characteristics. This reveals that higher-order interactions exist between the input variables. Although some characteristics do not have high importance individually, their interaction with other inputs can impact the model performance.

In addition, FIGURE 5 presents the results of the sensitivity analysis without differentiating between the time of day. It can be noticed that both first-order and total-order indexes can identify some important characteristics for modeling such as x_{23} , x_{27} , x_{28} , x_5 and x_{25} .

B. REGRESSION ANALYSIS

Regression methods assist with obtaining a model whose output is described by a linear combination of its input parameters. Therefore, the regression coefficient of a given characteristic can be considered as a sensitivity measure [48]. Backward elimination is the regression technique that has been used in this study. It is applied to each model through the following steps. Initially, each model is calibrated by using all the variables. Afterwards, the variable with the highest p -value is identified and compared with a significance level, $\alpha = 0.05$. If p -value $> \alpha$, the variable is eliminated and the model is re-calibrated with the remaining characteristics. This process is repeated until all the input variables have p -value lower than α .

TABLE 2 summarizes the ten most important model characteristics according to the results of the feature selection process. It can be noted that these features are consistent with the ones, captured by the variance-based method. In fact, the age, (x_5) and the main activity of respondent, (x_{23}) can be presented as the two most important characteristics. Conversely, x_8 and x_{10} can be reported as the least significant input variables.

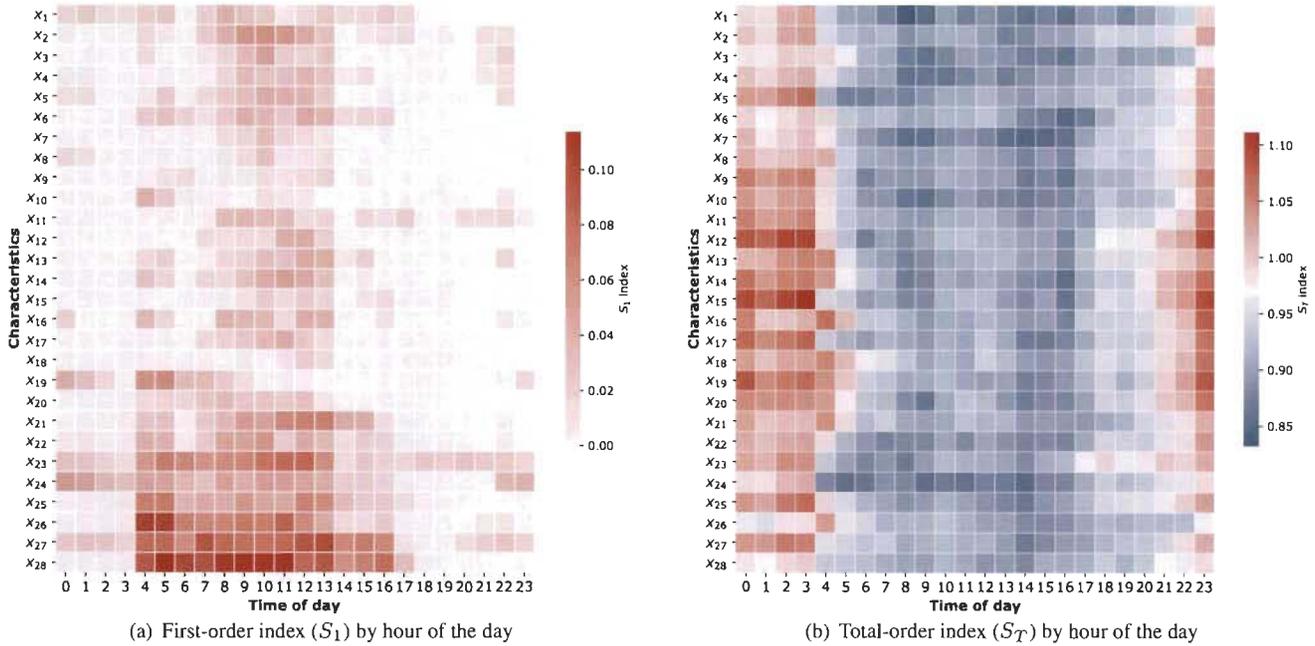


FIGURE 4. Mean estimation with 45000 samples from 10 retrials of first-order and total-order Sobol' indexes.

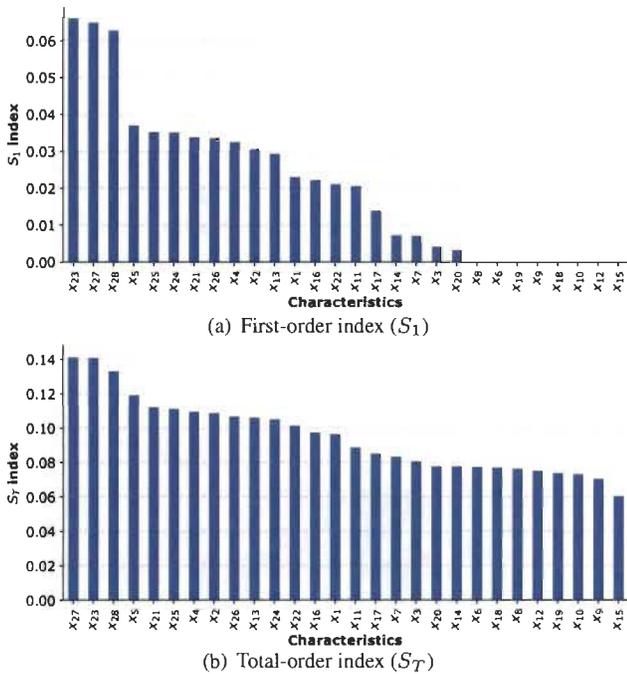


FIGURE 5. First-order and total-order Sobol' indexes without differentiation between time of day.

In order to verify the impact of the feature selection on the model performance, a comparison has been made between the model, fitted with the full set of characteristics and the one, obtained from the feature selection process. In order to reduce the risk of overfitting, a k -fold cross-validation technique has been utilized. Accordingly, the TUS data set has been divided into $k = 10$ subsets with equal sizes. For each subset, the individuals have been chosen randomly based on a uniform probability distribution. These subsets have

TABLE 2. Most significant individuals' characteristics.

	Description
x_5	Age of respondent
x_{23}	Main activity - Last 12 months
x_{26}	Educational attainment - Highest degree
x_1	Weekday
x_{16}	Age group of respondent's child(ren) in household
x_{13}	Age of respondent's youngest child in household
x_{28}	Household income
x_2	Type of day
x_{21}	Number of members in respondent's household (≥ 15 years)
x_{27}	Personal income

been divided into one validation set and $k - 1$ training sets to calibrate the model. This process has been repeated k times for a different validation set at each time.

The comparison results, presented in FIGURE 6, show that the backward elimination method has not substantially affected the model performance. Furthermore, it has reduced the number of model parameters from 1344 to 220 without jeopardizing the model effectiveness in explaining sub-populations' occupancy variability. It should be noted since two hazard functions (one for the presence and one for the absence) have been used for every hour, the resultant reduction is notable.

V. VALIDATION

The proposed model is validated by comparing its performance with the following methods, which have been studied in the literature.

- 1) First-order Markov model: This method has the Markov property and thus, it presents a system in which, the transition probability between the states only depends on the current state. In this model,

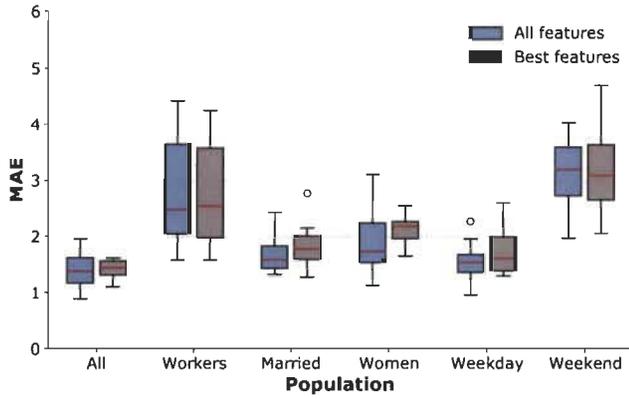


FIGURE 6. Comparison of the model performance with and without the feature selection phase.

the transition probabilities are calibrated at hourly intervals and behavior variability between individuals is integrated by using a logistic regression, defined by,

$$p(\mathbf{x}, t, s) = \frac{1}{1 + e^{-(\beta_0(x,t,s) + \beta_1(x,t,s)x_1 + \dots + \beta_M(x,t,s)x_M)}} \quad (17)$$

where \mathbf{x} is the vector of the individuals' characteristics, t presents the hour of the day, and s stands for the system state.

Afterwards, to perform a simulation with a 10-minute time step, the Chapman-Kolmogorov equation is employed [36]. In this way, ten-minutes transition matrices can be computed based on the previously calculated hourly matrices ($M_{10 \text{ min}} = M_t^{1/6}$). Hence, the transition between the system states can be explained by the matrix M as,

$$M = \begin{bmatrix} 1 - p(\mathbf{x}, k, s = 0) & p(\mathbf{x}, k, s = 0) \\ p(\mathbf{x}, k, s = 1) & 1 - p(\mathbf{x}, k, s = 1) \end{bmatrix} \quad (18)$$

The simulation process, used for the first-order Markov model, is described by the pseudo-code in Algorithm 2. Therefore, the system state at time instant k is determined according to the transition signal τ_k . This signal is generated at every step of the simulation by sampling from a Bernoulli distribution with $\alpha = p(\mathbf{x}, k, z_{k-1})$.

- 2) Duration ranges model: This model is an improvement to the first-order Markov model based on Flett's approach [34], [38]. It is a time-dependent higher-order Markov model that generates multiple probability transition matrices in accordance with the duration, d_k of the existing state (equation (19)).

Therefore, the transition probability, used to evaluate the occurrence of a state transition, not only varies over time k but also depends on the duration of the current state. In this model, the system state, z_k and duration, d_k are updated according to the estimated transition signal, τ_k (see Algorithm 3).

Algorithm 2 First-Order Markov Model

Input: Individuals characteristics: x_1, x_2, \dots, x_M
Output: Discrete state sequence: z_1, z_2, \dots, z_N

```

1 begin
2   sampling  $z_1[\pi_0, \pi_1]$ 
3   initialise  $k = 1$ 
4   for  $k = 2, 3, \dots, N$  do
5     sampling  $\tau_k \sim \text{Bern}(p(\mathbf{x}, k, z_{k-1}))$ 
6     update  $z_k = z_{k-1} \oplus \tau_k$ 
7   end
8 end
```

Algorithm 3 Duration Ranges Model

Input: Individual characteristics: x_1, x_2, \dots, x_M
Output: Discrete state sequence: z_1, z_2, \dots, z_N

```

1 begin
2   sampling  $z_1[\pi_0, \pi_1]$ 
3   initialise  $d_1 = 1$ 
4   for  $k = 2, 3, \dots, N$  do
5     if  $d_k \leq 12$  then
6       sampling  $\tau_k \sim \text{Bern}(p_{0-2}(\mathbf{x}, k, z_{k-1}))$ 
7     else if  $12 < d_k \leq 24$  then
8       sampling  $\tau_k \sim \text{Bern}(p_{2-4}(\mathbf{x}, k, z_{k-1}))$ 
9     else if  $24 < d_k \leq 36$  then
10      sampling  $\tau_k \sim \text{Bern}(p_{4-6}(\mathbf{x}, k, z_{k-1}))$ 
11    else if  $36 < d_k \leq 48$  then
12      sampling  $\tau_k \sim \text{Bern}(p_{6-8}(\mathbf{x}, k, z_{k-1}))$ 
13    else
14      sampling  $\tau_k \sim \text{Bern}(p_{8+}(\mathbf{x}, k, z_{k-1}))$ 
15    end
16    update  $z_k = z_{k-1} \oplus \tau_k$ 
17    update  $d_k = (1 - \tau_k)d_{k-1} + 1$ 
18  end
19 end
```

$$p(\mathbf{x}, k, s | d_k) = \begin{cases} p_{0-2}(\mathbf{x}, k, s) & \text{if } d_k \leq 2 \text{ hrs} \\ p_{2-4}(\mathbf{x}, k, s) & \text{if } 2 < d_k \leq 4 \text{ hrs} \\ p_{4-6}(\mathbf{x}, k, s) & \text{if } 4 < d_k \leq 6 \text{ hrs} \\ p_{6-8}(\mathbf{x}, k, s) & \text{if } 6 < d_k \leq 8 \text{ hrs} \\ p_{8+}(\mathbf{x}, k, s) & \text{if } d_k > 8 \text{ hrs} \end{cases} \quad (19)$$

- 3) Event model: It is a probabilistic model, in which transitions between states are estimated by drawing a duration, Δk from the corresponding probability distribution function (PDF) of duration $f(d | \theta_s^{\mathbf{x}, t})$. In order to estimate individual-dependent PDFs, a binary tree structure is used to split duration data according to a chosen set of individuals' characteristics. The selected characteristics explain a significant statistical

Algorithm 4 Event Model

Input: Individuals characteristics: x_1, x_2, \dots, x_M
Output: Discrete state sequence: z_1, z_2, \dots, z_N

```

1 begin
2   sampling  $z_1 [\pi_0, \pi_1]$ 
3   initialise  $s = z_1$ 
4   initialise  $k = 1$ 
5   while  $k \leq N$  do
6     sampling  $\Delta k \sim f(d | \theta_s^{x,t_k})$ 
7     update  $z_{[k:k+\Delta k]} = s$ 
8     update  $s = s \oplus 1$ 
9     update  $k = k + \Delta k + 1$ 
10  end
11 end

```

difference between the subgroups' average duration, obtained in the branches at the end of the tree (more details can be found in [33], [36]). As a result, each individual corresponds only to one of the distinct subsets at the bottom of the tree. Afterwards, the PDFs are fitted at hourly intervals ($t \in [0, 1, \dots, 23]$) by using a Weibull distribution, defined by,

$$f(d | \theta_s^{x,t}) = e^{-\left(\frac{d}{\alpha(x,t,s)}\right)^{\beta(x,t,s)}} - e^{-\left(\frac{d+1}{\alpha(x,t,s)}\right)^{\beta(x,t,s)}} \quad (20)$$

where $\theta_s^{x,t}$ represents the scale and shape parameters of the Weibull distribution.

As illustrated in Algorithm 4, once a transition occurs at time-step k , the duration Δk of the new state is estimated (row 6). Therefore, a deterministic process takes place between k and $k + \Delta k$, where the state of the system remains unchanged (row 7). Once the simulation reaches the time-step $k + \Delta k$, the system is forced to change the state (row 8) and update the k value.

- 4) Improved event model: As discussed above, the event model use a binary tree structure to divide the duration data in order to estimate individual-dependent PDFs. However, as stated in the literature [37], the Weibull distribution has limitations in capturing the multimodal character of duration data. Accordingly, we propose the utilization of Gaussian Mixture Model (GMM) instead of Weibull distribution. GMM assists with a better representation of duration with multimodal shape. Therefore, the PDFs can be represented as the summation of several normal distributions $\mathcal{N}(d | \mu_i, \sigma_i)$, expressed by,

$$f(d | \theta_s^{x,t}) = \sum_{i=1}^K \phi_i \mathcal{N}(d | \mu_i(x, t, s), \sigma_i(x, t, s)) \quad (21)$$

$$= \sum_{i=1}^K \phi_i \mathcal{N}(d | \varphi_{i,s}^{x,t}) \quad (22)$$

Algorithm 5 Improved Event Model

Input: Individuals characteristics: x_1, x_2, \dots, x_M
Output: Discrete state sequence: z_1, z_2, \dots, z_N

```

1 begin
2   sampling  $z_1 [\pi_0, \pi_1]$ 
3   initialise  $s = z_1$ 
4   initialise  $k = 1$ 
5   while  $k \leq N$  do
6     update  $i = \min_i \left( \sum_{i'=1}^K p_{i'} \right) \geq r$ 
7     sampling  $\Delta k \sim \phi_i \mathcal{N}(d | \varphi_{i,s}^{x,t_k})$ 
8     update  $z_{[k:k+\Delta k]} = s$ 
9     update  $s = s \oplus 1$ 
10    update  $k = k + \Delta k + 1$ 
11  end
12 end

```

where K is the number of clusters, estimated by the silhiutte method,⁴ $\varphi_{i,s}^{x,t}$ represents the mean $\mu_i(x, t, s)$ and the variance $\sigma_i(x, t, s)$ of the i^{th} component, and ϕ_i presents the mixture component weight, which must satisfy the condition $\sum_{i=1}^K \phi_i = 1$.

Additionally, as a part of the simulation process (see Algorithm 5), it is necessary to determine the probability that an individual belongs to a given cluster. This probability can be estimated from (22) through,

$$p_i = \frac{\phi_i \mathcal{N}(d | \varphi_{i,s}^{x,t})}{\sum_{i=1}^K \phi_i \mathcal{N}(d | \varphi_{i,s}^{x,t})}, \quad \text{for } i \in [1, K] \quad (23)$$

where p_i is a vector of dimension K in which each element indicates the probability that an individual belongs to a cluster.

As illustrated in Algorithm 5, once a transition occurs at time-step k , each individual is assigned to one of the clusters, defined by the GMM. In order to do this, the vector p_i is compared with the uniformly distributed random variable $r \in [0, 1]$ (row 6). Subsequently, according to the duration probability distribution ($\phi_i \mathcal{N}(d | \varphi_{i,s}^{x,t})$), associated to the selected cluster, the duration Δk of the new state is estimated (row 7). Afterwards, the current occupancy state is maintained from k to $k + \Delta k$ (row 8). Once this simulation step is reached, the occupancy state changes and the simulation step k is updated.

A. WHOLE POPULATION

To validate the proposed approach, a comparison is initially made by evaluating the models performance in estimating the overall presence probability. The results, presented in

⁴The silhiutte method [49] allows to identify the appropriate number of distributions (clusters) in a clustering analysis.

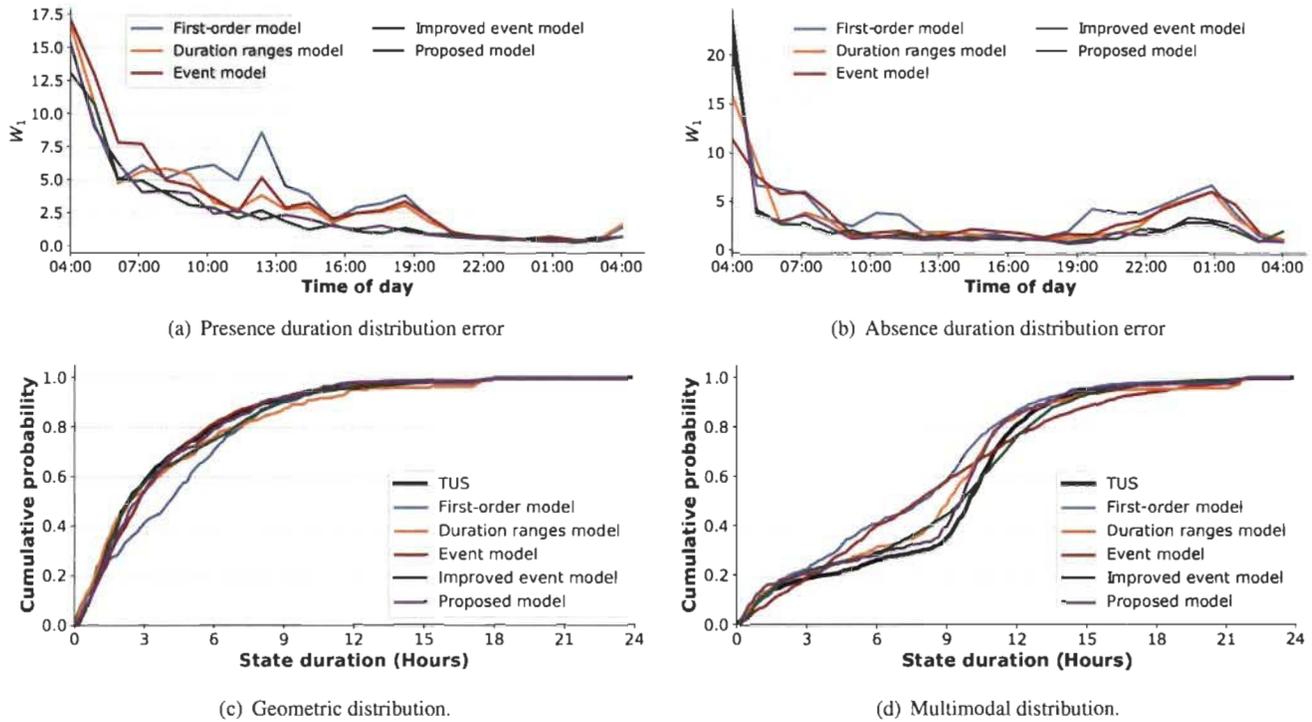


FIGURE 7. Comparison of duration distribution representation.

TABLE 3. Summary of the performance comparison of whole population occupancy predictions by the utilized models.

	MAE	W_1
First order model	3.07 ± 0.99	4.03/4.11
Duration ranges model	3.24 ± 0.59	3.39/3.44
Event model	3.30 ± 0.78	3.68/3.71
Improved event model	1.60 ± 0.94	2.47/2.78
Proposed model	1.44 ± 0.71	2.64/2.72

W_1 values corresponding to presence/absence states.

TABLE 3, show that among the baseline approaches, the first-order model has the best accuracy in estimating the presence probability of individuals. However, the duration range and event models can improve the first-order model performance in explaining the duration distribution of the occupancy states by up to 20%.

TABLE 3 indicates that the proposed model can outperform other approaches and reduce the error in estimating both the presence probability and the duration distribution by up to 56% and 34%, respectively. Furthermore, it can be observed that the recommended improvement to the event model, presented by Wilke *et al.* [33], has yielded a reduction of up to 50% in the MAE. This reduction has resulted in a performance close to that achieved by the proposed explicit duration model.

Moreover, FIGURE 7 presents a comparison of the duration distribution errors according to the time of day. FIGURE 7(a) and FIGURE 7(b) show that compared to other techniques, the proposed model and the improved event model have achieved better performances in both presence and absence duration probability estimation. This proves

the capability of these methods to efficiently evaluate events duration even in cases with multimodal distribution (FIGURE 7(d)). It should be noted that during different times of day, the performance of all models is similar. In fact, this similarity occurs when distributions have an exponential shape (FIGURE 7(c)).

B. SUB-POPULATION

Moreover, the models performance has been compared to realize their ability to explain the behavior of different individuals' sub-populations. The results of this comparison are presented in TABLE 4. It can be observed that the proposed approach and the improved event model result in a more accurate representation of sub-populations' occupancy patterns. Particularly, the proposed approach can accomplish an improvement of up to 55% and 29% in representing both the presence probability and the duration distribution of sub-populations, respectively.

FIGURE 8 provides an example that shows the capability of the model to represent the variability of occupancy profiles among different sub-populations. For this purpose, four groups of individuals with significantly different occupancy patterns have been selected.

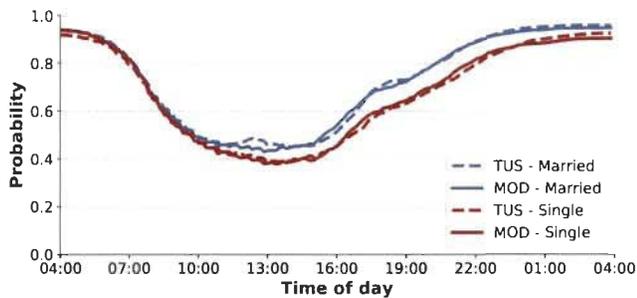
VI. DISCUSSION AND FUTURE PROSPECTS

A new probabilistic approach to modeling household occupancy has been proposed. It has been validated through a comparison with other techniques, reported in the literature. The results, obtained in Section V, demonstrate that the hazard-based model provides an appropriate methodology for modeling household occupancy. The comparative

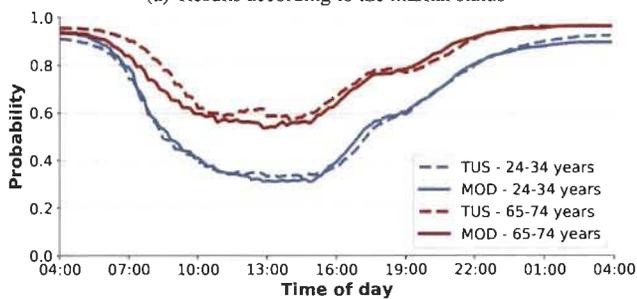
TABLE 4. Performance comparison in sub-populations' occupancy prediction.

Population	First order model		Duration ranges model		Event model		Improved event model		Proposed model	
	MEA	W_1	MAE	W_1	MAE	W_1	MAE	W_1	MAE	W_1
Workers	3.85 ± 1.02	5.06/4.78	4.44 ± 1.40	5.02/4.17	4.28 ± 0.96	5.07/4.64	2.42 ± 1.08	4.08/3.87	2.76 ± 1.32	4.19/4.07
Retirees	3.90 ± 1.58	7.05/7.14	5.18 ± 1.60	6.63/6.39	3.98 ± 2.14	6.53/5.26	2.67 ± 1.48	6.05/5.51	3.49 ± 1.16	6.22/6.14
Married	3.46 ± 1.20	5.35/5.52	3.75 ± 1.10	5.12/5.16	3.79 ± 1.12	5.11/5.00	2.23 ± 0.74	4.36/4.35	1.80 ± 0.82	4.29/4.27
Single	3.52 ± 1.74	6.92/6.25	3.31 ± 1.38	7.19/5.74	3.71 ± 1.28	7.21/5.35	2.81 ± 1.44	6.72/5.44	2.73 ± 1.56	6.81/5.49
Men	3.54 ± 1.02	5.22/5.28	3.41 ± 1.44	4.81/4.40	3.52 ± 0.94	5.07/4.89	2.09 ± 1.06	4.63/4.09	1.87 ± 0.56	4.95/4.39
Women	3.30 ± 1.14	5.40/5.15	3.62 ± 1.00	4.96/4.91	3.47 ± 0.58	5.40/4.42	2.22 ± 0.96	4.27/4.55	1.82 ± 0.90	4.82/4.15
Weekday	3.59 ± 1.28	4.85/5.16	3.70 ± 0.66	4.57/4.51	4.04 ± 0.88	4.58/4.88	1.91 ± 1.10	3.71/3.68	1.60 ± 1.16	3.59/3.68
Weekend	3.06 ± 1.22	6.69/6.07	3.46 ± 1.62	6.67/5.88	3.22 ± 1.36	6.32/5.45	2.08 ± 0.82	6.22/5.42	2.91 ± 1.34	6.33/5.52

W_1 values correspond to the results of presence/absence states.



(a) Results according to the marital status



(b) Results according to the age range

FIGURE 8. Simulation results of the occupancy patterns of some individuals sub-populations.

study validates that the proposed method is more accurate in examining the duration distribution of occupancy states. Furthermore, it has higher efficiency in representing occupancy variability among sub-populations. It should be noted that the cross-validation analysis has been also employed since it can help reduce the overfitting risk and preserve the model generalization capability.

Additionally, the event model proposed by Wilke et al. [33] has been improved by adding a GMM. In fact, the inclusion of this clustering method is advantageous since it has improved the MAE performance by up to 50% with respect to the initial model. Moreover, the improved event model has yielded a performance comparable to the proposed explicit duration approach.

Besides, the results of the parametric analysis, performed in Section IV, show that the parameter selection for model calibration is important to achieve an adequate representation of the occupancy variability. It should be mentioned that some of the selected parameters such as age, employment status (main activity), dwelling type, income, and weekdays

(TABLE 2) are consistent with those mentioned in previous studies [33], [34], [39]. Indeed, the utilized feature selection procedure can reduce the model complexity and result in a baseline for other researchers. This baseline can help i) formulate the most relevant questions to conduct a survey and ii) select the most appropriate variables for occupancy model calibration. Nevertheless, it should be noted that the parameters of interest can change according to the application context and the accuracy requirements.

Moreover, the robustness of the model is another topic to study in order to provide a more rigorous and detailed performance analysis. For this goal, the consistency of the model outcomes under significant variations of input data can be examined. The Monte Carlo simulation method can be used to perform such practice. Furthermore, the boundary of the prediction variance can be explored since input variables, explaining human behavior, have stochastic nature. Estimating the error of every single occupancy profile can be another concern to study. This is due to the fact that performance indicators usually explain the model ability to present the average behavior of a population. However, the primary output of occupancy models (including the one presented in this paper) is a time-series profile per individual. According to this issue, the consistency of the generated profiles considering the original ones can be also evaluated.

The number of the states of the proposed model can vary from two to N . This is beneficial to applications such as occupant activities prediction as an exercise that can significantly influence the performance of bottom-up models for electrical energy consumption and generation. Our approach can be also used to model household members interactions. For this purpose, the development of a multi-agent system with a leader has been advised in literature [31]. In such a system, the leader is an independent agent who conditions other agents' behavior according to their current state. On the other side, the integration of holidays, vacations, and seasons into the occupancy model is another aspect that requires significant consideration in order to generate annual profiles.

Moreover, the analysis of prolonged absences and interactions between occupants should be considered, albeit limitations related to surveys data. The model adaptability is another concern with regard to this type of data. Although TUS are useful methods for explaining occupants behavior,

the models, offered by them, are not adoptable due to the static nature of the data. Accordingly, the use of in-situ measurements based on the fusion of environmental sensors has been promoted as an efficient and cost-effective alternative to model real-life occupants behavior [10]. This assists with analyzing changes in occupants' habits and lifestyles as well as the influence of external factors such as weather. Furthermore, it can be used to evaluate the impact of behavioral information on actual energy management applications. Nevertheless, a future research direction can use surveys information as prior knowledge to speed up the convergence of algorithms and improve their accuracy.

VII. CONCLUSION

This paper presents an explicit-duration probabilistic model to predict individuals' occupancy profiles in their dwellings. The purpose of this work is to provide an approach that allows a more accurate representation of residential occupancy, which is significantly important to enhance the performance, design, and simulation of buildings. In this study, the influence of socio-demographic characteristics on model performance has been analyzed through a parametric method. As a result, the most significant individuals' characteristics has been identified that allows for using a reduced set of variables to develop the model. In this way, it is possible to reduce the complexity of the model without affecting its accuracy in predicting occupancy profiles. Besides, the effectiveness of the proposed model has been thoroughly examined within a comparative study. This study demonstrates that the recommended approach is able to reduce the error in describing both the presence probability and the duration distribution by up to 56% and 34%, respectively. The utilized method and the parametric analysis are applicable to other TUS and indoor activities prediction exercises. Furthermore, an event model, studied in the literature, has been improved in this study by using GMM. The suggested improvement has reduced the MAE of this model by up to 50% and made it competitive with the proposed explicit-duration probabilistic model. Literature has highlighted occupancy and occupant behavior as crucial factors in improving building energy performance. Indeed, several works have demonstrated that a significant amount of energy can be saved by adopting occupancy and activity-based control strategies. Therefore, a comprehensive analysis of residential energy-saving potentials based on these control strategies can bring about interesting subjects for future research. In this regard, providing realistic occupancy profiles becomes a primary concern to analyze the behavior diversity among people. The explicit-duration probabilistic model, presented in this article, is favored as an efficient method to address this issue.

ACKNOWLEDGMENT

The authors would like to thank the Laboratoire des technologies de l'énergie d'Hydro-Québec, the Natural Science and Engineering Research Council of Canada, the Foundation of

Université du Québec à Trois-Rivières, the Mitacs accelerate program, and Dr. Sayed Saeed Hosseini.

REFERENCES

- [1] S. Aman, Y. Simmhan, and V. K. Prasanna, "Energy management systems: State of the art and emerging trends," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 114–119, Jan. 2013.
- [2] Y. Zhang, X. Bai, F. P. Mills, and J. C. V. Pezzey, "Rethinking the role of occupant behavior in building energy performance: A review," *Energy Buildings*, vol. 172, pp. 279–294, Aug. 2018.
- [3] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, and X. Feng, "IEA EBC annex 66: Definition and simulation of occupant behavior in buildings," *Energy Buildings*, vol. 156, pp. 258–270, Dec. 2017.
- [4] H. Yoshino, T. Hong, and N. Nord, "IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods," *Energy Buildings*, vol. 152, pp. 124–136, Oct. 2017.
- [5] B. F. Balvedi, E. Ghisi, and R. Lamberts, "A review of occupant behaviour in residential buildings," *Energy Buildings*, vol. 174, pp. 495–505, Sep. 2018.
- [6] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, "The impact of occupants' behaviours on building energy analysis: A research review," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1061–1071, Dec. 2017.
- [7] B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng, and H. Fontenot, "Modeling occupancy and behavior for better building design and operation—A critical review," *Building Simul.*, vol. 11, no. 5, pp. 899–921, Oct. 2018.
- [8] F. Amara, K. Agbossou, Y. Dubé, S. Kelouwani, A. Cardenas, and S. S. Hosseini, "A residual load modeling approach for household short-term load forecasting application," *Energy Buildings*, vol. 187, pp. 132–143, Mar. 2019.
- [9] S. Sansregret, K. Lavigne, B. Le Lostec, L. Francois, and F. Guay, "High resolution bottom-up residential electrical model for distribution networks planning," in *Proc. 16th Conf. IBPSA*, 2019, pp. 3540–3547.
- [10] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," *Building Environ.*, vol. 180, Aug. 2020, Art. no. 106966.
- [11] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy Buildings*, vol. 56, pp. 244–257, Jan. 2013.
- [12] I. Georgievski, T. A. Nguyen, F. Nizamic, B. Setz, A. Lazovik, and M. Aiello, "Planning meets activity recognition: Service coordination for intelligent buildings," *Pervas. Mobile Comput.*, vol. 38, pp. 110–139, Jul. 2017.
- [13] M. Milenkovic and O. Amft, "An opportunistic activity-sensing approach to save energy in office buildings," in *Proc. 4th Int. Conf. Future Energy Syst.*, 2013, pp. 247–258.
- [14] J. Scott, A. J. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar, "PreHeat: Controlling home heating using occupancy prediction," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, 2011, pp. 281–290.
- [15] P. J. Boait and R. M. Rylatt, "A method for fully automatic operation of domestic heating," *Energy Buildings*, vol. 42, no. 1, pp. 11–16, Jan. 2010.
- [16] C. D. Korkas, S. Baldi, I. Michailidis, and E. B. Kosmatopoulos, "Intelligent energy and thermal comfort management in grid-connected microgrids with heterogeneous occupancy schedule," *Appl. Energy*, vol. 149, pp. 194–203, Jul. 2015.
- [17] C. D. Korkas, S. Baldi, I. Michailidis, and E. B. Kosmatopoulos, "Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage," *Appl. Energy*, vol. 163, pp. 93–104, Feb. 2016.
- [18] S. Baldi, C. D. Korkas, M. Lv, and E. B. Kosmatopoulos, "Automating occupant-building interaction via smart zoning of thermostatic loads: A switched self-tuning approach," *Appl. Energy*, vol. 231, pp. 1246–1258, Dec. 2018.
- [19] S. Srinivasan and C. R. Bhat, "Modeling household interactions in daily in-home and out-of-home maintenance activity participation," *Transportation*, vol. 32, no. 5, pp. 523–544, Sep. 2005.
- [20] J. Torriti, "A review of time use models of residential electricity demand," *Renew. Sustain. Energy Rev.*, vol. 37, pp. 265–272, Sep. 2014.
- [21] M. J. Roorda, E. J. Miller, and K. M. N. Habib, "Validation of TASHA: A 24-H activity scheduling microsimulation model," *Transp. Res. A. Policy Pract.*, vol. 42, no. 2, pp. 360–375, Feb. 2008.
- [22] G. Buttitta, O. Neu, W. J. Turner, and D. Finn, "Modelling household occupancy profiles using data mining clustering techniques on time use data," in *Proc. Building Simulation*, San Francisco, CA, USA, Aug. 2017, pp. 1788–1797.
- [23] C. F. Walker and J. L. Pokoski, "Residential load shape modelling based on customer behavior," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 7, pp. 1703–1711, Jul. 1985.

- [24] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations." *Energy Buildings*, vol. 40, no. 8, pp. 1560–1566, Jan. 2008.
- [25] Office for National Statistics and Ipsos-RSL. (2003). *United kingdom time use survey, 2000*. [Online]. Available: <http://doi.org/10.5255/UKDA-SN-4504-1>
- [26] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model." *Energy Buildings*, vol. 42, no. 10, pp. 1878–1887, Oct. 2010.
- [27] J. Widén, A. M. Nilsson, and E. Wäckelgård, "A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand." *Energy Buildings*, vol. 41, no. 10, pp. 1001–1012, Oct. 2009.
- [28] J. Widén and E. Wäckelgård, "A high-resolution stochastic model of domestic activity patterns and electricity demand." *Appl. Energy*, vol. 87, no. 6, pp. 1880–1892, Jun. 2010.
- [29] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, "A highly resolved modeling technique to simulate residential power demand." *Appl. Energy*, vol. 107, pp. 465–473, Jul. 2013.
- [30] A. J. Collin, G. Tsagarakis, A. E. Kiprakis, and S. McLaughlin, "Development of low-voltage load models for the residential load sector." *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2180–2188, Sep. 2014.
- [31] M. L. Baptista, H. Prendinger, R. Prada, and Y. Yamaguchi, "A cooperative multi-agent system to accurately estimate residential energy demand." in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, 2014, pp. 1405–1406.
- [32] J. Conlisk, "Interactive Markov chains." *J. Math. Sociol.*, vol. 4, no. 2, pp. 157–185, 1976.
- [33] U. Wilke, F. Haldi, J.-L. Scartezzini, and D. Robinson, "A bottom-up stochastic model to predict building occupants' time-dependent activities." *Building Environ.*, vol. 60, pp. 254–264, Feb. 2013.
- [34] G. Flett and N. Kelly, "An occupant-differentiated, higher-order Markov chain method for prediction of domestic occupancy." *Energy Buildings*, vol. 125, pp. 219–230, Aug. 2016.
- [35] J. Tanimoto, A. Hagishima, and H. Sagara, "A methodology for peak energy requirement considering actual variation of occupants' behavior schedules." *Building Environ.*, vol. 43, no. 4, pp. 610–619, Apr. 2008.
- [36] U. Wilke, "Probabilistic bottom-up modelling of occupancy and activities to predict electricity demand in residential buildings," Ph.D. dissertation, École Polytechnique Fédérale De Lausanne, Lausanne, Switzerland, 2013.
- [37] E. Vorger, "Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment," Ph.D. dissertation, l'École nationale supérieure des mines de Paris, Paris, France, 2014.
- [38] G. Flett, "Modelling and analysis of energy demand variation and uncertainty in small-scale domestic energy systems," Ph.D. dissertation, Dept. Mech. Aerosp. Eng., Univ. Strathclyde, Scotland, U.K., 2017.
- [39] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, and F. Descamps, "A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison." *Building Environ.*, vol. 75, pp. 67–78, May 2014.
- [40] D. Aerts, "Occupancy and activity modelling for building energy demand simulations, comparative feedback and residential electricity demand characterisation," Ph.D. dissertation, Vrije Univ. Brussel, Brussels, Belgium, 2015.
- [41] H. Hou, J. Pawlak, A. Sivakumar, B. Howard, and J. Polak, "An approach for building occupancy modelling considering the urban context." *Building Environ.*, vol. 183, Oct. 2020, Art. no. 107126.
- [42] F. Haldi and D. Robinson, "Interactions with window openings by office occupants." *Building Environ.*, vol. 44, no. 12, pp. 2378–2395, Dec. 2009.
- [43] S. P. Jenkins, "Survival analysis," Unpublished manuscript, Inst. for Social Econ. Res., Univ. Essex, Colchester, U.K., Tech. Rep., 2005, pp. 54–56.
- [44] S. Canada. *General Social Survey—Time Use (GSS)* Accessed: May 1, 2020. [Online]. Available: <https://www.statcan.gc.ca>
- [45] R. J. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, vol. 793. Hoboken, NJ, USA: Wiley, 2019.
- [46] M. Dewar, C. Wiggins, and F. Wood, "Inference in hidden Markov models with explicit state duration distributions." *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 235–238, Apr. 2012.
- [47] V. Barbu and N. Limnios, "Empirical estimation for discrete-time semi-Markov processes with applications in reliability." *J. Nonparametric Statist.*, vol. 18, nos. 7–8, pp. 483–498, Oct. 2006.
- [48] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, vol. 1. Hoboken, NJ, USA: Wiley, 2004.
- [49] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.



LUIS RUEDA received the B.S. degree in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2013, and the master's degree in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2017. He is currently pursuing the Ph.D. degree in electrical engineering with the Smart Energy Research and Innovation Laboratory, Université du Québec à Trois-Rivières, QC, Canada. His research interests include residential energy management systems, artificial intelligence applications for smart grids, embedded systems, renewable energies, and occupant behavior modeling in residential buildings.



SIMON SANSREGRET received the B.A.Sc. and M.A.Sc. degrees in mechanical engineering with specialization in energy from the University of Sherbrooke. He is a member of the Ordre des ingénieurs du Québec. He has been a Researcher with the laboratoire des technologies de l'énergie (LTE), Hydro-Québec Research Institute, since 2001. His expertise is related to energy efficiency and demand response in building sector. In recent years, he has been devoted to the development of simulation tools in order to improve the energy efficiency of commercial and institutional buildings. He was responsible for the development of simulation software called SIMEB, an interface to EnergyPlus Simulation Engine. He has published several scientific papers in connection with the building energy simulation, model calibration, and visualization of building performance data. He also contributed to various projects related to energy consumption and demand response in the residential sector. He was on the board of directors of the Canadian Chapter of International Building Performance Association (IBPSA-Canada) from 2010 to 2016.



BRICE LE LOSTEC received the M.S. degree in science and technologies from the Université de Savoie, France, in 2005, and the Ph.D. degree in mechanical engineering from Sherbrooke University, Canada, in 2010. His research interests include building stock modeling, solar energy, power generation from low grade heat, electrically driven compression heat pump, absorption heat pump, and refrigeration.



KODJO AGBOSSOU (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic measurements from the Université de Nancy I, France, in 1987, 1989, and 1992, respectively. He is currently the Hydro-Québec Research Chairholder on Transactive Management of Power and Energy in the Residential Sector, and the Chair of the Smart Energy Research and Innovation Laboratory of Université du Québec à Trois-Rivières (UQTR). He was the Head of Engineering School, UQTR, from 2011 to 2017. He was the Head of the Department of Electrical and Computer Engineering Department, UQTR, from 2007 to 2011. He was also the Director of Graduate Studies in Electrical Engineering, UQTR, from 2002 to 2004. He was a Postdoctoral Researcher (1993–1994) with the Electrical Engineering Department, UQTR, and was a Lecturer (1997–1998) at the same department. He is the author of more than 325 publications and has four patents and two Patent Pending. His present research activities are in the areas of renewable energy, the use of hydrogen, Home demand side management (HDSM), integration of energy production, storage and electrical energy generation system, connection of electrical vehicle to the grid, control and measurements. He is a member of the Hydrogen Research Institute and Research group "GREI" of UQTR. Since 2015, he has been the Sub-Committee Chair on Home and Building Energy Management of Smart Grid Technical Committee." IEEE Industrial Electronics Society (IES).



NILSON HENAO received the B.S. degree in electronics engineering from the Universidad de los Llanos, Villavicencio, Colombia, in 2010, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Quebec at Trois Rivières (UQTR), Trois-Rivières, QC, Canada, in 2013 and 2018, respectively. His research interests include statistical and machine learning methods with applications to energy management in the residential sector, distributed optimization, multi-

agent control, smart grid technologies, intelligent energy planning, smart energy storage, and load monitoring.



SOUSSO KELOUWANI (Senior Member, IEEE) received the Ph.D. degree in robotics systems from Ecole Polytechnique de Montreal, in 2011, and completed a postdoctoral internship on fuel cell hybrid electric vehicles at the Université du Québec à Trois-Rivières (UQTR), in 2012.

He is currently a Full Professor of Mechatronics with the Department of Mechanical Engineering since 2017 and a member of the Hydrogen Research Institute. He holds four patents in USA and Canada, in addition to having published more than 100 scientific articles. He is the Holder of the Canada Research Chair in Energy Optimization of Intelligent Transport Systems and the Holder of the Noovelia Research Chair in Intelligent Navigation of Autonomous Industrial Vehicles. He developed expertise in the optimization and the intelligent control of vehicular applications. In 2019, his team received the 1st Innovation Prize in partnership with DIVEL, awarded by the Association des Manufacturiers de la Mauricie et Center-du-Québec for the development of an autonomous and natural navigation system. In 2017, he received the Environment Prize at the Gala des Grands Prix d'excellence en transport from the Association québécoise du Transport (AQTr) for the development of hydrogen range extenders for electric vehicles. He was the Co-President and President of the technical committee of the IEEE International Conferences on Vehicular Power and Propulsion in Chicago (USA, 2018) and in Hanoi (Vietnam, 2019). His research interests focus on optimizing energy systems for vehicle applications, advanced driver assistance techniques, and intelligent vehicle navigation taking into account Canadian climatic conditions. He is a member of the Order of Engineers of Quebec. He is the Winner of the Canada General Governor Gold Medal, in 2003.

3.3 Synthèse du chapitre

L'objectif de développer une approche probabiliste permettant la modélisation des profils individuels d'occupation dans le contexte résidentiel a été atteint à travers la proposition présentée dans ce chapitre. Nous avons proposé un modèle probabiliste à deux états qui combine les modèles de Markov traditionnels avec la régression de Cox (modèle à risque proportionnel). L'approche proposée prend en compte la dynamique temporelle et la variabilité des profils de présence des individus à leur logement en utilisant une loi de transition conditionnée explicitement à la durée des états d'occupation et aux caractéristiques sociodémographiques des individus. Cela permet de générer des profils d'occupation quotidiens réalistes. Toutefois, le nombre d'états du modèle peut être augmenté, permettant son application pour la génération des profils d'activité. Ceux-ci peuvent être utilisés dans des modèles « bottom-up » pour simuler la consommation d'énergie d'un parc résidentiel, pour analyser la performance énergétique des bâtiments ou pour prédire la demande.

La méthode proposée a été évaluée par une comparaison de la performance par rapport à d'autres approches déjà présentées dans la littérature. Les résultats mettent en évidence le potentiel de notre approche, en démontrant que pour le cas étudié l'erreur dans la représentation de la probabilité de présence et de la distribution de la durée pour différentes populations a été réduite par rapport aux méthodes de la littérature utilisée comme base de comparaison. Par ailleurs, l'approche proposée peut-être aussi appliquée pour d'autres analyses comme la modélisation des activités et l'utilisation d'appareils électroménagers ou d'autres charges domestiques. De plus, l'application de la méthode pour modéliser le comportement dans les bâtiments commerciaux et institutionnels est un sujet de recherche qui peut être exploré ultérieurement.

Chapitre 4 - Approche pour la gestion de la demande résidentielle basée sur la détection et la prévision de l'occupation

4.1 Introduction

Au cours des dernières années, les défis énergétiques et climatiques ont attiré l'attention du public et des milieux scientifiques. Particulièrement, dans le secteur énergétique, les bâtiments étaient responsables de plus de 40% de la consommation mondiale d'énergie et de 33% des émissions mondiales de gaz à effet de serre en 2016 [63]. Par conséquent, des politiques visant à combler la demande croissante d'énergie et à atteindre les objectifs de réduction de l'empreinte écologique sont nécessaires. À cet égard, la gestion intelligente de la demande d'électricité dans le secteur résidentiel, l'amélioration de l'efficacité des systèmes énergétiques actuels et l'intégration des sources d'énergie renouvelables sont autant d'avenues possibles pour atteindre ces objectifs [5].

Conformément à ces tendances en matière de gestion de l'énergie et du réseau intelligent, à partir de l'hiver 2019-2020, Hydro-Québec a lancé l'option de crédit hivernal pour la clientèle au tarif D et le tarif Flex D [144]. Ces offres tarifaires permettent aux clients inscrits de réaliser des économies en réduisant leur consommation lors de période de forte demande d'électricité. À jour, environ 61 000 clients sont inscrits dans cette stratégie de tarification dynamique, permettant de réduire la demande d'électricité de 65 MW pendant les périodes de pointe de l'hiver 2020-2021 [145]. Cela représente environ la puissance totale générée par la centrale au fil de l'eau Rapide-2 sur la rivière des Outaouais.

Compte tenu de ce qui précède, il est évident que pour la mise en œuvre efficace de méthodes de gestion de la demande résidentielle, la participation des clients est essentielle.

Cependant, le manque d'information, l'engagement des consommateurs et leur comportement énergétique font partie du débat dans le domaine de l'énergie et des sciences sociales comme autant des limites lors de la mise en œuvre des actions visant l'économie d'électricité [146], [147]. Par conséquent, l'utilisation des méthodes d'apprentissage automatique pour modéliser les habitudes et les préférences des individus tout en minimisant l'intervention directe des utilisateurs lors de la prise de décision et de l'application des stratégies de gestion s'avère nécessaire. Dans cette perspective, les sections 4.2 et 4.3 présentent sur forme d'articles scientifiques, une étude approfondie des technologies et des méthodes utilisées pour la détection de l'occupation, ainsi que le développement d'une approche de gestion locale de la demande basée sur la détection et la prévision de la présence des individus respectivement. Par ailleurs, nous décrivons le développement d'une architecture permettant de valider expérimentalement les méthodes proposées.

4.2 Article 2 : A comprehensive review of approaches to building occupancy detection

Journal : Building and Environment (Elsevier) (Volume : 180)

Auteurs : Luis Rueda, Kodjo Agbossou, Alben cardenas, Nilson Henao and Sousso Kelouwani

Date de publication : 12 juin 2020

DOI : <https://doi.org/10.1016/j.buildenv.2020.106966>

4.2.1 Contexte

Comme présenté dans le chapitre précédent, les modèles basés sur des données d'enquête ont montré leur potentiel pour des applications dans des environnements simulés. Cependant, pour les applications dans des environnements réels, les mesures *in situ* ont été mises en avant comme une solution pour acquérir des informations permettant de modéliser la présence et le comportement des individus. De ce fait, les modèles d'occupation, décrivant la présence, l'absence et le déplacement des occupants dans les espaces fermés ont attiré l'attention des chercheurs comme un moyen prometteur pour contribuer au développement de bâtiments plus efficaces du point de vue énergétique [87]. Par exemple, certains auteurs ont montré que la présence/absence des occupants peut être utilisée pour contrôler le fonctionnement des systèmes CVC dans les bâtiments, obtenant ainsi des économies d'énergie d'environ 30% dans

ces systèmes [49]. Par ailleurs, en raison de l'intérêt croissant suscité par le RÉIs, la domotique et l'IoT, de nombreuses technologies et méthodologies pour la détection de l'occupation ont été explorées dans des environnements résidentiels, commerciaux et institutionnels. Cependant, étant donné le grand nombre d'approches existantes, nous avons mis en évidence la nécessité de réaliser une étude approfondie de la littérature dans le domaine de la détection et la modélisation de l'occupation. Ainsi, nous avons identifié les avantages et les inconvénients des différents capteurs, les leçons apprises, les tendances, les opportunités et les défis dans ce domaine de recherche.

4.2.2 Méthodologie

Une revue de la littérature a été menée pour rassembler les informations relatives à la détection de l'occupation des bâtiments à partir de mesures sur place. Cette revue est consacrée particulièrement aux systèmes utilisant de technologies à faible coût et plus respectueuses de la vie privée des clients. La recherche a été effectuée en utilisant Scopus comme base de données principale et en limitant l'exploration aux publications scientifiques dans des journaux et des conférences en anglais depuis 1998. Pour cette étude, « détection de l'occupation », « estimation de l'occupation », « surveillance de l'occupation » et « énergie des bâtiments » ont été les principaux mots-clés. Par ailleurs, les articles sélectionnés ont été classés selon le type d'algorithme utilisé et une analyse détaillée de chaque article est effectuée en utilisant comme critères d'évaluation la précision, le type de bâtiment, le niveau de résolution sur l'occupation et le type de charges qui ont été contrôlés. De même, le potentiel d'économie d'énergie obtenu avec les systèmes de gestion centrés sur l'occupant a été analysé.

4.2.3 Résultats

Les résultats principaux de l'étude approfondie des méthodes et des technologies utilisées pour les applications de détection de l'occupation sont présentés ci-dessous :

- Une compréhension approfondie des approches utilisées pour les systèmes de détection de l'occupation utilisant des mesures *in situ*. Cette analyse nous a permis de classer les approches en trois catégories, les méthodes analytiques, les méthodes basées sur les données et les méthodes basées sur les connaissances.

- Nous avons identifié que les méthodes basées sur les données, principalement les techniques d'apprentissage automatique telles que les réseaux neuronaux artificiels (ANN), les machines à vecteurs de support (SVM), K-nearest neighbors (KNN) et Hidden Markov Models (HMM) sont parmi les approches les plus utilisées pour la détection d'occupation.
- Il est souligné que la plupart des approches sont basées sur des méthodes d'apprentissage supervisé. Ces méthodes nécessitent un ensemble de données comprenant la valeur réelle d'occupation. Cependant, la mesure explicite de l'occupation est difficile à effectuer en raison de contraintes telles que le respect de la vie privée, faisant difficile l'utilisation de ces algorithmes dans la pratique. De plus, en raison de la variabilité du comportement humain, il est nécessaire de disposer de modèles qui puissent être adaptés dans le temps et aux conditions particulières de chaque application sans avoir besoin de la valeur réelle d'occupation pour son apprentissage. De ce fait, l'apprentissage non supervisé est présenté comme une alternative pour aborder ce problème.
- Les capteurs CO₂ et les capteurs PIR sont mis en évidence comme étant les capteurs les plus utilisés. Cependant, des variables environnementales telles que la température, l'humidité et la lumière, ainsi que des capteurs de sons et de vibrations ont également été étudiés. En outre, l'utilisation de nouvelles technologies telles que le WiFi, le Bluetooth et les compteurs intelligents suscite une attention croissante en raison de leur grand potentiel pour ce type d'applications.
- Il a été présenté une analyse des approches de gestion centrées sur l'occupant rapportées dans la littérature étudiée, mettant en évidence qu'une économie d'énergie moyenne d'environ 30% est possible en contrôlant le système CVC en fonction de la présence des individus.
- Cette étude a permis d'identifier les tendances de la recherche, ainsi que de présenter certaines des limites, opportunités et barrières dans ce domaine de recherche.

4.3 Article 3 : Online unsupervised occupancy anticipation system applied to residential heat load management

Journal : IEEE Access

Auteurs : Luis Rueda, Kodjo Agbossou, Nilson Henao, Souso Kelouwani, Juan C. Oviedo-Cepeda, Brice Le Lostec, Simon Sansregret, Michael Fournier.

Date d'acceptation : 08 juillet 2021

4.3.1 Contexte

Comme mentionné précédemment dans cette thèse, dans le contexte québécois, le chauffage des espaces est un facteur déterminant dans la consommation électrique des logements. Par ailleurs, le comportement des occupants, ainsi que l'acceptation, l'engagement et l'adoption des solutions de DR par les clients influencent la performance et la flexibilité énergétique des bâtiments [22], [148]. Compte tenu du rôle déterminant que jouent les personnes dans la GDR, il a été mis en évidence que les approches de gestion de la demande doivent prendre en compte non seulement des facteurs tels que les conditions météorologiques, la disponibilité des sources d'énergie et les caractéristiques des bâtiments, mais aussi le confort, les besoins et les préférences particuliers des clients. Cependant, la littérature scientifique a montré la diversité qui existe dans les patrons d'occupation et les activités des utilisateurs, ainsi que la difficulté à mesurer explicitement la présence et le comportement humains [56], [149]. Pour cette raison, il est nécessaire de développer des méthodes permettant aux HEMS d'apprendre et de s'adapter automatiquement aux caractéristiques uniques de chaque résidence (p. ex. le climat, l'enveloppe, la taille et le type de ménage) et au comportement de ses occupants.

Par conséquent, les stratégies de contrôle basées sur les occupants sont devenues une alternative intéressante pour aborder ce problème. En particulier, comme évoquée à plusieurs reprises dans cette thèse, la modélisation de l'occupation a été mise en évidence comme un facteur indispensable pour l'application de stratégies de gestion de la demande. Notamment pour la gestion des systèmes de chauffage, où la dynamique thermique lente des bâtiments doit être prise en compte. De ce fait, il est important d'anticiper les périodes de présence/absence des individus, afin de maximiser leur confort thermique et optimiser la consommation d'énergie.

4.3.2 Méthodologie

Cette section est consacrée à la proposition d'un modèle pour détecter et anticiper des états d'occupation dans une résidence, ainsi qu'au développement d'une architecture de gestion de la demande résidentielle. Nous proposons une stratégie de gestion prédictive et adaptative permettant de répondre aux besoins du client en termes du confort thermique des occupants et à ceux du distributeur en termes de réduction de la puissance en fonction d'une tarification dynamique. Il faut noter que l'étude proposée est menée dans un logement n'ayant pas de système de ventilation et d'échangeur d'air, et les portes (intérieures et extérieures) et fenêtres sont fermées la plupart du temps. De plus, à des fins de validation, la valeur réelle d'occupation de chaque zone thermique de la maison est disponible avec une résolution de 15 min. Ces informations sont collectées à partir d'un formulaire qui est rempli quotidiennement par les occupants (Annexe A). L'ensemble de l'étude a été réalisé pendant l'hiver 2020-2021 dans un appartement situé à Trois-Rivières, Québec, Canada.

4.3.2.1 Hypothèses et considérations de la proposition

L'approche proposée pour la détection et prévision de la présence, ainsi que l'architecture de gestion développée, considèrent les hypothèses suivantes :

- Les bâtiments résidentiels utilisent un système de chauffage électrique équipé de thermostats électroniques communicants (par exemple via WiFi, ZigBee ou d'autres technologies) pour contrôler indépendamment la température de chacune des pièces.
- La puissance consommée par le système de chauffage et par les autres appareils domestiques sont des variables observables, obtenues à partir d'un système de mesurage qui communique avec le contrôleur.
- Le contrôleur connaît à l'avance le prix de l'énergie et les variables climatiques pour l'horizon de contrôle, ainsi que la puissance nominale de chaque plinthe chauffante.
- Pour générer les profils synthétiques utilisés pour la validation par simulation, il est considéré que les variables observées dépendent de l'état d'occupation de la pièce et suivent une distribution normale conditionnelle.

4.3.2.2 Étapes de l'approche proposée

L'approche proposée pour la gestion de la demande résidentielle basée sur la détection et la prévision de l'occupation est illustrée par la FIGURE 4-1 et comporte les étapes suivantes :

- Premièrement, un système de mesurage est mis en place pour collecter des informations sur l'environnement intérieur de la résidence, ainsi que sur la consommation électrique de différents appareils domestiques. De plus, des services web sont utilisés pour l'acquisition de variables climatiques.
- Deuxièmement, une approche probabiliste non supervisée pour la détection en ligne de l'occupation a été proposée. Cette approche est fondée sur les modèles de Markov cachés à durée explicite (Explicit Duration Hidden Markov Model (EDHMM), en anglais) et utilise des données mesurées sur places. Pour ce faire, des variables telles que le CO₂, l'humidité, le mouvement et la consommation d'éclairage ont été collectés. Il faut noter que la résolution de l'approche proposée est limitée à l'état d'occupation, de sorte que le nombre de personnes et l'identité des individus ne sont pas abordés par la méthode.
- Troisièmement, une approche pour la prévision de la présence des individus est présente. Pour cela, une stratégie basée sur les modèles de hasard et la simulation de Monte-Carlo est utilisée. De cette manière, il est estimé la probabilité de présence à un horizon de 24

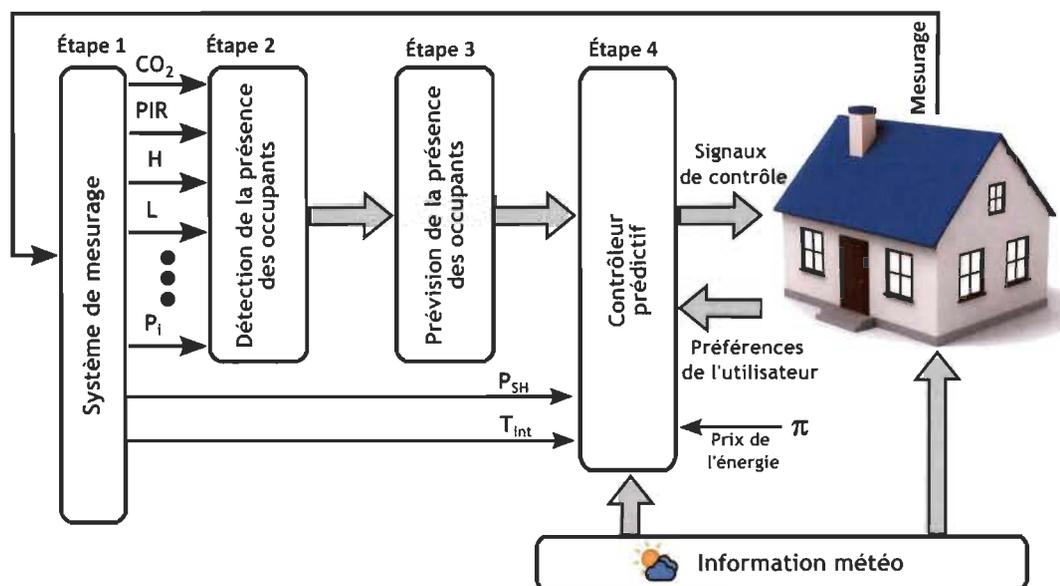


FIGURE 4-1 Structure de la stratégie de gestion centrée sur l'occupant.

heures, en tenant compte explicitement de la durée des états d'occupation et de l'heure de la journée.

- Finalement, une stratégie de contrôle prédictif à deux objectifs est appliquée. Pour ce faire, un objectif de coût dépendant d'une tarification dynamique et un objectif pour respecter le confort thermique, qui dépend de la présence des utilisateurs, sont formulés. De plus, un modèle linéaire est utilisé pour représenter la dynamique thermique de la maison (Annexe A). Les analyses sont effectuées dans un contexte sans sources d'énergie alternatives ni systèmes de stockage d'énergie, par conséquent les utilisateurs de type « producteur/consommateur » ne font pas partie du cadre de cette thèse. Toutefois, dans un scénario comportant des systèmes de production d'énergie, l'approche de gestion proposée peut être élargie pour prendre en considération ces apports énergétiques.

4.3.3 Résultats

À la suite de cette partie de la thèse, une méthode permettant de détecter et anticiper la présence des occupants, ainsi qu'une approche pour la gestion en ligne de la puissance pour les clients résidentiels avec chauffage par plinthe électrique sont proposées. Un sommaire des résultats est présenté ci-dessous :

- Nous avons présenté les développements d'une approche probabiliste à deux états pour la détection en ligne de l'occupation. L'approche proposée est non supervisée et intègre explicitement la dynamique temporelle des profils de présence/absence des individus. Les résultats de simulation montrent la consistance de la méthode proposée face aux variations de ses paramètres d'entrée. De plus, le potentiel de notre approche pour les applications de détection de l'occupation est mis en évidence, démontrant pour l'étude de cas analysée une performance supérieure aux méthodes de référence issues de la littérature.
- Nous avons proposé une stratégie de gestion centrée sur l'occupant basée sur des méthodes d'apprentissage automatique, pour réaliser la gestion du chauffage électrique de chacune des zones de la maison. Pour ce faire, une approche pour la prévision de la présence à un horizon de 24 heures a été utilisée comme mécanisme pour anticiper les

arrivées/départs des individus, permettant par exemple de préchauffer une pièce si le système prévoit que les occupants vont partir prochainement. Les résultats expérimentaux obtenus pour l'étude de cas analysée montrent le potentiel de l'approche proposée à atteindre un compromis entre la demande d'énergie et le confort thermique des clients, permettant également de réduire le coût de la facture de 31.8%.

- Une infrastructure expérimentale a été mise en place pour la validation des techniques proposées. Ce système utilise des équipements commerciaux et des logiciels open source, tandis que son déploiement s'effectue à partir d'un système embarqué à faible coût. De plus, nous avons utilisés des protocoles de communication couramment employés dans les applications IoT, ce qui rend le système extensible et peut être facilement répliqué. Cette infrastructure a permis de réaliser une preuve de concept des approches proposées dans un appartement à Trois-Rivières, pendant environ trois mois de la saison hivernale 2020-2021.



A comprehensive review of approaches to building occupancy detection

Luis Rueda^{a,*}, Kodjo Agbossou^a, Alben Cardenas^a, Nilson Henao^a, Souso Kelouwani^b

^a Department of Electrical and Computer Engineering, Hydrogen Research Institute, Smart Energy Research and Innovation Laboratory, Université du Québec à Trois-Rivières, QC, Canada

^b Department of Mechanical Engineering, Hydrogen Research Institute, Smart Energy Research and Innovation Laboratory, Université du Québec à Trois-Rivières, QC, Canada

ARTICLE INFO

Keywords:

Occupancy detection
Analytical methods
Data-driven methods
Knowledge-based methods
Sensor

ABSTRACT

Detailed occupancy information in buildings is useful to improve the performance of energy management systems in order to enable energy consumption savings and maintain occupants' comfort. Different technologies employed to provide occupancy information account for high-precision devices such as optical and thermal cameras, and environmental or specialized sensors like carbon dioxide (CO₂) and passive infrared (PIR). Although the latter systems have lower accuracy, they have received significant interest due to their affordable and less-intrusive nature. Accordingly, various studies have been conducted to explore the various elements of these technologies. Nevertheless, the algorithmic aspect of the occupancy detection process has not been adequately taken into consideration. This paper presents an extensive review of the techniques that have been exploited to process the information provided by the sensors and carry out occupancy information detection. In this study, a complete set of comparison criteria, comprising the performance, the occupancy resolution, the type of sensors used, the type of buildings, and the energy saving potentials has been considered in order to perform an in-depth analysis of the occupancy detection systems. Through its examination, this paper elaborates significant remarks on occupancy detection algorithms in order to realize a method that is not only efficient in processing sensors' data but also effective in providing accurate occupancy information.

1. Introduction

Smart grid is the key enabler of revolutionary power grids with a flexible, efficient, durable, and safe infrastructure. The emerging trends in communication, sensing, and metering technologies under smart grid archetype can enable an optimal management of all the units constituting power systems. Among the power grid sectors, the building sector receives significant attention due to its highest share of electricity consumption and potential of energy saving. Several researchers have shown the importance of occupancy information to improve building energy performance and reduce building energy consumption [1,2]. Agarwal et al. [3] proposed a heating, ventilating, air-conditioning (HVAC) control strategy that turns ON or OFF the system based on the occupancy detection in offices. The simulation results indicated that by using the proposed approach, the HVAC energy consumption was reduced between 10% and 15%. Jin et al. [4,5] reported a saving potential of 55% of the energy consumed by the ventilation system, by using a control ventilation strategy based on occupancy information.

Leephakpreeda [6] proposed an occupancy-based lighting control, and showed that the energy consumption of the system can be reduced between 35% and 75%. Yokoishi et al. [7] showed that it was possible to save 3.5 h of lighting power per day in a campus room, by using a network of PIR and illumination sensors for occupancy detection. Scott et al. [8] used information collected by PIR and RFID sensors to predict home occupancy patterns for heating control. The authors analyzed three different control techniques (based on scheduled, always-on, and PreHeat algorithms), and consequently obtained an energy saving potential up to 35%. Pend et al. [9] proposed a strategy with learning capacity for the control of the cooling system of six offices. The results showed that 20.3% of energy could be reduced by using the proposed control strategy. In this regard, Building Energy Management Systems (BEMS) are promoted in order to facilitate the utilization of occupancy information to minimize the power grid stress and maintain users' comfort [10,11].

Occupancy is not the only factor that influences electricity consumption, there are physical factors including building characteristics,

* Corresponding author.

E-mail addresses: luis.rueda@uqtr.ca (L. Rueda), kodjo.agbossou@uqtr.ca (K. Agbossou), alben.cardenasgonzalez@uqtr.ca (A. Cardenas), nilson.henao@uqtr.ca (N. Henao), souso.kelouwani@uqtr.ca (S. Kelouwani).

equipment efficiency and weather conditions that can change the electricity consumption behavior of a building [12,13]. However, these factors cannot be easily handled or modified by humans during the building utilization. Actually, occupancy is an important part of human factors that describe the occupants' presence, their consumption patterns, and the indoor environmental conditions. The concept of "occupancy" is regarded as the primary level of occupant behavior modeling [14]. Occupancy information can account for three main features that are explained in different resolution levels regarding the targeted applications [15]:

- Temporal resolution: it indicates the frequency at which events occur (e.g., hours, minutes, and seconds).
- Spatial resolution: it expresses the building properties in terms of number of floors, rooms, and other relevant data.
- Occupancy resolution: it explains the presence/absence status, the number of people, as well as their identification and activity in a building's zone.

The technologies and methodologies used to estimate buildings' occupancy information have been explored in literature. However, the algorithmic aspect of the building occupancy detection has not been completely analyzed in these works. Zhang et al. [2], Jia et al. [14], Kjaergaard et al. [16] and Balvedi et al. [17] summarized and compared different techniques for the acquisition of occupancy information. Likewise, these reviews presented a general analysis of the methods used to model the presence and behavior of occupants, as well as their influence on the energy consumption of buildings. Yang et al. [18] presented the opportunities and challenges of the occupancy sensing systems and occupancy modeling methodologies applied to institutional buildings. Mane et al. [19] briefly discussed the research trends and gaps in occupancy sensing. Labeodan et al. [20] published a survey of the occupancy measurement systems that are used in office buildings. The authors make a classification of the occupancy detection systems, using their spatial-temporal properties as comparison criteria. Saha et al. [21] presented a brief description of popular algorithms used for the detection, counting and tracking of occupants in buildings. Chen et al. [22] and Sun et al. [23] conducted an analysis of occupancy estimation and detection systems based on the type of the utilizing sensor. Likewise, they compared the sensors in order to identify their advantages and limitations for occupancy detection.

As presented above, it can be comprehended that the review papers about occupancy detection and estimation systems have explored the elements such as sensor type, cost, privacy issues, performance, energy savings, limitations, and infrastructure, as the essential prerequisite for occupancy detection. However, the analysis of the type of occupancy detection methods and their influence on building occupancy detection performance has not been adequately into consideration. Accordingly, this paper presents a categorization of the techniques that have been exploited to estimate building occupancy information. Additionally, a complete set of comparison criteria, comprising the performance, the occupancy resolution, the types of sensors used, the type of building, and the energy saving potential has been employed in order to perform an in-depth analysis of the occupancy detection methods.

The rest of this study is organized as follows: Section 2 presents the search method that has been used for the selection of the papers, discussed in the review. Section 3 presents the classification and analysis of the occupancy detection algorithms presented in the state of the art. Section 4 describes the advantages and disadvantages of sensors and algorithms, and benchmarks the performance of occupancy detection systems. Section 5 describes the existing trends and presents future research directions to encourage more improvements to this field. Finally, the conclusions are presented in Section 6.

2. Search method

A literature review was conducted to gather information related to building occupancy detection. The search was performed by using Scopus as the main database and restricting the exploration to scientific publications in English journals and proceedings of conferences since 1998.

For this study, 'occupancy detection', 'occupancy estimation', 'occupancy monitoring', and 'building energy' were the primary keywords. The first search in Scopus yielded 1746 documents. In order to limit this large amount, another search was made throughout the existing database by using the secondary keywords.

Additionally, the challenges related to privacy, costs, and high computing resources of video cameras technology have been highlighted in the literature. Due to high accuracy, this technology is promoted for automated generation of ground truth occupancy data for performance validation of more privacy-friendly and cost-effective occupancy detection systems [24,25]. Accordingly, this paper provides an exclusive review based on occupancy detection systems that use technologies other than video cameras.

According to the selected articles, Fig. 1 demonstrates a growth in the publications over the last few years. It can be observed that the development of building occupancy detection systems has attracted the attention of researchers. Indeed, the occupancy information can assist with energy management applications in buildings.

3. Occupancy detection algorithms

The importance of energy management in smart grid trends presents building occupancy as an interesting matter for research studies with different perspectives. As a result, a variety of algorithms and techniques have been used for this purpose. According to the basic strategies, the problem of occupancy detection can be explored through analytical, data-driven and knowledge-based methods.

3.1. Analytical methods

Analytical techniques study the physical interactions between occupants and environment. These methods use interpretable and causal information of the relationship between sensing variables and human presence. Many studies have reported the use of analytical techniques and CO₂ sensors to detect and count occupant numbers. In these studies, the dynamic and steady-state application of the mass balance equation has been the main approach used to obtain occupancy information.

Cali et al. [26] and Wang et al. [27,28] used a dynamic algorithm (dC) to detect occupants' presence and identify the number of people. In Ref. [26], the occupancy detection was performed in three offices and two rooms in a house, showing an occupants' presence accuracy up to 95.8%, and a number of occupants identification with an accuracy of 80.6%. Likewise, Wang et al. [27,28] compared the dynamic algorithm with the steady-state algorithm (ST ST), and concluded that dC algorithm provide a fast and accurate estimation of occupancy. Moreover,

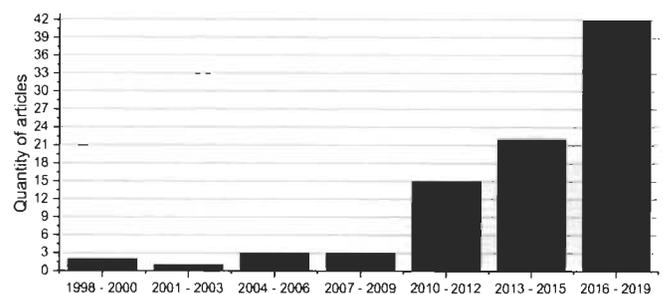


Fig. 1. Number of articles, published over time.

the authors evaluated the energy saving potential that can be obtained by using occupancy information to control an office ventilation system, reporting a reduction of between 8% and 13% depending on the season of the year.

Zuraimi et al. [29] present a comparison between dynamic and steady-state mass balance models. They used CO₂ data collected in a theater for its validation, and a root-mean-square error (RMSE) between 12.8 and 26.2 was obtained for these methods. Mumma [30] used the steady-state algorithm and transient equations to identify the real-time number of occupants in a classroom with a maximum of 45 students. The results of the occupancy count were accurate to ± 2 people. Sowa [31] compared the performance of steady state, dynamic and proportional (PROP) algorithms in on-line occupancy detection tasks. The authors used information collected from a 30-person conference room for their validation, and according to the results, the dynamic algorithm exhibited the best performance. Jin et al. [4,5] propose an occupancy detection system based on the sensing by proxy (SbP) methodology. This methodology is based on partial differential equations coupled with the ordinary differential equation presented in Ref. [32]. The authors used CO₂ data collected from a conference room for validation, and the results indicated that this approach achieves an RMSE between 0.5156 and 0.7331 with units of fractional people. In addition, the authors reported a saving potential of 55% of the energy consumed by the ventilation system, by using a control ventilation strategy based on occupancy information.

A summary of the occupancy detection works based on analytical methods is presented in Table 1.

3.2. Data-driven methods

Data-driven approaches are mainly investigated by means of machine learning methods that aim to detect hidden patterns in data without an explicit programming [33]. Several data-driven techniques have been used to detect the occupancy in buildings. Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) are the most popular methods.

3.2.1. Artificial neural networks

ANNs have been used by several studies for the detection of occupancy information in buildings. A combination of feed forward neural networks (FFNNs) and CO₂ data was used by Zuraimi et al. to estimate the number of occupants in a theater [29]. The results of that work showed that the ANN achieved an average accuracy of 70%. Likewise, a description of an environmental sensor network test-bed and its application to occupancy number detection in an office building was presented by Ref. [34,35]. In these papers, the authors used a neural network to detect the number of occupants, reporting an accuracy of 75%.

Kraipeerapun et al. [36] presented two methods for occupancy detection. The first method applied stacking and a multiclass neural network, whereas the second method combined stacking with a duo

output neural network for occupancy detection. The validation results showed an accuracy between 68.87% and 91.18%. Yang et al. [37,38] used the information collected with multiple sensors in combination with FFNNs to detect occupancy. The first of these works analyzed the effectiveness of each sensor in occupancy estimation, and compared the performance of a single hidden layer ANN and a multi hidden layer ANN. The test results reported a detection rate between 62.41% and 97.97%. In the second work, the authors used the ANN to detect presence and count occupant numbers in an office building, reporting a binary occupancy detection accuracy between 92.5% and 97.1%, and a RMSE between 0.139 and 0.758 for the occupancy number estimation.

Villariba [39] showed the implementation and validation of a feed forward neural network to infer the occupancy in a home based on electricity consumption data. The related information for three homes was collected and used for the experimental validation. A combination of time information and statistical measurements of electricity data was used as a feature set to train the ANN. According to the results, this approach was able to identify the residence occupancy state with an accuracy between 80.98% and 91.36%. Hobson et al. [25] used different combinations of sensors and ANNs to estimate occupancy number in an academic office building and achieved a mean coefficient of determination (R^2) of 83%. They showed that associated WiFi devices could be highly beneficial in the estimation of the number of occupants.

Rahman and Han [40] used time-delay neural networks (TDNNs) to estimate the number of occupants in an office room. This algorithm was tested and compared under different input conditions, yielding results with a coefficient of variation (CV) between 0.51% and 15.52%. Alam et al. [41] simulated indoor CO₂ data corresponding to various occupancy schedules and airflow patterns in order to investigate uncertainties that are caused by the parameters of TDNNs model. They concluded that accuracy estimation was significantly influenced by the frequency of occupancy variation rather than the fluctuation in airflow rate. Furthermore, Yang et al. [42] used a radial basis function (RBF) neural network to estimate the number of occupants in two laboratories. Field tests were carried out in two laboratory spaces with the results showing an overall detection rate between 63.23% and 88.74% when using a tolerance of one occupant.

Masood et al. [43] proposed the use of wrapper model feature selection and extreme learning machines (ELM) to examine the number of occupants in a tutorial room. The authors stated that their proposed method outperformed the filter methods by achieving an accuracy between 74.06% and 81.37%. A similar approach was also presented by Masood et al. in Ref. [44]. The authors performed a wrapper and hybrid feature selection method for occupancy estimation based on environmental parameters. Experimental data was collected in an office space, and the occupancy estimation results offered an accuracy between 75.63% and 79.17% for the WRANK-ELM algorithm and 76.88% and 77.92% for the RIG-ELM method. In another work, Masood et al. [45] proposed a filter-wrapper hybrid feature selection technique with ELM to estimate the range of occupants. In accordance with the experimental results, HFS-ELM algorithm carried out an accuracy between 94.37%

Table 1
Overview of the occupancy detection works that have applied analytical methods.

Algorithm	Sensors	Resolution	Building type	Performance
dC [26]	CO ₂	State* and quantity [†]	Residential and office	*72.3 – 95.8% [†] 45.8 – 80.6%
ST ST and dC [27,28]	CO ₂	Quantity	Office	Indicative graphs
ST ST and dC [29]	CO ₂	Quantity	Theater	12.8 < RMSE < 26.2
ST ST and transient equations [30]	CO ₂	Quantity	Classroom	± 2 people
ST ST, dC and PROP [31]	CO ₂	Quantity	Conference room	Indicative graphs
Sensing by proxy [4,5]	CO ₂	Quantity	Conference room	0.51 < RMSE < 0.73

dC: dynamic algorithm, ST ST: steady-state algorithm, PROP: proportional algorithm.

Table 2
Overview of the occupancy detection works that have applied artificial neural networks.

Type of NN	Sensors	Resolution	Building type	Performance
FFNN [29]	CO ₂	Quantity	Theater	70%
FFNN [34]	CO ₂ , L, T, CO, dew point, etc.	Quantity	Office	75%
FFNN [35]	CO ₂ , T, RH, motion, TVOC, etc.	Quantity	Office	75%
FFNN [36]	CO ₂ , L, T and H	State	Office	68.87 – 91.18%
FFNN [37]	CO ₂ , L, T, RH, sound, motion, etc.	Quantity	Laboratory	62.41 – 97.97%
FFNN [38]	CO ₂ , L, T, RH, sound, motion, etc.	State* and quantity ⁺	Office	*92.5 – 97.1% *0.139 < RMSE < 0.758
FFNN [39]	Smart meter	State	Residential	80.98 – 91.36%
FFNN [25]	WiFi, CO ₂ , PIR, etc	Quantity	Office	R ² : 83%
TDNN [40]	CO ₂	Quantity	Office	0.51%(CV)15.52%
TDNN [41]	CO ₂	Quantity	Office	Indicative graphs
RBF-NN [42]	CO ₂ , L, T, RH, sound and motion	Quantity	Laboratory	63.23 – 88.74%
ELM [43]	CO ₂ , T, RH and P	Quantity* and range ⁺	Tutorial room	*74.06% and *81.37%
WRANK-ELM* and RIG-ELM ⁺ [44]	CO ₂ , T, RH and P	Range	Office	*75.63 – 79.17% *76.88 – 77.92%
HFS-ELM [45]	CO ₂ , T, RH and P	State* and range ⁺	Office	*94.37 – 98.12% *78.12 – 81.67%
OS-ELM [46]	WiFi	Tracking	Laboratory	1.973m
OS-ELM [47]	WiFi	State*, quantity ⁺ and tracking ^o	Commercial	*98.85%, *NRMSD : 0.096, °1.385m
RELM [48]	CO ₂ , L, T and H	State	Office	99.02 – 99.58%
FS-ELM [50]	CO ₂	Quantity	Office	94%

and 98.12% for the presence detection and 78.12% and 81.67% for the occupancy range estimation.

Zou et al. [46] proposed an indoor localization algorithm based on an online sequential extreme learning machine (OS-ELM). The authors conducted experiments in a laboratory (580 m²) to demonstrate that OS-ELM could provide a localization accuracy of 1.973 m. Moreover, in Ref. [47], Zou et al. presented WinOSS, a novel WiFi-based non-intrusive occupancy sensing system. WinOSS used OS-ELM as a fingerprinting-based machine learning localization algorithm. The system was implemented in the entire floor (1500 m²) of a commercial building, and accomplished a localization accuracy of 1.385 m. The system also proved to be effective in detecting and counting occupants.

Ertugrul et al. [48] presented the utilization of a recurrent extreme learning machine (RELM) for building occupancy detection. The authors compared the performance of RELM with other methods presented in literature. Experimental validation was performed using the dataset generated by Candanedo and Feldheim in Ref. [49], and the results of the comparative study showed that RELM outperformed the others methods with an accuracy between 99.02% and 99.58%. Jiang et al. proposed a feature scaled extreme learning machine (FS-ELM) algorithm to estimate the number of occupants in real time using CO₂ measurements [50]. The experimental data used to validate the proposed algorithm were collected in an office room with a capacity of 35 occupants, and the results obtained show an accuracy up to 94% with a tolerance of four occupants.

The occupancy detection works based on artificial neural networks are summarized in Table 2.

FFNN: feed-forward neural network, TDNN: time-delay neural network, RBF-NN: radial basis function neural network, ELM: extreme learning machine, WRANK-ELM: Wrapper-based ranking using ELM, RIG-ELM: Relative information gain and ELM, HFS-ELM: Hybrid Feature-Scaled ELM, OS-ELM: online sequential extreme learning machine, RELM: recurrent extreme learning machine, FS-ELM: feature scaled extreme learning machine, T: temperature, H: humidity, RH: relative humidity, L: light, CO: carbon-monoxide, TVOC: total volatile organic compounds.

3.2.2. Support vector machines

Literature shows that support vector machines (SVMs) have been used by several works for building occupancy detection. Zuraimi et al. used a SVM in combination with the data collected from CO₂ sensors to predict the number of occupants in a theatre [29]. The authors performed a cross-validation analysis to search the optimal parameters for a radial basis function (RBF) kernel, in order to optimize SVM performance. The results show that this approach can achieve an occupation detection accuracy of approximately 76%.

A combination of SVM with a sensor network to estimate the occupancy in an office building is presented in Refs. [34,35]. These studies reported an average accuracy of 75% for the occupancy number detection. Yang et al. [38] showed the utilization of SVM with twelve ambient sensor variables for the detection of occupancy status and number of occupants. The authors reported an occupancy status detection accuracy between 95.1% and 97.5%, while the number of occupants is estimated with a RMSE between 0.141 and 1.192. Han et al. used a SVM with the RBF kernel to determine occupancy using a sensor network conformed by PIR sensors, CO₂ concentration sensors, and relative humidity (RH) sensors [51]. Experimental validation was carried out in a research laboratory yielding an average accuracy of 77.6%.

Akbar et al. [52] investigated a non-intrusive approach based on the pattern recognition of electricity consumption data to detect the occupancy state. That work presented a performance comparison of SVM with three different kernels (linear, polynomial and radial basis functions) to show the importance of the proper kernel function selection. The electricity data used to validate this approach were collected in a research center, and the accuracy of the results was between 55.37% and 79.12%. Otherwise, Kleiminger et al. [53,54] explored the utilization of SVM and electricity consumption data to estimate the occupancy state of a household. In these works, the authors showed the importance of proper selection of characteristics used for learning process. The authors

Table 3
Overview of the occupancy detection works that have applied support vector machines.

Kernel	Sensors	Resolution	Building type	Performance
RBF [29]	CO ₂	Quantity	Theater	76%
RBF [34]	CO ₂ , L, T, CO, etc.	Quantity	Office	75%
RBF [35]	CO ₂ , T, RH, motion, etc.	Quantity	Office	75%
RBF [38]	CO ₂ , T, L, RH, sound, etc.	State* and quantity ⁺	Office	*95.1 – 97.5% + 0.141 < RMSE < 1.192
RBF [51]	CO ₂ , RH, PIR	Quantity	Laboratory	77.6%
Linear, Poly and RBF [52]	Smart meter	State	Office	55.37 – 79.12%
Not specified [53,54]	Smart meter	State	Residential	68 – 94%
Linear [56]	CO ₂ and T	State	Laboratory	97.57%
Domain-invariant kernel [57]	WiFi	Quantity	Conference room	92.8%

RBF: radial basis function kernel, Linear: linear kernel, Poly: polynomial kernel.

Table 4
Overview of the occupancy detection works that have applied K-nearest neighbor.

K-value	Distance function	Sensors	Resolution	Building type	Performance
k = 5 [38]	Euclidean	CO ₂ , T, L, RH, sound, etc.	State* and quantity ⁺	Office	*95.4 – 97.5% + 0.121 < RMSE < 0.79
K = 4 [59]	Euclidean	RFID	Quantity	University	62 – 88%
K = 10 [52]	Uniform	Smart meter	State	Office	93.67%
Not specified [53]	Not specified	Smart meter	State	Residential	71 – 89%
K = 1 [54]	Euclidean	Smart meter	State	Residential	71 – 92%

used the public ECO dataset¹ [55] for validation, and the obtained occupancy detection accuracy was between 68% and 94%.

Liu et al. [56] proposed a two-stage approach for occupancy estimation by using CO₂ and temperature sensors. In the first stage, an ELM algorithm was employed to obtain preliminary detection results. In the second stage, a linear SVM was exploited to generate final detection results. The outcomes showed that this approach could detect occupancy with an accuracy of approximately 97.57%. Zou et al. [57] proposed WiFree, a WiFi-based system to estimate the number of occupants. The authors used transfer kernel learning (TKL) to evaluate a domain-invariant kernel, which was later adopted as the SVM kernel. Experimental results suggested that WiFree can achieve 92.8% crowd counting accuracy.

An overview of the occupancy detection works based on support vector machines is presented in Table 3.

3.2.3. K-nearest neighbors

The K-nearest neighbors (KNN) algorithm is a machine learning algorithm based on a distance function that measures the difference or similarity between two instances [58]. Yang et al. [38] presented an approach that uses KNN and multiple environmental and specialized sensors to detect and count occupant numbers. The results show that it is possible to achieve an accuracy between 95.4% and 97.5% for binary occupancy detection, while the number of occupants is estimated with a RMSE between 0.121 and 0.79. Li et al. in Ref. [59] proposed an individualized occupancy detection system based on KNN and RFID technology. The system is used for the detection of the number of stationary and mobile occupants in multiple spaces simultaneously. The field tests were carried out in an education building at the University of Southern California, and the results obtained showed an average zone level detection accuracy of 88% for stationary and 62% for mobile occupants.

Akbar et al. [52] and Kleiminger et al. [53,54] have proposed the

utilization of KNN with electricity consumption data to detect building occupancy information. The work of Akbar et al. [52] showed that an approach using only electricity data and KNN was able to detect the occupancy state with an accuracy up to 93.67%. Furthermore, Kleiminger et al. used this same approach in Ref. [53,54]. In the former study, the authors' attempt yielded an occupancy detection accuracy between 71% and 89%. However, in the latter, the authors improved the results and achieved an accuracy of 92%.

We summarize the occupancy detection works based on K-nearest neighbors in Table 4.

3.2.4. Bayesian inference

Bayesian inference is a method of statistical inference that uses Bayes' theorem to update the probability for a hypothesis in the evidence of new data. Dodier et al. [60] used Bayesian methods to determine the occupancy in 2 private offices. The authors described the development of an occupancy detection system based on a belief network and the information collected by three PIR sensors and one telephone sensor per office. Boait and Rylatt [61] presented the results of the implementation of a prototype to control the heating system in a house in the UK. The prototype utilized Bayesian inference to identify the occupancy state in the house based on the electrical consumption data and the temperature of hot water. The authors reported an energy saving potential up to 14.1% in the gas consumed by the boiler (19.2 kWh/day).

Rahman and Han [40] used a Bayesian Markov chain Monte Carlo (MCMC) algorithm in combination with CO₂ concentration to estimate the number of occupants in an office room. The authors reported a coefficient of variation between 4.04% and 47.12%. Ebadat et al. [62] proposed the utilization of blind identification methods to estimate the number of occupants. The authors used measurements of CO₂ concentration and ventilation levels in combination with a Bayesian marginal likelihood method to estimate the number of occupants in a laboratory. The experimental validation showed that the proposed approach can achieve an accuracy of 82.1%.

Yang et al. used ambient sensors with naive Bayesian (NB) and tree augmented naive Bayes (TAN) network algorithms to build occupancy

¹ The ECO dataset includes records of the occupancy information and the electricity consumption of five different households for a period of approximately eight months [55].

Table 5
Overview of the occupancy detection works that have applied Bayesian inference.

Algorithm	Sensors	Resolution	Building type	Performance
Belief networks [60]	PIR	State	Office	Indicative graphs
Bayes' theorem [61]	Smart meter	State	Residential	Indicative graphs
Bayesian MCMC [40]	CO ₂	Quantity	Office	4.04% < CV > 47.12%
Bayesian ML [62]	CO ₂	Quantity	Laboratory	82.1%
NB [38]	CO ₂ , T, L, RH, sound, etc.	State* and quantity ⁺	Office	*88.9 – 94.3% † 0.189 < RMSE < 0.757
TAN [38]	CO ₂ , T, L, RH, sound, etc.	State* and quantity ⁺	Office	*95.3 – 98% ‡ 0.117 < RMSE < 0.651

MCMC: Markov chain Monte Carlo, Bayesian ML: Bayesian marginal likelihood, NB: naive Bayesian, TAN: tree augmented naive Bayes.

models to support demand-response HVAC controls [38]. The results for binary occupancy detection showed an accuracy between 88.9% and 94.3% for the NB algorithm and an accuracy between 95.3% and 98% for the TAN algorithm. In addition, the number of occupants is estimated with a RMSE between 0.189 and 0.757 for the NB algorithm and a RMSE between 0.117 and 0.651 for the TAN algorithm.

Table 5 summarizes the occupancy detection works based on Bayesian inference.

3.2.5. Hidden Markov Models

Due to the random nature of building occupancy, hidden Markov models (HMMs) have proved to be an effective method for occupancy detection. HMMs are stochastic process in which the occupancy is assumed to be a Markov chain with hidden states. In this model, the transition to a future state depends only on the current state. Since the states are latent variables, they are inferred by estimating the most likely sequence given the emissions (measurements from sensors) and the specified model [63]. In Ref. [64,65], a HMM has been used in combination with PIR sensors to monitor building occupancy. Liu et al. [64] validated their approach with data collected from an office and an open space, achieved an occupancy detection accuracy between 81% and 97%. On the other hand, Lu et al. [65] presented an automatic On/Off control of the HVAC system based on the estimation values of occupancy and sleep patterns. The system was tested with the information collected from 8 homes, indicating that the HMM approach can yield an occupancy status accuracy of 88%. Moreover, the authors reported an energy saving potential by up to 28%.

Candanedo et al. [66] presented a methodology based on HMM that was tested with three different datasets. The authors evaluated the influence of different environmental variables in the occupancy detection performance, obtaining as a result an accuracy that can vary between 19.1% and 92.54%, according to the combination of variables used. Likewise, Dong et al. [34] and Lam et al. [35] proposed to use a HMM with the information collected by an environmental sensor network to estimate the occupancy in an office building. The results reported by Dong et al. show that their approach was able to estimate the number of occupants with an average accuracy of 73%. On the other hand, Lam et al. achieved an average occupancy number detection accuracy of 80%.

Chaney et al. [67] introduces an approach for the estimation of home occupancy profile based on the combination of HMM with environmental sensors and smart meters. The authors showed that the proposed approach is able to estimate occupation adequately. In addition, they used occupancy information and consumption behaviors to estimate a households' potential to participate in demand response programs. Ryo and Moon [68] developed an approach based on the combination of a decision tree algorithm and HMM for the estimation of the number of occupants in a test-bed building. The experimental validation presented by the authors shows that the proposed approach can achieve an occupancy prediction accuracy between 85% and 93.2% [68].

Kleiminger et al. [53,54] developed a households' occupancy

detection technique based on HMM and electricity consumption data. In these works, the authors showed that with an adequate selection of the characteristics used for the HMM training, it is possible to increase the performance of the occupancy detection from 87% to 92%. Furthermore, Becker and Kleiminger [69] explored the utilization of unsupervised occupancy detection algorithms based on HMM and electricity consumption data. The authors used three publicly available datasets to validate the algorithm, and the results showed an occupancy detection accuracy of between 69% and 89%.

A comparative study between HMM and autoregressive HMM (ARHMM) was presented in Ref. [51,70]. These works deployed a sensor network in a research laboratory to collect data for their experimental studies. The results reported by these works showed an occupancy detection accuracy by up to 79.63% for the HMM and 80.78% for the ARHMM, demonstrating that ARHMM was more effective for estimating the number of occupants than HMM. Raykov et al. [71] presented the potential of using infinite HMM (iHMM) and statistical regression methods with PIR sensors for occupancy estimation. An experimental validation was performed by using data from seven different conference rooms with low occupancy (less than eight occupants) and high occupancy (equal or more than eight occupants). The results of this approach showed the ability to count the number of occupants in the room with a mean absolute error (MAE) between 0.64 and 0.99.

Graphical models such as layered HMM (LHMMs) and deterministic finite state machines (FSMs) were used in Refs. [72] to identify desk-related activities and to count people in an office room. The authors used PIR, plug-in power meter and ultrasound sensor data to evaluate the performance of both algorithms. The overall accuracy for activity recognition, such as computer and desk work, was 95%. The best performance obtained for the people count was between 78% and 87%. Additionally, the authors compared different control strategies based on the activity and occupancy information detected, reporting an energy saving potential by up to 78.5% and 39.9% per year for lighting and HVAC systems, respectively. Chen et al. [63] proposed an occupancy range estimation architecture (ranges of occupancy: zero, low, medium and high) based on an inhomogeneous HMM with multinomial logistic regression (IHMM-MLR). Experimental validation was carried out with the information collected from an office, and the results showed that the proposed method can achieve an estimation accuracy between 75.21% and 78.13%.

An overview of the occupancy detection works based on hidden Markov models is presented in Table 6.

3.2.6. Other data-driven methods

In the previous sections, we presented studies that have used some of the most popular data-driven methods. In this section, we analyze the works that have used other data-driven methods. Zuraimi et al. [29] presented a prediction error minimization (PEM) algorithm to predict the occupant numbers in a theater. Using the PEM algorithm, the occupancy counts estimation performance was 87.2% false positive and 3.46% false negative, which gives a negative predictive value (NPV) of

Table 6
Overview of the occupancy detection works that have applied Hidden Markov Models.

Algorithm	Sensors	Resolution	Building type	Performance
HMM [64]	PIR	State	Office	81 – 97%
HMM [65]	PIR	State	Residential	88%
HMM [66]	CO ₂ , L, T and H	State	Office	19.1 – 90.24%
HMM [34]	CO ₂ , L, T, CO, dew point, etc.	Quantity	Office	73%
HMM [35]	CO ₂ , T, RH, motion, TVOC, etc.	Quantity	Office	80%
HMM [67]	CO ₂ , T, smart meter and dew point	State	Residential	Indicative graphs
HMM [68]	CO ₂ , smart meter	Quantity	Test-bed	85 – 93.2%
HMM [53,54]	Smart meter	State	Residential	70 – 92%
HMM [69]	Smart meter	State	Residential	69–89%
HMM and ARHMM [51]	CO ₂ , RH, PIR	Quantity	Laboratory	HMM: 79.63% , ARHMM: 80.78%
HMM and ARHMM [70]	CO ₂ , T, RH, PIR, etc.	Quantity	Laboratory	HMM: 25.2 – 76.2% , ARHMM: 80.1 – 84%
iHMM [71]	PIR	Quantity	Meeting room	0.64 < MAE < 0.99
LHMM and FSM [72]	PIR, Smart meter, etc.	Quantity	Office	78 – 87%
IHMM-MLR [63]	CO ₂ , H, T and P	State* and range ⁺	Office	*94.37%, ⁺ 75.2 – 78.1%

ARHMM: autoregressive hidden Markov model, iHMM: infinite Hidden Markov model, LHMM: layered hidden Markov model, FSM: finite state machine, IHMM-MLR: inhomogeneous hidden Markov model with multinomial logistic regression.

96.18%. Yang et al. [38] used a decision tree (DT) algorithm with ambient sensors to build occupancy models to support demand-response HVAC controls. This approach achieved an accuracy between 96% and 98.2% for binary occupancy detection, and a RMSE between 0.109 and 0.688 for occupancy number estimation. In addition, the authors reported that an energy saving potential by up to 21.3% (22.3% of gas and 19.6% of electricity) can be achieved, by using an occupancy-based HVAC control.

Candanedo and Feldheim [49] compared the performance of four occupancy detection algorithms: random forest (RF), gradient boosting machines (GBM), linear discriminant analysis (LDA) and classification and regression trees (CART). The authors used environmental data collected from an office, and the results showed a variation in classification performance between 32.68% and 99.33%, highlighting the importance of a proper method and feature selection. Gao et al. presented the concept of a self-programming thermostat that uses occupancy statistics to control an HVAC system [73]. The authors used motion sensors (PIR) and magnetic reed switches to register leaving and arrival information of the inhabitants, and with that data, HVAC system control automatically creates an optimal setback schedule. The approach was validated with the occupancy information of two homes, and the results indicated that this approach was able to reduce HVAC electricity consumption up to 15%. Cardenas et al. presented a correlation analysis between the occupancy and several environmental variables, in order to identify the relationship between the signals, as well as their phase difference in the temporal domain. Additionally, the authors used a linear regression model to estimate occupancy in a laboratory and a residence, reporting a RMSE between 0.113 and 0.358.

Hobson et al. [25] compared the performance of multiple linear regression models with different combinations of sensors in order to examine the occupancy number in an academic office building. The authors reported a mean R^2 of 80.1%. Becker and Kleiminger [69] explored the utilization of the geometric moving average (GeoMA) algorithm and the Page-Hinkley test (PHT) algorithm with electricity consumption data to detect the occupancy status. These unsupervised methods were validated using three different databases, and the results showed that GeoMA is able to achieve an occupancy detection accuracy of between 70% and 91%, while PHT obtained an accuracy of between 68% and 89%. Ebadat et al. [62] used measurements of CO₂ concentration and ventilation levels to infer the number of occupants. In this work, the authors proposed the utilization of a frequentist maximum likelihood (FML) method for occupancy number estimation. The experimental validation was carried out using the public dataset presented in Ref. [74], and the results showed that the proposed approach

can achieve an accuracy of 83%.

Yokoishi et al. [7] proposed using a particle filter in combination with the information collected with a network of PIR and illumination sensors for occupancy detection. The principal advantage of particle filtering is that it can eliminate incorrect occupancy state transition without incurring large delays. Experimental data were collected in a campus room (15.3m x 6.4m) outfitted with four PIR sensors. The results showed an occupancy detection accuracy between 65% and 98.3%. In addition, an occupancy-based lighting control system was tested over 3 months, showing that it was possible to save 3.5 h of lighting power per day.

Szczurek et al. [75] proposed to use pattern matching in CO₂ concentration time series to detect the occupancy and their influence on indoor air quality (IOQ). Experimental data were collected in a church and the best occupancy detection result achieved with this approach was 82% true positive and 22% false positive, which gave a positive predictive value (PPV) of 78.8%. Ebadat et al. [76] estimated the number of people in a room using an approach divided into two phases: 1) the authors designed a model to capture the dynamic relations between occupancy levels and the measurements of CO₂ concentration, temperature, fresh air inflow and door opening/closing events, 2) the model was used to formulate the occupancy estimation task as a regularized deconvolution problem. The experimental validation evidence an occupancy estimation accuracy between 82.2% and 88.8%.

Krumm and Brush [77] combined the drive-time algorithm derived from GPS data with a probabilistic schedule (PS) to estimate when a person would be at home or away. The results that the proposed approach can achieve an occupancy estimation accuracy between 64% and 68%. Park et al. [78] proposed an occupancy estimator based on Bluetooth (BT) signals. The authors used the capture and recapture (CRC) technique in different simulated environments by changing population size, BT ratio, and transition ratio. The results indicated that the system calculated accurate occupancy profiles and could be implemented in building environments for future experimental studies.

Deatla et al. [79] estimate the number of people walking in an area based on only WiFi received signal strength indicator (RSSI) measurements. The authors developed a mathematical expression for the probability distribution of the received signal amplitude as a function of the total number of occupants. Consequently, the occupancy estimation was achieved by using Kullback–Leibler divergence with the formulated mathematical expression. Longo et al. [80] proposed an occupancy estimation system based on the capture of both WiFi and Bluetooth management frames, transmitted from users' devices. With this information, the system created different sets of features that were used by

Table 7
Overview of the occupancy detection works that have applied other data-driven methods.

Algorithm	Sensors	Resolution	Building type	Performance
PEM [29]	CO ₂	Quantity	Theater	NPV : 96.18%
DT [38]	CO ₂ , T, L, RH, sound, etc.	State* and quantity ⁺	Office	*96 – 98.2% *0.109 < RMSE < 0.688
RF, GBM, LDA and CART [49]	CO ₂ , T, L and RH	State	Office	32.68 – 99.33%
Historical statistics [73]	PIR	State	Residential	–
Linear regression [82]	CO ₂ , T, RH, L, etc.	State	Residential and laboratory	0.113 < RMSE < 0.358
Linear regression [25]	WiFi, CO ₂ , PIR, etc	Quantity	Office	R ² : 80.1%
GeoMA and PHT [69]	Smart meter	State	Residential	GeoMA : 70 – 91%, PHT : 68 – 89%
FML [62]	CO ₂	Quantity	Laboratory	83%
Particle filtering [7]	PIR	State	Campus room	65% – 98.3%
Pattern matching [75]	CO ₂	Quantity	Church	PPV : 78.8%
Regularized deconvolution [76]	CO ₂ , T, etc.	Quantity	Laboratory	82.2 – 88.8%
Drive-time algorithm and PS [77]	GPS	State	Residential	64 – 68%
CRC [78]	Bluetooth	Occupancy density	Academic building	RMSE < 0.22
Kullback–Leibler divergence [79]	WiFi	Quantity	Indoor and outdoor site	Indicative graphs
RLR and MLR [80]	WiFi and Bluetooth	Quantity	Indoor and outdoor site	Indoor : 74 – 97% Outdoor : 87%
NL and TL [81]	Smart meter	State	Residential and office	NL : 74 – 89%, TL : 90%

PEM: prediction error minimization, DT: decision tree, RF: random forest, GBM: gradient boosting machines, LDA: linear discriminant analysis, CART: classification and regression trees, PHT: Page-Hinkley test, FML: frequentist maximum likelihood, PS: probabilistic schedule, CRC: capture and recapture technique, RLR: regularized linear regression, MLR: multinomial logistic regression, NL: non-intrusive learning, TL: transfer learning.

two supervised learning models for performing either precise occupancy estimation or occupancy level classification.

In [81], Jin et al. focused on the problems for which no/limited labeled data is available for model learning. The authors proposed a non-intrusive learning (NL) algorithm to detect occupancy with only power consumption data. They suggested a transfer learning (TL) algorithm for tackling the cases, where data from other sources were used in the current learning task. Experimental validation of these methods was performed with two different datasets. The results presented occupancy detection accuracy of approximately 74%–89% in homes and 90% in offices.

Table 7 summarizes the occupancy detection works based on other machine learning techniques.

3.3. Knowledge-based methods

Knowledge-based methods also known as expert system methods use specialized knowledge represented by rules to solve complex problems. Some authors have explored the utilization of knowledge-based methods to detect occupancy information. The design and implementation of a presence sensor platform for occupancy detection and applications in HVAC control are presented by Agarwal et al. in Ref. [3]. Their system uses a rule-based system in combination with PIR and magnetic reed switch door sensor data to detect occupancy. Experimental data were collected from ten offices over a two week period, and the results showed that the algorithm accurately detected occupancy. In addition, the authors used an occupancy-based HVAC control system and they reported an energy saving potential between 10% and 15%.

A plug-and-play occupancy detection method based on the trajectory of indoor environmental sensors was presented by Pedersen et al. [83]. The proposed method used data collected by sensors to apply a set of rules to determine the probability of occupancy. The authors tested the system in an office room and a three-room apartment, obtaining maximum accuracies of 98% and 78%, respectively. Ansanay [84] used a thresholding (THR) algorithm to estimate the occupancy state based on CO₂ levels. The experimental data, used for this study were collected from an office building using a time step of 10 min. The utilization of smart meters to detect when a household is occupied by its residents is presented by Kleiminger et al. in Ref. [53,54]. In the first work, the authors used a thresholding algorithm with data from five households,

giving a classification accuracy between 57% and 79%. In the second paper, the authors used the ECO dataset with the same THR algorithm and a Gaussian mixture model (GMM). In that work, the occupancy detection accuracy achieved for the THR and GMM algorithms was between 66% to 90% and 59%–88%, respectively.

Chen et al. [85] presented a non-intrusive occupancy monitoring (NIOM) system based on aggregate electricity consumption data from smart meters to infer occupancy. The authors proposed a threshold-based NIOM algorithm capable of detecting changes in three statistical metrics (average power, standard deviation and power range) to generate an occupancy profile. The authors validated their method using real data collected from two homes during the summer. The results showed that occupancy detection accuracy was between 73.27% and 90.63%. Soltanaghaei et al. [86] presented an off-line Walksense methodology based on motion sensors installed in walkways to infer zone occupancy states. This system was tested in six houses and the results showed an average occupancy detection accuracy by up to 96%. In addition, the authors controlled the temperature set-point of the HVAC system based on occupancy information, showing that HVAC energy consumption can be reduced by 30%–47%.

Wahl et al. [87,88] proposed two algorithms that use motion events from PIR sensors to estimate people count in an office. The first was a simple direction-based algorithm (DBA), and the second was a probabilistic distance-based algorithm (PDBA). The authors placed a pair of PIR sensors at the entrance and exit locations in the office space. Based on this topology, it was possible to use the timing between the motion events reported by the PIRs to differentiate an occupant's movement direction.

In [89], Tang et al. proposed an occupancy detection framework called sparse human action recovery with knowledge of appliances (SHARK). The SHARK method required no training process and, instead, used only the load curve data from a single meter in a residential house. This approach was evaluated with real data and the results showed that the proposed method achieved a precision between 0.733 and 1 evidencing a similar performance to the supervised approaches presented in literature. The identification of occupant presence in an office based on the information collected by ultrasonic and acceleration sensors, WiFi access points (to detect the presence of smartphones) and electricity consumption was presented by Jin et al. in Ref. [90]. In the first part of the work, a decision rule algorithm and a thresholding

Table 8
Overview of the occupancy detection works that have applied knowledge-based methods.

Algorithm	Sensors	Resolution	Building type	Performance
Rule-based system [3]	PIR	State	Office	–
Rule-based system [83]	CO ₂ , PIR, T, RH, etc.	State	Residential and office	78 – 98%
Thresholding [84]	CO ₂	State	Office	–
Thresholding [53]	Smart meter	State	Residential	57 – 79%
Thresholding and GMMs [54]	Smart meter	State	Residential	59 – 90%
Thresholding and clustering [85]	Smart meter	State	Residential	73.27 – 90.63%
Walksense method [86]	PIR	State	Residential	95 – 96%
DBA and PDBA [87,88]	PIR	Quantity	Office	–
SHARK [89]	Smart meter	State	Residential	0.733 < Precision < 1
PresenceSense algorithm [90]	Smart meter	State	Office	87 – 93%
WiFi sniffing [91]	WiFi	Zone and quantity	Office	90%
Sentinel [92]	WiFi	Zone and identity	Commercial	86%
STI [57]	WiFi	State	Conference room	99.1%

algorithm were used to obtain the presence states, and the results showed an accuracy between 70.62% and 93.71%. The second part of the paper presented a semisupervised learning algorithm called PresenceSense, which allows to infer presence with an accuracy between 87% and 93%.

Li et al. [91] derived coarse-grained and fine-grained occupancy information using WiFi signals. The authors used two approaches, the first approach (WiFi Sniffing) monitored access points in a building and their communications with WiFi devices. This method allowed for estimation of the number of people in the building and their approximate locations. The second approach added an indoor positioning system (IPS) to improve the accuracy. Experimental validation was done in an office area, yielding an accuracy of 90%. Balaji et al. [92] presented Sentinel, a system that leverages the existing WiFi infrastructure in commercial buildings along with smartphones with WiFi connectivity to infer occupancy and uses that information to control the HVAC system. The system used information from the authentication, authorization and accounting (AAA) logs of WiFi clients' devices to provide sufficient information for occupancy inference. This work showed that Sentinel was able to accurately determine the occupancy in office spaces 86% of the time. For one day, the authors tested an ON/OFF HVAC control based on occupancy information, obtaining as a result a savings of 17.8% in terms of HVAC electricity consumption.

Zou et al. [57] proposed WiFree, a novel device-free occupancy detection scheme by using only commercial WiFi. The authors used the signal tendency index (STI) to perform occupation detection in three indoor environments with different sizes. Experimental results revealed that WiFree could achieve 99.1% occupancy detection accuracy.

We summarize the occupancy detection works that studied knowledge-based methods in Table 8.

GMMs: Gaussian mixture models, DBA: direction-based algorithm,

PDBA: probabilistic distance-based algorithm.

4. Discussion

In this section, we present a discussion about advantages and limitations of occupancy detection technologies and algorithms. We provide a comparison framework to help researchers select sensors and algorithms that are more appropriate to implement according to the application context and accuracy requirements.

In order to enable a comparison framework of occupancy detection systems, it is important to consider a set of characteristics such as sensor (s) types, data processing techniques, occupancy resolution, and performance metrics.

4.1. Occupancy detection sensors

Occupancy detection systems are mainly based on the deployment of multiple environmental sensors (e.g., carbon dioxide (CO₂), temperature, humidity, and light sensors) or other specialized devices (e.g., PIR sensors, smart meters, and cameras) [93]. In Table 9, we present a brief comparison of these technologies. For this purpose, we have applied a set of comparison criteria, comprising the cost, the privacy, and the type of occupancy information they offer (occupancy resolution). Moreover, the main advantages and limitations of each technology are highlighted. This comparison is partially inspired by previous works on occupancy detection that have been derived from the conducted literature review [22,23,94].

As presented in Table 9, each technology has some strengths and weaknesses for the development of occupancy detection systems. Accordingly, a combination of multiple environmental and specialized sensors has been considered as an alternative to enhance the

Table 9
Comparison of different occupancy detection technologies.

Sensor	Occupancy information resolution					Cost	Privacy issue	Advantages	Limitations	
	State	Quantity	Activity	Identity	Track					
Environmental	PIR	Yes	Yes	No	No	No	Low	No	Low computational cost	Not capture static occupants
		Yes	Yes	Yes	No	No	Medium	No	Widely available and non-intrusive	Time lag and sensitive to the environment
Smart meter		Yes	No	Yes	No	No	NO	Partial	Existing infrastructure	Miss occupants not using appliance
	WiFi/Bluetooth	Yes	Yes	No	Yes	Yes	Medium	Partial	Widely available	Occupants with multiple smartphones or they may not turn on WiFi/Bluetooth
Camera	Yes	Yes	Yes	Yes	Yes	Hight	Hight	Yes	Hight accuracy	Hight computational complexity and illumination conditions

performance of occupancy detection [95].

4.2. Occupancy detection algorithms

According to our analysis, three major families of algorithms have been identified that consist of analytical, data-driven, and knowledge-based methods. Based on this classification, we summarize the pros and cons of each method in the following.

Analytical techniques are based on physical models that necessitate acquiring a detailed knowledge of the processes that occur in the targeted system. Due to the dynamic equations knowledge of the system, these types of models provide the closest match to the real system and thus, result in good accuracy. However, they are one of the most complex types of techniques to implement due to the need to have a deep understanding of all variables, involved in the system. For occupancy detection, these methods have been mainly used for the modeling of HVAC systems in order to identify the number of people in an indoor space based on variations in CO₂ concentration.

Data-driven methods are mainly based on machine learning algorithms and statistical and probabilistic approaches, which can scrutinize data and infer knowledge for making predictions or classifications. Among the advantages of this type of method, we can mention the ability to identify nonlinear relationships between independent and dependent variables, a wide range of applications, and the capability to learn, predict, and improve knowledge over time without human intervention. However, one of the major challenges of data-driven methods is that they require a massive amount of inclusive/unbiased and qualified data for the learning process, which has a significant impact on the accuracy of the results.

Knowledge-based methods are able to solve complex problems by using knowledge, provided by human experts on the domain. Expert systems offer good performance, easy maintenance and do not require training process or model calibration. However, the process of extracting, structuring, and organizing domain knowledge from experts for an application (knowledge acquisition) is a difficult and time-consuming task. For occupancy detection applications, these methods have reported a similar performance to existing analytical techniques and data-driven methods [83].

4.3. Performance comparison

Regarding the analysis, made in Section 3, the performance of occupancy detection systems is analyzed using various metrics such as accuracy, RMSE, precision, recall, and F-measure. Among these metrics, accuracy is one of the most used scores. It is defined as the number of correctly estimated instances divided by their total number [35,38]. However, modified versions of the accuracy have been also used. Jian et al. [50] introduced the concept of x – tolerance accuracy. This metric provides the percentage of instances for which the estimated result (\hat{o}) has an error of less than x (tolerated error), and is defined by:

$$Acc(\hat{o}, x) = \frac{\sum_{k=1}^N X(|o_k - \hat{o}_k|, x)}{N} \quad (1)$$

where

$$X(|o_k - \hat{o}_k|, x) = \begin{cases} 1, & \text{if } |o_k - \hat{o}_k| \leq x \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

According to our review, many occupancy detection techniques have been exploited to process the information provided by a particular type of sensor. Therefore, in Table 10, we compare the performance of occupancy detection systems according to the sensors, the resolution (occupancy status, and the number of occupants) and the type of algorithm. For this comparison, we use the articles that have employed the accuracy as the performance metric. Additionally, to simplify the comparison, we compute the average accuracy of the performance reported by the papers in each category.

Based on Table 10, we can observe that the occupancy state has been calculated with an average accuracy between 73.8% and 98.9%, and the number of occupants has been computed with an accuracy between 67.8% and 92.8%. Moreover, this comparison shows that the efficiency of an occupancy detection system depends not only on the type of sensors, but also the algorithm that is used to process the data, collected by the sensors. In fact, Table 10 shows that for a certain sensor type and resolution level, an efficient algorithm can improve the accuracy of the results up to 23.2%.

4.4. Energy saving potential

In this article, we associate the use of occupancy detection systems as a tool with improving building energy performance, reducing building energy consumption, and maintaining occupants' comfort. According to the review, presented in Section 3, some works have been done in this direction and demonstrated that up to 55% of the HVAC electricity can be saved by adopting occupancy-based control strategies (see Table 11).

Accordingly, Table 11 presents a summary of the energy saving potential that has been reported in the studies, analyzed in this review. This information shows that most of the studies have been conducted with a focus on HVAC systems. Likewise, Table 11 outlines a high energy saving through the control of lighting systems.

5. Research trends and future prospects

Numerous studies have investigated the utilization of various technologies and algorithms to detect occupancy information. According to the above discussions, we present the research trends and future prospects hereunder.

In general, Fig. 2 shows a growing trend over time in the type of sensors and algorithms used in occupancy detection systems, that can be divided into two phases. The first period occurred during 1998–2006,

Table 10

Average accuracy based on the occupancy resolution, the type of sensors, and the utilizing algorithm.

	Occupancy state					Quantity of occupants		
	PIR	Environmental sensors	Smart meters	WiFi and Bluetooth	Sensor fusion	Environmental sensors	WiFi and Bluetooth	Sensor fusion
Analytical methods	-	86%	-	-	-	67.8%	-	-
Data-driven methods	ANN	83.2%	86.5%	98.9%	95.4%	81.7%	-	79.7%
	SVM	-	97.6%	82.1%	-	96.2%	92.8%	75.9%
	KNN	-	-	83.2%	-	96.2%	-	-
	Bayesian inference	-	-	-	-	94.2%	82.1%	-
	HMM	88.7%	74.4%	81.6%	-	94.4%	-	76.6%
Others	96.3%	89.4%	81.2%	-	97.3%	83%	87.8%	87%
Knowledge-based methods	81.7%	-	73.8%	92.7%	77.6%	-	-	-

ANN: artificial neural network, SVM: support vector machine, KNN: K-nearest neighbors, HMM: hidden Markov model.

Table 11
Overview of energy saving potential reported in literature.

Author	Sensors	Resolution	Building type	Application system	Energy saved
Boait and Rylatt [61]	Smart meter	State	Residential	HVAC	19.2kWh/day
Lu et al. [65]	PIR	State	Residential	HVAC	28%
Gao et al. [73]	PIR	State	Residential	HVAC	15%
Soltanaghaei et al. [86]	PIR	State	Residential	HVAC	30 – 47%
Milenkovic et al. [72]	PIR, Smart meter, etc.	Activity and quantity	Office	Lighting and HVAC	Light = 78.5% HVAC = 39.9%
Wang et al. [27,28]	CO ₂	Quantity	Office	HVAC	8 – 13%
Yang et al. [38]	CO ₂ , T, L, RH etc.	State and quantity	Office	HVAC	21.3%
Agarwal et al. [3]	PIR	State	Office	HVAC	10 – 15%
Jin et al. [4,5]	CO ₂	Quantity	Conference room	HVAC	55%
Yokoishi et al. [7]	PIR	State	Campus room	Lighting	3.5h/day
Balaji et al. [92]	WiFi	Zone and identity	Commercial	HVAC	17.8%

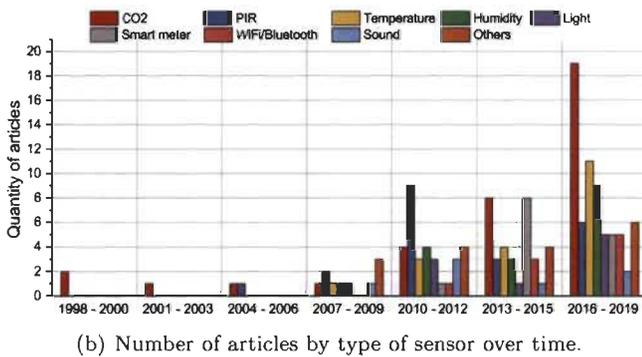
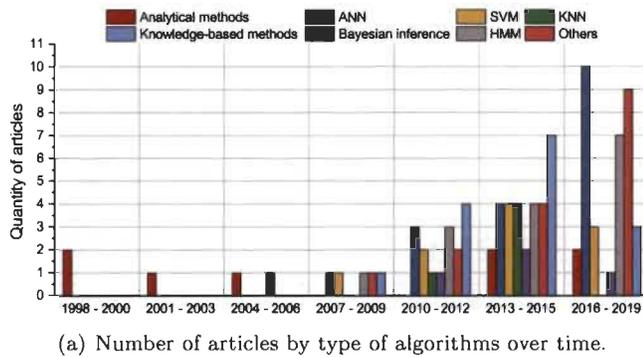
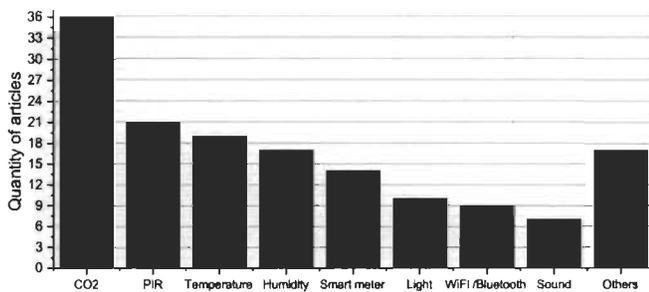
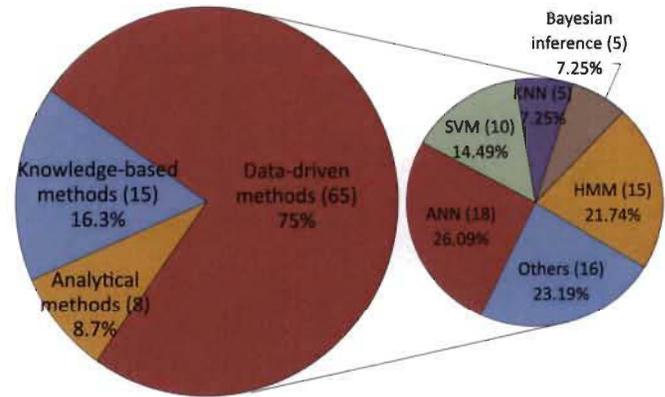


Fig. 2. Temporal analysis.



when most of the publications were based on the use of analytical methods in combination with environmental sensors (mainly CO₂ sensors). The second phase has started after the first period, where a great variety of new technologies and algorithms have been used and subsequently caused an increase in the number of published articles.



According to the literature presented in this review, we can highlight some lessons learned and various factors that have driven the development of occupancy detection systems. Several works have shown that each type of sensor has unique advantages and limitations for occupancy detection applications [20,38,96]. Therefore, sensor fusion has been promoted as an efficient way to leverage the strengths of different sensors to improve system performance. This fact can be evidenced in Fig. 2, where it is observed that since 2007, the use of technologies like motion detectors, CO₂, temperature, and humidity sensors have been an approach widely employed in occupancy detection. Indeed, as shown in Fig. 3, PIR, and environmental sensors are the technologies that have been mostly employed in this field.

Another factor that has driven interest in detection systems is the increasing importance of understanding the impact of occupants' behavior on buildings' energy performance. This has motivated researchers to develop occupancy-based control strategies, as well as the use of new technologies such as smart meters, which have been used in applications for occupancy detection systems since 2010.

The growing interest in occupancy detection can also be related to technological advances and the proliferation of the Internet of Things (IoT). Over the last decade, the IoT has led to the development of new low-cost sensors to enable smarter sensing and detection [19]. Therefore, as a result of the various data sources available, several techniques have been exploited to process the data provided by the sensors, in order to estimate reliable building occupancy information (Fig. 2). In this regard, the popularity of data-driven methods such as ANN and HMM should be highlighted. This fact is evidenced in Fig. 4, where it can be observed that within the literature, studied in this review, data-driven methods are the most popular type of algorithms for occupancy detection, followed by knowledge-based methods.

Further, literature reports a rising need for accurate and reliable

crowd counting, e.g., to improve the evacuation procedure in emergencies or to enhance customer satisfaction in restaurants or shopping malls [57]. Consequently, the use of new types of technologies, such as WiFi and Bluetooth, has emerged as a possible solution. Among the advantages of these technologies, we can highlight the fact that its infrastructure is widely available in interior spaces, which reduces development costs. Besides, a large variety of devices such as smartphones or laptops have built-in WiFi and Bluetooth connectivity.

5.1. Future research opportunities

5.1.1. Occupancy detection from smart meters

Different detection technologies have been employed to provide occupancy information, mostly based on environmental and specialized sensors. It is highlighting the fact that CO₂ and PIR sensors have been the most commonly used sensors for occupancy detection systems. However, literature has shown the potential of electricity consumption data to be used in occupancy detection applications. In addition, one advantage of this approach is that the measure can be performed in a non-intrusive manner (non-intrusive load monitoring), which reduces the complexity of necessary infrastructure and the associated costs; while maintaining the privacy of the occupants. Accordingly, we believe that this approach is a potential future research direction. Even though some initial works have been done in these directions, more efforts to investigate its advantages and disadvantages over approaches that use traditional detection technologies such as CO₂ can be conducted.

5.1.2. Occupancy detection systems in real-world settings

One of the main drawbacks of occupancy detection methods is the need of a training set comprising ground truth occupancy data. This is a problem because the collection of these measurements can be expensive or in some cases unfeasible, making it difficult to apply these methods in practice. To address this problem, some studies [62,69,81,89,90] have explored the use of unsupervised methods as a new alternative for occupancy detection, because these algorithms do not need to know the occupancy status of the buildings for their learning process. Accordingly, we believe that this is another potential future research direction that must be encouraged in order to overcome the existing limitations for the application of occupancy detection methods in real-world settings.

6. Conclusion

This paper presents an extensive review of the techniques that have been exploited to estimate building occupancy information. The researches examined in this work were categorized based on the type of algorithm as analytical, data-driven, and knowledge-based methods. We presented a discussion about the advantages and limitations of occupancy detection technologies and techniques to provide a comparison framework with the aim of helping researchers with their choice of sensors and algorithms.

We demonstrated the temporal evolution of sensors and algorithms, used for occupancy detection systems, as well as the factors that have motivated the development of these systems. Our analysis allows to observe that data-driven methods such as HMM and ANN are the most common family of algorithms for occupancy detection. Additionally, potential fields of research and improvement have been presented.

Occupancy information is crucial to improving building energy performance and reducing building energy consumption while maintaining occupants' comfort. Indeed, several works have demonstrated that a significant amount of energy can be saved by adopting occupancy-based control strategies. Although a considerable progress has been made in occupancy detection systems, further work is required to achieve viable applications.

Acknowledgment

The authors would like to thank the Laboratoire des technologies de l'énergie d'Hydro-Québec, the Natural Science and Engineering Research Council of Canada, and the Foundation of Université du Québec à Trois-Rivières.

References

- [1] T.A. Nguyen, M. Aiello, Energy intelligent buildings based on user activity: a survey, *Energy Build.* 56 (2013) 244–257.
- [2] Y. Zhang, X. Bai, F. P. Mills, J. C. Pezzey, Rethinking the Role of Occupant Behavior in Building Energy Performance: A Review, *Energy and Buildings*.
- [3] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, T. Weng, Occupancy-driven energy management for smart building automation, in: *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ACM, 2010, pp. 1–6.
- [4] M. Jin, N. Bekiaris-Liberis, K. Weekly, C.J. Spanos, A.M. Bayen, Occupancy detection via environmental sensing, *IEEE Trans. Autom. Sci. Eng.* 15 (2) (2018) 443–455.
- [5] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, A. Bayen, Sensing by proxy: occupancy detection based on indoor CO₂ concentration, *UBICOMM* 14 (2015).
- [6] T. Leephakpreeda, Adaptive occupancy-based lighting control via grey prediction, *Build. Environ.* 40 (7) (2005) 881–886.
- [7] T. Yokoiishi, J. Mitsugi, O. Nakamura, J. Murai, Room occupancy determination with particle filtering of networked pyroelectric infrared (pir) sensor data, in: *Sensors*, IEEE, 2012, pp. 1–4. IEEE, 2012.
- [8] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, N. Villar, Preheat: controlling home heating using occupancy prediction, in: *Proceedings of the 13th International Conference on Ubiquitous Computing*, ACM, 2011, pp. 281–290.
- [9] Y. Peng, A. Rysanek, Z. Nagy, A. Schlüter, Occupancy learning-based demand-driven cooling control for office spaces, *Build. Environ.* 122 (2017) 145–160.
- [10] C. Guzmán, K. Agbossou, A. Cardenas, Real-time emulation of residential buildings by hardware solution of multi-layer model, *IEEE Transactions on Smart Grid* 10 (4) (2018) 4037–4047.
- [11] H.N. Rafsanjani, C.R. Ahn, M. Alahmad, A review of approaches for sensing, understanding, and improving occupancy-related energy-use behaviors in commercial buildings, *Energies* 8 (10) (2015) 10996–11029.
- [12] J. Conti, P. Holtberg, J. Diefenderfer, A. LaRose, J.T. Turnure, L. Westfall, *International Energy Outlook 2016 with Projections to 2040*, Tech. rep., USDOE Energy Information Administration (EIA), Office of Energy Analysis, Washington, DC (United States), 2016, 2016.
- [13] H. Yoshino, T. Hong, N. Nord, Iea ebc annex 53: total energy use in buildings—analysis and evaluation methods, *Energy Build.* 152 (2017) 124–136.
- [14] M. Jia, R.S. Srinivasan, A.A. Raheem, From occupancy to occupant behavior: an analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency, *Renew. Sustain. Energy Rev.* 68 (2017) 525–540.
- [15] R. Melfi, B. Rosenblum, B. Nordman, K. Christensen, Measuring building occupancy using existing network infrastructure, in: *Green Computing Conference and Workshops (IGCC)*, 2011 International, IEEE, 2011, pp. 1–8.
- [16] M.B. Kjaergaard, F.C. Sangogboye, Categorization framework and survey of occupancy sensing systems, *Pervasive Mob. Comput.* 38 (2017) 1–13.
- [17] B. F. Valvedi, E. Ghisi, R. Lamberts, A Review of Occupant Behaviour in Residential Buildings, *Energy and Buildings*.
- [18] J. Yang, M. Santamouris, S.E. Lee, Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings, *Energy Build.* 121 (2016) 344–349.
- [19] P.K. Mane, K.N. Rao, Review of research progress, trends and gap in occupancy sensing for sophisticated sensory operation, in: *Computer Science On-Line Conference*, Springer, 2018, pp. 212–222.
- [20] T. Labeodan, W. Zeiler, G. Boxem, Y. Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, *Energy Build.* 93 (2015) 303–314.
- [21] H. Saha, A. R. Florita, G. P. Henze, S. Sarkar, Occupancy Sensing in Buildings: A Review of Data Analytics Approaches, *Energy and Buildings*.
- [22] Z. Chen, C. Jiang, L. Xie, Building Occupancy Estimation and Detection: A Review, *Energy and Buildings*.
- [23] K. Sun, Q. Zhao, J. Zou, A Review of Building Occupancy Measurement Systems, *Energy and Buildings*, 2020, 109965.
- [24] S. Petersen, T.H. Pedersen, K.U. Nielsen, M.D. Knudsen, Establishing an image-based ground truth for validation of sensor data-based room occupancy detection, *Energy Build.* 130 (2016) 787–793.
- [25] B.W. Hobson, D. Lowcay, H.B. Gunay, A. Ashouri, G.R. Newsham, Opportunistic occupancy-count estimation using sensor fusion: a case study, *Build. Environ.* 159 (2019) 106154.
- [26] D. Cali, P. Matthes, K. Huchtemann, R. Streblow, D. Müller, CO₂ based occupancy detection algorithm: experimental analysis and validation for office and residential buildings, *Build. Environ.* 86 (2015) 39–49.
- [27] S. Wang, X. Jin, CO₂-based occupancy detection for on-line outdoor air flow control, *Indoor Built Environ.* 7 (3) (1998) 165–181.

- [28] S. Wang, J. Burnett, H. Chong, Experimental validation of co₂-based occupancy detection for demand-controlled ventilation, *Indoor Built Environ.* 8 (6) (1999) 377–391.
- [29] M. Zuraimi, A. Pantazaras, K. Chaturvedi, J. Yang, K. Tham, S. Lee, Predicting occupancy counts using physical and statistical co₂-based modeling methodologies, *Build. Environ.* 123 (2017) 517–528.
- [30] S.A. Mumma, Transient occupancy ventilation by monitoring co₂, *ASHRAE IAQ Appl.* 5 (1) (2004) 21–23.
- [31] J. Sowa, in: *Co₂-based Occupancy Detection for On-Line Demand Controlled Ventilation Systems*, Proceedings of Indoor Air, 2002, pp. 334–339. Monterey, CA, USA.
- [32] K. Weekly, N. Bekiaris-Liberis, M. Jin, A.M. Bayen, Modeling and estimation of the humans' effect on the co₂ dynamics inside a conference room, *IEEE Trans. Contr. Syst. Technol.* 23 (5) (2015) 1770–1781.
- [33] T.D. Buskirk, An Introduction to Machine Learning Methods for Survey Researchers, Ph.D. thesis, University of Rochester, 2018.
- [34] B. Dong, B. Andrews, K.P. Lam, M. Hoyneck, R. Zhang, Y.-S. Chiou, D. Benitez, An information technology enabled sustainability test-bed (itest) for occupancy detection through an environmental sensing network, *Energy Build.* 42 (7) (2010) 1038–1046.
- [35] K.P. Lam, M. Hoyneck, B. Dong, B. Andrews, Y.-S. Chiou, R. Zhang, D. Benitez, J. Choi, et al., Occupancy detection through an extensive environmental sensor network in an open-plan office building, *IBPSA Build. Simulat.* 145 (2009) 1452–1459.
- [36] P. Kraipeerapun, S. Amornsamankul, Room occupancy detection using modified stacking, in: Proceedings of the 9th International Conference on Machine Learning and Computing, ACM, 2017, pp. 162–166.
- [37] Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A non-intrusive occupancy monitoring system for demand driven hvac operations, in: *Construction Research Congress 2012: Construction Challenges in a Flat World*, 2012, pp. 828–837.
- [38] Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A systematic approach to occupancy modeling in ambient sensor-rich buildings, *Simulation* 90 (8) (2014) 960–977.
- [39] K. Villarriba, A Neural Network Approach to Detecting the Occupancy State of a Home from Electricity Usage, Master's Thesis, 2014. Mills College.
- [40] H. Rahman, H. Han, Occupancy estimation based on indoor co₂ concentration: comparison of neural network and bayesian methods, *Int. J. Air-Cond. Refrige.* 25 (2017) 1750021, 03.
- [41] A.G. Alam, H. Rahman, J.-K. Kim, H. Han, Uncertainties in neural network model based on carbon dioxide concentration for occupancy estimation, *J. Mech. Sci. Technol.* 31 (5) (2017) 2573–2580.
- [42] Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A multi-sensor based occupancy estimation model for supporting demand driven hvac operations, in: Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design, Society for Computer Simulation International, 2012, p. 2.
- [43] M.K. Masood, Y.C. Soh, V.W.-C. Chang, Real-time occupancy estimation using environmental parameters, in: *Neural Networks (IJCNN)*, 2015 International Joint Conference on, IEEE, 2015, pp. 1–8.
- [44] M.K. Masood, Y.C. Soh, C. Jiang, Occupancy estimation from environmental parameters using wrapper and hybrid feature selection, *Appl. Soft Comput.* 60 (2017) 482–494.
- [45] M.K. Masood, C. Jiang, Y.C. Soh, A novel feature selection framework with hybrid feature-scaled extreme learning machine (hfs-elm) for indoor occupancy estimation, *Energy Build.* 158 (2018) 1139–1151.
- [46] H. Zou, X. Lu, H. Jiang, L. Xie, A fast and precise indoor localization algorithm based on an online sequential extreme learning machine, *Sensors* 15 (1) (2015) 1804–1824.
- [47] H. Zou, H. Jiang, J. Yang, L. Xie, C. Spanos, Non-intrusive occupancy sensing in commercial buildings, *Energy Build.* 154 (2017) 633–643.
- [48] Ö.F. Ertugrul, Y. Kaya, M. EminTagluk, Detecting occupancy of an office room by recurrent extreme learning machines, *Trees (Berl.)* 2 (4) (2016) 17.
- [49] L.M. Candanedo, V. Feldheim, Accurate occupancy detection of an office room from light, temperature, humidity and co₂ measurements using statistical learning models, *Energy Build.* 112 (2016) 28–39.
- [50] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, *Energy Build.* 131 (2016) 132–141.
- [51] Z. Han, R.X. Gao, Z. Fan, Occupancy and indoor environment quality sensing for smart buildings, in: *Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE International, IEEE, 2012, pp. 882–887, 2012.
- [52] A. Akbar, M. Nati, F. Carrez, K. Moessner, Contextual occupancy detection for smart office by pattern recognition of electricity consumption data, in: *Communications (ICC)*, 2015 IEEE International Conference on, IEEE, 2015, pp. 561–566.
- [53] W. Kleiminger, C. Beckel, T. Staake, S. Santini, Occupancy detection from electricity consumption data, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, ACM, 2013, pp. 1–8.
- [54] W. Kleiminger, C. Beckel, S. Santini, Household occupancy monitoring using electricity meters, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2015, pp. 975–986.
- [55] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, S. Santini, The eco data set and the performance of non-intrusive load monitoring algorithms, in: Proceedings of the 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys 2014), Memphis, TN, USA, 2014, pp. 80–89. ACM.
- [56] T. Liu, Y. Li, Z. Bai, J. De, C.V. Le, Z. Lin, S.-H. Lin, G.-B. Huang, D. Cui, Two-stage structured learning approach for stable occupancy detection, in: *Neural Networks (IJCNN)*, 2016 International Joint Conference on, IEEE, 2016, pp. 2306–2312.
- [57] H. Zou, Y. Zhou, J. Yang, C.J. Spanos, Device-free occupancy detection and crowd counting in smart buildings with wifi-enabled iot, *Energy Build.* 174 (2018) 309–322.
- [58] L. Jiang, Z. Cai, D. Wang, S. Jiang, Survey of improving k-nearest-neighbor for classification, in: *Fuzzy Systems and Knowledge Discovery, 2007*, pp. 679–683. FSKD 2007. Fourth International Conference on, Vol. 1, IEEE, 2007.
- [59] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an rfid based system for demand-driven hvac operations, *Autom. Construct.* 24 (2012) 89–99.
- [60] R.H. Dodier, G.P. Henze, D.K. Tiller, X. Guo, Building occupancy detection through sensor belief networks, *Energy Build.* 38 (9) (2006) 1033–1043.
- [61] P.J. Boait, R. Rylatt, A method for fully automatic operation of domestic heating, *Energy Build.* 42 (1) (2010) 11–16.
- [62] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, H. Hjalmarrsson, K. H. Johansson, Blind identification strategies for room occupancy estimation, in: *Control Conference (ECC)*, 2015 European, IEEE, 2015, pp. 1315–1320.
- [63] Z. Chen, Q. Zhu, M.K. Masood, Y.C. Soh, Environmental sensors-based occupancy estimation in buildings via ihmm-mlr, *IEEE Trans. Ind. Informat.* 13 (5) (2017) 2184–2193.
- [64] P. Liu, S.-K. Nguang, A. Partridge, Occupancy inference using pyroelectric infrared sensors through hidden markov models, *IEEE Sensor. J.* 16 (4) (2016) 1062–1068.
- [65] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, K. Whitehouse, The smart thermostat: using occupancy sensors to save energy in homes, in: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, 2010, pp. 211–224.
- [66] L.M. Candanedo, V. Feldheim, D. Deramaix, A methodology based on hidden markov models for occupancy detection and a case study in a low energy residential building, *Energy Build.* 148 (2017) 327–341.
- [67] J. Chaney, E.H. Owens, A.D. Peacock, An evidence based approach to determining residential occupancy and its role in demand response management, *Energy Build.* 125 (2016) 254–266.
- [68] S.H. Ryu, H.J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, *Build. Environ.* 107 (2016) 1–9.
- [69] V. Becker, W. Kleiminger, Exploring zero-training algorithms for occupancy detection based on smart meter measurements, *Comput. Sci. Res. Dev.* 33 (1–2) (2018) 25–36.
- [70] B. Ai, Z. Fan, R.X. Gao, Occupancy estimation for smart buildings by an auto-regressive hidden markov model, in: *American Control Conference (ACC)* 2014, IEEE, 2014, pp. 2234–2239.
- [71] Y.P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, M.A. Little, Predicting room occupancy with a single passive infrared (pir) sensor through behavior extraction, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 1016–1027.
- [72] M. Miflenkovic, O. Amft, An opportunistic activity-sensing approach to save energy in office buildings, in: Proceedings of the Fourth International Conference on Future Energy Systems, ACM, 2013, pp. 247–258.
- [73] G. Gao, K. Whitehouse, The self-programming thermostat: optimizing setback schedules based on home occupancy patterns, in: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, ACM, 2009, pp. 67–72.
- [74] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, K.H. Johansson, Estimation of building occupancy levels through environmental signals deconvolution, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, ACM, 2013, pp. 1–8.
- [75] A. Szczyrek, M. Maciejewska, A. Wylomanska, R. Zimroz, G. Zak, A. Dolega, Detection of occupancy profile based on carbon dioxide concentration pattern matching, *Measurement* 93 (2016) 265–271.
- [76] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, K.H. Johansson, Regularized deconvolution-based approaches for estimating room occupancies, *IEEE Trans. Autom. Sci. Eng.* 12 (4) (2015) 1157–1168.
- [77] J. Krumm, A.B. Brush, Learning time-based presence probabilities, in: *International Conference on Pervasive Computing*, Springer, 2011, pp. 79–96.
- [78] J.Y. Park, E. Mbata, Z. Nagy, Good to see you again: capture and recapture method on mobile devices to estimate occupancy profiles, in: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Transportation, Cities, 2019, pp. 145–148.
- [79] S. Depatla, A. Muralidharan, Y. Mostofi, Occupancy estimation using only wifi power measurements, *IEEE J. Sel. Area. Commun.* 33 (2015) 1381–1393.
- [80] E. Longo, A.E. Redondi, M. Cesana, Accurate occupancy estimation with wifi and bluetooth/ble packet capture, *Comput. Network.* 163 (2019) 106876.
- [81] M. Jin, R. Jia, C.J. Spanos, Virtual occupancy sensing: using smart meters to indicate your presence, *IEEE Trans. Mobile Comput.* 16 (11) (2017) 3264–3277.
- [82] A. Cardenas, S. Piche, D. Meunier, L. Rueda, K. Agbossou, On the occupancy measurement and analysis for residential applications, in: 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), IEEE, 2019.
- [83] T.H. Pedersen, K.U. Nielsen, S. Petersen, Method for room occupancy detection based on trajectory of indoor climate sensor data, *Build. Environ.* 115 (2017) 147–156.
- [84] G. Ansanay-Alex, Estimating occupancy using indoor carbon dioxide concentrations only in an office building: a method and qualitative assessment, in: REHVA World Congress on Energy Efficient, Smart and Healthy Buildings, CLIMA), 2013, pp. 1–8.
- [85] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, P. Shenoy, Non-intrusive occupancy monitoring using smart meters, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, ACM, 2013, pp. 1–8.

- [86] E. Soltanaghaei, K. Whitehouse, Walksense: classifying home occupancy states using walkway sensing, in: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, ACM, 2016, pp. 167–176.
- [87] F. Wahl, M. Milenkovic, O. Amft, A distributed pir-based approach for estimating people count in office environments, in: 2012 IEEE 15th International Conference on Computational Science and Engineering, IEEE, 2012, pp. 640–647.
- [88] F. Wahl, M. Milenkovic, O. Amft, A green autonomous self-sustaining sensor node for counting people in office environments, in: Education and Research Conference (EDERC), 2012 5th European DSP, IEEE, 2012, pp. 203–207.
- [89] G. Tang, K. Wu, J. Lei, W. Xiao, The meter tells you are at home! non-intrusive occupancy detection via load curve data, in: Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on, IEEE, 2015, pp. 897–902.
- [90] M. Jin, R. Jia, Z. Kang, I.C. Konstantakopoulos, C.J. Spanos, Presencesense: zero-training algorithm for individual presence detection based on power monitoring, in: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, ACM, 2014, pp. 1–10.
- [91] D. Li, B. Balaji, Y. Jiang, K. Singh, A wi-fi based occupancy sensing approach to smart energy in commercial office buildings, in: Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, ACM, 2012, pp. 197–198.
- [92] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: occupancy based hvac actuation using existing wifi infrastructure within commercial buildings, in: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, ACM, 2013, p. 17.
- [93] B. Dong, M.B. Kjærsgaard, M. De Simone, H.B. Gunay, W. O'Brien, D. Mora, J. Dziedzic, J. Zhao, *Sensing and Data Acquisition*, Springer International Publishing, Cham (2018), https://doi.org/10.1007/978-3-319-61464-9_4, 77–105.
- [94] A. Wagner, W. O'Brien, B. Dong, *Exploring Occupant Behavior in Buildings: Methods and Challenges*, Springer, 2018.
- [95] Z. Chen, R. Zhao, Q. Zhu, M.K. Masood, Y.C. Soh, K. Mao, Building occupancy estimation with environmental sensors via cdblstm, *IEEE Trans. Ind. Electron.* 64 (12) (2017) 9549–9559.
- [96] X. Guo, D. Tiller, G. Henze, C. Waters, The performance of occupancy-based lighting control systems: a review, *Light. Res. Technol.* 42 (4) (2010) 415–431.

Online Unsupervised Occupancy Anticipation System Applied to Residential Heat Load Management

LUIS RUEDA¹, KODJO AGBOSSOU¹, (Senior Member, IEEE), NILSON HENAO¹, SOUSSO KELOUWANI², (Senior Member, IEEE), JUAN C. OVIEDO-CEPEDA³, BRICE LE LOSTEC³, SIMON SANSREGRET³, AND MICHAËL FOURNIER³

¹Department of Electrical and Computer Engineering, Hydrogen Research Institute, University of Quebec at Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada

²Department of Mechanical Engineering, Hydrogen Research Institute, University of Quebec at Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada

³Laboratoire des Technologies de l'Énergie, Institut de Recherche Hydro-Québec, Shawinigan, QC G9N 7N5, Canada

Corresponding author: Luis Rueda (luis.rueda@uqtr.ca)

This work was supported in part by the Laboratoire des technologies de l'énergie d'Hydro-Québec, in part by the Natural Science and Engineering Research Council of Canada, and in part by the Foundation of Université du Québec à Trois-Rivières.

ABSTRACT Human preferences and lifestyles significantly impact buildings' energy consumption. Consequently, a better understanding of occupants' behavior is crucial to decrease energy consumption and maintain occupants' comfort. Occupant-centric control (OCC) strategies are effective approaches to fulfil such a purpose. As such, occupancy detection and prediction are of prime importance, particularly to manage Electric Space Heating (ESH) systems, due to the relatively slow dynamics of the temperature in dwellings. This paper proposes an Explicit Duration Hidden Markov Model (EDHMM) for unsupervised online presence detection and a hazard-based approach for occupancy prediction. Moreover, a control strategy using a cost function, weighted by occupancy predictions, and a load-shifting strategy based on time-varying electricity price are put forward. This work initially validates the consistency of the proposed approach by using synthetic data generated by a Monte Carlo simulation. Subsequently, the performance of our framework is compared with previous methods presented in the literature through experimental validation. Results demonstrate that the proposed EDHMM approach is efficient in detecting occupancy states. Besides, the results of the field implementation show the potential of the proposed control strategy to preserve occupants' thermal comfort while decreasing the heating energy consumption.

INDEX TERMS Occupant behavior, occupant-centric control, hidden markov models, survival analysis.

NOMENCLATURE

ABBREVIATIONS

ANN	Artificial Neural Networks
DSM	Demand-Side Management
DSO	Distribution System Operator
EDHMM	Explicit Duration Hidden Markov Model
ESH	Electric Space Heating
GMM	Gaussian Mixture Model
HEMS	Home Energy Management System
HMM	Hidden Markov Models
HVAC	Heating, Ventilation, and Air Conditioning
IML	Interpretable Machine Learning
IoT	Internet of Things

KME	Kaplan-Meier Estimator
KNN	K-nearest Neighbors
KS	Kernel Smoother
MAP	Maximum a Posteriori
MPC	Model Predictive Control
OBC	Occupancy-Based Control
OCC	Occupant-Centric Control
PCA	Principal Component Analysis
SCS	Splitting Conic Solver
SVM	Support Vector Machines
ToU	Time-of-Use
WT	Wavelet Transform

VARIABLES

z	Sequence of latent variable
v	Sequence of measurements from the sensors

The associate editor coordinating the review of this manuscript and approving it for publication was Alon Kuperman.

k	Discrete-time index
s	Occupancy state $\{0, 1\}$
d	Duration
G	Number of Gaussian distributions
π_s	Initial probability distribution
N	Number of days for sliding window
t	Time of day
τ_k	Transition signal at instant k
δ	Kronecker delta
C_{in}	Thermal capacitance of the room
T_{in}	Internal room temperature
T_{ext}	Temperature outside of the house
T_{avg}	Average temperature of the surrounding rooms
R_{ext}	Thermal resistance heat transfer to the exterior
R_{wt}	Thermal resistance of internal walls
Φ_{in}	Entering heat flux due to the heating appliance
Φ_{db}	Thermal gains and losses of the room
β	Predicted occupancy probability
Ω_i	Price of the energy
\tilde{x}_i	Predicted internal room temperature
x_i	Reference temperature (preset by the user)
u_i	Predicted power consumption of ESH system
w_i	Climate variables
x_{min}	Minimum allowable internal temperature
x_{max}	Maximum allowable internal temperature
u_{max}	Maximum power capacity of the heating system

I. INTRODUCTION

Due to the increasing energy consumption of the residential sector, energy efficiency, and grid reliability have become significant concerns for the Distribution System Operators (DSOs), especially in Nordic countries like Canada. In Quebec, where this study is conducted, ESH systems account for about 62% of the yearly household energy consumption [1]. These systems can significantly increase power demand during peak load hours as a great number of customers simultaneously heat their homes [2]. Over the last decades, many electricity suppliers have promoted Demand-Side Management (DSM) strategies based on price [3] and incentive-penalty [4] to exploit energy demand flexibility [5]. These conditions alongside the growing availability of smart appliances have led to the rapid development of Home Energy Management Systems (HEMSs) [6]. Energy consumption in residential buildings is influenced by building characteristics, climate, indoor environment quality, and occupants' behavior [7]. Among these factors, human behavior has been revealed as a major issue towards ensuring the effectiveness of the deployed residential energy management strategies [8]. Occupants' behavior refers to the interaction between occupants and buildings in order to preserve a healthy indoor environment and acquire the desired comfort and security [9], [10]. Occupancy is an essential aspect of the occupants' behavior concept, which deals with tracking the presence/absence of people in a building space [11]. Several occupancy detection methods have been integrated with

control strategies to mainly manage lighting and HVAC systems. These OCC strategies focused on the presence/absence of occupants (occupancy-based controls) have demonstrated up to 30% of energy savings [12]. However, instantaneous occupancy detection is not adequate for ESH systems as they require some time to affect the temperature of a room. Moreover, occupancy is a stochastic process complex to be anticipated and difficult to be directly measured. Therefore, the combination of indoor temperature slow dynamic and the stochastic nature of the occupants' presence is revealed as a challenging problem to be solved.

A. LITERATURE STUDY

Several techniques have been proposed for occupancy detection systems. Rueda *et al.* [13] categorized them into analytical, data-driven, and knowledge-based methods. Among them, data-driven methods, such as Artificial Neural Networks (ANN) [14], Support Vector Machines (SVM) [15], K-nearest neighbors (KNN) [16], and Hidden Markov Models (HMM) [17], are considered as the most utilized ones for occupancy detection [13]. However, supervised learning algorithms, like ANN, SVM, and KNN, need a training data set comprising ground truth occupancy evidence which limits their application in practice. As an alternative approach, some studies [17]–[23] have explored the use of unsupervised learning methods for occupancy detection in which HMM has been one of the most popular ones, demonstrating an accuracy between 80% and 90% [13]. Most of the used HMMs are nonetheless first order which means that they are memoryless stochastic processes and their inherent state duration follows a geometrical distribution that is inadequate to model real-world occupants' behavior [12]. Moreover, the use of time-variant transition probabilities on HMM has been mostly neglected, restricting the temporal dependency of occupancy dynamics representation [24].

Concerning the occupancy prediction methods, Kleiminger *et al.* [25] grouped them into schedule-based and context-aware approaches. *Schedule-based algorithms* predict occupancy using historical data on the household occupancy state. Kleiminger *et al.* [25] and Li and Dong [26] present a comparative performance analysis of the state-of-the-art of these approaches, stressing that predictions of up to 90% accuracy can be achieved. Moreover, algorithms, such as ANN [27], KNN [28], k-means, expectation-maximization [23], and Markov models [26], [29] have been explored as alternative to predict occupancy at horizons from 15 minutes up to 24 hours ahead. *Context-aware approaches* estimate future occupancy states of a home according to its occupants' current context (e.g., location or activity). For instance, Krumm and Brush [30] and Gupta *et al.* [31] use information from GPS to estimate when a person would be at home or away and dynamically control the thermostat of the rooms. Although context-aware approaches are promoted as one of the best-performing techniques for occupancy prediction, they depend on the permanent operation of specific devices by individuals (e.g., smartphone and GPS). As a

result, these methods are more intrusive regarding individuals' lifestyles, and are more sensitive to human errors, such as forgetting the phone at home. Therefore, in this work, schedule-based algorithms are preferred.

Several occupancy detection and prediction methods have been integrated with control strategies, such as model predictive control (MPC), to manage HVAC systems [32]. For instance, in [29], Dobbs *et al.* present an occupancy-predicting control algorithm for HVAC systems of a building and compare their approach with a scheduled controller and an occupancy-triggered controller. It is shown through simulation that the proposed approach can achieve up to 19% of energy savings compared to the scheduled controller. Furthermore, other approaches have reported energy savings between 6% to 17% [25] by combining occupancy information and MPC strategies. However, most of the occupancy-based controls (OBCs) field implementations have been done in academic and office buildings. In fact, only 10% of OCC studies have been performed in residential buildings [33]. Another worth noting aspect is that the main objective of OBC strategies is to reduce energy consumption while few studies have integrated load-shifting demand management strategies to improve user's flexibility [32]. Therefore, more efforts to integrate OBC strategies and time-of-use (ToU) pricing programs are still required.

B. CONTRIBUTION AND ORGANIZATION

In light of the discussed matters, this paper proposes an online unsupervised technique for automatic occupancy detection and prediction in residential buildings. This study falls within the overall objective to optimize residential energy consumption by intelligently managing ESH system (baseboard heaters) concerning the presence of the occupants, preserving their thermal comfort. Considering this, the main contributions of this work are summarized as follows:

- Firstly, an Explicit Duration Hidden Markov Model (EDHMM) is developed for unsupervised online occupancy detection.
- Secondly, a hazard-based model is proposed to predict the occupancy probability of each zone of the residence with a 24-hour horizon.
- Finally, a finite-horizon adaptive optimization approach based on a cost function, weighted by occupancy predictions, and a load-shifting strategy for TOU price is formulated.

Unlike the other existing papers in the literature, this work considers the limitations of Markov models for occupancy detection to consistently capture presence/absence duration and realistically predict individuals' presence. Moreover, this study utilizes an adaptive learning strategy based on a sliding window of fixed size (last N days) to daily retrain the system. It should be noted that only a few studies have integrated OBC techniques with load-shifting strategies to manage ESH systems in dwellings. The proposed approach minimizes human intervention during the system learning process and

provides an online occupancy detection and prediction by thermal zone. Therefore, the HEMS is empowered to detect and anticipate the presence/absence of occupants, allowing it to automatically adjust the thermal preferences of each room, maintaining a balance between comfort and energy cost. An experimental proof of concept of the proposed framework has also been carried out. In this regard, the proposed method has been thoroughly tested for three months during Winter in an apartment located in the province of Quebec in Canada.

The rest of this paper is organized as follows. Section II presents the combination of an EDHMM formulation and a Hazard-based model to detect/predict occupancy. Section III illustrates the use of occupancy predictions, an Interpretable Machine Learning model, and a Finite-Horizon controller to improve the performance of the HEMS in residential environments. Section IV describes the design of the case study for the validation purposes. Section V presents the obtained results through simulation and experiments. Finally, the conclusions of the work are drawn in Section VI.

II. OCCUPANCY DETECTION AND PREDICTION STRATEGY

Environmental variables inside a closed space and power consumption are highly correlated with humans' presence [10]. In this sense, due to the difficulty in directly measuring occupancy, Hidden Markov Models (HMMs), are stressed as a suitable method for modeling household/room occupancy through unsupervised learning [20]. HMMs can effectively model the temporal structure of data. Moreover, they allow the use of *a priori* knowledge and supports the explicit integration of duration as well as the time dependency of states. The latter is indispensable for a realistic representation of the occupancy profiles [24], [34].

Let $\mathbf{v} = v_1, v_2, \dots, v_K$ be a sequence of measurements from the sensors and let $\mathbf{z} = z_1, z_2, \dots, z_K$ be a sequence of latent variable taking values from two possible occupancy states $s = \{0, 1\}$, where K is the length of the sequence. In a HMM, two assumptions are made. First, the probability of a particular state only depends on the previous state, $P(z_{k+1}|z_1, \dots, z_k) = P(z_{k+1}|z_k)$, where k is the discrete-time index. Second, the probability of an observation, only depends on the state that produced the observation, $P(v_k|z_1, \dots, z_k, \dots, z_K, v_1, \dots, v_k, \dots, v_K) = P(v_k|z_k)$.

Nonetheless, literature has revealed that traditional first-order Markov models are not able to consistently predict state duration and thus, their generated occupancy profiles often fail to reflect a realistic behavior [35], [36]. In this regard, this work puts forward an Explicit Duration Hidden Markov Model to effectively handle this issue. In an EDHMM, a hypothesis of medium-term independence is introduced. Such hypothesis implies that the probability of changing a state depends not only on the current state, but also on its duration, as shown in (1).

$$P(z_{k+1}|z_1, z_2, \dots, z_{k-1}, z_k) = P(z_{k+1}|z_k, d_k) \quad (1)$$

where d_k is the duration of the current state. It is considered as a deterministic incremental counter restarted when a

transition is made, expressed by,

$$d_k = \begin{cases} d_{k-1} + 1 & \text{if } z_k = z_{k-1} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Furthermore, as presented in [12], the transition probability between the states can be formulated as,

$$P(z_{k+1}|z_k = s, d_k = d') = \begin{cases} h_s(d') & \text{if } z_{k+1} \neq s \\ 1 - h_s(d') & \text{otherwise} \end{cases} \quad (3)$$

where $h_s(d')$ is the discrete hazard function of state s , given by equation (4), which can be estimated either using a parametric or a non-parametric approach.

$$h_s(d') = \frac{f_s(d')}{S_s(d')} = 1 - \frac{S_s(d' + 1)}{S_s(d')} \quad (4)$$

where $f_s(d')$ and $S_s(d')$ are the state duration probability and the survival function, associated with state s , respectively. Besides, due to the time-dependent dynamics of occupancy [24], the use of a time-variant transition probability A is proposed. Therefore, the hazard function is computed in hourly intervals $t \in [0, 1, \dots, 23]$. This results in a hazard function with the form $h'_s(d')$. The transition probability matrix A for the two-state EDHMM can be represented by,

$$A = \begin{bmatrix} 1 - h'_0(d') & h'_0(d') \\ h'_1(d') & 1 - h'_1(d') \end{bmatrix} \quad (5)$$

where $h'_0(d')$ and $h'_1(d')$ are the discrete hazard functions for the absence ($s = 0$) and presence ($s = 1$) states, respectively.

Following the above-presented mathematical formulation, this work proposes an occupancy detection and prediction approach through a three-step process. First, Section II-A explains the start-up process that solves a cold-start problem to collect the necessary information to initialize the EDHMM learning. Second, Section II-B describes an online detection procedure to detect occupancy at five-minute intervals. Third, Section II-C presents the utilized approach to predict the occupancy probability with a 24-hour horizon.

A. START-UP PROCESS

Since the hidden states are unknown, measured data are used to estimate the transition matrix and emission distributions. For this purpose, the start-up process depicted in FIGURE 1 is proposed.

1) DATA PRE-PROCESSING

Measured signals can contain noise from different sources (e.g., the sensor itself, EMI, and acquisition/communication errors). Therefore, a denoising process based on a kernel smoother (KS) and a decomposition process using the Wavelet transform (WT) are proposed in this work. Kernel smoother is a statistical technique widely used to reduce signals noise [37]. In this work, a Gaussian kernel is used in order to give less weight to the more distant neighbours. Wavelet transform has been widely employed for image denoising, edge, and transient detection [34], [38]. Accordingly, WT is

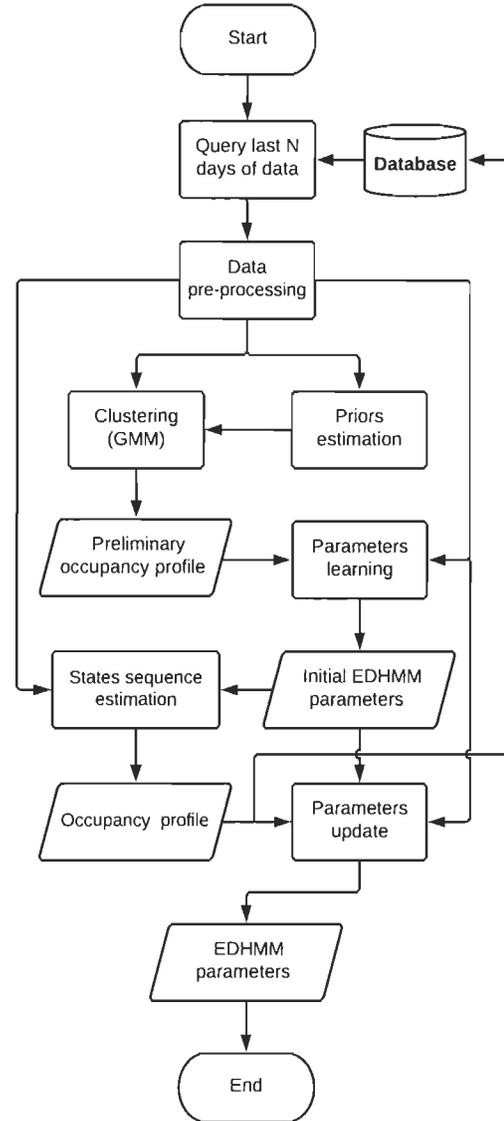


FIGURE 1. Block diagram of the start-up process.

used to decompose the denoised signal into low and high-frequency components, represented by the approximation and the details of the measured variables. Principal Component Analysis (PCA) is also performed to ensure a new set of decorrelated variables [39].

2) CLUSTERING AND PARAMETER LEARNING

Due to the lack of prior knowledge and hidden states information, a Gaussian Mixture Model (GMM) is utilized to estimate a preliminary occupancy profile from the set of observations. GMM is an unsupervised probabilistic model that assumes that all the data points are generated from a mixture of a finite number of Gaussian distributions $\mathcal{N}(v_k | \mu_i, \sigma_i)$, expressed by,

$$f(v_k | \theta) = \sum_{i=1}^G (\phi_i \mathcal{N}(v_k | \mu_i, \sigma_i)) \quad (6)$$

where $\theta = \{[\mu_1, \dots, \mu_G], [\sigma_1, \dots, \sigma_G], [\phi_1, \dots, \phi_G]\}$ denotes the set of parameters, μ_i and σ_i are the mean and variance of the i^{th} Gaussian distribution, G is the number of Gaussian distributions, and ϕ is the mixture component weight, which must satisfy the condition $\sum_{i=1}^G \phi_i = 1$. The probability that an observation belongs to a given distribution is estimated from (7) through,

$$p_i = \frac{\phi_i \mathcal{N}(v_k | \mu_i, \sigma_i)}{\sum_{i=1}^G \phi_i \mathcal{N}(v_k | \mu_i, \sigma_i)}, \quad \text{for } i \in [1, G] \quad (7)$$

where p_i is the probability vector in which each element indicates the probability that an individual belongs to a cluster and G is the number of clusters.

Before applying the GMM, the number of clusters G should be fixed. In this respect, approach based on minimizing the Root-Mean-Square Error (RMSE) between the model estimation and a prior average occupancy profile is employed. Due to the close correlation between the motion information captured by the PIR sensors and occupancy, these data are used as a reference to estimate an initial average profile for each zone. In accordance with the above-explained case, this work assumes that the conditional probability of PIR events given the time of day converges to the likelihood of occupancy at that instant over a long analysis period. Therefore, the average daily motion profile detected by the PIR sensors converges to the actual daily occupancy profile of the dwelling/room. Furthermore, this work presumes a stationary hypothesis which leads to discarding the effects of seasons, vacations, and other long-term behavioral changes. Therefore, as depicted in FIGURE 2, a threshold on the average daily motion for each zone is used to automatically decide on the reference average profile that will be used as the prior information to determine the value of G .

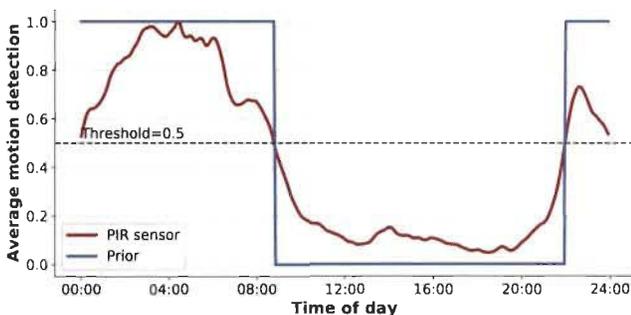


FIGURE 2. Zone 2 priors estimation based on two weeks of motion data.

Once the prior information for each zone is determined, the performance of the GMM is evaluated for a maximum number of 12 clusters, and the value of G that minimizes the RMSE is selected to estimate the preliminary occupancy profile, \tilde{z} .

After estimating \tilde{z} , this profile is used to fit the emission distributions and compute the survival functions. To do so, maximum likelihood estimation algorithm and the Kaplan-Meier estimator are utilized. In the case of emission distributions, three distributions (Normal, Weibull and LogLogistic)

are used to fit the observations, \mathbf{v} , corresponding to each occupancy state, $\mathcal{M} = \{f(v_k | \Theta_s^1), \dots, f(v_k | \Theta_s^n)\}$. Subsequently, the distribution with the best fit $\mathcal{M}_{\hat{m}}$ is calculated using the MLE algorithm, as follows,

$$\hat{m} = \arg \max_m (\mathcal{L}(\mathcal{M}_m)) \quad (8)$$

where \mathcal{L} is the log-likelihood operator, and \mathcal{M}_m represents each of the distributions with their parameters.

FIGURE 3 shows an example of the estimated emission distribution for the first principal component of living room observation. For the case of absent state, the three tested distributions are presented to illustrate the suitability of using MLE algorithm to achieve a good distribution fit to the data.

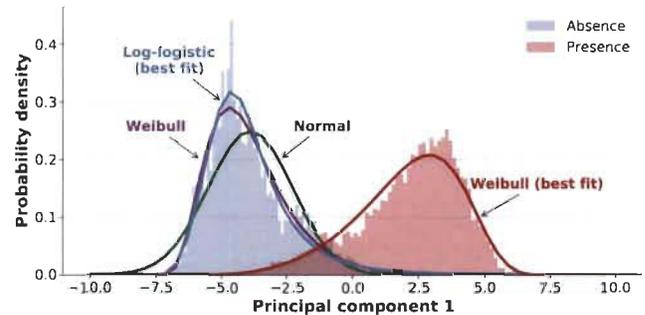


FIGURE 3. Example of estimated emission distribution for zone 1 data.

On the other hand, the non-parametric Kaplan–Meier estimator (KME) is used to compute the survival functions. Unlike the parametric models, KME, expressed by (9), does not restrict the survival function shape to a particular distribution. Therefore, KME assists with a better representation of duration distributions with multimodal shape.

$$S(d') = \prod_{i:d_i < d'} \frac{n_i - e_i}{n_i} \quad (9)$$

where e_i is the number of events (arrive/depart) that happened, d' is the duration, and n_i is the number of presence/absence periods that are still active within d' .

3) STATE SEQUENCE ESTIMATION

Once the learning procedure of the parameters is performed, the most probable occupancy sequence, \mathbf{z} , must be estimated considering the observation sequence, \mathbf{v} . For this purpose, a dynamic programming approach is used that follows the classical procedure of the Viterbi algorithm while explicitly including the state duration to compute the transition probabilities. The proposed decoding algorithm is divided into two stages:

- *Recursive Step:* To estimate the most probable sequence of hidden states at each time-step, the next hypothesis is followed: if the state $s = \{0, 1\}$ is part of the optimal sequence, what would be its best previous state? In this regard, if $\psi_{q,k-1}$ represents the highest probability along a single path for the first $k - 1$ observations, which ends at state $q \in s$, and the transition probabilities from state

q to s are given by the matrix $A_{q,s}$ (equation (5)), then, the best previous state q for the current state s is estimated by solving the following optimization problem,

$$Q_{s,k} = \arg \max_{q \in [0,1]} (\ln(A_{q,s}) + \psi_{q,k-1}) \quad (10)$$

where the initial values $\psi_{s,1}$ for all the states s are estimated by using the initial probability distribution π_s and the likelihood of the first observation v_1 of each state, as follows,

$$\psi_{s,1} = \ln(\pi_s) + \ln(f(v_1 | \Theta_s)) \quad (11)$$

Therefore, knowing the best previous state $\hat{q} = Q_{s,k}$, the probability of the sequence for each state is updated by including the probability of the observation at time-step k , which can be expressed as,

$$\psi_{s,k} = \ln(f(v_k | \Theta_s)) + \ln(A_{\hat{q},s}) + \psi_{\hat{q},k-1} \quad (12)$$

Additionally, in order to estimate transition probabilities, at each time-step k , duration needs to be computed. For this purpose, an incremental counter d is used. This counter is restarted ($d_{s,k} = 1$) when a transition between the best previous state \hat{q} and the current state s is made, based on,

$$d_{s,k} = \delta(\hat{q}, s) d_{\hat{q},k-1} + 1 \quad (13)$$

where $\delta(\hat{q}, s)$ is the Kronecker delta, defined by,

$$\delta(\hat{q}, s) = \begin{cases} 1 & \text{if } \hat{q} = s \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

- **Backtrack the Most Probable Path:** Once the recursive step is completed, the last hidden state is calculated by maximum likelihood, based on,

$$z_K = \arg \max_{s \in [0,1]} (\psi_{s,K}) \quad (15)$$

Then, the most likely sequence is selected by taking the most probable previous state, expressed as,

$$z_k = Q_{z_{k+1},k} \quad k = K-1, K-2, \dots, 2, 1 \quad (16)$$

Algorithm 1 summarizes the decoding procedure to estimate the most probable sequence of states for a finite state machine (FSM) with two-states.

4) PARAMETERS UPDATE

Once the new occupancy sequence \mathbf{z} is estimated, it is used to update the emission and duration distributions by applying the same procedure as presented in Equation 8.

B. ON-LINE PROCESS

After the start-up procedure, as shown in FIGURE 4, two processes are executed to estimate the real-time occupancy state and to perform the model parameters adaptation over time.

Algorithm 1: Occupancy Detection

Input: - Observations sequence: v_1, v_2, \dots, v_K
 - Emission distribution parameters Θ_s
 - Duration distributions

Output: Hidden states sequence: z_1, z_2, \dots, z_K

```

1 begin
2   Define number of states  $s = \{0, 1\}$ 
3   Define initial probability distribution ( $\pi_s$ )
4   Define matrices  $\psi$ ,  $Q$  and  $d$  of dimension  $|s| \times K$ 
5   for  $s \in [0, 1]$  do
6      $\psi_{s,1} = \ln(\pi_s) + \ln(f(v_1 | \Theta_s))$ 
7      $d_{s,1} = 1$ 
8   end
9   for  $k = 2, 3, \dots, K$  do
10    Update matrix  $A$ 
11    for  $s \in [0, 1]$  do
12       $Q_{s,k} = \arg \max_{q \in S} (\ln(A_{q,s}) + \psi_{q,k-1})$ 
13       $\hat{q} = Q_{s,k}$ 
14       $\psi_{s,k} = \ln(f(v_k | \Theta_s)) + \ln(A_{\hat{q},s}) + \psi_{\hat{q},k-1}$ 
15       $d_{s,k} = \delta(\hat{q}, s) d_{\hat{q},k-1} + 1$ 
16    end
17  end
18   $z_K = \arg \max_{s \in [0,1]} (\psi_{s,K})$ 
19  for  $k = K-1, K-2, \dots, 2, 1$  do
20     $z_k = Q_{z_{k+1},k}$ 
21  end
22 end

```

1) ONLINE OCCUPANCY DETECTION

Maximum a posteriori (MAP) estimation is computed at 5-minutes intervals to determine at instant k the most probable hidden state z_k of each of the house zones, expressed as,

$$z_k = \arg \max_{s \in [0,1]} (\ln(f(v_k | \Theta_s)) + \ln(A_{\hat{q},s})) \quad (17)$$

where $\hat{q} = z_{k-1}$ is the estimated occupancy state at the previous instant (5 minutes ago), and v_k is the observations at instant k .

2) PARAMETERS UPDATE

HEMS and occupancy detection systems operate in time-varying environments. Therefore, adaptive capabilities are essential for them. As shown in FIGURE 4, every day at midnight, the system performs an update of the model parameters. To do so, a sliding window with a fixed size of N days is utilized, in which the most recent information is included in the model and the oldest is discarded. It should be noted that studying the appropriate size of the learning window and when new data could be relevant for retraining the model is outside of the scope of this work [40].

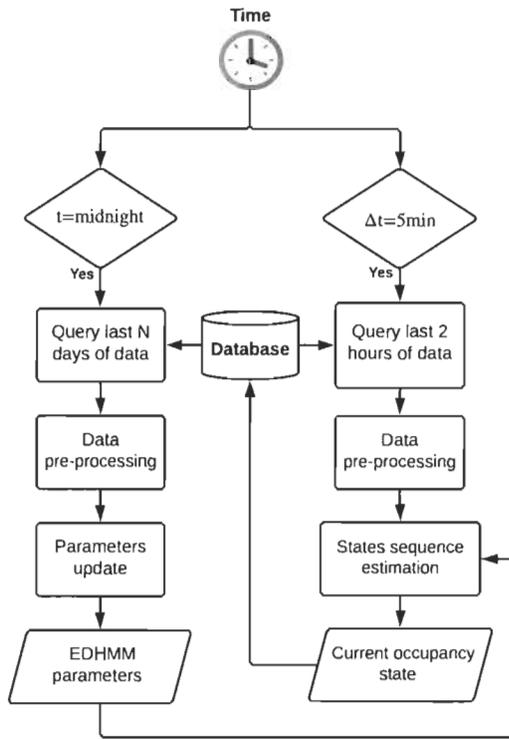


FIGURE 4. Block diagram of the online process.

C. OCCUPANCY PREDICTION

This step aims to predict the presence probability of each zone of the dwelling at a 24-hour horizon, N . This predictive model is an essential step for the control strategy to determine optimal control actions considering future observation forecasts. Accordingly, a hazard-based model and a Monte-Carlo simulation are utilized. Literature study shows that hazard-based models are a suitable generative approach for modeling occupancy profiles and occupants' behavior [12]. In fact, this method has proved its potential for applications, such as in-home and out-of-home activities generation [41], and windows opening and closing behavior modeling [42].

The prediction technique takes advantage of the duration distributions estimated during the learning process of the EDHMM. Likewise, the matrix A (equation (5)) is used to determine the transition probability of the states during the prediction horizon. Algorithm 2 describes the process used for the prediction of the zonal occupancy probability β . In this procedure, the transition signal, $\tau_k \in \{0, 1\}$ is sampled in every time-step using a Bernoulli distribution with $\alpha = h_{\hat{z}_{i-1}}^i(\hat{d})$ (see line 7 of the Algorithm 2). According to this transition signal, the algorithm updates the probable occupancy state, \hat{z}_i by using an XOR function, and restarts the duration counter, \hat{d}_i , which determines the elapsed time in each state. The prediction technique takes as initial state the current occupancy status, estimated by the EDHMM, and computes the current duration according to the information of the last detected transition (arrival/departure). The process is repeated M times, and the average of the generated profiles

Algorithm 2: Occupancy Prediction

Input: - Transition probability matrix A
 - Current occupancy state and duration
Output: Predicted occupancy probability sequence β

```

1 begin
2   Define  $\hat{z}$  as a  $M \times N$  matrix
3   for  $m = 1, 2, \dots, M$  do
4     initialize  $\hat{z}_{m,i=k} = z_k$ 
5     initialize  $\hat{d}_{i=k} = \text{current duration}$ 
6     for  $i = k + 1, k + 2, \dots, k + N$  do
7       sampling  $\tau_i \sim \text{Bern}(h_{\hat{z}_{i-1}}^i(\hat{d} = \hat{d}_{i-1}))$ 
8       update  $\hat{z}_{m,i} = \hat{z}_{m,i-1} \oplus \tau_i$ 
9       update  $\hat{d}_i = (1 - \tau_i)\hat{d}_{i-1} + 1$ 
10    end
11  end
12   $\beta = \frac{1}{M} \sum_{i=k+1}^{k+N} \sum_{m=1}^M \hat{z}_{m,i}$ 
13 end
  
```

is used as the presence probability during the optimization process.

III. FINITE-HORIZON ADAPTIVE CONTROLLER

This work proposes a control strategy to adjust the thermostat set-points of each thermal zone in the residence. By adjusting the thermostat set-points, the control strategy pre-heats the thermal mass of the residence before peak hours, which implicitly modifies the energy consumption patterns. Moreover, it utilizes the thermal preferences of the users and the occupancy predictions to improve the overall performance. The control strategy has two steps of learning and optimizing the ESH management. The first step uses an Interpretable Machine Learning (IML) model to learn the thermal dynamics of the residence using historical data. The second step employs the results of the IML model and the occupancy predictions to formulate an optimization problem that minimizes the costs and guarantees the occupants' comfort within a finite horizon. Section III-A explains the learning procedure of the thermal dynamics, and Section III-B presents the optimization formulation of the proposed controller. Another worth noting feature of the proposed control strategy is the deployment of an event-based adaptation approach which allows the re-estimation of control signals when changes in occupancy states are detected. Section III-C describes the event-based adaptation method.

A. LEARNING OF THE RESIDENCE THERMAL DYNAMICS

In this work, a linear model based on an equivalent Resistance-Capacitance network (2R1C) is used to estimate the thermal dynamics of the house in each thermal zone [43]. Equation (18) presents the energy balance between the heat flux imposed into the environment by the space heaters and the thermal losses of the rooms [2].

$$C_{in}\dot{T}_{in} = \frac{T_{ext} - T_{in}}{R_{ext}} + \frac{T_{avg} - T_{in}}{R_{wl}} + \Phi_{in} + \Phi_{db} \quad (18)$$

where T_{in} is the internal temperature, T_{ext} is the outside temperature, and T_{avg} is the average temperature of the surrounding zones, which allows to account for internal heat exchange between the house rooms. Φ_{in} is the rated electrical power consumption of the heating system (applied thermal flux), and Φ_{db} are the thermal gains and losses of the room (e.g., solar gains and air infiltration). The latter have not been considered in this study. Additionally, C_{in} , R_{ext} and R_{wl} represent the internal thermal mass, the thermal resistance that isolates the building, and the thermal resistance of internal walls (thermal zones divisions), respectively. Therefore, the mathematical expression of the state-space model for a residence with two thermal zones (η_1 and η_2) can be expressed as,

$$\begin{bmatrix} T_{in\eta_1}^{(k+1)} \\ T_{in\eta_2}^{(k+1)} \end{bmatrix} = A \begin{bmatrix} T_{in\eta_1}^{(k)} \\ T_{in\eta_2}^{(k)} \end{bmatrix} + B \begin{bmatrix} \Phi_{in\eta_1}^{(k)} \\ \Phi_{in\eta_2}^{(k)} \end{bmatrix} + C \begin{bmatrix} T_{ext}^{(k)} \end{bmatrix} \quad (19)$$

where $A \in R^{2 \times 1}$, $B \in R^{2 \times 1}$ and $C \in R^{1 \times 1}$ are the parameter matrices that the IML model finds by minimizing the sum-of-squares loss function between Equation (19) and the actual measured output Y over a set of historical data.

B. OPTIMIZATION

To maximize the individual welfare of the customers, a bi-objective cost function is introduced in this study to perform the optimization process. The first term of the cost function guarantees the occupants’ thermal comfort, and the second one minimizes the customer’ payments for any type of dynamic tariff. Occupants’ thermal comfort is expressed by the difference between the predicted temperature $\tilde{x}_i^{\eta a}$ (obtained by Equation (19)), and the reference temperature $x_i^{\eta a}$ (preset by the user). Moreover, the occupancy predictions, $\beta_i^{\eta a}$ are used as weights for the comfort term. This allows the control strategy to guarantee the thermal comfort of the customers when they are actually present in a specific room of the residence. Consequently, a seamlessly integration of the occupancy predictions into the control strategy, as shown in Equation (20), is achieved using this formulation.

$$J(u) = \sum_{i=k}^{k+N} \sum_{a=1}^2 \left(\beta_i^{\eta a} (\tilde{x}_i^{\eta a} - x_i^{\eta a})^2 - \Omega_i u_i^{\eta a} \right) \quad (20)$$

In Equation (20), Ω_i represents the dynamic tariff and $u_i^{\eta a}$ is the predicted power consumption of the ESH. Thereby, considering the cost function and system constraints, Equation (21) shows the complete formulation of the optimization problem.

$$\begin{aligned} & \text{Maximize } J(u) \\ & \text{subject to } \tilde{x}_{i+1} = g(\tilde{x}_i, u_i, w_i) \\ & \tilde{x}_i \in [x_{min}, x_{max}] \end{aligned}$$

$$\begin{aligned} & u_i \in [0, u_{max}] \\ & |\Delta u_i| \leq 0.25 u_{max} \\ & \tilde{x}_0 = \text{Initial state} \\ & u_0 = \text{Initial state} \end{aligned} \quad (21)$$

where x_{min} and x_{max} are the possible allowed temperature interval in the rooms of the house (user-defined), u_{max} is the maximum output power of the ESHs at each thermal zone, and $|\Delta u_i|$ denotes the changes in heating power which is restricted to 25% of u_{max} with the aim of avoiding discomfort to users. Equation (21) follows the rules of Disciplined Convex Programming approach that ensures the convexity of the formulation. It guarantees that if a solution is found, it will be unique and globally optimal [44].

C. EVENT-BASED ADAPTATION

An event-based adaptation strategy is developed in this paper to reduce the risk of discomfort to users when unexpected events occur. This strategy utilizes the changes of the occupancy state detected by the EDHMM as a trigger to re-estimate control actions. FIGURE 5 shows an example of how the adaptation process works. The system plans a series of control actions based on the energy price and the predicted occupancy probability for the control horizon. Accordingly, the system seeks to provide a temperature close to the pre-defined setpoint when the occupancy probability indicates a high likelihood of presence in the house. However, there may be cases in which occupants arrive to/leave the residence earlier than expected. These can cause discomfort since the temperature will be below the predicted value. Hence, to reduce the users’ discomfort and energy waste, this work takes advantage of the information provided by the online occupancy detection system. Accordingly, once an

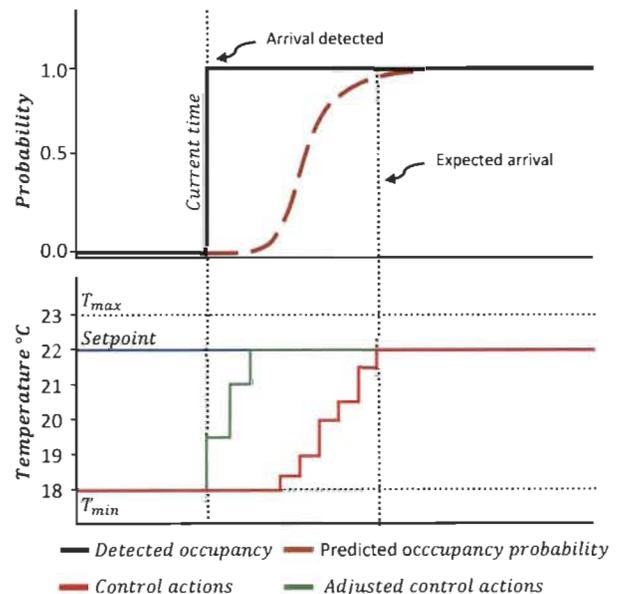


FIGURE 5. Control action adaptation according to occupancy changes detected.

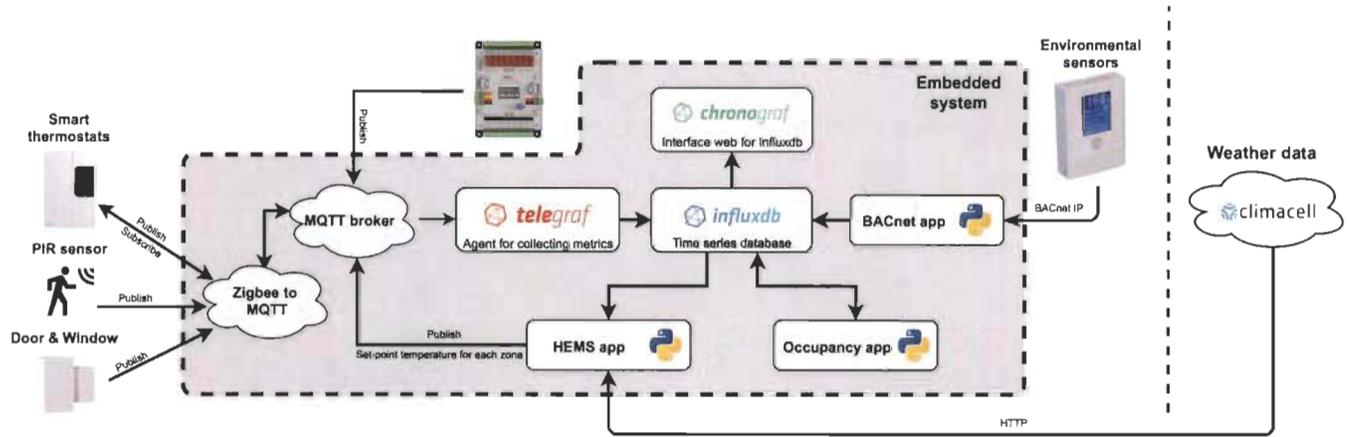


FIGURE 6. Architecture of the sensor network deployed.

arrival/departure is detected, the control system automatically updates its plan.

IV. DESCRIPTION OF THE CASE STUDY

The EDHMM, the IML model, and the finite-horizon adaptive controller are data-driven models. For this reason, the proposed framework uses a network of sensors to monitor the environmental variables and the energy consumption of appliances. From the collected data, the proposed occupancy detection and prediction approach are trained and validated. Moreover, the gathered information is used to tune the thermal model of the house employed on the optimization problem. Hereinafter, Section IV-A presents the design of the experimental setup, and Section IV-B describes the performance metrics of the study.

A. EXPERIMENTAL SETUP

Occupancy detection systems are mainly based on the deployment of environmental sensors (e.g., carbon dioxide (CO₂), temperature, humidity, and light sensors), specialized devices (e.g., PIR sensors, smart meters, and cameras), and other technologies such as WiFi and Bluetooth [13], [45]. Literature reveals that each of these sensors has unique advantages and limitations for occupancy detection applications [16], [46], [47]. Consequently, sensor fusion has been promoted as an efficient way to leverage the strengths of different sensors to improve system performance. In this work, information on CO₂ concentration, movement, relative humidity, temperature, heating, and lighting consumption are collected. For this purpose, the sensors network described in FIGURE 6 has been deployed in an apartment located in the province of Quebec, Canada. TABLE 1 presents the list of the sensors used and a brief description of their specifications. Furthermore, FIGURE 7 shows a floor plan of the apartment with the location of the sensors.

The deployed sensors network leverages the WiFi communication infrastructure already available in place. Message Queuing Telemetry Transport (MQTT) and Hypertext

TABLE 1. Sensors description.

Sensor	Variable	Range	Accuracy
AirLab	CO ₂	3000 ppm	±70 ppm
	Temperature	-30 ~ 70 °C	< ±0.5 °C
	Relative humidity	0 ~ 100 %	5% RH
SNZB-03	Motion	0 ~ 1	NA
SNZB-04	Door/Window contact	0 ~ 1	NA
Fleox Q5	Current	4 ~ 20 mA	-
	Voltage	±60 V	0.05 mV

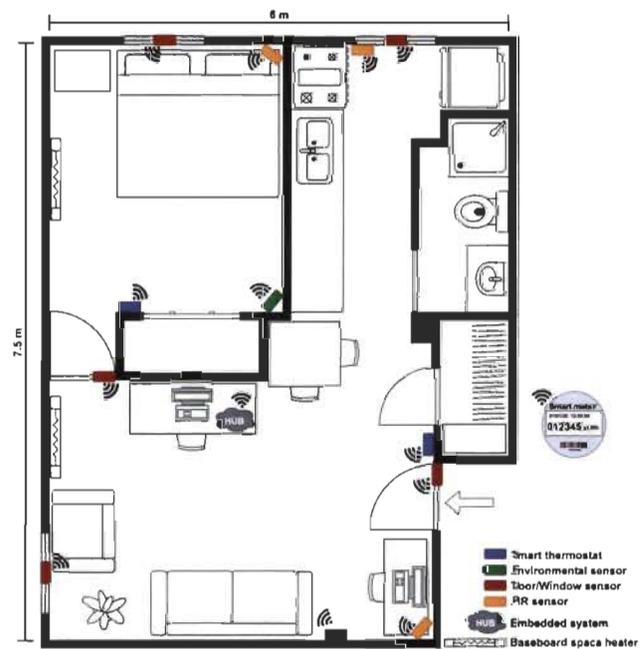


FIGURE 7. Location of the sensors and the smart hub (Raspberry pi 3 B+).

Transfer Protocol (HTTP) are the main communication protocols used for data exchange between the devices and weather data acquisition. Furthermore, Telegraf, Influxdb,

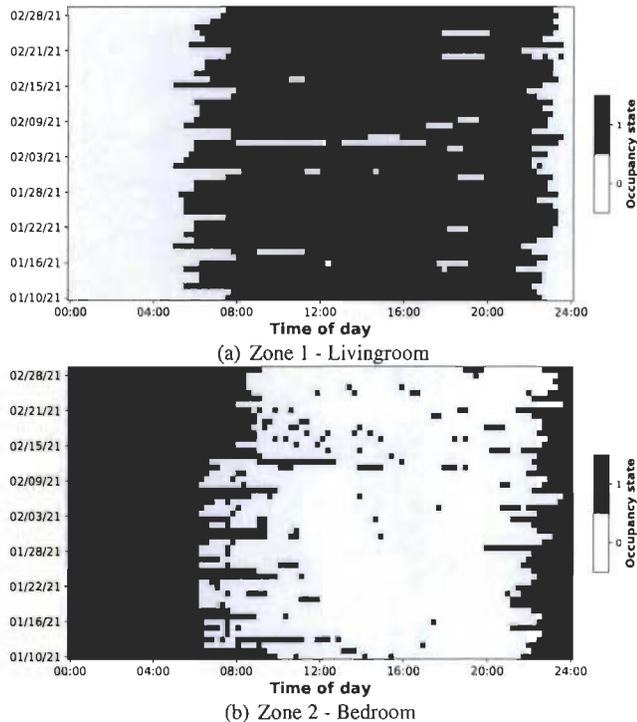


FIGURE 8. Ground truth occupancy data used for validation.

and Chronograf¹ are used for collecting the information transmitted through the MQTT broker, time-series data storage, and as a graphical interface to monitoring the system, respectively. A Python module utilizing the BACnet IP² protocol is also used to get information from the environmental variable sensors. Occupancy app and HEMS app modules, presented in FIGURE 6, correspond to the occupancy detection/prediction model and the controller discussed in Sections II and III, respectively. To validate the performance of the proposed methods, the occupants of the apartment were asked to fill out a form with the actual occupancy. FIGURE 8 illustrates a portion of the ground truth occupancy data for the two zones of the apartment.

It should be noted that the proposed experimental setup is based entirely on open source development tools. Besides, a container-based infrastructure has been adopted to deploy the framework on the embedded system. A Raspberry Pi 3 B+ along with the open platform Docker [50] are used in this regard.

B. PERFORMANCE METRICS

In order to evaluate and compare the performance of the proposed system, the following metrics have been used.

¹Influxdb, Telegraf, and Chronograf are all open-source developments of the Influxdata company, which provide a platform for developing time series applications [48].

²Building Automation and Control Networks (BACnet) is a data communication protocol for building automation and control networks. In a BACnet IP network, its devices use the TCP/IP family of protocols for their communication [49].

- **Accuracy:** This metric is used to quantify the proportion of correct estimations performed by the model. The accuracy can be expressed as,

$$ACC = \frac{1}{K} \sum_{k=1}^K [\hat{z}_k = z_k] \quad (22)$$

where, \hat{z}_k and z_k are the estimated and actual occupancy states sequences, respectively. Moreover, $[\hat{z}_k = z_k]$ is a Iverson bracket, defined by,

$$[\hat{z}_k = z_k] = \begin{cases} 1 & \text{if } \hat{z}_k = z_k \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

- **Precision and Recall:** Precision, also called positive predictive value (PPV), is the fraction of retrieved positive instances among all retrieved instances. Recall, also known as true positive rate (TPR), is the fraction of positive instances retrieved from all the positive cases. These metrics can be expressed as,

$$PPV = \frac{TP}{TP + FP} \quad (24)$$

$$TPR = \frac{TP}{TP + FN} \quad (25)$$

where TP , FP , and FN are the true positives, the false positives, and the false negatives, respectively.

- **F1-Score:** This metric is the harmonic mean of precision and recall, and is approximately the average of the two when they are close. The F1-score can be expressed as,

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} \quad (26)$$

- **MCC:** The Matthews correlation coefficient (MCC) is a balanced measure that can be used even if the classes have very different sizes. The MCC returns a value between -1 and $+1$, where $+1$ represents a perfect prediction, 0 no better than a random prediction, and -1 indicates total disagreement between prediction and observation. This indicator is defined by,

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (27)$$

V. SIMULATION AND EXPERIMENTAL VALIDATION

In this section, numerical examples are provided to illustrate the performance of the proposed framework. In this regard, the collected data over three months of cold weather during 2020-2021 are utilized. Initially, the consistency of the EDHMM under variations in the input data is evaluated through simulation in Section V-A. Afterwards, in Section V-B, the effectiveness of the proposed occupancy detection approach and the control strategy is investigated using an experimental case study.

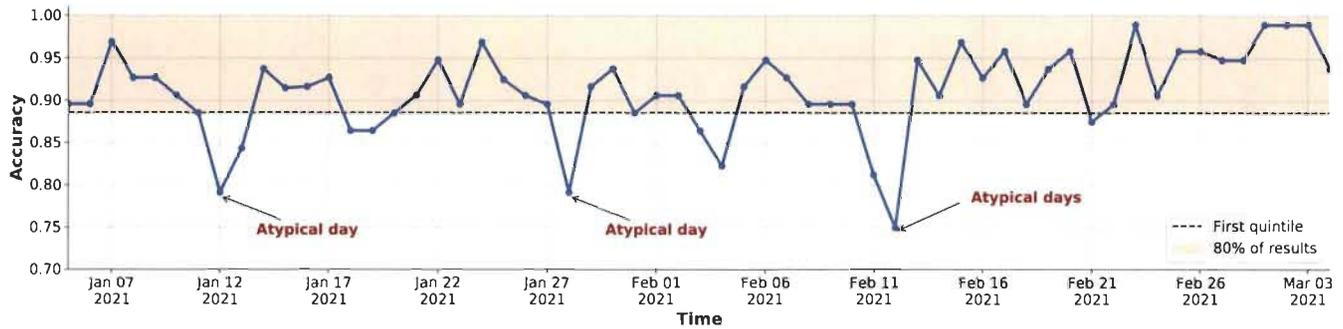


FIGURE 9. Occupancy detection accuracy over time - Zone 2.

TABLE 2. EDHMM performance over 200 simulations.

		Mean	Std	Min	Max
Zone 1	Accuracy	0.958	0.023	0.894	0.971
	Precision	0.963	0.007	0.947	0.979
	Recall	0.973	0.033	0.885	0.991
	F1-score	0.968	0.019	0.915	0.978
	MCC	0.912	0.047	0.781	0.937
Zone 2	Accuracy	0.945	0.013	0.914	0.964
	Precision	0.911	0.025	0.876	0.957
	Recall	0.972	0.017	0.893	0.982
	F1-score	0.940	0.013	0.910	0.959
	MCC	0.898	0.024	0.829	0.927

TABLE 3. Performance comparison.

		HMM	iHMM-MLR	EDHMM
Zone 1	Accuracy	0.8408	0.9056	0.9573
	Precision	0.9161	0.8772	0.9694
	Recall	0.8166	0.9904	0.9647
	F1-score	0.8635	0.9303	0.9671
	MCC	0.6804	0.7998	0.9063
Zone 2	Accuracy	0.8594	0.8787	0.9279
	Precision	0.9311	0.9171	0.8756
	Recall	0.7669	0.8078	0.9766
	F1-score	0.8410	0.8590	0.9233
	MCC	0.7279	0.7579	0.8602

A. SIMULATION ANALYSIS

This work uses the Monte Carlo simulation method to generate 200 synthetic lighting consumption and environmental variable profiles in order to evaluate the consistency of the EDHMM. Thus, the in situ data collected by the sensor network is modeled, assuming they follow a normal distribution. The simulation analysis is carried out for a period of two months of systematic data. Once the data is generated, the EDHMM is used to estimate the occupants' presence. Then, for each of the simulated scenarios, the model performance is evaluated using the metrics described in Section IV-B. TABLE 2 summarizes the obtained results. The results show that the model efficiency for detecting the presence of individuals yields an average accuracy of 95.8% and 94.5% over the 200 cases evaluated for zone 1 and 2, respectively. Furthermore, TABLE 2 illustrates that the difference between the maximum and minimum performance values and the standard deviation is slight. This slight difference confirms the consistency of the model performance under variations in input data.

B. EXPERIMENTAL VALIDATION

1) OCCUPANCY DETECTION

In order to validate the performance of the proposed EDHMM, it is compared with two other generative methods in the literature. The first one is an homogeneous first-order HMM, presented by Candanedo *et al.* [17], and the second one is an in-homogeneous first-order HMM with

multinomial logistic regression (iHMM-MLR), introduced by Chen *et al.* [24]. To perform the validation analysis, this work utilizes the data collected over a period in which windows and internal doors were usually closed. Furthermore, in this case study, the system uses a time window of $N = 21$ days for the learning process, and the occupancy of each zone is detected at 5-minute intervals.

TABLE 3 shows the results of the three approaches under five evaluation criteria. These results illustrate that among the baseline approaches, the iHMM-MLR has a better estimation of occupancy states. It can be attributed to the in-homogeneous character of its transition matrix which captures the occupancy time-variant dependency. Moreover, it is seen that the proposed EDHMM outperforms the state-of-the-art methods under the five criteria, achieving an accuracy up to 95.7%. In fact, this improvement highlights the potential of explicitly integrating the state duration and temporal variability in HMM for applications, such as occupancy state detection. It should be noted that the number of the states of the proposed model can exceed two. Accordingly, the EDHMM can be exploited in other applications, such as activity detection or the estimation of the number of people in residential and commercial buildings.

FIGURE 9 shows the EDHMM daily average accuracy for about two months of data. These findings reveal that for 80% of the testing days, the model accuracy is higher than 88.5%. Furthermore, over the analysis period, some outlier days have been identified in which a decrease in model performance can be observed. On these atypical days, the EDHMM accuracy

has been between 75% and 80%. Such findings are expected since the model has little or no information about these anomalous patterns in the learning data, making it difficult to anticipate these sporadic changes in behavior.

2) CONTROL STRATEGY

As presented in Section III, an occupancy-based control strategy is proposed to control the ESH systems. In fact, for the considered case study, two electric baseboard heaters with nominal power of 1.5kW and 1kW are controlled for zone 1 and 2, respectively. Moreover, to deploy the proposed strategy, this work uses the Python-embedded modeling language for convex optimization (CVXPY) [51] combined with the Splitting Conic Solver (SCS). The latter is selected due to its suitability to solve very large convex cone programs [52]. Furthermore, this work defines a fixed optimization horizon of 24 hours with a discrete time-step of 15 minutes.

Occupancy prediction is a nontrivial part of the control strategy. FIGURE 10 shows an example of the results obtained by the proposed hazard-based approach and the Monte-Carlo procedure while using 500 iterations to predict occupancy. In that figure, the actual occupancy value for zone 2, the predicted occupancy probability (β) for a 24-hour horizon, and the uncertainty of the prediction ($\mu \pm \sigma$) are illustrated. These results show that occupancy probability is a time series that can take values between 0 and 1. However, due to the binary nature of occupancy, a threshold to transform occupancy probability values into a binary signal is often used in the literature [23]. Nevertheless, selecting a suitable threshold value to minimize the prediction error is necessary.

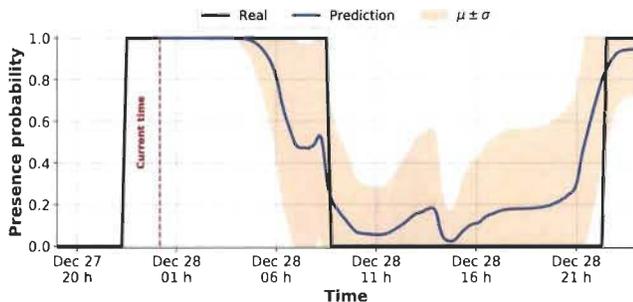


FIGURE 10. Zone 2 occupancy prediction.

An analysis to evaluate the impact of the threshold value on the prediction error is performed. As shown in TABLE 4 and FIGURE 11, the prediction accuracy is considerably affected by the selected threshold value. Furthermore, it is observed that the RMSE of the predicted presence probability is similar or even lower than the one obtained by the optimal threshold value. Consequently, this work uses β to put more or less pressure on the comfort objective of the cost function.

Furthermore, to validate the performance of the proposed control strategy, it is compared with an always-on strategy, which seeks to maintain the reference setpoint at all times, regardless of whether the room is occupied or not. Besides,

TABLE 4. Average RMSE of the occupancy prediction.

	Average	th=0.3	th=0.4	th=0.5	th=0.6	th=0.7
Zone 1	10.81	11.35	10.78	11.68	13.35	15.23
Zone 2	10.03	14.79	12.98	11.27	10.28	10.07

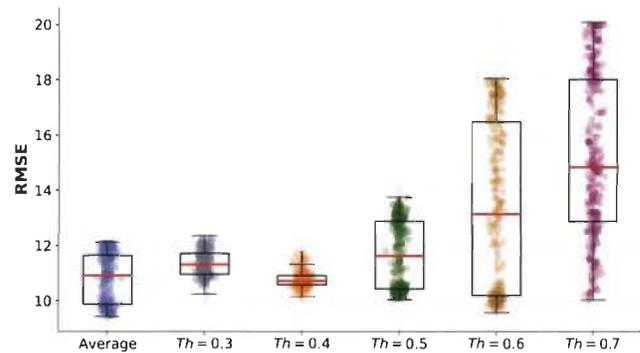


FIGURE 11. Zone 1 occupancy prediction error.

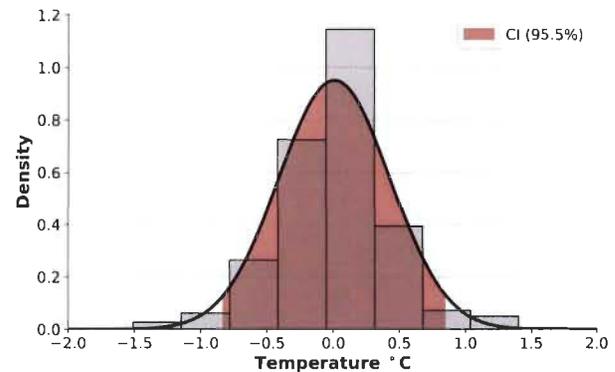


FIGURE 12. Internal temperature error distribution when using the occupancy-based control strategy.

it should be noted that a deterministic control is performed for this work. However, the uncertainty of the occupancy prediction and weather variables are not integrated into this study.

To compare the two control strategies, each of them is applied for several days in the testing apartment. In order to make a fair performance comparison, a similar day strategy is used. Since the focus of this study is thermal comfort and the heating system energy consumption, five days on which the outside temperature maintains a similar daily average (in this case -1°C) are selected for each of the control strategies. Likewise, for each of the analysis periods, weekdays and weekends are included.

TABLE 5 presents the obtained results by the two control strategies. As expected, the minimum discomfort is obtained in the always-on strategy. Although the discomfort MAE and deviation are higher in the proposed OBC strategy, as presented in FIGURE 12, the internal temperature difference is less than $\pm 0.85^\circ\text{C}$ at 95.5% of the time. This result confirms that the proposed approach can guarantee people's comfort.

TABLE 5. Summary of the results according to the control strategy.

	Discomfort (°C)				Energy (kWh)	Final cost (CAD)	Cost reduction (%)
	MAE ($z_1 z_2$)		Variance ($z_1 z_2$)				
Always-on	0.049	0.185	0.091	0.273	33.51	3.08	0%
OBC	0.095	0.292	0.213	0.419	21.78	2.10	31.8%

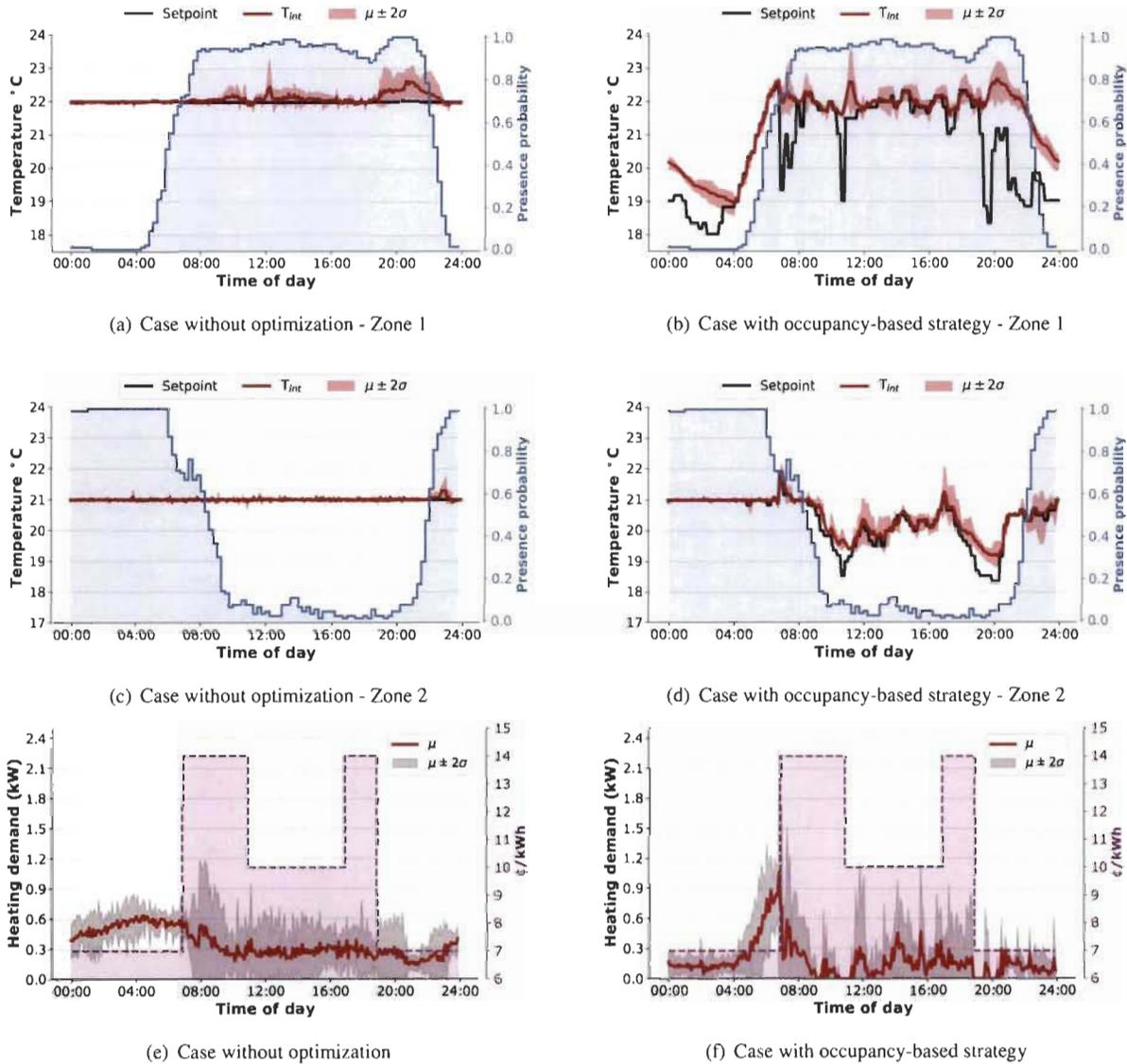


FIGURE 13. Comparison of results without and with occupancy-based optimization strategy.

From TABLE 5, the proposed OBC strategy can reduce the consumed energy for heating by 31.8%. FIGURE 13(a) to FIGURE 13(d) depict that these savings are achieved because the system reduces the room temperature during periods of absence. Moreover, as shown in FIGURE 13(e) and FIGURE 13(f), some of the energy consumption is shifted away from the periods when energy has the highest price. Therefore, the preheating of rooms is performed when it is estimated that a period of presence may begin shortly or if a period of peak cost starts soon. Likewise, the setpoint is reduced when it is predicted that the occupancy will end in

a short time. In this way, energy reduction is achieved while the occupant’s comfort is ensured.

VI. CONCLUSION

This paper presents an unsupervised online system for automatic occupancy detection and prediction applied to a residential space heating control scheme. To do so, an EDHMM is developed for unsupervised online presence detection, and a hazard-based approach is devised for occupancy prediction. Furthermore, a control strategy based on a weighted cost function and a load-shifting strategy based on dynamic

pricing are suggested to enhance the energy and cost economies. Compared to other similar methods, the strengths of the proposed EDHMM lie in its unsupervised learning, the integration of occupancy state duration, and its real-time capability, which are of prime importance for accurate occupancy modeling and its application in real-life scenarios. The deployment of the proposed occupancy-based control strategy has resulted in 35% of energy savings and 31.8% of cost reduction in the analyzed case study. Furthermore, the experimental tests demonstrate that the system has ensured the thermal comfort of the occupants over 95.5% of the time. It is worth noting that the experimental setup has been developed using low-cost embedded systems, open source development tools, and common communication standards for IoT implementations to be conveniently replicable and scalable. The obtained results through simulation and experiments indicate that the features of the proposed occupancy-based control strategy regarding the anticipation of the occupants' presence and use of a dynamic energy price are highly effective in proactively managing ESH systems and minimizing the costs as well as users' discomfort. Moreover, up to three months of simulation and experimental results confirm the model consistency.

In future, it is expected to expand the analysis performed in this article to a scenario with multiple residences, for example, a neighborhood. Therefore, it will be essential to study the design of tariff schemes and identify the number of customers needed to change the load profile in the distribution network without creating new peaks.

REFERENCES

- [1] "Energy fact book 2020–2021," Natural Resour. Canada, Canada, Tech. Rep., 2020. [Online]. Available: <https://www.nrcan.gc.ca>
- [2] W. Devia, K. Agbossou, and A. Cardenas, "An evolutionary approach to modeling and control of space heating and thermal storage systems," *Energy Buildings*, vol. 234, Mar. 2021, Art. no. 110674.
- [3] Z. Chen, L. Wu, and Y. Fu, "Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1822–1831, Dec. 2012.
- [4] A. Fraija, K. Agbossou, N. Henao, and S. Kelouwani, "Peak-to-average ratio analysis of a load aggregator for incentive-based demand response," in *Proc. IEEE 29th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2020, pp. 953–958.
- [5] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.
- [6] B. Zhou, W. Li, K. W. Chan, Y. Cao, Y. Kuang, X. Liu, and X. Wang, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renew. Sustain. Energy Rev.*, vol. 61, pp. 30–40, Aug. 2016.
- [7] H. Yoshino, T. Hong, and N. Nord, "IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods," *Energy Buildings*, vol. 152, pp. 124–136, Oct. 2017.
- [8] Y. Zhang, X. Bai, F. P. Mills, and J. C. V. Pezzey, "Rethinking the role of occupant behavior in building energy performance: A review," *Energy Buildings*, vol. 172, pp. 279–294, Aug. 2018.
- [9] B. F. Balvedi, E. Ghusi, and R. Lamberts, "A review of occupant behaviour in residential buildings," *Energy Buildings*, vol. 174, pp. 495–505, Sep. 2018.
- [10] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, "The impact of occupants' behaviours on building energy analysis: A research review," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1061–1071, Dec. 2017.
- [11] M. Jia, R. S. Srinivasan, and A. A. Raheem, "From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency," *Renew. Sustain. Energy Rev.*, vol. 68, pp. 525–540, Feb. 2017.
- [12] L. Rueda, S. Sansregret, B. Le Lostec, K. Agbossou, N. Henao, and S. Kelouwani, "A probabilistic model to predict household occupancy profiles for home energy management applications," *IEEE Access*, vol. 9, pp. 38187–38201, 2021.
- [13] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," *Building Environ.*, vol. 180, Aug. 2020, Art. no. 106966.
- [14] B. W. Hobson, D. Lowcay, H. B. Gunay, A. Ashouri, and G. R. Newsham, "Opportunistic occupancy-count estimation using sensor fusion: A case study," *Building Environ.*, vol. 159, Jul. 2019, Art. no. 106154.
- [15] W. Kleiminger, C. Beckel, and S. Santini, "Household occupancy monitoring using electricity meters," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2015, pp. 975–986.
- [16] Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A systematic approach to occupancy modeling in ambient sensor-rich buildings," *Simulation*, vol. 90, no. 8, pp. 960–977, Aug. 2014.
- [17] L. M. Candanedo, V. Feldheim, and D. Deramaix, "A methodology based on hidden Markov models for occupancy detection and a case study in a low energy residential building," *Energy Buildings*, vol. 148, pp. 327–341, Aug. 2017.
- [18] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, H. Hjalmarsson, and K. H. Johansson, "Blind identification strategies for room occupancy estimation," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2015, pp. 1315–1320.
- [19] V. Becker and W. Kleiminger, "Exploring zero-training algorithms for occupancy detection based on smart meter measurements," *Comput. Sci.-Res. Develop.*, vol. 33, nos. 1–2, pp. 25–36, Feb. 2018.
- [20] M. Jin, R. Jia, and C. Spanos, "Virtual occupancy sensing: Using smart meters to indicate your presence," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3264–3277, Nov. 2017.
- [21] G. Tang, K. Wu, J. Lei, and W. Xiao, "The meter tells you are at home! Non-intrusive occupancy detection via load curve data," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2015, pp. 897–902.
- [22] M. Jin, R. Jia, Z. Kang, I. C. Konstantakopoulos, and C. J. Spanos, "PresenceSense: Zero-training algorithm for individual presence detection based on power monitoring," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings*, Nov. 2014, pp. 1–10.
- [23] A. Nacer, B. Marhic, L. Delahoche, and J.-B. Masson, "ALOS: Automatic learning of an occupancy schedule based on a new prediction model for a smart heating management system," *Building Environ.*, vol. 142, pp. 484–501, Sep. 2018.
- [24] Z. Chen, Q. Zhu, M. K. Masood, and Y. C. Soh, "Environmental sensors-based occupancy estimation in buildings via IHMM-MLR," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2184–2193, Oct. 2017.
- [25] W. Kleiminger, F. Mattern, and S. Santini, "Predicting household occupancy for smart heating control: A comparative performance analysis of state-of-the-art approaches," *Energy Buildings*, vol. 85, pp. 493–505, Dec. 2014.
- [26] Z. Li and B. Dong, "Short term predictions of occupancy in commercial buildings—Performance analysis for stochastic models and machine learning approaches," *Energy Buildings*, vol. 158, pp. 268–281, Jan. 2018.
- [27] M. C. Mozer, L. Vidmar, and R. H. Dodier, "The neurothermostat: Predictive optimal control of residential heating systems," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 953–959.
- [28] J. Scott, A. J. B. Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar, "PreHeat: Controlling home heating using occupancy prediction," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, 2011, pp. 281–290.
- [29] J. R. Dobbs and B. M. Hency, "Model predictive HVAC control with online occupancy model," *Energy Buildings*, vol. 82, pp. 675–684, Oct. 2014.
- [30] J. Krumm and A. B. Brush, "Learning time-based presence probabilities," in *Proc. Int. Conf. Pervas. Comput.*, Berlin, Germany: Springer, 2011, pp. 79–96.
- [31] M. Gupta, S. S. Intille, and K. Larson, "Adding GPS-control to traditional thermostats: An exploration of potential energy savings and design challenges," in *Proc. Int. Conf. Pervas. Comput.*, Berlin, Germany: Springer, 2009, pp. 95–114.
- [32] H. Stopps, B. Huchuk, M. F. Touchie, and W. O'Brien, "Is anyone home? A critical review of occupant-centric smart HVAC controls implementations in residential buildings," *Building Environ.*, vol. 187, Jan. 2021, Art. no. 107369.

- [33] J. Y. Park, M. M. Ouf, B. Gunay, Y. Peng, W. O'Brien, M. B. Kjergaard, and Z. Nagy, "A critical review of field implementations of occupant-centric building controls," *Building Environ.*, vol. 165, Nov. 2019, Art. no. 106351.
- [34] L. Rueda, A. Cardenas, S. Kelouwani, and K. Agbossou, "Transient event classification based on wavelet neuronal network and matched filters," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 832–837.
- [35] U. Wilke, F. Haldi, J.-L. Scartezzini, and D. Robinson, "A bottom-up stochastic model to predict building occupants' time-dependent activities," *Building Environ.*, vol. 60, pp. 254–264, Feb. 2013.
- [36] G. Flett and N. Kelly, "An occupant-differentiated, higher-order Markov chain method for prediction of domestic occupancy," *Energy Buildings*, vol. 125, pp. 219–230, Aug. 2016.
- [37] M. A. Z. Alvarez, K. Agbossou, A. Cardenas, S. Kelouwani, and L. Boulon, "Demand response strategy applied to residential electric water heaters using dynamic programming and K-means clustering," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 524–533, Jan. 2020.
- [38] X. Wang, "Moving window-based double Haar wavelet transform for image processing," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2771–2779, Sep. 2006.
- [39] I. T. Jolliffe, "Principal components in regression analysis," in *Principal Component Analysis*. New York, NY, USA: Springer, 1986, pp. 129–155.
- [40] D. Toquica, K. Agbossou, R. Malhamé, N. Henao, S. Kelouwani, and A. Cardenas, "Adaptive machine learning for automated modeling of residential prosumer agents," *Energies*, vol. 13, no. 9, p. 2250, May 2020.
- [41] S. Srinivasan and C. R. Bhat, "Modeling household interactions in daily in-home and out-of-home maintenance activity participation," *Transportation*, vol. 32, no. 5, pp. 523–544, 2005.
- [42] F. Haldi and D. Robinson, "Interactions with window openings by office occupants," *Building Environ.*, vol. 44, no. 12, pp. 2378–2395, Dec. 2009.
- [43] E. Atam and L. Helsen, "Control-oriented thermal modeling of multizone buildings: Methods and issues: Intelligent control of a building system," *IEEE Control Syst. Mag.*, vol. 36, no. 3, pp. 86–111, Jun. 2016.
- [44] L. Liberti and N. Maculan, *Global Optimization: From Theory to Implementation*, vol. 84. Springer, Feb. 2006.
- [45] B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng, and H. Fontenot, "Modeling occupancy and behavior for better building design and operation—A critical review," *Building Simul.*, vol. 11, no. 5, pp. 899–921, Oct. 2018.
- [46] T. Labeodan, W. Zeiler, G. Boxem, and Y. Zhao, "Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation," *Energy Buildings*, vol. 93, pp. 303–314, Apr. 2015.
- [47] X. Guo, D. Tiller, G. Henze, and C. Waters, "The performance of occupancy-based lighting control systems: A review," *Lighting Res. Technol.*, vol. 42, no. 4, pp. 415–431, Aug. 2010.
- [48] *InfluxData*. Accessed: Mar. 10, 2021. [Online]. Available: <https://www.influxdata.com/>
- [49] *BACnet, Official Website of ASHRAE SSPC 135*. Accessed: Mar. 10, 2021. [Online]. Available: <http://www.bacnet.org/>
- [50] *Docker*. Accessed: Mar. 10, 2021. [Online]. Available: <https://www.docker.com/>
- [51] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, Jan. 2016.
- [52] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016.



LUIS RUEDA received the B.S. and master's degrees in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2013 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with the Smart Energy Research and Innovation Laboratory, Université du Québec à Trois-Rivières, QC, Canada. His research interests include residential energy management systems, artificial intelligence applications for smart grids,

embedded systems, renewable energies, and occupant behavior modeling in residential buildings.



KODJO AGBOSSOU (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic measurements from the Université de Nancy I, France, in 1987, 1989, and 1992, respectively. He was a Postdoctoral Researcher with the Electrical Engineering Department, Université du Québec à Trois-Rivières (UQTR), from 1993 to 1994, and a Lecturer with the Electrical Engineering Department, from 1997 to 1998. He was the Director of graduate studies in electrical engineering with UQTR, from 2002 to 2004. He was the Head of the Department of Electrical and Computer Engineering, UQTR, from 2007 to 2011. He was the Head of the Engineering School, UQTR, from 2011 to 2017. He is currently the Hydro-Québec Research Chairholder on Transactive Management of Power and Energy in the Residential Sector and the Chair of the Smart Energy Research and Innovation Laboratory, UQTR, where he is a member of the Hydrogen Research Institute and Research Group "GREI." He is the author of more than 325 publications and has four patents and two patent pending. His current research interests include the areas of renewable energy, the use of hydrogen, home demand side management (HDSM), integration of energy production, storage and electrical energy generation systems, connection of electrical vehicle to the grid, and control and measurements. Since 2015, he has been the Sub-Committee Chair on home and building energy management of the Smart Grid Technical Committee, IEEE Industrial Electronics Society (IES).



NILSON HENAO received the B.S. degree in electronics engineering from the Universidad de los Llanos, Villavicencio, Colombia, in 2010, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Quebec at Trois Rivières (UQTR), Trois-Rivières, QC, Canada, in 2013 and 2018, respectively. His research interests include statistical and machine learning methods with applications to energy management in the residential sector, distributed optimization, multi-agent control, smart grid technologies, intelligent energy planning, smart energy storage, and load monitoring.



SOUSSO KELOUWANI (Senior Member, IEEE) received the Ph.D. degree in robotics systems from the Ecole Polytechnique de Montreal, in 2011.

He completed a postdoctoral internship on fuel cell hybrid electric vehicles at the Université du Québec à Trois-Rivières (UQTR), in 2012. Since 2017, he has been a Full Professor of mechatronics with the Department of Mechanical Engineering. He is currently a member of the Hydrogen Research Institute. He holds four patents in USA and Canada, in addition to having published more than 100 scientific articles. He is the Holder of Canada Research Chair in energy optimization of intelligent transport systems and the Holder of the Noovelia Research Chair in intelligent navigation of autonomous industrial vehicles. He developed expertise in the optimization and the intelligent control of vehicular applications. His research interests include optimizing energy systems for vehicle applications, advanced driver assistance techniques, and intelligent vehicle navigation taking into account Canadian climatic conditions. He was the Co-President and the President of the Technical Committee of the IEEE International Conferences on Vehicular Power and Propulsion in Chicago, USA, in 2018, and Hanoi, Vietnam, in 2019. He is a member of the Order of Engineers of Quebec. He is the Winner of Canada General Governor Gold Medal, in 2003. In 2017, he received the Environment Prize at the Gala des Grands Prix d'excellence en transport from the Association québécoise du Transport (AQTr) for the development of hydrogen range extenders for electric vehicles. In 2019, his team received the 1st Innovation Prize in partnership with DIVEL, awarded by the Association des Manufacturiers de la Mauricie et Center-du-Québec for the development of an autonomous and natural navigation system.



JUAN C. OVIEDO-CEPEDA received the Ph.D. degree in electrical engineering from the Universidad Industrial de Santander. During his doctoral studies, he completed two research internships at the Université du Québec à Trois-Rivières, Hydrogen Research Institute, Canada, Université de Technologie Belfort-Montbéliard, La Fédération de Recherche FCLAB, France, and collaboration with Ontario Tech University, Canada. His Ph.D. thesis contributed to the sus-

tainable energy topic of the Misión de Sabios organized by the Government of Colombia. He is currently working as an Industry Research Scientist at the Laboratoire des technologies de l'énergie (LTE), Hydro-Quebec. He is an engineer. His research interests include the integration of distributed energy resources, optimal sizing and operation of autonomous microgrids, transactive energy markets, and machine learning applications for smart grids.



BRICE LE LOSTEC received the M.S. degree in science and technologies from the Université de Savoie, France, in 2005, and the Ph.D. degree in mechanical engineering from Sherbrooke University, Canada, in 2010. His research interests include building stock modeling, solar energy, power generation from low grade heat, electrically driven compression heat pump, absorption heat pump, and refrigeration.



SIMON SANSREGRET received the B.A.Sc. and M.A.Sc. degrees in mechanical engineering from the University of Sherbrooke, with a focus on energy. He has been a Researcher with the Laboratoire des technologies de l'énergie (LTE), Hydro-Québec Research Institute, since 2001. He is currently a member of the Ordre des ingénieurs du Québec. His expertise is related to energy efficiency and demand respond in building sector. In recent years, he has been devoted to the

development of simulation tools in order to improve the energy efficiency of commercial and institutional buildings. He was responsible for the development of simulation software called SIMEB, an interface to EnergyPlus Simulation Engine. He has published several scientific articles in connection with the building energy simulation, model calibration, and visualization of building performance data. He contributed to various projects related to energy consumption and demand response in the residential sector. He was on the Board of Directors of the Canadian Chapter of International Building Performance Association (IBPSA-Canada), from 2010 to 2016.



MICHAËL FOURNIER received the B.S. degree in physical engineering from the Ecole Polytechnique de Montréal, Canada, in 2000, and the master's degree in electrical engineering from the Université du Québec a Trois-Rivières, Trois-Rivières, Canada, in 2003. He has been with Greenlight Power Technologies (now Greenlight Innovation) and the Hydrogen Research Institute, where he worked as a Professional Researcher in fuel cell modeling and integration. Since 2006,

he has been a Researcher at the Laboratoire des technologies de l'énergie, Institut de recherche d'Hydro-Québec, Shawinigan, Canada. His current research interests include demand response, power load profile analysis, sustainable energy vectors, energy efficiency, and management of residential loads. He is a member of the l'Ordre des ingénieurs du Québec.

4.4 Synthèse du chapitre

L'objectif de réaliser une étude approfondie de la littérature scientifique en lien avec les techniques de la détection et la modélisation de la présence des occupants dans les bâtiments à partir des mesures sur place a été atteint par l'article de revue présenté dans ce chapitre. À cet égard, nous avons proposé une approche non supervisée pour la détection de l'occupation des logements en utilisant un modèle de Markov caché à durée explicite basée sur des données mesurées sur place. Le modèle s'appuie sur les HMM traditionnels, en incluant explicitement des informations sur la durée à partir de l'utilisation des analyses de survie non paramétriques, notamment l'utilisation de la technique de Kaplan-Meier pour estimer les fonctions de survie. De plus, le modèle est non homogène, faisant varier sa probabilité de transition non seulement en fonction de la durée, mais aussi de l'heure de la journée.

Pour aborder le problème de l'apprentissage non supervisé, nous exploitons la forte corrélation entre les capteurs PIR et la présence humaine. Nous nous basons sur l'hypothèse que la probabilité conditionnelle des événements PIR étant donné l'heure de la journée converge vers la vraie probabilité d'occupation à ce même instant sur une longue période d'analyse. Par conséquent, le profil de mouvement quotidien moyen détecté par les capteurs PIR converge vers le vrai profil d'occupation quotidien du logement ou de la pièce. De plus, nous utilisons une hypothèse de stationnarité, écartant ainsi les effets des saisons, des vacances et d'autres changements de comportement à long terme. De cette façon, les informations de mouvement recueillies par les capteurs PIR sont utilisées pour la génération automatique de connaissance préalable. Ensuite, ces derniers sont utilisés avec un GMM pour estimer une première séquence d'états d'occupation qui est utilisée pour l'initialisation des paramètres du EDHMM. Une fois cette initialisation effectuée, une étape d'optimisation basée sur des méthodes de programmation dynamique est réalisée, en particulier la méthode de Viterbi est utilisée pour réestimer la séquence d'occupation et ainsi mettre à jour les paramètres du modèle.

Une fois l'étape d'apprentissage complétée, l'estimateur du maximum a posteriori (MAP) est utilisé pour effectuer la détection en ligne de l'occupation. De plus, un modèle basé sur le hasard est utilisé pour anticiper la présence d'individus sur un horizon de 24 heures, ce modèle

se sert des distributions de durée dérivées de la construction du EDHMM. Les approches proposées ont été évaluées par simulation et expérimentalement, en comparant sa performance avec d'autres méthodes déjà présentées dans la littérature. Il faut noter que le pas de temps utilisé pour les analyses de validation était de 15 minutes. Ainsi, avec la proposition du modèle de Markov caché à durée explicite et du modèle pour la prévision de l'occupation, le troisième objectif de cette thèse est atteint.

Par ailleurs, le quatrième objectif de la thèse est satisfait à travers le développement d'une stratégie de gestion de la demande résidentielle basée sur les prévisions de la présence d'occupants, pour la gestion des systèmes de chauffage par plinthe électrique en tenant compte d'un prix variable de l'énergie. Les résultats expérimentaux de l'approche proposée mettent en évidence leur potentiel pour réaliser une gestion prédictive du système de chauffage. Ceci est démontré par la capacité du système à préchauffer les pièces lorsqu'une période de présence est anticipée, ou si le prix de l'énergie est élevé. De même, la consigne est réduite lorsqu'il est prévu que l'occupation se termine dans peu de temps. De cette manière, une réduction de l'énergie est atteinte tandis que le confort thermique des occupants est respecté.

Enfin, la conception et la mise en œuvre d'une architecture expérimentale, grâce à laquelle la validation expérimentale des approches proposées a été réalisée sont décrites.

Chapitre 5 - Conclusions et recommandations

Les travaux entrepris dans cette thèse ont été centrés sur la modélisation du comportement des occupants, en particulier l'occupation. Notamment, nous avons abordé ce problème à partir de deux optiques différentes. La première était orientée vers le développement d'un modèle génératif axé sur les applications pour la simulation énergétique des bâtiments résidentiels. La seconde portait sur le développement de modèles d'occupation pour des applications de gestion de la demande dans des environnements réels.

Dans la première partie de ce travail, nous répondons au besoin soulevé à plusieurs reprises dans la littérature de modéliser de manière réaliste le comportement des occupants dans des environnements simulés axés sur l'analyse énergétique des bâtiments. Concrètement, nous abordons les limites des méthodes existantes dans l'état de l'art, concernant la modélisation de la dynamique temporelle des profils d'occupation et d'activités, ainsi que leur variabilité en fonction des caractéristiques de chaque individu. Pour parvenir aux résultats, nous présentons une méthode générative issue de l'analyse statistique bayésienne pour la modélisation des profils de présence des individus acquis à partir des enquêtes d'emploi du temps. Plus particulièrement, dans cette thèse nous avons proposé un modèle de Markov à durée explicite dans lequel les probabilités de transition entre les états sont basées sur la fonction de hasard discrète (fonction de risque) issue des analyses de survie.

De plus, étant donné la variabilité du comportement humaine, les caractéristiques sociodémographiques des individus sont incorporées dans l'analyse de la durée des états. En effet, nous avons formulé une approche pour estimer à chaque pas de temps la probabilité de transition entre les états d'occupation, en la conditionnant non seulement à la durée mais aussi aux variables calendaires et aux caractéristiques de chaque individu. De cette façon, une

modélisation plus consistante de la dynamique temporelle des profils de présence est effectuée, ce qui est constaté par les résultats obtenus lors de l'analyse comparative avec d'autres méthodes existantes. Par ailleurs, grâce à la nature semi-paramétrique de la méthode proposée, les résultats de validation ont montré qu'il est possible d'obtenir un meilleur ajustement des distributions de la durée, par rapport aux autres méthodes de référence de la littérature. Cela est particulièrement évident dans les cas où les distributions ont une forme multimodale.

Compte tenu de ce qui précède, la méthode proposée s'est avérée être une alternative ayant le potentiel de fournir une représentation réaliste des profils d'occupation. Ceci est d'une grande importance au moment de réaliser par exemple la simulation d'un parc résidentiel, où l'hétérogénéité de la consommation énergétique est largement influencée par la diversité du comportement humain.

La deuxième partie de cette thèse met l'accent sur le défi de modéliser l'occupation d'un espace intérieur dans des environnements réels. Pour répondre à cette problématique, nous avons d'abord effectué une revue détaillée de la littérature scientifique reliée à la détection et à la modélisation de l'occupation à partir de mesures sur place. Comme résultats, ce travail nous a permis de contribuer à l'état de l'art à travers un article de revue, qui met en évidence les avantages, les difficultés, les leçons apprises, les applications et les perspectives de travaux futurs qui peuvent être explorés dans ce domaine. De ce fait, le travail réalisé sert non seulement de référence pour notre travail de recherche, mais aussi pour les personnes intéressées par le domaine de la modélisation du comportement.

Suite au travail de revue effectué, nous avons axé la suite de la thèse sur la difficulté d'obtenir explicitement et en temps réel l'information sur l'occupation d'un espace intérieur, ainsi que du défi d'anticiper les heures d'arrivée et de départ des individus. Cette partie du travail est basée sur des mesures *in situ* de l'environnement interne de la maison, de la consommation électrique et du mouvement des personnes, en tant que mesures indirectes de la présence et du comportement humains. En conséquence, nous exploitons le potentiel des chaînes de Markov cachées et de l'analyse de survie comme base pour proposer un EDHMM qui permet d'effectuer la détection non supervisée de l'occupation. Ainsi, nous présentons un modèle basé sur le hasard pour estimer la probabilité de présence dans un horizon de 24 heures. Le EDHMM proposé effectue l'estimation de l'occupation à partir des observations

obtenues d'un réseau de capteurs, en conditionnant explicitement à la durée de l'état actuel et à l'heure de la journée, les probabilités de transition entre les états. Les résultats de la validation du EDHMM ont été présentés dans la section 4.3 (article 3). Ces résultats montrent que l'inclusion de la dynamique temporelle (durée et heure de la journée) dans les HMMs traditionnels est une approche qui a du potentiel pour des applications telles que la détection de l'occupation. Pour le cas étudié, la performance du EDHMM améliore celle d'autres méthodes de référence issues de la littérature. De plus, le modèle a démontré la consistance de ses résultats face aux variations de ses variables d'entrée, qui a été validée par simulation sur 200 scénarios synthétiques générés.

Par ailleurs, en s'appuyant sur le travail effectué dans la première partie de la thèse, les modèles basés sur le hasard sont exploités comme mécanisme pour anticiper la présence des occupants. Cette analyse est d'une grande importance pour la gestion des systèmes de chauffage, en raison de la dynamique lente de la température intérieure des bâtiments. En fonction de ces prévisions, nous proposons une architecture de gestion basée sur l'occupant pour faire la gestion du système de chauffage par plinthe électrique. Ce type de système de chauffage est sélectionné en raison de sa grande importance dans le contexte québécois, ainsi que de sa flexibilité de contrôle à partir de thermostats communicants, qui permettent de moduler la consommation électrique. Les résultats du cas d'étude analysé ont révélé la possibilité de réduire la consommation d'énergie tout en gardant le confort des occupants dans un contexte sans sources d'énergie alternatives ni systèmes de stockage thermique. L'architecture de gestion a la capacité de déplacer la charge en fonction d'un tarif dynamique, ce qui pourrait contribuer à la réduction de la pointe de puissance vue par Hydro-Québec. Cependant, dans un scénario avec un groupe de résidences, une étude sur les profils tarifaires et la coordination de la gestion des maisons doit être réalisée à l'avenir pour éviter la création d'un nouveau pic de puissance.

Une validation exhaustive de chacune des propositions a permis de corroborer leur performance individuellement et dans l'ensemble de l'architecture de gestion. Pour effectuer cette validation, le système de mesurage et de gestion proposé a été installé dans un logement situé dans la province du Québec, ce qui nous a permis d'obtenir les informations nécessaires pour effectuer les analyses dans un cas avec des conditions réelles. Nous considérons que ce

type de développements expérimentaux à faible coût et basés sur des logiciels open source ont un grand potentiel pour l'utilisation lors de campagnes de mesurage impliquant plusieurs ménages.

Ce projet de recherche contribue à une utilisation plus efficace des ressources énergétiques dans les milieux résidentiels, avec une possible application aux systèmes institutionnels, dans le nouveau contexte des REI et par la voie d'une participation active et automatisée des clients.

5.1 Recommandations

D'après les résultats et les limites des travaux accomplis, des possibles travaux à venir dans les domaines reliés à cette thèse peuvent être :

- En raison de la nature stochastique du comportement humain et compte tenu du fait que les méthodes proposées ont été testées dans des études de cas spécifiques, nous suggérons d'explorer les limites des méthodes proposées. De plus, étant donné la difficulté d'accéder à des données détaillées provenant d'un nombre représentatif de clients, l'utilisation de données synthétiques générées par des techniques telles que Monte-Carlo pourrait être une alternative.
- Études conduisant à la proposition et à l'évaluation de nouvelles plate-formes de simulation/cosimulation orientées vers les analyses transactionnelles, incluant simultanément des modèles comportementaux tels que celles proposées dans le cadre de cette thèse.
- Développement de programmes permettant la réalisation de campagnes de mesurage impliquant plusieurs ménages. À travers ces initiatives, il serait possible de valider l'efficacité des HEMSs qui prennent en compte les préférences et les habitudes des utilisateurs, en plus d'autres tendances actuelles du REI comme l'énergie transactionnelle, et l'intégration de la blockchain.
- Études conduisant à la proposition des stratégies pour élargir l'analyse de l'occupation proposée dans cette thèse afin de modéliser d'autres composantes du comportement humain telles que les activités, l'interaction avec les fenêtres et les portes, et les habitudes de consommation. Ces études sont nécessaires afin de mieux gérer la consommation

d'énergie tout en assurant le confort et la qualité optimale de l'air intérieur pour les occupants.

- Analyse de la taille optimale de la fenêtre d'information à utiliser pour l'apprentissage des modèles. De plus, afin d'assurer la robustesse dans l'adaptabilité du modèle, une étude des techniques d'adaptation, des mécanismes d'oubli et des stratégies d'apprentissage est recommandée.
- Développement des outils permettant l'apprentissage automatique et adaptatif des préférences des utilisateurs (par exemple, la température de confort désirée), en réduisant l'utilisation de paramètres prédéfinis qui peuvent conduire à des actions de contrôle susceptibles d'entraîner un inconfort pour les usagers. Cette approche permettrait au système de gestion d'apprendre et de s'adapter en fonction du feedback des utilisateurs, par exemple dans les cas où ceux-ci modifient manuellement les actions de contrôle estimées par le contrôleur.
- Réalisation d'une étude pour déterminer le placement optimal des capteurs environnementaux, ainsi que des capteurs PIR, afin d'analyser l'impact que cela peut avoir sur les performances du modèle de détection et de prévision de la présence.
- Études conduisant à la proposition et évaluation de stratégies visant à élargir l'approche de gestion centrée sur l'occupant, en permettant l'intégration d'utilisateurs de type « producteur/consommateur ». De plus, développez des stratégies de gestion visant plusieurs charges domestiques importantes telles que le chauffage, le chauffe-eau et les véhicules électriques, ainsi que l'intégration de technologies telles que le stockage électrique et thermique.
- Il faut noter qu'en raison des contraintes de temps, la performance de la stratégie de gestion centrée sur l'occupant proposé dans la Section 4.3 n'a été comparée qu'à un scénario de référence avec une température de consigne constante. Toutefois, il est recommandé de réaliser une étude comparative avec d'autres cas de référence plus représentatifs des réglages typiques (p. ex. scheduled setpoints). De cette manière, la performance en termes de coût et de confort de chacune des méthodes pourrait être évaluée de manière plus exhaustive.

5.2 Annexe

Finalement, au-delà du travail déjà discuté dans cette thèse, l'annexe B présente l'analyse du signal de puissance totale et de quelques charges domestiques, à savoir, le chauffage, le chauffe-eau et l'éclairage. Le but de ce travail est d'identifier les événements récurrents et les habitudes de consommation des utilisateurs, ainsi que leur relation avec les périodes de plus forte consommation d'énergie d'une résidence. Les résultats préliminaires de cette analyse révèlent le potentiel de ce type d'approches pour modéliser les habitudes de consommation des utilisateurs. Cela pourrait contribuer à améliorer l'efficacité des HEMSs, pouvant aussi servir par exemple à la détection des événements déclencheurs permettant d'anticiper l'utilisation de certains appareils ou l'arrivée d'une période de forte consommation d'énergie du résidence. De plus, une meilleure compréhension du comportement et des habitudes des utilisateurs en matière de consommation peut aider les utilités et les distributeurs à mettre en œuvre des programmes de réponse à la demande plus efficaces et des services plus personnalisés.

Bibliographie

- [1] S. TATSA et M. en ÉCONOMIQUE, « Modélisation et prévision de la consommation horaire d'électricité au Québec, » 2013.
- [2] S. AMAN, Y. SIMMHAN et V. K. PRASANNA, « Energy management systems : state of the art and emerging trends, » *IEEE Communications Magazine*, t. 51, n° 1, p. 114-119, 2013.
- [3] W. TUSHAR, C. YUEN, B. CHAI, S. HUANG, K. L. WOOD, S. G. KERK et Z. YANG, « Smart grid testbed for demand focused energy management in end user environments, » *IEEE Wireless Communications*, t. 23, n° 6, p. 70-80, 2016.
- [4] HYDRO-QUÉBEC, *SIMEB Simulation énergétique des bâtiments, Données Météo : "Available on line"*, 2019.
- [5] M. NEUKOMM, V. NUBBE et R. FARES, « Grid-interactive efficient buildings technical report series : Overview of research challenges and gaps, » 2019.
- [6] A. FARUQUI, S. SERGICI et A. SHARIF, « The impact of informational feedback on energy consumption—A survey of the experimental evidence, » *Energy*, t. 35, n° 4, p. 1598-1608, 2010.
- [7] ENERDATA, *Global Energy Statistical Yearbook 2018*, <https://yearbook.enerdata.net/>, [Online accessed 27-March-2019].
- [8] Natural Resources Canada (NRCAN), *Faits sur l'électricité*, <https://www.nrcan.gc.ca/energie/faits/electricite/20079>, [Accédé en ligne le 27 mars 2019].
- [9] ———, « Energy Fact Book 2020 - 2021, » Tech. Rep. 2020, [Online]. Available : <https://www.nrcan.gc.ca>.

- [10] W. DEVIA, K. AGBOSSOU et A. CARDENAS, « An Evolutionary Approach to Modeling and Control of Space Heating and Thermal Storage Systems, » *Energy and Buildings*, p. 110 674, 2020.
- [11] HYDRO-QUÉBEC, « Rapport annuel 2010, » rapp. tech., 2010.
- [12] —, « Rapport annuel 2011, » rapp. tech., 2011.
- [13] —, « Rapport annuel 2012, » rapp. tech., 2012.
- [14] —, « Rapport annuel 2013, » rapp. tech., 2013.
- [15] —, « Rapport annuel 2014, » rapp. tech., 2014.
- [16] —, « Rapport annuel 2015, » rapp. tech., 2015.
- [17] —, « Rapport annuel 2016, » rapp. tech., 2016.
- [18] —, « Rapport annuel 2017, » rapp. tech., 2017.
- [19] —, « Rapport annuel 2018, » rapp. tech., 2018.
- [20] —, *Mieux gérer sa consommation par grand froid*, <http://www.hydroquebec.com/residentiel/espace-clients/consommation/consommation-electrique-hiver.html>, [Accédé en ligne le 27 mars 2019].
- [21] P.-O. PINEAU, « L'électricité au Québec : Faire entrer le secteur dans le 21e siècle, » 2014.
- [22] S. Ø. JENSEN, H. MADSEN, R. LOPES, R. G. JUNKER, D. AELENEI, R. LI, S. METZGER, K. B. LINDBERG, A. J. MARSZAL, M. KUMMERT et al., « Annex 67 : Energy Flexible Buildings-Energy Flexibility as a key asset in a smart building future, » 2017.
- [23] M. HU et F. XIAO, « Quantifying uncertainty in the aggregate energy flexibility of high-rise residential building clusters considering stochastic occupancy and occupant behavior, » *Energy*, t. 194, p. 116 838, 2020.

- [24] K. ZHANG, S. TROITZSCH, S.-Y. ZHANG, E. S. P. TEH, L. SUBRAMANIAN et T. MASSIER, « Transactive Energy in an Urban Environment : A Case Study of Local Generation and Flexibility Potentials in a Singaporean Urban District, » *Energy Res*, t. 9, p. 632-992, 2021.
- [25] C. M. CLEVENGER, J. R. HAYMAKER et M. JALILI, « Demonstrating the impact of the occupant on building performance, » *Journal of computing in civil engineering*, t. 28, n° 1, p. 99-102, 2014.
- [26] B. F. BALVEDI, E. GHISI et R. LAMBERTS, « A review of occupant behaviour in residential buildings, » *Energy and Buildings*, 2018.
- [27] E. DELZENDEH, S. WU, A. LEE et Y. ZHOU, « The impact of occupants' behaviours on building energy analysis : A research review, » *Renewable and Sustainable Energy Reviews*, t. 80, p. 1061-1071, 2017.
- [28] B. DONG, D. YAN, Z. LI, Y. JIN, X. FENG et H. FONTENOT, « Modeling occupancy and behavior for better building design and operation—A critical review, » *Building Simulation*, t. 11, n° 5, p. 899-921, 2018.
- [29] D. YAN, W. O'BRIEN, T. HONG, X. FENG, H. B. GUNAY, F. TAHMASEBI et A. MAHDAVI, « Occupant behavior modeling for building performance simulation : Current state and future challenges, » *Energy and Buildings*, t. 107, p. 264-278, 2015.
- [30] International Energy Agency – Energy in BUILDINGS et COMMUNITIES, « Definition and Simulation of Occupant Behavior in Buildings - Annex 66 Final Report, » Tech. Rep. 2018, [Online]. Available : <http://www.annex66.org/>.
- [31] F. STAZI et F. NASPI, « Occupants' Behaviours Impact on Indoor Environment, » in *Impact of Occupants' Behaviour on Zero-Energy Buildings*, Springer, 2018, p. 13-18.
- [32] T. HONG, D. YAN, S. D'OCA et C.-f. CHEN, « Ten questions concerning occupant behavior in buildings : The big picture, » *Building and Environment*, t. 114, p. 518-530, 2017.

- [33] W. O'BRIEN et H. B. GUNAY, « The contextual factors contributing to occupants' adaptive comfort behaviors in offices—A review and proposed modeling framework, » *Building and Environment*, t. 77, p. 77-87, 2014.
- [34] T. HONG, S. D'OCA, W. J. TURNER et S. C. TAYLOR-LANGE, « An ontology to represent energy-related occupant behavior in buildings. Part I : Introduction to the DNAs framework, » *Building and Environment*, t. 92, p. 764-777, 2015.
- [35] T. ZARAKET, « Stochastic activity-based approach of occupant-related energy consumption in residential buildings, » thèse de doct., Ecole Centrale Paris, 2014.
- [36] D. ROBINSON, « Some trends and research needs in energy and comfort prediction, » rapp. tech., 2006.
- [37] U. WILKE, « Probabilistic bottom-up modelling of occupancy and activities to predict electricity demand in residential buildings, » PhD thesis, École Polytechnique Fédérale de Lausanne, 2013.
- [38] J. SERYAK et K. KISSOCK, « Occupancy and behavioral affects on residential energy use, » in *Proceedings of the Solar conference*, American Solar Energy Society ; American Institute of Architects, 2003, p. 717-722.
- [39] A. EMERY et C. KIPPENHAN, « A long term study of residential home heating consumption and the effect of occupant behavior on homes in the Pacific Northwest constructed according to improved thermal standards, » *Energy*, t. 31, n° 5, p. 677-693, 2006.
- [40] D. YAN, T. HONG, B. DONG, A. MAHDAVI, S. D'OCA, I. GAETANI et X. FENG, « IEA EBC Annex 66 : Definition and simulation of occupant behavior in buildings, » *Energy and Buildings*, t. 156, p. 258-270, 2017.
- [41] W. O'BRIEN, A. WAGNER, M. SCHWEIKER, A. MAHDAVI, J. DAY, M. B. KJÆRGAARD, S. CARLUCCI, B. DONG, F. TAHMASEBI, D. YAN et al., « Introducing IEA EBC Annex 79 : Key challenges and opportunities in the field of occupant-centric building design and operation, » *Building and Environment*, t. 178, p. 106-138, 2020.

- [42] T. BUSO, V. FABI, R. K. ANDERSEN et S. P. CORGNATI, « Occupant behaviour and robustness of building design, » *Building and Environment*, t. 94, p. 694-703, 2015.
- [43] M. M. OUF, W. O'BRIEN et H. B. GUNAY, « Optimization of electricity use in office buildings under occupant uncertainty, » *Journal of Building Performance Simulation*, t. 13, n° 1, p. 13-25, 2020.
- [44] W. O'BRIEN et H. B. GUNAY, « Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design, » in *Building Simulation*, Springer, t. 8, 2015, p. 621-636.
- [45] S. SINGH et A. YASSINE, « Mining energy consumption behavior patterns for households in smart grid, » *IEEE Transactions on Emerging Topics in Computing*, t. 7, n° 3, p. 404-419, 2017.
- [46] M. JIA, R. S. SRINIVASAN et A. A. RAHEEM, « From occupancy to occupant behavior : An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency, » *Renewable and Sustainable Energy Reviews*, t. 68, p. 525-540, 2017.
- [47] H. S. CHO, T. YAMAZAKI et M. HAHN, « AERO : extraction of user's activities from electric power consumption data, » *IEEE Transactions on Consumer Electronics*, t. 56, n° 3, p. 2011-2018, 2010.
- [48] L. RUEDA, S. SANSREGRET, B. LE LOSTEC, K. AGBOSSOU, N. HENAO et S. KELOUWANI, « A Probabilistic Model to Predict Household Occupancy Profiles for Home Energy Management Applications, » *IEEE Access*, t. 9, p. 38 187-38 201, 2021.
- [49] T. A. NGUYEN et M. AIELLO, « Energy intelligent buildings based on user activity : A survey, » *Energy and buildings*, t. 56, p. 244-257, 2013.
- [50] J. F. NICOL et M. A. HUMPHREYS, « A stochastic approach to thermal comfort-occupant behavior and energy use in buildings/discussion, » *ASHRAE transactions*, t. 110, p. 554, 2004.

- [51] Y CHIOU, « An occupant-driven integrative household system for residential building energy simulation, » thèse de doct., Ph. D. dissertation, School of Architecture, Carnegie Mellon Univ . . . , 2009.
- [52] D. TORSTENSSON et F. WALLIN, « Potential and barriers for demand response at household customers, » *Energy Procedia*, t. 75, p. 1189-1196, 2015.
- [53] J. WHITMORE et P.-O. PINEAU, « État de l'énergie au Québec 2021, » Chaire de gestion du secteur de l'énergie, HEC Montréal, préparé pour Transition énergétique Québec.
- [54] M. OSMAN et M. OUF, « A comprehensive review of time use surveys in modelling occupant presence and behavior : Data, methods, and applications, » *Building and Environment*, p. 107785, 2021.
- [55] É. VORGER, « Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment, » PhD thesis, École nationale supérieure des mines de Paris, 2014.
- [56] L. RUEDA, K. AGBOSSOU, A. CARDENAS, N. HENAO et S. KELOUWANI, « A comprehensive review of approaches to building occupancy detection, » *Building and Environment*, p. 106966, 2020.
- [57] F. AMARA, K. AGBOSSOU, Y. DUBÉ, S. KELOUWANI, A. CARDENAS et S. S. HOSSEINI, « A residual load modeling approach for household short-term load forecasting application, » *Energy and Buildings*, t. 187, p. 132-143, 2019.
- [58] S. SANSREGRET, K. LAVIGNE, B. LE LOSTEC, F. LAURENCELLE et F. GUAY, « High-resolution bottom-up residential electrical model for distribution networks planning, » in *Proceedings of eSim Conference 2019*, 2019.
- [59] V. FABI, « Influence of Occupant's Behaviour on Indoor Environmental Quality and Energy Consumptions. A Roadmap to predict the unpredictable Energy-related occupant behaviour, » thèse de doct., Tesi di Dottorato, Politecnico di Torino, Torino, 2013.

- [60] D. AERTS, « Occupancy and activity modelling for building energy demand simulations, comparative feedback and residential electricity demand characterisation, » PhD thesis, Vrije Universiteit Brussel, 2015.
- [61] S. GUY et E. SHOVE, *The sociology of energy, buildings and the environment : Constructing knowledge, designing practice*. Routledge, 2014.
- [62] H. HOU, J. PAWLAK, A. SIVAKUMAR, B. HOWARD et J. POLAK, « An approach for building occupancy modelling considering the urban context, » *Building and Environment*, t. 183, p. 107-126, 2020.
- [63] Y. ZHANG, X. BAI, F. P. MILLS et J. C. PEZZEY, « Rethinking the role of occupant behavior in building energy performance : A review, » *Energy and Buildings*, 2018.
- [64] S. SRINIVASAN et C. R. BHAT, « Modeling household interactions in daily in-home and out-of-home maintenance activity participation, » *Transportation*, t. 32, n° 5, p. 523-544, 2005.
- [65] J. TORRITI, « A review of time use models of residential electricity demand, » *Renewable and Sustainable Energy Reviews*, t. 37, p. 265-272, 2014.
- [66] M. J. ROORDA, E. J. MILLER et K. M. HABIB, « Validation of TASHA : A 24-h activity scheduling microsimulation model, » *Transportation Research Part A : Policy and Practice*, t. 42, n° 2, p. 360-375, 2008.
- [67] G. BUTTITTA, O. NEU, W. J. TURNER et D. FINN, « Modelling household occupancy profiles using data mining clustering techniques on time use data, » in *Building Simulation 2017, San Francisco, CA, August 7-9 2017*, 2017.
- [68] STATISTICS CANADA, *General Social Survey - Time Use (GSS)*, <https://www.statcan.gc.ca>, accessed : 2020-05-01.
- [69] I. RICHARDSON, M. THOMSON et D. INFELD, « A high-resolution domestic building occupancy model for energy demand simulations, » *Energy and buildings*, t. 40, n° 8, p. 1560-1566, 2008.
- [70] OFFICE FOR NATIONAL STATISTICS et IPSOS-RSL, *United Kingdom Time Use Survey, 2000*, 2003. DOI : 10.5255/UKDA-SN-4504-1.

- [71] I. RICHARDSON, M. THOMSON, D. INFELD et C. CLIFFORD, « Domestic electricity use : A high-resolution energy demand model, » *Energy and buildings*, t. 42, n° 10, p. 1878-1887, 2010.
- [72] J. WIDÉN, A. M. NILSSON et E. WÄCKELGÅRD, « A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand, » *Energy and Buildings*, t. 41, n° 10, p. 1001-1012, 2009.
- [73] J. WIDÉN et E. WÄCKELGÅRD, « A high-resolution stochastic model of domestic activity patterns and electricity demand, » *Applied energy*, t. 87, n° 6, p. 1880-1892, 2010.
- [74] M. MURATORI, M. C. ROBERTS, R. SIOSHANSI, V. MARANO et G. RIZZONI, « A highly resolved modeling technique to simulate residential power demand, » *Applied Energy*, t. 107, p. 465-473, 2013.
- [75] A. J. COLLIN, G. TSAGARAKIS, A. E. KIPRAKIS et S. MCLAUGHLIN, « Development of low-voltage load models for the residential load sector, » *IEEE Transactions on Power Systems*, t. 29, n° 5, p. 2180-2188, 2014.
- [76] M. L. BAPTISTA, H. PRENDINGER, R. PRADA et Y. YAMAGUCHI, « A cooperative multi-agent system to accurately estimate residential energy demand, » in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, p. 1405-1406.
- [77] J. CONLISK, « Interactive markov chains, » *Journal of Mathematical Sociology*, t. 4, n° 2, p. 157-185, 1976.
- [78] U. WILKE, F. HALDI, J.-L. SCARTEZZINI et D. ROBINSON, « A bottom-up stochastic model to predict building occupants' time-dependent activities, » *Building and Environment*, t. 60, p. 254-264, 2013.
- [79] G. FLETT et N. KELLY, « An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy, » *Energy and Buildings*, t. 125, p. 219-230, 2016.

- [80] J. TANIMOTO, A. HAGISHIMA et H. SAGARA, « A methodology for peak energy requirement considering actual variation of occupants' behavior schedules, » *Building and Environment*, t. 43, n° 4, p. 610-619, 2008.
- [81] G. FLETT, « Modelling and analysis of energy demand variation and uncertainty in small-scale domestic energy systems, » PhD thesis, University of Strathclyde, 2017.
- [82] D AERTS, J MINNEN, I GLORIEUX, I WOUTERS et F DESCAMPS, « A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, » *Building and environment*, t. 75, p. 67-78, 2014.
- [83] M. MILENKOVIC et O. AMFT, « An opportunistic activity-sensing approach to save energy in office buildings, » in *Proceedings of the fourth international conference on Future energy systems*, ACM, 2013, p. 247-258.
- [84] F. HALDI et D. ROBINSON, « Interactions with window openings by office occupants, » *Building and Environment*, t. 44, n° 12, p. 2378-2395, 2009.
- [85] S. PETERSEN, T. H. PEDERSEN, K. U. NIELSEN et M. D. KNUDSEN, « Establishing an image-based ground truth for validation of sensor data-based room occupancy detection, » *Energy and Buildings*, t. 130, p. 787-793, 2016.
- [86] B. W. HOBSON, D. LOWCAY, H. B. GUNAY, A. ASHOURI et G. R. NEWSHAM, « Opportunistic occupancy-count estimation using sensor fusion : A case study, » *Building and environment*, t. 159, p. 106-154, 2019.
- [87] Z. CHEN, C. JIANG et L. XIE, « Building occupancy estimation and detection : A review, » *Energy and Buildings*, t. 169, p. 260-270, 2018.
- [88] T. LABEODAN, W. ZEILER, G. BOXEM et Y. ZHAO, « Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation, » *Energy and Buildings*, t. 93, p. 303-314, 2015.
- [89] T. H. PEDERSEN, K. U. NIELSEN et S. PETERSEN, « Method for room occupancy detection based on trajectory of indoor climate sensor data, » *Building and Environment*, t. 115, p. 147-156, 2017.

- [90] D. T. DELANEY, G. M. O'HARE et A. G. RUZZELLI, « Evaluation of energy-efficiency in lighting systems using sensor networks, » in *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ACM, 2009, p. 61-66.
- [91] H. ZOU, Y. ZHOU, J. YANG et C. J. SPANOS, « Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT, » *Energy and Buildings*, t. 174, p. 309-322, 2018.
- [92] Z. YANG, N. LI, B. BECERIK-GERBER et M. OROSZ, « A systematic approach to occupancy modeling in ambient sensor-rich buildings, » *Simulation*, t. 90, n° 8, p. 960-977, 2014.
- [93] X GUO, D. TILLER, G. HENZE et C. WATERS, « The performance of occupancy-based lighting control systems : A review, » *Lighting Research & Technology*, t. 42, n° 4, p. 415-431, 2010.
- [94] H. STOPPS, B. HUCHUK, M. F. TOUCHIE et W. O'BRIEN, « Is anyone home? A critical review of occupant-centric smart HVAC controls implementations in residential buildings, » *Building and Environment*, p. 107 369, 2020.
- [95] W. KLEIMINGER, C. BECKEL et S. SANTINI, « Household occupancy monitoring using electricity meters, » in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2015, p. 975-986.
- [96] L. M. CANDANEDO, V. FELDHEIM et D. DERAMAIX, « A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building, » *Energy and Buildings*, t. 148, p. 327-341, 2017.
- [97] M. JIN, R. JIA et C. J. SPANOS, « Virtual occupancy sensing : Using smart meters to indicate your presence, » *IEEE Transactions on Mobile Computing*, t. 16, n° 11, p. 3264-3277, 2017.
- [98] V. BECKER et W. KLEIMINGER, « Exploring zero-training algorithms for occupancy detection based on smart meter measurements, » *Computer Science-Research and Development*, t. 33, n° 1, p. 25-36, 2018.

- [99] A. NACER, B. MARHIC, L. DELAHOUCHE et J.-b. MASSON, « ALOS : Automatic learning of an occupancy schedule based on a new prediction model for a smart heating management system, » *Building and Environment*, t. 142, p. 484-501, 2018.
- [100] M. JIN, R. JIA, Z. KANG, I. C. KONSTANTAKOPOULOS et C. J. SPANOS, « Presencesense : Zero-training algorithm for individual presence detection based on power monitoring, » in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, ACM, 2014, p. 1-10.
- [101] G. TANG, K. WU, J. LEI et W. XIAO, « The meter tells you are at home! non-intrusive occupancy detection via load curve data, » in *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on*, IEEE, 2015, p. 897-902.
- [102] J. PAGE, D. ROBINSON, N. MOREL et J.-L. SCARTEZZINI, « A generalised stochastic model for the simulation of occupant presence, » *Energy and buildings*, t. 40, n° 2, p. 83-98, 2008.
- [103] C. W. GELLINGS, « The concept of demand-side management for electric utilities, » *Proceedings of the IEEE*, t. 73, n° 10, p. 1468-1470, 1985.
- [104] L. GELAZANSKAS et K. A. GAMAGE, « Demand side management in smart grid : A review and proposals for future direction, » *Sustainable Cities and Society*, t. 11, p. 22-30, 2014.
- [105] F. E. R. COMMISSION, « Assessment of Demand Response and Advanced Metering, » rapp. tech., 2016.
- [106] R. SHARIFI, S. FATHI et V. VAHIDINASAB, « A review on Demand-side tools in electricity market, » *Renewable and Sustainable Energy Reviews*, t. 72, p. 565-572, 2017.
- [107] ———, « Customer baseline load models for residential sector in a smart-grid environment, » *Energy Reports*, t. 2, p. 74-81, 2016.

- [108] A. NILSSON, M. WESTER, D. LAZAREVIC et N. BRANDT, « Smart homes, home energy management systems and real-time feedback : Lessons for influencing household energy consumption from a Swedish field study, » *Energy and Buildings*, t. 179, p. 15-25, 2018.
- [109] H. SHAREEF, M. S. AHMED, A. MOHAMED et E. AL HASSAN, « Review on home energy management system considering demand responses, smart technologies, and intelligent controllers, » *IEEE Access*, t. 6, p. 24 498-24 509, 2018.
- [110] M. SHAKERI, M. SHAYESTEGAN, H. ABUNIMA, S. S. REZA, M AKHTARUZZAMAN, A. ALAMOUD, K. SOPIAN et N. AMIN, « An intelligent system architecture in home energy management systems (HEMS) for efficient demand response in smart grid, » *Energy and Buildings*, t. 138, p. 154-164, 2017.
- [111] SMA SOLAR TECHNOLOGY AG, « SMA Smart Home, » rapp. tech., 2013.
- [112] G. SERALE, M. FIORENTINI, A. CAPOZZOLI, D. BERNARDINI et A. BEMPORAD, « Model predictive control (MPC) for enhancing building and HVAC system energy efficiency : Problem formulation, applications and opportunities, » *Energies*, t. 11, n° 3, p. 631, 2018.
- [113] P. H. SHAIKH, N. B. M. NOR, P. NALLAGOWNDEN, I. ELAMVAZUTHI et T. IBRAHIM, « A review on optimized control systems for building energy and comfort management of smart sustainable buildings, » *Renewable and Sustainable Energy Reviews*, t. 34, p. 409-429, 2014.
- [114] A. I. DOUNIS et C. CARASCOS, « Advanced control systems engineering for energy and comfort management in a building environment—A review, » *Renewable and Sustainable Energy Reviews*, t. 13, n° 6-7, p. 1246-1261, 2009.
- [115] Z. WU, S. ZHOU, J. LI et X.-P. ZHANG, « Real-time scheduling of residential appliances via conditional risk-at-value, » *IEEE Transactions on Smart Grid*, t. 5, n° 3, p. 1282-1291, 2014.
- [116] Y. DOTE, « Introduction to fuzzy logic, » in *Proceedings of IECON'95-21st Annual Conference on IEEE Industrial Electronics*, IEEE, t. 1, 1995, p. 50-56.

- [117] C. ELMAS, O. DEPERLIOGLU et H. H. SAYAN, « Adaptive fuzzy logic controller for DC–DC converters, » *Expert Systems with Applications*, t. 36, n° 2, p. 1540-1548, 2009.
- [118] C. GUZMAN, « Contrôle et gestion intelligents d'énergie et de puissance dans les systèmes électriques, » PhD thesis, Université du Québec à Trois-Rivières, 2019.
- [119] M. KILLIAN et M. KOZEK, « Ten questions concerning model predictive control for energy efficient buildings, » *Building and Environment*, t. 105, p. 403-412, 2016.
- [120] H. FONTENOT et B. DONG, « Modeling and control of building-integrated microgrids for optimal energy management—a review, » *Applied Energy*, t. 254, p. 113 689, 2019.
- [121] B LOKESHGUPTA et S SIVASUBRAMANI, « Multi-objective home energy management with battery energy storage systems, » *Sustainable Cities and Society*, t. 47, p. 101 458, 2019.
- [122] F. DE ANGELIS, M. BOARO, D. FUSELLI, S. SQUARTINI, F. PIAZZA et Q. WEI, « Optimal home energy management under dynamic electrical and thermal constraints, » *IEEE Transactions on Industrial Informatics*, t. 9, n° 3, p. 1518-1527, 2012.
- [123] J. DOMINGUEZ, A. DANTE, K AGBOSSOU, N HENAO, J CAMPILLO, A CARDENAS et S KELOUWANI, « Optimal Charging Scheduling of Electric Vehicles based on Principal Component Analysis and Convex Optimization, » in *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE)*, IEEE, 2020, p. 935-940.
- [124] J. OVIEDO-CEPEDA, I. SERNA-SUÁREZ, G. OSMA-PINTO, C. DUARTE, J. SOLANO et H. A. GABBAR, « Design of tariff schemes as demand response mechanisms for stand-alone microgrids planning, » *Energy*, t. 211, p. 119 028, 2020.
- [125] A. KHALID, N. JAVAID, M. GUIZANI, M. ALHUSSEIN, K. AURANGZEB et M. ILAHI, « Towards dynamic coordination among home appliances using multi-objective energy optimization for demand side management in smart buildings, » *IEEE access*, t. 6, p. 19 509-19 529, 2018.

- [126] M. A. Z. ALVAREZ, K. AGBOSSOU, A. CARDENAS, S. KELOUWANI et L. BOULON, « Demand response strategy applied to residential electric water heaters using dynamic programming and K-means clustering, » *IEEE Transactions on Sustainable Energy*, t. 11, n° 1, p. 524-533, 2019.
- [127] X. WU, X. HU, X. YIN et S. J. MOURA, « Stochastic optimal energy management of smart home with PEV energy storage, » *IEEE Transactions on Smart Grid*, t. 9, n° 3, p. 2065-2075, 2016.
- [128] D. O'NEILL, M. LEVORATO, A. GOLDSMITH et U. MITRA, « Residential demand response using reinforcement learning, » in *2010 First IEEE international conference on smart grid communications*, IEEE, 2010, p. 409-414.
- [129] A.-H. MOHSENIAN-RAD, V. W. WONG, J. JATSKEVICH, R. SCHOBER et A. LEON-GARCIA, « Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid, » *IEEE transactions on Smart Grid*, t. 1, n° 3, p. 320-331, 2010.
- [130] R. DENG, Z. YANG, M.-Y. CHOW et J. CHEN, « A survey on demand response in smart grids : Mathematical models and approaches, » *IEEE Transactions on Industrial Informatics*, t. 11, n° 3, p. 570-582, 2015.
- [131] J. Y. PARK, M. M. OUF, B. GUNAY, Y. PENG, W. O'BRIEN, M. B. KJÆRGAARD et Z. NAGY, « A critical review of field implementations of occupant-centric building controls, » *Building and Environment*, t. 165, p. 106351, 2019.
- [132] S. BALDI, C. D. KORKAS, M. LV et E. B. KOSMATOPOULOS, « Automating occupant-building interaction via smart zoning of thermostatic loads : A switched self-tuning approach, » *Applied energy*, t. 231, p. 1246-1258, 2018.
- [133] W. KLEIMINGER, F. MATTERN et S. SANTINI, « Predicting household occupancy for smart heating control : A comparative performance analysis of state-of-the-art approaches, » *Energy and Buildings*, t. 85, p. 493-505, 2014.

- [134] J. SCOTT, A. BERNHEIM BRUSH, J. KRUMM, B. MEYERS, M. HAZAS, S. HODGES et N. VILLAR, « PreHeat : controlling home heating using occupancy prediction, » in *Proceedings of the 13th international conference on Ubiquitous computing*, ACM, 2011, p. 281-290.
- [135] M. C. MOZER, L. VIDMAR et R. H. DODIER, « The neurothermostat : Predictive optimal control of residential heating systems, » in *Advances in Neural Information Processing Systems*, 1997, p. 953-959.
- [136] Z. LI et B. DONG, « Short term predictions of occupancy in commercial buildings—Performance analysis for stochastic models and machine learning approaches, » *Energy and Buildings*, t. 158, p. 268-281, 2018.
- [137] J. R. DOBBS et B. M. HENCEY, « Model predictive HVAC control with online occupancy model, » *Energy and Buildings*, t. 82, p. 675-684, 2014.
- [138] J. KRUMM et A. B. BRUSH, « Learning time-based presence probabilities, » in *International Conference on Pervasive Computing*, Springer, 2011, p. 79-96.
- [139] M. GUPTA, S. S. INTILLE et K. LARSON, « Adding gps-control to traditional thermostats : An exploration of potential energy savings and design challenges, » in *International Conference on Pervasive Computing*, Springer, 2009, p. 95-114.
- [140] L. G. SWAN et V. I. UGURSAL, « Modeling of end-use energy consumption in the residential sector : A review of modeling techniques, » *Renewable and sustainable energy reviews*, t. 13, n° 8, p. 1819-1835, 2009.
- [141] A. GRANDJEAN, J. ADNOT et G. BINET, « A review and an analysis of the residential electric load curve models, » *Renewable and Sustainable energy reviews*, t. 16, n° 9, p. 6539-6565, 2012.
- [142] L. G. SWAN, « Residential sector energy and GHG emissions model for the assessment of new technologies, » PhD thesis, Dalhousie University, 2010.
- [143] G. HAPPLE, J. A. FONSECA et A. SCHLUETER, « A review on occupant behavior in urban building energy models, » *Energy and Buildings*, t. 174, p. 276-292, 2018.

- [144] HYDRO-QUÉBEC, *Tarifification dynamique*, <https://www.hydroquebec.com/residentiel/espace-clients/tarifs/tarifification-dynamique.html>, [Accédé en ligne le 14 mai 2021].
- [145] —, *Bilan de la tarification dynamique*, <https://www.hydroquebec.com/residentiel/espace-clients/tarifs/tarifification-dynamique-bilan-collectif.html>, [Accédé en ligne le 14 mai 2021].
- [146] K. GAMMA, R. MAI, C. COMETTA et M. LOOCK, « Engaging customers in demand response programs : The role of reward and punishment in customer adoption in Switzerland, » *Energy Research & Social Science*, t. 74, p. 101-927, 2021.
- [147] OBSERVATOIRE DE LA CONSOMMATION RESPONSABLE ESG UQAM POUR TEQ, « Portrait des citoyens québécois face à la transition énergétique, » rapp. tech., 2020.
- [148] S. Ø. JENSEN, A. MARSZAL-POMIANOWSKA, R. LOLLINI, W. PASUT, A. KNOTZER, P. ENGELMANN, A. STAFFORD et G. REYNDERS, « IEA EBC annex 67 energy flexible buildings, » *Energy and Buildings*, t. 155, p. 25-34, 2017.
- [149] K. SUN et T. HONG, « A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures, » *Energy and Buildings*, t. 146, p. 383-396, 2017.
- [150] CLIMACELL, <https://www.tomorrow.io/>, accessed : 2021-09-12.
- [151] E. ATAM et L. HELSEN, « Control-oriented thermal modeling of multizone buildings : methods and issues : intelligent control of a building system, » *IEEE Control Systems Magazine*, t. 36, n° 3, p. 86-111, 2016.
- [152] C. GUZMAN, L. RUEDA, G. ROMERO, S. BISCANS, K. AGBOSSOU et A. CARDENAS, « Enabling Winter Behavior Analysis on Electrically Heated Residential Buildings by Smart Sub-metering, » *2018 IEEE International Conference on Industrial Technology (ICIT)*, p. 1213-1218, 2018.
- [153] G. W. HART, « Nonintrusive appliance load monitoring, » *Proceedings of the IEEE*, t. 80, n° 12, p. 1870-1891, 1992.

- [154] G. E. BOX, G. M. JENKINS, G. C. REINSEL et G. M. LJUNG, *Time series analysis : forecasting and control*. John Wiley & Sons, 2015.
- [155] S. B. LEEB et J. L. KIRTLEY JR, *Transient event detector for use in nonintrusive load monitoring systems*, US Patent 5,483,153, 1996.
- [156] HYDRO-QUÉBEC, *Répartition de la consommation d'un ménage*, <http://www.hydroquebec.com/residentiel/espace-clients/consommation/consommation-electrique-sources.html>, [Accédé en ligne le 24 juin 2019].
- [157] N. R. CANADA, « Water Heater Guide, » rapp. tech., 2012.
- [158] A. ZUNIGA, « Gestion du stockage thermique résidentiel dans le contexte de la gestion de la demande : chauffe-eau, » mém. de mast., Université du Québec à Trois-Rivières, 2018.
- [159] Y. WANG, Q. CHEN, T. HONG et C. KANG, « Review of smart meter data analytics : Applications, methodologies, and challenges, » *IEEE Transactions on Smart Grid*, t. 10, n° 3, p. 3125-3148, 2018.
- [160] G.W. HART, « Residential energy monitoring and computerized surveillance via utility power flows, » *IEEE Technology and Society Magazine*, t. 8, n° 2, p. 12-16,
- [161] K. D. ANDERSON, M. E. BERGÉS, A. OCNEANU, D. BENITEZ et J. M. MOURA, « Event detection for non intrusive load monitoring, » in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*, IEEE, 2012, p. 3312-3317.

Annexe A - Système de mesure et de contrôle déployé

Cette annexe présente des informations complémentaires sur le système de mesure et l'architecture de contrôle utilisée dans cette thèse. De plus, le formulaire utilisé pour recueillir les valeurs réelles d'occupation pendant la période d'analyse y est décrit.

A.1 Système de mesure

La FIGURE A-1 montre l'architecture du système de mesure et de contrôle qui a été mise en place pour la validation expérimentale réalisée dans cette thèse. L'architecture présentée permet de mettre en évidence chacun de ses composants :

- Capteurs : Le système déployé est composé de capteurs PIR, de capteurs d'ouverture/fermeture de portes et de fenêtres, ainsi que de capteurs permettant de mesurer des variables environnementales telles que la concentration en CO₂, l'humidité relative et la

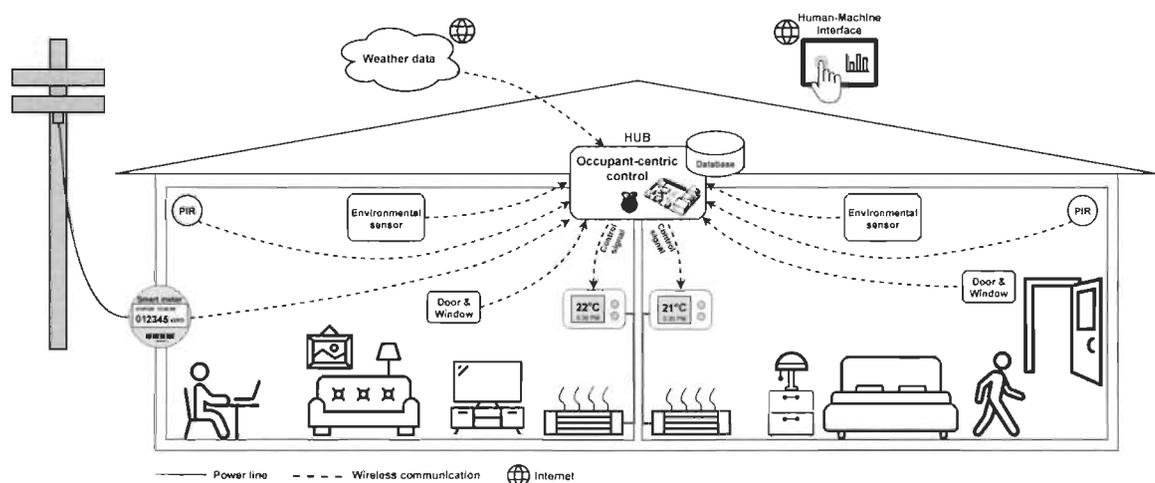
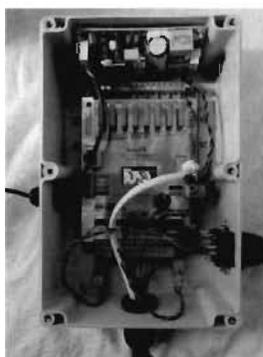


FIGURE A-1 Schéma de l'architecture proposée.

température intérieure (FIGURE A-2c). Un compteur intelligent (FIGURE A-2a) a également été installé pour surveiller la consommation électrique des systèmes d'éclairage et de chauffage. De plus, les thermostats de chacune des pièces ont été remplacés par des thermostats communicants (FIGURE A-2e), lesquels fournissent des informations sur la température intérieure de chaque zone thermique et permettent de contrôler à distance la température de consigne.

- Concentrateur (HUB) : Le concentrateur est constitué d'un Raspberry Pi 3 (FIGURE A-2e), doté des fonctionnalités suivantes.
 - Réalise l'acquisition et le stockage des données mesurées par chacun des capteurs. Le HUB dispose d'une base de données locale.
 - Réalise des requêtes auprès du serveur de données météorologiques pour obtenir les valeurs actuelles et les prévisions de variables telles que la température extérieure.



(a) Flexs Q5



(b) Capteurs de courant



(c) Capteur AirLab



(d) Thermostat



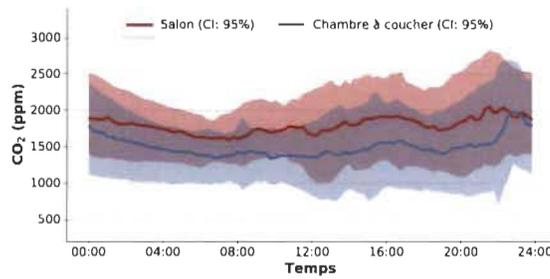
(e) Concentrateur (Raspberry pi)

FIGURE A-2 Photos du montage expérimental.

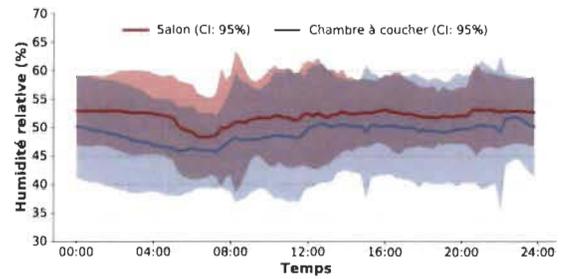
- Exécute le modèle pour effectuer la détection et la prévision de l'occupation, qui a été présentée à la Section 4.3.
 - Exécute le contrôleur prédictif et envoie les actions de contrôle résultantes à chacun des thermostats.
 - Utilisé comme une passerelle pour communiquer avec les dispositifs utilisant le protocole Zigbee, en combinaison avec le module USB CC2531 de Texas Instruments.
- Interface utilisateur : Elle est composée d'une interface web, à travers laquelle les utilisateurs peuvent visualiser les données stockées dans la base de données.
 - Serveur avec des données météorologiques : Les données météorologiques sont recueillies via la plateforme web Climacell [150]. Elle fournit des informations sur les valeurs actuelles, les prévisions et l'historique des variables telles que la température extérieure, le rayonnement solaire, etc.

A.1.1 L'impact des portes et fenêtres sur les variables mesurées

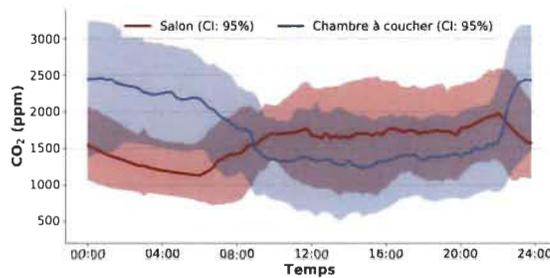
L'étude de cas analysée dans cette thèse a été réalisée dans un appartement comportant deux zones, le salon et la chambre, lesquels sont séparés par une porte. Une fois le système installé, des tests préliminaires ont été réalisés afin d'analyser l'impact des portes intérieures sur les variables environnementales de chaque zone thermique. À titre d'exemple, la FIGURE A-3 montre une comparaison de la concentration en CO₂ et l'humidité relative de l'air dans chacune des pièces de la résidence. Les figures A-3a, A-3b présentent le cas où la porte séparant les zones est laissée ouverte en tout temps, et les Figures A-3c et A-3d montrent le cas où la porte est fermée pendant la nuit. Ces résultats révèlent que lorsque la porte reste ouverte, les variables environnementales ne présentent pas de variations considérables tout au long de la journée, présentant une dynamique similaire dans toutes les zones de la maison. Cependant, lorsque la porte est fermée, les espaces sont « isolés », ce qui permet de mieux observer les variations produites par le comportement humain. Par exemple, la FIGURE A-3c montre l'augmentation de la concentration en CO₂ dans la chambre à coucher pendant la nuit lorsque les utilisateurs vont dormir, et montre la diminution de la concentration en CO₂ dans le salon lorsqu'il n'y a personne dans la pièce pendant cette période de temps.



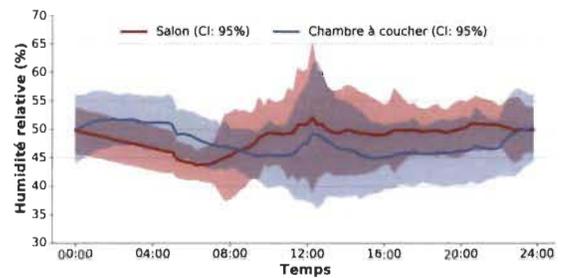
(a) Concentration en CO₂ avec la porte ouverte tout le temps



(b) Humidité relative avec la porte ouverte tout le temps



(c) Concentration en CO₂ avec la porte fermée pendant la nuit

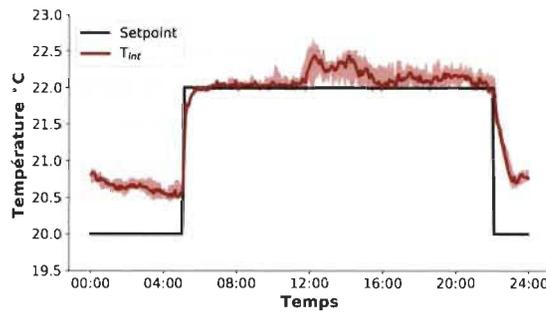


(d) Humidité relative avec la porte fermée pendant la nuit

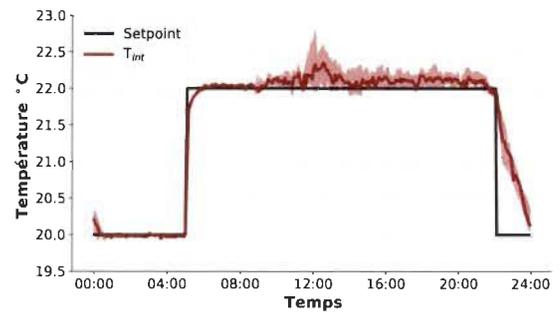
FIGURE A-3 Exemple de l'impact des portes intérieures sur les variables mesurées.

Par ailleurs, la FIGURE A-4 montre qu'en laissant les portes intérieures de la maison ouvertes, il se produit un échange de chaleur entre les zones, ce qui peut à son tour affecter la consommation électrique de chaque plinthe chauffante. Cela est illustré par les Figures A-4a, A-4c et A-4e, où il est évident qu'en gardant la porte ouverte, la plinthe chauffante du salon n'est pas utilisée pendant la nuit, car la chaleur provenant de la chambre à coucher maintient une température supérieure à la valeur de consigne du salon. Contrairement à ce qui se passe lorsque la porte est fermée, où chacune des plinthes chauffantes doit entrer en fonctionnement pour maintenir la température de consigne dans sa propre zone.

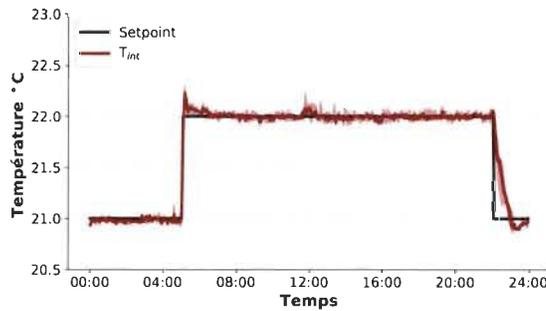
Suite à cette analyse, pour le travail présenté à la section 4.3, les portes et les fenêtres de la résidence ont été maintenues fermées la plupart du temps. De cette manière, l'objectif est de favoriser l'indépendance entre les variables environnementales de chaque zone de la résidence.



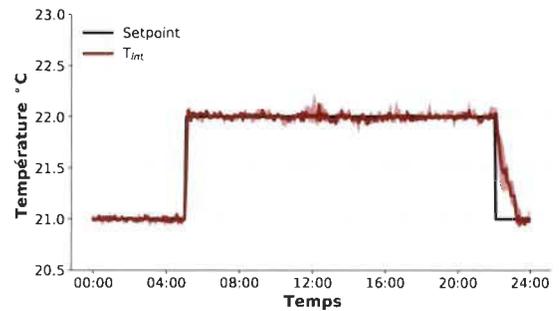
(a) Température du salon avec la porte ouverte tout le temps



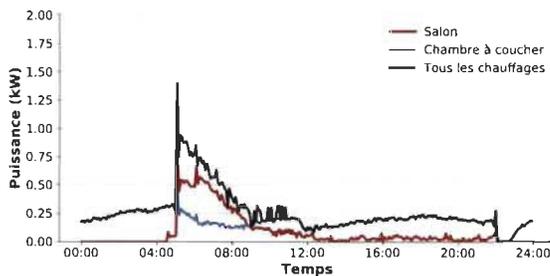
(b) Température du salon avec la porte fermée pendant la nuit



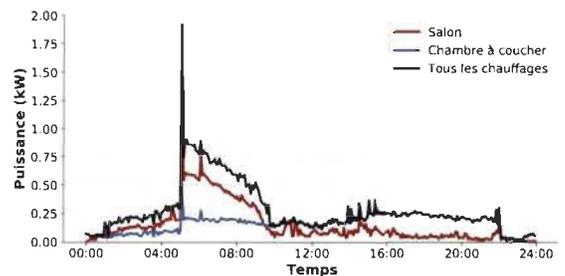
(c) Température de la chambre avec la porte ouverte tout le temps



(d) Température de la chambre avec la porte fermée pendant la nuit



(e) Puissance de chauffage avec la porte ouverte tout le temps



(f) Puissance de chauffage avec la porte fermée pendant la nuit

FIGURE A-4 Exemple de l'impact des portes intérieures sur le système de chauffage.

A.2 Modélisation de la dynamique thermique du bâtiment

Le contrôleur prédictif comporte trois parties essentielles : le modèle thermique du bâtiment, l'algorithme d'optimisation, et la fonction de coût. Les deux dernières ont été discutées en détail dans la Section 4.3 (Article 3), cependant, le développement du modèle thermique mérite quelques clarifications et informations complémentaires, qui sont présentées ci-dessous.

Pour reproduire la dynamique thermique de la maison, un modèle linéaire basé sur la

représentation semi-physique par analogie électrique-thermique de type nRC [151] est utilisé. Cette analogie permet d'utiliser les notions d'analyse de circuits dans le domaine thermique de la même manière que dans le domaine électrique [118]. Comme illustré dans FIGURE A-5 et formulé sur l'équation (A.1), une analogie de circuit 2R1C est utilisée pour représenter l'équilibre énergétique entre la chaleur introduite dans l'environnement et les pertes thermiques des pièces [10].

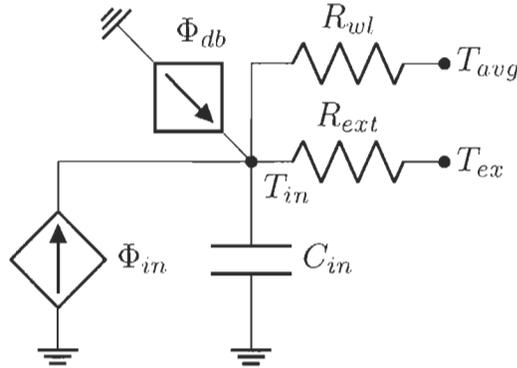


FIGURE A-5 Circuit équivalent RC de la dynamique thermique des pièces.

$$C_{in}\dot{T}_{in} = \frac{T_{ext} - T_{in}}{R_{ext}} + \frac{T_{avg} - T_{in}}{R_{wl}} + \Phi_{in} + \Phi_{db} \quad (\text{A.1})$$

où T_{in} est la température intérieure, T_{ext} est la température extérieure, T_{avg} est la température moyenne des zones adjacentes, Φ_{in} est la consommation électrique du système de chauffage et Φ_{db} sont les gains et les pertes thermiques de la pièce (par exemple, gains solaires, infiltration d'air, occupation), qui n'ont pas été pris en compte dans cette étude. De plus, C_{in} , R_{ext} et R_{wl} représentent respectivement la masse thermique interne, la résistance thermique qui isole le bâtiment et la résistance thermique du mur interne divisant les zones.

Ensuite, l'expression mathématique du modèle d'espace d'état pour une résidence avec deux zones thermiques (η_1 et η_2) peut être exprimée comme suit,

$$\begin{bmatrix} T_{in\eta_1}^{(k+1)} \\ T_{in\eta_2}^{(k+1)} \end{bmatrix} = A \begin{bmatrix} T_{in\eta_1}^{(k)} \\ T_{in\eta_2}^{(k)} \end{bmatrix} + B \begin{bmatrix} \Phi_{in\eta_1}^{(k)} \\ \Phi_{in\eta_2}^{(k)} \end{bmatrix} + C \begin{bmatrix} T_{ext}^{(k)} \end{bmatrix} \quad (\text{A.2})$$

où $A \in R^{2 \times 2}$, $B \in R^{2 \times 2}$ et $C \in R^{1 \times 1}$ sont les matrices de paramètres que le modèle trouve en minimisant la norme ℓ_2 au carré entre l'équation (A.2) et la sortie réelle mesurée Y sur un ensemble de données historiques.

La FIGURE A-6 montre un exemple des résultats obtenus avec le modèle thermique construit. Il est évident que le modèle est capable de reproduire la dynamique thermique de la maison. Cependant, il est important de noter que, vu qu'il s'agit d'un modèle linéaire, son application est limitée aux cas où il n'y a pas de transfert de chaleur par convection forcée.

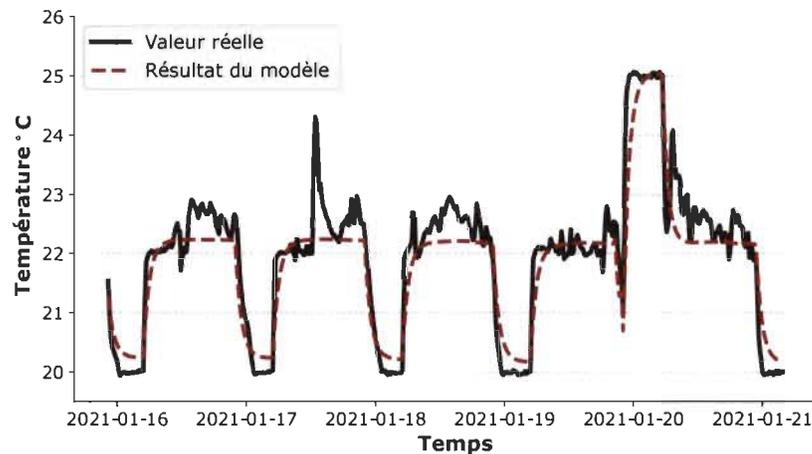


FIGURE A-6 Circuit équivalent RC de la dynamique thermique des pièces.

A.3 Formulaire pour collecter la valeur d'occupation réelle du logement

Pour valider le modèle de détection de présence proposé dans cette thèse, la valeur réelle d'occupation est nécessaire. C'est pourquoi le formulaire présenté dans le Tableau A-1 a été conçu pour que les utilisateurs puissent noter les périodes de présence/absence quotidiennes dans chacune des zones thermiques de la résidence. La résolution du formulaire est de 15 minutes, ce qui fournit des informations détaillées sur la présence des occupants. La période de collecte des données s'est déroulée entre décembre 2020 et mars 2021.

Les indications suivantes ont été données aux utilisateurs pour remplir le formulaire :

- Les périodes de présence sont enregistrées comme 1 et les périodes d'absence comme 0.
- La pièce est considérée comme occupée si au moins une personne est présente.

- Il est considéré qu'il y a présence sur la pièce pendant un intervalle de 15 minutes du formulaire si au moins une personne est présente pendant une période égale ou supérieure à 5 minutes.

Tableau A-1 Formulaire pour collecter la valeur d'occupation réelle du logement

Temps	20/12/08	20/12/09	20/12/10	...	21/03/18	21/03/19	21/03/20
00 :00h	-	-	-	-	-	-	-
00 :15h	-	-	-	-	-	-	-
00 :30h	-	-	-	-	-	-	-
00 :45h	-	-	-	-	-	-	-
01 :00h	-	-	-	-	-	-	-
01 :15h	-	-	-	-	-	-	-
⋮	-	-	-	-	-	-	-
22 :30h	-	-	-	-	-	-	-
22 :45h	-	-	-	-	-	-	-
23 :00h	-	-	-	-	-	-	-
23 :15h	-	-	-	-	-	-	-
23 :30h	-	-	-	-	-	-	-
23 :45h	-	-	-	-	-	-	-

Annexe B - Caractérisation du signal de puissance

Cette annexe présente la caractérisation du signal de puissance qui a été réalisé en tant qu'une étude complémentaire de cette thèse. Différents travaux tels que [45], [46] et [47] ont montré qu'en identifiant les habitudes de consommation et la relation entre les différentes charges domestiques, il est possible de détecter les appareils qui ont un impact plus important sur la consommation de la résidence, ainsi que de modéliser le comportement des occupants et leurs variations dans le temps, ce qui peut être utilisé pour améliorer l'efficacité énergétique des logements. À travers cette étude exploratoire, nous utilisons un système permettant le sous-mesurage de la puissance, pour analyser le signal de puissance de certaines charges domestiques et la puissance totale d'un logement situé à Trois-Rivières, dans le but d'identifier les événements récurrents et les habitudes de consommation des utilisateurs. Tout d'abord, nous analysons le signal de puissance totale, ainsi que les habitudes d'utilisation du système de chauffage, du chauffe-eau et de l'éclairage. Par la suite, la relation entre les habitudes d'utilisation des charges domestiques et les périodes de forte consommation d'énergie d'une résidence sont étudiées en utilisant une analyse de corrélation croisée. Enfin, nous présentons une approche pour la détection d'évènements transitoires de charge électrique

B.1 Caractérisation du signal de puissance totale

La caractérisation du signal de puissance agrégé est effectuée en suivant un processus en trois étapes, conformément à la FIGURE B-1. Premièrement, une phase de prétraitement est effectuée pour améliorer la qualité des données, afin d'obtenir des résultats plus précis et significatifs. Dans la deuxième étape, une analyse d'autocorrélation est effectuée pour identifier des patrons périodiques dans le signal de puissance totale. Finalement, la troisième

étape permet la détection de périodes de forte consommation d'énergie de la résidence en utilisant une méthode basée sur des filtres à convolution.

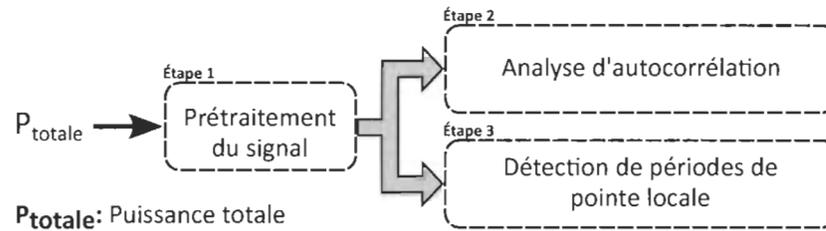


FIGURE B-1 Schéma-bloc de l'approche proposée pour la caractérisation du signal de puissance.

La FIGURE B-2 présente sous forme d'exemple la température extérieure et le signal de puissance totale de la résidence utilisée pour les analyses effectuées dans cette annexe. Cette maison est située dans la ville de Trois-Rivières, Québec, Canada et l'information a été collectée pendant la saison d'hiver 2017-2018, en utilisant le système de mesurage présenté dans [152].

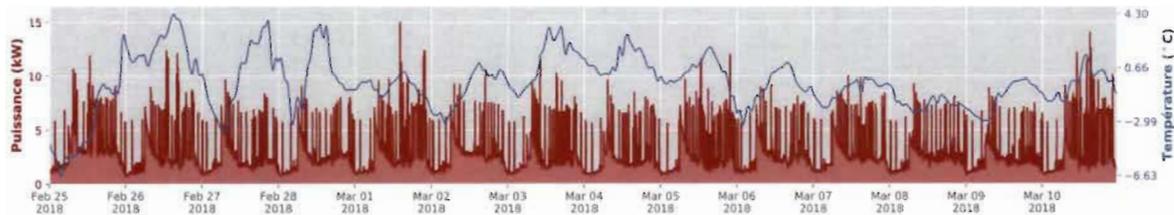


FIGURE B-2 Puissance totale d'une maison réelle pendant la saison d'hiver 2017-2018.

B.1.1 Prétraitement du signal

Premièrement, une étape de détection et de correction des données manquantes a été effectuée. Ensuite, tel que proposé par Hart [153], l'équation B.1 est utilisée pour réduire les fluctuations du réseau électrique et obtenir un signal de puissance normalisé à 120V.

$$P_{Norm}(t) = 120^2 Y(t) = \left(\frac{120}{V(t)} \right)^2 P(t) \quad (\text{B.1})$$

où $Y(t)$ est l'admittance, $P(t)$ est la puissance et $V(t)$ est la tension.

Par la suite, une étape de lissage a été réalisée. Pour cela, une méthode basée sur l'estimation par noyau est utilisée (Kernel smoothing, en anglais). Cette approche permet

de calculer la moyenne pondérée de chaque valeur de signal avec ses voisins. Pour ce faire, le poids de chaque échantillon est défini par le noyau (K), de telle sorte que les points les plus proches reçoivent des poids plus élevés. Ainsi, le lissage d'une fonction Y est obtenu comme décrit dans l'équation suivante :

$$\hat{Y}(x^*) = \frac{\sum_{i=1}^N K(x^*, x_i) Y(x_i)}{\sum_{i=1}^N K(x^*, x_i)} \quad (\text{B.2})$$

où $\hat{Y}(x^*)$ est le signal lisse, K est le noyau, x_i correspond au $i^{\text{ème}}$ point observé, x^* correspond au point estimé et N est le nombre de points observés.

Dans la littérature, les noyaux les plus populaires utilisés pour le lissage sont les noyaux paraboliques (Epanechnikov), les tricubes et les gaussiens. Pour ce travail, nous avons utilisé le noyau gaussien, lequel est défini de la façon suivante :

$$K(x^*, x_i) = e^{-\frac{(x^* - x_i)^2}{2b^2}} \quad (\text{B.3})$$

où b est l'échelle de longueur de l'espace d'entrée.

La FIGURE B-3 présente le résultat obtenu avec l'étape de lissage pour le signal de puissance totale d'une journée hivernale.

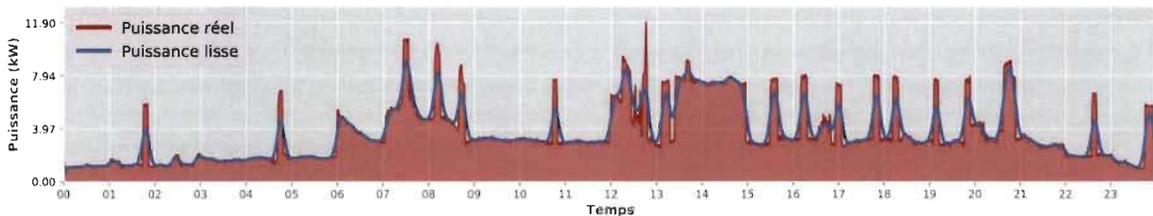


FIGURE B-3 Exemple du résultat obtenu avec l'étape de lissage.

B.1.2 Analyse d'autocorrélation

L'analyse de corrélation croisée est utilisée comme mécanisme pour identifier la relation entre les variables et leur différence de phase dans le domaine temporel. La corrélation croisée normalisée (NCC) est défini à partir de l'expression suivante :

$$NCC(x) = \frac{\sum_{i=1}^N S_1(x+i) \cdot S_2(i)}{\sqrt{\sum_{i=1}^N S_1(x+i)^2} \cdot \sqrt{\sum_{i=1}^N S_2(i)^2}} \quad (\text{B.4})$$

où S_1 et S_2 sont des séries temporelles de longueur N et x représentent le temps de latence (lag) entre S_1 et S_2 .

L'autocorrélation est un cas particulier de l'analyse de corrélation qui permet de détecter des patrons périodiques dans un signal [154]. Cette analyse consiste à réaliser la corrélation d'un signal par lui-même, c'est-à-dire que dans l'équation B.4, il faut rendre S_1 égal à S_2 . La FIGURE B-4 montre le résultat de l'analyse d'autocorrélation sur deux semaines du signal de puissance totale de la maison de test. Cette analyse permet d'observer que la consommation électrique présente une périodicité toutes les 12 et 24 heures. La périodicité de 24 heures indique que le comportement de la consommation électrique de cette résidence est très similaire chaque jour. Quant à la périodicité de 12 heures, elle peut être attribuée à la présence de la consommation de pointe du matin (6h - 9h) et de celle de l'après-midi (16h - 21h), qui sont les périodes habituelles avec le plus d'activité dans les maisons.

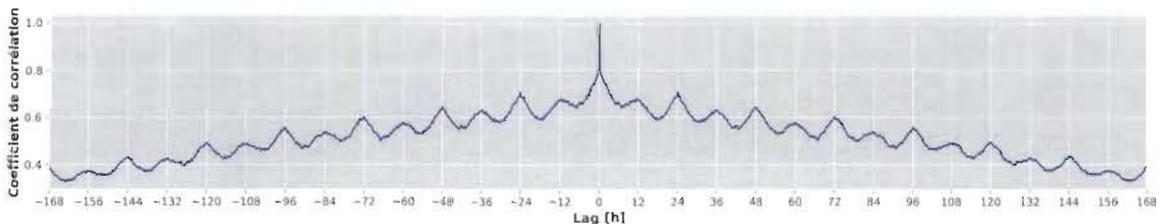


FIGURE B-4 Autocorrélation du signal de puissance totale présenté dans la FIGURE B-2.

B.1.3 Détection des périodes de pointe locale

B.1.3.1 Analyse multi-échelle

Pour ce travail, nous considérerons une période de pointe locale à une plage de temps de plus de 30 minutes pendant laquelle la consommation d'énergie est supérieure à la moyenne. Afin d'identifier ces périodes de forte consommation d'énergie d'une résidence, l'utilisation d'une analyse de convolution multi-échelle basée sur des noyaux est proposé. Le noyau sélectionné est l'ondelette de Ricker (chapeau mexicain), et comme présenté à la FIGURE

B-5, il a été utilisé avec quatre échelles de temps : 30 minutes, 60 minutes, 90 minutes et 120 minutes.

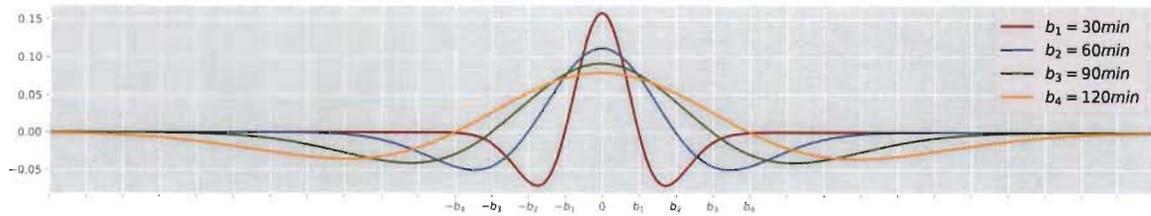


FIGURE B-5 Noyaux multi-échelles utilisés.

De plus, en se basant sur [155], la signal de puissance totale a été normalisé en utilisant l'équation B.5. Cette normalisation permet d'éliminer toute composante DC du signal, produisant un signal couplé AC et normalisé en amplitude avec une moyenne nulle.

$$x_{norm} = \frac{x - \frac{1}{N} \sum_{i=0}^{N-1} x[i]}{\left\| x - \frac{1}{N} \sum_{i=0}^{N-1} x[i] \right\|} \quad (\text{B.5})$$

où x est une série temporelle de longueur $N - 1$ et le dénominateur est la norme 2 du numérateur.

Il faut noter que lorsqu'on travaille avec des signaux ayant une moyenne égale à zéro, le résultat de la convolution nous permettra d'identifier facilement les valeurs de consommation supérieures ou inférieures à la moyenne. La FIGURE B-6 montre le résultat de l'analyse multi-échelles pour une journée d'hiver de la maison de test. La figure ne montre que des valeurs de convolution supérieures à zéro, c'est-à-dire tout ce qui est au-dessus de la moyenne, car les résultats ont été tronqués. Ces résultats révèlent que les échelles de temps plus petites permettent d'identifier les périodes de consommation élevée de courte durée (événements à haute fréquence). Cependant, en augmentant l'échelle de temps, les pointes de courte durée sont éliminées, ne laissant que des événements de basse fréquence, qui peuvent se produire en raison de l'utilisation de plusieurs charges à faible ou forte consommation d'énergie.

B.1.3.2 Analyse du début et de la fin des périodes de pointe locale

Une fois l'analyse multi-échelle effectuée, la distribution de probabilité pour qu'une période de pointe commence ou finisse pendant la journée est déterminé. La FIGURE B-7

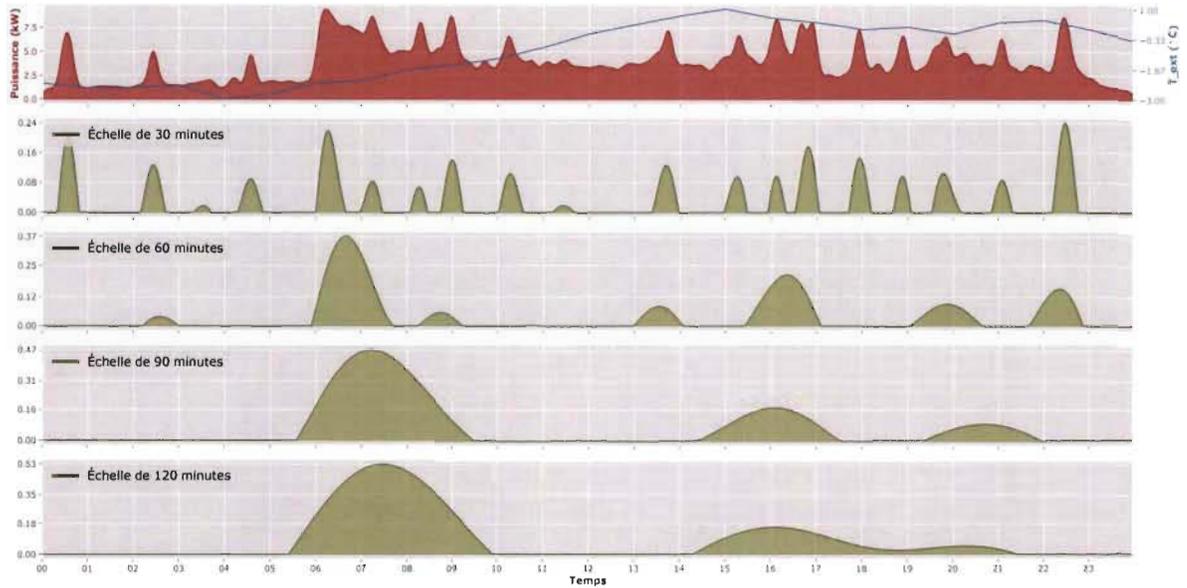
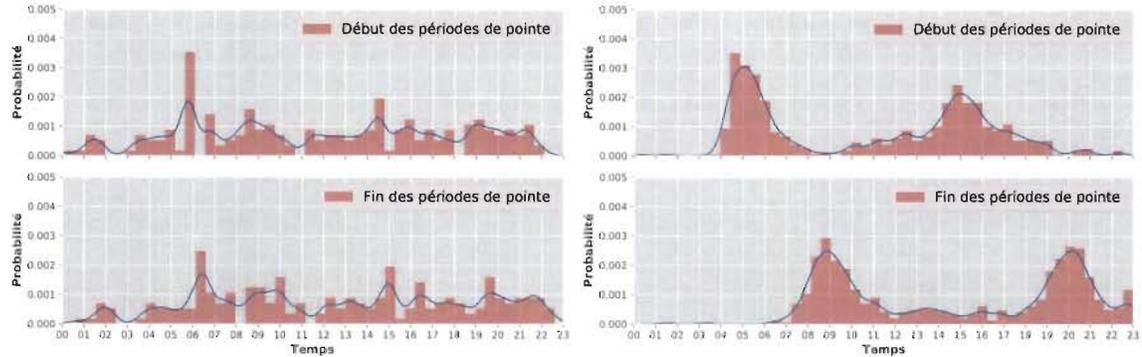


FIGURE B-6 Exemple des résultats obtenus avec l'analyse multi-échelle.

présente la distribution de probabilité pour les échelles de 30 et 120 minutes afin de comparer leurs résultats.

Dans le cas de l'échelle de 30 minutes (FIGURE B-7a), on observe que la probabilité du début et de la fin d'une période de pointe est similaire tout au long de la journée. Ce résultat est dû au fait que l'échelle de 30 minutes permet d'identifier les périodes de consommation élevée de courte durée, qui peuvent être présentes à tout moment de la journée en raison de charges telles que le chauffe-eau. Cependant, pour les résultats de l'échelle de 120 minutes (FIGURE B-7b) on observe qu'entre 4h et 6h du matin il y a une forte probabilité d'avoir le début d'une période de forte consommation d'énergie. De plus, il est évident que cette période de pointe du matin peut être prolongée jusqu'à 8h ou 10h.

Par ailleurs, on peut aussi identifier la présence d'une période de pointe dans l'après-midi, qui commence vers 14h - 16h et peut continuer jusqu'à 19h ou 21h. Ce résultat présente une concordance avec les périodes de pointe définies par Hydro-Québec [20], qui indique que les pointes quotidiennes se produisent le matin entre 6h et 9h et à la fin de la journée entre 16h et 20h en fonction de grand nombre de clients actifs dans les logements. Pour cette raison, pour la suite de ce travail, nous avons décidé d'utiliser les résultats obtenus avec l'échelle de 120 minutes comme référence pour les périodes de pointe locale.



(a) Résultats pour l'échelle de 30 minutes.

(b) Résultats pour l'échelle de 120 minutes.

FIGURE B-7 Distribution de probabilité pour le début et la fin des périodes de pointe obtenues avec l'échelle de 30 minutes et 120 minutes.

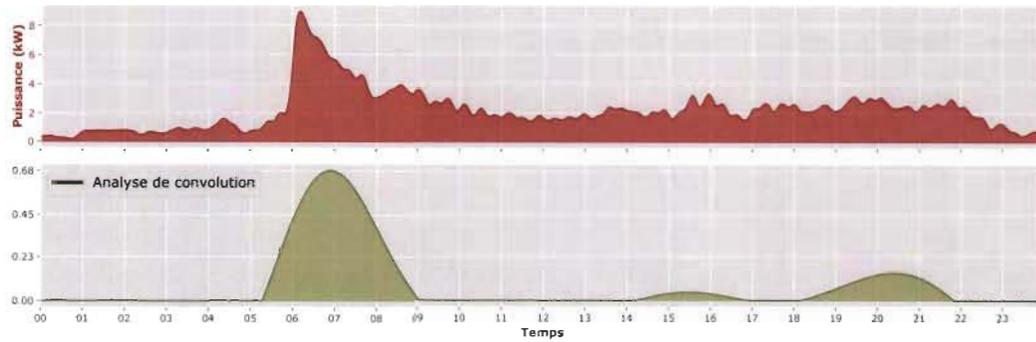
B.2 Analyse de charges domestiques

Au Québec, le chauffage électrique des espaces et de l'eau sont des charges affectant considérablement le profil de consommation d'électricité des résidences. Par ailleurs, bien que l'éclairage ne représente qu'environ 5% de la consommation totale d'un ménage [156], ceci est un charge fortement liée à la présence et au comportement des occupants. Par conséquent, l'analyse des habitudes d'utilisation de ces trois charges est présentée ci-dessous.

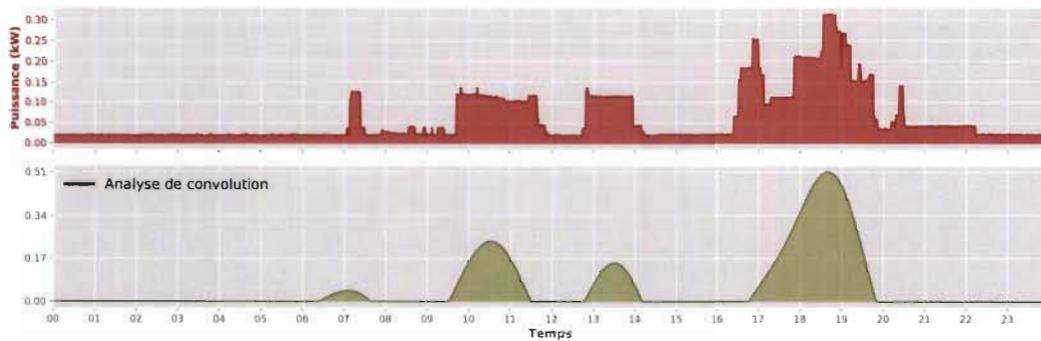
B.2.1 Chauffage des espace et éclairage

Pour identifier les patrons d'utilisation du chauffage et de l'éclairage, la même méthodologie présentée dans la section B.1 est utilisée. Ainsi, on commence par appliquer l'analyse de convolution basée sur des noyaux pour identifier les périodes de la journée où la demande d'énergie est la plus élevée. Un exemple des résultats obtenus avec cette analyse peut être observé à la FIGURE B-8.

Ensuite, les moments de début et de fin de chacune des périodes de forte consommation sont identifiés. Avec ces événements, les distributions de probabilité sont ensuite calculées, permettant ainsi de connaître les moments de la journée et la durée pendant laquelle les charges sont plus sollicitées. D'un côté, les Figures B-9a et B-9c montrent que le chauffage des espace a une consommation d'énergie importante et récurrente entre 4-5h et 8-9h, qui peut être liée au moment de la journée où les personnes se réveillent et augmentent la température de consigne



(a) Résultats de l'analyse de la consommation de chauffage des espace.



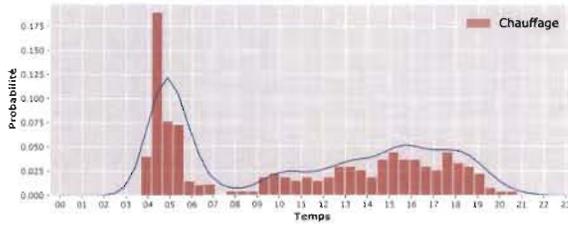
(b) Résultats de l'analyse de la consommation de d'éclairage.

FIGURE B-8 Exemple des résultats obtenus avec l'analyse de convolution.

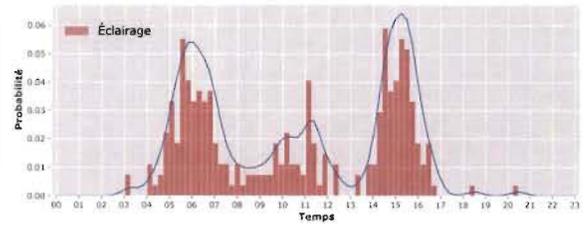
pour maintenir le confort thermique. De plus, le soir, vers 21-22h, une réduction de l'utilisation du chauffage est évidente, ce qui peut être attribuée à une réduction de la température de consigne pendant la période où les gens dorment. D'un autre côté, les Figures B-9b et B-9d montrent que l'utilisation de l'éclairage est clairement liée à la période de présence et d'activité des personnes dans la résidence (entre 5 h et 21 h). Par ailleurs, la relation de ce système avec les périodes d'obscurité de la journée est évidente, ce qui se reflète par une fréquence d'utilisation plus élevée le soir ou tôt le matin.

B.2.2 Chauffe-eau

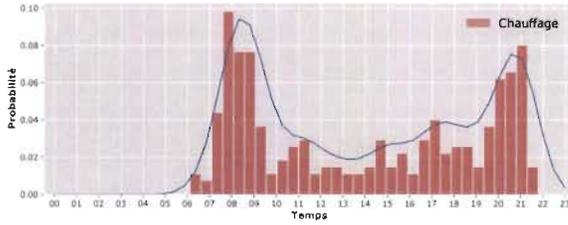
Parmi la répartition de la consommation d'énergie dans le secteur résidentiel du Québec, les systèmes de chauffage de l'eau représentent environ 17 à 20% de la consommation [156], [157]. Le chauffe-eau électrique est le système de chauffage de l'eau le plus utilisé au Québec et l'une des charges électriques les plus énergivores dans les maisons canadiennes après le chauffage des locaux [158].



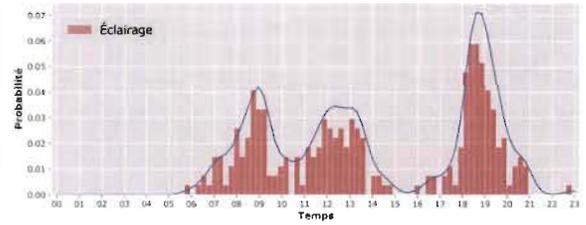
(a) Début des périodes de forte demande d'énergie.



(b) Début des périodes de forte demande d'énergie.



(c) Fin des périodes de forte demande d'énergie.

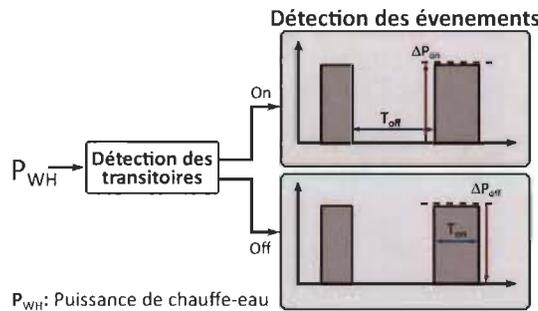


(d) Fin des périodes de forte demande d'énergie.

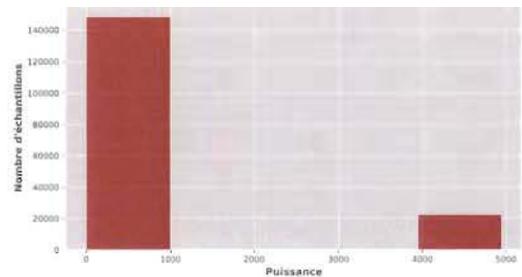
FIGURE B-9 Distribution de probabilité pour le début et la fin des périodes de plus forte demande d'énergie pour le chauffage des espace et l'éclairage.

B.2.2.1 Détection des événements

En raison de la nature résistive du chauffe-eau et de sa forte consommation électrique, il s'agit d'une charge relativement facile à identifier. Pour cette raison, tel que présenté à la FIGURE B-10a, la détection de ses événements transitoires de connexion et de déconnexion a été effectuée à partir d'une analyse basée sur la dérivée de son signal de puissance. De cette façon, il est possible d'identifier les variations soudaines de puissance, lesquelles sont comparées à une valeur K constant, afin de déterminer si le transitoire détecté correspond à l'un des événements recherchés.



(a) Schéma-bloc de l'approche proposée.



(b) Histogramme des valeurs de puissance.

FIGURE B-10 Approche proposée pour la détection des événements du chauffe-eau.

La sélection du constant K s'effectue en fonction de l'analyse de la distribution des valeurs de puissance (FIGURE B-10b). Selon l'histogramme obtenu on observe que la puissance prend des valeurs comprises entre 0 et 5000 w, en observant deux grands groupes, l'un avec une consommation inférieure à 1000 w et l'autre supérieure à 4000 w. Alors, parce que la consommation du chauffe-eau ne prend pas une grande nombre de mesures intermédiaires, K est choisi égal à 3000. Un exemple de résultats obtenus à partir de cette approche de détection d'événements est présenté à la FIGURE B-11.

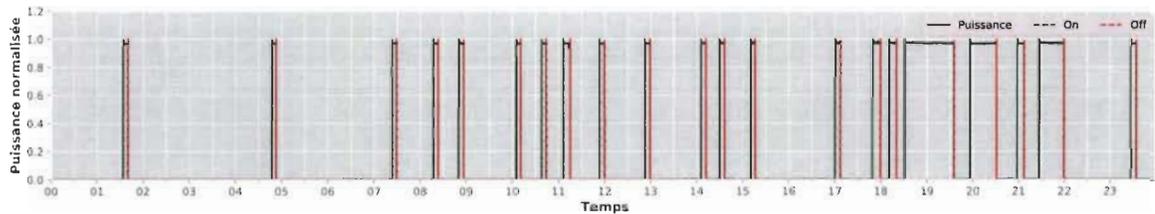
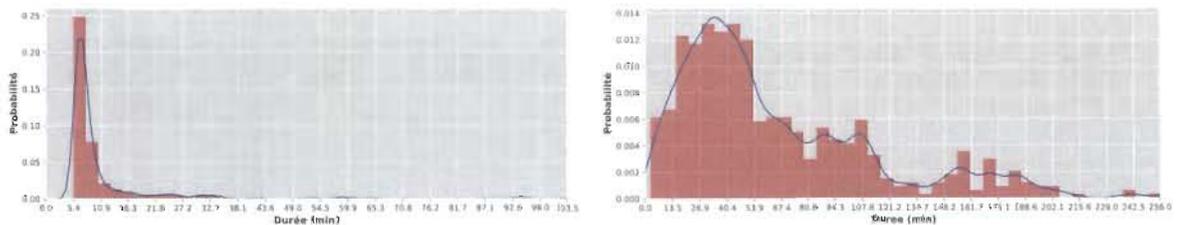


FIGURE B-11 Détection des événements du chauffe-eau.

À partir d'événements détecté, il est effectué une analyse de la durée des états de fonctionnement du chauffe-eau («on-off»). La FIGURE B-12a permet voir que la durée de l'état «on» se trouve généralement entre 5 et 11 minutes, cependant, des périodes de fonctionnement plus prolonges sont possibles comme conséquence d'une utilisation intensive de l'eau chaude. Par contre, la FIGURE B-12b révèle que l'état «off» a une durée qui, dans la plupart des cas, prend des valeurs allant jusqu'à 1 heure et demie. Néanmoins, il peut aussi y avoir des cas où le temps de repos est plus long.



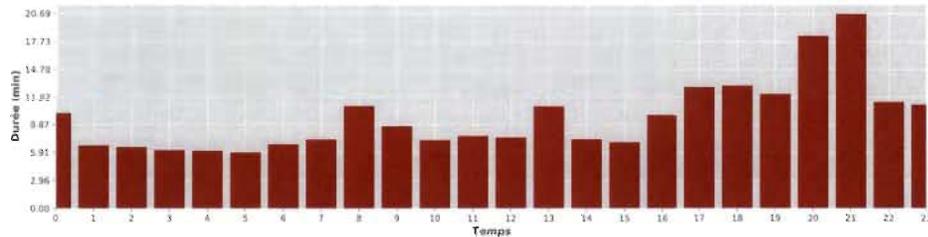
(a) Durée de l'état «on».

(b) Durée de l'état «off».

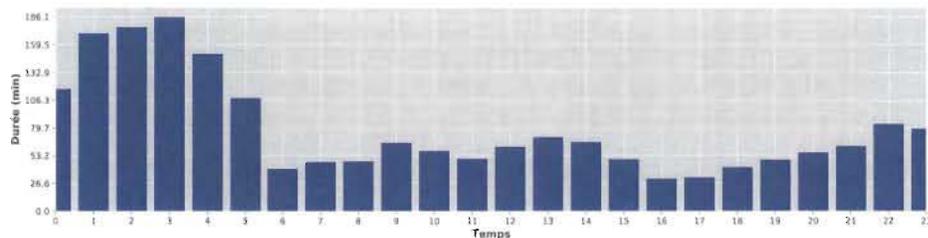
FIGURE B-12 Analyse de la durée des états de fonctionnement du chauffe-eau.

De plus, l'influence du moment de la journée sur l'utilisation du chauffe-eau a été analysée. Pour ce faire, la durée moyenne des états de fonctionnement du chauffe-eau pendant chaque heure de la journée est calculée. Les résultats de la FIGURE B-13b montrent qu'à 6 heures du

matin, la durée de l'état «off» du chauffe-eau diminue, ce qui signifie que la consommation d'eau chaude dans la maison augmente. De même, la FIGURE B-13a indique que la période d'utilisation plus intensive de l'eau chaude se situe entre 16 et 21 h, ce qui se reflète dans l'augmentation du temps de fonctionnement du chauffe-eau.



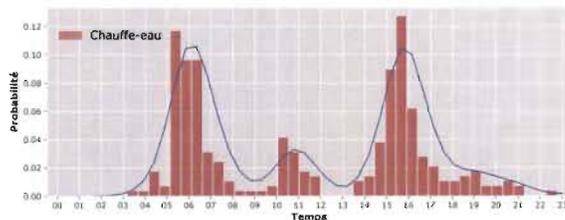
(a) Durée moyenne de l'état «on» pendant la journée.



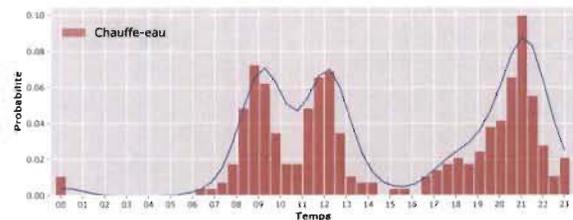
(b) Durée moyenne de l'état «off» pendant la journée.

FIGURE B-13 Analyse de la durée des états de fonctionnement du chauffe-eau pendant la journée.

Par ailleurs, de la même façon que dans les sections précédentes, l'analyse de convolution est utilisée pour identifier les périodes de plus forte consommation d'énergie du chauffe-eau. La FIGURE B-14 montre les distributions de probabilité résultant de cette analyse.



(a) Début des périodes de forte demande d'énergie.



(b) Fin des périodes de plus demande d'énergie.

FIGURE B-14 Distribution de probabilité pour le début et la fin des périodes de plus forte demande d'énergie pour le chauffe-eau.

B.3 Analyse de corrélation croisée entre les habitudes de consommation

Dans cette section, le but est d'identifier la relation entre les habitudes d'utilisation des charges domestiques et les périodes de plus forte consommation d'une résidence. Pour ce faire, nous effectuons une analyse de corrélation croisée entre la probabilité d'avoir le début d'une période de forte consommation et le début de chacun des événements de charge identifiés.

Les résultats de cette analyse, présentés à la FIGURE B-15, nous permettent d'identifier que le début des événements de forte consommation de chauffage des espace est directement corrélé au début des périodes de pointe de la résidence. De plus, il est relevé que le début d'utilisation de l'éclairage et le chauffe-eau est décalé d'environ 30 minutes et 1 heure, respectivement. Cela signifie que ces appareils commencent à être utilisés plus souvent peu après le début de la période de pointe.

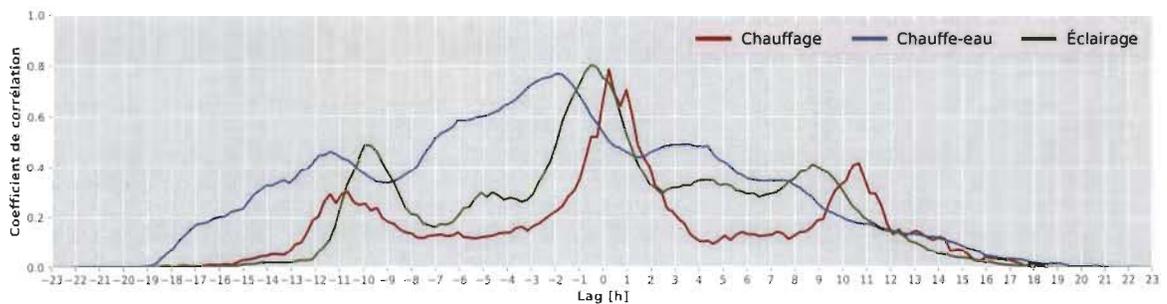


FIGURE B-15 Analyse de corrélation croisée normalisée.

La littérature montre que les données des compteurs intelligents fournissent de nombreuses informations sur les comportements et les modes de vie des consommateurs en matière de consommation d'électricité [159]. En fait, tel que présenté dans cette annexe, la caractérisation et le profilage de la consommation des charges domestiques et de la consommation totale du logement est une approche ayant le potentiel pour extraire et modéliser des informations sur le comportement et les préférences des utilisateurs. Par ailleurs, ce type d'analyse met en évidence la possibilité d'identifier la relation existant entre l'utilisation des appareils ménagers. Cela constitue une alternative pour anticiper la demande ou pour aider les utilités et les distributeurs à mettre en œuvre des programmes de réponse à la demande plus efficaces et des services plus personnalisés.

B.4 Détection des évènements de charge électrique

La détection des évènements de charge électrique est une étape importante pour la surveillance non intrusive des charges (NILM, pour Non-intrusive load monitoring) et la modélisation des habitudes de consommation. Pour le cas de NILM, Hart [153], [160] a été le premier à montrer comment les changements échelonnés (step-like changes) présents dans le signal de consommation d'énergie électrique peuvent être utilisés pour distinguer les appareils [161]. Cependant, pour le cas de la modélisation comportementale, tel que discuté précédemment dans cette annexe et dans la littérature scientifique, à partir de l'analyse du signal de puissance et la détection des évènements de charge électrique, il est possible d'extraire des informations reliées aux préférences de consommation et du comportement des occupants [45]. De cet fait, une analyse exploratoire sur la détection des évènements transitoire a été réalisée. Ce travail est présenté sous la forme d'un article scientifique, comme précisé ci-dessous.

Titre : Transient event classification based on wavelet neuronal network and matched filters

Journal : IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society

Auteurs : Luis Rueda, Alben Cardenas, Sousso Kelouwani and Kodjo Agbossou

Date de publication : 21-23 octobre 2018

DOI : <https://doi.org/10.1109/IECON.2018.8591684>

Transient Event Classification based on Wavelet Neuronal Network and Matched Filters

Luis Rueda, Alben Cardenas, Souso Kelouwani and Kodjo Agbossou
Electrical and Computer Engineering Department, Hydrogen Research Institute
Université du Québec à Trois-Rivières
Trois-Rivières, QC, Canada G9A 5H7

Email: {luis.rueda, alben.cardenasgonzalez, souso.kelouwani, kodjo.agbossou}@uqtr.ca

Abstract—Detailed information about load behavior and home's occupancy is important to implement Home Energy Management Systems (HEMS) capable of reducing energy consumption while maintaining user comfort. This is why Non-intrusive Appliance Load Monitoring (NIALM) and Non-intrusive Occupancy Monitoring (NIOM) have an important role to play in the new context of smart grid. This paper shows the implementation of two algorithms for transient event detection and classification, which is the first key step of a NIOM process. The first method employs Wavelet transform for the feature extraction and Artificial Neural Networks for the classification problem. The second method is based on the theory of Matched Filters to achieve the transient event detection and classification. Experiments permitted to validate the proposed methods using a dataset of occupied residential building.

Index Terms—Non-intrusive load monitoring, Non-intrusive Occupancy Monitoring, Artificial Neural Networks, Wavelet, Matched Filters, Home Energy Management.

I. INTRODUCTION

Electricity service is a fundamental factor in modern economic activity [1]. The electric grid is radically evolving and transforming into the Smart Grid (SG), which is characterized by improved energy efficiency and manageability of available resources [2].

The Smart Grid is a modern power grid that permits bi-directional communication between energy providers and consumers, the integration of distributed energy resources and the electric vehicles, real-time demand response and monitoring of consumers' energy usage behavior [2], [3].

Energy management (EM) systems, often integrated with home automation systems, play an important function in the control of home energy consumption and enable increased consumer participation. These systems provide to consumers information about their energy consumption patterns and help them to adopt energy-efficient behavior [2].

In this new paradigm, the management of residential power consumption is one of the most interesting alternatives to permit the balance of demand and production. Thus, the concepts of Demand Side Management (DSM) and Home Energy Management (HEM) appear and take relevant places in research and development efforts of academia and industry. The implementation and success of such programs depend on advanced knowledge of the loads characteristics and their utilization. Thus the modeling and prediction of their behavior could be an important support for DSM and HEM. In this

trend, researchers and industry worked over the last decades looking for efficient and non-intrusive load monitoring and identification techniques [4], [5].

Non-intrusive Appliance Load Monitoring is a concept, which takes an important role in the new context of smart grid. This idea, with origins in the early 1980s [4] and continuously evolving can support the implementation and enhance the performance of Demand Response and Home Energy Management strategies. NIALM comprises at least two main parts: the modeling of loads and the load disaggregation from the total power. Traditional approaches perform the modeling task from measurements of each load to obtain individual signatures that are employed after for disaggregation. However, it might be more convenient to obtain the models from the global power measurement and perform the load disaggregation in an adaptive manner.

In the same way, occupancy detection could be an efficient tool to favor energy savings in building [6]. Most current approaches for monitoring occupancy are intrusives, typically requiring the deployment of multiple environmental sensors or other dedicated devices [7]. Recent research looks for Non-intrusive Occupancy Monitoring using electricity data from smart meters to infer or detect occupancy in residential buildings and to improve the energy efficiency and reduce energy consumption [8].

Event detection is important for the occupant detection and modeling. From the power consumption information, it is possible to identify some events that trigger a series of activities that require a characteristic energy consumption (energy consumption profile). Following this, a good detection of events and a proper identification of the consumption profiles will make possible the development of an accurate occupancy model. This later allows us to anticipate the demand of energy in the residence and to improve the performance of residential management systems.

This article shows the comparison between two methods used for transient event classification. This comparison allows to identify the advantages and disadvantages of each approach. The first method is based on Wavelet decomposition and Artificial Neural Network (ANN) and the second method is based on Matched Filters. These approaches can be employed to build the models of residential loads from the aggregated measure of power consumption at the electric board panel

level. These models will enable the real-time Non-Intrusive Load and Occupancy Monitoring (NILM and NIOM) [7], [9].

The structure of this paper is organized as follows: Section II gives an overview of event detection approaches used for NILM in literature. Section III explains the methodology of the event detection approaches and their implementation. Section IV shows the experiments and results of the proposed approaches. Finally, Section V presents the conclusions of the research.

II. RESEARCH BACKGROUND

Hart was the pioneer in the work of NILM [4], [9]. He was the first to show how step-like changes present in the power consumption signal can be used to distinguish between appliances [10]. Our interest in this paper is the events detection and classification as the first step for the modeling of occupancy, energy consumption and load utilization in a residence.

After the work realized by Hart, in the literature, it is possible to find different event detection approaches. These publications have been classified by categorizing the approaches into four categories: 1) Expert Heuristics, 2) Probabilistic approach, 3) Transversal filters approach and 4) Other approaches.

1) *Expert Heuristics*: In the Expert Heuristic method, the changes of power consumption signal in steps enable us to detect the electrical appliances corresponding to these changes [11]. In this methodology, the first step of the algorithm is set the pre-event window and the post-event window. After that, the average value of power consumption samples in the two windows is calculated. Then, the absolute value of the difference between average value of power consumption in the pre-event window and post-event window is calculate. Finally, the difference value is compared to pre-determined threshold values (x) for the on and off power changes associated with each appliance. If the difference is larger than the x , an event is detected [10], [11].

2) *Probabilistic approach*: The problem of detecting abrupt changes in time series data is known as change detection [10]. In the literature, there are many techniques that have been applied to the problem of detecting events in power signals. A comparative study of event detection methods is presented in [11], that paper provides a comparison between a statistical approach, namely Goodness-of-Fit (GoF) and a Expert Heuristic method. In [12] a Generalized Likelihood Ratio (GLR) approach is introduced, that approach, calculates a “decision statistic from the natural log of a ratio of probability distributions before and after a potential change in mean value”. Similar methods are proposed in [10] and [13].

3) *Transversal filter approach*: Transversal filter or Matched filtering is a process for detecting the presence of a known piece of signal (mask) in the unknown signal [10]. Transients are identified by matching events in the incoming aggregated power stream to previously defined transient signatures (start-up or turn-off transients of different appliances). The work related in [14] presents the details of a transient

event detection algorithm based on transversal filters approach that identifies turn-on transients by examining spectral envelopes. This same approach was used for the works presented in [15] and [16].

4) *Other approaches*: In the literature it is possible to find other different approaches for event detection. For example, the classification of different transient appliances based on the utilization of Stockwell Transform (S-Transform) and Support Vectors Machines (SVM) has been investigated in [17]. In [18] a grid-based clustering algorithm is utilized in order to segment the power signals into transient and steady-state sections. Also, in [19] the Wavelet Transform (WT) is employed to analyze and detect the transient physical behavior of loads during the load identification. Finally, in [20] it is possible to see a last example. In that work the WT is used to identify simpler electrical consumption patterns as a part of total consumption curve.

III. EVENT DETECTION IMPLEMENTATION

A. Data-set

To evaluate the performance and robustness of the studied approaches, we used the sub-metering system presented in [21] as a tool for data collection. Fig. 1 presents the architecture of the sub-metering system. In the proposed architecture, two main parts can be differentiated: an embedded and electronics part and a microprocessor and software part.

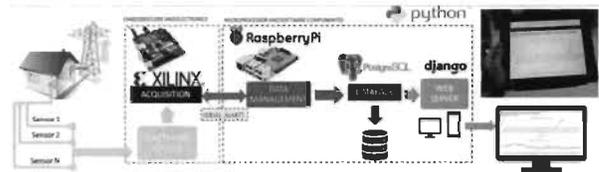


Fig. 1. Simplified diagram of proposed sub-metering system [21].

For this work, we used one electricity load data-set (time resolution of 10 samples per second) of ten (10) different weeks from an occupied house in Trois-Rivières (Canada). The data-set is divided into three non-overlapped subsets as detailed in Table I. The training subset is used for feature selection, training of the ANN and to calculate the template for the Matched Filters. The validation subset is used to verify the performance of the ANN during training and also it is used to stop the training if the validation error increases or does not change for 10 consecutive epoch in order to selecting the best setting of the ANN. In the same way, the tests’ subset is used to generalize the classification and to evaluate the classification accuracy with a different data set than the one used for training and validation [21].

TABLE I
DIVISION OF THE DATASET.

Training subset	Validation subset	Tests’ subset
70% of the first three weeks	30% remaining of the first three weeks	The last seven weeks

In the literature, there are different ways to perform the signature extraction. However, in this work, the events of interest are identified from the sub-metering information obtained with the system presented in [21] and [22]. For the classification, we identified the events of interest and then we extracted a window of 300 samples for each event as illustrated in Fig. 2.

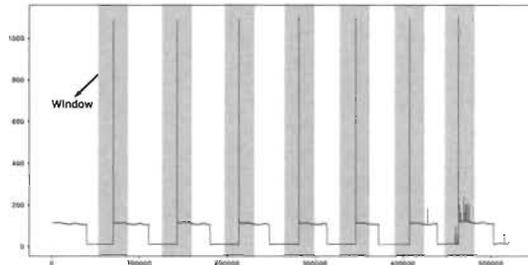


Fig. 2. Selection of the windows of each event.

B. Wavelet Neural Network (WNN)

The methodology used for the implementation of this method is shown in Fig. 3.

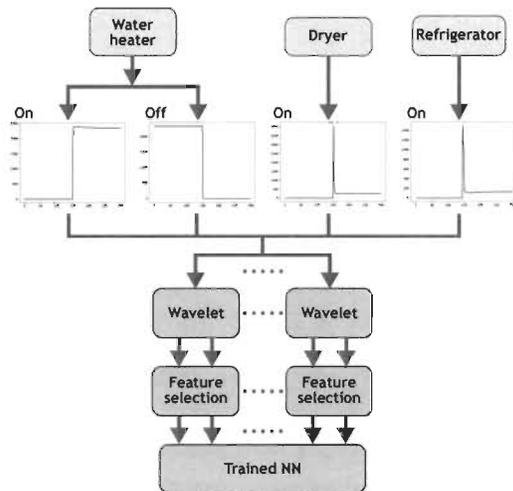


Fig. 3. Methodology used for the implementation of Wavelet Neural Network.

1) *Discrete Wavelet Transform (DWT)*: It allows to decompose the signal into a set of low frequency components called approximations and high frequency components called details. The approximations represent the general trend of the signal; the details represent irregular fluctuations [23].

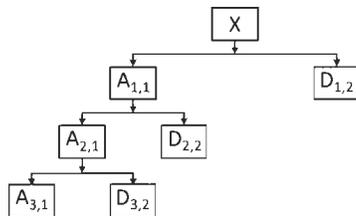


Fig. 4. DWT multilevel decomposition tree.

DWT can be implemented efficiently by repeated filtering of the signal to build a multilevel decomposition tree. Fig. 4 shows the DWT decomposition tree until depth 3. At each level l the approximation A is further decomposed to approximation and details at level $l + 1$, while the details D are not decomposed.

For the DWT decomposition we used the Haar wavelet as mother wavelet defined by (1). This is a wavelet commonly used for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines and electrical transients.

$$\varphi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Once the DWT decomposition tree has been obtained, it is necessary to identify the approximations and details that contain the most important information for the analysis to be performed. In the literature, there are different cost functions that we can use to select the best DWT decomposition sub-tree: 1) entropy [23], 2) l^1 norm [24], 3) logarithm of energy [25] and 4) the rate-distortion framework [26].

For this article, the first detail ($D_{1,2}$) and the last approximation ($A_{3,1}$) were arbitrarily selected to continue with the implementation of the method. Fig. 5 shows the components $A_{3,1}$ and $D_{1,2}$ obtained from the DWT decomposition for a refrigerator connection window. For future work, the analysis and comparison of the different ways to select the best DWT decomposition sub-tree will be carried out.

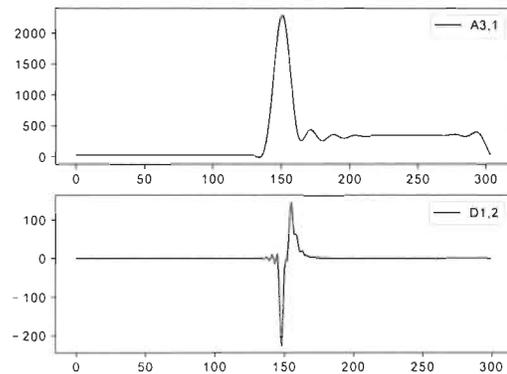


Fig. 5. DWT decomposition for a refrigerator connection window.

2) *Feature selection*: Feature selection is the process of choosing a small set of informative variables that are necessary and sufficient for accurate event classification [23]. Appropriate feature selection is one of the critical factors for successful prediction and has been an active area of research [27].

In this study the features used for the classification are shown in Table II. The feature selection is conducted separately for each wavelet component.

3) *Classification model*: Artificial neural networks (ANNs) are computing systems inspired by the biological neural networks that constitute animal brains [28]. ANNs are suitable

TABLE II
SELECTED FEATURES FOR EACH WAVELET COMPONENT.

Wavelet Component	Feature number	Feature description
A3,1	F1	Standard deviation
	F2	Maximum power
	F3	Variance
	F4	Skewness
D1,2	F5	Standard deviation
	F6	Maximum power
	F7	Variance
	F8	Skewness

as they can model complex linear and non-linear relationships between the predictor variables and the target variable, and can learn these relationships from examples, instead of fitting a model based on prior assumption about the input-output relationship [23].

We used a multi-layer perceptron neural network with one hidden layer as depicted in Fig. 6 for the event classification. The NN has eight (8) input neurons, corresponding to the features from Table II, and four (4) outputs neurons, corresponding to the results of classification for the four (4) events. The number of neurons in the hidden layer is set to ten (10).

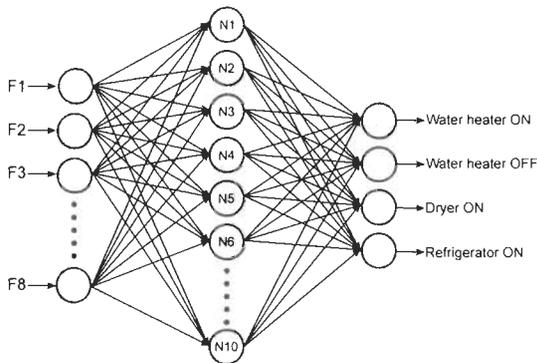


Fig. 6. Result of training and validation of the ANN.

As training algorithm we employed the Momentum algorithm and the error function used was the Mean Squared Error (MSE). The ANN training process stops when there is no improvement in the error of validation during 10 consecutive epochs or if the error of validation increases for 10 consecutive epochs or a maximum number of 1000 epochs is reached.

C. Matched filters approach

In signal processing, a matched filter is equivalent to convolving the unknown signal with a conjugated time-reversed version of a template. The matched filter will maximize the signal-to-noise ratio (SNR) in the presence of additive stochastic noise [29].

Fig. 7 shows the methodology used for the implementation of this method.

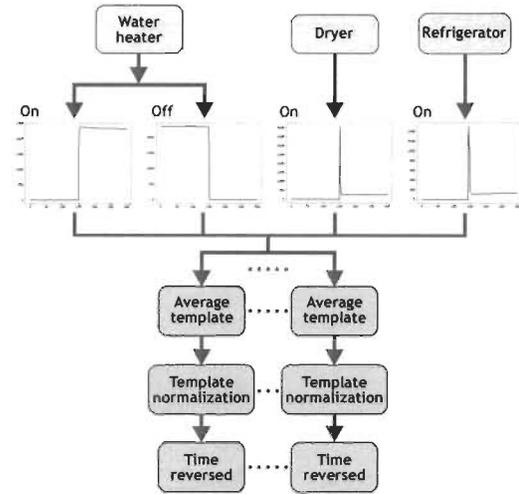


Fig. 7. Methodology used for the implementation of Matched Filters

1) *Template extraction:* We manually identify all templates in the training and validation subsets and then we calculate an average template as illustrated in Fig. 8.

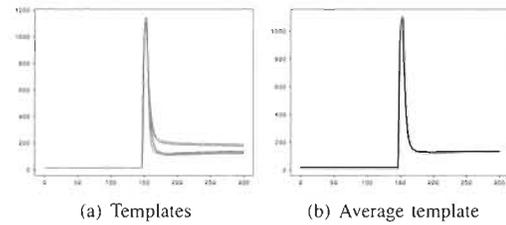


Fig. 8. Example of the calculation of the average template.

2) *Normalization:* Normalization is important to avoid adding DC components to the signal. An “AC coupled” and normalized version of the template, can be computed as:

$$x_{norm} = \frac{x - \frac{1}{N} \sum_{i=0}^{N-1} x[i]}{\left\| x - \frac{1}{N} \sum_{i=0}^{N-1} x[i] \right\|} \quad (2)$$

where the denominator in (2) is the 2-norm of the numerator [14]. Thus, x_{norm} is a unit length vector with zero mean (Fig. 9a). After that, it is necessary to calculate the time-reversed version of the normalized version of the template (Fig. 9b).

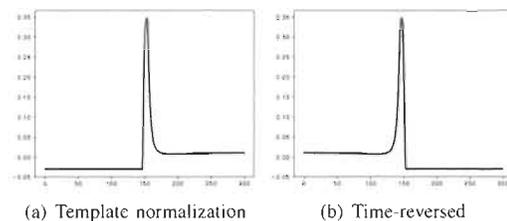


Fig. 9. Example of the normalized and time-reversed template.

3) *Classification model:* For the classification, it is necessary to do the convolution into the input signal with a time-

reversed version of all four templates simultaneously (Fig. 10). After that, the results of the convolution are compared to the better fitting. Notice that, the maximum value among the results corresponds to the better match between the input and the filter.

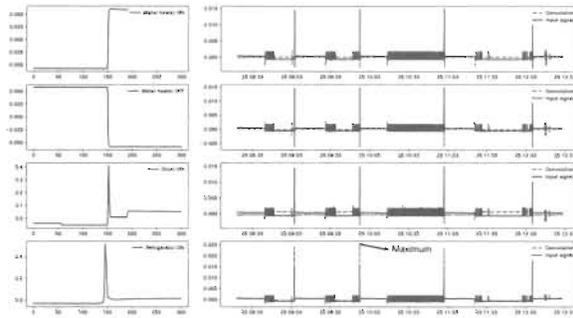


Fig. 10. Classification model with the Matched filters approach.

IV. EXPERIMENTS AND RESULTS

The Fig. 11 shows the result of the training and validation of the ANN. In order to avoid neural network overfitting, at the end of each epoch, the validation database is used to verify that the new ANN architecture is working properly. We can see that the validation error decreases during the whole training process and because of this the training is only stopped at the last epoch.

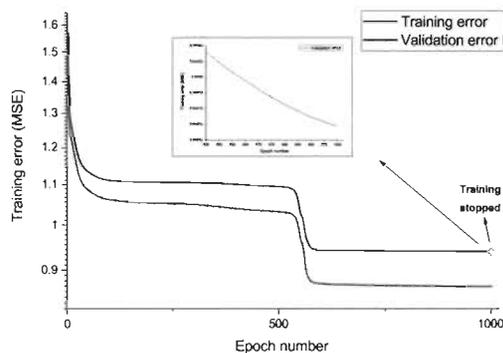


Fig. 11. Result of neural network training.

To verify the performance of the methods studied, a first validation with the original data is performed. Then, to verify the robustness of the classifiers methods, random white noise with a normal (Gaussian) distribution is added to the tests' subset (Fig. 12) and a new event classification is performed.

Fig. 13 shows the results of the classification (confusion matrix) obtained with the tests' subset. Figs. 13a and 13b show the results of the classification using the original raw data. The results in Fig. 13c and 13d are obtained after adding white noise to the tests' subset (Fig. 12).

The Fig. 13a shows an 99.45% of accuracy in the classification with the Wavelet Neural Network approach. Likewise, with the Matched Filters approach an accuracy of 99.33% was

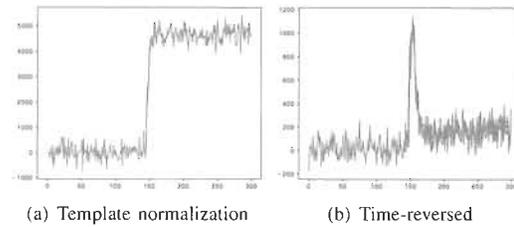


Fig. 12. Information with White Noise.

obtained (Fig. 13b). A total of 2547 events were used for validation.

The results obtained with the noisy signals shows that under the Wavelet Neural Network approach the accuracy decreases to 96.89%. This is due to the fact that the template of the dryer connection is very similar to the template of the refrigerator connection, which causes the mistakes in the classification (Fig. 13c).

Similarly, Fig. 13d shows that with the noisy signals the accuracy of Matched Filters approach decreases to 98.74%. In this case, the confusion between the dryer connection and the refrigerator connection is less than with the Wavelet Neural Network approach.

V. CONCLUSIONS

This paper shows the comparison between two different methods used for the classification of transient events as the first step to build a model of occupancy and residential loads from the aggregated measure of consumption at the electric board panel level. We described the step by step process for the implementation of the two approaches and the validation was done using a laboratory test bench and data of occupied residential building.

Comparison of experimental results with the original data shows that Wavelet Neural Network approach and Matched Filter approach provide similar performance (99.45% and 99.33% of accuracy respectively). However, the results obtained with the experimental database with noise show that although the number of errors increases with the two methods analyzed, the classification is better with Matched Filter approach (98.74% of accuracy). Nevertheless, in both cases the accuracy of classification remains higher than 96.89%.

With the wavelet analysis it is possible to break down the signals into different frequency components, which allows a deeper analysis of the signals and also makes possible to extract discriminant parameters that allow the classification of events in an easier and precise way. Also, Matching Filter is an interesting technique that it is easy to implement and it is also possible to parallelize. Which allows to increase the calculation speed with less complexity compared to ANNs.

The utilization of detailed information of power consumption makes possible to perform deeper analysis of the appliances behavior utilization. Also, the sub-metering system permitted the implementation of the methods analyses in this paper, in order to achieve on-line experimental validation.

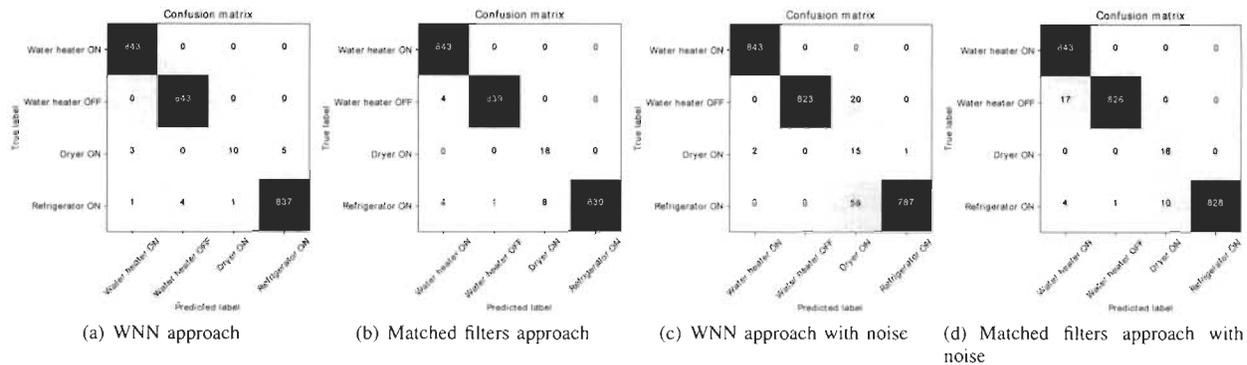


Fig. 13. Results of the classification.

ACKNOWLEDGMENT

This work was supported in part by the Laboratoire des technologies de l'énergie (LTE) d'Hydro-Québec, the Natural Science and Engineering Research Council of Canada and the Fondation UQTR.

REFERENCES

- [1] S. Tatsa and M. en Économique, "Modélisation et prévision de la consommation horaire d'électricité au québec," Master's thesis, Université Laval, 2013.
- [2] S. Aman, Y. Simmhan, and V. K. Prasanna, "Energy management systems: state of the art and emerging trends," *IEEE Communications Magazine*, vol. 51, no. 1, pp. 114–119, 2013.
- [3] W. Tushar, C. Yuen, B. Chai, S. Huang, K. L. Wood, S. G. Kerk, and Z. Yang, "Smart grid testbed for demand focused energy management in end user environments," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 70–80, 2016.
- [4] G.W. Hart, "Residential energy monitoring and computerized surveillance via utility power flows," *IEEE Technology and Society Magazine*, vol. 8, no. 2, pp. 12–16.
- [5] S. B. Leeb, "A conjoint pattern recognition approach to nonintrusive load monitoring," Ph.D. dissertation, Massachusetts Institute of Technology, 1993.
- [6] G. Tang, K. Wu, J. Lei, and W. Xiao, "The meter tells you are at home! non-intrusive occupancy detection via load curve data," in *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on*. IEEE, 2015, pp. 897–902.
- [7] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-intrusive occupancy monitoring using smart meters," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.
- [8] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.
- [9] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [10] K. D. Anderson, M. E. Bergés, A. Océanu, D. Benitez, and J. M. Moura, "Event detection for non intrusive load monitoring," in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2012, pp. 3312–3317.
- [11] C. C. Yang, C. S. Soh, and V. V. Yap, "Comparative study of event detection methods for non-intrusive appliance load monitoring," *Energy Procedia*, vol. 61, pp. 1840–1843, 2014.
- [12] D. Luo, L. K. Norford, S. R. Shaw, and S. B. Leeb, "Monitoring hvac equipment electrical loads from a centralized location—methods and field test results/discussion," *ASHRAE Transactions*, vol. 108, p. 841, 2002.
- [13] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, and K. Anderson, "User-centered nonintrusive electricity load monitoring for residential buildings," *Journal of computing in civil engineering*, vol. 25, no. 6, pp. 471–480, 2011.
- [14] S. Leeb, S. Shaw, and J. K. Jr, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, no. 3, pp. 1200–1210.
- [15] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1445–1454, 2008.
- [16] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE power and energy magazine*, vol. 99, no. 2, pp. 56–63, 2003.
- [17] Y. Jimenez, C. Duarte, J. Petit, and G. Carrillo, "Feature extraction for nonintrusive load monitoring based on s-transform," in *Power Systems Conference (PSC), 2014 Clemson University*. IEEE, 2014, pp. 1–5.
- [18] K. S. Barsim, R. Streubel, and B. Yang, "An approach for unsupervised non-intrusive load monitoring of residential appliances," in *Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring*, 2014.
- [19] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, 2012.
- [20] F.-J. Ferrández-Pastor, J.-M. García-Chamizo, M. Nieto-Hidalgo, V. Romacho-Agud, and F. Flórez-Revuelta, "Using wavelet transform to disaggregate electrical power consumption into the major end-uses," in *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer, 2014, pp. 272–279.
- [21] C. Guzman, L. Rueda, G. Romero, S. Biscans, K. Agbossou, and A. Cardenas, "Enabling winter behavior analysis on electrically heated residential buildings by smart sub-metering," *2018 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1213–1218, 2018.
- [22] A. Cardenas, K. Agbossou, and C. Guzmán, "Development of real-time admittance analysis system for residential load monitoring," in *Industrial Electronics (ISIE), 2016 IEEE 25th International Symposium on*. IEEE, 2016, pp. 696–701.
- [23] M. Rana and I. Koprinska, "Forecasting electricity load with advanced wavelet neural networks," *Neurocomputing*, vol. 182, pp. 118–132, 2016.
- [24] L. Brechet, M.-F. Lucas, C. Doncarli, and D. Farina, "Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2186–2192, 2007.
- [25] M. V. Wickerhauser, "Lectures on wavelet packet algorithms," in *Lecture notes, INRIA*. Citeseer, 1991.
- [26] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 160–175, 1993.
- [27] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [28] F. B.-M. Verleysen and F. BLAYO, "Les réseaux de neurones artificiels," *Presses Universitaire De France*, 1996.
- [29] J. C. Bancroft, "Introduction to matched filters," *CREWES Research*, 2002.