



HAL
open science

Archétypes de profils d'occupation dans le résidentiel : entre emploi du temps et consommation énergétique

Alexis Wagner, Marie Ruellan, Romain Bourdais

► To cite this version:

Alexis Wagner, Marie Ruellan, Romain Bourdais. Archétypes de profils d'occupation dans le résidentiel : entre emploi du temps et consommation énergétique. Conférence Francophone de l'International Building Performance Simulation Association (IPBSA), May 2024, La Rochelle, France. hal-04738525

HAL Id: hal-04738525

<https://univ-rennes.hal.science/hal-04738525v1>

Submitted on 15 Oct 2024

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

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

Archétypes de profils d'occupation dans le résidentiel : entre emploi du temps et consommation énergétique

Alexis Wagner^{1,*}, Marie Ruellan², Romain Bourdais¹

¹Institut d'Electronique et des Technologies du numéRique, CentraleSupélec
Avenue De Boulaie, Cesson-Sévigné, 35700, France

² Systèmes et Applications des Technologies, de l'Information et de l'Electricité, CY Cergy Paris
Université

Boulevard du Port, Cergy, 95000, France

*alexis.wagner@centralesupelec.fr

RÉSUMÉ. La classification d'enquêtes emploi du temps est une méthodologie permettant de proposer des modèles d'occupation représentatifs de la diversité des individus occupant le parc de logements. La distance de classification est un outil influençant fortement le résultat de la classification. Dans la littérature, c'est principalement la distance de Hamming qui est utilisée. Cependant, cette distance est très peu questionnée. Ce travail propose une méthodologie basée sur une distance de classification multicritère. Cette distance est constituée d'une pondération entre la distance de Hamming et une distance énergétique basée sur la consommation du poste de chauffage de chaque individu composant la base de données. Cette consommation est obtenue à partir d'une simulation thermique par gestion prédictive du chauffage. Finalement, la classification des individus issus de l'enquête emploi du temps française de 2010 a montré que l'intégration du critère énergétique facilite le choix du nombre de clusters et réduit la dispersion des consommations en chauffage au sein d'un cluster de 20 %.

MOTS-CLÉS. archétypes, commande prédictive, enquêtes emploi du temps, consommations, k-modes

1. CONTEXTE

Pour donner suite à la Stratégie nationale de réduction des émissions de gaz à effet de serre, la France s'engage à devenir neutre en carbone d'ici 2050. Dans cette optique, le secteur du bâtiment offre des perspectives d'économies d'énergie et de réduction des émissions particulièrement prometteuses puisqu'il représente environ 20 % des émissions nationales de GES et 40% de la consommation nationale d'énergie primaire (MTES,2015). Néanmoins, la mise en oeuvre de mesures de rénovation centralise des moyens humains et économiques conséquents. Il est nécessaire d'intégrer, en amont du processus, l'évaluation de leur efficacité environnementale et économique. L'enjeu central est donc de calculer la consommation des logements avant et après mise en oeuvre de la mesure de rénovation.

2. ETAT DE L'ART

2.1. L'ÉVALUATION DE MESURES DE RÉNOVATION À L'ÉCHELLE DU PARC DE LOGEMENTS

Dans la littérature, plusieurs travaux ont quantifié l'impact environnemental de solutions techniques de rénovation à l'échelle du parc. Certains de ces travaux utilisent des modèles basés sur la physique pour prédire la consommation des logements avant et après la mise en oeuvre de ces solutions techniques (Yang et al., 2022).

Les modèles physiques se basent sur les simulations thermique et électrique du logement. On observe deux inconvénients majeurs freinant leur exploitation pour l'évaluation à grande échelle du parc de logements. D'une part, les simulateurs physiques nécessitent d'intégrer des modèles réalistes de comportements de ménages. Pour éclaircir la suite de notre propos, nous définissons une situation d'habitation, comme l'association d'un bâtiment, de conditions extérieures et d'un ménage (Heinrich, 2024). Le second inconvénient des modèles physiques dans le cadre de l'évaluation de mesures de rénovation à grande échelle est que, chaque situation d'habitation traitée nécessite une simulation, avant et après mise en oeuvre de la rénovation. Traiter, au cas par cas, chaque situation d'habitation du parc français demeure un challenge quasiment impossible. Pour lever ce verrou, plusieurs travaux ont restreint leur étude à la simulation d'archétypes (Ben and Steemers, 2020; Ohene et al., 2022; Torabi Moghadam et al., 2019; Dascalaki et al., 2011). Les archétypes représentent des bâtiments, ménages ou conditions extérieures fictifs mais représentatifs de segments homogènes. L'hypothèse centrale est que la simulation d'un archétype fournit des informations sur l'ensemble des éléments composant son segment. Le nombre de simulations à réaliser est donc nettement réduit. Ainsi, Les bâtiments français constituant le parc peuvent être étudiés par l'intermédiaire des archétypes proposés par le projet TABULA (Loga et al., 2016). L'étude de l'influence du climat peut être restreinte à l'étude des zones climatiques françaises utilisées dans les réglementations thermiques (Ministère de la Transition Écologique, 2020). Cependant, à notre connaissance, il n'existe pas d'archétypes de comportement représentatifs des ménages français intégrables dans des simulations de modèles physiques.

2.2. LA CLASSIFICATION DE DONNÉES D'ENQUÊTE EMPLOI DU TEMPS

L'occupation du logement par un ménage influence fortement sa consommation. Pour prédire cette dernière avec précision, les simulateurs doivent intégrer des modèles réalistes d'occupation. Le challenge majeur dans la construction de ces modèles réside dans la variabilité du comportement d'un même individu. Ainsi, pour capter cette variabilité, des travaux ont développé des modèles d'occupation stochastiques à partir des données issues d'Enquête Emploi du Temps (Institut National de la Statistique et des Études Économiques (INSEE), 2010). Ces données sont constituées du carnet d'activité sur une journée de milliers d'individus. Ce type d'enquête a été mis en place dans de nombreux pays et leur utilisation est largement répandue pour l'étude des comportements de ménages (Osman and Ouf, 2021). A partir de ces données, les auteurs de (Wilke et al., 2011) ont donc proposé un modèle d'occupation de ménage stochastique défini par une loi de Bernoulli. Un tirage est réalisé pour déterminer l'état d'occupation à l'instant t_k et un autre pour estimer la durée de l'état. Néanmoins, l'enchaînement des activités présente des incohérences. En effet, dans ce travail, l'état d'occupation à l'instant t_k ne dépend pas de l'état à l'instant précédent t_{k-1} . Améliorant ce modèle, les auteurs de (Richardson et al., 2008) ont intégré cet effet mémoire en définissant l'occupation comme une chaîne de modèles de Markov d'ordre 1. La probabilité de rester ou de changer d'état dépend directement de l'état précédent. Il en résulte des enchaînements d'activités plus cohérents.

Néanmoins, le modèle développé par les auteurs de (Richardson et al., 2008) n'est pas représentatif de la diversité des individus occupant le parc de logements. En effet, tous les individus ont servi à l'estimation des paramètres de la chaîne de Markov. Pour améliorer la représentativité des modèles d'occupations, les auteurs de (Vorger, 2014) ont développé une méthodologie dont les entrées sont les variables descriptives du ménage étudié et dont la sortie est la chaîne de Markov représentative de son occupation. Cependant, bien qu'adapté pour des études au cas par cas, le nombre important de paramètres d'entrée ne permet pas d'intégrer

ces modèles pour l'évaluation à grande échelles de mesures de rénovation. D'autres travaux nombreux travaux ont ajouté avant l'étape d'identification une étape de classification. L'étape de classification a pour objectif de regrouper les profils d'occupation similaires entre eux. Pour chaque groupe, les paramètres de la chaîne de Markov sont estimés à partir des profils d'occupation des individus le composant. Ces travaux ont montré des résultats notables. En effet, les auteurs de (Sood et al., 2023) ont montré des écarts de consommation allant de 25 % à 35 % entre des modèles d'archétypes d'occupation stochastiques basées sur des données réelles et déterministes (ASHRAE 90.1 (Abushakra et al., 2001)). Les auteurs de (Buttitta et al., 2019) ont associé des archétypes d'occupation à des archétypes de logement et ont montré des écarts de consommation avec des scénarios d'occupation déterministes allant jusqu'à 30 %. Les auteurs de (Mitra et al., 2020) ont quantifié, quant à eux, des écarts allant jusqu'à 41 % pour des archétypes basés sur les données d'enquête emploi du temps américaines. Principalement, deux algorithmes de classification sont utilisés pour regrouper les profils d'occupation: l'algorithme k-modes (Diao et al., 2017; Buttitta et al., 2017) et les algorithmes de clustering hiérarchique (Aerts et al., 2014). Cependant, bien que considéré comme un facteur déterminant, la distance caractérisant la similitude entre deux scénarios est peu questionnée dans les travaux précédents. Majoritairement, ces travaux utilisent la distance de Hamming, quantifiant le temps d'activité commun entre deux profils. On retrouve une distance différente uniquement dans les travaux (Diao et al., 2017) avec l'utilisation de la distance de Levenshtein. Ces travaux supposent donc, à travers leur distance de classification, que la similarité des comportements des individus d'un même groupe engendre une consommation homogène au sein de ce groupe.

Cette observation nous questionne sur l'utilisation de la distance de Hamming pour la classification. **Est-il possible de définir une distance offrant un meilleur regroupement des comportements permettant d'estimer de façon robuste les consommations énergétiques après simulation ?** Ce travail se distingue en deux parties. Dans un premier temps, à partir des profils d'activités des individus interrogés lors de l'enquête emploi du temps réalisée par l'INSEE en 2010, d'estimer pour chaque profil la consommation énergétique du chauffage par commande prédictive linéaire. Cette consommation est par la suite intégrée dans un algorithme de classification multicritère dont la distance entre deux individus est définie par une pondération entre leur similitude en termes d'activité et la proximité de leur consommation en chauffage.

3. MÉTHODOLOGIE

La schéma de principe de la méthodologie développée est présenté figure 1. Un premier tri de la base de données est réalisé. Pour chaque profil d'occupation la consommation de chauffage est évaluée. Le chauffage est piloté par commande prédictive linéaire avec gestion de l'intermittence. La consommation est estimée pour un unique cas d'étude composé d'un modèle thermique de bâtiment et de conditions extérieures. Les consommations énergétiques et les profils d'occupation sont associés, par méthode k-modes. Les individus composant chaque groupe permettent d'identifier les paramètres d'une chaîne de Markov d'ordre 1.

3.1. JEU DE DONNÉES ET TRI

Les données utilisées sont issues de l'enquête française emploi du temps réalisée en 2010. Cette enquête a interrogé 12 000 ménages pour obtenir le déroulement des activités qu'ils ont réalisé avec un pas de temps de 10 minutes. Pour ce travail, ces activités ont été classifiées en 3 états : présent, sommeil et absent. De plus, uniquement les carnets des individus habitant dans une maison individuelle ont été sélectionnés dans ce travail. La figure 2 présente la distribution moyenne des profils d'activité des individus sélectionnés dans la base de données.

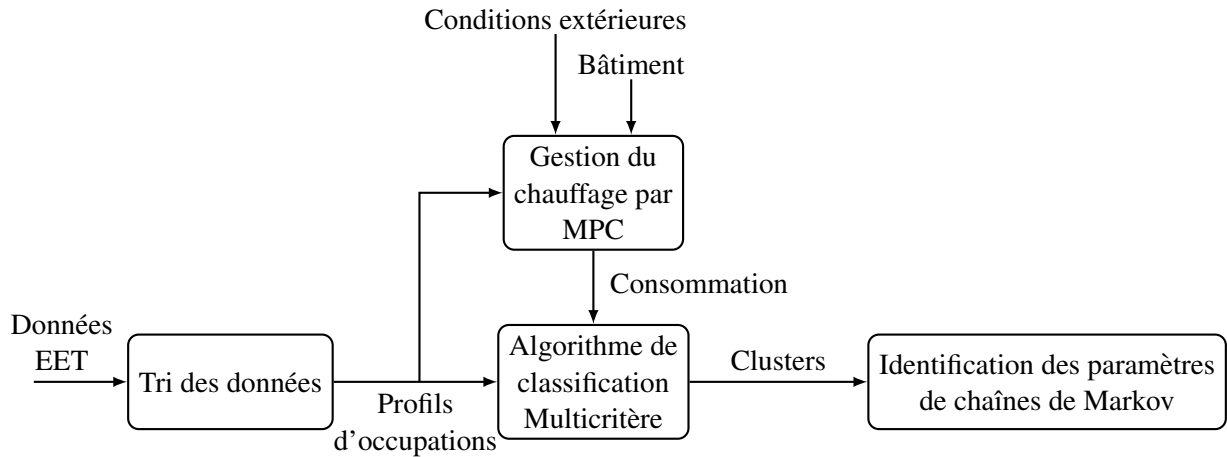


Figure 1: Méthodologie de classification de profils d'occupation avec intégration d'un critère énergétique

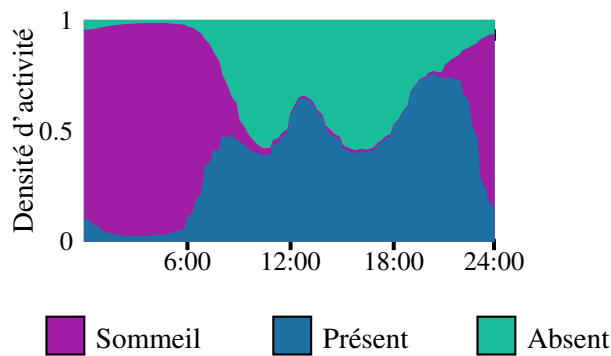


Figure 2: Densité moyenne d'activité des individus habitant une maison individuelle

3.2. GESTION DU CHAUFFAGE PAR COMMANDE PRÉDICTIVE LINÉAIRE (MPC)

3.2.1. Cas d'étude

Le logement individuel étudié correspond à un archétype de maison individuelle moyennement isolée. Ce modèle a été construit à partir des données du projet Tabula belge. Ce projet européen classe le parc de logements de 20 pays en segments homogènes (Loga et al., 2016). Notre choix s'est porté sur un modèle d'archétype de bâtiment belge car, à notre connaissance, il n'existe pas de modèles thermiques d'archétypes de bâtiments français validés par la communauté scientifique. De plus, le parc de logements belges a subi une évolution de sa réglementation thermique similaire à la France. Les logements inclus dans un même segment sont caractérisés par leur année de construction et leur type d'habitation (maison individuelle, mitoyenne, logement collectif, immeuble). Le modèle utilisé est explicité dans (Reynders et al., 2014). Sa particularité est qu'il est bi-zone, la partie 'jour' du logement est séparée de la partie 'nuit' par un étage (figure 3). Pour chaque zone, la température de l'air intérieur est désignée par T_{air} , la température des murs en contact avec l'extérieur par T_m , celle des murs intérieurs T_{mi} , et d'étage T_{e0} et T_{e1} . La partie jour échange de la chaleur avec le sol dont la température est désignée par T_s . Le renouvellement de l'air intérieur est modélisé à travers la résistance R_{vent} . Cette résistance thermique inclut également la résistance thermique des fenêtres. La

température de l'air extérieure est notée T_{ext} .

De plus, l'air intérieur des deux zones est directement chauffé par convection par le système de chauffage. Les autres murs et sols sont chauffés par rayonnement. Les consignes en chauffage de la partie 'jour' et de la partie 'nuit' peuvent être différentes soit:

$$u = \begin{bmatrix} u_j \\ u_n \end{bmatrix} \quad (1)$$

On observe les températures de l'air intérieur, à la fois de la zone 'jour' et de la zone 'nuit', soit:

$$Y = \begin{bmatrix} T_{air_j} \\ T_{air_n} \end{bmatrix} \quad (2)$$

Les fenêtres et les murs absorbent les rayonnements solaires décomposés selon les 4 points cardinaux tel que:

$$\Phi_s = \begin{bmatrix} \Phi_{nord} \\ \Phi_{est} \\ \Phi_{sud} \\ \Phi_{ouest} \end{bmatrix} \quad (3)$$

Le modèle d'état linéaire associé peut s'écrire sous sa forme discrète à partir de la figure 3:

$$X_{k+1} = AX_k + Bu_k + E_T T + E_\Phi \Phi_s \quad (4)$$

$$Y_k = CX_k \quad (5)$$

- $X \in \mathbf{R}^9$: vecteur d'état
- $A \in \mathbf{R}^{9 \times 9}$: matrice d'état
- $E_T \in \mathbf{R}^{9 \times 2}$: matrice de perturbation (T_{ext}, T_s)
- $E_\Phi \in \mathbf{R}^{9 \times 2}$: matrice de perturbation (Φ_s)

La série temporelle de température extérieure sélectionnée est représentative d'une journée de janvier de l'archétype de la zone climatique H2B (La Rochelle) (figure 4a). Les apports solaires sont obtenus à partir de PVSYS pour cette même zone.

3.2.2. Commande prédictive linéaire

L'algorithme de gestion de chauffage étudié est une commande prédictive linéaire avec gestion de l'intermittence (Morosan et al., 2010). A chaque instant k , le contrôleur détermine la séquence de commande de la puissance de chauffe $u_{K|k}$ sur un horizon de prédiction de longueur K , séquence minimisant un critère d'optimisation. Seul le premier élément de la séquence est appliqué au système et un nouveau calcul est effectué au pas de temps suivant. La contrainte à respecter, $W_{T_{K+1|k}}$, est liée au confort lorsque l'occupant est présent dans le logement, indication fournie par la matrice d'occupation $OCC_{K+1|k}$. Pour garantir la faisabilité du problème d'optimisation, cette contrainte est relâchée par une variable d'écart, notée $\epsilon_{T_{K+1|k}}$ et pénalisée

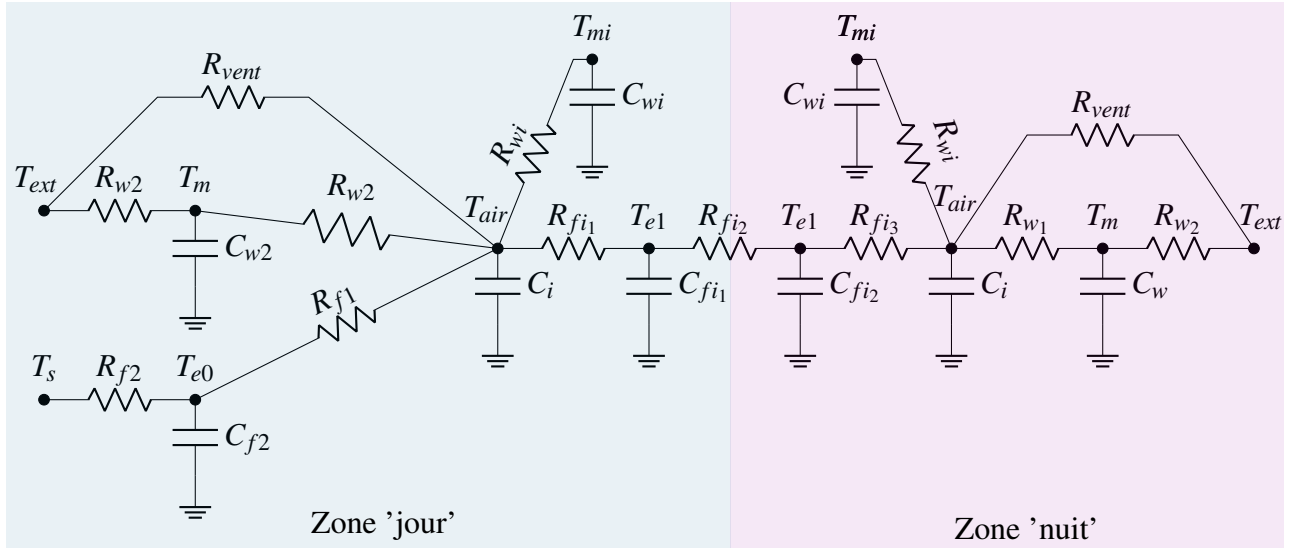


Figure 3: Modèle thermique bi-zone issu de (Reynders et al., 2014)

si la contrainte n'est pas respectée. Le problème d'optimisation prend la forme suivante :

$$\begin{aligned} \min \quad & J(u_{jK|k}, u_{nK|k}, \varepsilon_{T_{jK|k}}, \varepsilon_{T_{nK|k}}) = \mathbf{1}_{N_p}^T u_{jK|k} + \mathbf{1}_{N_p}^T u_{nK|k} + \text{pen}_T \mathbf{1}_{N_p}^T \varepsilon_{T_{jK|k}} + \text{pen}_T \mathbf{1}_{N_p}^T \varepsilon_{T_{nK|k}} \\ \text{soumis à:} \quad & 0 \leq u_{jK|k} \leq u_{j\max} \\ & 0 \leq u_{nK|k} \leq u_{n\max} \\ & T_{\text{air}_{jK+1|k}} + \varepsilon_{T_{jK+1|k}} \geq \text{OCC}_{jK+1|k} W_{T_{jK+1|k}} \\ & T_{\text{air}_{nK+1|k}} + \varepsilon_{T_{nK+1|k}} \geq \text{OCC}_{nK+1|k} W_{T_{nK+1|k}} \\ & \varepsilon_{T_{jK|k}} \geq 0 \\ & \varepsilon_{T_{nK|k}} \geq 0 \end{aligned}$$

De manière identique à la commande prédictive présentée chapitre 3, le modèle de prédiction choisit est le même modèle d'état que celui utilisé pour la simulation thermique. La propagation du modèle d'état (1) permet d'exprimer les sorties prédites en fonction de l'état courant X_k . Soit la température de l'air sur l'ensemble de l'horizon de prédiction :

$$Y_{K|k} = FX_k + H_u u_{K|k} + H_T T_{K|k} + H_\Phi \Phi_{K|k} \quad (6)$$

$$F = \begin{bmatrix} CA \\ \vdots \\ CA^{N_p} \end{bmatrix}, \quad H_u = \begin{bmatrix} CB_u & 0 & 0 & \cdots & 0 \\ CAB_u & CB_u & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N_p-1}B_u & CA^{N_p-2}B_u & CA^{N_p-3}B_u & \cdots & CB_u \end{bmatrix}$$

$$H_T = \begin{bmatrix} CB_t & 0 & 0 & \cdots & 0 \\ CAB_t & CB_t & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N_p-1}B_t & CA^{N_p-2}B_t & CA^{N_p-3}B_t & \cdots & CB_t \end{bmatrix}$$

3.3. SIMULATION THERMIQUE

Une simulation thermique de la commande prédictive précédente est appliquée à l'ensemble des profils d'occupation de la base de données. Chaque simulation est réalisée pour le même bâtiment et les mêmes conditions extérieures. Les consignes en température de chaque zone sont regroupées dans le tableau suivant:

| | Contrainte | $OCC = 1$ | $OCC = 0$ |
|-------------|------------|-----------|-----------|
| Zone 'jour' | W_{T_j} | 20 | 16 |
| Zone 'nuit' | W_{T_n} | 18 | 16 |

On définit la consommation totale associée au i^{eme} profil d'individus conséquence de la gestion du chauffage par la commande prédictive linéaire précédente :

$$E_i = \int u_{i_j} dt + \int u_{i_n} dt \quad (7)$$

La figure 4 présente la distribution de la consommation de la zone 'jour', la zone 'nuit' et la consommation totale pour l'ensemble de la base de données.

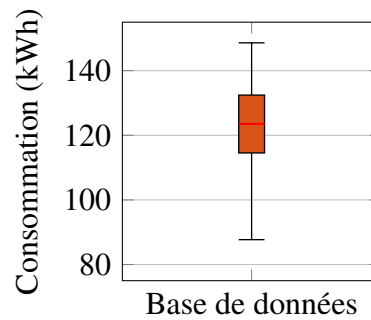


Figure 4: Distribution de la consommation totale en chauffage des individus composant la base de données

3.4. ALGORITHME DE CLASSIFICATION MULTICRITÈRE

3.4.1. *k-modes++*

L'algorithme K-modes ++ est une amélioration à l'initialisation de l'algorithme K-modes, largement répandu pour la classification de données. Son objectif est de diviser un jeu de données noté X en n_c groupes, minimisant l'inertie intra-cluster du jeu de données :

$$W(X, C) = \sum_{k=1}^{n_c} \sum_{Z_i \in C_k} d(Z_i, C_k) \quad (8)$$

- X : Jeu de données constitué des profils d'occupations sur une journée des 12600 individus
- n_c : nombre de clusters
- C : Vecteurs constitué des centres des n_c clusters
- C_k : centre du cluster k
- Z_i : i^{eme} individu
- d : distance de classification entre deux profils d'occupation à trois états

3.4.2. Distance entre deux individus

La distance choisie est l'élément central pour intégrer le potentiel d'économie d'énergie au travail de classification. L'objectif est d'obtenir des profils d'occupation à trois états distants entre eux mais minimisant également la dispersion de la consommation d'énergie des individus appartenant au même cluster. Ainsi, la distance de classification développée se décompose en deux termes pondérés par le coefficient $\lambda \in [0, 1]$: la distance de Hamming d_h et la distance "énergétique" d_E . Le profil de chaque individu est constitué de l'ensemble de ses activités (présent, sommeil et absent) sur une journée avec un pas de temps de 10 minutes et de sa consommation obtenue par simulation thermique. Soit la distance entre le profil Z_i et Z_j :

$$d(Z_i, Z_j) = (1 - \lambda) \cdot d_H(Z_i, Z_j) + \lambda \cdot d_E(Z_i, Z_j) \quad (9)$$

Distance de Hamming:

La distance de Hamming compte le nombre de dissemblances entre deux vecteurs de même taille. Pour l'application proposée, cette distance compare, à chaque pas de temps t , les activités a de deux individus considérant les m pas de temps composant les différents profils Z_i et Z_j , soit:

$$d_H(Z_i, Z_j) = \frac{1}{m} \sum_{t=1}^m \delta(a_{i_t}, a_{j_t}) \quad (10)$$

Où:

$$\delta(a_{i_t}, a_{j_t}) = \begin{cases} 1 & \text{si } a_{i_t} = a_{j_t} \\ 0 & \text{si } a_{i_t} \neq a_{j_t} \end{cases}$$

Distance énergétique:

Le critère énergétique se base sur la consommation obtenue par simulation thermique présentée dans la sous-section précédente. On définit la distance énergétique entre deux profils :

$$d_E(Z_i, Z_j) = \frac{|E(Z_i) - E(Z_j)|}{E_m - \bar{E}} \quad (11)$$

- $E(Z_i)$: consommation totale de l'individu Z_i
- E_m : consommation maximale
- \bar{E} : consommation moyenne

3.4.3. Critère de décision

Une métrique couramment utilisée pour estimer le nombre de clusters le plus adapté à la classification est l'inertie intra-cluster, caractérisant la dispersion des individus par rapport au centre du cluster. La distance de classification a deux critères distincts : la similitude des profils et leur consommation. Ainsi, on définit deux inerties pour évaluer la performance de la classification.

La variance intra-cluster associée à la distance de Hamming, W_h , pour un nombre de clusters n_c :

$$IIC_{E,N}(n_c) = \sum_{k=1}^{n_c} \sum_{Z_i \in C_k} d_H^2(Z_i, C_k) \quad (12)$$

La variance intra-cluster associée à la distance énergétique, W_E , pour un nombre de clusters n_c :

$$W_E(n_c) = \sum_{k=1}^{n_c} \sum_{Z_i \in C_k} d_E^2(Z_i, C_k) \quad (13)$$

Ces deux critères vont permettre à la fois d'estimer le nombre de clusters n_c mais aussi un coefficient de pondération λ adapté.

3.5. IDENTIFICATION DES PARAMÈTRES DE CHAÎNES DE MARKOV

Dans le cadre de notre travail, on suppose qu'un individu peut uniquement être dans un état 'sommeil', 'présent' ou 'absent'. Une fois les individus regroupés, pour chaque groupe et à chaque pas de temps, les fréquences de changement d'état sont décomptées. Ainsi, de l'instant $t - 1$ à t , l'individu peut conserver son état ou passer à un autre état. La chaîne de Markov est donc dépendante du temps. On définit la matrice de transition associée à l'instant t , $\forall t > 1$:

$$T_t = \begin{bmatrix} p_t^{11} & p_t^{12} & p_t^{13} \\ p_t^{21} & p_t^{22} & p_t^{23} \\ p_t^{31} & p_t^{32} & p_t^{33} \end{bmatrix} \quad (14)$$

Où P_t^{ji} désigne la probabilité de passer de l'état i sachant $t - 1$ à l'état j à l'instant t , soit:

$$P_t^{ji} = P(a(t) = j \mid a(t-1) = i) = \frac{n_t^{ji}}{n_t^{tot}} \quad (15)$$

$a(t)$: état de l'individu à l'instant t

n_t^{ji} : nombre de passage de l'état j à i entre l'instant $t - 1$ et t

n_t^{tot} : nombre total d'observations

4. RÉSULTATS ET DISCUSSIONS

Les données présentées précédemment sont regroupées suivant cinq valeurs de $\lambda \in \{0, 0.1, 0.15, 0.2, 0.5\}$ et pour $K \in [1, \dots, 9]$. La figure 5 représente l'inertie intra cluster énergétique normalisée, $W_{E,N}$ en fonction de l'inertie intra cluster normalisée associée à la distance de Hamming, $W_{H,N}$. La normalisation permet d'observer les deux grandeurs à la même échelle.

Pour $\lambda = 0$, la variance énergétique n'est pas systématiquement décroissante lorsque le nombre de clusters augmente. Ce résultat met en lumière que sans critère de classification énergétique, l'augmentation de K ne garantit pas une meilleure distinction énergétique des clusters.

De plus, pour $\lambda = 0.5$, la variance de Hamming diminue très peu en comparaison des autres valeurs. Les regroupements résultants sont uniquement réalisés suivant le critère énergétique. Bien qu'intéressant énergétiquement, les modèles de Markov s'en retrouvent donc dégradés.

Pour $\lambda \in \{0.1, 0.15, 0.2\}$, une enveloppe apparait représentant un compromis intéressant entre similitude des profils d'un même groupe et dispersion énergétique des consommations. L'intégration du critère énergétique stabilise également la classification pour $\lambda \in \{0.15, 0.2\}$. En effet, les deux variances sont décroissantes. La variance énergétique permet également de choisir plus facilement le nombre de clusters résultant. En effet, un nombre de clusters acceptable correspond au nombre de clusters minimal, à partir duquel l'augmentation de K diminue

peu les deux variances. Pour la suite de l'analyse, nous fixons le nombre de clusters à 5. En effet, au-delà de 5 clusters, les inerties intra-clusters de Hamming et énergétiques stagnent. L'ajout d'un cluster supplémentaire est donc peu intéressant.

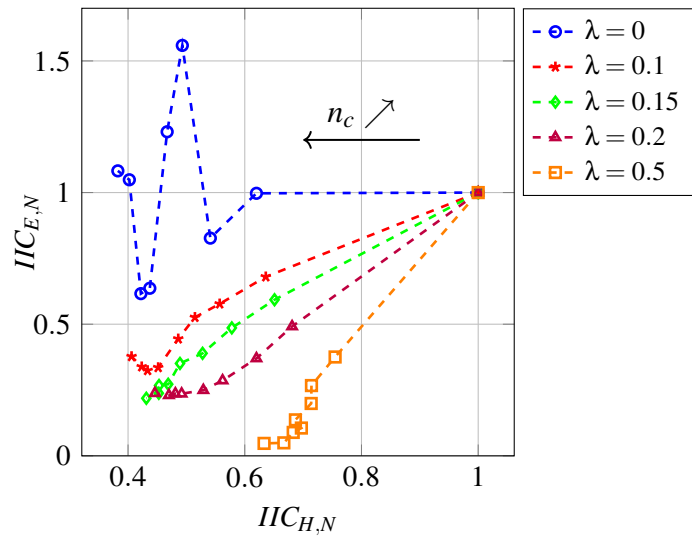


Figure 5: Inertie Intra cluster énergétique, $W_{E,N}$, en fonction de l'Inertie Intra cluster de Hamming, $W_{H,N}$

La figure 6 présente les densités moyennes des clusters résultant pour $\lambda \in \{0, 0.15\}$ pour $K = 5$. On identifie:

Pour $\lambda = 0$ pour 12663 individus:

- Cluster 1 (9.4 %):** départ en journée avec un retour tardif au logement
- Cluster 2 (23.3 %):** départ en journée retour en début de soirée
- Cluster 3 (17.5 %):** absence longue en après-midi
- Cluster 4 (14.2 %):** absence en journée avec intermittence
- Cluster 5 (35.6 %):** présence au logement toute la journée

Pour $\lambda = 0.15$ pour 12663 individus:

- Cluster 1 (10.8 %):** départ en journée avec un retour tardif au logement
- Cluster 2 (30.9 %):** départ en journée retour en début de soirée
- Cluster 3 (18.8 %):** absence longue en après-midi
- Cluster 4 (18 %):** absence le matin
- Cluster 5 (21 %):** présence au logement toute la journée

L'observation de ces clusters montre que les principaux changements lorsque λ passe de 0 à 0.15 se distinguent au niveau des clusters 3-4-5. En effet, le cluster 4 pour $\lambda = 0$ est un cluster caractérisant l'intermittence. Cependant, l'algorithme de classification pour $\lambda = 0.15$ considère qu'énergétiquement, il est plus pertinent de diviser ce cluster en 1 cluster de départ le matin (cluster 4) et un de départ l'après-midi (cluster 5). On observe également une meilleure définition du cluster 5.

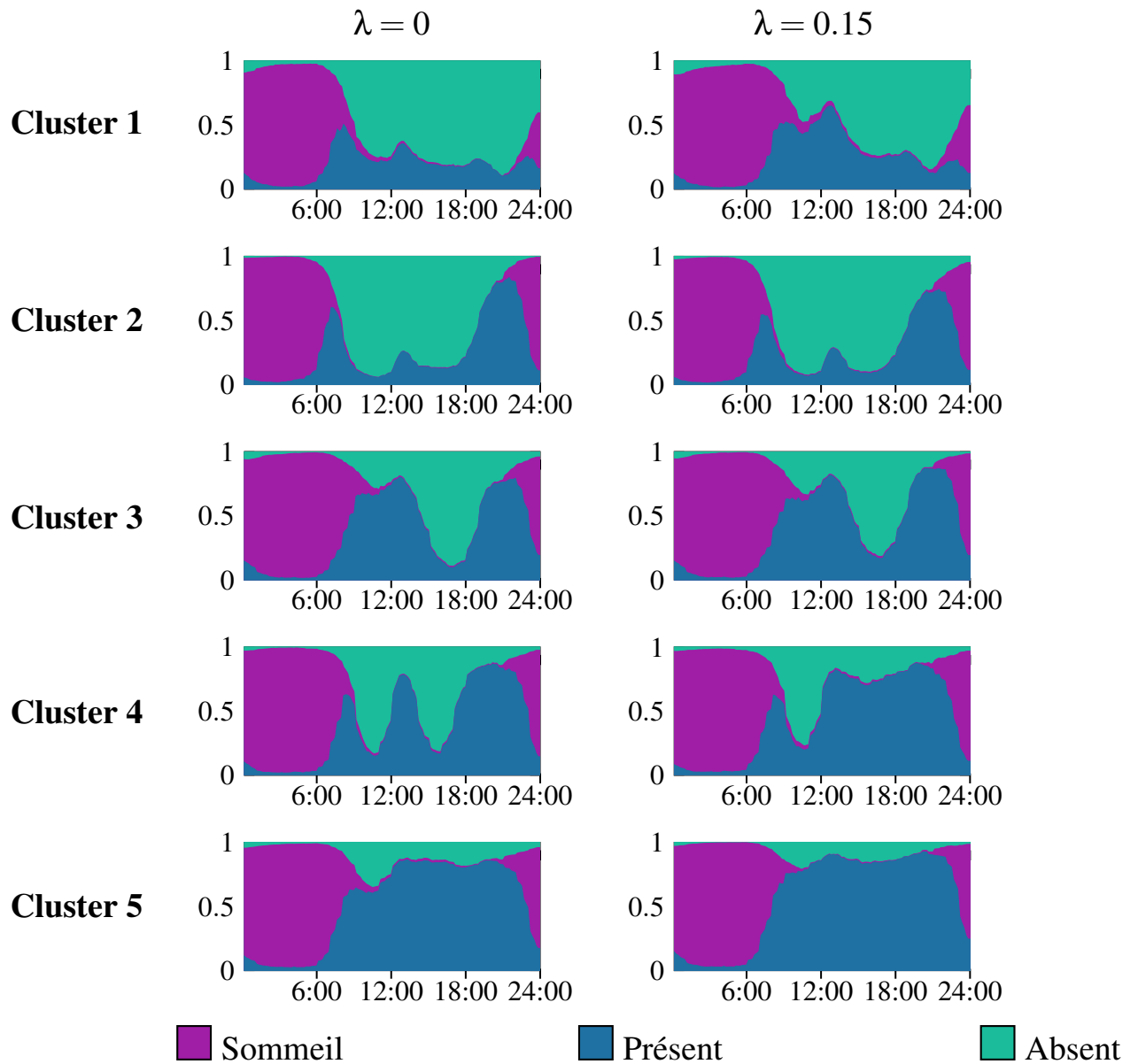


Figure 6: Densité de probabilité d'activité des 5 clusters pour $\lambda = 0$ et $\lambda = 0.15$. Les clusters sont numérotés selon leur consommation en chauffage croissante

La figure 7 présente les boîtes à moustaches des consommations des profils composant chaque cluster. Dans un premier temps, l'écart type moyen en consommation des groupes est de 8.5 kWh pour $\lambda = 0$ et de 6.8 kWh pour $\lambda = 0.15$ soit une réduction de 20 %. Il y a donc en moyenne moins de dispersion énergétique au sein des groupes.

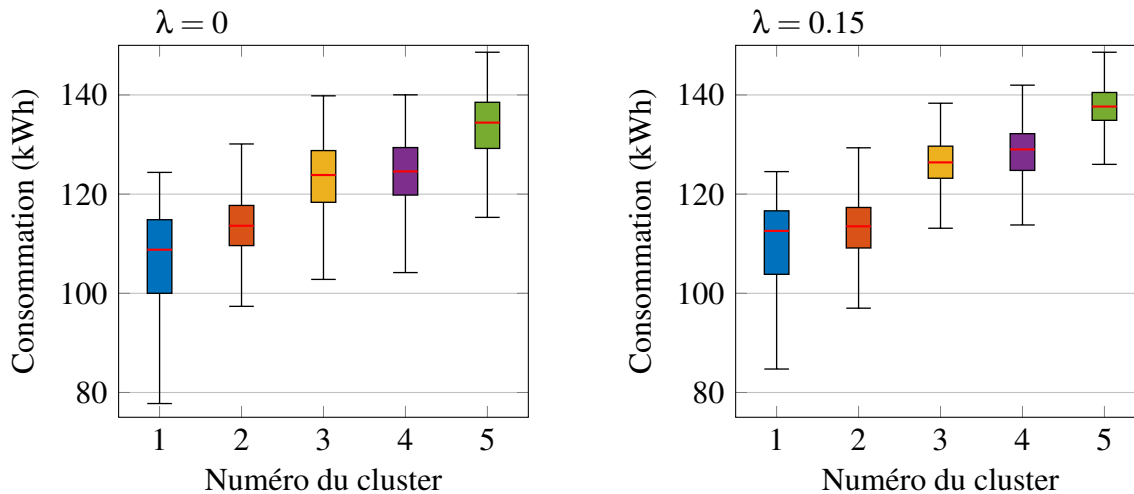


Figure 7: Boîtes à moustaches de la consommation énergétiques des individus composant les clusters

5. CONCLUSION

Ce travail propose une méthodologie de classification d'enquêtes emploi du temps basée sur une distance de classification multicritère. En effet, pour chaque profil, leur consommation est estimée par simulation thermique du chauffage piloté par commande prédictive linéaire avec gestion de l'intermittence. Les différents profils et leur consommation définissent ainsi une distance de classification multicritère. Les profils sont regroupés selon leur similitude et leur différence de consommation. Dans un premier temps, on observe que l'intégration du critère énergétique stabilise la classification facilitant la détermination du nombre de clusters. De plus, l'application aux enquêtes emploi du temps françaises a montré une réduction moyenne de la dispersion des consommations de 20 %. Ce travail apporte donc des premières perspectives intéressantes. Cette idée de construction d'un multicritère pour la classification peut être enrichie, en intégrant d'autres éléments, tels les algorithmes de pilotage, les consommations électriques, les capacités d'autoconsommation, ...

Cependant, un travail conséquent reste à réaliser pour consolider les apports méthodologiques et l'analyse des résultats. Il est notamment nécessaire de valider la méthodologie vis à vis d'autres situations d'habitation (bâtiments, conditions extérieures).

BIBLIOGRAPHIE

- Abushakra, B., Sreshthaputra, A., Haberl, J., and Claridge, D. (2001). Compilation of diversity factors and schedules for energy and cooling load calculations. Technical report, ASHRAE Research Project 1093-RP, Final Report.
- Aerts, D., Minnen, J., Glorieux, I., Wouters, I., and Descamps, F. (2014). A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and Environment*, 75:67–78.
- Ben, H. and Steemers, K. (2020). Modelling energy retrofit using household archetypes. *Energy and Buildings*, 224:110224.
- Buttitta, G., Turner, W., and Finn, D. (2017). Clustering of Household Occupancy Profiles for Archetype Building Models. *Energy Procedia*, 111:161–170.
- Buttitta, G., Turner, W. J. N., Neu, O., and Finn, D. P. (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings*, 198:84–99.
- Dascalaki, E. G., Droutsa, K. G., Balaras, C. A., and Kontoyiannidis, S. (2011). Building typologies as a tool for assessing the energy performance of residential buildings – A case study for the Hellenic building stock. *Energy and Buildings*, 43(12):3400–3409.
- Diao, L., Sun, Y., Chen, Z., and Chen, J. (2017). Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings*, 147:47–66.
- Heinrich, M. (2024). *Habitats, Habitants et Pratiques énergétiques : Intégrer les situations d’habitation dans la modélisation de la consommation d’énergie domestique*. PhD thesis, Ecole des Ponts, ParisTech.
- Institut National de la Statistique et des Études Économiques (INSEE) (2010). Enquête emploi du temps (eet). Données d’enquête.
- Loga, T., Stein, B., and Diefenbach, N. (2016). TABULA Building Typologies in 20 European countries making energy-related features of residential building stocks comparable. *Energy and Buildings*, 132.
- Ministère de la Transition Écologique (2020). Règlement Environnemental 2020. Disponible sur : <https://www.ecologie.gouv.fr/reglement-environnemental-re-2020>.
- Mitra, D., Steinmetz, N., Chu, Y., and Cetin, K. S. (2020). Typical occupancy profiles and behaviors in residential buildings in the united states. *Energy and Buildings*, 210:109713.
- Morosan, D.-P., Bourdais, R., Dumur, D., and Buisson, J. (2010). Building temperature regulation using a distributed model predictive control. *Energy and Buildings*, 42(9):1445.
- Ohene, E., Hsu, S.-C., and Chan, A. P. C. (2022). Feasibility and retrofit guidelines towards net-zero energy buildings in tropical climates: A case of Ghana. *Energy and Buildings*, 269:112252.

- Osman, M. and Ouf, M. (2021). A comprehensive review of time use surveys in modelling occupant presence and behavior: Data, methods, and applications. *Building and Environment*, 196:107785.
- Reynders, G., Diriken, J., and Saelens, D. (2014). Bottom-up modeling of the belgian residential building stock: Influence of model complexity. *International conference on system simulation in buildings edition 9, Liège, Belgium*;, pages 574–592.
- Richardson, I., Thomson, M., and Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40(8):1560–1566.
- Sood, D., Alhindawi, I., Ali, U., McGrath, J. A., Byrne, M. A., Finn, D., and O’Donnell, J. (2023). Simulation-based evaluation of occupancy on energy consumption of multi-scale residential building archetypes. *Journal of Building Engineering*, 75:106872.
- Torabi Moghadam, S., Coccolo, S., Mutani, G., Lombardi, P., Scartezzini, J.-L., and Mauree, D. (2019). A new clustering and visualization method to evaluate urban heat energy planning scenarios. *Cities*, 88:19–36.
- Vorger, (2014). *Étude de l’influence du comportement des habitants sur la performance énergétique du bâtiment*. These de doctorat, Paris, ENMP.
- Wilke, U., Haldi, F., and Robinson, D. (2011). A model of occupant’ activities based on time use survey data.
- Yang, X., Liu, S., Zou, Y., Ji, W., Zhang, Q., Ahmed, A., Han, X., Shen, Y., and Zhang, S. (2022). Energy-saving potential prediction models for large-scale building: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 156:111992.