



HAL
open science

Modélisation et simulation des comportements en lien avec les consommations électriques dans le secteur résidentiel

Eric Ferreri

► **To cite this version:**

Eric Ferreri. Modélisation et simulation des comportements en lien avec les consommations électriques dans le secteur résidentiel. Psychologie et comportements. Université de Bordeaux, 2017. Français. NNT: . tel-03151186

HAL Id: tel-03151186

<https://hal.science/tel-03151186>

Submitted on 24 Feb 2021

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

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

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
**DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE SCIENCES PHYSIQUES ET DE L'INGENIEUR
SPÉCIALITÉ COGNITIVE

Par Eric FERRERI

**MODELISATION ET SIMULATION DES COMPORTEMENTS
EN LIEN AVEC LES CONSOMMATIONS ELECTRIQUES
DANS LE SECTEUR RESIDENTIEL**

Sous la direction de : Jean-Marc SALOTTI
(Co-directeur : Pierre-Alexandre FAVIER)

Soutenue le 5 septembre 2017

Membres du jury :

M. CARDOT, Hubert	Université de Tours,	Rapporteur
M. SALEMBIER, Pascal	Université de Technologie de Troyes,	Rapporteur
M. DEHAIS, Frédéric	ISAE-Supaéro,	Examineur
M. SALOTTI, Jean-Marc	Ecole Nationale Supérieure de Cognitique,	Directeur
M. FAVIER, Pierre-Alexandre	Ecole Nationale Supérieure de Cognitique,	Co-Directeur

RESUME

Cette thèse s'inscrit dans le domaine de la modélisation des comportements humains associés à la consommation électrique dans l'habitat. De nombreuses études ont montré que le comportement humain joue un rôle clé. La plupart du temps la prise en compte du facteur humain se limite à établir un comportement moyen qui est répliqué afin de simuler une population. Il existe des méthodes probabilistes ou stochastiques qui ne tiennent pas compte des facteurs humains.

Nous présentons dans cette thèse une méthode hybride inspirée du modèle psychologique de Fiske et Neuberg, capable de tirer parti de l'information issue de bases de données. Grâce à l'exploitation de « cartes de chaleur » permettant d'appréhender plus facilement les comportements et d'un modèle agentique, les comportements habituels ou planifiés sont mieux pris en compte. De cette manière tout en affichant une dépendance moins forte aux bases de données que les méthodes stochastiques, nous diminuons fortement le besoin d'établir des règles comportementales et des scénarii ad hoc.

Afin de mesurer l'efficacité de notre méthode nous proposons également deux métriques permettant d'estimer la ressemblance entre un profil généré de consommation et un profil issu d'une base de données.

Mots-clés : consommation électrique, comportement, habitudes

ABSTRACT

We address the problem of simulating domestic activities related to electrical consumption. Continuously growing electrical energy requirements have become an issue as resources to produce it are limited. Human factors have to be taken into account in order to better predict electric consumption and optimize its production. We propose a framework based on a psychological scheme and on stochastic features. The decision process is based on the representation of habits and the use of heat maps. Inclusion of planning-related behavior is performed by a non-parametric voting method using relational properties among domestic activities. The results show a good fitting between the predictions and ground truth.

Keywords : electric consumption, behavior, habits.

REMERCIEMENTS

Je tiens d'abord à remercier le directeur de cette thèse, le Professeur Jean-Marc SALOTTI , pour m'avoir guidé tout au long de ma thèse. En plus de m'avoir beaucoup enseigné sur la manière de conduire une activité de recherche, il m'a accordé de nombreux rendez-vous, a su me prodiguer des conseils avisés et m'a permis de m'acquitter de certaines responsabilités pédagogiques et scientifiques. Je lui dois donc une grande part du progrès que j'ai effectué dans le domaine scientifique. A titre personnel je dois aussi dire que la patience dont il a fait preuve à mon égard m'a touché, je ne sais comment lui exprimer ma gratitude si ce n'est en faisant de même à l'égard des stagiaires et étudiants qui se retrouveraient dans ma situation. Je tiens aussi à remercier Mme LESPINET-NAJIB et Mme CLERMONT qui m'ont donné l'occasion de prendre en charge certaines activités pédagogiques que j'ai beaucoup apprécié et qui ont contribué à ma formation.

Je remercie tous ceux sans qui cette thèse ne serait pas ce qu'elle est , par leurs suggestions ou contributions. Je pense ici en particulier à M. Pascal SALEMBIER ainsi qu'à M. Hubert CARDOT qui m'ont fait l'honneur de participer au jury de cette thèse . Je remercie également l'équipe de cognitive de l'Ecole Nationale Supérieure de Cognitive, leur compagnie et les conversations intéressantes que nous avons eues ont su me motiver pendant les années de thèse.

Je souhaite faire mention de ma Famille et amis, Cristian CHIANALE un ami de longue date qui a su m'encourager à poursuivre mes études dans des moments difficiles. Pour finir, je souligne l'inestimable valeur du soutien de ma Famille et sa présence tout au long de ce travail, qui mérite un respect particulier, Merci.

TABLE DES MATIERES

RESUME	3
ABSTRACT	5
REMERCIEMENTS	7
TABLE DES MATIERES	9
TABLE DES ILLUSTRATIONS	13
TABLE DES TABLEAUX.....	15
CHAPITRE 1 : INTRODUCTION	17
1.1 Problématique.....	18
1.2 Contribution	19
1.3 Description des sous-modèles	21
1.3.1 Le sous-modèle de représentation des habitudes.....	21
1.3.2 Le sous-modèle d'interactions entre activités	22
1.3.3 Le sous-modèle de prise de décision	22
1.3.4 Suggestion de protocoles expérimentaux	22
1.3.5 Plan de lecture.....	23
CHAPITRE 2 : CONTEXTE DE RECHERCHE	25
2.1 Modélisation.....	27
2.2 Présentation des différentes familles de modèles	28
2.2.1 Modèles non cognitifs	28
2.2.1.1 Chaines de Markov	29
2.2.1.2 Réseaux Bayésiens	31
2.2.1.3 Réseaux de neurones	33
2.2.2 Modèles cognitifs	34
CHAPITRE 3 : MODELISATION COMPORTEMENTALE.....	37
3.1 Introduction.....	38
3.2 Bases biologiques et psychologiques du comportement	38
3.2.1 Principes fondateurs	38
3.2.2 Ganglions de la base et conditionnement opérant	40
3.2.3 Apprentissage par renforcement : présentation et bases opérationnelles.....	42
3.3 Modèles psychologiques du comportement	45

3.3.1 Objectif	45
3.3.2 Approches centrées sur l'individu : modèles issus de la psychologie cognitive et sociale.....	45
3.3.2.1 Retour sur les modèles multi-agent : modèle BDI	45
3.3.2.2 Autres méthodes.....	46
3.3.2.3 Alternative envisagée.....	46
3.4 Implémentation de la première version du modèle	47
3.4.1 Obtention des données.....	47
3.4.2 Présentation du cadre bayésien.....	47
3.4.2.1 Principe général.....	47
3.4.2.2 Formalisme bayésien.....	49
3.5 Apport de la psychologie cognitive au modèle bayésien de prise de décision.....	50
3.5.1 Classification des modèles	50
3.5.2 Modèles à deux processus	51
3.5.2.1 Principe des deux processus	51
3.5.2.2 Modèles sélectifs.....	51
3.5.2.3 Modèles compétitifs	52
3.5.2.4 Modèles consolidatifs	52
3.5.2.5 Modèles correctifs.....	53
3.5.3 Implémentation du modèle de Fiske et Neuberg	54
3.5.3.1 Quels comportements considérer ?.....	54
3.5.3.2 Au-delà du paradigme des deux processus	55
3.5.3.3 Catégorisation initiale	56
3.5.3.3 Catégorisation confirmatoire	58
3.5.3.4 Recatégorisation	59
3.6 Résultats et conclusions obtenus à partir de la méthode proposée	61
CHAPITRE 4 : MODELE APPROFONDI	66
4.1 Présentation générale de la méthode.....	67
4.1.1 Description de l'approche	67
4.1.2 Représentation par une suite de delta de Dirac	67
4.1.3 Décomposition en ondelettes de Haar	68
4.1.4 Première étape.....	71

4.1.5 Deuxième étape	72
4.1.6 Cartes de chaleur.....	72
4.2 Exploitation des cartes de chaleur	75
4.2.1 Localisation de caractéristiques des activités sur les cartes de chaleur	75
4.2.2 Liens entre les cartes de chaleur et le modèle de Fiske et Neuberg	75
4.2.3 Interprétation des caractéristiques d'activités	76
4.2.4 Extraction des paramètres comportementaux	79
4.2.4.1 Principe général.....	79
4.2.4.2 Segmentation	80
4.2.4.3 Mise en relation des régions segmentées de plusieurs cartes de chaleur	81
4.3 Simulation.....	83
4.3.1 Mise en œuvre	83
4.3.2 Résultats et étude comparative	85
CONCLUSION GENERALE	88
REFERENCES	92
VALORISATION SCIENTIFIQUE	102

TABLE DES ILLUSTRATIONS

Figure 1.1: Architecture du modèle : le sous-modèle de représentations des habitudes est le pivot entre la perception des besoins et le choix d'une action adaptée.....	p21
Figure 2.1: Chaîne de Markov à 2 états.....	p30
Figure 2.2 : Courbe représentant l'heure et la durée d'utilisation de la cuisine.....	p31
Figure 2.3 : Schéma du réseau bayésien proposé par Hawarah 2010.....	p33
Figure 2.4 : Réseau bayésien prenant en compte l'interaction entre utilisation de divers appareils.....	p34
Figure 2.4 : Réseau bayésien prenant en compte l'interaction entre utilisation de divers appareils.....	p34
Figure 3.1 : localisation anatomique des noyaux gris centraux (aussi connu comme ganglions de la base).....	p40
Figure 3.2 : illustration du schéma de PDM correspondant à l'exemple présenté dans le texte.....	p43
Figure 3.3 : les habitudes servent de pivot entre perception et action.....	p48
Figure 3.4 : Schéma général du modèle sélectif.....	p51
Figure 3.5 : Schéma général d'un modèle compétitif.....	p52
Figure 3.6 : Schéma général du modèle consolidatif.....	p53
Figure 3.6 : Schéma général du modèle correctif.....	p54
Figure 3.7 : Diagramme de flux du modèle original de formation des connaissances et de décision proposé par Fiske et Neuberg (1990).....	p56
Figure 3.8 courbes de Densité de Probabilité décrivant la probabilité d'utiliser un appareil en lien avec les besoins physiques et de récréation	p57
Figure 3.8 : schéma d'agent bio-inspiré basé sur le modèle de Fiske et Neuberg.....	p61
Figure 3.7 : exemplaire de courbe de consommation utilisé dans la génération de nouvelles courbes.....	p63
Figure 3.8 : comparaison d'une courbe d'utilisation générée à une courbe présente dans la base de données.....	p65

Figure 4.1: Décomposition d'une série chronologique de consommations avec ses composantes.....	p68
Figure 4.2 : décomposition en ondelettes de la suite chronologique du peigne de Dirac....	p71
Figure 4.3: décomposition en ondelettes Haar.....	p72
Figure 4.4 : Exemple de représentation temps vs période de chaque utilisation.....	p73
Figure 4.6 : Cas 1 activité résultant d'une habitude régulière.....	p78
Figure 4.7: Cas 2, activité dont la période de répétition dépend de l'heure de réalisation...	p80
Figure 4.8 : Cas 3, représentation d'un comportement plus complexe.....	p82
Figure 4.9 : Séparation en régions chaudes et tièdes de la carte de chaleur initiale.....	p83
Figure 4.10 : Cartes de chaleur représentant différentes étapes du traitement.....	p84
Figure 4.11: Matrice de pondération du WPSNR.....	p86
Figure 4.12: Résultats de la simulation pour l'usage de l'ordinateur.....	p87

TABLE DES TABLEAUX

Tableau 2.1 : Probabilités de transitions.....	p30
Tableau 2.2 : Distribution de probabilités décrivant les durées d'utilisation de la télévision.....	p33
Tableau 2.3 : Probabilités horaires d'utilisation de la télévision.....	p33
Tableau 3.1 : Classification des appareils électroménagers.....	p57
Tableau 3.2 : récapitulatif des distances entre profils de consommation.....	p62
Figure 4.5 : Carte de chaleur représentant l'utilisation du lave-vaisselle.....	p75
Tableau 4.1 : évaluation des performances.....	p88

CHAPITRE 1 : INTRODUCTION

1.1 Problématique

Les consommations électriques du secteur domestique sont particulièrement complexes à modéliser. En effet, l'équipement dont dispose chaque foyer est très variable et l'influence du facteur humain reste quantitativement mal connue. Dans ce travail nous nous sommes intéressés aux moyens mis en œuvre pour modéliser et simuler l'effet du facteur humain sur les consommations.

Si on vous demande de décrire les activités que vous réalisez habituellement lorsque vous êtes chez vous, vous allez effectuer cette tâche très facilement. En effet, les activités présentent des caractéristiques de régularité qui vous permettent de rapidement décrire, en moyenne, vos activités domestiques au quotidien. Mais que se passe-t-il si on demande de réaliser la même tâche à votre voisin, et qu'on compare vos habitudes ? On pourrait s'attendre à de légères variations dans les habitudes, puisque les besoins de base, le climat ainsi que le quartier restent inchangés. Pourtant ce n'est pas le cas, même lorsque deux foyers sont dans le même quartier et que leur équipement et le nombre de personnes sont identiques, leurs habitudes peuvent présenter de grandes différences.

En plus de cette variation inter-foyers importante, on s'aperçoit qu'à l'intérieur d'un même foyer, les habitudes sont parfois amenées à évoluer ou à fluctuer en fonction de facteurs difficilement prévisibles. Il peut s'agir par exemple du suivi d'événements sportifs, d'une grève, d'une maladie, de circonstances climatiques particulières, ou de la panne d'un équipement. Si on ne tient pas compte de ces facteurs particuliers, il existe toujours des variations dans les activités qui sont dues notamment à des systèmes de valeurs différents ainsi qu'à des représentations différentes de l'environnement et des tâches domestiques. Ces différences soulèvent une première interrogation :

Quels sont les mécanismes qui permettent d'expliquer une telle variation dans les habitudes, pourtant censées être une réponse stable dans le temps ?

Sous l'influence historique du béhaviorisme, la formulation scientifique de ce qu'est une habitude a été réduite à une probabilité d'apparition d'une réponse motrice lors de la présentation d'un stimulus (Watson 2013). Dans le domaine plus récent des neurosciences computationnelles, les habitudes sont conçues comme un comportement réactif issu d'une association probabiliste entre une catégorie de stimuli et des décisions. Cette association est construite grâce à ce que nous percevons dans l'environnement, ainsi que de ce que nous percevons comme nos besoins. Lorsque nous ressentons la nécessité de réaliser une action habituelle, par exemple laver nos vêtements, nous n'avons pas tous la même perception de ce qu'est une utilisation adéquate du lave-linge (attendre d'avoir une charge pleine vs

lavages fréquents afin de maintenir un maximum de linge propre). Cette constatation soulève une seconde interrogation :

Comment prendre en compte les différences liées aux facteurs humains et propres à chaque foyer afin de mieux comprendre et prédire l'utilisation d'appareils électriques ?

Dans les définitions précédemment énoncées de ce que sont les habitudes, nous avons fait allusion à un lien probabiliste entre environnement et action. Donc, par définition, une habitude correspond à la réponse la plus probable dans une circonstance donnée. A partir de ce constat, plusieurs méthodes stochastiques ont été proposées. Parmi les plus simples figure une méthode basée sur l'utilisation de chaînes de Markov. Malgré la simplicité de la méthode, il devient possible d'obtenir un comportement moyen et des écarts de comportements semblables à ceux mesurés sur une population (Widen 2009). Par la suite d'autres méthodes statistiques permettant d'améliorer la précision des consommations ont vu le jour. Plus la base de données des consommations électriques est représentative de la population et plus la simulation stochastique est précise. Malheureusement, au fil du temps s'accumulent diverses variations dans les habitudes et l'équipement des habitants. Cet écart entre les données historiques de consommation électrique et les données de consommation actuelle conduit à l'obsolescence progressive des modèles stochastiques et l'erreur de prédiction reste difficile à mesurer.

1.2 Contribution

Afin d'apporter davantage d'éléments de réponses aux deux questions précédentes, nous avons modélisé un système qui propose de quantifier l'influence du comportement humain en prenant en compte les particularités de celui-ci. Ce système a pour but de simuler, de manière simplifiée, les processus cognitifs qui aboutissent à la réalisation d'une action consommatrice d'électricité (par exemple mise en marche de la machine à laver).

Nous avons d'abord défini sa structure globale en prenant en compte des observations et des modèles existants dans la littérature. Cela nous a permis de définir un ensemble d'hypothèses et de contraintes fondatrices du modèle. Comme nous pouvons le voir figure 1.1, le modèle proposé est formé de trois sous-parties : le sous-modèle de représentation des habitudes, le sous-modèle d'interaction des habitudes, et celui des décisions.

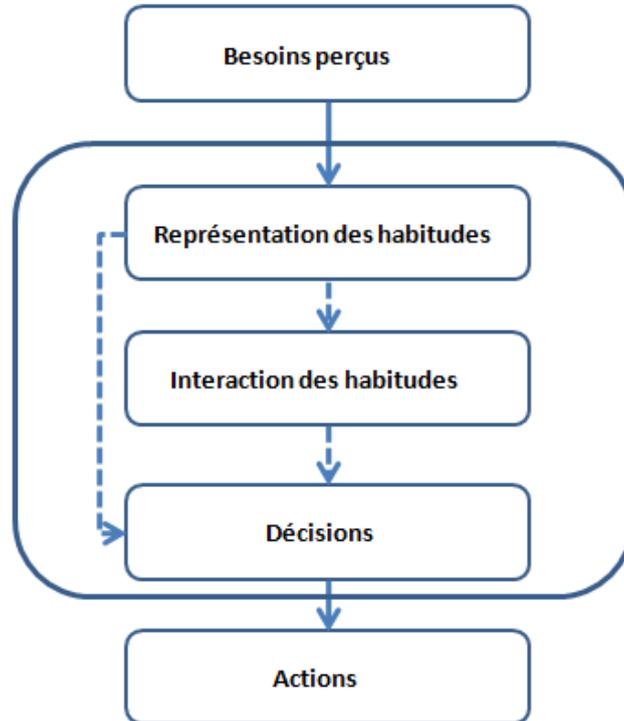


Figure 1.1 - Architecture du modèle : le sous-modèle de représentation des habitudes est le pivot entre la perception des besoins et le choix d'une action adaptée.

Ce modèle cognitif est basé sur 3 prémisses :

- 1- Au sein des processus de décision, l'analyse globale d'une situation a la priorité sur l'analyse basée sur des caractéristiques. Ainsi, nous postulons que dans un premier temps les besoins perçus par les habitants sont catégoriels. Cela revient d'abord à savoir si le besoin à satisfaire est orienté au divertissement ou à un besoin physique. Ce n'est qu'ensuite qu'une action adéquate est choisie, basée sur une caractéristique telle que l'heure.
- 2- Plusieurs stratégies sont possibles lors du choix d'une action. Une action peut être notamment guidée par une habitude ou par un but planifié.
- 3- Le choix de la stratégie orientée vers un but dépend des ressources attentionnelles disponibles.

Nous avons utilisé un formalisme probabiliste pour décrire la formation et l'exploitation des habitudes. L'apprentissage par renforcement, souvent utilisé en neurosciences et en intelligence artificielle pour décrire les processus de décision (Doya 2000), est également proposé ici pour expliquer la formation des habitudes. Le schéma général des modèles

d'apprentissage par renforcement est très proche de ceux qui exploitent les chaînes de Markov.

Une fois que les paramètres du modèle probabiliste représentant les habitudes sont connus, il reste à implémenter la stratégie orientée vers un but. En effet, certaines actions sont décidées à l'avance et requièrent un raisonnement et une planification. Nous choisissons de décrire la planification comme une interaction entre les choix de réalisation de diverses activités. Par exemple le repassage du linge n'est pas isolé du lavage de celui-ci, il n'est pas possible de repasser du linge sale.

A partir de bases de données, il est possible d'avoir non seulement une suite d'activités type, mais également de déduire la manière dont se construisent les habitudes ainsi que les interactions entre activités.

Connaissant ces données, et étant capable de simuler dans une certaine mesure l'évolution des habitudes et du matériel, le modèle est capable de répondre à plusieurs questions concernant les consommations induites par la réalisation d'activités domestiques.

Dans la suite de ce chapitre, nous allons décrire chacune des parties du modèle cognitif proposé et une des utilisations possibles de ce modèle : la simulation de changements d'habitudes au sein d'une population.

1.3 Description des sous-modèles

Comme nous l'avons vu précédemment, notre modèle est composé de 3 parties : le sous-modèle de représentation des habitudes, le sous-modèle d'interaction des activités, et celui des décisions. Dans la suite de ce chapitre, nous allons présenter la fonction de chacun de ces trois sous-modèles.

1.3.1 Le sous-modèle de représentation des habitudes

La représentation interne des habitudes joue un rôle de pivot entre la perception des besoins de base et la réalisation d'une activité lorsque celle-ci est automatique. Elle intervient donc logiquement dans ces deux processus. La représentation choisie doit alors permettre aisément le transfert d'informations entre ces deux opérations. Dans le modèle cognitif, nous avons supposé que chaque habitude est représentée par deux mesures particulières : l'intervalle modal entre répétitions d'une même action et l'heure plus courante à laquelle elle est réalisée.

1.3.2 Le sous-modèle d'interactions entre activités

Les différents besoins interagissent entre eux, et de ce fait requièrent une planification. Cette représentation intervient donc logiquement dans les processus de perception et de réalisation d'activités. La représentation choisie doit alors permettre aisément la visualisation et le transfert d'informations entre ces deux opérations. Dans le modèle cognitif, nous avons supposé que chaque habitude est représentée par le moyen de deux mesures particulières : l'intervalle modal entre répétitions d'une même action et l'heure plus courante à laquelle elle est réalisée.

1.3.3 Le sous-modèle de prise de décision

Le sous-modèle de prise de décision permet de déterminer si le type d'action requis peut être choisi automatiquement en respectant les habitudes, ou bien s'il requiert la prise en compte de plusieurs activités se succédant dans le temps. Il permet donc un arbitrage entre réaction et planification lors du choix de l'action. Ce sous-modèle permet aussi de répondre à la question inverse, c'est-à-dire lorsque l'ensemble des activités d'une habitation est connu, quelles décisions ont été prises par habitude ? Pour répondre à cette question, ce sous-modèle exploite les informations issues du sous-modèle de représentation des habitudes ainsi que celui d'interactions entre activités.

1.3.4 Suggestion de protocoles expérimentaux

Le système est bâti sur une structure bayésienne lui permettant de résoudre plusieurs tâches : la simulation d'activités domestiques lorsque l'on ne considère que la part d'activité liée aux habitudes, la prédiction de consommations électriques, l'influence de prises de décisions basées plus sur la planification que sur les habitudes, *etc.* Nous pouvons donc simuler le modèle proposé sous plusieurs conditions, ce qui nous permet de réaliser des prédictions expérimentales. Cependant, les données expérimentales n'étant pas toujours existantes, nous avons dû utiliser des bases de données issues d'habitations sur une résolution variable dans le temps. Dans certains cas les données étaient moyennées à l'année et dans d'autres cas elles avaient une résolution temporelle de 10 minutes. De plus, nous avons suggéré de nouvelles voies de recherche de terrain.

Il est important de noter que nous considérons ces suggestions comme des amorces de protocoles. En effet la rédaction d'un protocole expérimental est très délicate. Il faut que sa description soit suffisamment claire pour que l'expérience puisse être reproduite à l'identique et il doit faire l'objet d'une analyse critique pour notamment détecter

d'éventuels biais. Or cela dépasse le cadre de la présente étude. C'est pourquoi nous fournissons uniquement des pistes de protocole.

1.3.5 Plan de lecture

Chapitre 2 : le chapitre 2 présente l'état de l'art concernant les diverses méthodes, non-cognitives et cognitives permettant de simuler des activités domestiques consommant de l'électricité.

Chapitre 3 : le chapitre 3 introduit le formalisme utilisé, basé sur un système multi agent bio inspiré. Les bases biologiques, psychologiques et probabilistes sont expliquées puis leur implémentation informatique est décrite. Une première version du modèle est présentée et les premiers résultats et conclusions sont montrés.

Chapitre 4 : le chapitre 4 montre une version avancée du modèle, propose et exploite une nouvelle représentation temps-période des données comportementales permettant d'améliorer clairement la puissance explicative et prédictive du modèle proposé.

CHAPITRE 2 : CONTEXTE DE RECHERCHE

Les méthodes numériques et outils informatiques permettant de simuler des consommations électriques ont atteint un point critique dans leur développement. Lorsque l'on s'intéresse à la modélisation au niveau d'un quartier, le nombre de données et de traitements à effectuer sur celles-ci atteint très rapidement un grand volume. Par exemple pour un ensemble d'une centaine d'habitations où la consommation des principaux appareils électroménagers est mesurée à intervalle régulier (en principe toutes les 10 minutes), il est nécessaire de traiter plusieurs dizaines de milliers de données. Le traitement des données implique non seulement de les « débruiter », mais surtout de leur attribuer un sens dans le cadre des comportements humains. En effet, détecter la mise en route d'un lave-vaisselle n'est pas en soi une information importante, ce qui devient intéressant c'est de savoir dans quelle mesure ce comportement est habituel, et s'il y a un lien entre ce comportement et d'autres, aussi bien au sein de l'habitation que du quartier. La mise en place et l'exploitation d'outils informatisés sont actuellement possibles, en effet la modélisation des appareils électriques requiert la génération d'une masse de données importantes. Les capacités de calcul des ordinateurs modernes permettent d'estimer rapidement les futures dépenses énergétiques d'un ensemble de bâtiments.

Dans le domaine de la prédiction de consommation énergétique liée au chauffage, on exploite des modèles thermiques sophistiqués décrivant précisément l'évolution des températures au sein d'un bâtiment. Cependant, la bonne précision des estimations concernant la thermique des bâtiments ne conduit pas à des simulations exactes concernant les consommations électriques, essentiellement à cause d'incertitudes liées aux facteurs humains. En effet, les variables liées au comportement humain sont devenues la principale source d'inexactitudes dans les résultats lors de simulations de consommations énergétiques. L'influence des habitudes et usages est particulièrement évidente à l'échelle du quartier. A ce niveau, la demande d'énergie électrique ainsi que sa variance est importante, en effet, pour une population plus nombreuse, à l'échelle d'une ville par exemple, les fluctuations dans la demande sont mieux connues, en conformité avec ce qu'établit le théorème centrale limite. En effet, pour un échantillon suffisamment important, la distribution globale des consommations suit une loi Gaussienne.

Pour ce qui concerne la consommation électrique, jusqu'à présent, plusieurs approches ont été utilisées pour prendre en compte le rôle des habitants. Parmi les solutions retenues, les plus simples définissent un comportement « moyen ». Par exemple, si dans un quartier certains habitants utilisent le four électrique deux fois par semaine alors que d'autres l'utilisent quatre fois par semaine, à des fins de simulation on considérera que le quartier est composé de répliques d'un habitant type qui se servira du four trois fois par semaine. Cette approche bien que produisant peu d'erreurs pour une population de plusieurs centaines de personnes sur une période de plusieurs mois, s'avère trop limitée et biaisée à l'échelle d'un quartier. Une inspection attentive des données de consommation révèle que les pics sont plutôt dus à la variabilité dans les habitudes qu'aux habitudes elles-mêmes. Une moyenne

est insuffisante car elle ne parvient pas à résumer la variabilité intra et inter habitations qui est à l'origine de variations brusques dans la demande énergétique.

Les modèles stochastiques, qui permettent de définir adéquatement des systèmes dynamiques complexes, semblent donc bien plus adaptés que les modèles déterministes précédemment cités. L'utilisation de modèles probabilistes représente un réel progrès, cependant il est difficile de résumer un comportement à des distributions de probabilité. Parmi les difficultés rencontrées figure l'interprétation des paramètres du modèle qui semble déconnectée des mécanismes comportementaux. Afin d'avoir une approche plus naturelle, récemment des modèles basés sur des systèmes multi agents ont vu le jour. Ce deuxième chapitre donnera un aperçu de l'essentiel des techniques existantes concernant la simulation de comportements en lien avec les consommations électriques.

2.1 Modélisation

Méthodologie générale

La consommation électrique d'une habitation dépend d'un grand nombre de variables, la plupart de celles-ci sont liées à des comportements et des usages. Certains comportements sont réguliers et prévisibles, par exemple la fréquence d'utilisation de la machine-à-laver est stable. Mais pour d'autres activités il est très difficile de trouver à priori une régularité quelconque, par exemple l'utilisation de l'ordinateur semble difficilement prévisible. Afin de proposer une représentation simplifiée permettant de prédire et simuler adéquatement les consommations électriques domestiques. Pour ce faire nous chercherons à définir un système qui prend en entrée un ensemble d'équipements électroménagers et électriques ainsi qu'un nombre d'habitants et qui génère en sortie une variété de profils de consommations possibles. La complexité et la variabilité du comportement humain, rend impossible de déterminer par un calcul direct l'évolution temporelle des consommations électriques.

D'autres phénomènes sont difficiles à prévoir de manière déterministe, l'intervalle de temps entre deux appels téléphoniques, la valeur d'actions cotées en bourse ou encore la quantité d'énergie électrique demandée par un domicile ou un quartier sont des exemples de variables stochastiques. Il est intéressant de déterminer les lois de probabilités sous-jacentes à ces processus car grâce aux lois obtenues il est possible de décrire l'évolution dans le temps d'une variable et même de prédire une valeur approximative de celle-ci à un instant donné. Afin d'induire (on utilise des faits pour remonter aux lois) des lois de probabilités, il est nécessaire de disposer d'une base de données qui prend souvent la forme d'une série temporelle. Une fois qu'une loi de probabilités a été estimée, celle-ci doit être validée à partir d'autres données du même type.

Une fois la distribution de probabilités décrivant les consommations électriques pour un groupe d'habitations connu, il devient possible d'estimer un volume de consommations prévues pour un instant donné. Différentes familles de modèles permettent de décrire une grande variété de processus stochastiques. Le choix dépend de certaines caractéristiques de la variable d'intérêt. Dans le présent travail, en plus de modèles stochastiques, nous avons surtout exploité des modèles comportementaux issus de la psychologie. Dans ce chapitre, nous allons décrire les modèles couramment utilisés. Nous commencerons par les modèles stochastiques qui sont les plus répandus, puis nous aborderons l'utilisation de modèles psychologiques. Pour conclure, nous expliciterons le lien entre ces différentes approches.

2.2 Présentation des différentes familles de modèles

2.2.1 Modèles non cognitifs

Pour modéliser les processus de décision auxquels nous nous intéressons, c'est à dire ceux qui concernent l'utilisation d'appareils électroménagers, diverses approches sont possibles. Prenons l'exemple de la machine à laver, comment déterminer l'instant de son démarrage ? Une première approche consiste à s'intéresser exclusivement aux observations, dans ce cas on analysera les données afin de connaître l'heure de mise en route la plus fréquente. Cette analyse de fréquences revient à dire que décision et comportement sont étroitement liés sans que l'on connaisse la nature précise de ce lien, on parle d'approches de type « boîte noire ». Ces méthodes cherchent à reproduire aussi fidèlement que possible les propriétés statistiques concernant les données comportementales mesurées, Pour des entrées données (contexte et appareils disponibles), la sortie (décision d'action associée au fonctionnement d'un appareil) doit être aussi proche que possible de la sortie espérée au sens statistique, le mécanisme cognitif sous-jacent étant totalement ignoré.

Généralement cette approche est employée lorsqu'il n'est pas possible ou trop complexe de connaître un lien entre les données observées et d'autres variables propres au système modélisé. La collectivité d'utilisateurs résidentiels est souvent vue comme une « boîte noire » par les fournisseurs d'énergie. Afin de montrer un aperçu de chaque méthode appartenant à l'approche non cognitive ou boîte-noire, nous proposons de les appliquer sur des exemples.

Nous utilisons pour cela des données simulées portant sur une habitation très simple, composée d'un seul habitant qui réalisera deux activités : cuisiner ou se reposer.

L'habitant commencera à cuisiner tous les jours vers 9 heures (Heure initiale : $H_i=9$) et réalisera cette activité pendant une période comprise entre 15 minutes et une heure ($15mn < \Delta T_c < 60mn$). Lorsque l'habitant ne cuisine pas, il se repose. Pour les heures de début de cuisine, les données ont été générées d'après une distribution normale $H_i \sim N(\mu =$

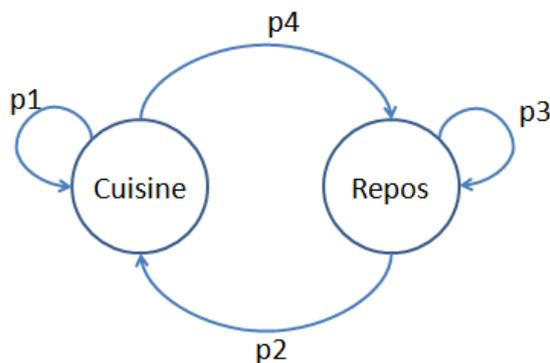
9 heures, $\sigma = 1$ heures), la durée des temps de cuisine suit une distribution uniforme : $\Delta T_c \sim U(15mn, 60mn)$.

2.2.1.1 Chaines de Markov

Parmi les modèles les plus couramment utilisés figurent les chaînes de Markov (Widen 2009 b). Dans le cadre de processus Markoviens, les activités domestiques sont vues comme un système en temps discret, ayant N états possibles notés S_1, S_2, \dots, S_n , où chaque état représente une activité. A un instant t , le système ne peut être que dans un état. De plus, nous ne nous intéressons qu'aux modèles markoviens du premier ordre, c'est-à-dire où l'on considère que les probabilités de transition entre états à l'instant t dépendent uniquement de l'état observé à l'instant $t-1$ (voir équation 1). Le but principal est donc de connaître les transitions possibles entre états ainsi que les probabilités associées à ces transitions (voir figure 2.1).

Equation 1

$$P(q_t = S_i) = \alpha(q_{t-1})$$



Valable pour l'heure H1	C, T+1	R, T+1
C, T	p1	p4
R, T	p2	p3

Figure 2.1: Chaîne de Markov à 2 états

Tableau 2.1 : Probabilités de transitions

On suppose que la probabilité de transition entre états est constante sur une période d'une heure, la propriété d'homogénéité temporelle (voir équation 2) est valable sur une heure.

Equation 2

$$P(q_t = S_i | q_{t-1} = S_j) = P(q_{t+1} = S_i | q_t = S_j) \quad \forall i, j \in \{1, 2, \dots, N\}$$

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} = P\{X_{n+1} = j \mid X_n = i\} = P_{i,j}$$

Equation 3

$$P(q_t = S_i | q_{t-1} = S_j, q_{t-2} = S_k, \dots)$$

Pour obtenir les probabilités de transition, les auteurs (Widén 2010) utilisent une méthode classique qui repose sur un registre des activités domestiques. L'utilisation de chaînes de Markov telles que décrites, vise à modéliser chaque habitant, en utilisant une chaîne de Markov par heure. Une journée type est donc décrite grâce à 24 chaînes de Markov par habitant. Pour illustrer leur méthode, nous l'appliquons aux données synthétiques décrites dans l'introduction. Dans un premier temps l'habitant note chaque fois l'heure à laquelle il cuisine, ensuite ces données sont mises dans un format matriciel simplifiant le calcul des probabilités de transitions entre les états observés (S_1 =cuisiner, S_2 =se reposer).

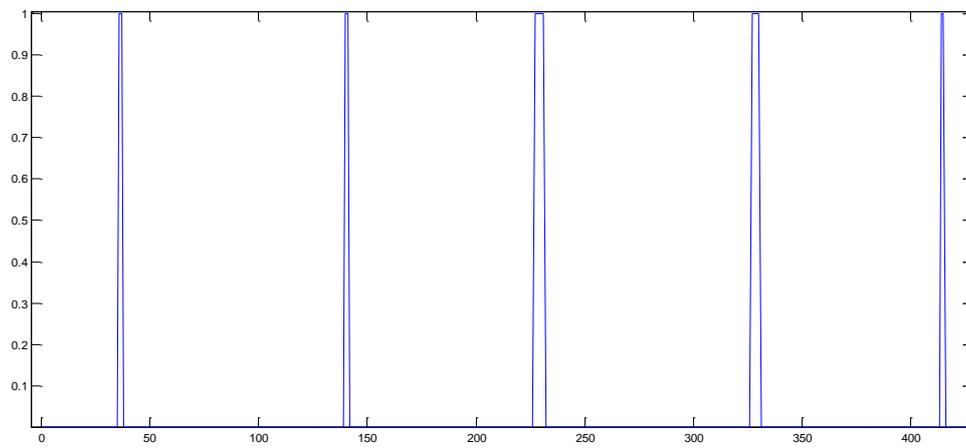


Figure 2.2 : Courbe montrant les données concernant l'utilisation de la cuisine. Sur l'axe vertical '1' indique que la cuisine est utilisée, '0' qu'il s'agit d'une période de repos. Sur l'axe horizontal le temps est indiqué en quart d'heures. Ce graphique permet de représenter l'heure et la durée d'utilisation de la cuisine.

Il s'agit de chaînes de Markov en temps discret, avec $\Delta t=15$ minutes et valable sur une période d'une heure. Les probabilités de transitions sont déterminées à partir des données de consommation rapportées par les habitants eux-mêmes. Pour chaque Δt de la journée, les transitions entre états sont comptées puis divisées par le nombre total de transitions. Les auteurs proposent d'utiliser une période plus large lorsque le nombre de transitions observées en une heure est proche de zéro.

$$P(q_t = S_i | q_{t-1} = S_i) \quad (3)$$

Le comportement observé a un lien direct avec la période considérée (matin/soir) et l'activité réalisée à l'instant $t-1$. Ce modèle compte parmi les plus simples (très peu de paramètres, simplicité des calculs, quasi absence de base comportementale) et les plus

efficaces (faible taux d'erreur lorsque l'échantillon représente bien une importante population). Cependant les performances d'une chaîne de Markov dépendent très étroitement de la précision et l'exhaustivité de la base de données utilisée.

De plus, les probabilités de transitions entre états ne peuvent être ajustées en fonction de l'évolution des usages, ce qui implique que pour maintenir un certain niveau de fiabilité, il faut périodiquement remettre à jour l'intégralité des bases de données et recalculer tout le modèle.

Il est possible d'améliorer encore cette méthode en prenant en compte la durée habituelle de permanence dans un état. En effet une pré-enquête où nous demandions de décrire librement puis en termes d'heures d'utilisation et de fréquence a révélé qu'il est plus naturel et moins contraignant pour les habitants de donner l'heure de début ainsi que la durée approximative d'utilisation d'un appareil, que de décrire qualitativement les usages de chaque appareil électrique. De plus, pour les périodes où peu de transitions entre états surviennent, la durée de permanence dans un état peut être utilisée pour estimer une probabilité de transition. Par exemple si l'on considère que la durée d'utilisation de la cuisinière est indépendante de l'heure, par le calcul il est possible de relier le temps espéré de permanence dans l'état C avec la probabilité p_1 (voir tableau 2.1)

2.2.1.2 Réseaux Bayésiens

D'autres modèles de comportement sont basés sur des réseaux bayésiens (Hawarah 2010). Un réseau bayésien est un modèle graphique orienté qui permet de représenter des liens probabilistes et causaux entre un ensemble de variables. Chaque nœud représente une variable aléatoire, un lien qui relie le nœud A au nœud B signifie que A (nœud causal) a une influence directe sur B (nœud-conséquence). Cette influence peut être quantifiée, un tableau de probabilités conditionnelles permet de résumer l'information apportée par chaque nœud. Lesdites probabilités peuvent provenir d'estimations réalisées par un expert ou être issues de bases de données. La construction adéquate d'un réseau bayésien requiert des informations portant sur la structure du réseau ainsi que des informations portant sur les variables.

Une fois le réseau bayésien défini et les probabilités conditionnelles connues, il est possible de calculer la probabilité d'occurrence d'un événement (nœud terminal) pour certaines valeurs des nœuds causaux. Inversement, il est aussi possible de connaître la cause la plus probable d'un événement. Hawarah et al. (Hawarah 2010) ont réalisé un travail, basé sur un réseau Bayésien visant à prédire le comportement des occupants dans le but d'anticiper les demandes énergétiques.

Leur approche permet d'organiser les tâches domestiques afin d'éviter certains problèmes tels que les pics de surconsommation. Pour illustrer leur méthode, ils prennent un exemple portant sur l'utilisation du four électrique. La probabilité d'utiliser le four électrique dépend

de l'heure, du jour (week-end ou semaine) et du mois. En plus d'estimer à tout instant la probabilité d'utiliser le four électrique, leur réseau permet de prédire la durée d'utilisation et l'énergie consommée lorsqu'une mise en marche du four est détectée (voir figure 2.3).

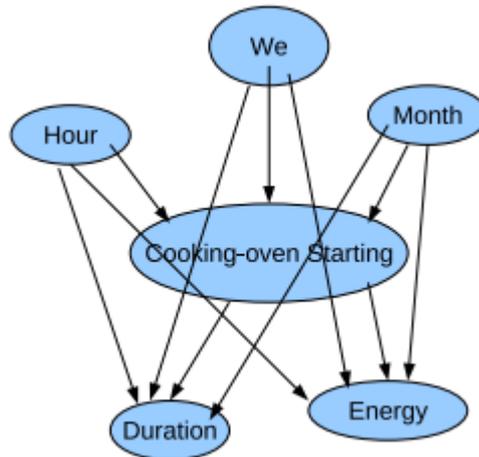


Figure 2.3 : Schémas du réseau bayésien proposé par Hawarah 2010.

L'utilisation d'un réseau bayésien permet donc une modélisation plus fine que les chaînes de Markov. En effet, dans ce cas, le lien causal entre facteurs environnementaux et comportement est bien plus riche. Cette méthode requiert un apprentissage préalable afin d'associer aux probabilités conjointes des divers facteurs environnementaux, la probabilité d'observer un comportement donné. Pour ce faire, les auteurs font une estimation de la distribution de probabilités conditionnelles pour chaque nœud du réseau puis les utilisent comme des 'a priori' pour retrouver des fréquences relatives de comportements observés. Ce qui leur permet de dire –par exemple- qu'au mois de décembre, le four est utilisé au dîner, en particulier pendant les week-ends. Dans le formalisme bayésien, cela revient à écrire :

$$\text{MaxArg}(P(\text{four}|\text{mois}, \text{période}, \text{heure})) \equiv (\text{four} = \text{oui}, \text{mois} = \text{décembre}, \text{période} = \text{WE}, \text{heure} = \text{après midi})$$

Durée	P(ΔT utilisation)
$\Delta t < 30\text{min}$	0.089
$30\text{min} < \Delta t < 1 : 30\text{h}$...
$\Delta t > 1 : 30\text{h}$	0.1

Tableau 2.2 : Distribution de probabilités décrivant les durées d'utilisation de la télévision.

Heure	P(Télévision=ON)
H=1	0.089
...	...
...	...
H=24	0.1

Tableau 2.3 : Probabilités horaires d'utilisation de la télévision.

La méthode telle qu'elle est présentée et proposée permet d'obtenir de meilleurs résultats que ceux obtenus à partir des chaînes de Markov.

Il est cependant regrettable que parmi les facteurs causaux, ne figurent pas les activités elles-mêmes. Par exemple, l'utilisation du lave-linge peut permettre de prédire efficacement une prochaine utilisation du sèche-linge. Les auteurs ont choisi de rester sur une modélisation d'habitudes portant uniquement sur un seul appareil électroménager plutôt que d'utiliser l'inter corrélation entre usages. Ce choix est toutefois compréhensible : dans le cadre d'un modèle non-cognitif, inclure des informations comportementales complique excessivement les calculs. En effet, inclure des informations portant sur l'inter corrélation aurait signifié augmenter la complexité du modèle de $N!$, où N représente le nombre d'appareils électroménagers.

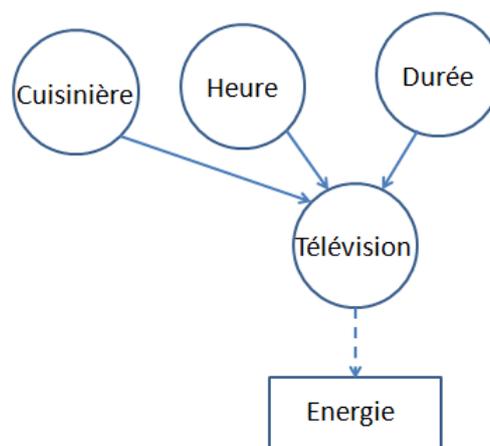


Figure 2.4 : Réseau bayésien prenant en compte l'interaction entre utilisation de divers appareils.

2.2.1.3 Réseaux de neurones

D'après Buratti 2014, un réseau de neurones artificiels permet de se rapprocher des fonctions réalisées par des neurones biologiques. Comme les auteurs l'indiquent, l'avantage de ce type d'approche est que les paramètres du réseau de neurones artificiels ne proviennent pas d'une modélisation et d'une programmation préalable comme c'était le cas des autres méthodes (chaînes de Markov par exemple). En effet, les opérations réalisées proviennent d'un apprentissage autonome réalisé grâce à l'algorithme bien connu de rétropropagation du gradient de l'erreur. Il s'agit cependant d'un fonctionnement de type *boîte noire*, les paramètres du réseau restent difficiles à interpréter en termes comportementaux. L'article cité décrit l'utilisation d'un réseau de neurones artificiels avec N entrées, M couches et 1 sortie afin de modéliser le comportement thermique d'un bâtiment, en prenant en compte les usagers. Les résultats publiés sont encourageants lorsqu'il s'agit

de simulations concernant l'utilisation de chauffage : la décision d'allumer un radiateur est largement expliquée par des facteurs externes, tels que la température extérieure ainsi que le climat. Une autre étude, plus détaillée réalisée par (Aydinalp 2003) propose d'utiliser un réseau de neurones qui prend en compte quatre variables : le type de construction, la fréquence d'utilisation de divers appareils électroménagers, les caractéristiques socio-économiques des habitants et le climat. Il est dommage que les auteurs ne montrent pas les courbes de consommation obtenues et évaluent l'efficacité de leur méthode en comparant la consommation totale d'un quartier simulée pour une année aux mesures directes.

Le fait que les réseaux réalisent un apprentissage automatique permet d'ajouter des variables d'entrée de manière plus aisée que dans le cas des réseaux bayésiens. Il serait intéressant de considérer une approche similaire en prenant en compte les corrélations entre activités comme entrées complémentaires au sein de leur réseau de neurones récurrents. Le réseau de neurones proposé dans la figure 2.5 utilise une information concernant le poste de télévision pour mieux prédire la consommation associée à la cuisinière. Le nombre de poids du réseau passe de 25 à 29, l'augmentation du temps d'apprentissage et des exemples requis est négligeable.

Dans (Aydinalp 2008), une comparaison entre les performances de prédiction d'un réseau de neurones à deux autres approches statistiques est proposée. Cette étude met en évidence l'influence du comportement humain dans les consommations électriques. Le réseau de neurones proposé dans cet article utilise l'inter corrélation entre habitudes d'usages de divers appareils électriques. La couche d'entrée du réseau de neurones est prévue pour recevoir les données correspondant aux usages de divers appareils, et la sortie porte sur un seul appareil. La principale différence avec un réseau bayésien réside dans le fait que les opérations réalisées par les couches cachées ne proviennent pas d'estimations statistiques, mais d'un apprentissage automatique. Cette approche permet des prédictions plus précises que celles des réseaux bayésiens en partie parce que la structure interne d'un réseau de neurones possède plus de degrés de liberté et peut mieux s'ajuster aux données. Bien que cette technique ait de très bonnes capacités de prédiction, les coefficients décrivant les opérations des neurones n'ont pas de sens physique. Il s'agit d'un sérieux défaut car, comme pour les chaînes de Markov, en cas de changement dans les usages, le matériel ou la population, il est impossible de modifier adéquatement les poids des connexions entre neurones sans devoir recalculer tout le modèle en utilisant une nouvelle base de données.

2.2.2 Modèles cognitifs

En cas de modification importante du contexte, avec les modèles non cognitifs il n'y a aucun moyen de prédire l'impact car le modèle n'intègre aucune information comportementale. En réalité, cette méthode comme la plupart des modélisations statistiques résume le comportement humain à une suite d'activités où la probabilité de réaliser une certaine activité à l'instant t , dépend au mieux de l'activité réalisée à l'instant $t-1$. D'autres méthodes parviennent à obtenir de bons résultats en requérant des bases de données moins

exhaustives. Par exemple l'application '*Courbotree*' proposée par Stephan et *al.* (Stephan 2009) se base sur une technique d'arbre de régression multivariée.

Une méthode d'apprentissage automatique permet de regrouper les consommateurs en classes, grâce à un ensemble de variables explicatives (par exemple la proportion d'énergie électrique consommée pendant les heures creuses, le type de logement...). Il devient ainsi possible d'associer une courbe espérée de consommation à un individu inconnu à partir d'un minimum d'information. Mais une fois de plus, il est difficile d'établir un lien causal entre le comportement des usagers et les valeurs des paramètres du modèle.

Une autre famille de modèles que nous appellerons cognitifs par opposition aux modèles précédents, utilise une connaissance des processus mentaux afin d'inférer davantage d'informations à partir des observations. Offrant des solutions aux problèmes cités précédemment, les modélisations multi agents suscitent un grand intérêt. En s'appuyant sur une structure qui cherche à reproduire les processus comportementaux des habitants, les paramètres du modèle deviennent interprétables en termes d'intentions et d'habitudes. D'autre part, il devient possible d'étudier l'influence de certaines habitudes sur la consommation électrique d'un logement. Dans cet article, nous proposons un nouveau schéma d'agent basé sur un modèle cognitif ayant fait ses preuves. Notre démarche se déroule en deux étapes. Une première étape vise à utiliser des données comportementales afin de paramétrer le processus de formation des décisions de l'agent, et une seconde étape cherche à utiliser ce processus de formation des décisions pour générer des comportements et des habitudes plus variés que ceux de la base de paramétrage. Nos contributions portent donc sur le nouveau schéma de l'agent cognitif, et ses capacités à reproduire dans une certaine mesure, un comportement motivé. Dans la suite de cet article, nous présentons dans la section 2, le principe de la modélisation multi agent appliqué au domaine des activités domestiques. Notre méthode de modélisation basée sur un modèle bio-inspiré est exposée dans la section 3. Nous proposons dans la section 4, une validation de notre approche en l'appliquant à la simulation de consommation électrique pour divers appareils électroménager et divers comportements d'usagers. Les conclusions et perspectives de nos travaux sont décrites dans la section 5.

Lorsqu'un processus est bien connu, sa modélisation devient plus précise, nous avons donc choisi d'utiliser, à chaque fois que cela est possible un modèle de type cognitif.

CHAPITRE 3 : MODELISATION COMPORTEMENTALE

3.1 Introduction

En exploitant, les modèles précédemment présentés, les auteurs cherchent à reproduire des comportements dans un contexte donné. Malgré cette grande variété de moyens mis en œuvre, peu se sont intéressés aux bases théoriques du comportement. Dans le présent chapitre nous centrons notre attention sur les bases biologiques et psychologiques du comportement. Plus précisément nous nous intéressons à la formation d'habitudes. Ensuite nous proposons un schéma décrivant l'interaction habitant-environnement, qui montre que modéliser et simuler des comportements revient à trouver un modèle inverse. Finalement, nous présentons un modèle inspiré de la psychologie sociale récemment développé qui constitue le point central de cette thèse.

3.2 Bases biologiques et psychologiques du comportement

3.2.1 Principes fondateurs

Pour conduire adéquatement une simulation concernant les comportements domestiques, deux solutions sont possibles : soit utiliser une base de données reprenant le plus précisément possible stimuli et réactions, soit s'inspirer de structures biologiques et de modèles issus de la psychologie sociale et cognitive, afin de simuler sur système informatique des facultés d'apprentissage et de réaction semblables à celles des habitants. Les approches faisant intervenir des bases de données ont été vues dans le précédent chapitre et nous avons expliqué leurs limites, en particulier concernant la gêne occasionnée aux habitants et la rigidité des systèmes. Dans cette partie nous présentons les principaux modèles issus des neurosciences et des sciences humaines.

En 1908 un psychologue français, Henri Piéron utilise le terme '*comportement*' pour la première fois (Piéron 1908). Ce terme désigne alors l'activité des êtres et leurs rapports sensorimoteurs avec le milieu. Dans le cadre qui nous intéresse, le milieu est constitué par l'environnement domestique et les rapports sensorimoteurs correspondent à un ensemble d'activités domestiques consommant de l'électricité. Des recherches plus récentes en psychologie ont montré qu'en plus de définir une action par son résultat comme le suggérait Piéron, il faut aussi prendre en compte l'intention. Plus précisément, il s'agit de préciser si une action est le résultat d'un processus guidé par l'environnement ou bien par un but supra ordonné. Cette première distinction effectuée il y a plus d'un siècle s'est révélée très juste, et a servi de point de départ à des recherches dans des domaines très variés. Le comportement est la partie observable d'une décision issue après un traitement très complexe. La perception de l'environnement physique est modulée par nos habitudes. Par exemple, regarder une montre indiquant midi n'a pas le même sens ni la même importance pour une personne habituée à sauter des repas ou pour une personne devant impérativement s'occuper d'enfants en bas âge.

Cet exemple permet aussi de voir que nos actions dépendent aussi des autres, certaines règles sociales nous empêchent de satisfaire à tout instant nos besoins, et d'autres règles obligent la prise en compte d'autres habitants partageant notre environnement. Des contraintes liées aux horaires de travail ont une forte répercussion dans le choix des activités, par exemple le lavage et repassage du linge le week-end car c'est à ce moment que l'on a le temps de les réaliser.

Des chercheurs en psychologie sociale, Aarts et Dijksterhuis (Aarts 2000) appellent '*habitudes*' les actions dirigées par l'environnement et montrent qu'à l'origine ces habitudes correspondaient à des comportements guidés par un but. Yin (Yin 2006) montre l'importance et le rôle de certaines structures cérébrales étroitement impliquées dans la formation d'habitudes.

Intuitivement, les actions dirigées par les buts sont contrôlées par leurs conséquences, alors que les habitudes le sont par des stimuli. Afin de traduire efficacement ces concepts intuitifs en éléments opérationnels et en hypothèses pouvant être testées, les auteurs définissent un cadre de travail mixte. D'une part, ils prennent en compte des comportements habituels et non-habituels, et les mettent en lien avec des découvertes récentes impliquant des structures du cerveau. Un ensemble de noyaux appelés ganglions de la base (voir figure 3.1) sont classiquement associés à la réalisation volontaire d'activités ainsi qu'à l'apprentissage d'habitudes.

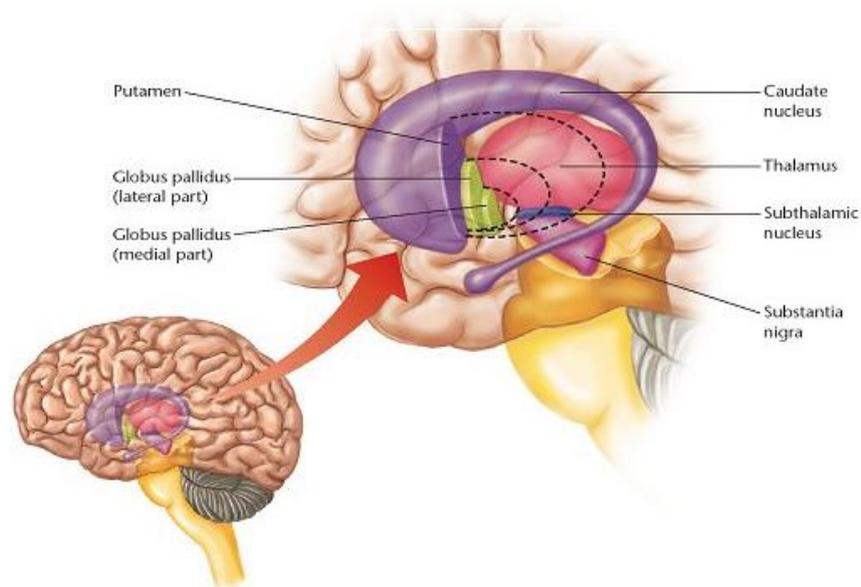


Figure 3.1 : localisation anatomique des noyaux gris centraux (aussi connu comme ganglions de la base) (source de la figure: Wikipédia)

Une meilleure compréhension du fonctionnement de cette structure permettrait de savoir comment les habitudes sont codées au niveau neuronal.

Dans le cadre du présent travail, il est particulièrement intéressant d'utiliser cette connaissance afin de proposer une implémentation informatique reproduisant certains mécanismes neurologiques à la base de l'apprentissage des habitudes.

3.2.2 Ganglions de la base et conditionnement opérant

Il a été vu précédemment que les habitudes constituent l'aboutissement d'un conditionnement. En effet, une habitude est essentiellement une action –ou plutôt une réaction- fréquente en réponse à des stimuli environnementaux relativement stables. Il existe essentiellement deux types de conditionnement: le conditionnement Pavlovien et le conditionnement opérant.

Le conditionnement Pavlovien est un réflexe conditionné qui consiste à associer un stimulus neutre (par exemple une image, une heure lue) au stimulus inconditionné (SI) (par exemple un repas servi). A l'origine, le stimulus neutre n'entraîne pas de réponse inconditionnée (RI). La simple vue d'une horloge indiquant midi ne donne pas faim ni ne donne envie de manger. Ce n'est que lorsque de manière constante le repas est servi à midi que le stimulus (l'horloge indiquant midi) est relié à une réaction organique, (sensation de faim, salivation). Le conditionnement pavlovien permet donc de se préparer à prendre une décision adaptée rapidement et sans réfléchir.

Le conditionnement opérant explique que nos comportements soient nettement plus influencés par les conséquences que provoquent ces comportements que par un stimulus présent dans l'environnement. Par exemple l'habitude de préparer ou servir le repas lorsque l'on a faim et que l'on voit l'horloge indiquer midi peut être expliquée par le conditionnement opérant, puisque servir un repas aura pour conséquence de pouvoir manger et donc assouvir sa faim. Le conditionnement opérant aussi appelé conditionnement instrumental, car contrairement au conditionnement pavlovien, une réponse organique ne suffit pas, un instrument est nécessaire pour que le comportement adéquat devienne une décision automatique.

Cette décision automatique n'étant pas innée, son apprentissage requiert une intervention active de diverses aires du cerveau. Il s'agit plus précisément d'une structure appelée ganglions de la base. Afin de traduire efficacement les concepts intuitifs d'apprentissage, de prise en compte des conséquences de l'action habituelle –ce qui constitue une forme de planification-, nous reprenons les travaux développés par l'équipe de Yin (Yin 2006). Leur travail montre que les ganglions de la base font partie de plusieurs réseaux neuronaux et que ceux-ci sont très impliqués dans les processus aboutissant à un conditionnement opérant permettant la formation d'habitudes et de décisions portant sur des actions immédiates.

Les ganglions de la base correspondent à un ensemble de noyaux situés à l'intérieur de l'encéphale (voir figure 3.1). Les principaux composants de cette structure sont : le striatum, le noyau sous-thalamique ainsi que le *Globus Pallidus*. Dans ce qui suit, nous allons décrire les ganglions de la base et montrer comment chacune de ses parties participe à l'apprentissage par renforcement.

Le fonctionnement de cette structure est directement lié aux principes de sélection : face à une situation donnée, on dispose d'un ensemble de réactions possibles ; par la suite, le système doit apprendre à préférer celles qui ont conduit à des résultats perçus comme positifs. Le *striatum* est la principale structure permettant l'entrée d'informations en provenance du cortex cérébral vers les ganglions de la base. Les connexions entrantes véhiculent des informations associées aux descriptions de l'environnement actuel. Grâce à celles-ci, le *striatum* génère un ensemble d'actions possibles pour une situation perçue et attribue des poids à chacune de ces actions en fonction de la récompense attendue. Ensuite interviennent les noyaux sous-thalamiques. Ceux-ci sont chargés d'effectuer le choix final de l'action. Par exemple, si deux actions possibles ont une récompense espérée très semblable, d'après (Zaghloul 2012), ils résoudront le conflit en prenant en compte la motivation interne du sujet à réaliser chaque action. En plus de favoriser les actions en fonction de leur motivation, les noyaux sous-thalamiques coordonnent la planification temporelle des mouvements composant l'action choisie, ce qui a pour résultat de rendre la réalisation de la tâche plus fluide dans le temps.

L'étape finale est constituée par le *Globus Pallidus* qui constitue la principale voie de sortie des ganglions de la base. Le *Globus Pallidus* a un rôle complémentaire à celui du cervelet : la coordination entre ces deux structures permet le contrôle et le bon déroulement des gestes –notamment l'absence d'à-coups ou de tremblements (Brotchie 1991) ainsi que l'organisation générale des mouvements (Gittis 2014).

Plusieurs modèles computationnels des ganglions de la base ont pu être implémentés en robotique. En s'inspirant de structures au sein du cerveau Doya réussit à programmer des robots afin qu'ils effectuent diverses tâches motrices, notamment en lien avec la locomotion (Doya 2000). Dans le cas qui intéresse cette thèse, nous souhaitons utiliser l'apprentissage par renforcement afin d'obtenir un système informatique capable de simuler des décisions, comportements et habitudes humaines. L'exploitation d'une structure d'apprentissage semblable à celle du cerveau est donc une approche particulièrement intéressante, mais qui doit être complétée. En effet, l'apprentissage par renforcement est un processus nécessaire mais pas suffisant pour expliquer le comportement domestique, puisque celui-ci n'est pas réductible à des habitudes immuables dans un environnement parfaitement stable.

L'approche dominante, qui permet de modéliser le conditionnement opérant en se basant sur les phénomènes neurologiques précédemment décrits, porte le nom d'apprentissage par renforcement. Cette approche a été implémentée informatiquement à de nombreuses reprises afin de modéliser des comportements d'apprentissage et de décision de

nombreuses espèces, par exemples la souris, le singe et l'homme (Sutton et Barto 1998, Meyer 2005, Salotti 2008).

Afin de définir l'approche par renforcement adéquatement, nous allons présenter les modèles et bases mathématiques permettant de simuler et reproduire les comportements habituels dans un environnement domestique.

3.2.3 Apprentissage par renforcement : présentation et bases opérationnelles

Dans la plupart des tâches de conditionnement opérant, le problème peut être mis sous la forme d'un processus de décision Markovien (PDM). Dans les PDM, l'environnement correspond à un 'état' parmi un ensemble d'états possibles décrivant des caractéristiques d'intérêt. Par exemple, dans le domaine des habitudes domestiques, prenons le cas du lavage du linge. Un environnement est décrit par un état où beaucoup de linge est sale. A partir de cet état, plusieurs actions sont possibles :

- l'action 'faire une lessive' qui provoque un passage vers l'état 'peu de linge sale' si le linge à laver requiert les mêmes conditions de lavage.
- l'action 'remettre le lavage à plus tard', qui a pour effet de laisser l'environnement actuel dans le même état 'beaucoup de linge sale'.

Le schéma de PDM correspondant à cet exemple est présenté ci-dessous.

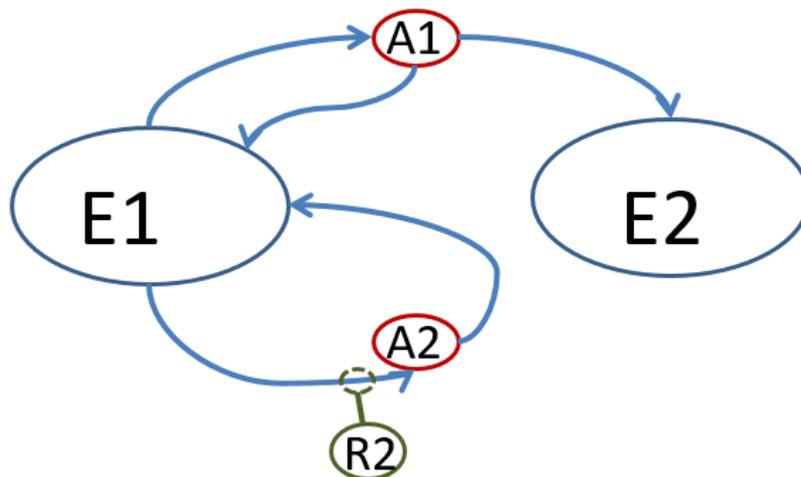


Figure 3.2 : illustration du schéma de PDM correspondant à l'exemple présenté dans le texte. E1 correspond à l'état 'beaucoup de linge sale', E2 correspond à l'état peu de linge sale. A1 correspond à l'action 'faire une lessive', A2 correspond à mettre en attente le lavage du linge. R2 est une des récompenses prévues lorsque le choix de se reposer est fait alors qu'il y a beaucoup de linge sale.

Une fois qu'un habitant a déterminé son environnement et lui a attribué un état, il est en mesure de choisir une action dont l'issue paraît agréable ou positive. Cette notion d'agrément est définie en fonction de la récompense associée à une action ou un état. Dans l'exemple présenté, la récompense peut être associée au bien-être ressenti lorsque le linge est propre, ou au bien-être ressenti lorsque l'habitant choisit de se reposer. Dans une situation donnée (état actuel), plusieurs actions sont possibles et peuvent être adéquates. La manière dont on choisit l'action à mener dans une situation donnée porte le nom de 'politique de décision' (Policy en anglais). Dans l'exemple du linge sale, une politique de décision pourrait être de privilégier deux fois sur trois le lavage du linge sur le repos ou l'attente. Connaître la politique de décision optimale pour une situation donnée, ou connaître la politique de décision adoptée par des habitants d'une résidence est un problème complexe car le choix des actions peut avoir des répercussions à la fois sur le court et sur le long terme.

L'un des algorithmes permettant un apprentissage par renforcement aussi bien des habitudes que de l'environnement est l'algorithme *Q-Learning*. La méthode se base sur des erreurs de prédiction entre la récompense attendue et la récompense effectivement obtenue après une action. Les erreurs de prédiction permettent une prise en compte plus fine des conséquences des choix d'action, et à la longue favorisent certaines décisions dans des états précis. L'algorithme est présenté à la page suivante.

Initialisation des valeurs $Q(s, a)$ arbitrairement

Pour tous les épisodes

Initialisation de s

Tant que s n'est pas un état final et/ou absorbant réaliser :

Choisir une action, observer r et s'

Mise à jour de $Q(s, a)$:

$$Q(s, a) = Q(s, a) + \eta(r + \gamma \max_{a'} Q(s', a') - Q(s, a))$$

$$s = s'$$

Fin (Tant que)

Fin

Algorithme 1 : Q-Learning

$Q(s, a)$ est une fonction d'association permettant d'attribuer à chaque choix d'action a de chaque état s une valeur de récompense.

L'algorithme *Q-Learning* est couramment utilisé en robotique et en neurosciences pour modéliser l'apprentissage par renforcement (Sutton 1998). Cet algorithme exploite les observations d'états et les récompenses pour définir un PDM très proche des données comportementales. Cette approche est intéressante en raison de son inspiration neurobiologique, qui permet d'expliquer en grande partie la qualité des résultats. Certains auteurs suggèrent implicitement que l'efficacité des chaînes de Markov est liée au fait que les activités domestiques correspondent à un phénomène stochastique au même titre que – par exemple- le nombre de pièces défectives produites par une machine. En complément, il nous semble important de préciser que l'efficacité des chaînes de Markov pour modéliser les conduites domestiques –en particulier les conduites habituelles- vient du fait qu'il s'agit de comportements appris par renforcement et qui peuvent donc être adéquatement représentés par un PDM.

Malgré le succès rencontré par différents algorithmes d'apprentissage par renforcement pour expliquer et simuler certains comportements habituels, certaines décisions et actions restent difficiles à expliquer. Par exemple, après avoir reçu une récompense inattendue, les gens n'ont pas tendance à répéter l'action à l'origine de la récompense, mais à alterner les

actions choisies (Shteingart 2014). De plus, les modèles actuels d'apprentissage par renforcement ont des difficultés à reproduire la variabilité des résultats observés lors d'études portant sur des perceptions et des comportements pourtant très simples (Worthy2013). Une des difficultés majeures est de prendre en considération toutes les variables physiologiques, socioprofessionnelles, environnementales, culturelles et contextuelles qui définissent les comportements.

3.3 Modèles psychologiques du comportement

3.3.1 Objectif

Pour modéliser adéquatement le comportement, en plus de s'intéresser aux processus neuronaux, il est particulièrement intéressant de considérer une approche psychologique afin de prendre en compte des processus cognitifs de plus haut niveau. Par ailleurs le choix de se tourner vers des modèles psychologiques semble tout à fait justifié : nous souhaitons simuler des processus de pensée dont l'abstraction est difficilement réductible à des traitements réalisés par des structures biologiques du cerveau. A partir d'un certain niveau de décision, l'échelle d'analyse change, on passe du groupe de neurones à l'individu et parfois même à l'interaction entre individus.

3.3.2 Approches centrées sur l'individu : modèles issus de la psychologie cognitive et sociale

3.3.2.1 Retour sur les modèles multi-agent : modèle BDI

Les méthodes issues des sciences humaines et sociales permettent de considérer l'être humain comme un agent agissant et interagissant au sein d'un environnement physique mais aussi social. Ce point de vue nous amène à considérer un schéma cognitif en plus d'une description basée sur des données statistiques. Ainsi, il devient possible d'établir un lien plus explicite entre les consommations électriques et les variables comportementales. Dans leur article Le et al (Le2010) utilisent une architecture d'agent belief-desire-intention (BDI) qui leur permet de reconstruire des données à partir de règles comportementales et environnementales issues de bases de données adéquates. Cependant, le schéma d'agent BDI, bien que très utile, n'est pas totalement adapté pour simuler des comportements d'habitants. En effet, dans le cadre qui nous intéresse, le comportement ne peut être réduit à une liste de moyens à utiliser de manière optimale pour satisfaire des besoins. Un comportement est l'aboutissement d'un processus de réflexion de niveau variable qui n'a pas nécessairement d'utilité associée.

3.3.2.2 Autres méthodes

Dans le domaine de la modélisation comportementale bio-inspirée, plusieurs approches sont possibles afin d'obtenir des résultats plus représentatifs que ceux obtenus uniquement avec des PDM. Jusqu'à présent, les approches publiées appartiennent à l'un des deux groupes ci-dessus présentés :

- Un premier groupe qui rassemble les méthodes qui s'inscrivent dans la continuité des travaux réalisés en neurosciences, en particulier des techniques d'apprentissage par renforcement précédemment expliquées. Ces recherches visent l'utilisation de modèles computationnels plus évolués. En particulier il devient possible d'intégrer des systèmes d'apprentissage supervisé et non-supervisé. Dans ce cas, l'échelle d'analyse reste inchangée et se situe au niveau de structures cérébrales, on met l'accent sur l'utilisation de modèles apparentés à ceux utilisés en IA.
- Un deuxième groupe de méthodes se basant sur un niveau d'analyse différent. Au lieu de réaliser des études portant sur des structures du cerveau, il est possible de se baser sur des processus de décision. A la place d'une approche centrée sur les neurosciences, on modélise le traitement de l'information tel qu'il est appréhendé en psychologie cognitive. Dans ce second groupe on met l'accent sur les mécanismes généraux de traitement de l'information opérés par le cerveau, en particuliers sur les déterminants (Grégoire 2013), ainsi que sur l'émergence des habitudes (Judah 2013). L'échelle d'analyse est donc différente. Dans ce cas, une modélisation computationnelle ad hoc doit être développée. En effet la psychologie computationnelle est nettement moins développée que les neurosciences computationnelles, qui font l'objet de nombreuses études (Bennett 2015).

3.3.2.3 Alternative envisagée

Afin d'exploiter le meilleur de chacun de ces groupes, nous proposons une approche hybride permettant à la fois de reprendre la vue d'ensemble des processus cognitifs propres à la psychologie et le formalisme mathématique propre aux neurosciences computationnelles.

C'est ce qui a motivé notre utilisation d'un schéma de raisonnement basé sur le modèle développé par Fiske et Neuberg (Fiske1990, Fiske 1999) en psychologie sociale. Leur modèle définit d'une part des réponses schématiques, basées sur les habitudes et d'autre part des réponses spécifiques basées sur une réflexion plus approfondie et propre à la situation. Les habitudes permettent de savoir rapidement et sans effort cognitif quel type d'action peut être adapté à une situation courante. Dans le cas où les habitudes ne suffisent pas à décider, des informations plus spécifiques à la situation et un raisonnement approfondi permettent de choisir une action appropriée.

Dans le cadre qui nous intéresse, cette vision hiérarchique du raisonnement est très adaptée. En effet, elle permet d'expliquer que l'on trouve une certaine régularité horaire dans les activités quotidiennes, mais elle permet aussi d'expliquer pourquoi ces habitudes ne sont parfois pas respectées : d'autres envies peuvent prendre le dessus. Toutefois, lorsqu'il s'agit d'un modèle issu de la psychologie, certaines adéquations au formalisme informatique seront nécessaires afin de le rendre opérationnel au sein d'un système multi agents.

3.4 Implémentation de la première version du modèle

3.4.1 Obtention des données

Pour l'obtention des données comportementales, nous avons décidé d'utiliser des données de consommation électrique. Il est en effet possible de déduire que, par exemple, si un pic de consommation électrique est enregistré de 11h à 12h30 au niveau du four, cela signifie qu'un repas était probablement en préparation. De la même manière, il est possible de retrouver des actions portant sur le lavage du linge, de la vaisselle ainsi que sur l'utilisation de la télévision et l'ordinateur et d'autres appareils.

Notre première hypothèse concerne le lien à établir entre perception, action et représentation interne.

Nous postulons qu'une représentation interne I est associée à chaque action domestique A faisant intervenir un appareil électrique au sein d'une habitation H . Nous utilisons le formalisme bayésien pour désigner et opérer sur les représentations internes. Par exemple, $P(A|I,H)$, représente la probabilité que les habitants de H réalisent l'action A , sachant que la représentation interne de la tâche A correspond à I . Nous considérons pour l'instant que les représentations internes peuvent adéquatement être représentées comme des distributions de probabilité définissant un 'a priori'. Par exemple dans le cas d'habitants attachant une grande importance à la propreté, la probabilité d'observer une mise en marche du lave-linge sera *a priori* plus élevée que la moyenne.

3.4.2 Présentation du cadre bayésien

3.4.2.1 Principe général

Nous utilisons des représentations probabilistes comme des pivots entre la perception d'un besoin N et la réalisation de l'action A impliquant l'utilisation d'un appareil électrique. En d'autres termes, l'hypothèse est qu'il existe un codage interne des tâches, commun à la perception et à l'action.

Précédemment, nous avons présenté plusieurs modélisations. Bien qu'issues de familles différentes, la plupart représentent les activités domestiques dans un espace de

représentation des tâches, donc basé surtout sur des observations. Bien que permettant de reproduire des comportements avec une certaine fidélité, ces modèles permettent difficilement d'expliquer la différence entre activités habituelles, activités planifiées et activités non habituelles. Ces approches fréquemment utilisées sont justifiées dans un cadre industriel, par exemple pour l'obtention d'une courbe de demande énergétique pour une population stable et à très court terme. Cependant ces approches sont limitées car elles ne permettent pas d'aboutir à un modèle comportemental représentatif du système cognitif humain. Nous soutenons que l'utilisation d'un cadre bayésien bio-inspiré, permettant de représenter et différencier plusieurs types de comportements, constitue une étape indispensable pour réduire la dépendance des prédictions aux bases de données de consommation électrique. Grâce au cadre bayésien proposé, il est possible de reproduire une variété de comportements portant sur des activités, mais aussi d'en savoir plus sur leurs interactions. Grâce à une meilleure connaissance des interactions entre activités, il devient par exemple possible d'utiliser ce que la base de données révèle sur les habitudes d'utilisation de la cuisinière électrique afin de mieux prédire celles concernant le lave-vaisselle.

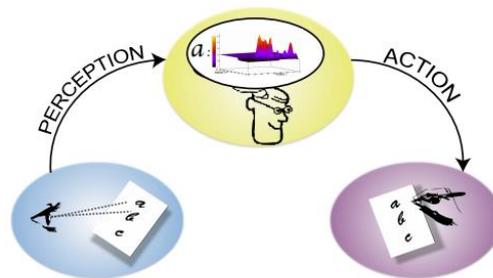


Figure 3.3 : les habitudes servent de pivot entre perception et action

Le formalisme bayésien permet d'associer à une représentation une distribution de probabilité conjointe concernant des paramètres définissant les habitudes tels que l'heure d'utilisation habituelle d'un appareil ainsi que sa fréquence d'utilisation. Connaissant cette probabilité conjointe, le modèle est capable de résoudre plusieurs tâches cognitives. Chacune de ces tâches correspond à une question probabiliste qui est résolue automatiquement par inférence bayésienne :

- 1- la perception des besoins en lien avec les besoins physiques (l'hygiène ou alimentation par exemple).
- 2- la réalisation des activités à l'échelle domestique.

Par ailleurs le cadre proposé permet également de :

3– reconnaître une maisonnée grâce aux données de consommation

4– simuler des consommations domestiques lorsque l'équipement électroménager varie.

3.4.2.2 Formalisme bayésien

La formulation bayésienne permet d'effectuer des prédictions et de plus suggère de nouveaux protocoles expérimentaux, afin de non seulement simuler des activités mais aussi des besoins.

Dans la présente thèse, nous nous focalisons essentiellement sur les points 1 et 2, et donc essentiellement sur la simulation prédictive de consommation électrique. La première hypothèse que nous faisons afin de définir le cadre bayésien concerne la représentation interne des tâches domestiques. Nous postulons que pour une habitation H faisant intervenir un appareil A chaque tâche ou action domestique possède une représentation I_A qui peut être approximée par une fonction de distribution de probabilité. Ainsi, il est possible de représenter sous une forme mathématique compacte des questions portant sur les consommations, les activités des habitants ou des caractéristiques portant sur leurs représentations des tâches. Par exemple le calcul de $P(I_A|AH)$ permet de connaître la distribution de probabilités décrivant la représentation d'une tâche connue pour une habitation en particulier.

De plus, nous faisons l'hypothèse que ces représentations I_A agissent comme des pivots entre la perception d'un besoin N et une action A impliquant l'utilisation d'un appareil électrique. L'utilisation de propriétés de dépendance marginale ou conditionnelle permet de simplifier les formules en réduisant les facteurs nécessaires au calcul de probabilités conjointes.

Il y a indépendance marginale lorsque l'on a :

$$\forall b \in D_b, P(A | [B = b]) = P(A)$$

On dit que A et B sont indépendants. Nous pouvons alors décomposer la distribution de probabilité conjointe de la façon suivante :

$$P(A B) = P(A) P(B)$$

Il y a indépendance conditionnelle lorsque l'on a :

$$P(A|B C) = P(A|B)$$

On dit que A est indépendante de C sachant B. On notera que l'indépendance conditionnelle n'implique pas l'indépendance marginale, et réciproquement.

Si la représentation est connue, connaître la probabilité de réalisation de la tâche n'apporte pas d'information supplémentaire. La probabilité que les habitants d'une maison aient une représentation des tâches I_A , un besoin N à satisfaire et réalisent la tâche A s'écrit :

$$P(A | H, I_A, N) = P(A) P(H) P(I_A | A, H) P(N | I_A)$$

Ce cadre probabiliste permet de poser et définir le problème, mais ne suffit pas à lui seul pour réaliser une simulation de consommations centrée sur le comportement humain. En effet, nous souhaitons dépasser les représentations purement probabilistes d'actions. En conséquence, nous proposons un modèle agentique bio-inspiré permettant de prendre en compte une dynamique 'humaine'. Ce modèle d'agent basé sur un schéma de prise de décision issu de la psychologie permet de moduler les probabilités d'actions en fonctions de paramètres cognitifs. Dans ce qui suit, nous allons décrire le modèle qui complète le cadre bayésien et explique le choix des paramètres.

3.5 Apport de la psychologie cognitive au modèle bayésien de prise de décision

3.5.1 Classification des modèles

La psychologie cognitive propose différents modèles concernant la représentation des connaissances et la prise de décision. La tâche de modélisation de processus cognitifs de haut niveau est particulièrement ardue. En effet les différents facteurs intervenant dans la prise de décision (environnement physique et humain, goûts et histoire personnelle...) interagissent de manière complexe entre eux. Afin de faire face à ces difficultés, plusieurs modèles sont proposés par la psychologie cognitive et sociale. Ce qui distingue ces différents modèles c'est non seulement leur champ d'application, mais surtout la manière dont les processus portant sur la prise de décision sont traités. Nous allons présenter les deux familles regroupant les principales approches utilisées en psychologie.

Nous avons précédemment vu l'importance de l'apprentissage par renforcement. En particulier, cet apprentissage permet sans effort conscient de parvenir à une certaine régularité dans les actions réalisées à domicile. Cet apprentissage des habitudes s'effectue donc automatiquement. D'autres décisions requièrent cependant un effort conscient. Il peut s'agir de décisions prises dans un cadre exceptionnel. Par exemple, si le vendredi après-midi est l'instant habituel pour réaliser le repassage et que des invités passent à l'improviste, les règles sociales incitent à un report de l'activité de repassage. Dans ce cas, il faut réaliser un effort conscient de planification afin de savoir réagir à une situation imprévue. Les modèles à

double flux postulent que les processus automatiques et les processus dirigés sont dissociés. Les modèles à simple flux, expliquent que les décisions spécifiques et les habitudes ont pour base un même processus sous-jacent. Dans la partie qui suit, nous allons développer les idées propres à chacune des familles de modèles et nous concluons en expliquant le choix retenu dans le cadre de nos travaux.

3.5.2 Modèles à deux processus

3.5.2.1 Principe des deux processus

Afin de mieux comprendre comment l'être humain est capable de percevoir et de donner un sens à l'environnement (c'est-à-dire l'ensemble des facteurs aussi bien psychologiques que physiques et sociologiques) et d'avoir un comportement adéquat avec ses percepts, il est possible de raisonner en termes de processus automatiques ou volontaires. Il semble naturel de séparer les processus conduisant à des actions inconscientes et automatiques des processus permettant de prendre des décisions planifiées et plus complexes, qui requièrent une volonté et une motivation. Cette hypothèse simple permet d'expliquer qu'à partir de deux *processus ou flux* de traitement d'information qualitativement distincts, on puisse aboutir à des processus de décision différents. Bien que très intuitif, cette représentation des opérations mentales s'avère d'une grande richesse ; par exemple elle permet d'expliquer que l'on puisse passer d'un mode de pensée superficielle à une réflexion profonde en un laps de temps relativement court. Cependant, la manière dont coexistent ces deux flux est loin d'être évidente, afin de mieux cerner la manière dont les deux flux peuvent interagir nous allons considérer 4 exemples.

3.5.2.2 Modèles sélectifs

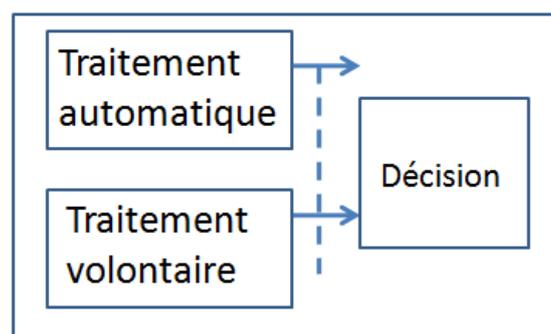


Figure 3.4 : Schéma général du modèle sélectif

Le premier cas concerne les modèles sélectifs se conformant aux préceptes behavioristes. Il s'agit de modèles relativement simples à interpréter. Ils permettent d'expliquer pourquoi les gens se comportent si différemment en fonction de l'environnement et du contexte. Il y

aurait dans ce cas une activation du processus (volontaire ou automatique) le plus adapté à la situation en cours (Chaiken 1987, Schwarzmoood 1986). Dans chaque situation, deux processus qualitativement différents coexistent mais un seul est activé en fonction de stimuli externes et dans une moindre mesure de l'histoire personnelle du sujet.

3.5.2.3 Modèles compétitifs

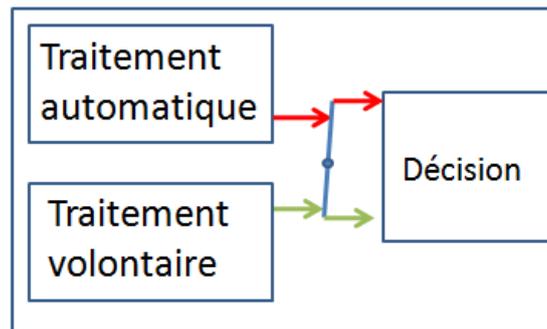


Figure 3.5 : Schéma général du modèle compétitif

Le deuxième cas regroupe **les modèles compétitifs**. Ici les processus automatiques et volontaires sont activés simultanément, mais la décision dépend uniquement de celui qui est plus rapide ou dont la réponse est plus « forte », il s'agit de traitements de type « winner-takes-all ». On peut noter que les cartes auto-organisatrices et les colonnes corticales fonctionnent précisément de cette manière avec des connexions latérales fortement inhibitrices (Alexandre 1990). Dans le cas de la prise de décision contextuelle, les psychologues sociaux choisissent rarement ce schéma, qui pour eux est une simple variante du cas sélectif. En effet, dans les deux cas, un seul processus contrôle le comportement. Dans les schémas sélectifs, l'un des processus est au repos et dans les schémas compétitifs l'un des processus, bien qu'actif, n'a pas d'effet sur la décision finale. Des psychologues ainsi que des neuroscientifiques proposent de voir le cerveau comme un système massivement parallèle capable de générer une réalité en intégrant des résultats issus de différents modules (Fodor 1983, Gazzaniga 1988, Marenla 2004). Il apparaît que les opérations menées par ces modules ainsi que les résultats partiels échappent à notre conscience, ce qui évite que nous soyons submergés par une masse colossale d'informations non pertinentes.

3.5.2.4 Modèles consolidatifs

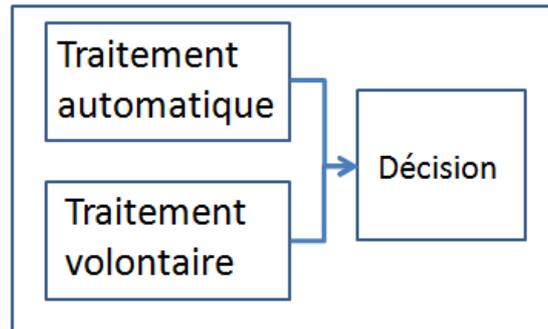


Figure 3.6 : Schéma général du modèle consolidatif

Les modèles sélectifs et compétitifs peuvent expliquer la diversité de résultats comportementaux observés. Dans le cas des modèles sélectifs, le choix entre les traitements automatiques et volontaires est effectué à un stade précoce, tandis que dans le cas compétitif le choix entre mécanismes est réalisé juste avant l'apparition du comportement observable. Cependant cette 'bataille' entre processus internes n'est pas indispensable. En effet, un stimulus peut simultanément activer les deux processus qui co-construiront la perception et le comportement. En psychologie sociale, les modèles consolidatifs sont très couramment utilisés et dans une variété si grande qu'ils ne constituent pas une catégorie unique. Par exemple, aucune théorie ne suggère que notre perception des autres se base soit sur leur comportement verbal, leur comportement non-verbal ou leur catégorie sociale. Notre perception des autres est construite simultanément à partir de considérations plus ou moins conscientes de chacune de ces informations. Une impression est formée lorsque différents éléments semblent avoir suffisamment en commun pour conduire à une représentation cohérente : chaque module doit contribuer de manière constructive (consolider) à l'évaluation des autres modules (Carlston 1992). Parfois nos impressions sur une situation en particulier se basent plus sur ce que nous entendons, voyons ou encore sur le degré de danger ressenti. Les modèles consolidatifs sont particulièrement utiles lorsqu'il s'agit d'expliquer que les comportements observés résultent d'une pondération variable et adaptative des conclusions issues de divers processus, qu'ils soient automatiques ou volontaires.

3.5.2.5 Modèles correctifs

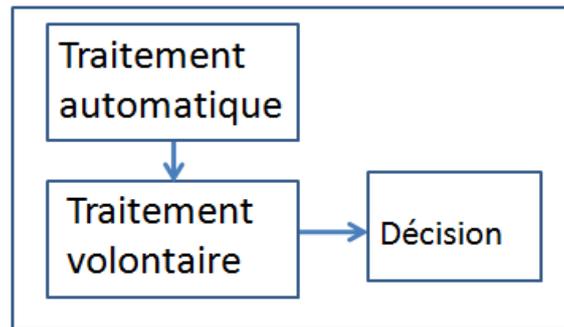


Figure 3.6 : Schéma général du modèle correctif

Dans le modèle correctif comme dans le précédent modèle consolidatif, les sorties des processus automatiques et volontaires contribuent conjointement à l'élaboration de la décision finale. Cependant le résultat final du modèle correctif semble être capable d'apporter davantage de finesse dans le résultat construit.

Tout comme le modèle correctif est une variante du modèle sélectif, le modèle correctif peut être vu comme une variante du modèle consolidatif. Il convient toutefois de montrer les deux principales différences qui les distinguent. Premièrement, dans le cas du *design* consolidatif, un stimulus unique peut activer simultanément les processus automatique et volontaire, et leurs contributions à la décision finale sont déterminées par un système de pondération variable. Deuxièmement, le système de pondération variable ne permet pas de corriger des erreurs propres au processus, mais simplement de s'adapter à l'environnement en considérant qu'il n'y a pas d'incertitude liée aux processus internes. Dans le cas du modèle correctif, un stimulus met en route un seul processus à la fois, c'est le résultat du premier processus qui active le deuxième processus. Le deuxième processus modifie, ajuste ou corrige le résultat du premier, et de plus prend en compte d'éventuelles erreurs du premier processus lors de son action. Le modèle correctif est donc a priori plus robuste que le modèle consolidatif puisqu'il est capable de corriger certaines erreurs dues au système.

Ce type de modèle correctif est le plus couramment utilisé en psychologie sociale. En particulier, Fiske et Neberg utilisent une architecture proche pour expliquer comment se construisent les connaissances et les décisions concernant autrui (Fiske et Neberg 1990).

3.5.3 Implémentation du modèle de Fiske et Neberg

3.5.3.1 Quels comportements considérer ?

Nous nous intéressons à la consommation électrique de l'habitat, et plus particulièrement à celle qui est liée à l'utilisation des appareils électroménagers. La base de données dont nous disposons recense les usages de la télévision, de la chaîne hifi, du four, du lave-linge, du lave-vaisselle et de la friteuse. Nous l'avons déjà indiqué, l'usage de ces appareils dépend de multiples facteurs qu'il n'est pas possible de considérer de façon exhaustive. De manière

générale, il existe des comportements guidés par des besoins physiologiques (par exemple se nourrir), par des règles sociales (maintien de la propreté) et par des envies récréatives ou culturelles (ordinateur, télévision et autres équipements multimédia). Certains de ces comportements deviennent rapidement des habitudes journalières ou suivent des rythmes particuliers sur la semaine. Cependant, certains sont relativement indépendants des facteurs qui déclenchent les autres comportements, alors que d'autres sont très liés. Par exemple, lorsque l'heure du déjeuner arrive, la préparation du repas est en général prioritaire sur toutes les autres activités. Démarrer le sèche-linge, en revanche, est temporellement très corrélé à l'action de laver le linge. Dans ce cas, le processus de décision est sans doute plus délibératif. La prise en compte de ces dépendances est fondamentale pour caractériser les comportements associés à la consommation électrique. Notre problème est de distinguer ces comportements en exploitant les différents mécanismes de catégorisation du modèle de Fiske et Neuberg et de paramétrer en même temps le processus de décision en fonction des données comportementales existantes associées à la consommation électrique.

3.5.3.2 Au-delà du paradigme des deux processus

Le modèle de Fiske et Neuberg se rapproche du modèle correctif précédemment présenté. Cependant, une nette différence le distingue de celui-ci. En effet, Fiske et Neuberg expliquent qu'en réalité la distinction entre processus automatiques et processus volontaires n'est pas qualitative. La différence entre les deux processus réside essentiellement dans les ressources utilisées, donc la distinction entre processus automatiques et volontaires serait purement quantitative.

Cette analyse est particulièrement intéressante pour deux raisons. Premièrement, elle permet de simplifier l'hypothèse de départ portant sur deux processus. A présent, un seul flux permet de rendre compte aussi bien de comportements considérés automatiques que de comportements volontaires. Deuxièmement, dans le présent travail, nous souhaitons proposer une modélisation informatique agentique. De prime abord, définir un agent à flux unique semble diminuer la complexité du travail de programmation ainsi que la puissance de calcul requise.

Nous exposons le modèle de Fiske et Neuberg (figure 3.4) ainsi que notre approche permettant d'utiliser dans un système informatique un schéma d'agent proche de leur schéma cognitif. Nous procédons ainsi pour chaque étape.

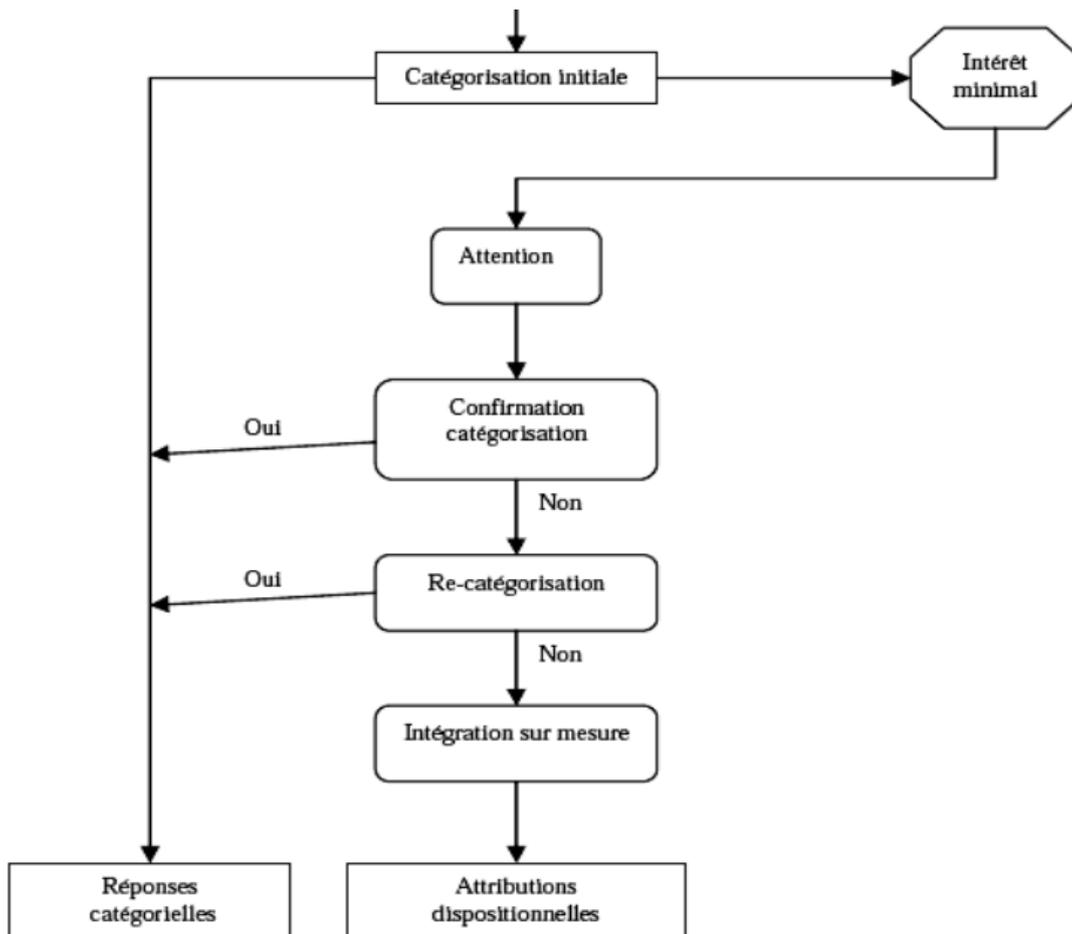


Figure 3.7 : Diagramme de flux du modèle original de formation des connaissances et de décision proposé par Fiske et Neuberg (1990)

3.5.3.3 Catégorisation initiale

L'objectif de cette étape est de proposer un type d'action à un instant t de la journée en cours. Si nous nous en tenons au modèle psychologique, la première étape de traitement doit être faite grâce à un descripteur global, c'est-à-dire portant sur des classes de comportement vastes. Nous proposons que la catégorisation différencie globalement deux classes d'actions bien différentes : celles qui concernent les besoins physiques (cuisiner, laver...) et celles qui concernent les activités récréatives (chaîne hifi, télévision).

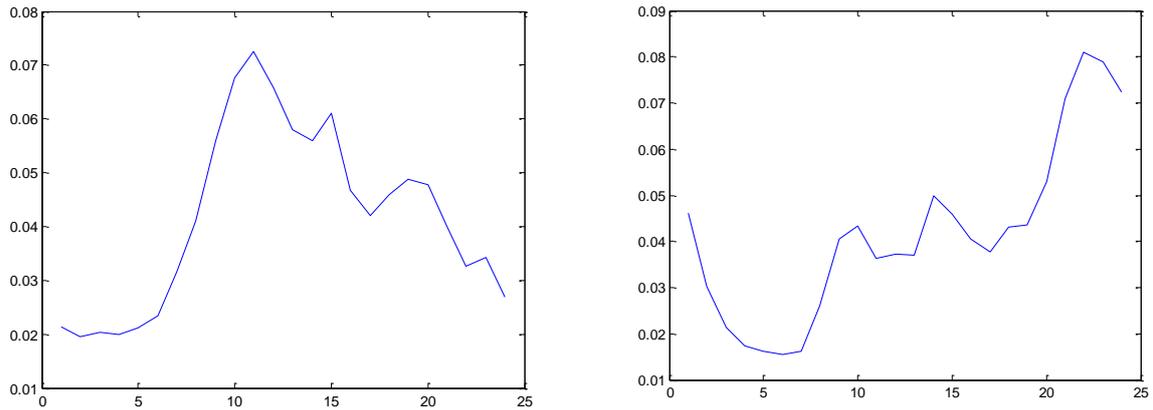


Figure 3.8 courbes de Densité de Probabilité décrivant la probabilité d'utiliser un appareil en lien avec les besoins physiques (à gauche) et de récréation (à droite).

Dans le cas qui nous intéresse, il faut classifier les usages des 6 appareils précédemment cités. La classification par mixtures gaussiennes permet d'obtenir deux prototypes, chacun d'eux représentant fidèlement sa classe d'appartenance. La figure 3.5 montre une représentation graphique des prototypes obtenus pour le groupe 1 et 2 après normalisation.

Il est possible de classifier les activités de la base de données en fonction de leur distance de répartition horaire à ces prototypes par la méthode du plus proche voisin. La classification retenue ne prendra pas en compte le four, les données concernant son utilisation se sont avérées insuffisantes. Le tableau 3.1 montre les résultats de la classification finale:

Tableau 3.1 : Classification des appareils électroménagers

Appareil électrique	Groupe attribué
Télévision	2
Lave-vaisselle	1
Lave-Linge	1
Chaine Hi-Fi	2
Friteuse	1

Pour décider quel type d'action doit être entrepris, on calcule la probabilité de réaliser une activité appartenant au groupe i à l'instant t notée $P_t(A_i)$. Dans ce premier niveau d'analyse, on considère que :

$$P_t(A_i) = P(A_i|H) \quad (3)$$

où H correspond à l'heure actuelle de la journée.

La règle de décision qui permet de savoir s'il est nécessaire de passer au niveau délibératif suivant correspond à :

$$\sum_{i=1}^2 P(A_i|H) > 0.5 \quad (4)$$

L'équation 3 permet de modéliser la formation d'une intention en lien avec les habitudes. L'équation 4 permet de savoir si le processus de réflexion ira plus loin ou s'il s'agira uniquement d'une intention restée sans suite. L'utilisation des prototypes de classification, ainsi que des règles de décision basées sur (3) et (4) constituent le bloc 'Catégorisation Initiale' de la figure 1.

3.5.3.4 Catégorisation confirmatoire

L'étape de traitement qui suit la catégorisation initiale se caractérise par une attention plus soutenue et une évaluation plus précise de la situation que celle effectuée précédemment. Ce qui nous intéresse à ce stade, n'est pas de reproduire des patterns d'utilisation d'appareils ou d'habitudes observés, mais d'en générer de nouveaux. En effet, même en absence d'imprévus, notre comportement s'éloigne parfois de nos habitudes. Par exemple, l'instant où nous prenons un repas n'est pas seulement conditionné par l'habitude de dîner à heure fixe, mais aussi par l'heure du repas précédent. Une loi de distribution de probabilité peut adéquatement rendre compte de ce phénomène. En raison de certaines régularités horaires liées aux besoins physiques et aux habitudes, une loi de Poisson décrit adéquatement la probabilité d'utiliser k fois dans un intervalle de temps un même appareil électrique. Lorsque le nombre moyen λ d'utilisations dans cet intervalle de temps de cet appareil est connu, la fonction de distribution de probabilité correspond à :

$$P(X_a = k) = e^{-\lambda_A} \frac{\lambda_A^k}{k!} \quad (5)$$

Le paramètre λ_A de cette distribution peut être directement estimé à partir du profil moyen de consommation de l'appareil a , il suffit de compter le nombre d'utilisations et de le diviser par le nombre d'unités de temps de la période considérée.

En plus de $P(A_i|H)$ qui est obtenue directement à partir de la lecture du profil journalier de consommation, nous disposons de la loi de Poisson (5). Une technique appelée Méthode de la Fonction Inverse (MFI) permet de générer des suites aléatoires et non répétitives qui suivent la loi de distribution de l'équation (5). Cette méthode est décrite de manière très détaillée dans un article de Page (Page 2008).

En utilisant la MFI, on génère une décision concernant l'appareil a_i d'après la fréquence observée d'utilisation de celui-ci. En utilisant toutes les informations dont nous disposons à présent, et en sachant que n_{A_i} représente le nombre de fois que l'appareil A_i a été utilisé, nous pouvons donner une nouvelle définition de $P_t(A_i)$:

$$P_t(A_i) = P(A_i|H, n_{A_i}) \quad (6)$$

L'utilisation de l'équation (6) pour chaque appareil électrique donne un ensemble de probabilités d'utilisation de chaque appareil A_i . Dans le cadre d'une simulation, pour définir l'utilisation de chacun des appareils, on génère un nombre aléatoire $S \sim \mathcal{U}(0,1)$. Si $S < P(A_i|H, n_{A_i})$, alors la décision de réaliser l'action A_i est prise.

Afin de tirer pleinement parti du schéma d'agent bio-inspiré utilisé, nous proposons l'inclusion d'une variable comportementale décrivant la tendance à passer à un niveau supérieur de délibération. Afin de simplifier l'utilisation de ce modèle, nous choisissons de représenter cette tendance simplement comme une probabilité constante de passer au niveau supérieur. En psychologie, cette tendance peut être expliquée par la motivation du sujet et possède une influence sur le comportement. Par la suite, le modèle pourra être amélioré, par exemple, en proposant une variable basée sur des réponses à un test psychométrique de mesure d'autodétermination. D'autres éléments ayant une influence marquée sur la motivation pourront également être pris en compte, tels que la fatigue ou le stress.

Dans ce module de décision, nous avons pris en compte la période entre réalisations d'une même activité. A présent, nous allons proposer un module de décision qui intègre un autre ensemble de règles, afin de prendre en compte plus finement l'information contenue dans la base de données de profils journaliers d'utilisation.

3.5.3. Recatégorisation

A notre connaissance, la plupart des études portant sur les consommations électriques domestiques ont très peu traité du lien existant entre utilisations de divers appareils électriques (Wilke 2013). Pourtant, l'information contenue dans la corrélation entre activités possède un double intérêt que nous allons expliquer par le biais d'un exemple. Le repassage du linge est normalement effectué une fois que celui-ci a été lavé et séché, il est donc naturel de prévoir qu'un certain temps après la mise en marche de la machine à laver, suivra une utilisation du fer à repasser. Le délai habituellement écoulé entre ces deux activités peut être calculé par corrélation croisée normalisée des profils de consommation de la machine à laver et du fer à repasser. Ce délai révèle deux liens : un premier lien statistique entre deux variables aléatoires de répartition connue, et un lien plus profond qui correspond à une planification d'activités.

Les processus de décision, tels que décrits en psychologie sociale, confirment que lors d'évaluations attentionnelles, des interactions de haut niveau sont prises en compte dans la formulation de décisions (Turner 1986).

Lors de l'analyse par corrélation croisée de notre base de données, nous avons constaté un délai significatif séparant l'utilisation du lave-linge et celui du lave-vaisselle. Dans ce cas, la cause n'est pas à rechercher dans la planification, mais plutôt dans la prise en compte de restrictions matérielles. En effet, il n'est parfois pas possible d'allumer le lave-vaisselle et le lave-linge en même temps, car le débit d'eau ou de courant requis peut dépasser les capacités de l'installation domestique. Afin de prendre en compte les corrélations entre activités, nous proposons d'utiliser à nouveau la méthode de la fonction inverse sur une distribution de Poisson. Cette fois, le calcul du paramètre λ est quelque peu différent :

$$\lambda_{i,j} = \frac{24}{\text{argmax}(A_i * A_j)}, i \neq j \quad (7)$$

Le symbole * est utilisé pour représenter l'opération d'autocorrélation croisée normalisée, ainsi, $\text{argmax}(A_i * A_j)$ correspond à l'intervalle de temps le plus fréquemment observé entre utilisations entre l'utilisation des appareils A_i et A_j (aussi appelé intervalle de temps modal). En procédant de cette manière pour chaque appareil électrique de notre base de données, nous disposons d'une distribution décrivant sa périodicité, et de 5 distributions de probabilité décrivant l'intervalle de temps entre utilisations d'appareils pour chaque paire i, j d'appareils différents. Cette nouvelle information permet de proposer un nouveau moyen de calcul pour $P_t(A_i)$:

$$P_t(A_i) = P(A_i | H, n_{A_i}, \lambda_{i,j}), i \neq j \quad (8)$$

La règle de décision utilise les probabilités $P_t(A_i)$ de la même manière que lors de la catégorisation confirmatoire : pour chaque $P_t(A_i)$, on génère un nombre aléatoire $S \sim \mathcal{U}(0,1)$, et si $s < P(A_i | H, n_{A_i}, \lambda_{i,j})$, alors l'appareil A_i sera utilisé.

Lorsque l'on prend en compte les simplifications introduites dans le modèle de Fiske et Neuberg, en particulier concernant l'étape d'intégration sur mesure qu'ils proposent afin de pouvoir affronter une situation très différente de celles rencontrées jusqu'à présent, nous obtenons le schéma de la figure 3.6 :

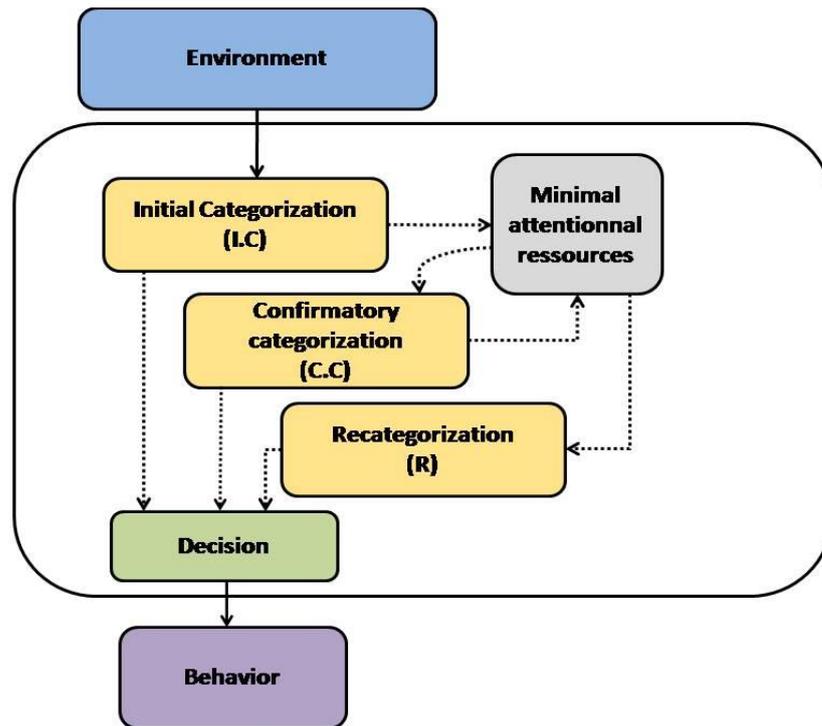


Figure 3.8 : schéma d'agent bio-inspiré basé sur le modèle de Fiske et Neuberg.

3.6 Résultats et conclusions obtenus à partir de la méthode proposée

Nous avons présenté un schéma de simulation basé sur un système d'agents afin d'améliorer les deux principaux défauts des modélisations statistiques, à savoir une dépendance trop grande aux bases de données et une difficulté à fournir une interprétation comportementale aux paramètres utilisés par le modèle. Ce système d'agents permet de reproduire une part de variance des comportements plus importante que si l'on avait uniquement recours aux méthodes d'analyse statistique. Dans ce qui suit, nous montrons comment il est possible d'exploiter notre modèle afin de générer des profils de consommation cohérents mais différents de ceux qui sont présents dans la base de données.

Pour mettre en œuvre cette méthode, nous avons utilisé une procédure de type 'leave k-out'. Parmi les 200 profils d'utilisation d'appareils électriques répartis par habitation, nous sélectionnons 20 profils pour générer automatiquement un nouveau groupe correspondant à la consommation de 5 habitations. Nous avons choisi de générer 5 habitations car c'est à cette échelle que la simulation semble la plus intéressante. En effet, lorsqu'il y a beaucoup plus d'habitations, un phénomène de moyennage se produit et les fluctuations brusques sont moins observées, ce qui rend les résultats des simulations artificiellement meilleurs. Lorsque moins d'habitations sont simulées, l'influence du comportement propre au caractère des habitants devient prépondérante; ce qui signifie que ce n'est pas "une" habitation qui doit être simulée, mais plutôt "cette habitation". Le choix

de simuler 5 habitations paraît donc justifié. Par la suite, nous évaluons la pertinence et l'originalité des profils générés en les comparant à l'ensemble des profils de la base de données. Ainsi, il est possible de savoir si les nouveaux profils de consommation horaire générés sont comparables à des comportements humains observés, mais aussi de vérifier que les profils générés sont bien différents des profils de départ.

Un exemple de profil initial est présenté figure 3.7. Dans la Figure 3.8, une comparaison est proposée entre le profil d'utilisation généré et un profil présent dans la base de données qui n'a pas été utilisé (appartenant aux profils délaissés). Le tableau récapitulatif 3.2, présente les résultats obtenus. Pour comparer deux profils d'utilisation journalière, nous avons considéré chaque profil comme un vecteur dans un espace à 24 dimensions, et nous avons calculé leur distance euclidienne.

Le tableau montre d'une part l'espérance du carré des distances (noté D) entre profils générés (notés NP) et les profils de consommation réels (notés PBD), ainsi que le carré des distances entre les profils utilisés par le programme (notés AP) et les nouveaux profils générés.

Tableau 3.2 : récapitulatif des distances entre profils de consommation

	Télévision	Chaine Hi-Fi	Lave-linge	Lave-vaisselle	Four	Friteuse
D (NP, PBD)	2.9	2.4	2.3	4.1	6.3	5.3
D (NP, AP)	4.1	4.8	5.9	9.9	10.0	8.9

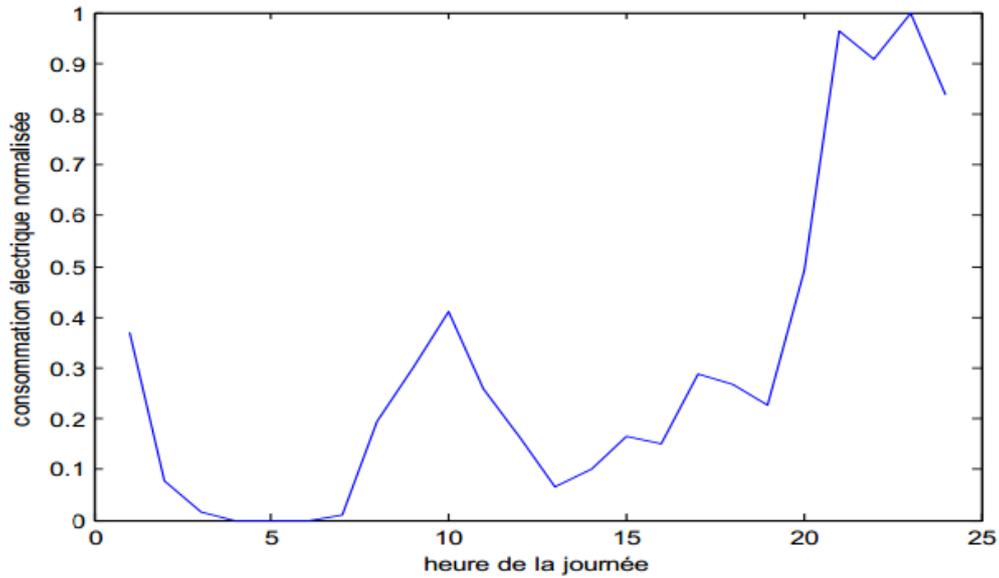


Figure 3.7 : exemplaire de courbe de consommation utilisé dans la génération de nouvelles courbes

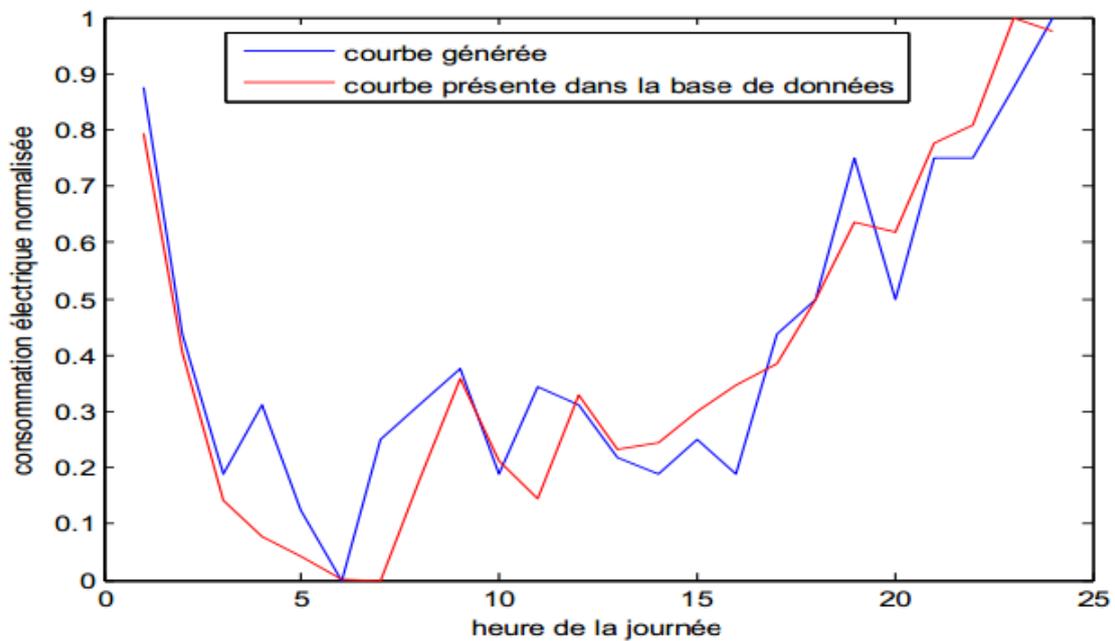


Figure 3.8 : comparaison d'une courbe d'utilisation générée à une courbe présente dans la base de données.

Le tableau 3.2 ainsi que les figures 3.7 et 3.8 montrent que les profils générés sont proches des profils réels. La possibilité de générer des profils nouveaux et réalistes suggère que le système proposé est capable de dépasser les possibilités des méthodes basées uniquement sur des PDM.

Dans le chapitre suivant, cette première version du modèle est l'objet d'améliorations rendant plus aisée la génération de profils et améliorant l'exploitation d'informations contenues dans les séries chronologiques des consommations électriques. L'interaction entre activités nécessite en particulier une étude plus poussée, notamment en établissant son lien avec des fonctions cognitives de plus haut niveau : par exemple la planification.

CHAPITRE 4 : MODELE APPROFONDI

4.1 Présentation générale de la méthode

4.1.1 Description de l'approche

Dans le chapitre précédent nous avons introduit une approche novatrice consistant à s'inspirer d'un modèle issu de la psychologie sociale afin de simuler des activités –habituelles ou non– au sein d'une habitation. Le modèle du continuum réussit à expliquer la manière dont les représentations internes aboutissant à des décisions se forment. Une des bases intéressantes du modèle provient du fait qu'il appartient à une famille de modèles psychologiques dits 'à simple flux'. Cette famille de modèles permet d'expliquer que la différence entre des actions habituelles (ou stéréotypées) et des actions planifiées réside essentiellement dans la quantité de ressources cognitives attribuées. Par ailleurs, le modèle de Fiske et Neberg nous a permis de relier besoins, perceptions et actions.

Modéliser la demande d'électricité du secteur domestique reste cependant une tâche particulièrement complexe. En effet bien que l'analyse temps-amplitude proposée précédemment se soit avérée utile pour générer de nouveaux profils de consommation plausibles du point de vue des facteurs humains, elle possède certaines limites. En premier lieu, l'analyse temps-amplitude des consommations oblige à prendre en compte simultanément le matériel utilisé (marque, modèle, performance énergétique, puissance consommée...) et les caractéristiques propres aux facteurs humains, en particulier ceux décrivant la représentation des tâches (régularité d'utilisation, habitudes et préférences).

4.1.2 Représentation par une suite de delta de Dirac

Après avoir normalisé les consommations en fonction des appareils électriques, le problème de prise en compte du matériel n'est que partiellement réglé puisqu'en cas de changements (nouvel achat, rénovation...) toute la normalisation est à recommencer. De plus, le problème concernant l'identification des paramètres décrivant le comportement et les décisions des habitants reste entier. Face à ces problèmes très différents et pourtant fortement liés, une première idée consiste à isoler deux composantes du signal, l'une décrivant le facteur humain et la prise de décision et l'autre liée au profil de consommation de l'appareil utilisé (puissance, durée typique d'utilisation...). En traitement du signal de manière simplifiée, on peut considérer deux types de mélange de signaux : le multiplexage et la convolution.

Le multiplexage consiste à faire passer sur un même canal deux informations différentes, en alternant dans le temps la transmission de chacune d'entre elles, ce n'est pas le cas ici. En effet le réseau électrique fournit continuellement l'énergie en fonction de la demande du consommateur, il n'y a aucune « alternance » dans la transmission d'énergie.

En réalité, lorsque l'on observe une suite temporelle de consommations électriques d'un appareil donné on possède une représentation du produit de convolution de deux signaux. Un premier signal représentant la prise de décision matérialisé par l'action d'allumer un

appareil et un deuxième signal constitué par le profil de consommation de l'appareil en question. Une décision portant sur la mise en marche d'un appareil est réalisée pendant un temps $\Delta t \rightarrow 0$ et existe uniquement à l'instant $T = H_i$, H_i étant l'heure à laquelle la mise en marche a commencé.

Le signal décrivant une décision peut adéquatement être représenté par une fonction delta de Dirac, l'ensemble des décisions relatives à la mise en marche d'un appareil sur une période de temps donnée, correspond donc à une suite de delta de Dirac (figure 4.1b).

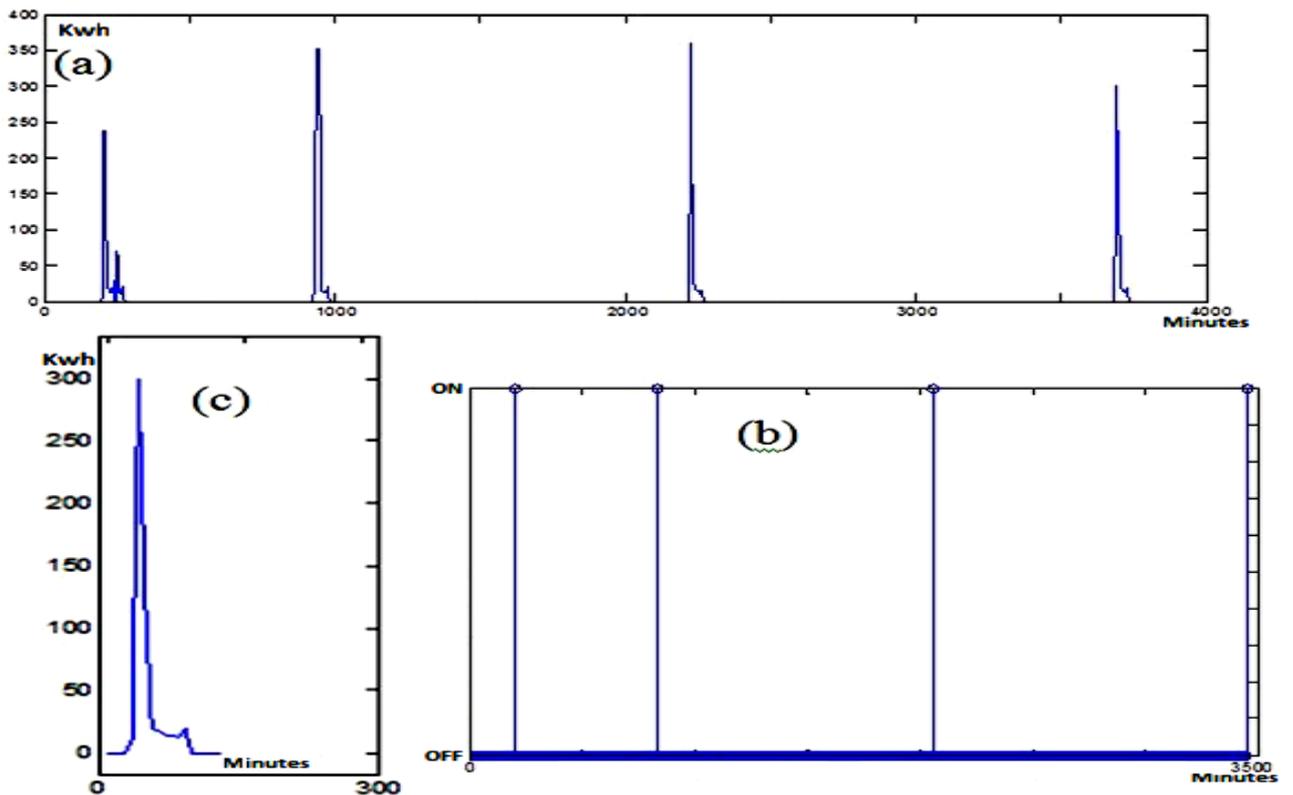


Figure 4.1: Décomposition d'une série chronologique de consommations avec ses composantes.

Le deuxième signal est complètement défini par la machine et ne dépend en aucune manière des coutumes des habitants. Ainsi isolé, le profil de consommation est une courte suite temporelle dont la variabilité due aux facteurs humains a été éliminée (figure 3c). Ce profil se maintient stable dans le temps et peut être assez facilement déduit des données constructeur (figure c).

4.1.3 Décomposition en ondelettes de Haar

Revenons à la composante correspondant aux facteurs humains, celle-ci a été isolée et représentée sous forme d'un peigne de Dirac (une suite de delta de Dirac). La suite d'intervalles de temps entre chaque Delta de Dirac constitue la série chronologique que

nous analyserons, en extrayant des caractéristiques structurelles afin de décrire les habitudes à l'aide d'un ensemble de coefficients. Pour ce faire nous avons choisi la transformée de Haar, il s'agit de l'une des plus anciennes transformées, elle a été proposée par le mathématicien hongrois Alfred Haar. Il s'agit d'une manière efficace d'analyser un signal. La transformée de Haar est calculée grâce à une multiplication matricielle qui permet d'obtenir la moyenne des valeurs de la série temporelle ainsi que le différentiel des valeurs sur plusieurs échelles de temps. De manière générale il est possible de voir cette analyse comme l'utilisation d'une moyenne glissante et de plusieurs différences glissantes afin d'analyser le niveau et les variations d'une série temporelle. Par exemple, la matrice H_4 permet grâce à une multiplication matricielle sur une fenêtre glissante contenant 4 éléments de la série chronologique d'obtenir 4 coefficients qui résument l'information contenue de manière plus parcimonieuse que les valeurs de la série temporelle (Eq. 4.1).

$$H_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \quad (\text{Eq. 4.1})$$

La transformée de Haar est la plus simple des transformées en ondelettes. Cette transformée est semblable à la transformée de Fourier. Les principales différences étant que la transformée de Fourier utilise une onde sinusoïdale au lieu d'une onde carrée et le résultat de la transformée de Fourier est une matrice de nombres complexes, alors que la transformée de Haar correspond à une matrice de nombre réels et nécessite moins d'opérations. A présent nous montrerons comment la transformée de Haar permet de caractériser des habitudes de consommation électrique d'abord en expliquant les calculs puis en donnant l'interprétation des coefficients obtenus.

Soit $s(n)$ une série chronologique représentant les intervalles de temps entre actions de mise en marche d'un appareil électrique donné. Il s'agit d'une série 1D de longueur N. La transformée Haar de $s(n)$ est notée $S_{haar}(j)$, il est possible de la décomposer et l'exprimer en termes de sa valeur absolue $|S_{haar}(j)|$, et de sa phase $S'_{haar}(j)$ voir équation 4.2.

$$S_{haar}(j) = |S_{haar}(j)| \cdot S'_{haar}(j) \quad (\text{Eq. 4.2})$$

$S'_{haar}(j)$ peut prendre deux valeurs : $S'_{haar}(j) \in \{+1, -1\}$, en effet la transformée de Haar n'a pas de composante complexe. La figure 4 montre le lien entre la structure des habitudes observées et des caractéristiques de $|S_{haar}(j)|$ et $S'_{haar}(j)$. En particulier, la séparation sur plusieurs échelles de temps permet à partir de la magnitude $|S_{haar}(j)|$ de trouver une composante dominante (1, 2 ou 3, voir figure 4.2). Si un appareil est utilisé fréquemment, ce

sera plutôt la composante 2 ou 3 qui aura une magnitude plus importante. Les composantes d'amplitude permettent de quantifier de manière globale les habitudes, c'est-à-dire si l'utilisation d'un appareil est peu ou très fréquente, mais ne permet pas de quantifier une caractéristique temporellement locale c'est-à-dire l'heure d'utilisation ou d'éventuelles alternances. Par exemple si la machine à laver est utilisée une fois par semaine, le vendredi, et que ce jour on réalise un lavage de linge blanc puis ensuite de linge de couleur, on verra uniquement deux composantes de fréquence, une hebdomadaire (dominante) et une correspondant à l'intervalle entre le lavage blanc et le lavage des couleurs. En utilisant $|S_{haar}(j)|$ Il ne sera pas possible de retrouver l'heure ou de savoir -par exemple- que c'est systématiquement le linge blanc qui est lavé d'abord.

La composante de phase permet de calculer l'heure d'utilisation, cependant l'information est plus clairement présentée sous forme d'histogrammes que sous forme d'une suite de valeurs (+1,-1....+1). En plus de quantifier les habitudes d'utilisation d'appareils, la transformée en ondelettes proposée permet de compresser des données de consommation. En effet, il est possible de laisser de côté certaines irrégularités négligeables de la série chronologique $s(n)$ de longueur N, puis de stocker les données comme une série $t(m)$ avec $m < n$.

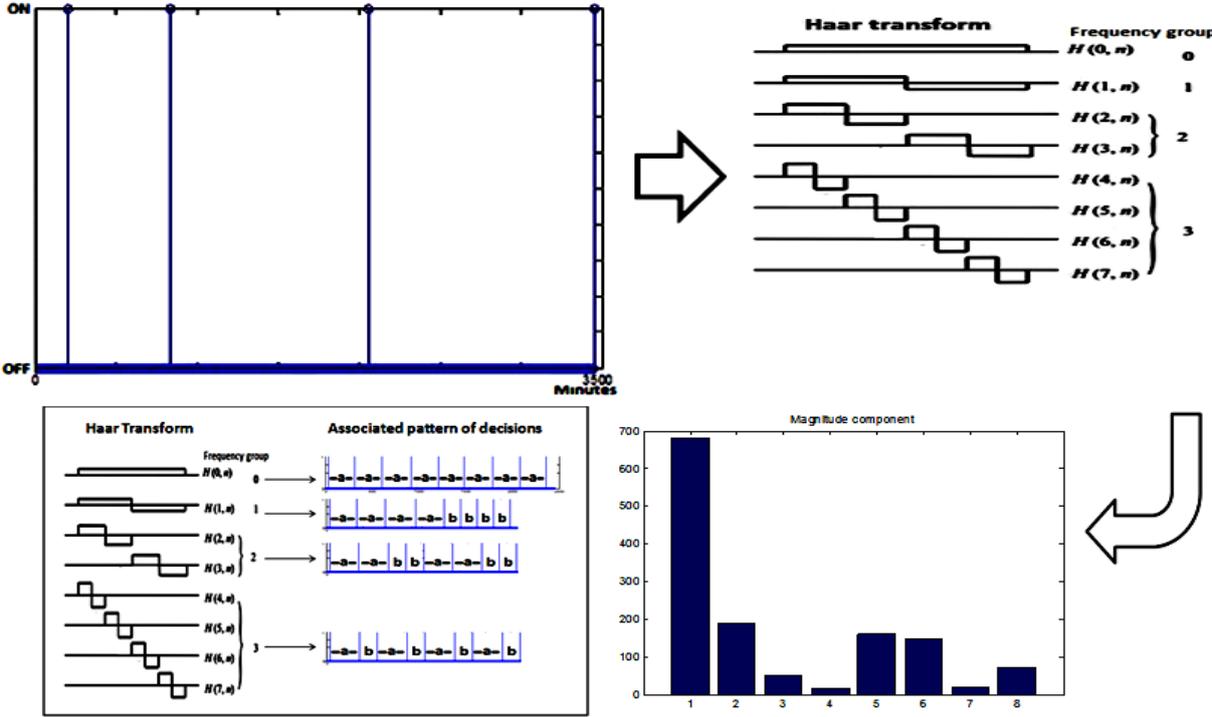


Figure 4.2 : décomposition en ondelettes de la suite chronologique de la suite de Delta Dirac.

L'utilisation de la transformée en ondelettes a permis d'attribuer un sens comportemental aux données concernant l'utilisation d'appareils électroménagers et ce grâce à une analyse

basée essentiellement sur deux dimensions : l'heure habituelle et la fréquence d'utilisation d'un appareil.

En nous basant sur l'analyse par ondelettes associée à une analyse statistique, il a été possible de proposer une représentation des données encore plus adéquate car elle permet une utilisation plus simple et directe du modèle psychologique proposé tout en augmentant la quantité d'informations disponibles par simple inspection visuelle. En effet, bien que la séparation en composantes fréquence et phase réalisée ait permis de quantifier les habitudes grâce à peu de coefficients, il reste difficile de comparer à l'œil nu les habitudes concernant l'utilisation de deux appareils pour une même habitation, ou l'utilisation d'un même appareil pour deux habitations différentes.

Afin de pallier cet inconvénient, nous proposons de décrire l'utilisation de chaque appareil par une fonction continue de distribution de probabilité de mise en route dans un espace bidimensionnel : heure d'utilisation (0...23h) et période écoulée depuis la dernière utilisation (0...24x7 h). L'obtention de cette distribution de probabilité est un peu plus laborieuse qu'une multiplication matricielle, nous décrivons celle-ci dans le détail, en 4 étapes.

4.1.4 Première étape

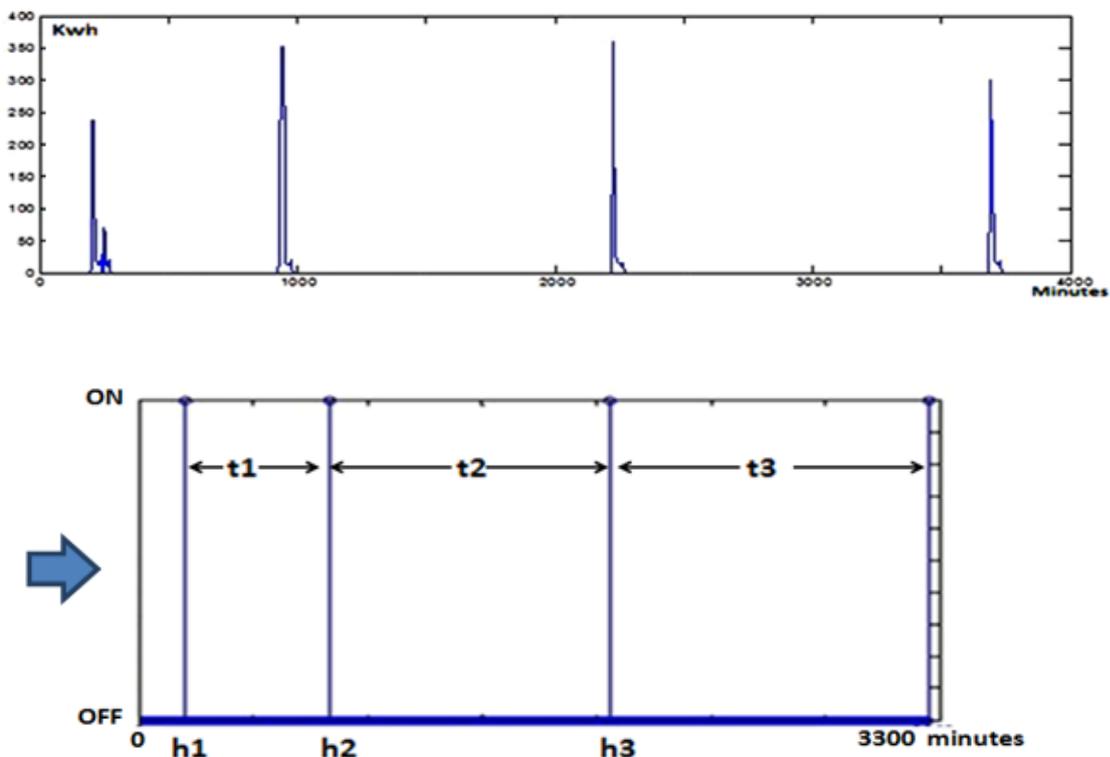


Figure 4.3: décomposition en ondelettes Haar

Il s'agit d'identifier le début de chaque période de consommation afin d'obtenir le peigne de Dirac représentant l'ensemble des prises de décision. Ensuite il faut relever chaque intervalle de temps survenu entre deux mises en route successives (suite $T(n)$ sur la figure 4.3) ainsi que l'heure de la journée correspondant auxdites heures de mise en route (suite $H(n)$ sur la figure 4.3).

4.1.5 Deuxième étape

A partir des suites $T(n)$ et $H(n)$ décrivant les mises en route d'un appareil donné on construit un ensemble de points dans un plan cartésien tel que $(T(i), H(i)) = (x_i, y_i)$. Chaque mise en route est à présent matérialisée par un carré noir de manière à rendre clairement visible dans le plan cartésien ci-dessous montré.

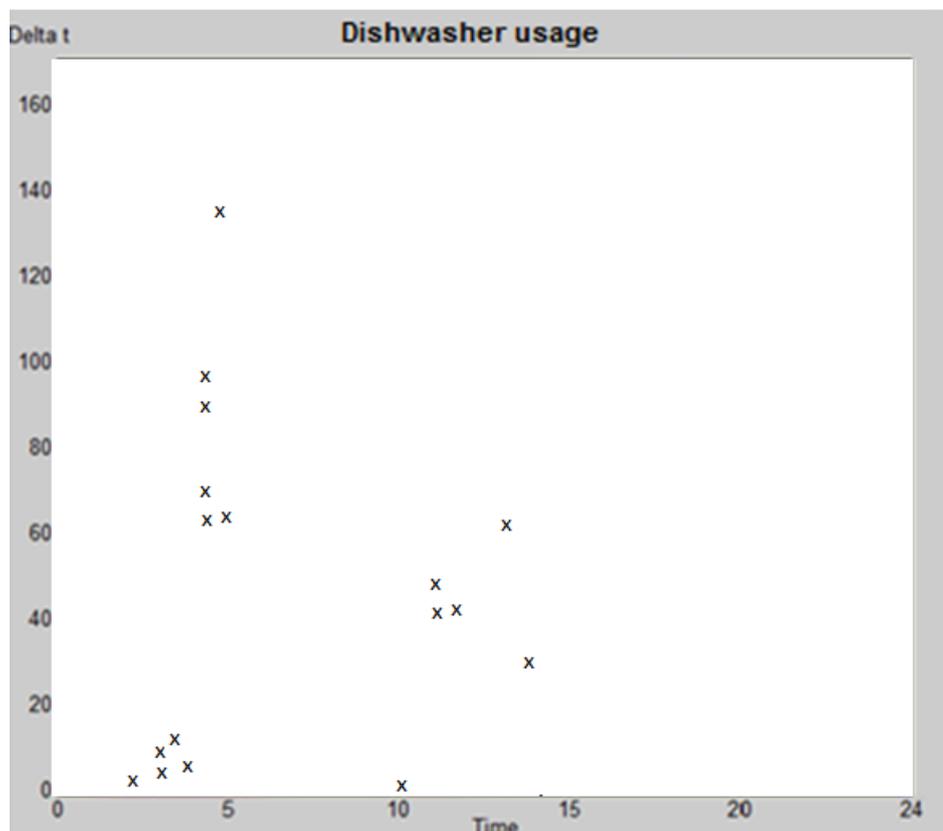


Figure 4.4 : Exemple de représentation temps vs période de chaque utilisation

A partir de cette représentation discrète des données, nous proposons d'établir une fonction de densité de probabilité continue afin d'obtenir une probabilité d'utilisation quel que soit le point de l'espace cartésien proposé.

4.1.6 Cartes de chaleur

Intuitivement, une carte de chaleur des usages peut être vue comme un histogramme bidimensionnel comptabilisant le nombre de fois qu'un foyer démarre un appareil spécifique à une heure donnée et avec un intervalle de temps entre utilisations donné. A partir du comptage, il est possible de déterminer des densités de probabilité d'usage. C'est la démarche que nous proposons de mettre en œuvre.

Afin d'obtenir cette fonction de densité de probabilité continue, nous avons utilisé une méthode non-paramétrique basée sur une estimation par noyau. Nous avons choisi d'exploiter la méthode de la fenêtre de Parzen. Cette approche a été proposée par Emanuel Parzen au début des années 1960 (Parzen 1962). Il s'agit d'une technique d'interpolation de données. Pour chaque donnée appartenant à un échantillon de valeurs, on cherche à déterminer la fonction de densité de probabilités qui est à l'origine des données collectées. Pour ce faire, une fonction noyau est superposée sur chaque donnée, de cette manière chaque observation contribue de manière continue à l'estimation de la fonction de densité. Afin de réduire la sensibilité au bruit dans les données, une largeur de fenêtre est définie.

De cette manière, plusieurs données d'un même voisinage contribuent au calcul d'une densité de probabilité locale. Afin de réaliser un choix adéquat, nous nous sommes basés sur des études précédentes notamment celle d'Archambeau et al. (Archambeau 2006 a). Lorsque les données sont peu nombreuses, ces derniers conseillent la méthode de Parzen-Rosenblatt dont nous présentons brièvement les équations pour le cas 1D pour plus de clarté.

Soit x_1, x_2, \dots, x_n un ensemble de données de la variable x , alors l'estimateur $\hat{f}(x)$ non-paramétrique par la méthode de Parzen-Rozenblatt est donné par l'équation 4.3.

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N K(x - x_i) \quad (\text{Eq 4.3})$$

N est le nombre de points pris en compte, K correspond à un noyau gaussien dont la formule est donnée par l'équation 4.4, facilement transposable en dimension 2.

$$K(w) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{w^2}{2\sigma^2}} \quad (\text{Eq 4.4})$$

Il n'est pas nécessaire de prendre en compte tous les points de données. En effet, les valeurs de la gaussienne sont proches de 0 quand on s'éloigne du point considéré. En pratique, comme cela est suggéré par les auteurs, il suffit de prendre les points qui sont à moins de 4 écart-types du point de donnée (Archambeau 2006 b). Les paramètres de largeur de fenêtre, de moyenne et de variance du noyau K ont été choisis de manière à respecter les conseils de choix proposés par les auteurs. Un écart type de 1 conduit à une densité de probabilité trop concentrée sur l'heure exacte et l'intervalle de temps, un écart type de 2 au contraire laisserait à penser qu'il y a peu de différence entre des horaires relativement distincts. Dans le cas présent, les données étant collectées toutes les 10 minutes, nous avons finalement choisi de manière empirique un écart type de 15 minutes pour les horaires et de 75 minutes

pour les intervalles de temps. En résumé, la matrice de covariance du noyau gaussien centré sur les données est $\begin{bmatrix} 15^2 & 0 \\ 0 & 75^2 \end{bmatrix}$. Des fenêtres de 7 pixels centrées sur chaque point suffisent à calculer l'estimateur de densité.

Après l'application de cette méthode sur des données de consommation telles que celles montrées dans l'illustration précédente nous obtenons un histogramme bidimensionnel. Afin de produire une carte de chaleur, nous proposons une étape de normalisation qui permet de ramener la somme des densités de probabilité à 1. Il suffit ensuite de convenir d'une échelle de couleurs pour afficher la carte. Voir figure 4.5.

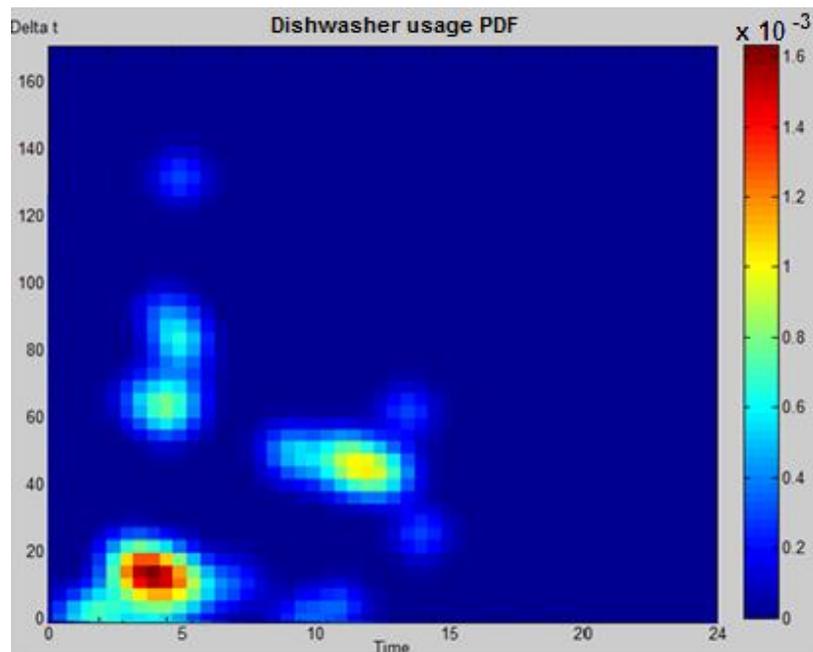


Figure 4.5 : Carte de chaleur représentant l'utilisation du lave-vaisselle

Cette représentation permet par simple inspection visuelle de connaître les habitudes concernant les usages des appareils électroménagers en termes de fréquence d'utilisation et d'heure d'utilisation. Cette description des habitudes est-elle naturelle ? Il est légitime de se demander si cette représentation bi-factorielle -donc simple- qui reprend une grande partie des informations d'une série exhaustive de données de consommations constitue un moyen adéquat du point de vue de la perception et de la représentation mentale des informations. Pour répondre à cette question, nous avons conduit une brève enquête demandant aux usagers de décrire librement leurs habitudes concernant l'utilisation de divers appareils électroménagers. Sur les 10 enquêtes complétées, l'intégralité des participants ont décrit leurs habitudes en termes de fréquence (nombre de fois par semaine que l'appareil est utilisé) et d'heure préférée d'utilisation. D'un point de vue traitement statistique du signal ou facteurs humains la représentation visuelle innovatrice des actions courantes que nous proposons semble justifiée.

4.2 Exploitation des cartes de chaleur

4.2.1 Localisation de caractéristiques des activités sur les cartes de chaleur

Une manière intéressante d'intégrer divers éléments qui doivent faire partie d'un processus de décision, est de considérer ceux-ci comme des contraintes afin de parvenir à une décision satisfaisante et réalisable. Lorsqu'il s'agit d'activités habituelles, il est aisé de les repérer et de connaître leur heure habituelle et fréquence de réalisation en lisant les coordonnées du point ayant la température la plus élevée sur la carte de chaleur. Cependant en dehors de ces zones attribuées aux habitudes, il subsiste d'autres régions qui - bien qu'éloignées des valeurs extrêmes- présentent un intérêt très particulier.

Les régions tièdes qui sont sur une zone de la carte de chaleur distincte de celles des habitudes, représentent des activités qui ont une certaine régularité mais qui dépendent non seulement de l'heure et la fréquence de réalisation mais d'autres facteurs. La régularité de ces comportements est insuffisante pour qu'il s'agisse d'habitudes, mais trop élevée pour qu'il s'agisse simplement d'un phénomène aléatoire. Dans la partie suivante, nous proposons une analyse multidisciplinaire permettant de remonter aux comportements à partir de différentes cartes de chaleur pour différentes activités d'une même habitation.

Sur la carte de chaleur le point le plus saillant correspond à une définition graphique de ce qu'est une habitude. Cependant les zones en dehors de ce point saillant contiennent des informations importantes concernant des activités moins régulières.

Nous montrerons à présent comment l'exploitation optimale des cartes de chaleur nécessite une information supplémentaire.

4.2.2 Liens entre les cartes de chaleur et le modèle de Fiske et Neuberg

Les habitants décident généralement de réaliser une action après une courte période d'évaluation du contexte et des besoins du moment. Ceci implique que la plupart des activités que nous réalisons dans un environnement connu sont essentiellement des habitudes et par conséquent sont fortement stéréotypées et font appel à des schèmes. Le niveau 2 de la catégorisation de Fiske et Neuberg prend précisément en compte des stéréotypes, c'est-à-dire des éléments de perception immédiate pour une réponse qui ne requiert pas de traitement complexe de l'information (Fiske 1990). Ramené au contexte de notre étude, les "stéréotypes" des activités correspondent aux habitudes, définies par un horaire et une fréquence particulière. C'est ce que nous proposons de déterminer dans une première étape de segmentation des cartes de chaleur. En dehors de ces habitudes, le processus de décision est plus complexe, comme l'est la catégorisation de niveau 3 de Fiske et Neuberg, qui nécessite de prendre en compte des informations complémentaires, parfois inhabituelles, et de procéder à une analyse plus fine des données disponibles. Dans le

contexte de notre étude, nous ne disposons malheureusement que de peu d'informations complémentaires. De manière générale, une personne ou un groupe de personnes peuvent être amenées à utiliser un appareil électroménager pour de nombreuses raisons personnelles ou circonstanciées dans un contexte social et familial donné, avec parfois un processus de négociation complexe. Il est impossible de considérer toutes les informations concernant l'environnement et l'ensemble des processus cognitifs qui font parties du processus.

Malgré tout, il existe d'autres informations à notre disposition, il s'agit des cartes de chaleur des autres activités. Pour le niveau 3 dit de recatégorisation de Fiske et Neuberg, nous proposons d'élaborer un processus de décision de déclenchement des activités en fonction d'une part de ces informations, et d'autre part, sur des règles simples portant soit sur les contraintes de l'installation (certaines installations ne permettent pas d'utiliser simultanément lave-linge et lave-vaisselle), soit sur une utilisation cohérente des appareils (on sèche du linge après un lavage, pas avant).

Afin de correctement définir un processus de décision orienté planification, il est nécessaire d'intégrer à la fois une analyse basée sur des caractéristiques spécifiques et une analyse décrivant les habitudes comme un tout. En ce qui concerne l'analyse centrée sur des caractéristiques locales elle prend en compte l'heure habituelle de réalisation d'une activité ainsi que le temps entre réalisations de cette même activité. L'analyse globale permet d'intégrer un début de planification en incluant des caractéristiques telles que la corrélation entre réalisation d'activités, ce qui permet de savoir dans quelle mesure le fait d'avoir réalisé l'activité A renseigne sur la probabilité de réaliser l'activité B par la suite.

Dans la partie suivante nous décrivons précisément comment les dimensions locales et globales ont été prises en compte. Nous expliquons également comment la méthode de visualisation des données par carte de chaleur permet de retrouver des régularités dans les habitudes et de détecter certaines planifications d'activités.

4.2.3 Interprétation des caractéristiques d'activités

Nous proposons à partir d'exemples simples de montrer comment lire une carte de chaleur afin de retrouver des informations concernant chaque activité domestique. Nous avons simulé plusieurs cas pédagogiques afin de montrer comment réaliser une lecture des cartes de chaleur. Cette démarche pédagogique a pour but de clarifier et représenter les différents comportements enregistrés dans les séries chronologiques de données de consommation.

-Cas 1:

Le cas le plus simple montré à la figure 4.6 montre une activité domestique réalisée régulièrement. De légères variations aléatoires dans la période entre répétitions et dans l'horaire sont à l'origine de la répartition. Pour des raisons de réalisme et de simplicité nous avons choisi de représenter ce cas par une distribution Gaussienne bidimensionnelle. En superposant plusieurs cartes de chaleur d'activités, on peut en apprendre davantage sur la manière dont chaque habitation gère les activités domestiques. Si la plupart des cartes sont constituées comme celle de la figure 4.7, en fonction de la variance dans les habitudes et du délai entre activités, le profil aura tendance à être constitué de pics très accusés ou au contraire comportera peu de différences de consommation au fil de la journée. La difficulté dans la prédiction du profil de consommation global résulte plus de facteurs aléatoires difficilement modélisables que de variables comportementales.

Dans ce cas, une analyse statistique classique effectuée sur les cartes de chaleurs constitue une solution efficace pour simuler ou prédire un profil de consommation.



Figure 4.6 : Cas 1 activité résultant d'une habitude régulière.

-Cas 2:

Dans ce cas, la représentation de la figure 4.7 montre une gaussienne 2D dont l'axe principal comporte une inclinaison. Cette inclinaison montre une claire corrélation entre l'heure à laquelle est effectuée une certaine tâche et la fréquence de répétition de celle-ci. Par exemple, lorsque le lave-vaisselle est mis en route avant un repas, il effectue souvent un cycle rapide à charge incomplète afin de disposer rapidement des couverts et de quelques assiettes. Ce lavage à une heure précoce de peu d'éléments signifiera donc un nouveau lavage sous peu. Alors qu'un lavage plus tardif qui aurait donc été effectué avec plus d'éléments aurait signifié un délai entre lavages nettement plus important. L'interaction entre plusieurs activités de ce genre est complexe, cependant elle dépend davantage de

facteurs humains liés à la prise de décision. Cette particularité permet de tirer profit de la méthode que nous proposons et expliquerons plus bas avec des données réelles.

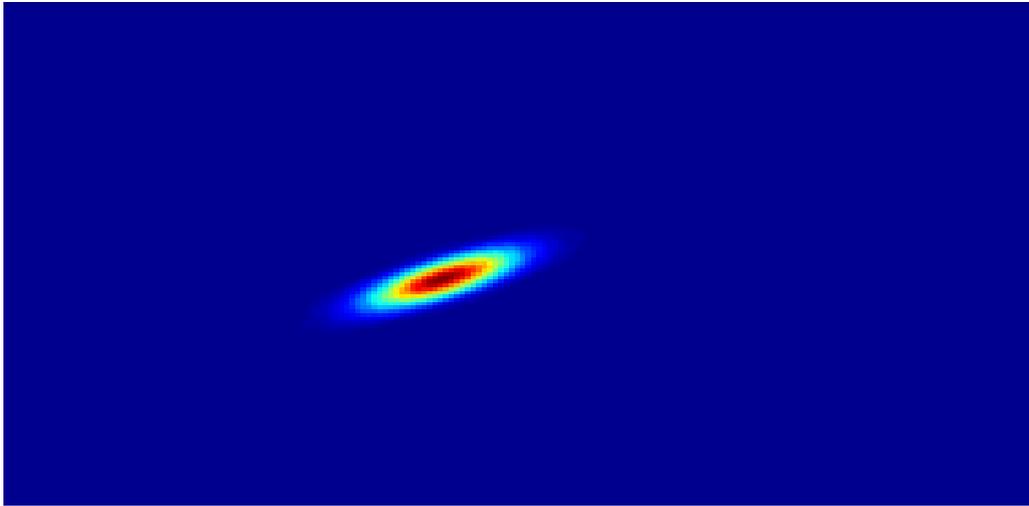


Figure 4.7 : Cas 2, activité dont la période de répétition dépend de l'heure de réalisation

Cas 3:

Dans la figure 4.8 la réalisation d'une activité n'est pas représentée par une gaussienne. En effet on observe plusieurs pics. De plus les pics ont des amplitudes bien différentes, ce qui suggère qu'en plus d'activités habituelles et du bruit il y a un autre facteur nettement plus dominant dans les mesures.

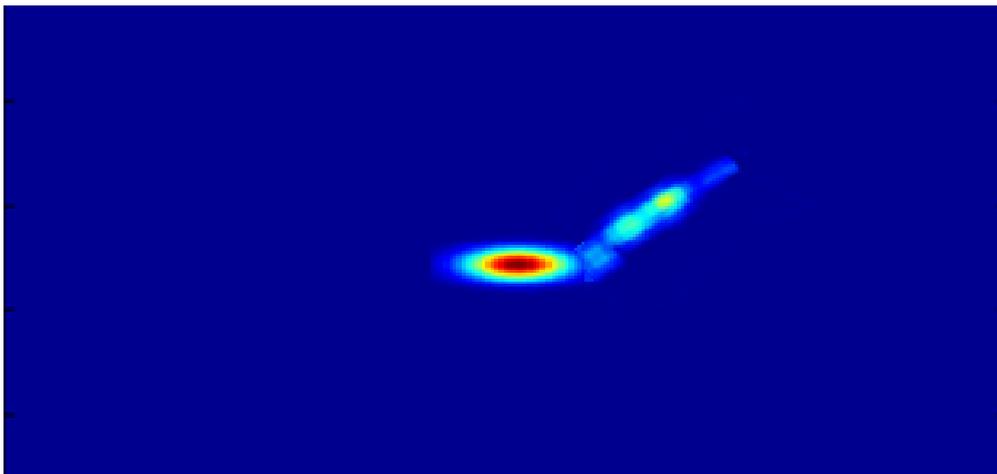


Figure 4.8 : Cas 3, représentation d'un comportement plus complexe.

Lors de l'analyse de la figure 4.7, la corrélation intra-activité (lien entre les composantes Delta t et heure de réalisation) a permis d'expliquer l'inclinaison de la gaussienne représentant la réalisation d'une activité. Dans le cas présent, une prise en compte plus fine

des facteurs humains est nécessaire afin de mieux savoir à quoi correspond le nuage "tiède" de la figure 4.8. Après analyse de l'enquête portant sur les habitudes de consommation, il a été manifeste que certaines activités dépendent d'autres activités. Ainsi le séchage du linge doit être réalisé après le lavage de celui-ci.

De plus, dans certains cas, des activités sont temporellement corrélées parce que cela correspond à une période de disponibilité dans la journée ou la semaine. Cet ordonnancement de tâches prévues sera appelé ici planification. En plus d'un lien parfois causal entre tâches, il existe un ordonnancement temporel de celles-ci. Puisqu'il existe des corrélations temporelles et que l'axe horizontal de la carte de chaleur permet de connaître l'heure de réalisation des activités, en superposant les cartes de chaleur de deux tâches il doit être possible de déceler une planification. Nous postulons donc que le nuage contenant la région "chaude" contient de l'information en lien avec les habitudes et l'amas contenant une région "tiède" représente une part de planification. Afin de continuer avec l'analyse comportementale par le biais de cartes de chaleurs, nous proposons une nouvelle démarche permettant de bien étudier le comportement en tant que composante multi-composentielle des facteurs humains. Dans le cas plus complexe de l'interaction entre activités, nous utiliserons un jeu de données réelles afin de conduire une analyse qui soit représentative du quotidien.

4.2.4 Extraction des paramètres comportementaux

4.2.4.1 Principe général

La carte de chaleur permet une représentation des activités bien plus précise qu'avec l'utilisation des méthodes statistiques classiques. En effet, que ce soit en modélisant avec des chaînes de Markov ou en représentant les habitudes avec une distribution classique (une distribution de Poisson par exemple), il est très compliqué de faire ressortir deux modes différents d'utilisation d'un appareil. La figure 4.9 montre une représentation de l'utilisation du lave-vaisselle à partir de données réelles. Il est aisé de déterminer les principaux modes. On constate en effet qu'en moyenne le lave-vaisselle est utilisé une fois par jour et pendant les heures "creuses", c'est-à-dire très tard la nuit. Cette première zone peut être délimitée par une segmentation obtenue par un seuillage. Par ailleurs, une importante zone "tiède" montre que le lave-vaisselle est parfois utilisé plusieurs fois par jour, en particulier après le goûter et le dîner. Afin d'exploiter l'information contenue dans la région "tiède" de cette carte de chaleur, on peut chercher à la mettre en relation avec une carte de chaleur d'une autre activité, dans le but de déceler d'éventuelles interactions.

Les différentes étapes d'analyse sont présentées dans les sections suivantes.

4.2.4.2 Segmentation

Les pixels de la carte de chaleur sont liés à des densités de probabilité. Les valeurs les plus fortes sont potentiellement associées aux habitudes. Dans un premier temps, nous proposons de déterminer le maximum et sa position sur la carte. Les pixels proches d'une zone chaude correspondent en général à la même habitude, qui s'étend un peu sur l'axe des heures et l'axe des intervalles de temps, du fait de l'utilisation de noyaux gaussiens et des tailles de fenêtre. Nous proposons de regrouper tous les pixels autour du maximum selon une méthode simple qui suffira en première approximation : nous agrégeons tous les pixels situés à une distance inférieure à un seuil. Ce seuil est logiquement corrélé à la taille du noyau gaussien. Dans le pire des cas, nos calculs de sommes de gaussiennes correspondent à une probabilité d'erreur de type I à 1% lorsque 2 points de données sont situés à 2,5 écarts-types. Les calculs ont alors un impact sur un voisinage de 5 écarts-types. Nous proposons donc de regrouper les pixels situés au plus loin à 5 écarts-types du maximum. Le groupe de pixels ainsi sélectionné est associé à une habitude seulement si la densité de probabilité associée est conséquente. Nous proposons en pratique un seuil à 0,3. Si ce seuil est dépassé, nous marquons tous les pixels pour les associer à l'habitude H1, puis, nous recommençons le même processus en cherchant le maximum suivant non marqué. Eventuellement, si la densité de probabilité du nouveau groupe de pixels est supérieure au seuil, une deuxième habitude H2 est identifiée et le processus est itéré une dernière fois pour la détection d'une éventuelle 3^{ème} habitude. Ce choix permet de repérer des habitudes quotidiennes comme la préparation des aliments avec un four juste avant les heures de repas.

La segmentation aboutit donc à une séparation en 3 types de région : une région chaude correspondant à une habitude, une zone froide qui ne correspond que peu ou même pas du tout à des activités, et une ou des régions tièdes regroupant des actions qui présentent une régularité insuffisante pour être considérées comme une habitude. Graphiquement, nous proposons dans un premier temps de séparer la zone des habitudes du reste de la carte de chaleur comme cela est proposé figure 4-9. Il reste à interpréter les pixels de la zone tiède.

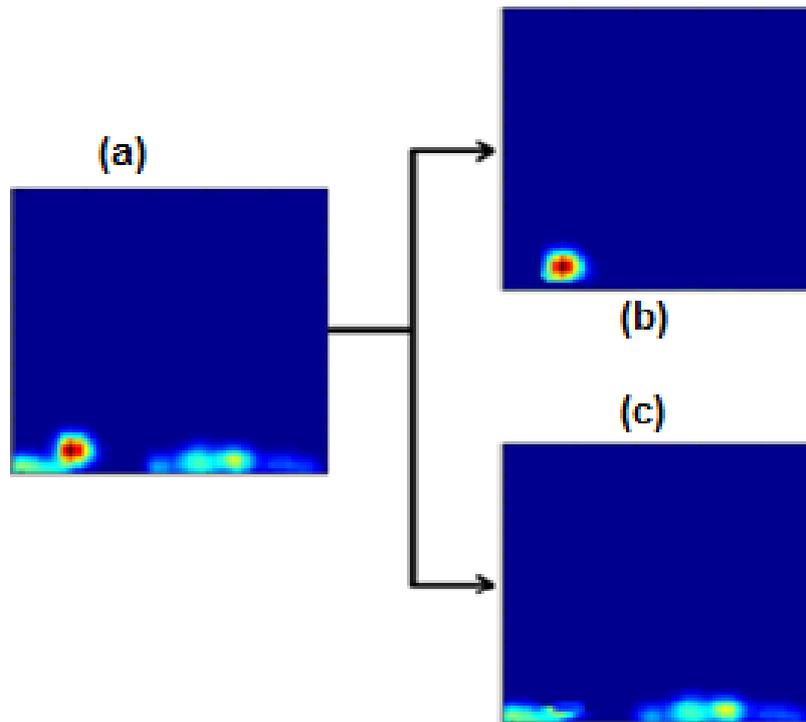


Figure 4.9 : Séparation en régions chaudes(b) et tièdes(c) de la carte de chaleur initiale (a).

4.2.4.3 Mise en relation des régions segmentées de plusieurs cartes de chaleur

Il est tentant de comparer toutes les cartes entre elles. L'éventuelle corrélation de plusieurs habitudes pourrait être détectée. Cela n'est toutefois pas nécessaire pour modéliser et simuler les comportements habituels, car ils ont déjà été traités au niveau 2 et d'éventuelles corrélations peuvent effectivement apparaître. Il reste à étudier les corrélations pour les autres zones de points après l'étape de segmentation.

Il pourrait théoriquement y avoir des corrélations liées à des interactions mal comprises entre activités. Pour simplifier le problème, comme expliqué section 4.2.2, nous posons au contraire l'hypothèse que ces interactions sont soit évidentes et peuvent être modélisées a priori grâce à des règles comportementales, soit correspondent à une association entre une activité ménagère et une activité récréative. Certaines tâches ménagères requièrent peu de ressources attentionnelles ou pendant une durée limitée (*e.g.* démarrer le lave-vaisselle). Il est donc fréquent que des activités récréatives soient réalisées en même temps (*e.g.* regarder la télévision). Les corrélations pertinentes du point de vue comportemental sont donc contraintes à différentes échelles de temps. Le critère naturel suggéré par les théories développées en sciences cognitives (Marendaz 2004), notamment en vision et reconnaissance des formes, nous ont conduit à exploiter la dichotomie global/local afin de prendre en compte simultanément des contraintes portant sur les liens entre activités et les horaires de réalisation de celles-ci. Concrètement, nous utilisons deux contraintes :

1- Une contrainte globale portant sur les liens fonctionnels entre différentes activités. Il s'agit typiquement du cas où une activité requiert en général la précédence d'une autre sans nécessairement être associée à un horaire précis. Il n'est donc pas possible d'attribuer une position fixe dans l'espace des tâches.

2- Une contrainte locale portant sur la temporalité de l'activité. Pour cela, nous proposons de chercher la corrélation uniquement avec les points de la carte de chaleur correspondant à une habitude de la première activité.

En pratique, à partir de la base de données disponible et des cartes de chaleur calculées, nous ²proposons d'appliquer la règle suivante : l'utilisation de la machine à laver dans un mode habituel conduit généralement à l'utilisation du sèche-linge. D'un point de vue formel, cela peut s'écrire de la manière suivante :

Soit $P(A_1(t))$ la probabilité d'avoir l'activité associée à l'utilisation de la machine à laver à la date t et $P(A_2(t))$ la probabilité d'avoir l'activité associée à l'utilisation du sèche-linge à la même date, alors :

habitude ($A_1(t)$) et non habitude ($A_2(t)$) $\Rightarrow P (A_1(t) | A_2(t)) = 1$

La difficulté est de déterminer la probabilité conjointe des deux activités, les horaires et les fréquences associées. Idéalement, il faudrait revenir aux données et calculer des probabilités conditionnelles. Toutefois, cela paraît difficile car les horaires fluctuent fortement. Nous proposons de poursuivre l'analyse des cartes de chaleur. La mise en relation de plusieurs cartes de chaleur peut mettre en évidence un phénomène d'attraction des zones tièdes d'une activité A_1 par la zone chaude d'une activité A_2 qui lui est reliée.

Pour identifier ces zones tièdes corrélées, une première idée est d'appliquer un masque sur la 2^{ème} carte avec les points de la 1^{ère} carte correspondant aux habitudes. Une autre idée est de réaliser une opération de logique floue avec un "et bruité". Ceci peut être mis en œuvre en procédant au calcul du produit de Hadamard sur les cartes de chaleur en question et de ne retenir que les points de valeur relativement forte, donc supérieure à un seuil donné. Ensuite, on utilise les formules de probabilités conditionnelles : la probabilité de déclencher une activité A_2 sachant l'activité A_1 est estimée en calculant $P(A_2)/P(A_1)$. Cette méthode est illustrée figure 4.10.

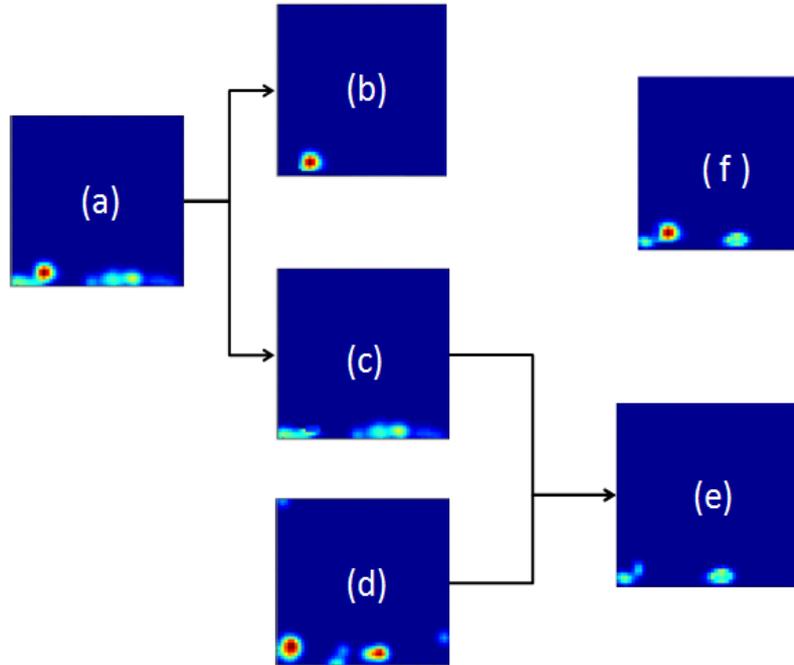


Figure 4.10 : Cartes de chaleur représentant différentes étapes du traitement. (a) : Données initiales d'utilisation du lave-vaisselle. (b) : Extraction des habitudes. (c) : Autres régions. (d) : Données initiales d'utilisation de la télévision.(e) : Données du lave-vaisselle corrélées aux habitudes d'utilisation de la télévision. (f) : Carte de chaleur représentant les habitudes et les utilisations corrélées avec l'utilisation du lave-vaisselle.

4.3 Simulation

4.3.1 Mise en œuvre

Pour effectuer une simulation, nous définissons dans un premier temps une liste d'activités dont les caractéristiques ont pu être déterminées après analyse des cartes de chaleur selon la méthode décrite précédemment. Les activités domestiques sont classées en deux groupes : un groupe avec des activités ménagères (lave-vaisselle, lave-linge, etc.), l'autre groupe est lié à des activités récréatives (télévision et utilisation de l'ordinateur). L'échelle temporelle choisie est la même que celle des données collectées, ici 10 minutes. De manière itérative, un processus de décision est mis en œuvre au sein d'un agent cognitif pour déterminer le démarrage des activités. Ce processus est basé sur différents niveaux de traitement et de catégorisation qui s'inspirent du schéma de Fiske et Neuberg (voir figure 3.8 reproduite ci-dessous).

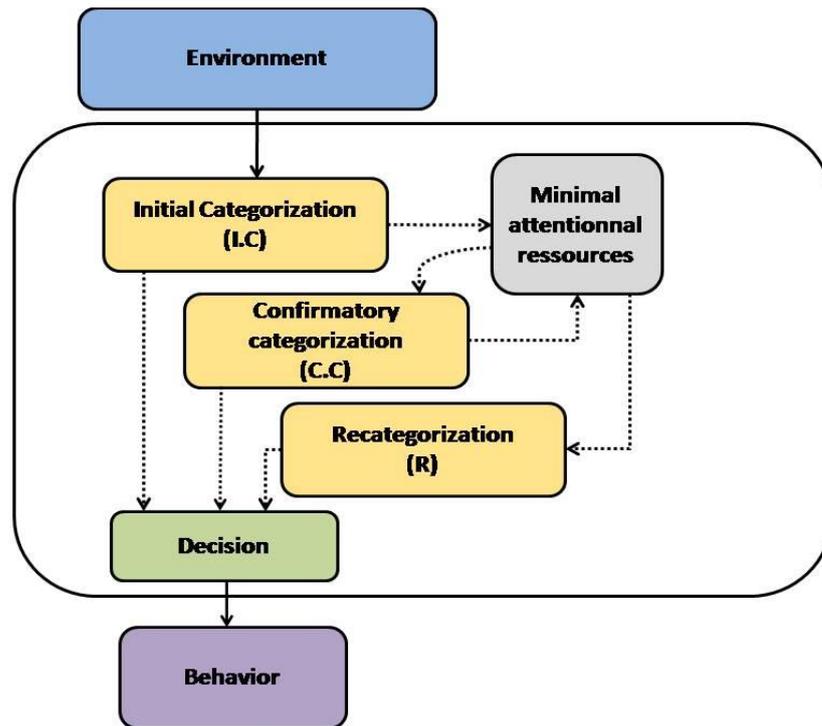


Figure 3.8 (bis) : schéma d'agent bio-inspiré basé sur le modèle de Fiske et Neuberg.

Pour exploiter les cartes de chaleur, nous proposons de déterminer la CDF (*cumulative density function*). Elle s'obtient en calculant les cumuls de densité de probabilité sur les deux axes, heure et DT.

La première étape consiste à s'assurer que l'agent a les ressources attentionnelles requises pour démarrer une activité. Cette étape paraît judicieuse car il arrive souvent que la maison soit inoccupée ou que les personnes qui y vivent ont une activité qui mobilise leur attention et qui est sans lien avec la consommation énergétique. Cela permet également d'effectuer une distinction entre les jours de la semaine. Pour prendre en compte cette éventualité, nous proposons de procéder à un tirage aléatoire (loi uniforme entre 0 et 1) basé sur la probabilité de démarrer une activité quelconque dans un contexte (heure, jour) donné. Le seuil à atteindre est déterminé de la manière suivante : pour chacun des deux groupes d'activité, on reprend les données brutes et on calcule pour chaque jour et pour chaque créneau horaire la probabilité qu'une activité soit démarrée. En pratique, si le tirage aléatoire est inférieur au seuil, aucune activité n'est démarrée, sinon, on passe à l'étape suivante.

La première étape peut être vue comme une catégorisation initiale. Le jour et l'heure suggère le démarrage d'une activité. Cependant, comme dans le schéma de Fiske et Neuberg, nous proposons à l'étape 2 une catégorisation confirmatoire. Celle-ci consiste, pour chaque activité, à prendre en compte l'heure et le temps écoulé depuis la dernière réalisation de l'activité pour estimer la probabilité de démarrage de celle-ci en exploitant les

cartes de CDF obtenues après la première segmentation, afin de ne prendre en compte que les habitudes. Le principe est de considérer le cumul des densités de probabilités (valeurs de la carte CDF) pour l'heure en cours et le DT qui lui correspond. La valeur est en général inférieure à 1 car elle ne recouvre qu'une partie de l'espace des probabilités. Un tirage aléatoire d'une loi uniforme entre 0 et 1 est effectué et comparé à la valeur précédente pour déterminer si cette heure est sélectionnée, ou si on ne fait rien pour cette heure là. Remarque : pour le cas particulier où aucune habitude n'a été détectée, nous prenons en compte la carte de chaleur de départ, avant segmentation.

Dans une troisième étape, qui correspond à la recatégorisation du schéma de Fiske et Neuberg, le processus de décision prend en compte les corrélations. Pour cela, nous proposons un nouveau tirage aléatoire et un nouveau seuil pour chaque activité, qui est déterminé par la probabilité conditionnelle vue au 4.2.4.3, si elle existe. C'est à ce niveau que les résultats de la segmentation et de la recherche de corrélations sont exploités. Cela ne concerne que les activités ayant un lien établi avec une activité qui a démarré à l'étape précédente.

4.3.2 Résultats et étude comparative

Nous avons utilisé la base de données ENERTECH pour tester la méthode de simulation. Les données de consommation ont été échantillonnées toutes les 10 minutes pour différents appareils électroménagers : machine-à-laver la vaisselle, machine-à-laver le linge, four, ordinateur et télévision. Les données ont été collectées à partir de 30 habitations pendant une période d'un an. Pour les besoins de l'évaluation, 20 jeux de données ont été exploités pour l'apprentissage et les 10 jeux restants ont été retenus pour comparer les données simulées avec les données réelles.

Différentes métriques ont été proposées dans la littérature pour évaluer la qualité des résultats de simulation (Page 2008). Les courbes moyennes des consommations prédites sont généralement comparées aux courbes moyennes des consommations observées. Cependant, cette méthode ne permet pas toujours d'obtenir des résultats précis. En effet, des composantes à haute fréquence, notamment des pics transitoires de consommation sont parfois manquants. Des métriques standard pour la comparaison de signaux, telles que *Peak Signal to Noise Ratio*, ou PSNR (équation 4.), ne peuvent pas être exploitées directement parce qu'elles ne prennent pas en compte la continuité jour/nuit. Par exemple, avec PSNR, une courbe de consommation qui a un maximum à 23h59 est considérée comme significativement différente d'une courbe qui aurait un maximum à 0h01. D'autres techniques qui permettent une analyse plus fine, par exemple basée sur les ondelettes (Ferreri 2014a), sont plus précises mais plus complexes à mettre en œuvre.

La méthode retenue ici consiste dans un premier temps à simuler les consommations comme indiqué dans la section précédente sur une période de 6 mois (mars jusqu'à août), puis à calculer la carte de chaleur. Ensuite, nous procédons à une comparaison des cartes de chaleur par la méthode PSNR pondérée (Weighted PSNR) qui prend en compte les similarités dans les habitudes à l'aide d'une matrice de pondération (voir figure 4.11 et les équations (5), (6), (7), (8)).

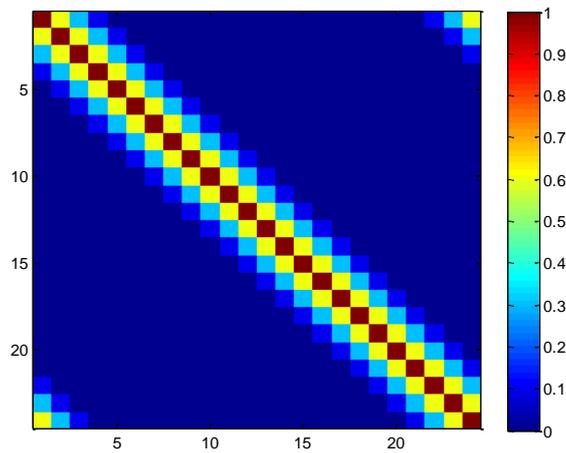


Figure 4.11: Matrice de pondération du WPSNR.

Soit S une matrice correspondant à une carte de chaleur de n lignes et m colonnes obtenues par simulation et M une carte de chaleur de référence obtenue à partir des mesures.

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=0}^n ((S(i, j) - M(i, j))^2$$

$$PSNR = 10 \cdot \log\left(\frac{Maxrange^2}{MSE}\right)$$

Soit $Diff = S - M$,

$$MSWE = \frac{1}{mn} \sum_{i=1}^n \{Diff[i] \cdot W \cdot Diff[i]^T\}$$

$$WPSNR = 10 \cdot \log\left(\frac{Maxrange^2}{MSWE}\right)$$

Le WPSNR fournit des résultats plus fiables que le PSNR et peut être calculé rapidement car la matrice de pondération est parcimonieuse. Les résultats obtenus sont présentés figure 4.12 pour ce qui concerne la simulation d'une carte de chaleur et tableau 4.1 pour l'évaluation des performances. L'estimateur de ressemblance des consommations proposé fournit un résultat relativement bon. On considère en général qu'une valeur supérieure à 30 dB caractérise une forte ressemblance du signal. Il convient cependant de rester prudent sur

l'interprétation des résultats, nous sommes partis de l'hypothèse que les cartes de chaleur étaient représentatives des comportements d'usage et qu'il était légitime de les coder de cette façon.

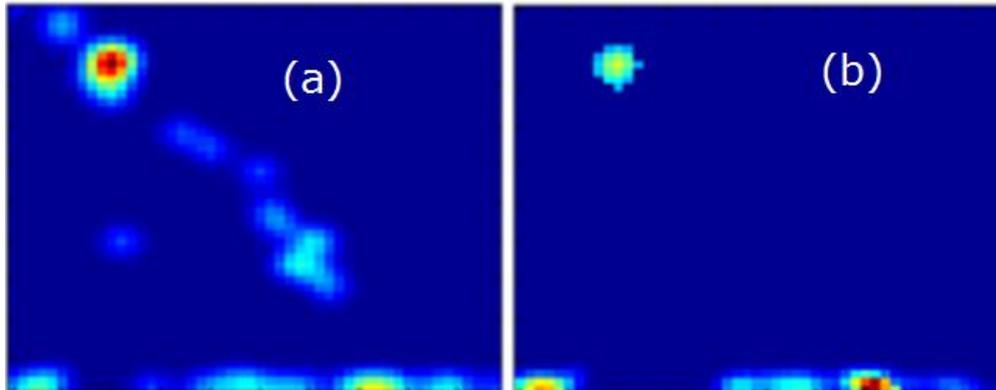


Figure 4.12: Résultats de la simulation pour l'usage de l'ordinateur. La carte de chaleur associée à la simulation (a) est semblable à celle obtenue à partir des données Enertech (b). La valeur de WPSNR est 38.56WdB.

Tableau 4.1 : évaluation des performances

	Appareils électroménagers			
	<i>Lave-vaisselle</i>	<i>Lave-linge</i>	<i>Télévision</i>	<i>Ordinateur</i>
WPSNR	40 WdB	37WdB	35WdB	34WdB

CONCLUSION GENERALE

Nous avons proposé une nouvelle méthode pour modéliser la consommation des appareils électriques d'habitations. Elle est basée d'une part sur un processus de décision inspiré d'un modèle psychologique et d'autre part sur l'exploitation de cartes de chaleur caractérisant les données de consommation. A partir de cette modélisation, nous avons mis en œuvre un processus de simulation permettant de reproduire des comportements typiques de consommation.

Les principales contributions de nos travaux de recherche sont les suivantes :

- Premièrement, les cartes de chaleur temps/période constituent une représentation intéressante alternative à la représentation temps/amplitude couramment utilisée. Elles sont obtenues par un processus d'interpolation et de votes dans un espace discret (fenêtres de Parzen). Visuellement, elles permettent de repérer et de quantifier rapidement les habitudes, qui correspondent à des points chauds de la carte. Ces points chauds sont détectés grâce à une méthode de segmentation. Des zones tièdes sont également exploitées pour la corrélation entre activités.
- Deuxièmement, le processus de décision intègre plusieurs niveaux d'analyse qui ont une interprétation comportementale. Le premier niveau consiste à démarrer une activité en fonction des habitudes connues. Le deuxième niveau est caractérisé par un processus plus complexe, dit de "planification" qui tient compte des liens entre activités. Cette interprétation n'existe pas dans les autres méthodes basées sur des analyses statistiques. Ces dernières fournissent en effet de très bons résultats en termes de moyennes et d'écart-types, mais avec parfois des incohérences au niveau comportemental, comme par exemple la préparation d'un même repas effectuée deux fois de suite.
- Troisièmement, les résultats obtenus en simulation sont proches des données observées. Cette étude gagnerait à être complétée avec une base de données plus conséquente, comportant en plus des consommations, des informations issues d'enquêtes comportementales sur les usages.
- Enfin, on peut noter également la réalisation d'un système de mesures temps réel de la consommation électrique (voir annexe). Ce système a été testé avec succès, mais nous n'avons pu exploiter les données collectées qui ne constituaient pas un échantillon assez représentatif.

Il convient cependant de rappeler les limites de notre approche et de nos résultats. Tout d'abord, la base de données que nous avons exploitée reste très insuffisante en regard du nombre de variables du problème. Comme nous l'avons indiqué, les profils de consommation dépendent de nombreux facteurs environnementaux, socioprofessionnels, en lien avec l'équipement, la situation familiale, etc. Les courbes de consommation fluctuent en fonction de tous ces paramètres. Il faudrait donc enrichir les bases de données avec un très grand nombre d'observations complémentaires.

Une autre limite de notre approche réside dans la relative simplicité du modèle psychologique de Fiske et Neuberg relativement à la complexité du processus de décision humain. De plus, notre méthode repose sur une interprétation computationnelle de ce modèle et comporte des seuils empiriques.

Perspectives :

Au-delà de la qualité des résultats, l'exploitation d'une méthode centrée sur un modèle psychologique est intéressante car elle offre de multiples perspectives. En effet, la connaissance extraite des bases de données n'est pas la seule source d'information, nous disposons également de celle décrivant le processus de décision. De cette manière, en plus d'une moindre dépendance aux bases de données, il devient possible de simuler des scénarii portant sur des profils d'habitants un peu différents ou sur certaines évolutions du matériel ou des habitudes, ce qui n'était pas possible avec les approches classiques. Prenons quelques exemples :

- Si la population cible a des habitudes différentes concernant la prise des repas ou l'usage de certains appareils, il est possible d'anticiper le changement de profil de consommation en changeant uniquement le processus de décision concernant les habitudes. Cette anticipation serait approximative, mais offrirait néanmoins une information exploitable pour une aide à la décision.
- De même, si les usages évoluent et qu'un appareil est utilisé de plus en plus fréquemment en association avec un autre appareil, il est possible d'anticiper le profil de consommation attendu en introduisant un lien temporel et en modifiant le processus de décision au niveau de la recatégorisation (activité déclenchée en fonction d'une probabilité conditionnelle).

Idéalement, il faudrait pouvoir disposer d'un processus de décision paramétré uniquement par le profil psychologique des usagers, par exemple défini par les 5 dimensions de personnalité, ouverture à l'expérience, caractère consciencieux, extraversion, agréabilité, névrosisme (Plaisant 2010), ainsi que des variables de contexte pour caractériser, notamment, l'environnement physique, le contexte socioprofessionnel et l'équipement de la maison. En effet, toutes choses étant égales par ailleurs, le comportement humain dépend essentiellement de la personnalité des usagers. Autrement dit, de façon théorique, si on connaissait le contexte général de l'usage et la personnalité attendue des usagers, il serait possible de prédire le profil de consommation. Pour le bureau d'études en charge de prédire la consommation électrique, l'échelle individuelle n'est pas forcément utile, mais il est possible de relever les tendances et d'extrapoler la somme des consommations au niveau d'un groupe d'habitations ou d'un quartier. Pour y parvenir, il faudrait dans un premier temps procéder à une enquête approfondie sur le profil de personnalité des usagers et établir les liens avec le processus de décision et les profils de consommation, en tenant compte de tous les autres facteurs. Schématiquement, cela revient à inverser partiellement le modèle pour trouver une estimation des dimensions psychologiques. Avec un processus

de décision comportant de tels paramètres, dans une certaine mesure, on pourrait prédire l'évolution des habitudes et des usages en fonction des actions entreprises au niveau de la communication, du conseil ou de l'introduction de nouveaux équipements. Il faut cependant émettre quelques réserves sur une telle approche, un individu ne se résumant pas à 5 dimensions psychologiques. De plus, il paraît difficile de constituer une base de données assez large pour prendre en compte la diversité des contextes et des usagers. Ces considérations offrent un vaste champ de recherche en perspectives à nos travaux.

REFERENCES

- (Aarts 2000) Aarts H., & Dijksterhuis A. (2000). Habits as knowledge structures: automaticity in goal-directed behavior. *Journal of personality and social psychology*, 78(1), 53.
- (Abrams 2006) Abrams S., Ploix S., Pesty S., and Jacomino M. (2006) A multi-agent design for a home automation system dedicated to power management.
- (Alexandre 1990) Alexandre A. (1990) Une modélisation fonctionnelle du cortex: la colonne corticale: aspects visuels et moteurs (Doctoral dissertation, Nancy 1).
- (Allerding 2012) Allerding F., Premm M., Shukla P., and Schmeck H. (2012). Electrical load management in smart homes using evolutionary algorithms. *Evolutionary Computation in Combinatorial Optimization*, pages 99–110.
- (Amouroux 2013) Amouroux E., Huraux T., Sempé F., Sabouret N., & Haradji Y. (2013). Simulating Human Activities to Investigate Household Energy Consumption. In *ICAART (2)* (pp. 71-80).
- (Archambeau 2006 a) Archambeau C., Valle M., Assenza A., & Verleysen M. (2006, May). Assessment of probability density estimation methods: Parzen window and finite Gaussian mixtures. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* (pp. 4-pp) IEEE.
- (Archambeau 2006 a) Archambeau C., Valle M., Assenza A., & Verleysen M. (2006, May). May). Assessment of probability density estimation methods: Parzen window and finite Gaussian mixtures. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* (pp. 4-pp) IEEE.
- (Aydinalp 2003) Aydinalp M., Ugursal V. I., & Fung A. S. (2003). Modelling of residential energy consumption at the national level. *International Journal of Energy Research*, 27(4), 441-453.

- (Aydinalp 2008) Aydinalp-Koksal M., & Ugursal V. I. (2008). Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy*, 85(4), 271-296.
- (Balaras 2005) Balaras C. A., Drousa K., Dascalaki E., & Kontoyiannidis S. (2005). Heating energy consumption and resulting environmental impact of European apartment buildings. *Energy and Buildings*, 37(5), 429-442.
- (Balota 1983) Balota D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of verbal learning and verbal behavior*, 22(1), 88-104.
- (Baptista 2014) Baptista M., Fang A., Prendinger H., Prada R., & Yamaguchi Y. (2014, July). Accurate household occupant behavior modeling based on data mining techniques. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (pp. 1164-1170).
- (Basu 2011) Basu K. (2011). Predicting home service demands from appliance usage data (Doctoral dissertation, GRENOBLE INSTITUTE OF TECHNOLOGY).
- (Bellugi 2014) Bellugi U., Blumstein S., Fodor J., Mann V., & Samuel A. (2014, January). Panel discussion: the modularity of speech and language. In *Modularity and the Motor Theory of Speech Perception: Proceedings of A Conference To Honor Alvin M. Liberman* (p. 359). Psychology Press.
- (Bennett 2015) Bennett M. R., & Hacker P. (2015). On explaining and understanding cognitive behaviour. *Australian Journal of Psychology*.
- (Brown 2008) Brown R. E. (2008, July). Impact of smart grid on distribution system design. In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE* (pp. 1-4). IEEE.
- (Brotchie 1991) Brotchie P., Iansek R., & Horne M. K. (1991). Motor function of the monkey globus pallidus. *Brain*, 114(4), 1685-1702.
- (Buratti 2014) Buratti C., Lascaro E., Palladino D., & Vergoni M. (2014). Building Behavior Simulation by Means of Artificial Neural Network in Summer Conditions. *Sustainability*, 6(8), 5339-5353.
- (Chaiken 1987) Chaiken S. (1987). The heuristic model of persuasion. In *Social influence: the ontario symposium* (Vol. 5, pp. 3-39).

- (Chan 1999) Chan K. P., & Fu A. W. C. (1999, March). Efficient time series matching by wavelets. In Data Engineering, 1999. Proceedings., 15th International Conference on (pp. 126-133). IEEE.
- (Conte 2003) Conte G., & Scaradozzi D. (2003) Viewing home automation systems as multiple agents systems. Multi-agent system for industrial and service robotics applications, RoboCUP2003, Padova, Italy.
- (Doya 2000) Doya K. (2000). Computational model of basal ganglia in TD learning. *Current opinion in neurobiology*, 10(6), 732-739.
- (Ferreri 2014a) Ferreri E., Salotti J.M., & Favier P.A. (2014). Prediction of Electrical Consumptions Using a Bio-Inspired Behavioral Model, BauSIM2014 Conference Proceedings, Aachen, Germany, September 22-24.
- (Ferreri 2014b) Ferreri E., Salotti J.M., & Favier P.A. (2014). Taking human behavior into account in energy consumption simulation. In Proceedings 2014 SESDE I3M International Conference.
- (Ferreri 2015) Ferreri E., Salotti J.M., & Favier P.A. (2014). (2015, March). Simulation of Habits and Planned Activities Related to Electrical Domestic Consumption. In Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation (pp. 171-175). IEEE Computer Society.
- (Ferreri 2016) Ferreri E., Salotti J.M., & Favier P.A. (2016). Behavioural Approach for the Simulation of Electrical Consumption at the Neighbourhood Level. *International Journal of Simulation: Systems, Science and Technology*, 16(3), 12-1.
- (Fiske 1989) Fiske S. T., & Neuberg S. L. (1989). Category-based and individuating processes as a function of information and motivation: Evidence from our laboratory. In *Stereotyping and prejudice* (pp. 83-103). Springer New York.
- (Fiske 1999) Fiske S. T., & Neuberg S. L. (1999). The continuum model. *Dual-process theories in social psychology*, 321-254.
- (Gazzaniga 1995) Gazzaniga M. S. (1995). Principles of human brain organization derived from split-brain studies. *Neuron*, 14(2), 217-228.
- (Gilbert 2005) Gilbert N., & Troitzsch K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

- (Gittis 2014) Gittis A. H., Berke J. D., Bevan M. D., Chan C. S., Mallet N., Morrow M. M., & Schmidt, R. (2014). New roles for the external globus pallidus in basal ganglia circuits and behavior. *The Journal of Neuroscience*, 34(46), 15178-15183.
- (Grégoire 2013) Grégoire L. (2013). L'émergence et l'évolution du caractère obligatoire des automatismes cognitifs (Doctoral dissertation, Université de Bourgogne).
- (Haldi 2010) Haldi F. (2010). Towards a Unified Model of Occupants Behaviour and Comfort for Building Energy Simulation. (Doctoral dissertation, Université de Bourgogne).
- (Hawarah2010) Hawarah L., Ploix S., & Jacomino M. (2010, June). User behavior prediction in energy consumption in housing using Bayesian networks. In *Artificial Intelligence and Soft Computing* (pp. 372-379). Springer Berlin Heidelberg.
- (Hérault 2004) Guyader N., Chauvin A., Peyrin C., Hérault J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, 327(4), 313-318.
- (Ivry 2008) Ivry R., Schlerf J., Dedicated and intrinsic models of time perception, *Trends in Cognitive Sciences*, 12 (7), July 2008, Pages 273-280.
- (Jessen 2007) Jessen P. (2007). Simulating occupant presence and behaviour in buildings (Doctoral dissertation, Ecole Polytechnique Fédérale de Lausanne).
- (Joumaa 2011) Joumaa H., Ploix S., Abras S., De Oliveira G., et al. (2011) An a.s integrated into home automation system, for the resolution of power management problem in smart homes. *Energy Procedia*, 6:786–794.
- (Judah 2013) Judah G., Gardner B., & Auger R. (2013). Forming a flossing habit: An exploratory study of the psychological determinants of habit formation. *British journal of health psychology*, 18(2), 338-353.
- (Kashif 2011) Kashif A., Le X. H. B., Dugdale J., & Ploix S. (2011, January). Agent based Framework to Simulate Inhabitants' Behaviour in Domestic Settings for Energy Management. In *ICAART* (2) (pp. 190-199).

- (Klein 2011) Klein L., Kavulya G., Jazizadeh F., Kwak J., Becerik-Gerber B., Varakantham P., and Tambe M. (June 2011), Towards optimization of building energy and occupant comfort using multi-agent simulation. In The 28th International Symposium on Automation and Robotics in Construction (ISARC) 2011.
- (Lally 2010) Lally R., Van Jaarsveld P., Potts C. H., & Wardle J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6), 998-1009.
- (Le 2010) Le X. H. B., Kashif A., Ploix S., Dugdale J., Di Mascolo M., & Abras S. (2010). Simulating inhabitant behaviour to manage energy at home. In International Building Performance Simulation Association Conference, Moret-sur-Loing, France.
- (Lin 2013) Guiqing L., Xiuying X., and Qiao Y. (2013) Building energy saving design based on multi-agentsystem. In *Industrial Electronics*.
- (Meyer 2005) Meyer J. A., Guillot A., Girard B., Khamassi M., Pirim P., & Berthoz A. (2005). The Psikharpax project: towards building an artificial rat. *Robotics and autonomous systems*, 50(4), 211-223.
- (Montgomery 2009) Montgomery D. C., Runger G. C., & Hubele N. F. (2009). *Engineering statistics*. John Wiley & Sons.
- (Neal 2006) Neal D. T., Wood W., & Quinn J. M. (2006). Habits—A repeat performance. *Current Directions in Psychological Science*, 15(4), 198-202.
- (Nicolle 2013) Nicolle J., Bailly K., Rapp V., & Chetouani, M. (2013, September). Locating facial landmarks with binary map cross-correlations. In *Image Processing (ICIP), 2013 20th IEEE International Conference on* (pp. 2978-2982).
- (Page 2008) Page J., Robinson D., Morel N., & Scartezzini J. L. (2008). A generalised stochastic model for the simulation of occupant presence. *Energy and buildings*, 40(2), 83-98. press.
- (Parzen 1962) Parzen E. (1962) On the estimation of a probability density function and the mode. *Annals of Math. Stats.*, 33:1065-1076, 1962.
- (Philippe 1992) Philippe B., Saad Y., & Stewart W. J. (1992). Numerical methods in Markov chain modeling. *Operations Research*, 40(6), 1156-1179.
- (Piéron 1908) Piéron H. (1908). Les problèmes actuels de l'instinct. *Revue Philosophique de la France et de l'Étranger*, 66, 329-369.

- (Plaisant 2010) Plaisant O., Courtois R., Réveillère C., Mendelsohn G. A., & John O. (2010, March). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. In *Annales Médico-psychologiques, revue psychiatrique* (Vol. 168, No. 2, pp. 97-106). Elsevier Masson.
- (Robinson 2009) Robinson D., Haldi F., Kämpf J., Leroux P., Perez D., Rasheed A., & Wilke U. (2009, July). CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning. In *Eleventh international IBPSA conference* (pp. 1083-1090).
- (Roy 2006) Roy N., Roy A., and Das S.K. (2006) Context-aware resource management in multi-inhabitant smart homes a nash h-learning based approach. In *Pervasive Computing and Communications. PerCom 2006. Fourth Annual IEEE International Conference on*, pages 11–pp. IEEE, 2006.
- (Russel 1995) Russell S., Norvig P., & Intelligence, A. (1995). A modern approach. *Artificial Intelligence*. Prentice-Hall, Englewood Cliffs, 25, 27.
- (Salotti 2008) Salotti J.M. and Lepretre F. (2008), Classical and Operant Conditioning as Roots of Interaction for Robots, proceedings of the Workshop From motor to interaction learning in robots, Conference on Intelligent Robotics Systems, Nice, September 22-26
- (Seidenberg 1985) Seidenberg M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, 19(1), 1-30.
- (Sharples 1999) Sharples S., Callaghan V., and Clarke G. (1999) A multi-agent architecture for intelligent buildings sensing and control. *Sensor Review*, 19(2):135–140.
- (Shteingart 2014) Shteingart H., & Loewenstein Y. (2014). Reinforcement learning and human behavior. *Current opinion in neurobiology*, 25, 93-98.
- (Stephan 2009) Stéphan V., & Cogordan F. Courbotree: application des arbres de regression multivariés pour la classification de courbes. *Revue MODULAD* juin 2009
- (Struzik 1999) Struzik Z. R., & Siebes A. (1999). The Haar wavelet transform in the time series similarity paradigm. In *Principles of Data Mining and Knowledge Discovery* (pp. 12-22). Springer Berlin Heidelberg.
- (Sutton 1998) Sutton R. S., & Barto A. G. (1998). Reinforcement learning: An introduction. MIT press.

- (Tso 2007) Tso G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761-1768.
- (Watson 2013) Watson J. B. (2013). *Behaviorism*. Read Books Ltd. (Réédition)
- (Widen 2009 a) Widén J., Lundh M., Vassileva I., Dahlquist E., Ellegård K., & Wäckelgård E. (2009). Constructing load profiles for household electricity and hot water from time-use data—Modelling approach and validation. *Energy and Buildings*, 41(7), 753-768.
- (Widen 2009 b) Widén J., Nilsson A. M., & Wäckelgård E. (2009). A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand. *Energy and Buildings*, 41(10), 1001-1012.
- (Widen 2010) Widén J., & Wäckelgård E. (2010). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, 87(6), 1880-1892.
- (Worthy 2013) Worthy D. A., Hawthorne M. J., & Otto A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic bulletin & review*, 20(2), 364-371.
- (Yin 2006) Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6), 464-476.
- (Ywood 2002) Wood Y., Quinn W., & Kashy D. A. (2002). Habits in everyday life: thought, emotion, and action. *Journal of personality and social psychology*, 83(6), 1281.
- (Zaghloul 2012) Zaghloul K. A., Weidemann C. T., Lega B. C., Jaggi J. L., Baltuch G. H., & Kahana M. J. (2012). Neuronal Activity in the Human Subthalamic Nucleus Encodes Decision Conflict during Action Selection. *The Journal of Neuroscience*, 32(7), 2453-2460.]

ANNEXE: DATALOGGER INTELLIGENT



Description fonctionnelle : Il s'agit d'un système conçu, testé et assemblé en totalité, se plaçant entre la prise murale et la prise d'un appareil électroménager et permettant de récupérer les données de consommation électrique en temps réel.

Il comporte les éléments suivants :

- un microcontrôleur ATMEL 328P
- un capteur de courant (30A maximum)
- un capteur de tension secteur
- un émetteur récepteur Bluetooth 3.0 (115200 bps)
- une carte microSD 2 Go

- une horloge temps réel
 - un écran graphique monochrome
 - un boîtier convivial protecteur pour intégrer le tout avec une prise sur secteur
 - une alimentation autonome
- + un programme sous Android permettant de collecter les données sur Smartphone et d'éteindre et allumer à distance

Tentative de valorisation

Nous avons présenté le système à Aquitaine Science Transfert pour étudier la possibilité de déposer un brevet. Après étude du dossier, le dépôt de brevet s'est avéré possible mais complexe, car il existait déjà des systèmes avec des fonctionnalités comparables, bien que bien plus coûteux.

Nous avons également discuté avec des spécialistes des entreprises My Olympe et Legrand qui nous ont expliqué la difficulté et les coûts pour qualifier le système du point de vue des normes de sécurité, ce qui nous a dissuadés d'aller plus loin.

VALORISATION SCIENTIFIQUE

LISTES DES PUBLICATIONS EXHAUSTIVE

* Ferreri, E., Salotti, J. M., & Favier, P. A. Behavioural Approach for the Simulation of Electrical Consumption at the Neighbourhood Level. *International Journal of Simulation: Systems, Science and Technology*, 16(3), 12-1, 2016.

* Ferreri, E., Salotti, J. M., & Favier, P. A.. Simulation of Habits and Planned Activities Related to Electrical Domestic Consumption. In *Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation* (pp. 171-175). IEEE Computer Society, March 2015.

* E. Ferreri, J.-M. Salotti, P.-A. Favier, Taking human behaviour into account in energy consumption simulation, *proceedings of the 2nd International Workshop on Simulation for Energy, Sustainable Development & Environment (SESDE)*, Bordeaux, September 10-12, 2014.

* E. Ferreri, J.-M. Salotti J.-M. and P.-A. Favier, Prediction of electrical consumptions using a bio-inspired behavioral model, *BauSIM2014 Conference Proceedings*, Aachen, Germany, September 22-24, 2014.

* Bona, D. Doucet, E. Ferreri, B. Ramarohetra, J.M. Salotti, Acceptabilité des automatismes pour le contrôle thermique de l'habitat, *actes du congrès IBPSA*, 7 pages, Arras, France, 20-21 mai 2014.

* E. Ferreri, J.M. Salotti, P.A. Favier, Simulation prédictive pour la gestion des consommations électriques dans un quartier, *actes du congrès IBPSA*, 8 pages, Arras, France, 20-21 mai 2014.

Ateliers de travail :

- Participation à plusieurs ateliers de travail du CREAHD (Construction, Ressources, Environnement et habitat Durable) de la région aquitaine.
- Séminaire des doctorants de l'école doctorale SPI
- Présentation des travaux à l'équipe Cognitive et ingénierie humaine du laboratoire IMS