Caractérisation aveugle de la courbe de charge électrique : Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire

Mabrouka El Guedri

▶ **To cite this version:**

Mabrouka El Guedri. Caractérisation aveugle de la courbe de charge électrique : Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire. Traitement du signal et de l'image [eess.SP]. Université Paris Sud - Paris XI, 2009. Français. NNT : . tel-00461671

HAL Id: tel-00461671

https://theses.hal.science/tel-00461671

Submitted on 5 Mar 2010

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.

Faculté des sciences d'Orsay

UNIVERSITÉ PARIS-SUD 11

THESE DE DOCTORAT

SPÉCIALITÉ : PHYSIQUE

ÉCOLE DOCTORALE « SCIENCES ET TECHNOLOGIES DE L'INFORMATION, DES TÉLÉCOMMUNICATIONS ET DES SYSTÈMES »

Présentée par : MABROUKA EL GUEDRI

Caractérisation aveugle de la courbe de charge électrique :

Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire

Soutenance prévue le 9 novembre 2009 devant les membres du jury

Monsieur	BÉNIDIR	MESSAOUD	Examinateur
Monsieur	D'URSO	GUY	Encadrant EDF
Monsieur	FLEURY	GILLES	Directeur de thèse
Monsieur	HEBRAIL	GEORGES	Rapporteur
Monsieur	IDIER	JÉRÔME	Examinateur
Monsieur	LAJAUNIE	CHRISTIAN	Co-directeur de thèse
Madame	MARTIN	NADINE	Examinatrice
Monsieur	TOURNERET	JEAN-YVES	Rapporteur

À Toi le Très-Constant,
À toi le guide, le modèle,
À ma famille, mes amis et mes collègues,
À mes grands-parents paix à leurs âmes,
A nous.

Table des matières

Table des matières i

Table des figures vi

Liste des tableaux x

Introduction générale 1

1 Problématique et état de l'art 7
Introduction . . . 7
1.1 Problématique industrielle . . . 8
1.1.1 Contexte et motivations . . . 8
1.1.2 Objectifs . . . 9
1.1.3 Contraintes industrielles et stratégie envisagée . . . 10
1.1.4 Données d'entrée et connaissances annexes . . . 11
1.1.5 Synthèse . . . 13
1.2 Formalisation du problème de la *caractérisation* de la CdC . . . 13
1.2.1 Notations et définitions utiles . . . 13
1.2.2 Une définition de la « *caractérisation de la CdC* » revisitée . . . 14
1.2.3 Les problèmes sous-jacents à la caractérisation de la CdC . . . 15
1.3 État de l'art . . . 19
1.3.1 Historique . . . 19
1.3.2 Critères de classification des études antérieures . . . 20
1.3.3 Méthodes et modèles proposés . . . 21
Synthèse : avantages et limites . . . 23
Conclusions et discussions . . . 23

2 Analyse des signaux en vue de l'extraction de connaissances sur les usages 25
Introduction . . . 25
2.1 Démarche expérimentale . . . 25
2.1.1 Description des données : cas du secteur résidentiel . . . 26
2.1.2 Description des données : cas du secteur tertiaire (hypermarché) . . . 27
2.2 Analyse des signaux . . . 28
2.2.1 Définitions et notations . . . 28
2.2.2 Le chauffage par convecteurs . . . 31
2.2.2.1 Généralités sur le chauffage par convecteurs . . . 31

2.2.2.2 Analyse des signaux . . . 31
2.2.3 L'eau chaude sanitaire (ECS) . . . 44
2.2.3.1 Loi de l'amplitude moyenne . . . 44
2.2.3.2 Lois des durées . . . 44
2.2.4 Le froid alimentaire . . . 47
2.2.4.1 Premier cas : le secteur résidentiel . . . 47
2.2.4.2 Deuxième cas : le secteur tertiaire - froid alimentaire dans un hypermarché . . . 50
2.2.5 Les « gros » appareils électroménagers . . . 52
2.2.5.1 Présentation des appareils électriques ciblés . . . 52
2.2.5.2 Analyse de signaux . . . 52
2.2.6 Les appareils de cuisson . . . 57
2.2.7 Analyse d'exemples de CdC . . . 58
2.2.7.1 Cartographie des principaux usages chez deux clients particuliers . . . 58
2.2.7.2 Propriétés de la CdC générale et représentations adaptées . . . 60
2.2.7.3 Analyse événementielle pour la reconnaissance du chauffage 64
2.2.7.4 Foisonnement du chauffage et du chauffe-eau . . . 66
Synthèse de l'analyse des signaux d'usages électriques . . . 67
Orientations méthodologiques . . . 67

3 Segmentation automatique de la courbe de charge 69
Introduction . . . 69
3.1 Méthodologie de segmentation de la CdC . . . 69
3.1.1 Position du problème . . . 69
3.1.2 Approche envisagée . . . 70
3.2 Segmentation « chauffage / autre » . . . 71
3.2.1 Domaines transformés adaptés . . . 71
3.2.1.1 Le spectrogramme : un domaine de représentation adapté au chauffage . . . 71
3.2.1.2 La densité des événements . . . 72
3.2.2 Détection du chauffage dans le domaine temps-fréquence . . . 72
3.2.2.1 Une « nouvelle » méthode d'interprétation du spectrogramme 72
3.2.2.2 Segmentation du spectrogramme . . . 75
3.2.2.3 Extraction des motifs spectraux « pertinents » . . . 78
3.2.2.4 Identification du support temporel du chauffage électrique . . . 80
3.2.2.5 Contributions et discussions . . . 83
3.2.3 Détection du chauffage par seuillage de la densité des événements . . . 83
3.2.3.1 Connaissances *a priori* . . . 83
3.2.3.2 Approche proposée . . . 84
3.2.3.3 Illustrations . . . 86
3.2.3.4 Amélioration du détecteur du chauffage . . . 88
3.2.3.5 Contributions et discussions . . . 89
3.3 Segmentation « ECS / autre » . . . 89
3.3.1 Principe . . . 89

3.3.2 Détection des événements . . . 92
3.3.2.1 Test de Page-Hinkley avec *a priori* . . . 92
3.3.2.2 Sélection des événements pertinents . . . 93
3.3.3 Détecteur séquentiel de l'ECS . . . 94
3.4 Évaluation des algorithmes mis en oeuvre . . . 95
3.4.1 Performances de la segmentation chauffage / autres usages . . . 95
3.4.1.1 Taux d'erreur de segmentation . . . 95
3.4.2 Performances de la segmentation du chauffe-eau / autres usages . . . 98
Synthèse . . . 98

4 Estimation des caractéristiques d'usages à partir de la CdC 103
Introduction . . . 103
4.1 Objectifs . . . 103
4.2 Estimation de l'énergie consommée par le chauffage . . . 104
4.2.1 Démarche . . . 104
4.2.2 Modèle de l'énergie consommée par le chauffage . . . 104
4.2.3 Estimation du modèle de l'énergie du chauffage . . . 105
4.2.3.1 Estimation des énergies e_r, e_H et e_B . . . 105
4.2.3.2 Estimation du nombre de convecteurs $N_{co}(t)$ . . . 109
4.2.3.3 Estimation des « constantes du profil-consommateur » . . . 112
4.2.4 Procédure d'estimation de l'énergie du chauffage . . . 112
4.2.5 Evaluation de l'estimation de l'énergie consommée par chauffage . . . 113
4.3 Estimation de l'énergie consommée par le chauffe-eau . . . 117
4.3.1 Principe . . . 117
4.3.2 Résultats et discussions . . . 117
4.4 Identification des paramètres d'un modèle . . . 119
4.4.1 Exemples de modèles d'amplitude . . . 119
4.4.2 Estimation du modèle de l'ECS . . . 119
4.4.3 Résultats . . . 119
4.5 Estimation des paramètres dans un cadre probabiliste . . . 121
4.5.1 Estimation par MV des amplitudes des compresseurs . . . 121
4.5.2 Estimation MAP des caractéristiques du chauffage électrique . . . 121
Synthèse . . . 121

5 Séparation des sources de la CdC générale 125
Introduction . . . 125
5.1 Approche probabiliste de décomposition de la CdC du chauffage . . . 126
5.1.1 Modèle stochastique . . . 126
5.1.1.1 Lois *a priori* . . . 127
5.1.1.2 Vraisemblance des données . . . 128
5.1.1.3 Loi *a posteriori* . . . 128
5.1.2 Estimation des paramètres . . . 128
5.1.2.1 Exploration de l'espace d'état de la CdC . . . 129
5.1.2.2 « Chaînage » de signaux de convecteurs . . . 130
5.1.3 Procédure de décomposition du chauffage . . . 132
5.1.4 Exemples de résultats . . . 132

5.1.4.1 Convecteurs en régime de régulation . . . 133
5.1.4.2 Convecteurs en régime de régulation et/ou de saturation . . . 133
Synthèse et discussions . . . 133
5.2 Approche du type « décodage » - application au froid alimentaire . . . 136
5.2.1 Principe . . . 136
5.2.2 Décomposition de la CdC du froid alimentaire dans le secteur résidentiel 138
5.2.2.1 Informations *a priori* . . . 138
5.2.2.2 Prétraitements des signaux et extraction des événements pertinents . . . 138
5.2.2.3 Algorithme de décomposition de la CdC du froid alimentaire 138
5.2.3 Décomposition de la CdC du froid alimentaire d'un hypermarché . . . 142
5.2.3.1 Informations *a priori* . . . 143
5.2.3.2 Estimation des événements des compresseurs à partir de la CdC générale . . . 144
5.2.3.3 Reconstruction des scénarios de gestion des compresseurs . . . 145
Synthèse . . . 147

Conclusion générale **148**

Perpsectives **151**

Publications **160**

Annexes **i**

A Description du compteur bleu électronique **v**

B Définitions de quelques grandeurs électriques **vii**

C Sythèse des travaux antérieurs sur la caractérisation des usages électriques **xi**

D Thermostatically Controlled Appliances **xxxiii**

E Approches pour la décomposition des signaux composites **xxxv**

F Filtrage morphologique **li**

Bibliographie **liii**

Liste des algorithmes

5.1 Algorithme stochastique de décomposition du chauffage 132
5.2 Algorithme itératif d'extraction des compresseurs 140

Table des figures

1.1 Répartition de la consommation en électricité (France 2006) . . . 8
1.2 Exemple de CdC générale du secteur résidentiel (sortie du CBE) . . . 12
1.3 Exemple de CdC générale du secteur tertiaire (mesurée par un wattmètre) . . . 12
1.4 Décomposition de la CdC générale vue comme un problème de cartographie des scénarios des usages . . . 16
1.7 Décomposition de la CdC générale vue comme un problème de description « physique » des usages . . . 16
1.5 Décomposition de la CdC générale vue comme un problème de séparation des séries des événements . . . 17
1.6 Décomposition de la CdC générale vue comme un problème de cartographie de la consommation . . . 17
1.8 Décomposition de la CdC générale vue comme un problème de reconstruction de signaux des usages . . . 18
1.9 Modèles des composantes de la CdC d'un hypermarché [Laj02] . . . 22

2.1 Illustration des séries des événements d'un appareil électrique . . . 30
2.2 Illustration des suites des durées caractéristiques des événements d'un appareil électrique . . . 30
2.3 Puissance appelée par un convecteur électrique ($T_e = 2$s, K_2) . . . 32
2.4 Symétrie de l'amplitude d'enclenchement et de l'amplitude de déclenchement d'un convecteur ($T_e = 2$sec, K_2) : Convecteurs test - tous les convecteurs de la Maison Domotique . . . 34
2.5 Quasi-périodicité de la puissance appelée par un convecteur . . . 34
2.6 Evolution « quotidenne » du rapport cyclique : Exemple de deux convecteurs . . . 36
2.7 Analyse statistique du rapport cyclique . . . 37
2.8 Foisonnement de convecteurs . . . 39
2.9 Exemple de CdC de chauffage - janvier 2008 . . . 40
2.10 Exemples de CdC du chauffage - avril et mai 2008 ($Te = 1$s, K_3) . . . 41
2.11 Nombre d'événements de la CdC chauffage en fonction de la période d'échantillonnage . . . 42
2.12 Exemple de CdC sous-échantillonnée . . . 43
2.13 Puissance appelée par l'ECS - un cycle de fonctionnement . . . 43
2.14 Asymétrie On-Off . . . 45
2.15 Histogrammes des durées *caractéristiques* de l'ECS : (Δt_{11}), (Δt_{00}) et (Δt_{10}) . . . 45
2.16 Exemple de CdC d'ECS . . . 46

2.17 Graphe d'intensité de l'ECS . . . 47
2.18 Puissance moyenne quotidienne appelée par les compresseurs . . . 48
2.19 Puissance active d'un réfrigérateur (essai de laboratoire - EDF R&D avril 2007) . . 49
2.20 Puissance instantanée d'un réfrigérateur mono-compresseur (Client A - mars 2009) 49
2.21 CdC du froid alimentaire d'un hypermarché . . . 51
2.22 Zoom sur la CdC générale d'un hypermarché . . . 51
2.23 Puissance moyenne appelée par un lave-linge ($T_e = 1\ s$) . . . 53
2.24 Puissance moyenne appelée par un lave-vaisselle (Maison Domotique -K_1,$T_e = 1\ s$) 55
2.25 Puissance moyenne appelée par un sèche-linge ($T_e = 1$ s) et son spectre . . . 55
2.26 a) Puissance moyenne appelée par le groupe des gros életroménagers : (Client B -K_2,$T_e = 1s$)
b) Puissance moyenne appelée par un lave-linge (K_4,$T_e = 1$ s) : plusieurs cycles de fonctionnement (Client B) . . . 56
2.27 Puissance appelée par les appareils de cuisson ($T_e = 1$ s) . . . 57
2.28 Histogramme des durées de l'usage micro-ondes (K_4) . . . 59
2.29 Consommation quotidienne du client A : janvier 2008 . . . 59
2.30 Part du chauffage et de l'ECS dans la consommation quotidienne chez le client A . 61
2.31 Part du froid alimentaire dans la consommation quotidienne chez le client B . . . 61
2.32 Exemple de CdC chez un client particulier . . . 62
2.33 Zoom - foisonnement d'usages . . . 63
2.34 CdC quotidienne - 12 juillet 2008 (Client A) . . . 63
2.35 Composantes de la CdC quotidienne - 12 juillet 2008 (Client A) . . . 65
2.36 CdC quotidienne - 30 septembre 2008 (Client A) . . . 65
2.37 Exemple de CdC quotidienne générale + ECS (Client A) . . . 66

3.1 Schéma général de la segmentation de la CdC quotidienne résidentielle . . . 71
3.2 Exemple de spectrogramme d'une CdC générale ($T_e = 2$s, $N_f = 512$ échantillons) . 74
3.3 Schéma général de la détection du chauffage dans le domaine temps-fréquence . . 74
3.4 Schéma général de la détection du chauffage dans le domaine temps-fréquence . . 77
3.5 Segmentation du spectrogramme d'une CdC électrique . . . 77
3.6 Illustration de l'extraction des motifs fréquents : **(a)** Matrice des données **(b)** Treillis **(c)** Motifs obtenus pour $f_c = 2$ . . . 79
3.7 Extraction des motifs du chauffage électrique : CdC du 5 au 9 Décembre 2004 ($T_e = 2$s) . . . 81
3.8 Détection du chauffage électrique à partir de l'extraction des motifs fréquents les plus longs . . . 82
3.9 Schéma général de l'approche événementielle pour la détection du chauffage électrique . . . 84
3.10 Extraction des régulations rapides à partir d'une CdC générale ($A_{c0} = 350$ W, $T_e = 1$ s, $t_{c0} = 1400$ s et $T_r = 80$ s) . . . 87
3.11 Illustration de la détection par seuillage de la densité des événements ($s_0 = 35$) . . 87
3.12 Signatures de convecteurs et d'un sèche-linge dans le domaine (P^+, P^-) . . . 88
3.13 Schéma général de détection de l'ECS à partir d'une CdC générale ($T_e \sim 1$s) . . . 90
3.14 Illustration du filtrage de la CdC générale en absence de chauffage électrique . . . 91
3.15 Illustration du filtrage de la CdC générale en présence chauffage électrique de faible consommation . . . 91

3.16 Illustration du filtrage de la CdC générale en chauffage en marche toute la journée (forte consommation) . . . 92
3.17 Taux de non-détection du chauffage (cas de la Maison Domotique) . . . 96
3.18 Taux de fausse-alarme du chauffage (cas de la Maison Domotique) . . . 96
3.19 Taux de fausse alarme et taux de non-détection du détecteur de chauffage . . . 97
3.20 Détection de l'ECS : CdC en absence du client . . . 99
3.21 Détection de l'ECS : CdC avec très peu de chauffage . . . 99
3.22 Détection de l'ECS : CdC avec une consommation moyenne du chauffage . . . 100
3.23 Détection de l'ECS : CdC avec une forte consommation du chauffage . . . 100
3.24 Illustration de non détection et fausse alarme du détecteur de l'ECS . . . 101
3.25 Détection de l'ECS : exemple de non-détection importante . . . 101

4.1 Filtrage : extraction de l'enveloppe haute de la CdC ($T_H = 700$ s) . . . 106
4.2 Filtrage : Illustration de l'extraction de l'enveloppe basse de la CdC ($T_B = 700$ s) . . . 107
4.3 Filtrage : Illustration de l'extraction de l'enveloppe haute de la CdC ($T_H = 700$ s) . . . 108
4.4 Filtrage : Illustration de l'extraction de l'enveloppe basse de la CdC ($T_B = 700$ s) . . . 108
4.5 Segmentation de la densité des évévenements en zones à N_{co} constant ($t_{c0} = 700$ s) 110
4.6 Détection de ruptures de la répartition de la composante "régulations " . . . 110
4.7 Estimation de l'énergie consommée par le chauffage (en kWh) . . . 114
4.8 Erreur relative (en points) d'estimation de l'énergie consommée par le chauffage . . . 115
4.9 Histogramme de l'erreur relative sur l'estimation de la part du chauffage dans l'énergie consommée par jour . . . 116
4.10 Estimation de l'énergie consommée par l'ECS . . . 117
4.11 Energie consommée par le chauffe-eau . . . 118
4.12 Erreur d'estimation de l'énergie consommée par le chauffe-eau . . . 120
4.13 Estimation du modèle de l'ECS :illustration du critère (d =2380 W) . . . 120
4.16 Estimation de l'amplitude moyenne des compresseurs à partir de la CdC générale d'un hypermarché . . . 121
4.14 Estimation du modèle de l'ECS à partir des données réelles . . . 123
4.15 Histogramme du résidu d'identification du modèle . . . 123

5.1 Signaux de puissance de convecteurs fonctionnant simultanément ($T_e = 2$ s) . . . 127
5.2 Configurations possible d'un convecteur sur deux fenêtres successives . . . 131
5.3 Décomposition d'un mélange de deux convecteurs - sans saturation . . . 133
5.4 Décomposition d'un convecteur en présence d'une plage de saturation . . . 134
5.5 Décomposition d'un mélange de deux convecteurs en présence d'une zone de saturation . . . 135
5.6 Schéma général de la méthode d'extraction des signaux de compresseurs . . . 137
5.7 Extraction des événements pertinents des compresseurs à partir de la CdC du froid alimentaire ($T_e = 2$ s) . . . 139
5.8 Zoom sur le signal estimé et sur le résidu- Première itération . . . 141
5.9 Zoom sur le signal estimé et sur le résidu- Deuxième itération . . . 141
5.10 Zoom sur le signal estimé et sur le résidu - Troisième itération . . . 142
5.11 Durées des cycles de fonctionnement des signaux extraits . . . 143
5.12 Estimation de l'amplitude moyenne des compresseurs à partir de la CdC générale d'un hypermarché . . . 144

5.13 Reconstruction de scénario de fonctionnement d'un nombre inconnu de compresseurs à partir de la CdC générale d'un hypermarché . . . 146
5.14 Reconstruction de scénario de fonctionnement d'un nombre inconnu de compresseurs à partir de la CdC du froid alimentaire seul . . . 146
5.15 Histogramme des variations d'amplitude *on-off* d'un convecteur et de l'ECS . . . 157
5.16 Cas test - mélange du froid alimentaire et de l'ECS . . . 159

A.1 Sortie Télé-info du CBE . . . v

B.1 Dipôle linéaire et dipôle non-linéaire . . . vii

C.1 Modèles des composantes de la CdC d'un hypermarché [Laj02] . . . xvi
C.8 Estimation des usages : première famille de méthodes . . . xxviii
C.2 Schéma de principe illustrant des exemples de signatures électriques dans le domaine $(\Delta P, \Delta Q)$ . . . xxix
C.3 Architecture du software du système NILM (régime permanent) de la société ENETICS . . . xxix
C.4 Puissance active d'une lampe et d'un moteur (120 Hz) (Extrait de [Lee93]) . . . xxx
C.5 Séparation d'usages basée sur les puissances harmoniques de rang 3 (Extrait de [LLC+03]) . . . xxx
C.6 $\nu - sections$ d'un moteur à induction (Extrait de [LMLS93]) . . . xxxi
C.7 Méthode hybride de décomposition de la CdC électrique . . . xxxi

D.1 Chauffe-eau à accumulation . . . xxxiii
D.2 Schéma de principe d'un système à compression . . . xxxiv

E.1 Schéma de séparation de sources . . . xlii
E.2 Séparation des sources à partir d'un seul capteur dans un domaine transformé . . . xlv
E.3 Schéma général de séparation basée sur des modèles statistitiques *a priori* . . . xlv

F.1 Ouverture et fermeture morphologiques . . . lii
F.2 Ouverture et fermeture morphologiques . . . lii

Liste des tableaux

1.1 Résumé des caractéristiques des méthodes non-intrusives 23

2.1 Caractéristiques des données d'entrée - CdC dans le secteur résidentiel 27
2.2 Cartographie du chauffage électrique en France par type (matériel) de chauffage (2000) 31
2.3 Caractéristiques de consommation d'un ballon d'ECS 43

4.1 Performances de l'estimation du nombre de convecteurs à partir de la CdC générale 111
4.2 Apprentissage des constantes du profil consommateur 112

5.1 Résultats d'estimation du taux horaire de fonctionnement 145
5.2 Lois a priori des paramètres des classes d'événements de la CdC 156

A.1 Pas de quantification des mesures électriques en sortie du CBE vi

B.1 Exemples de générateurs d'harmoniques ix

E.1 Exemples de problèmes de séparation de sources à partir d'un seul capteur xlii

Introduction générale

La persévérance, c'est ce qui rend l'impossible possible,

le possible probable et le probable réalisé.

Robert Half

Dans un contexte marqué par une sensibilité croissante aux enjeux environnementaux liés à l'énergie et où l'équilibre entre la demande et la consommation en électricité est accentué par l'ouverture du marché énergétique, Électricité de France (EDF) s'intéresse à la conception et au développement de services énergétiques à sa clientèle. C'est dans ce cadre général que s'inscrivent les travaux de cette thèse intitulée « C**aractérisation aveugle de la courbe de charge électrique : détection, classification et estimation des usages** [1] **dans les secteurs résidentiel et tertiaire** ».

La courbe de charge (**CdC**) générale désigne une grandeur électrique qui décrit l'évolution au cours du temps de la consommation en électricité d'un système donné. L'objectif de la *caractérisation* (dont nous proposons une définition dans ce mémoire) est d'extraire un ensemble de descripteurs à partir de la CdC électrique d'un client donné. Il peut s'agir de la détection de présence d'un usage électrique, de l'estimation de ses paramètres caractéristiques (énergie, puissance nominale, etc.) ou encore la de la reconstruction des signaux générés par chaque appareil. Ces estimées contribuent à la définition d'une offre énergétique variée et personnalisée (analyse détaillée de la consommation totale, détection de défaut, surveillance de systèmes complexes, audit énergétique). Ces services permettent au client d'identifier les sources potentielles d'optimisation de sa consommation.

Notre objectif se résume ainsi : étudier les possibilités et les méthodes permettant de caractériser la CdC générale de manière aveugle, non-intrusive et dans des conditions réelles. Par « *aveugle* » nous entendons qu'aucune information sur le client et ses appareils n'est accessible et qu'aucune procédure d'apprentissage à partir d'une base de signaux collectés chez le client n'est envisageable. Le caractère « non-intrusif » de la solution recherchée signifie que la solution ne doit pas nécessiter à l'intervention d'un personnel qualifié chez le client. La solution recherchée doit également être compatible avec les contraintes industrielles (infrastructure existante, robustesse et simplicité de mise en œuvre des algorithmes).

1. Un usage électrique désigne un ou plusieurs appareils électriques de la même famille, par exemple le froid alimentaire, le chauffage électrique, etc.

Trois sous-problématiques sont sous-jacentes à la caractérisation de la CdC :

1. *la détection d'usages* ;
2. *la classification des signaux du mélange* ;
3. *l'estimation de descripteurs des usages ou encore l'extraction de certaines classes de signaux à partir du mélange.*

La problématique industrielle considérée est relativement récente. À notre connaissance, il n'existe pas de solution théorique ou industrielle au problème posé. L'étude bibliolgraphique a permis d'identifier très peu de travaux de recherche et de solutions industrielles. Néanmoins, le sujet de caractérisation aveugle de la courbe de charge dans une optique de maîtrise de la demande en énergie (**MDE**) devient un sujet à la mode et très actif autant sur le plan industriel (*Maison intelligente*) que sur le plan académique. En effet, ce sujet fait l'objet de deux thèses récentes : la thèse de Mme Gäelle Darnis [2] [DR09] et celle Melle Hala Najmeddine [3] [Naj09]. Étant donnés le caractère confidentiel de ces travaux et les enjeux commerciaux associés, peu d'échanges ont pu avoir lieu.

Une étude comparative des trois solutions proposées serait intéressante sur le plan scientifique et sur le plan industriel. Nous pouvons d'ores et déjà préciser que les trois études s'intéressent au problème de caractérisation des usages électriques à partir de données d'entrée de propriétés (nombre d'observations et caractéristiques des mesures) différentes . Plus précisément, nous considérons un mélange mono-valué (la puissance active moyenne en sortie du compteur électrique échantillonnée à une fréquence de l'ordre du hertz), alors que les deux thèses étudient les solutions envisageables à partir de l'analyse de deux mélanges (intensité de courant et tension ou bien puissance active instantanée et puissance réactive instantanée à une fréquence d'échantillonnage assez élevée).

Mise à part ces deux études, les premières tentatives de décomposition de la CdC générale remontent aux années 80 notamment aux Etats-Unis [Har89, Lee93, LLC+03, Cox06], au Canada [Far99], en France [Bon96, Heb, Laj02] ainsi qu'au Japon [CRI] et plus récemment en Allemagne [BV03, BV04]. L'analyse de ces travaux a permis d'identifier quatre familles d'approches. La première approche s'inspire des modèles à machines d'états finis. La deuxième approche s'affranchit de la sous-détermination du problème posé en introduisant un modèle probabiliste du mélange et en intégrant des connaissances *a priori* sur les composantes à estimer. Les deux dernières approches s'inscrivent dans le cadre de la reconnaissance de formes et exploitent souvent une signature de l'appareil (signal électrique exact ou modèle du signal ou des paramètres caractéristiques) apprise pour la reconnaissance des usages.

Le dénominateur commun à ces approches est l'utilisation (formelle ou non) de connaissances apprises pour discriminer les usages de différentes manières : apprentissage de la signature exacte (ou d'un modèle) de l'appareil, estimation de quelques paramètres des usages étudiés à partir de données collectées sur le site ou utilisation d'informations fournies par le client à travers un questionnaire par exemple.

Nous nous appuyons sur l'analyse de ces travaux ainsi que sur l'état de l'art des méthodes de séparation de sources mono-capteur pour définir une méthodologie générale dédiée à la caractérisation de la CdC générale.

2. Thèse soutenue en Janvier 2009
3. Soutenance prévue pour fin 2009

Nous proposons de formaliser le problème dans le cadre générique de la décision en le décomposant en quatre sous-problématiques : *la détection*, *la classification* et *l'estimation* (estimation de paramètres ou recontruction de signaux) des usages. Afin de compléter notre connaissance du problème, une analyse approfondie des données nous permet de définir des *a priori* sur les usages étudiés et de choisir des domaines de représentations adaptés à certains signaux.

Ainsi, nous situons notre problématique générale soit dans le cadre de la théorie bayésienne qui offre la possiblité d'introduire des connaissances *a priori* imprécises ou encore dans un cadre de la séparation de sources mono-capteur où l'on exploite les propriétés des signaux dans des domaines de représentations dédiés.

Les principaux résultats de cette thèse concernent :

- ➢ L'extraction de connaissances *a priori* sur les usages électriques à partir de données réelles.
- ➢ La définition d'une méthodologie générale pour segmenter la CdC et son application au cas du secteur résidentiel. Nous verrons que les performances obtenues dans des conditions réelles sont satisfaisantes.
- ➢ Le choix et la mise en œuvre de différentes méthodes d'estimation des caractéristiques des usages.
- ➢ La définition de deux modèles stochastiques de la CdC estimés par les méthodes **MCMC** (Monte Carlo par Chaînes de Markov). Ces deux modèles sont respectivement utilisés pour décomposer la courbe de charge du chauffage électrique et pour reconstruire les principales classes de signaux de la CdC.

Les deux apports marquants de cette thèse sont les suivants

Apport industriel (applicatif) *Définition, mise en place et évaluation d'une méthode aveugle de cartographie de l'énergie totale consommée par usage à partir d'une CdC générale quelconque. Les algorithmes ont en partie été implantés sur un système embarqué en vue d'une intégration dans un système dédié. EDF envisage d'évaluer ces algorithmes en 2010 pour une cinquantaine de clients du secteur résidentiel en vue d'un déploiement à moyen termes.*

Apport scientifique *Définition d'un modèle stochastique de la CdC générale, plus précisément de la série des événements de la CdC, vue comme le mélange de cinq classes de signaux. Bien que ce travail ne soit pas achevé, il constitue un premier modèle générique qui ouvre des possibilités de reconstruction des signaux d'une CdC quelconque et qui peut être transposé à d'autres cas d'application* (*hypermarchés, auxilliaires d'une centrale nucléaire*). *Il s'agit d'une approche innovante dans le domaine de la séparation de sources mono-capteurs. En effet, le paradigme de séparation de sources mono-capteur est fondé sur un apprentissage à partir d'un exemple de signaux. Alors que notre approche s'appuie sur des connaissances a priori définies à partir de signaux d'entraînements, mais ne nécessite pas d'apprentissage.*

Organisation du manuscrit

Ce travail se décompose en six chapitres :

1. Problématique et état de l'art
2. Analyse des signaux en vue de l'extraction de connaissances sur les usages
3. Segmentation automatique de la courbe de charge générale
4. Estimation des caractéristiques d'usages à partir de la CdC
5. Séparation des sources de la courbe de charge générale
6. Conclusion générale et perspectives.

Problématique et état de l'art

Ce premier chapitre se décompose en trois sections. La première section présente le contexte et les motivations industrielles de cette étude. La deuxième section est dédiée à la formalisation du problème et définit le cadre général des sous-problématiques sous-jacentes à la caractérisation de la CdC dans le cadre de la décision.

Analyse des signaux en vue de l'extraction de connaissances sur les usages

Le deuxième chapitre présente les résultats de l'analyse détaillée des signaux des usages, les connaissances *a priori* extraites (modèles de signaux, loi *a priori* sur des paramètres, représentations adaptées aux usages). Cette analyse porte principalement sur les trois classes d'appreils suivantes : les convecteurs, le chauffe-eau et les compresseurs. Les données disponibles sont peu représentatives à cause des difficultés rencontrées pour la collecte de ces signaux. Néanmoins les concuslions de cette analyse sont forts utiles pour le choix de méthodes pertinenetes pour la caractérisation des signaux électriques.

Segmentation automatique de la courbe de charge générale

Ce chapitre présente une méthodologie de segmentation de la CdC, dénommée approche classe/non-classe. Cette approche est déclinée dans le cas du secteur résidentiel. Nous détaillons deux méthodes de segmentation chauffage/non-chauffage et les résultats obtenus à partir des données réelles. Un algorithme séquentiel de segmentation chauffe-eau / non-chauffe-eau fondée sur le test de Page-Hikley avec introduction d'*a priori* est également exposé. Nous présentons des illustrations des résultats obtenus, les performances de l'algorithme ainsi que les avantages et limites de cette méthode.

Estimation des caractéristiques d'usages à partir de la courbe de charge

Le problème d'estimation de paramètres d'usages (énergie, amplitude, nombre de composantes) est traité dans ce chapitre par différentes méthodes (filtrage non-linéaire, optimisation non-linéaire, approche probabiliste). Ce chapitre exploite souvent des résultats des chapitres dédiés à la segmentation et à la reconstruction des signaux.

Un résultat marquant décrit dans ce chapitre est l'estimation de l'énergie consommée par usage (chauffage et chauffe-eau). Les résultats issus des tests sur des données réelles sont détaillés et consituent une réponse concrète à un des objectif industriel (analyse détaillée de l'énergie totale consommée).

Séparation des sources de la CdC générale

Compte tenu de la diversité des signaux du mélange et du caractère mal posé du problème de reconstruction des signaux des usages, nous proposons trois méthodes différentes. Les deux premières méthodes sont définies pour la séparation des sources d'un mélange mono-capteur particulier : il s'agit de la CdC *quasi-mono-usage* c'est à dire une CdC contenant soit un usage unique soit plusieurs usages dont un prépondérant et présentant une forte parcimonie avec les autres usages. La première méthode de séparation utilise un modèle stochastique à base d'un dictionnaire de formes temporelles. Les paramètres du modèle sont estimées au sens du maximum *a posteriori* (**MAP**) par un algorithme du recuit simulé. La deuxième méthode consiste en deux étapes : décodage de la série des événements du mélange puis reconstruction les signaux étant donné un modèle d'amplitude.

La troisième méthode s'attache au problème de la séparation des sources d'une CdC générale quelconque. Nous proposons de poser ce problème comme étant un problème de séparation en cinq classes de signaux prédéfinies. Le choix des classes est orienté par l'ensemble des connaissances *a priori* sur les signaux. Etant donnée cette définition du problème, nous ramenons le problème de séparation de la CdC en un problème de séparation et d'appariement des événements de la CdC. Ce problème est considéré dans le cadre de la simulation bayésienne. Plus précisément, la décomposition de la série des événements esteffectuée au sens du maximum *a posteriori* et s'appuie sur des lois a priori définies sur les durées caractéristiques des usages et sur la morphologie des signaux.

Conclusions et persectives

Nous présentons les principales conclusions de ces travaux, une synthèse des possibilités et modalités de caractérisation de la CdC dans les conditions réelles définies par EDF.

Nous exposons quelques perspectives envisagées concernant les méthodes et algorithmes détaillés dans la thèse, une méthode de décomposition d'une CdC quelconque ayant fait l'objet d'une étude de faisabilité. des événements et enfin des perspectives fondées sur de nouvelles solutions technologiques permettant d'enrichir les données d'entrée par exemple.

Chapitre 1

Problématique et état de l'art

Une humble fleur est le labeur des siècles.

William Blake

Introduction

Les trois dernières décennies ont été marquées par l'émergence de nouveaux services, *les services énergétiques*, conçus dans une optique de maîtrise de la demande en énergie (**MDE**). Ce concept de services énergétique n'est pas nouveau. La nouveauté réside dans la richesse du bouquet de services proposés ainsi que dans la finesse des informations fournies au client. La définition de ces services était fondée sur l'exploitation de données contextuelles (données climatiques, caractéristiques thermiques du bâtiment, questionnaire du client, etc.). Bien que ces sources d'information offrent la possibilité d'une *description empirique et qualitative de la consommation électrique d'un client donné* (ou d'un groupe de clients), elles ne permettent pas de proposer une offre énergétique diversifiée, personnalisée et à grande échelle. Une démarche, initiée par George Hart [Har89] dans les années 80 et poursuivie dans le cadre de travaux décrits en (1.3), consiste à analyser la *courbe de charge générale* (une ou plusieurs grandeurs électriques) du système étudié pour décrire la consommation de chaque appareil.

C'est dans ce cadre général que s'inscrivent les travaux de cette thèse [1] intitulée **Caractérisation de la courbe de charge générale : détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire**, où la courbe de charge désigne une grandeur électrique qui décrit l'évolution au cours du temps de la consommation d'un système donné.

Dans ce premier chapitre, nous décrivons la problématique industrielle considérée et proposons une formalisation du problème considéré dans le cadre général des problèmes inverses. Nous présentons ensuite une synthèse des approches existantes, leurs avantages, leurs limites ainsi qu'une justification d'une nouvelle méthodologie au problème posé.

1. La thèse a été initiée par Électricité de France, l'École supérieure d'Électricité et l'École des Mines de Paris en 2006.

1.1 Problématique industrielle

1.1.1 Contexte et motivations

a) Contexte

L'énergie électrique représente la première énergie finale [2] utilisée en France. Omniprésente dans le quotidien de chacun, elle est devenue un vecteur énergétique employé à divers usages domestiques et industriels. La production de l'énergie électrique en France a triplé pendant les trente dernières années pour répondre à la demande croissante des différents secteurs, notamment les secteurs résidentiel et tertiaire. La cartographie de la consommation électrique par secteurs en France, présentée sur la figure (1.1), met en évidence la part importante de la consommation dans l'habitat tertiaire et résidentiel.

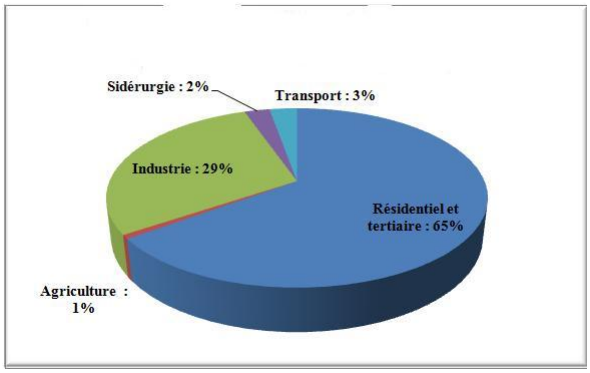


FIGURE 1.1: Répartition de la consommation en électricité (France 2006)

Une meilleure gestion des ressources énergétiques, particulièrement dans ces deux secteurs, s'impose, notamment dans un contexte où les défis environnementaux et l'épuisement à terme de certaines ressources d'énergie représentent un enjeu majeur. Pour celà, il est nécessaire d'avoir une bonne connaissance de la consommation d'électricité afin d'amener les utilisateurs finaux à modifier leur comportement énergétique.

b) Motivations

Les motivations pour avoir une meilleure connaissance de la consommation d'une installation donnée varient en fonction du secteur concerné et du point de vue (client / fournisseur / distributeur). Dans le secteur résidentiel, fournir au client une analyse détaillée de la consommation des différents usages permet de l'aider à optimiser sa consommation d'électricité et à modifier son *comportement énergétique* (travaux d'isolation de la maison, remplacement d'un ou de plusieurs appareils, etc). Précisons que l'intérêt que prête la clientèle à une telle analyse a été confirmé par des enquêtes [Wag05].

2. L'énergie finale est l'énergie distribuée aux différents consommateurs (industriels, particuliers, ...), sachant que l'énergie intialement produite s'appelle énergie primaire et que l'énergie obtenue en tenant compte des consommations internes à la branche énergie s'appelle énergie secondaire (ADEME).

Dans le secteur tertiaire, le client (exemple : hôtels, hypermarchés) est potentiellement motivé par une connaissance précise de la gestion (scénario d'utilisation) des usages ainsi qu'au diagnostic de défauts. Ces informations lui permettent d'optimiser le remplacement des machines défaillantes et d'éviter une sur-consommation coûteuse. Quant au fournisseur d'électricité, il souhaite avoir des informations plus précises sur les usages des clients pour mieux gérer ses moyens de production (notamment pendant les périodes de pointe) et optimiser les outils de surveillance de sytèmes électriques complexes (au sein des systèmes de production eux-mêmes par exemple les transformateurs des auxilliaires d'une centrale nucléaire).

En résumé, disposer d'informations fiables et précises sur la consommation d'électricité d'une installation donnée devient une nécessité dans les contextes économique et industriel, et des enjeux environnementaux actuels.

1.1.2 Objectifs

La finalité de ces travaux est *d'avoir une connaissance précise et fiable d'un système électrique donné afin d'optimiser sa consommation d'électricité*, et ce à travers l'analyse de la CdC générale du système étudié. Les objectifs industriels se déclinent de la manière suivante.

a) Analyse détaillée de l'énergie (non)-consommée par usages

Il s'agit d'avoir une décomposition de la consommation totale d'un système donné par usage électrique. Cette information permet de proposer par exemple au client du secteur résidentiel une analyse détaillée de sa consommation quotidienne, hebdomadaire et mensuelle, ce qui lui permet d'identifier les sources potentielles d'amélioration de sa gestion et de suivre l'impact de ses actions au cours du temps à travers une estimation de l'énergie non-consommée par certains usages (exemple : énergie non-consommée grâce à l'effacement du chauffage électrique sur des plages horaires précis).

b) Détection d'usage

L'objectif est de détecter la présence (ou non) d'un usage donné, par exemple le chauffage électrique chez un client. Cette information permet de mieux cibler des offres de nouveaux systèmes de chauffage par exemple.

c) Extraction des caractérisques des appareils électriques

Les caractéristiques recherchées sont de deux natures différentes :

- ➢ les paramètres intrinsèques d'un appareil donné tel que la puissance nominale.
- ➢ le nombre d'appareils d'un usage donné (exemple : nombre de compresseurs du parc de froid alimentaire d'un hypermarché).

Ces paramètres peuvent servir de données d'entrée à des logiciels d'audit énergétique ou d'appui au conseil de la clientèle en vue de l'optimisation de son installation. Dans le secteur résidentiel par exemple, disposer de la puissance nominale des différents convecteurs ou encore du chauffe-eau permet d'évaluer l'adéquation ou non du système installé respectivement par rapport aux caractéristiques du bâtiment et au besoin du client en consommation d'eau-chaude.

d) Détection d'anomalies

L'objectif est de détecter les défauts de dysfonctionnement d'un appareil donné (exemple : entartrage du chauffe-eau) ainsi que les défauts de « mauvaise » utilisation d'un usage (exemple : dysfonctionnement du chauffe-eau par rapport au tarif choisi, erreur de programmation du scénario de fonctionnement des appareils d'un parc de machines donné).

e) Estimation des scénarios d'usages

Deux grandeurs sont recherchées.

- ➢ Le support temporel de fonctionnement des usages d'intérêt, où le support temporel est l'union des intervalles définis par la mise en marche et l'arrêt de l'usage étudié. Dans le cas particulier du secteur résidentiel, notre objectif est de fournir une segmentation de la courbe de charge quotidienne pour les deux usages suivants : chauffage électrique et chauffe-eau.
- ➢ Le scénario de fonctionnement d'un usage donné, c'est à dire le séquencement de fonctionnement de différents appareils d'un même usage. Nous traitons le cas du froid alimentaire d'un hyper-marché. Cette grandeur peut être utile pour l'optimisation de la gestion des usages.

f) Suivi au cours du temps des usages électriques

Il s'agit de décomposer la puissance cumulée au cours du temps d'une installation électrique donnée par usages. Ainsi, nous disposons de la courbe de charge de chaque usage. Ces données d'analyse permettent d'envisager par exemple des offres telles que le suivi de vies des personnes âgées. Elles serviront également à améliorer la connaissance pour un fournisseur de sa clientèle pour mieux prédire par exemple les pics de consommation et d'avoir un profil dynamique (à l'échelle d'une journée ou d'une semaine) par usage, à une échelle régionale ou nationale.

1.1.3 Contraintes industrielles et stratégie envisagée

Afin d'obtenir les informations précises et pertinentes sur la consommation des usages électriques, différentes stratégies sont envisageables : exploiter des questionnaires remplis par les clients et les données climatiques, analyser des grandeurs électriques des différents appareils ou instrumenter les appareils. Or les sources d'informations contextuelles (données météo, questionnaires) ne sont pas nécessairement fiables ou ne sont pas disponibles. Malheureusement, une solution fondée sur l'instrumentation de l'ensemble des appareils n'est pas toujours envisageable en fonction de l'application étudiée. Différents obstacles peuvent être rencontrés. La stratégie envisagée dans le cadre de nos travaux est définie en fonction de ses contraintes et résulte d'un choix de l'entreprise.

a) Contraintes industrielles

➢ **Contraintes technologiques**

Idéalement, il suffit de mesurer la consommation des usages séparement. Mais, disposer d'un système de mesure centralisé et garantissant l'accès à l'ensemble des appareils constitue encore

un défi technologique à lever. Notons par ailleurs qu'un tel système de mesure (s'il existe) ne se prête pas à tous les problèmes industriels. Par exemple, instrumenter tous les appareils de froid alimentaire d'un hypermarché pour collecter les signaux nécessaires à leur caractérisation n'est pas possible à cause de la complexité de l'installation.

➢ **Contraintes commerciales**

Le retour d'expérience de différents pays (les Etats-Unis, la Finlande, etc.), ayant commercialisé des systèmes de caractérisation des usages électriques, montre qu'une solution fondée sur une mesure par appareil n'est pas économiquement viable dans les marchés de masse (*mass-market*). En effet, le client résidentiel ne souhaite pas d'intrusion chez lui pour installer un système de mesure. De plus, la solution proposée doit être à faible coût pour que la clientèle s'y intéresse réellement.

➢ **Contraintes juridiques**

Pour des raisons juridiques (en plus des contraintes de coût et de sécurité), une solution impliquant l'intervention technique au niveau du compteur électrique n'est pas envisageable.

b) Stratégie

Une source d'information permettant de s'affranchir des contraintes énoncées précédemment est la (les) grandeur(s) électrique(s), nommée(s) de l'installation globale disponible(s) au niveau du compteur électrique. Ces grandeurs électriques sont génériquement nommées **courbe de charge générale** (**CdC**) . En effet, ce signal véhicule l'*empreinte* (signature électrique) des différents appareils de l'installation.

1.1.4 Données d'entrée et connaissances annexes

Dans le cadre de cette thèse, EDF souhaite investiguer une solution compatible avec l'infrastructure existante c'est à dire utilisant les mesures dont la nature électrique et les caractéristiques (période d'échantillonnage, horizon d'observation) sont prédéfinies par la chaîne de mesure intégrée aux compteurs existants.

a) Données d'entrée : cas du secteur résidentiel

La CdC générale disponible en sortie du compteur bleu électronique (**CBE**) [LB06] (description détaillée en annexe (A)) est la puissance apparente échantillonnée à une période $T_e \sim 1$ s). C'est la puissance apparente cumulée par tous les usages sur un intervalle de temps qui correspond à la période d'échantillonnage. Le domaine temporel d'analyse choisi est la journée. Une illustration d'une CdC quotidienne d'une résidence d'un client particulier est présentée sur la figure (1.2).

b) Données d'entrée : cas des hypermarchés

Les données disponibles pour répondre à ce besoin sont la CdC générale (*puissance active et puissance réactive* $T_e = 100$ ms) ainsi que la CdC d'un ou de deux usages supplémentaires. La figure (1.3) représente un exemple de signaux électriques mesurés avec un wattmètre décrivant la puissance active totale d'un hypermarché, celle du froid alimentaire et celle des fours d'une boulangerie du site.

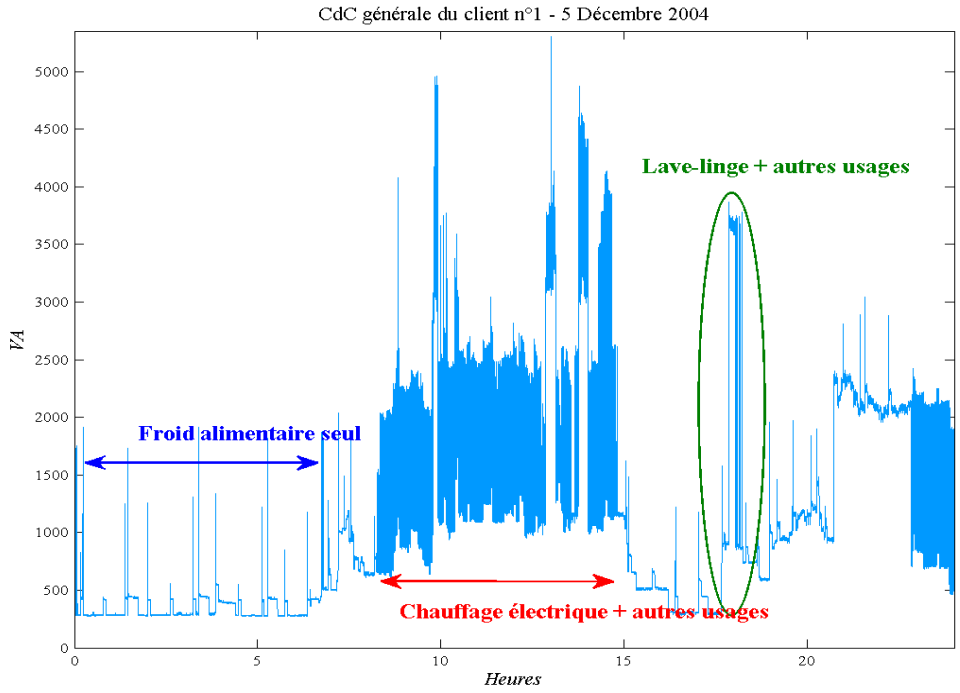


FIGURE 1.2: Exemple de CdC générale du secteur résidentiel (sortie du CBE)

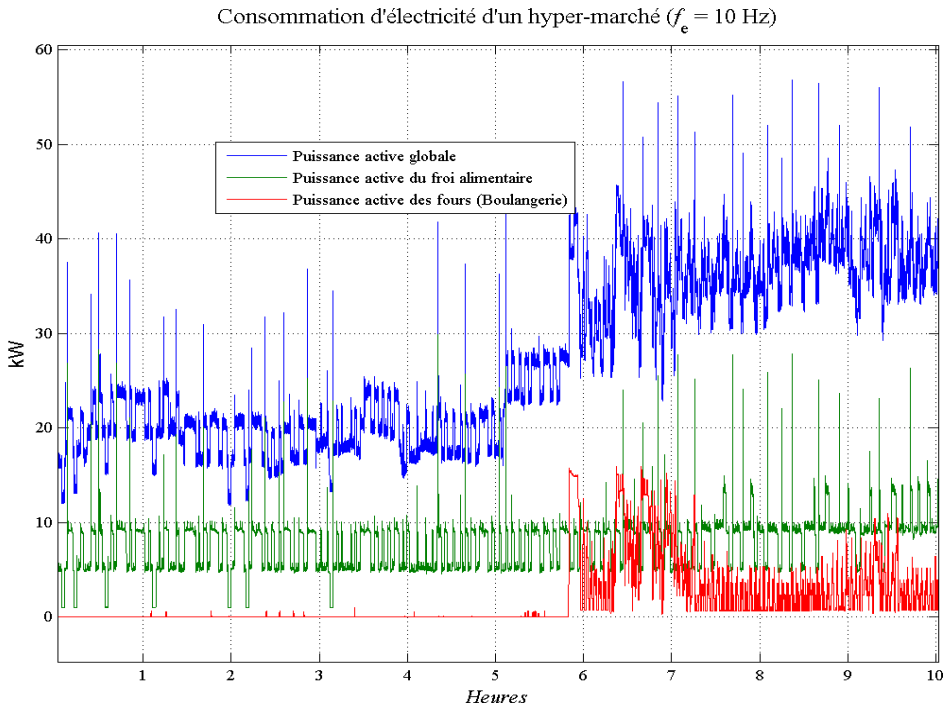


FIGURE 1.3: Exemple de CdC générale du secteur tertiaire (mesurée par un wattmètre)

c) Connaissances annexes

Aucune connaissance contextuelle n'est disponible pour répondre aux objectifs définis dans ce cas d'application. Mais des données issues de l'instrumentation des usages seuls sont disponibles, leur analyse est utile pour comprendre les signaux d'usages et définir les méthodes et modèles appropriés. De plus, nous montrerons lors de l'analyse de signaux (chapitre 2) que des caractéristiques intrinsèques (caractéristiques électrotechniques, modèle de fonctionnement) de certains appareils constituent des connaissances annexes précieuses pour la résolution du problème.

1.1.5 Synthèse

Le problème industriel considéré peut s'énoncer de la manière suivante : extraire un ensemble de descripteurs à partir d'une CdC générale (la puissance active quotidienne pour le secteur résidentiel et le couple (puissance active, puissance réactive) pour les hypermarchés). Les descripteurs recherchés sont de nature différentes : énergie, signaux élémentaires, scénario d'usage, etc. La solution envisagée doit être entièrement non-intrusive, automatique, compatible avec l'infrastructure existante et commercialement et techniquement viable à grande échelle.

1.2 Formalisation du problème de la *caractérisation* de la CdC

1.2.1 Notations et définitions utiles

Nous introduisons ici quelques notations et définitions.

- ➢ La CdC générale est notée $y(t)$ (de dimension 1 ou 2 en fonction du cas d'application), où t désigne les instants de mesure.
- ➢ La période d'échantillonnage est $T_e = 1/f_e$.
- ➢ Les CdCs élémentaires des appareils sont désignées par $x_i(t)$ et sont au nombre de N.
- ➢ L'énergie consommée (par jour) par l'installation globale est notée par la variable scalaire e_g, celle consommée par l'appareil n°i est notée e_i.
- ➢ Le nombre total des usages de la CdC est $K \leq N$.

Définition d'un événement de la CdC Un événement décrit la **transition** entre deux états électriques « stables », de durée Δt. En régime permanent, les évènements correspondent à des variations d'amplitude *significatives* de la CdC. Un événement est appelé « *on* » (« *off* ») quand la variation du signal est positive (négative).

Chaque événement est défini par l'instant (notée t_v) et la variation de l'amplitude signal (A_v) pendant la durée Δt. Le seuil de détection des événements A_{min} est choisi en fonction des usages étudiés et on a :

$$\begin{cases} \mid y(t+\Delta t) - y(t) \mid \geq A_{min} \\ A_v = y(t+\Delta t) - y(t) \end{cases} \tag{1.1}$$

1.2.2 Une définition de la « *caractérisation de la CdC* » revisitée

L'étude du besoin industriel (et l'analyse des travaux antérieurs présentée en (1.3)) montre(nt) :
- ➢ que la CdC désigne des signaux de nature électrique (tension, intensité, puissance moyenne, etc.) et de propriétés très différentes principalement dues au choix de la fréquence d'échantillonnage et au système de mesure ;
- ➢ que les informations recherchées présentent une grande diversité ;
- ➢ que de nombreux objectifs sont sous-jacents à la « caractérisation de la CdC » : problème de décomposition de la CdC, problème d'estimation de paramètres, problème de détection de présence d'usage, etc.

Nous proposons une définition du « problème de caractérisation d'une CdC » fondée sur une classification des grandeurs recherchées.

a) Classification des grandeurs recherchées

Nous proposons de regrouper les grandeurs pertinentes dans le cadre de notre problématique industrielle en cinq classes.
❶ *Les grandeurs énergétiques : énergie consommée par usage ou par appareil sur un horizon de temps donné* (*par exemple la journée dans le secteur résidentiel*).

❷ *Les grandeurs temporelles : support temporel de la courbe de charge de chaque appareil (ou usage), les scénarios de gestion* (*fonctionnement*) *des usages.*

❸ *Les grandeurs événementielles : série des événements des usages étudiés ;*

❹ *Les caractéristiques des appareils présents* : *puissance nominale de fonctionnement, modèle de fonctionnement, mode de régulation, nombre d'appareils d'un usage donné.*

❺ *Les signaux élémentaires des différents appareils électriques.*

b) Définition du problème de « caractérisation de la CdC »

Le problème de caractérisation de la CdC générale est un problème de décomposition d'un mélange monovalué (*un seul capteur*) *dans le domaine d'observation direct ou dans un domaine de projection pré-défini. Les « composantes constitutives » recherchées sont* : *les énergies consommées par les usages, les signaux temporels, les grandeurs événementielles, les caractéristiques de l'appareil ou de l'usage et les scénarios de fonctionnement.*

Cette défintion nous amène naturellement à poser le problème de la manière suivante.
Si $y(t)$ est la CdC générale d'une installation donnée mesurée à la fréquence d'échantillonnage f_e (pendant une journée par exemple) et de dimension $(k \times M)$ avec $k = 1$ ou 2, alors le modèle de mélange et la relation de conservation de l'énergie «consommée» sont décrits par le système suivant (1.2) :

$$\begin{cases} y(t) = \sum_{i=1}^{N} x_i(t) + b(t) \\ e_g = \sum_{i=1}^{N} e_i + e_b \end{cases} \tag{1.2}$$

où

- $x_i(t)$ est la CdC d'un appareil électrique, $b(t)$ est un bruit blanc gaussien centré de variance σ_b^2 (modélisant le bruit de mesure) ;
- e_g (respectivement e_i) est l'énergie consommée par le système global (respectivement l'appareil n°i) ;
- e_b est l'énergie associée au bruit de mesure $b(t)$.

Décomposer la CdC par usages dans un espace prédéfini consiste à définir un opérateur D tel que :

$$D(y(t)) = \Theta$$

où Θ désigne une des grandeurs définies précédemment,

1.2.3 Les problèmes sous-jacents à la caractérisation de la CdC

Le problème général de caractérisation de la CdC s'inscrit dans le cadre générique des problèmes inverses où les estimées recherchées sont, soit des paramètres des sources du mélange, soit les signaux élémentaires dans un domaine de représentation adapté. Dans la suite, nous situons notre problème initial dans les trois domaines suivants : la détection, la classification et l'estimation (estimation de paramètres ou reconstruction de signaux) que nous présentons ci-après.

a) Segmentation de la CdC

Dans le cas où l'objectif est de détecter la présence d'un usage et de déterminer le cas échéant les plages horaires de fonctionnement des usages d'intérêt : c'est un problème de détection. Un schéma de principe de cette sous-problématique ainsi qu'une illustration des entrées-sorties dans le cas du secteur résidentiel sont présentés sur la figure (1.4). Sur cette figure, $\mathscr{D}_b$ est l'opérateur de segmentation de la CdC. Les sorties sont des grandeurs vectorielles binaires (décrivant les plages horaires de fonctionnement) désignées par la matrice Θ_b $(M \times N)$ où M définit la taille du détecteur (nombre d'échantillons correspondant au domaine temporel de l'analyse de la CdC). Sur l'exemple présenté, les deux usages détectés sont le chauffage par convecteurs (**CHC**) et l'eau chaude sanitaire (**ECS**).

Nous montrerons lors de l'analyse des signaux que la « notion de plage de fonctionnement d'un usage » est difficile à définir de manière unique et dépend du mode de régulation de l'appareil. Par exemple, le signal d'un convecteur, de par sa régulation périodique, s'annule alors que l'appareil est en marche. C'est aussi le cas des compresseurs (**CMP**) du froid alimentaire ou de la pompe à chaleur (**PAC**). Ceci devient encore plus délicat quand l'usage étudié, par exemple le chauffage par convecteurs (**CHC**), est le foisonnement de plusieurs appareils (convecteurs) dont les scénarios d'usage sont indépendants entre eux. Dans le secteur résidentiel, nous introduisons une durée de fonctionnement minimale « détectable » par usage τ_{pres}^{min} . Ainsi, nous pouvons définir précisément la date de début et de fin de fonctionnement d'un usage. La justification et les choix effectués sont détaillés dans le chapitre 2.

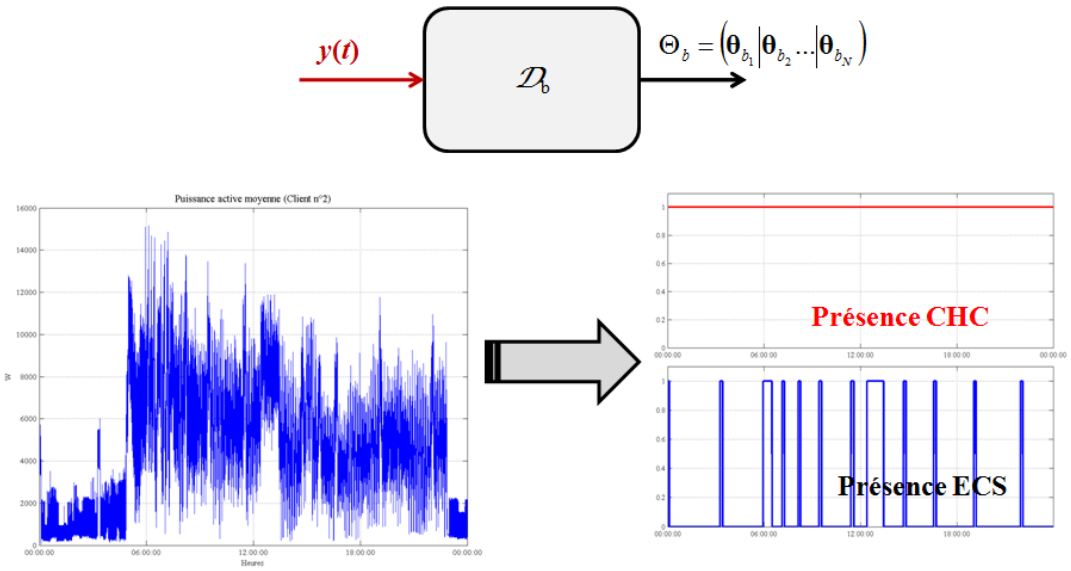


FIGURE 1.4: Décomposition de la CdC générale vue comme un problème de cartographie des scénarios des usages

b) Classification des signaux de la CdC

Le problème de reconnaissance des événements de la CdC par les usages est un problème de classification, où l'objectif est d'associer à chaque usage k la série d'événements $\theta_{v_k} = (t_v^{(k)}, A_v^{(k)})$ à partir de l'observation $y(t)$. Le schéma de principe de cette sous-problématique ainsi qu'une illustration des entrées-sorties (cas du secteur résidentiel) sont décrits sur la figure (1.5). Sur cette figure, $\mathcal{D}_v$ est l'opérateur de classification (et si possible d'appariement (*associer un « on » au « off » correspondant*) des événements les par usage.

c) Estimation de paramètres d'usages à partir de la CdC

Ici, nous considérons les questions relatives à l'estimation de l'énergie consommée par chaque usage, le nombre d'appareils d'un usage ainsi que les caractéristiques de l'appareil (amplitude par exemple). Les illustrations du problème d'estimation de l'énergie consommée par usage et de celui de l'estimation des caractéristiques des appareils sont respectivement présentées sur les figures (1.6) et (1.7).

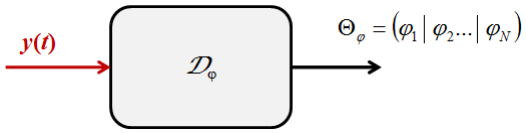


FIGURE 1.7: Décomposition de la CdC générale vue comme un problème de description « physique » des usages

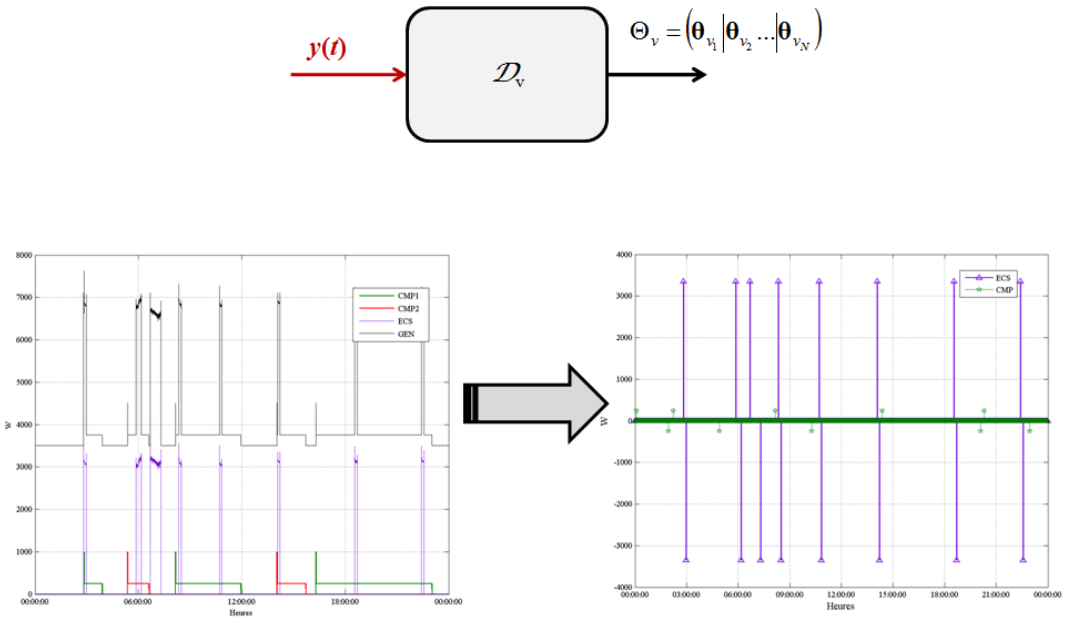


FIGURE 1.5: Décomposition de la CdC générale vue comme un problème de séparation des séries des événements

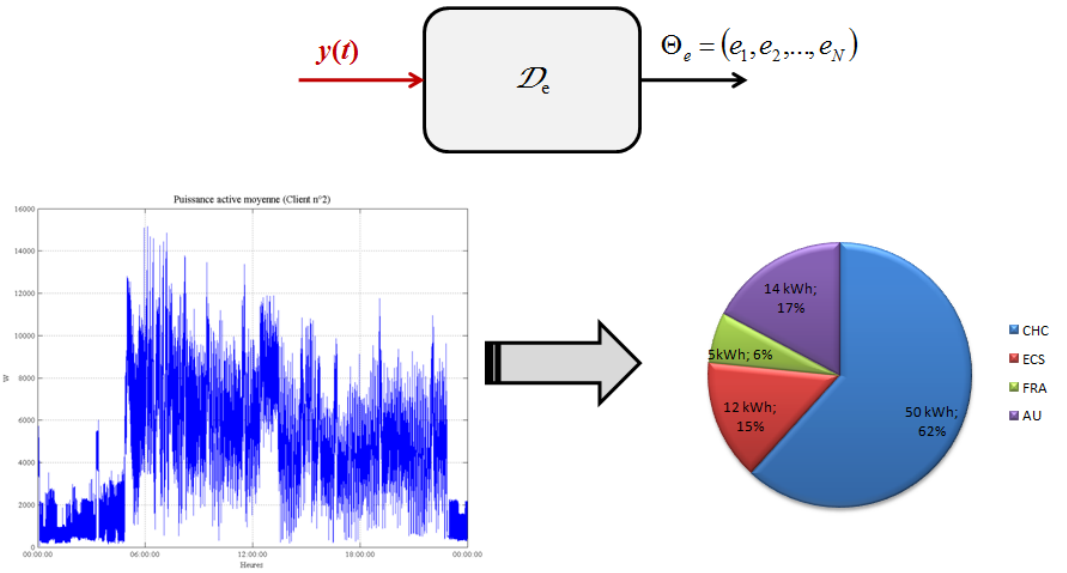


FIGURE 1.6: Décomposition de la CdC générale vue comme un problème de cartographie de la consommation

Sur la première figure, Θ_e $(1 \times N)$ où N est le nombre d'usages étudiés et $\mathscr{D}_e$ représente un (ou plusieurs) estimateur(s) du (des) descripteurs recherchés. Sur la seconde, $\mathscr{D}_\varphi$ représente schématiquement l'ensemble des fonctions et domaines transformés à définir pour répondre au problème. Dans le cadre de la thèse, nous traiterons :

- ➢ l'estimation des amplitudes des usages suivants : froid alimentaire (résidentiel ou dans un hypermarché), les convecteurs électriques et le chauffe-eau ;
- ➢ l'estimation du nombre de composantes des usages suivants : froid alimenatire (résidentiel ou dans un hypermarché) et les convecteurs électriques.

d) Reconstruction des signaux de la CdC

L'objectif est de reconstruire les CdCs (ou une partie) de la CdC générale. Les inconnues sont les signaux associés aux usages étudiés. La figure (1.8) synthétise schématiquement cette sous-problématique, où $\mathscr{D}_s$ représente un ou plusieurs opérateurs pour reconstruire les signaux du mélange. Un exemple de signal observé et des signaux à estimer sont également décrits sur la figure (1.8).

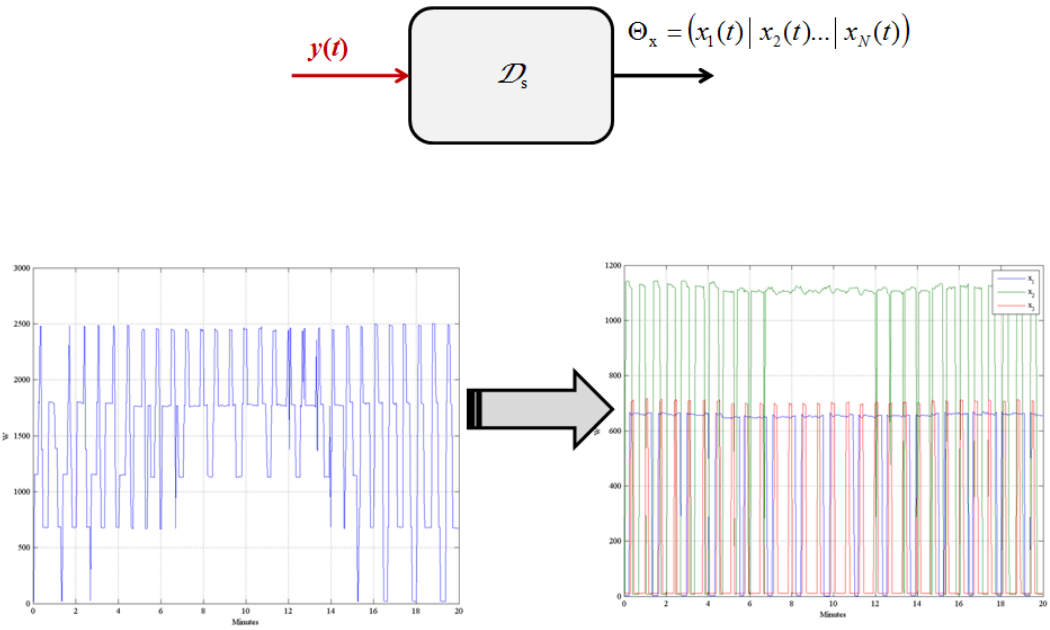


FIGURE 1.8: Décomposition de la CdC générale vue comme un problème de reconstruction de signaux des usages

1.3 État de l'art

La caractérisation de la CdC est un vaste sujet qui a fait l'objet de nombreuses études. Nous nous limitons ici à une présentation des méthodes et modèles dédiés à la caractérisation de la CdC telle que définie précédemment. La recherche bibliographique n'a pas permis d'identifier de travaux sur ce problème avec les mêmes spécifications et contraintes (données d'entrée et absence de phase d'apprentissage). Nous proposons de faire une synthèse des travaux les plus proches de notre problématique, au sens de la séparation de composantes de la CdC à partir d'un mélange unique. La présentation des approches existantes s'appuie sur des critères de classification de ces méthodes.

Les méthodes de caractérisation des usages électriques, sans exploiter la CdC, sont résumées en annexe C.

1.3.1 Historique

Afin de caractériser les usages domestiques ou tertiaires, **George Hart** [Har89] et ses collaborateurs ont proposé une démarche semi-intrusive de caractérisation des usages à partir de l'analyse de la CdC globale. En effet, les appareils se différencient par leur caractère résistif, inductif ou capacitif. Cette « signature électrique » de chaque appareil est également véhiculée par les grandeurs électriques du système.

Les premiers travaux sur la caractérisation de la CdC cherchaient à détecter les variations de courant lors de la mise en marche ou de l'arrêt d'un appareil électrodomestique.

Les principales mesures retenues au début de ces travaux étaient les variations de la puissance active $\Delta P(t)$ et les variations de la puissance réactive $\Delta Q(t)$. La représentation de ces signaux dans le domaine $(\Delta P, \Delta Q)$ permet de mettre en évidence la catégorie des charges électriques étudiées et d'estimer leurs puissances nominales de fonctionnement [Har92].

La dimension temporelle est prise en compte en introduisant la notion d'**événement** : en régime permanent il s'agit de la **transition** entre états électriques « stables » caractéristiques du fonctionnement d'un appareil. Par conséquent, un événement est défini par (1.3) :

$$e_v(t_r) = (t_r, \Delta P(t_r), \Delta Q(t_r)) \tag{1.3}$$

où t_r est l'instant auquel l'événement a eu lieu, $\Delta P(t_r)$ et $\Delta Q(t_r)$ sont respectivement la variation de la puissance active et la variation de la puissance réactive observées à cet instant.

Une approche aveugle de décomposition de la CdC électrique est fondée sur le décodage de la séquence des événements observés en régime permanent [Har89]. George Hart [3] a proposé certaines techniques de décodage en s'appuyant sur une modélisation par machines à états finis (Finite State Machines). Chaque appareil est modélisé par une machine à états, une recherche du plus court chemin permettant d'effectuer le décodage des événements [Har94].

C'est ainsi qu'est née la démarche de décomposition non-intrusive de la courbe de charge électrique en Amérique du Nord [Har92, Har94] et en France [BC93]. Cette démarche commune se résume en trois modules de traitement dont l'orchestration et la technique de mise en œuvre varient d'une méthode à l'autre :

3. et ses collègues Lawrence Carmichael de l'EPRI et Mark Malmendier de Telog Instrument

➢ **Etude des événements : détection et reconnaissance (classification) des événements ;**
➢ **Identification des appareils (ou classes d'appareils) (appariement des événements) ;**
➢ **Estimation de la courbe de charge par usages.**

Notons d'ores et déjà que cette démarche proposée par G. Hart est à notre connaissance la seule et unique démarche qui a été retenue malgré l'évolution des méthodes de mise en œuvre, quelle que soit la fréquence d'échantillonnage des grandeurs électriques (allant de quelques millisecondes à quelques minutes). Ce point sera mis en évidence à travers une présentation détaillée de quatre études sur la CdC du secteur résidentiel et du secteur tertiaire *(*à différentes périodes d'échantillonnage*)*. Les différences entre les méthodes concernent souvent la manière dont on compense le manque d'information pour résoudre le problème d'estimation des usages :

➢ *en intégrant des connaissances a priori sur la structure des signaux temporels, la forme des signaux, les propriétés de symétrie ou encore des connaissances a priori sur l'utilisation de certains appareils (fréquence d'utilisation, plage horaire)* [Bon96, Boz00, BV03, BV04, Laj02] ;
➢ *ou en exploitant une bibliothèque de signaux d'appareils seuls* [Lee93, Don02, LMLS93, Lee03].

Remarque

Les principaux travaux menés sur la caractérisation des composantes constitutives de la consommation générale d'une installation électrique à partir de la CdC générale (courant / tension / puissance) ont été menés aux Etats-Unis (MIT, EPRI, Université de Californie), en France (EDF-R & D) et au Canada (Université de Concordia). Ces travaux ont souvent donné lieu à des produits commercialisés. D'autres études en Allemagne (Université de Paderborn), au Japon (CRIEPI) et en Finlande (VTT Energy) ont abouti à des logiciels d'analyse et de décomposition de la consommation électrique ou à des appareils de mesure de consommation par usage (sous-comptages électriques). Les performances annoncées seront détaillées dans la section (1.3.3).

1.3.2 Critères de classification des études antérieures

Nous choisissons les quatre critères suivants pour classer les études portant sur la « caractérisation de la CdC générale »

➢ la classe de la méthode ;
➢ les données d'entrée ;
➢ le caractère supervisé / non-supervisé ;
➢ le degré d'intrusion toléré.

La performance et la robustesse des méthodes ne constituent pas un critère à part entière à cause du manque d'informations sur les performances des algorithmes développés.

Remarque : caractère intrusif, semi-intrusif ou non-intrusif d'une solution

Le dernier critère de classification étant subjectif, nous présentons des précisions sur la qualification des méthodes en fonction de ce critère.

Les approches intrusives

Les approches intrusives sont celles qui nécessitent d'instrumenter l'ensemble (ou une partie) des usages pour chaque installation étudiée.

Les approches semi-intrusives

Il s'agit des méthodes de caractérisation de la CdC pour lesquelles l'obtention d'une (ou plusieurs)
entrée(s), nécessaire(s) au bon fonctionnement de la méthode, induit une intrusion chez le client : information demandée directement au client (questionnaire par exemple), instrumentation dédiée mise en place chez le client pour réaliser une phase d'apprentissage (constituer une bibliothèque de signaux électriques) ou encore une intervention sur le site pour réaliser un audit du bâtiment.

Les approches non-intrusives

Les approches non-intruives exploitent uniquement les données de consommation globales et ne nécessitent ni une intervention sur site ni les mesures de sous-comptages de l'installation étudiée.

1.3.3 Méthodes et modèles proposés

Bien que les degrés d'avancement des travaux soient très hétérogènes, nous les regroupons selon les critères définis précédemment. En effet, ce qui nous intéresse en premier lieu sont les différentes méthodologies et démarches adoptées pour résoudre le problème de décomposition de la CdC par usage.

Quatre classes d'approches ont ainsi été identifiées. Un présentation succincte de chaque méthode est fournie dans les sections suivantes. Le lecteur peut se référer à une présentation détaillée de ces méthodes que nous proposons dans l'annexe (C).

a) Modèles à base de machines à états finis

La CdC est vue comme le foisonnement des événements générés par les différents appareils. Chaque usage est alors modélisé par une machine à états finis. Deux méthodes sont proposées pour décoder les événements. La première méthode, développée par Marc Bons dans le cadre de sa thèse [Bon96], consiste à modéliser la CdC résidentielle (puissance active et puissance réactive, $T_e = 1\ s$) par des Chaînes de Markov Cachées. Le lecteur pourra se référer à l'annexe (C) pour une description plus détaillée. La deuxième méthode, proposée par une équipe de l'Université de Paderborn en Allemagne, utilise un algorithme génétique pour résoudre le problème de décodage des événements de la CdC résidentielle (puissance active, $T_e = 1\ s$). Nous détaillons la méthode, ses avantages et ses limites dans l'annexe (C) de ce manuscrit.

b) Approche inverse probabiliste pour la décomposition de CdC des hypermarchés

Cette méthode a été développée dans le cadre d'une collaboration entre EDF/R&D et l'Ecole des Mines de Paris [Laj02]. L'objectif est de fournir une décomposition de la CdC d'un hypermarché donné à partir de CdC annuelle (Puissance active, $T_e = 10$ mn). Des données météorologiques sont également disponibles.

Un modèle stochastique de la CdC des hypermarchés est proposé [Laj02]. La méthode s'appuie sur une modélisation paramétrique des usages dans un domaine transformé (heures x semaines x années) à base d'un dictionnaire de formes géométriques. Une illustration est présentée sur la figure (1.9). L'espace des paramètres du modèle étant de très grande dimension, ce dernier

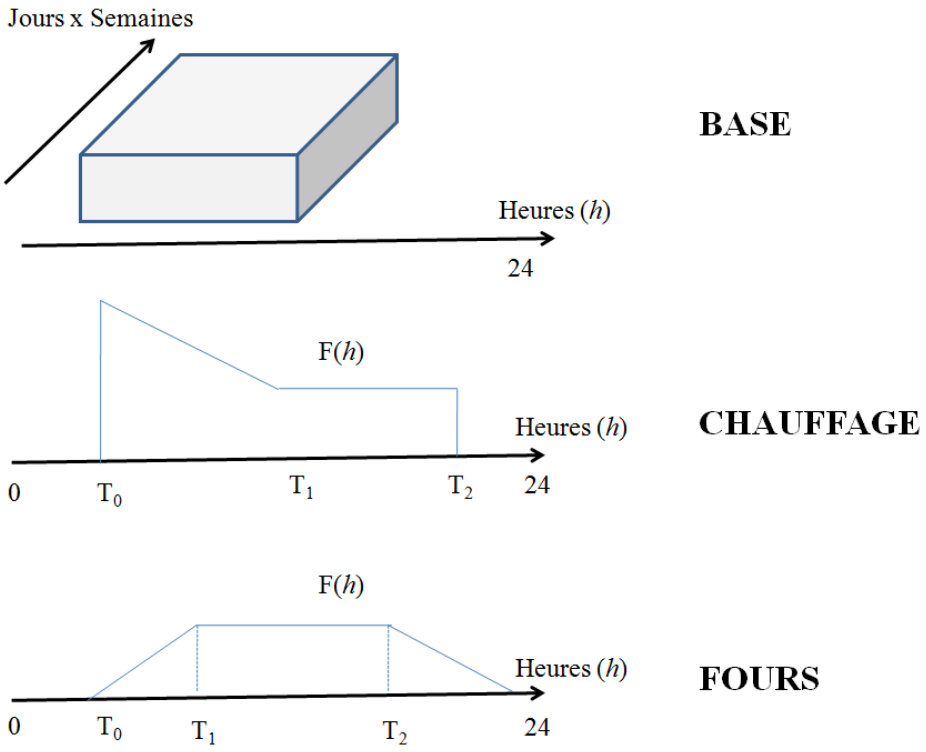


FIGURE 1.9: Modèles des composantes de la CdC d'un hypermarché [Laj02]

est exploré par une méthode d'échantillonnage stochastique issue du domaine de la simulation bayésienne. Nous détaillons le modèle proposé et l'algorithme mis en œuvre dans l'annexe (C).

c) Méthodes de reconnaissance de forme pour la séparation des usages

Des critères discriminants entre usages, ou des caractéristiques de certains usages, sont définis à partir d'une base de signaux d'apprentissage. Un algorithme utilisant des règles d'association permet d'identifier les usages à partir de l'analyse de la puissance active ($T_e = 1$ s) [Far99]. Une description plus détaillée est fournie dans l'annexe (C).

Ces méthodes principalement proposées par l'Université de Concordia (Canada) ont abouti au développement de deux systèmes :

➢ NALRA (Non-intrusive Appliance Load Recognition Algorithm) ;
➢ RPRA (Rule based Pattern Recognition Algorithm) .

d) Les méthodes NIALM (détection-classification)

Les méthodes NIALM (Non-Intrusive Appliances Load Monitoring) sont des méthodes de reconnaissance des usages développées au sein du LEES [4] au MIT [5] en collaboration avec l'EPRI [6]. « NIALM » est en effet un label déposé.

4. **L**aboratory for **E**lectromagnetic and **E**lectronic **S**ystems
5. Massachusets Institute of Technology
6. Electric Power Research Institute

	Données d'entrée	Période d'échantillonnage	Apprentissage	Méthodes	Secteur	Performances
Famille des NIALM	Courant et tension (+ éventuellement les puissances harmoniques)	quelques µs à quelques s	Oui	Détection, classification, reconnaissance de forme	Résidentiel et tertiaire	5% à 10% pour l'usage HVAC (Heating Ventilation and Air -Conditioning)
HELP	Audit du site (mesures) + Puissance active globale	1 h	Oui	Arbre de décision	Résidentiel	10%
Familles des NALRA et RPRA	Puissance active globale (éventuellement Q) + audit (questionnaire)	1s	Oui	Règles d'association	Résidentiel	-
Travaux de l'université de Paderborn	Puissance active en sortie du compteur	1s	Oui	Modélisation de la CdC par machines d'état + décodage par un algorithme génétique	Résidentiel	-
Modèles à sources markoviennes cachées (EDF)	Puissance active et puissance réactive	1s	Oui	HMM	Résidentiel	-
Décomposition de la CdC par approche statistique	Puissance active	10mn	Non	Modélisation paramétrique des usages + Algorithme de recuit simulé	Tertiaire	-

TABLE 1.1: Résumé des caractéristiques des méthodes non-intrusives

Ces méthodes consistent souvent en trois étapes communes : détection des événements, classification puis estimation de l'énergie consommée par usage. Les méthodes mises en œuvre sont classées en trois familles en fonction de la grandeur électrique étudiée, plus précisément en fonction du régime de fonctionnement des signaux exploités. La première méthode s'intéresse à l'analyse des variations des puissances active et réactive en régime permanent ($T_e = 1$ s). La seconde exploite les signaux de courant et de tension en régime transitoire ($T_e = 100$ µs). La troisième famille d'algorithmes est hybride : la détection des événements est souvent effectuée à partir des signaux en régime permanent, la reconnaissance des usages exploite les signaux en régime transitoire. Une présentation détaillée des trois familles de méthodes est fournie en annexe (C) de ce mémoire.

Récapitulatif des travaux antérieurs : méthodes non-intrusives ou faiblement intrusives
Le tableau (1.1) décrit les caractéristiques de l'ensemble des méthodes semi-intrusives et non-intrusives de décomposition de la CdC générale. Nous y précisons les informations suivantes : les données d'entrée utilisées et la période d'échantillonnage, la nécessité ou non d'une phase d'apprentissage, les algorithmes développés, le secteur étudié et les performances quand elles sont fournies.

Remarque

Le système **HELP** n'a pas été décrit lors de la présentation des méthodes existantes par manque d'informations précises sur les techniques mises en œuvre.

Synthèse

Notre problématique industrielle consiste à caractériser la CdC générale disponible en sortie du compteur électrique afin d'extraire un ensemble de descripteurs des principaux usages résidentiels ou tertaires (cas des hypermarchés). La solution doit être entièrement non-intrusive et compatible avec les choix du commercialisateur (simplificité de mise en oeuvre, fiabilité de la solution proposée). Nous avons identifié trois classes de problèmes sous-jacents au problème initial à savoir :

- ➢ la segmentation de la CdC générale ;
- ➢ l'estimation de paramètres caractéristiques des usages à partir du mélange observé ;
- ➢ la classification des composantes du mélange ;
- ➢ et enfin la reconstruction des sources élémentaires du mélange.

Quatre familles d'approches dédiées à la caractérisation de la CdC générale (pas nécessairement la même que nous traitons dans la thèse) ont été identifiées. La première classe de méthodes s'appuie sur des modèles de machines à états finis, où le modèle de fonctionnement des appareils et leurs modèles électrotechniques permettent de les séparer. La seconde classe de méthodes s'affranchit de la sous-détermination du problème posé en proposant un modèle probabiliste du mélange et en intégrant des connaissances *a priori* sur les composantes à estimer. Les deux dernières approches s'inscrivent dans le cadre de la reconnaissance de forme et exploitent souvent une signature apprise de l'appareil pour la reconnaissance des usages. Bien que ces méthodes ne soient pas applicables en l'état à notre problème, elles sont transposables à notre cas d'étude. Plus précisément, nous pouvons d'ores et déjà envisager d'adopter une approche bayésienne proposée par [Laj02] sous réserve de définir un modèle des signaux à extraire et de pouvoir construire des *a priori* sur les paramètres du modèle. Une deuxième approche fondée sur les méthodes de reconnaissance de forme non pas par apprentissage de la signature exacte de l'appareil, mais en définissant des représentations adaptées est également à étudier.

Le problème générique est celui de la caractérisation des composantes d'un signal à partir d'une seule observation. Il est nécessaire de compléter notre connaissance du problème à travers des données contextuelles, des données d'expertise ou encore des données d'apprentissage. L'extraction de ces connaissances à partir des données réelles est détaillée dans le chapitre (2). Les informations extraites à partir des signaux guideront le choix des méthodes adaptées au problème considéré dans cette étude.

Chapitre 2

Analyse des signaux en vue de l'extraction de connaissances sur les usages

Je ne cherche pas à connaître les réponses,
je cherche à comprendre les questions.
Confucius

Introduction

La caractérisation de la courbe de charge électrique à partir de grandeurs électriques mesurées au niveau du système global est un problème mal posé. Nous avons vu que pour répondre à cette question, il faut compléter notre connaissance du problème. C'est l'objectif de cette partie du mémoire. Nous présenterons en effet la démarche expérimentale adoptée afin de collecter le maximum d'information sur les signaux du mélange. Ces informations sont issues soit de mesures soit des experts dans le domaine des usages électriques ciblés. Une analyse des signaux élémentaires est également présentée. Les résultats de cette phase d'analyse guideront le choix d'approches possibles pour caractériser les composantes du mélange observé.

2.1 Démarche expérimentale

Rappelons :

- ➢ que le besoin d'EDF est de disposer d'information (énergétique, évenementielle, etc.) sur les composantes de la CdC ;
- ➢ que l'on s'intéresse à la CdC dans l'habitat résidentiel, mais également à la CdC d'un cas particulier du secteur tertaire (hypermarché) ;
- ➢ et enfin que les données d'entrée sont imposées par des contraintes de l'infrastructure existante (mesure disponible en sortie du compteur bleu électronique par exemple).

Nous avons vu que le problème est mal posé et qu'il faut compléter notre connaissance des signaux élémentaires et du mélange observé. La démarche adoptée dans le cadre de ce travail est la suivante :

- ➢ se restreindre aux données d'entrée disponibles ;
- ➢ résoudre le problème sous les contraintes liées aux besoins industriels ainsi qu'à l'infrastructure actuelle ;
- ➢ identifier les limites de performance des algorithmes ainsi que les limitations en terme d'extraction de descripteurs qui seraient liées aux caractéristiques des données mesurées ;
- ➢ identifier les améliorations possibles en termes de spécification des données d'entrée ainsi que les potentielles informations contextuelles utiles à la caractérisation de la CdC.

Cette démarche s'est concrétisée par de nombreuses campagnes de mesure afin de collecter des signaux d'usages électriques pour les deux applications envisagées. Dans la suite nous décrivons la base de données collectées ainsi que leur principales caractéristiques.

Précisons que l'on choisit de travailler sur des signaux réels au lieu de simuler des signaux à partir de modèles électrotechniques des appareils électriques ciblés.

La description des données est faite séparément afin de souligner les différences entre les deux secteurs ciblés.

2.1.1 Description des données : cas du secteur résidentiel

Nous nous intéressons aux courbes de charge mono-site (un client particulier). L'horizon d'observation est d'une journée.

Les autres propriétés de la CdC (grandeur électrique mesurée et période d'échantillonnage) sont résumées dans le tableau (2.1). Quatre campagnes de mesure sont décrites :

- ➢ K_1 : Puissance apparente d'une maison enregistrée pendant 5 semaines. Aucune information sur les puissances des appareils individuels n'est disponible. Ces données ont principalement servi dans le cadre du stage de master précédant ce travail de thèse [DWEG05].
- ➢ K_2 est une campagne de mesure spécifique dans une maison expérimentale (Maison Domotique) d'EDF-R&D. Tous les appareils de la maison ont été instrumentés. Des scénarios d'utilisation de certaines appareils (chauffe-eau, aspirateur, télévision, etc) ont été programmés. Cette campagne présente l'avantage de nous fournir des signaux des appareils individuels, mais les scénarios de consommation se sont révélés *in fine* pas réalistes.
- ➢ K_3 et K_4 sont deux campagnes de mesure lancées d'une part pour disposer de données pour valider les algorithmes développées (K_3) et d'autre part pour élargir la base des exemples de signaux des usages ciblés à la fréquence type de la téléInfo mais également à une fréquence plus élevée pour une meilleure caractérisation des signaux.

Ce tableau met en évidence deux problèmes :

- ➢ le nombre limité de cas-test :
  - – absence de retour d'expérience sur les données au démarrage de cette étude ;
  - – la complexité de mettre en place une instrumentation d'usages chez un client particulier ;
- ➢ la difficulté de disposer de courbes de charge électriques réelles des usages :
  - – difficulté technique (chaîne d'acquisition à définir) ;
  - – intrusion chez les clients expérimentaux pour disposer des sous-comptages électriques ;
  - – contraintes juridiques liées à la séparation entre le distributeur et le commercialisateur.

Remarque

EDF R&D envisage d'élargir la base de données à une cinquantaine de clients courant 2010 afin de disposer de données plus représentatives.

Source des données	Capteur	Grandeur (s) mesurée(s)	Période d'échantillonnage T_e	Remarques	Campagne de mesures
Site – clients particuliers	Compteur Bleu Electronique (CBE)	**Puissance apparente** (P_{app} de la sortie télé-Info)	1.3 sec (en moyenne)	* Seule Papp est disponible. * Nombre de clients : 1	K_1 : Décembre 2004 à Janvier 2005
Maison expérimentale – Maison Domotiques (EDF R&D – Les Renardières)	CBE + logiciel d'acquisition multivoies dédiée (PATERN)	•**Puissance moyenne** •Température ambiante (dans les chambres •Tension globale	• 2 sec • 2 min • 2 min	* Tous les appareils de la maison ont été instrumentés. * PATERN est un système développé par EDF R&D	K_2 : Janvier à Février 2006
Site – clients particuliers « sites-pilotes »	CBE + Un appareil de mesure dédié conçu par EDF R&D	•**Puissance apparente** •**Intensité et Tension instantanées** du général et de sept usages	• 1.3 sec en moyenne • 100 µsec	Nombre de clients : 2	K_3 : Décembre 2007 à aujourd'hui
Mesures « ponctuelles » chez des clients particuliers	Un appareil de mesure dédié conçu par EDF R&D	**Intensité et Tension instantanées d'un ou de plusieurs appareils**	• 100 µsec ou • 200 µsec	Une vingtaine de clients (agents EDF)	K_4 : Avril 2007 à aujourd'hui

TABLE 2.1: Caractéristiques des données d'entrée - CdC dans le secteur résidentiel

2.1.2 Description des données : cas du secteur tertiaire (hypermarché)

Spécificités du secteur tertiaire

Bien que l'objectif de caractérisation de la CdC soit commun aux deux secteurs, les contraintes liées à l'infrastructure sont beaucoup moins importantes dans le secteur tertiaire. En effet, les données d'entrée sont la CdC globale ainsi que deux à trois usages prédéfinis en fonction des objectifs recherchés. Notons deux particularités de ce secteur :

➢ Imprécisions sur les usages instrumentés :
- Compte tenu de la complexité des installations électriques dans ce secteur, des appareils autres que l'usage ciblé par l'instrumentation peuvent se trouver sur la même voie (au niveau du compteur électrique). Par conséquent, la CdC d'un sous-comptage est entachée d'une imprécision quant au contenu.
- Des informations sur les caractéristiques des sous-comptages ou leur gestion sont souvent fournies par l'électricien. Mais, elles sont souvent imprécises.

➢ Durée de la collecte des données :
- La durée d'enregistrement des données est souvent de quelques mois à quelques années. Mais, ces CdC ne sont pas échantillonnées à une période suffisamment faible pour en déduire des informations sur les usages telles que la puissance nominale ou le scénario de gestion.
- Un système de mesure dédié (wattmètre) étant nécessaire à la collecte des données, l'auditeur du site intervient ponctuellement pour collecter des mesures de CdC globales et de quelques sous-comptages sur une durée de quelques jours.

Choix de la période d'échantillonnage

La période d'échantillonnage des signaux a été choisie de manière empirique compte tenu des deux contraintes suivantes :

- ➢ on doit disposer d'une signature électrique « fine » des appareils pour détecter notamment tous les enclenchements et les déclenchements des appareils électriques ;
- ➢ le volume de données acquises doit être facile à gérer (acquisition avec un PC portable).

La période d'échantillonnage choisie est $T_e = 100$ ms.

Usages étudiés et grandeurs électriques mesurées

Etant donné que les usages ciblés sont de natures différentes (Fours : dominante résistive, Compresseurs : dominante inductive), les grandeurs mesurées sont la puissance active et la puissance réactive, moyennées sur une durée égale à T_e.

Informations complémentaires

Nous avons des informations complémentaires sur les usages : plages horaires de fonctionnement, le processus définissant le fonctionnement de l'usage ciblé.

2.2 Analyse des signaux

Cette section est dédiée à la présentation des signaux électriques des appareils électro-domestiques et ceux d'un hypermarché. Le choix d'une présentation du type « catalogue » est fait d'une part pour que le lecteur se familiarise avec ces signaux et d'autre part pour mettre en évidence les caractéristiques de chaque classe d'appareils. Nous présenterons d'abord les signaux issus de différentes expérimentations par type d'usage, puis des exemples de mélanges que nous aurons à caractériser.

2.2.1 Définitions et notations

Enclenchement - Déclenchement

Pour un appareil fonctionnant en «tout ou rien», nous désignerons par *enclenchement* (respectivement *déclenchement*), le passage de l'état de puissance moyenne minimale nulle (respectivement puissance moyenne non nulle) à une puissance moyenne non nulle (respectivement à une puissance moyenne quasiment nulle). Nous utiliserons souvent les raccourcis respectifs suivants *on* et *off*. Nous distinguerons ces terminologies de la mise en marche et de l'arrêt d'un appareil qui ne coïncident pas nécessairement avec les *on* et *off*.

Événement

Rappelons qu'un événement (A_i, t_i) ayant lieu à l'instant t_i et d'amplitude signée A_i définit soit un enclenchement soit un déclenchement. Afin de décrire les événements de la CdC, nous

introduisons les trois séries suivantes : la série des «*on*», la série des «*off*» et la série globale des évènements dénommée série des «*on-off*». La définition de chaque série est donnée ci-après.

La série des évènements «*on*»

Cette série est décrite par la donnée de deux suites : la suite $(t_i^+)_{i\in\mathbb{N}}$ et la suite $(A_i^+)_{i\in\mathbb{N}}$ sont respectivement la suite des instants (consécutifs) des événements du type *on* et la suite des amplitudes associées.

La série des évènements «*off*»

Cette série est décrite par la donnée de deux suites : la suite $(t_i^-)_{i\in\mathbb{N}}$ et la suite $(A_i^-)_{i\in\mathbb{N}}$ sont respectivement la suite des instants (consécutifs) des événements du type off et la suite des amplitudes associées.

La série globale des évènements «*on-off*»

Cette série est décrite par la donnée de deux suites : la suite $(t_i^\dagger)_{i\in\mathbb{N}}$ est la suite des instants issus de la réunion des deux suites $(t_i^+)_{i\in\mathbb{N}}$ et $(t_i^-)_{i\in\mathbb{N}}$; la suite $(A_i^\dagger)_{i\in\mathbb{N}}$ est la suite des amplitudes associées.
Une illustration de ces séries est présentée sur la figure (2.1).

Etant données ces séries décrivant les événements de la CdC, nous définissons les suites suivantes utiles pour l'extraction d' *a priori* sur les appareils électriques. Il s'agit des suites définies ci-après et dont une illustration est présentée sur la figure (2.2) :

➢ la suite $(a_i^+)_{i\in\mathbb{N}}$ et la suite $(\tau_i^+)_{i\in\mathbb{N}}$, dont les termes définis par l'équation 2.1, décrivent la variation d'amplitude et la durée entre deux événements successifs du type « *on* ». La série des durées $(\tau_i^+)_{i\in\mathbb{N}}$ décrit la durée du cycle de fonctionnement de l'appareil étudié.

$$\begin{cases} a_i^+ = A_{i+1}^+ - A_i^+ \\ \tau_i^+ = t_{i+1}^+ - t_i^+ \end{cases} \tag{2.1}$$

➢ la suite $(a_i^-)_{i\in\mathbb{N}}$ et la suite $(\tau_i^-)_{i\in\mathbb{N}}$, dont les termes définis par l'équation 2.2, décrivent la variation d'amplitude et la durée entre deux événements successifs du type « *off* ».

$$\begin{cases} a_i^- = A_{i+1}^- - A_i^- \\ \tau_i^- = t_{i+1}^- - t_i^- \end{cases} \tag{2.2}$$

➢ la suite $(m_i)_{i\in\mathbb{N}}$ et la suite $(\tau_i^m)_{i\in\mathbb{N}}$, dont les termes sont définis par l'équation 2.3, décrivent respectivement la variation d'amplitude et la durée entre un événement du type « *on* » et un événement du type « *off* » consécutifs. La série des durées décrit la durée effective de fonctionnement pendant un cycle de fonctionnement.

$$\begin{cases} m_i = A_i^- + A_i^+ \\ \tau_i^m = t_i^- - t_i^+ \end{cases} \tag{2.3}$$

➢ la suite $(r_i)_{i\in\mathbb{N}}$ et la suite $(\tau_i^r)_{i\in\mathbb{N}}$, dont les termes sont définis par l'équation 2.4, décrivent respectivement la variation d'amplitude et la durée entre un événement du type « *off* » et un événement du type « *on* » consécutifs. La série des durées $(\tau_i^r)_{i\in\mathbb{N}}$ décrit la durée d'arrêt effective (de repos effectif) de l'appareil étudié.

$$\begin{cases} r_i = A_{i+1}^+ + A_i^- \\ \tau_i^r = t_{i+1}^+ - t_i^- \end{cases} \tag{2.4}$$

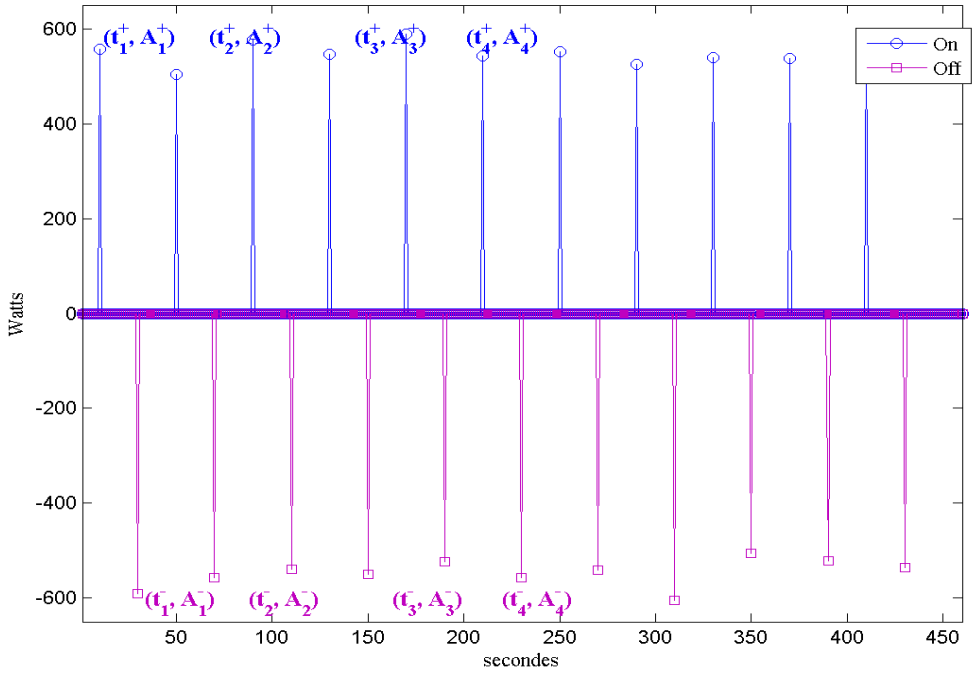


Figure 2.1: Illustration des séries des événements d'un appareil électrique

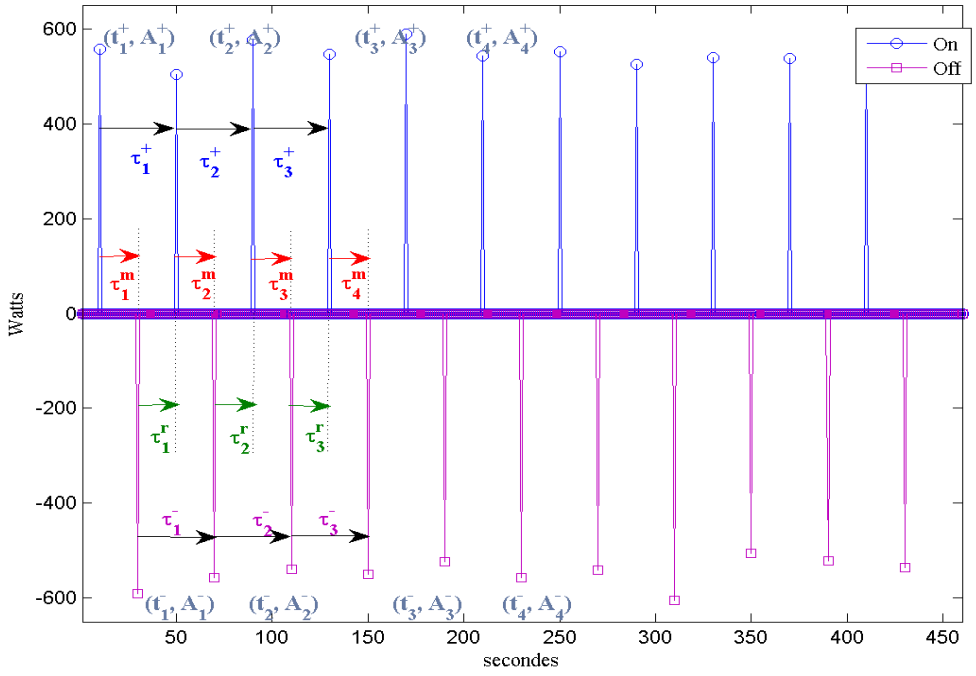


Figure 2.2: Illustration des suites des durées caractéristiques des événements d'un appareil électrique

Matériel de chauffage	Part dans le parc de chauffage résidentiel
Convecteurs	64 %
Panneaux rayonnants	23 %
Planchers rayonnants	12 %
Planfonds	1%

TABLE 2.2: Cartographie du chauffage électrique en France par type (matériel) de chauffage (2000)

Remarque

On distingue :

➢ l'intervalle de temps séparant un *on* et un *off*, dans le cas d'un appareil mono-charge (exemple : un convecteur, un four).
  – il est caractérisé par la durée Δt et la variation d'amplitude ΔA.

➢ l'intervalle de temps définissant le fonctionnement d'une charge (exemple : moteur, résistance), dans le cas d'un appareil multi-charges (exemple : un sèche-linge, un lave-linge, etc.).

2.2.2 Le chauffage par convecteurs

2.2.2.1 Généralités sur le chauffage par convecteurs

Le chauffage électrique est connu comme une spécificité française. En effet, depuis le choc pétrolier de 1973 et l'avènement du nucléaire, la part du chauffage électrique dans la consommation intérieure n'a cessé de croître. D'après des rapports de l'**AERE** [1], le parc français du chauffage électrique représentait 42% du parc européen. En 2000, la répartition de cet usage par type de chauffage pour les ménages français est détaillée dans le tableau (2.2), où l'on constate qu'environ deux tiers correspondent aux convecteurs électriques.

Dans la suite, on se limite au cas des convecteurs électriques. Le chauffage est réalisé par convection de l'air. Précisons que le convecteur a une régulation électronique par train d'ondes quasi-périodiques, et qu'il est asservi par rapport à une température de consigne.

2.2.2.2 Analyse des signaux

Les données exploitées durant la thèse pour caractériser les signaux de convecteurs sont celles de la campagne K_2. Les conclusions obtenues ont été confirmées et complétées grâce à la campagne de mesure K_3.

Un exemple de CdC d'un convecteur est illustré sur la figure (2.3). Il s'agit de la puissance moyenne échantillonnée à la période $T_e = 2$ s. On note que la puissance est quasiment constante sur une plage horaire d'environ huit heures. Les zooms (1) et (2) illustrent la présence de régulations. Plus précisément, sur le zoom (1), on observe une alternance entre des régulations et des zones de saturation (puissance quasiment constante).

1. **A**lternatives pour l'énergie, les **E**nergies **R**enouvelables et l'**E**nvironnement

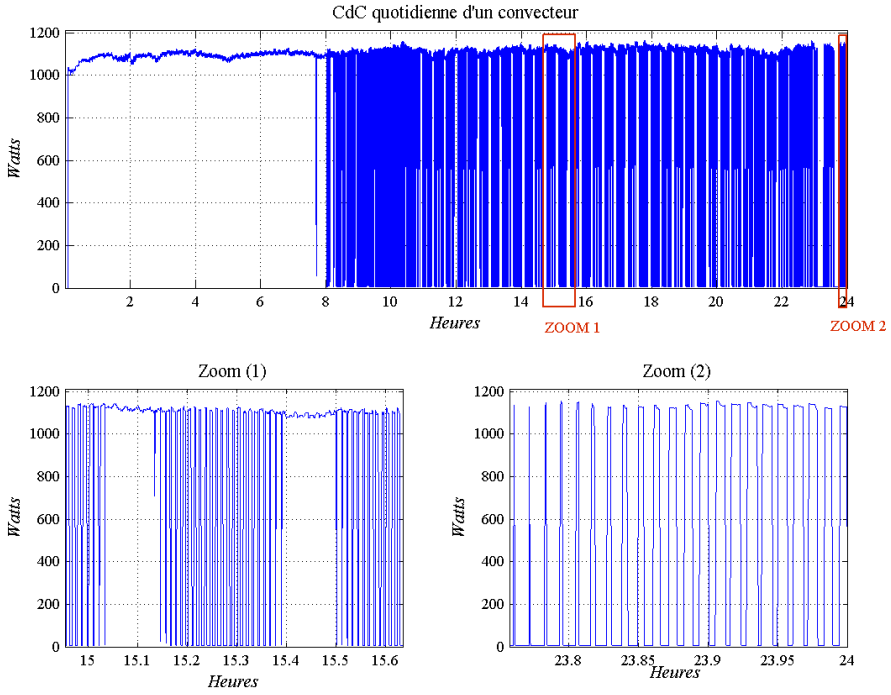


FIGURE 2.3: Puissance appelée par un convecteur électrique ($T_e = 2$s, K_2)

On en déduit la présence de deux régimes de fonctionnement pour cet appareil : un régime de régulation (à définir) et un régime de saturation. Le zoom (2) sur une zone où le convecteur ne sature pas, nous montre :

- ➢ que la durée Δt évolue au cours du temps ;
- ➢ et que les amplitudes des différents événements (enclenchement ou déclenchement) sont quasiment identiques.

On s'est alors intéressé à l'analyse de ces paramètres sur l'ensemble des convecteurs.

Séries des amplitudes

La figure (2.4) présente les histogrammes des variations d'amplitude obtenues respectivement pour (a_i^+), (a_i^-)et (m_i). Ces histogrammes ont été construits à partir de l'ensemble des signaux de la campagne K_2. La symétrie des histogrammes et la faible variance permettent d'établir la propriété suivante :

La série des on et la série des off issus de la puissance moyenne d'un convecteur, pour une période d'échantillonnage autour d'un hertz, présentent des amplitudes quasiment identiques. Cette propriété sera appelée « propriété de symétrie des on-off d'un convecteur ».

Remarque : modélisation des séries d'amplitudes

Soient $E_v(i)$ et $E_v(i+1)$ deux événements consécutifs correspondant respectivement à un *on* et à un *off*. On obtient alors le modèle décrit par (2.5) où ε_a décrit un bruit centré d'écart type σ_a évalué à partir des échantillons observés.

$$\begin{cases} A_{i+1}^{+} = A_i^{+} + \varepsilon_a(i) \\ A_{i+1}^{-} = A_i^{-} + \varepsilon_a(i) \\ A_i^{+} = -A_i^{-} + \varepsilon_a(i) \end{cases} \tag{2.5}$$

Régulation quasi-périodique des convecteurs électriques

On note sur le zoom (2) de la figure (2.3) une certaine régulation des événements de la CdC d'un convecteur. Une analyse spectrale de plusieurs exemples de convecteurs montre la présence d'une quasi-périodicité de la puissance appelée par un convecteur. Cette propriété est prévisible étant donné que la régulation de l'appareil est gérée par une régulation électronique périodique.

Afin de vérifier cette observation mono-appareil, nous avons calculé la durée séparant deux *on* consécutifs.

Les résultats (en valeur absolue) obtenus à partir de l'ensemble des données de la campagne K_2 sont présentés sur la figure (2.5).

On en conclut que la période séparant deux *on* consécutifs est quasiment constante. La faible variance de l'histogramme et le pic prononcé autour de la valeur « 0 » nous permettent d'établir la propriété suivante :

La régulation d'un convecteur électrique est quasi-périodique. Si T_0 est la période de régulation théorique, *deux événements consécutifs du type « on » vérifient la relation suivante* (2.6).

$$t_{i+1}^{+} = t_i^{+} + T_0 + \varepsilon_{T_0}(i) \tag{2.6}$$

où ε_{T_0} est un bruit gaussien centré de variance $\sigma_{T_0}^2$.

Périodes de régulation

Suite à l'analyse des signaux de l'ensemble des convecteurs de la Maison Domotique et des deux sites réels, nous avons constaté la présence de deux périodes de régulation uniquement. Les valeurs obtenues sont 40 s et 80 s.

Evolution de la durée on-off («rapport cyclique») d'un convecteur

Nous avons vu que la puissance appelée par un convecteur est un mélange de régime de saturation et de régulation. De plus, nous avons noté que la durée *on-off* évolue au cours du temps.

Par analogie avec les signaux périodiques, nous introduisons un paramètre appelé *rapport cyclique* d'un convecteur défini par l'équation (2.7) :

$$\rho_{T_0}(i) = \frac{\tau_i^m}{t_{i+1}^{+} - t_i^{+}} \sim \frac{\tau_i^m}{T_0} \tag{2.7}$$

L'évolution au cours du temps de ce paramètre pour deux convecteurs de la Maison Domotique (même journée) présentant des zones de saturation est décrite par la figure (2.6). Dans les deux cas, on note la présence de valeurs supérieures ou égales à 100, indicateur de la saturation du convecteur. Remarquons que ces phases de saturation peuvent durer quelques heures.

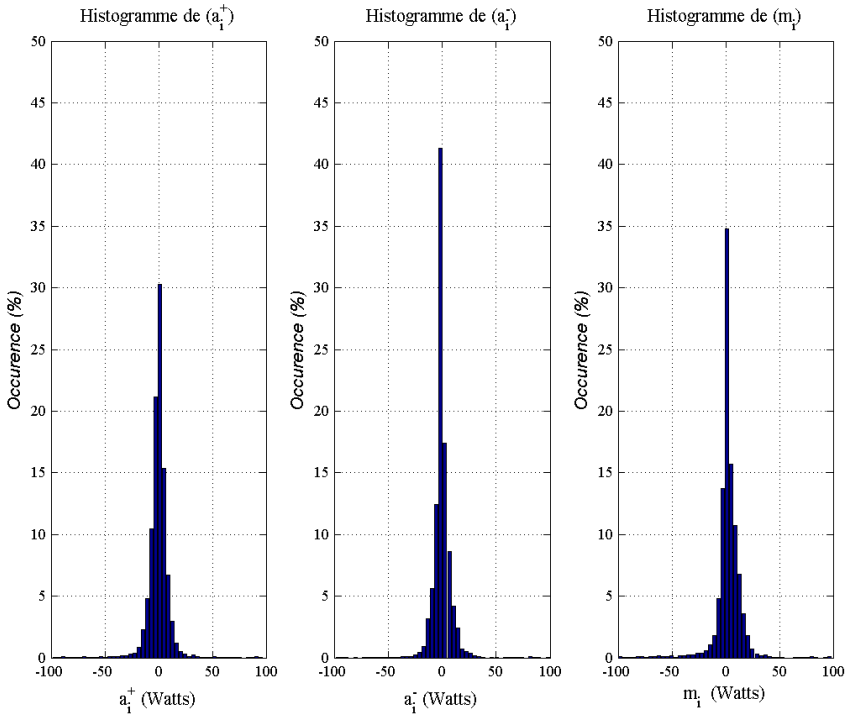


FIGURE 2.4: Symétrie de l'amplitude d'enclenchement et de l'amplitude de déclenchement d'un convecteur ($T_e = 2$sec, K_2) : Convecteurs test - tous les convecteurs de la Maison Domotique

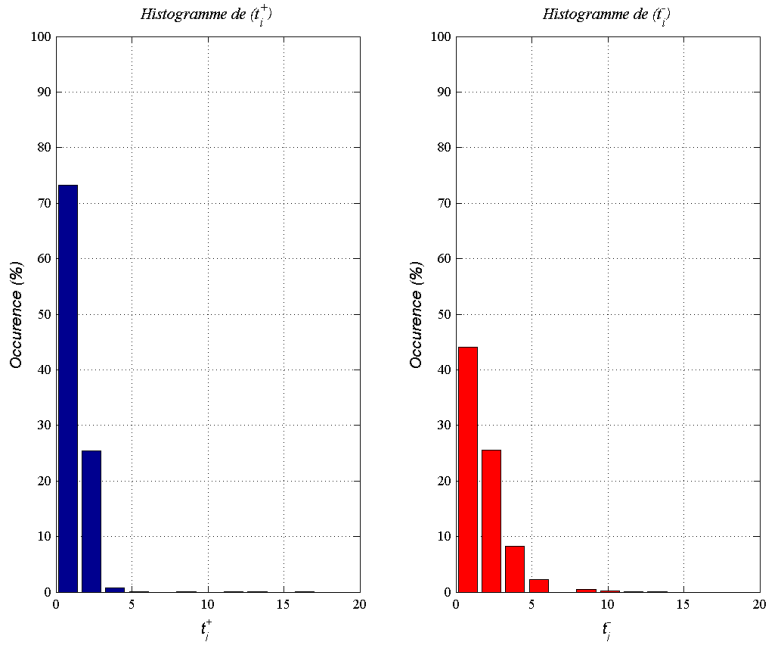


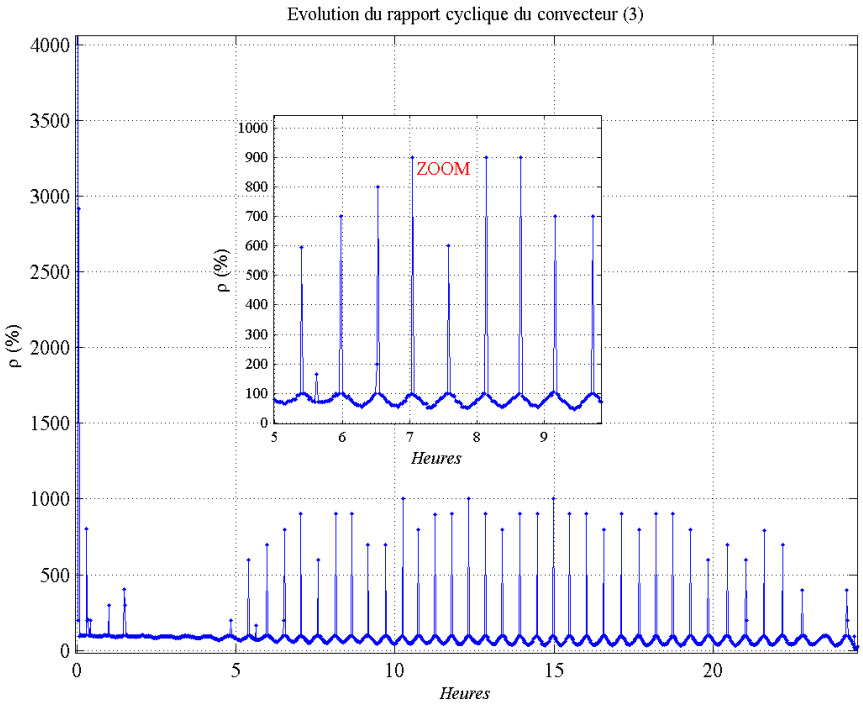
FIGURE 2.5: Quasi-périodicité de la puissance appelée par un convecteur

Les deux cas présentés correspondent à deux exemples d'évolution différents. Dans le cas du convecteur (3), un zoom montre que le rapport cyclique oscille de manière quasi-sinusoïdale avec l'apparition périodique de phases de saturation. Dans le cas du convecteur (5), le rapport cyclique décroît globalement et ne présente pas de structure particulière d'évolution temporelle.

La fonction de répartition empirique de ρ_{T_0} ainsi que l'histogramme normalisé de la variation du rapport cyclique obtenus pour les deux convecteurs étudiés sont représentés sur la figure (2.7). On constate que les variations du rapport cyclique sont faibles (multiples d'un pas constant commun à tous les appareils).

Au regard de ces observations, on peut considérer que le rapport cyclique (ramené à 1) évolue lentement dans l'intervalle [0, 1]. La valeur nulle correspond à l'état «d'arrêt» de l'appareil. La valeur 1 correspond à un régime de saturation.

Précisons que l'évolution du rapport cyclique dépend de plusieurs paramètres dont la température extérieure, la température de consigne et les caractéristiques thermiques du bâtiment.



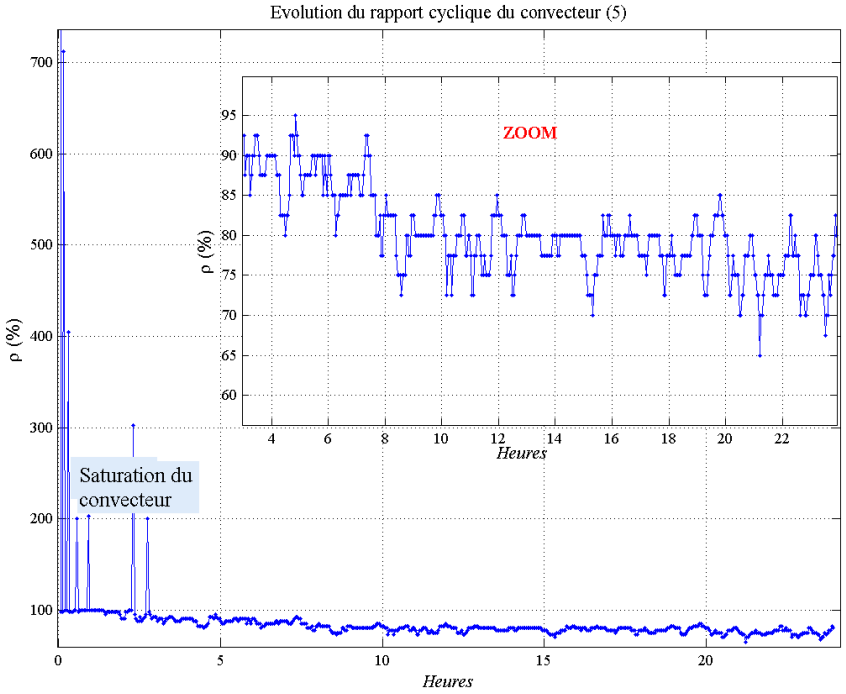
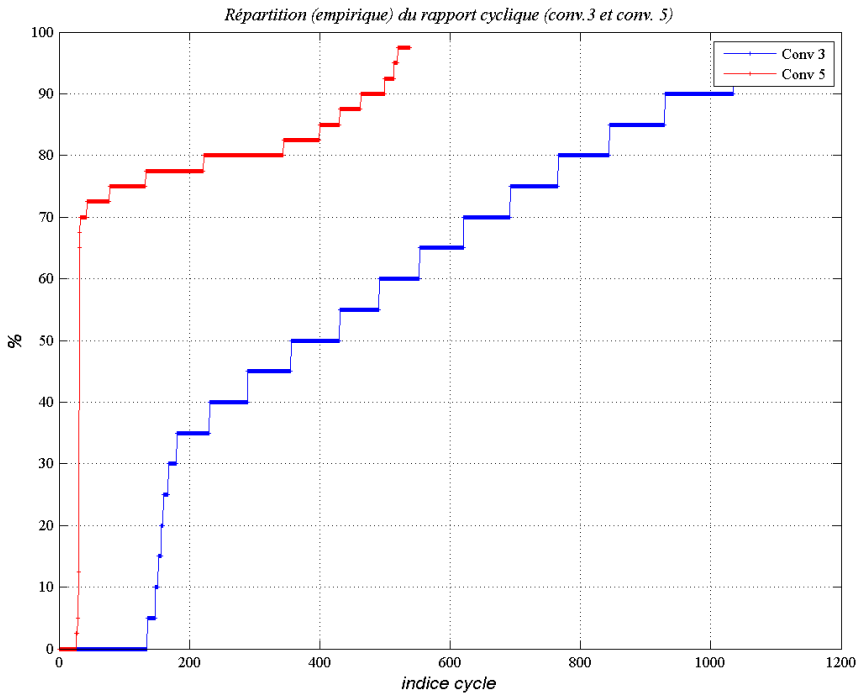


FIGURE 2.6: Evolution « quotidenne » du rapport cyclique : Exemple de deux convecteurs



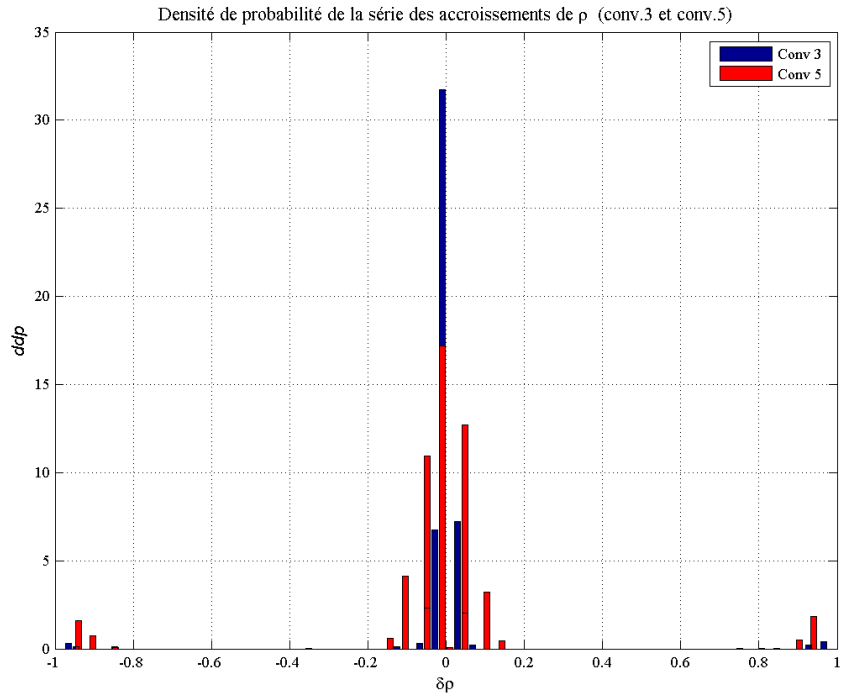


FIGURE 2.7: Analyse statistique du rapport cyclique

Foisonnement des convecteurs

La CdC chauffage est un mélange linéaire des courbes de charge élémentaires des différents convecteurs dont on ignore le nombre, les périodes de régulation et les amplitudes.

Si l'on se restreignait à la CdC du chauffage, nous aurions un problème de décomposition de la CdC à part entière. Ce problème est plus ou moins complexe en fonction des caractéristiques des appareils et de leur foisonnement temporel.

La figure (2.8) illustre deux exemples de foisonnement de signaux de convecteurs sur une faible durée (environ un quart d'heure).

Le premier cas est un mélange de deux convecteurs de puissances nominales identiques et de périodes de régulation différentes (40 s et 80 s). La décomposition de mélange simple à première vue n'est pas immédiate par une approche aveugle.

Le second mélange est plus complexe. Les amplitudes des signaux sont différentes. On a en effet les 4 cas de figure possibles qui se produisent simultanément :

- ➢ un convecteur en régime de saturation ;
- ➢ un convecteur qui alterne des régulations et une phase de saturation ;
- ➢ un convecteur dont le rapport cyclique décroît jusqu'à atteindre la valeur nulle (arrêt) ;
- ➢ un convecteur en régime saturé, dont le rapport cyclique commence à descendre en dessous de 1.

Le foisonnement de ces signaux, sans prendre en compte l'interférence avec les autres appareils, montre la complexité des mélanges réels que l'on peut avoir et la grande hétérogénéité de ces mélanges.

Exemples de CdC de chauffage

Dans la suite, trois exemples de CdC quotidiennes de l'usage chauffage chez un client particulier sont présentés.

La figure (2.9) correspond à une journée du mois de janvier, et met en évidence :

- ➢ d'une part le niveau élevé de consommation de cet usage ;
- ➢ d'autre part le foisonnement important des signaux élémentaires qui peut donner lieu à l'apparition d'une composante basse fréquence illustrée sur le zoom de la figure (2.9).

La figure (2.10) correspond à une journée du mois d'avril. Le zoom réalisé montre que le foisonnement des appareils persiste, mais apparaît moins important qu'au mois de janvier. La courbe de consommation d'une journée du mois de mai est plus simple, peu de convecteurs sont en marche, le foisonnement est nettement moins important que dans les deux cas précédents. Ce cas devrait *a priori* être plus simple à étudier, mais n'oublions pas que cet usage sera aussi superposé aux courbes de charge des autres appareils.

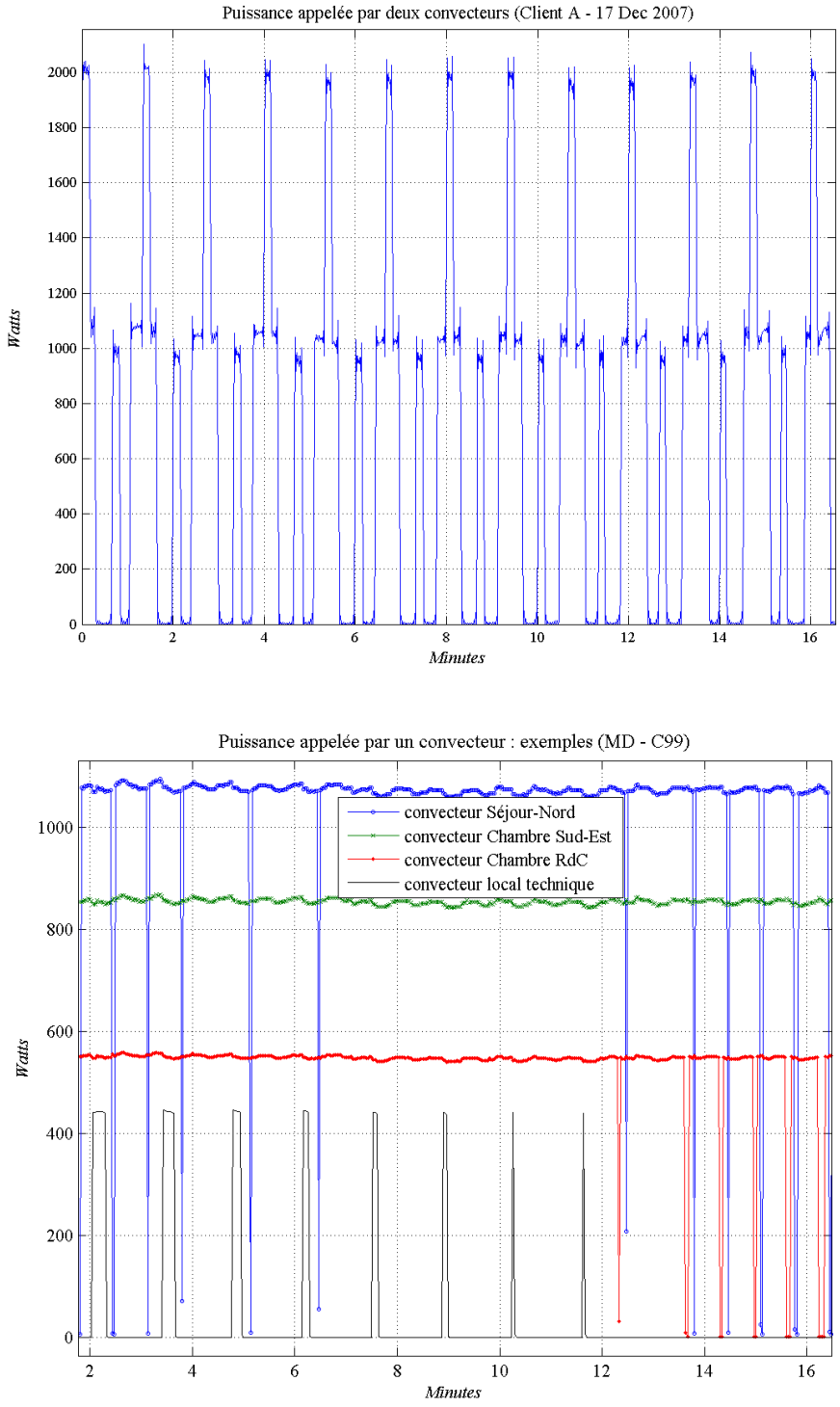


FIGURE 2.8: Foisonnement de convecteurs

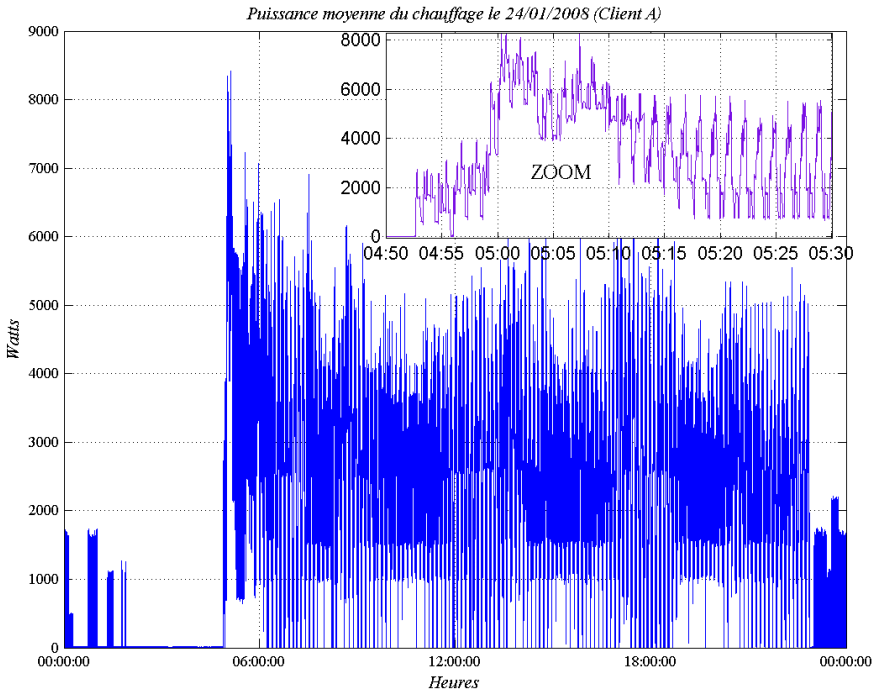


FIGURE 2.9: Exemple de CdC de chauffage - janvier 2008

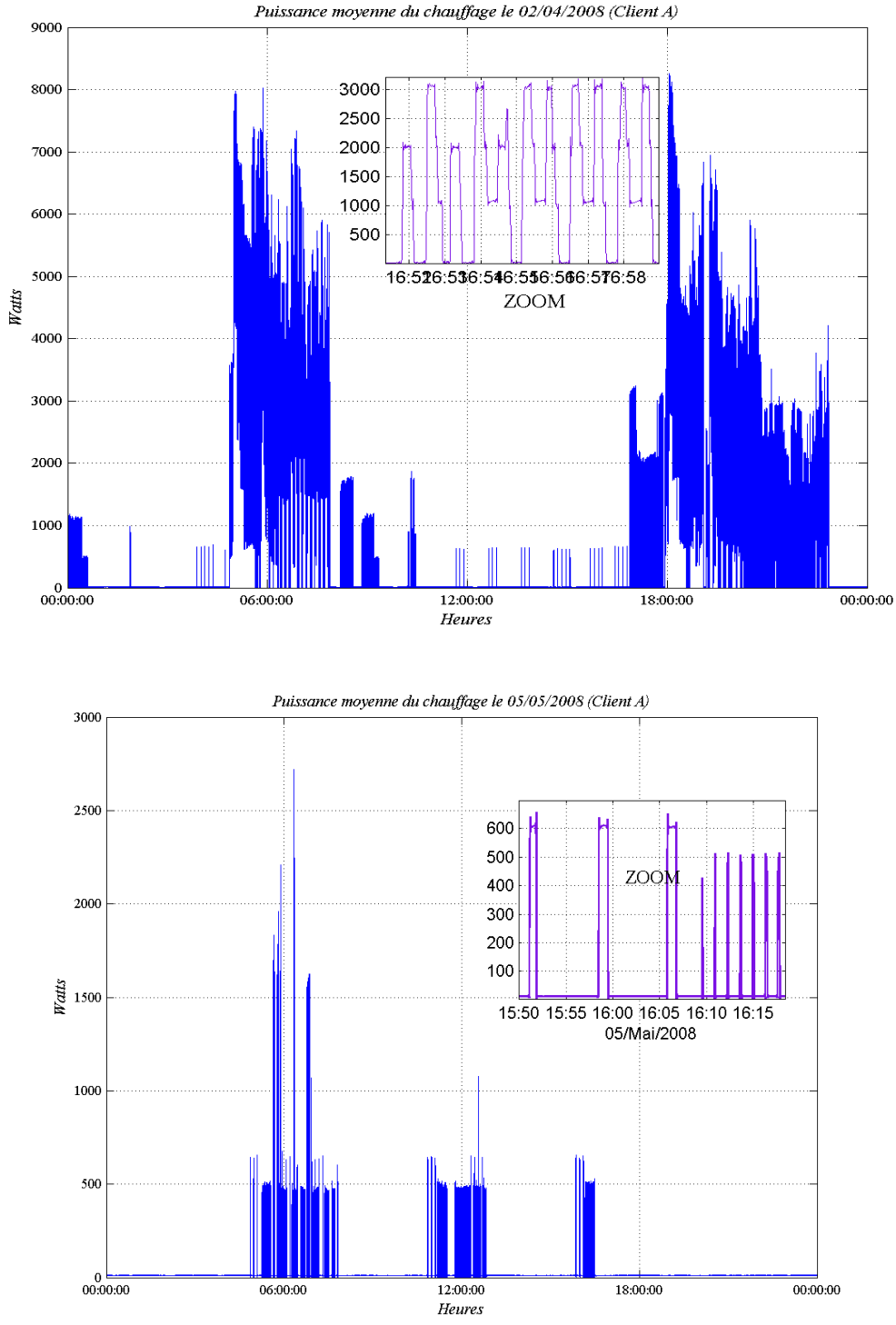


FIGURE 2.10: Exemples de CdC du chauffage - avril et mai 2008 ($Te = 1$s, K_3)

Lien période d'échantillonnage - nombre d'événements

La figure (2.11) met en évidence l'influence du choix de la période d'échantillonnage sur le nombre d'événements générés par chauffage électrique. Ce nombre est d'autant plus élevé que T_e est faible, et devient quasi-nul dès que T_e s'approche de la période de régulation minimale d'un convecteur. En effet, à partir de ce seuil la puissance moyenne d'un convecteur est quasiment constante en permanence. La CdC du chauffage présente alors très peu de variations qui correspondent simplement à la mise en marche et à l'arrêt des différents convecteurs. Une période d'échantillonnage trop élevée masque les événements des différents cycles de fonctionnement des convecteurs. Cependant, elle met en évidence d'autres propriétés telle que l'évolution du rapport cyclique.

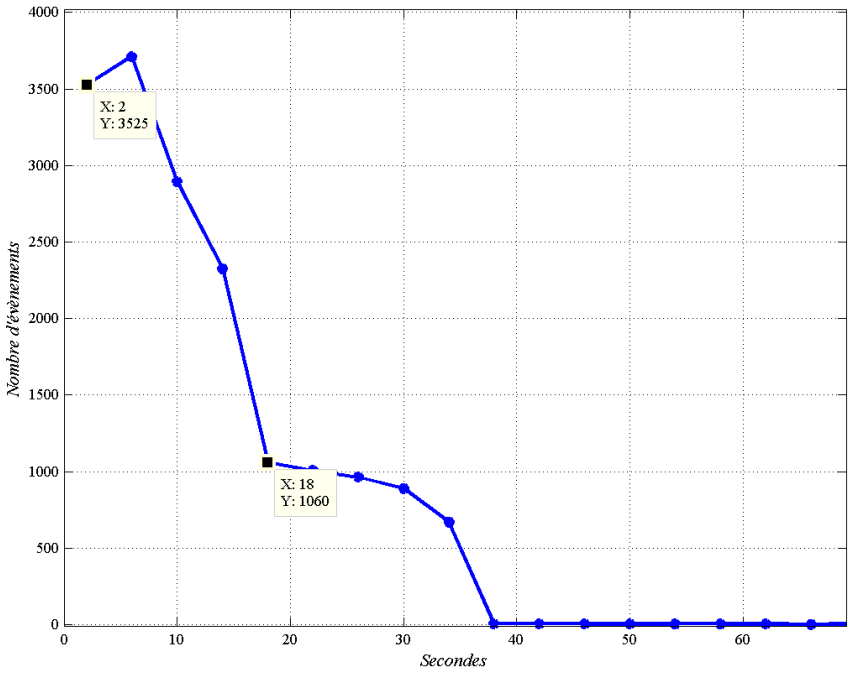


FIGURE 2.11: Nombre d'événements de la CdC chauffage en fonction de la période d'échantillonnage

La figure 2.12 (ellipse rouge) met en évidence l'apport de l'échantillonnage pour étudier cette dernière propriété, dont l'observation n'est pas immédiate à l'échelle d'une journée en présence de plusieurs convecteurs pour une période d'échantillonnage de l'ordre de la seconde.

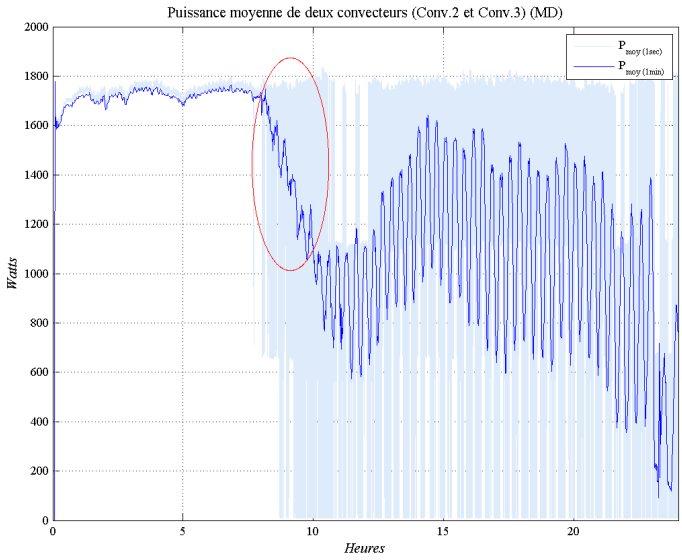


FIGURE 2.12: Exemple de CdC sous-échantillonnée

Type de logement	F1	F2	F3	F4	F5
Volume du chauffe-eau (litre)	100	150	200	250	300
Puissance nominale (kW)	1.2	1.8	2.2	3	3

Table 2.3: Caractéristiques de consommation d'un ballon d'ECS

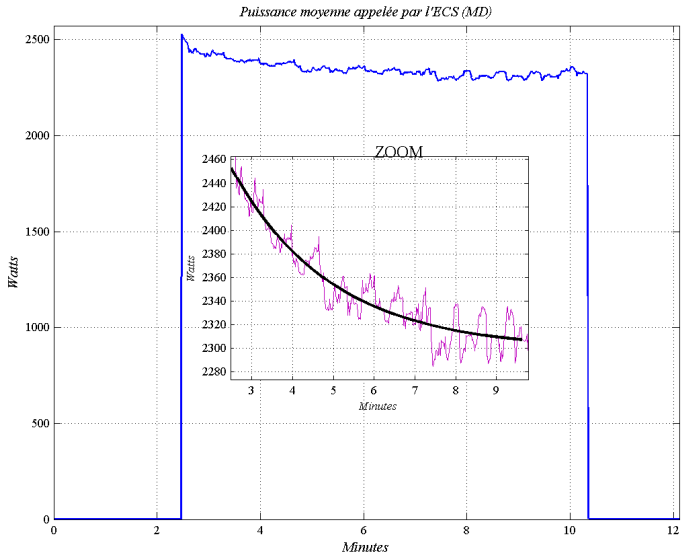


FIGURE 2.13: Puissance appelée par l'ECS - un cycle de fonctionnement

2.2.3 L'eau chaude sanitaire (ECS)

L'eau chaude sanitaire (ECS) est un usage à dominante résistive. Une description du chauffe-eau à accumulation est fournie en annexe (D). Avant de présenter les signaux de cet usage, on s'est intéressé à l'existence d'un lien entre la taille d'un logement, le volume du chauffe-eau et sa puissance nominale. Des liens précis ne sont pas évidents à établir, mais des ordres de grandeur peuvent être établis tel que le résume le tableau (2.3).

Pour caractériser cet usage, on s'appuiera sur les données de la Maison Domotique (issues de 4 scénarios de consommation simulés) et les données d'un client particulier (K_4). Ici, CdC désignera la puissance moyenne appelée par l'appareil, échantillonnée à $T_e = 1$ s. Un exemple de cycle de fonctionnement est illustré par la figure (2.13). Ce signal est extrait des signaux de la MD, la puissance nominale de l'appareil est de 2500 W. On constate :

- ➢ que l'appareil fonctionne en «tout ou rien» ;
- ➢ que la puissance décroît légèrement de manière quasiment exponentielle durant le cycle de fonctionnement (voir zoom sur la figure 2.13).

Ces observations ont été confirmées en analysant plusieurs cycles de fonctionnement. Tout comme pour le chauffage électrique, nous nous sommes intéressés aux propriétés de la série des événements de la CdC de l'ECS.

2.2.3.1 Loi de l'amplitude moyenne

Série des amplitudes « on à off » (m_i)

Précisons d'ores et déjà que la puissance d'un chauffe-eau n'est pas une grandeur spécifique à cet appareil. L'objectif est de caractériser l'asymétrie observée entre l'enclenchement et le déclenchement du chauffe-eau. Nous avons alors évalué les accroissements de la série des amplitudes (m_i) sur l'ensemble des données valides disponibles. Les histogrammes obtenus sont présentés sur la figure (2.14) sachant que la puissance nominale du chauffe-eau du client A est de 3500 W. Ces graphiques confirment l'observation précédente, et permettent de quantifier en moyenne l'écart absolu entre l'amplitude d'un *on* et celle du off correspondant. Cet ordre de grandeur devra être revu si des données supplémentaires sont disponibles.

2.2.3.2 Lois des durées

Analyse des durées

L'analyse des durées (on-on, off-off et on-off) des deux ballons d'eau chaude disponibles est résumée par les histogrammes de la figure (2.15). Le nombre d'appareils étant limité, on se limite à en déduire les bornes inférieure et supérieure de chacune des durées. La comparaison avec l'analyse effectuée pour les convecteurs, l'ECS est un usage lent au sens où la durée séparant deux cycles consécutifs (de l'ordre de l'heure) est largement supérieure à celle des convecteurs (de l'ordre d'une minute).

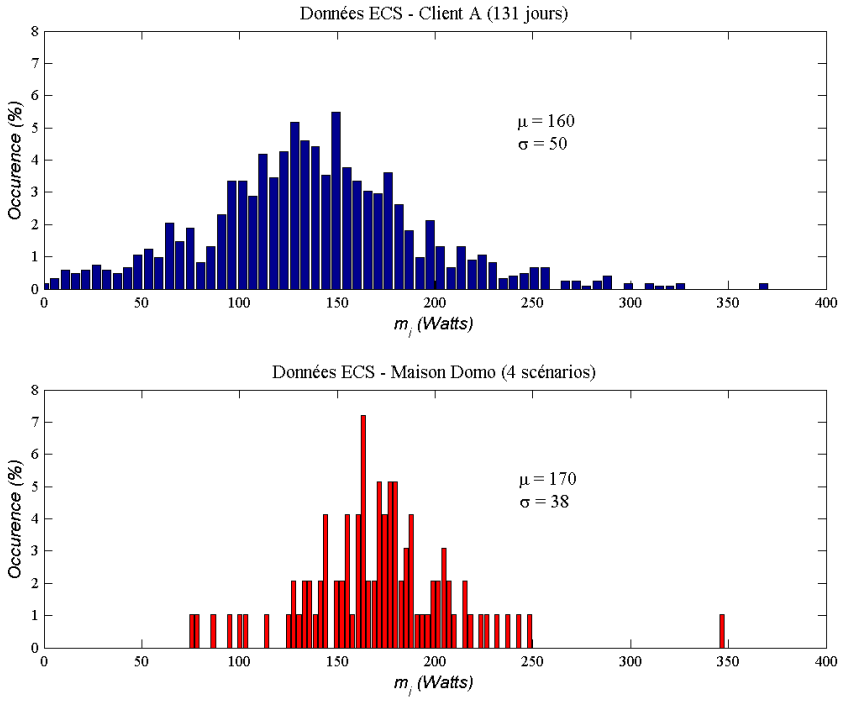


FIGURE 2.14: Asymétrie On-Off

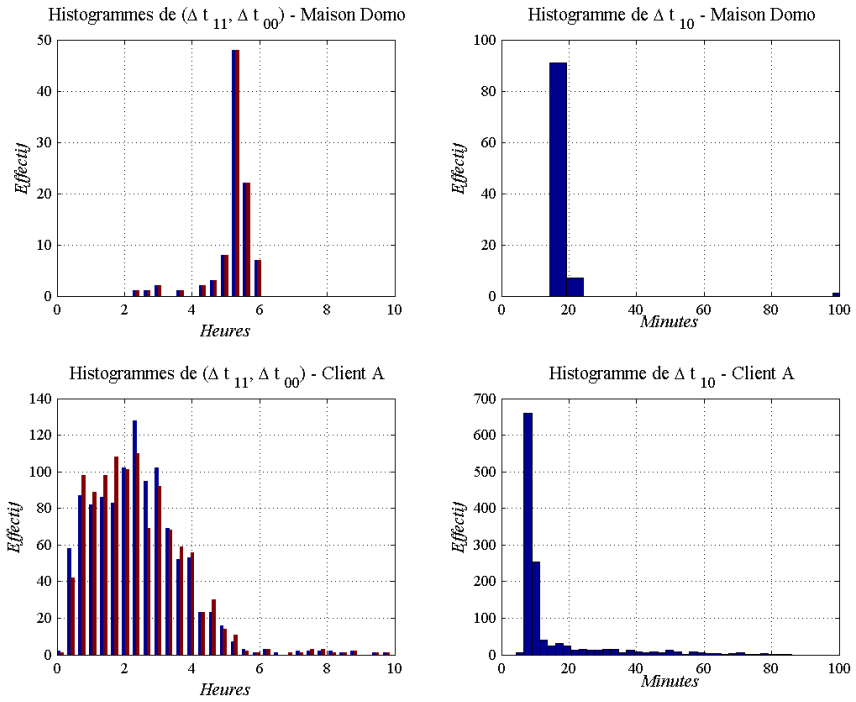


FIGURE 2.15: Histogrammes des durées *caractéristiques* de l'ECS : (Δt_{11}), (Δt_{00}) et (Δt_{10})

Exemple de CdC quotidienne d'ECS

Un exemple de CdC quotidienne est présenté sur la figure (2.16), et met en évidence les deux propriétés suivantes de l'ECS :

- ➢ *le caractère «tout ou rien» de cet usage* ;
- ➢ *le nombre limité de cycles de fonctionnement (et donc d'événements) par jour*
  - *en comparaison avec l'usage chauffage, la CdC quotidienne de l'ECS correspondrait en nombre d'événements générés à un convecteur de période de régulation de* 80 s *fonctionnant pendant environ un quart d'heure seulement.*

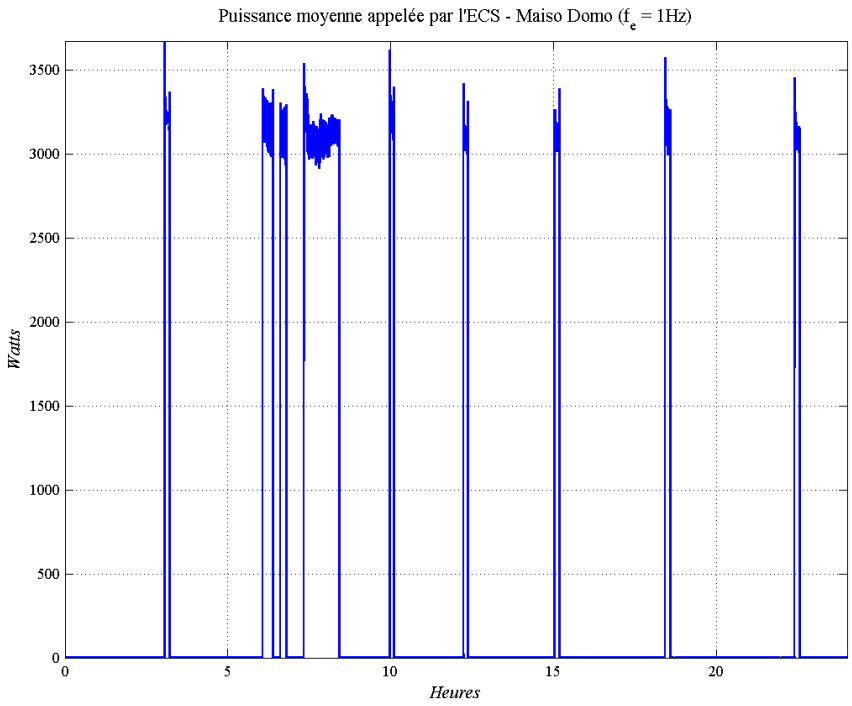


FIGURE 2.16: Exemple de CdC d'ECS

Un mode d'analyse simple de consommation sur un horizon de plusieurs jours est le graphe d'intensité. Cette représentation dans le plan (*Heures, jours*) permet d'identifier des spécificités de la CdC tels que la régularité de consommation, le profil de consommation jour de la semaine/week-end. La figure 2.17 présente deux graphes d'intensité de l'ECS de la Maison Domotique et d'un client particulier. La régularité observée dans le premier cas s'explique par le scénario de tirage d'eau simulé : un volume identique d'eau est tiré à des moments précis de la journée. Cette régularité disparaît dans le second cas, mais on observe certains motifs qui ne sont pas nécessairement réguliers sur tout le domaine temporel de l'analyse de la CdC. Le graphe d'intensité permet de détecter les jours à faible consommation d'eau chaude.

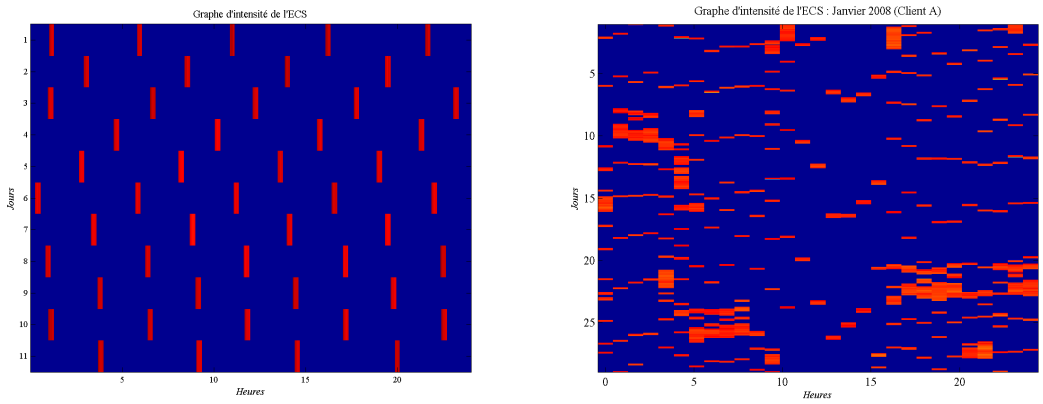


FIGURE 2.17: Graphe d'intensité de l'ECS

2.2.4 Le froid alimentaire

L'usage froid alimentaire est le foisonnement d'un nombre inconnu de courbes de charge de compresseurs [2]. Il s'agit d'un système multi-charges, asservi par rapport à une température de consigne. La CdC globale a une dominante inductive. Dans la suite, on présentera séparément les signaux du secteur résidentiel et ceux du secteur tertiaire.

Précisons qu'il est difficile d'acquérir la puissance appelée par une composante (un compresseur seul) afin de caractériser le système global.

2.2.4.1 Premier cas : le secteur résidentiel

Les signaux de compresseurs issus des deux premières campagnes de mesure sont eux-mêmes des signaux composites. Chez le client particulier, nous avons isolé la puissance appelée par le froid alimentaire par extraction d'une journée d'absence des résidents. La puissance globale mesurée en sortie du CBE est illustrée sur la figure (2.18). On y observe un foisonnement de créneaux de faible amplitude et présentant souvent des pics de forte amplitude au démarrage.

Il s'agit en réalité d'un mélange d'un nombre inconnu de compresseurs associé à la consommation de l'usage froid alimentaire et d'une pompe à chaleur (**PAC**). Le démarrage et l'arrêt de la PAC sont indiqués par une double flèche sur la figure (2.18).

Le zoom de la figure (2.18) met en évidence les deux propriétés suivantes de la CdC observée :

- ➢ *tous les compresseurs ont a priori la même puissance nominale* ;
- ➢ *il existe* a priori ***au minimum trois compresseurs (en dehors de la PAC) dans le mélange, au regard de la durée des cycles d'événements observés***. *(cet a priori a été validé par le client)* ;
- ➢ *le nombre de cycles de fonctionnement (et donc d'événements) par jour est supérieur à celui de l'ECS, mais très inférieur à celui généré par un convecteur en régime de régulation* :
  - *La CdC de compresseurs présentée est équivalente en nombre d'événements par jour à un convecteur de période de régulation* 80 s *fonctionnant pendant environ une heure seulement.*

2. Un compresseur est une pompe qui comprime un gaz particulier dont la détente ultérieure génère le froid.

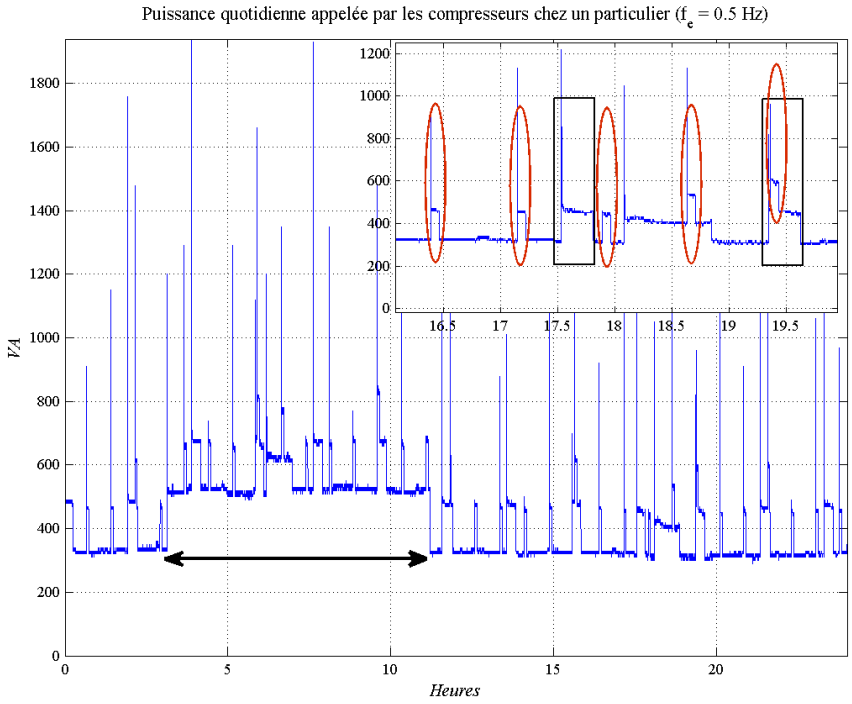


FIGURE 2.18: Puissance moyenne quotidienne appelée par les compresseurs

Remarque

Contrairement aux deux usages précédents (chauffage et ECS), on ne peut pas définir des *a priori* quantitatifs sur les amplitudes des événements et sur les lois temporelles à partir des signaux élémentaires. Mais, on peut formuler des *a priori* à partir de connaissances d'experts ou d'analyses qualitatives des signaux composites.

Des signaux collectés ultérieurement dans le cadre de la campagne K_4 permettront une caractérisation plus fine de cet usage. Nous présentons ici à titre d'illustration deux exemples de la puissance instantanée d'un réfrigérateur ayant un seul compresseur (*a priori*). Le premier cas illustre la puissance instantanée sous-échantillonnée à $T_e = 10$ms (2.19). Le zoom autour de l'enclenchement du compresseur met en évidence la présence d'une forme particulière du signal au démarrage du compresseur : il s'agit soit d'un pic prononcé soit d'une forme quasi-parabolique pendant le régime transitoire de l'appareil. Cette forme a été observée sur un échantillon d'une dizaine d'essais sur des réfrigérateurs différents.

Le second zoom met en évidence la présence d'une seconde composante d'amplitude moyenne proche de 10 W. Il est quasiment impossible d'avoir un signal isolé pour cet usage.

Le second cas (2.20) illustre la puissance instantanée ($T_e = 1$s) d'un réfrigérateur d'un client particulier (client A - K_4). Il s'agit d'un réfrigérateur ayant *a priori* un seul compresseur. La figure montre une signature particulière du signal durant un cycle de fonctionnement, qui traduit le caractère inductif de cet appareil. Ce modèle de décroissance du signal est similaire à celui de l'ECS.

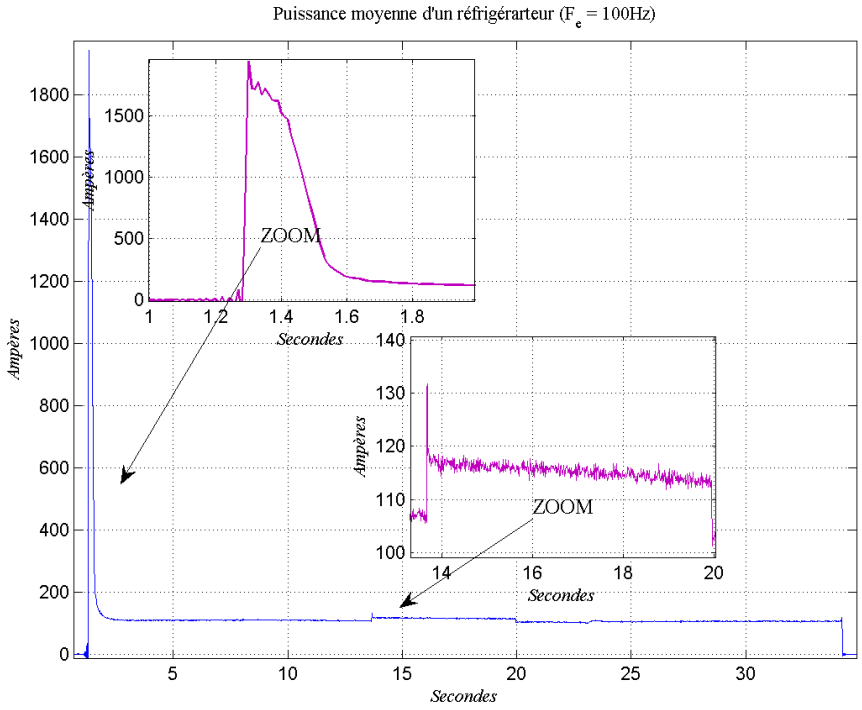


Figure 2.19: Puissance active d'un réfrigérateur (essai de laboratoire - EDF R&D avril 2007)

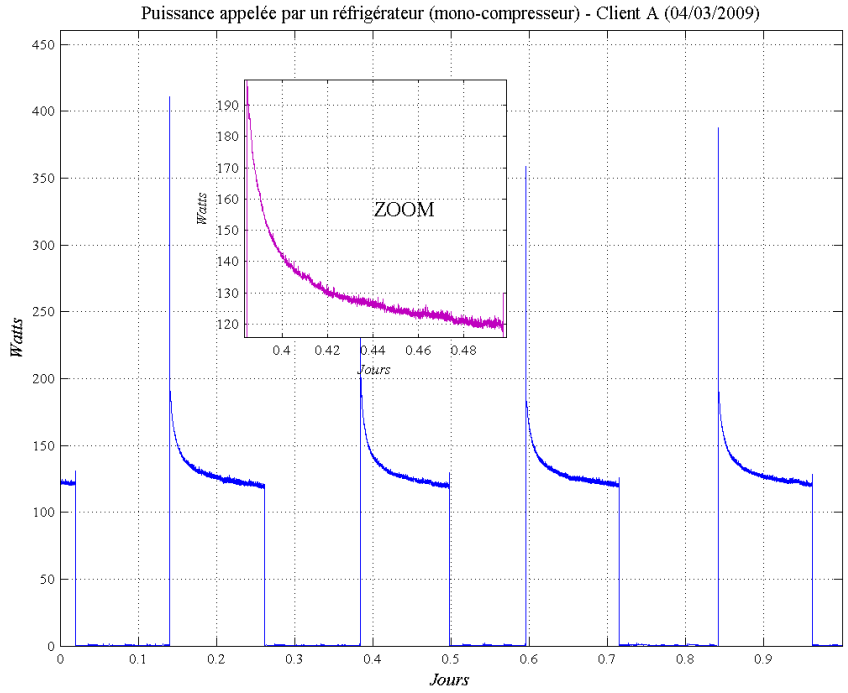


Figure 2.20: Puissance instantanée d'un réfrigérateur mono-compresseur (Client A - mars 2009)

2.2.4.2 Deuxième cas : le secteur tertiaire - froid alimentaire dans un hypermarché

Ici, « courbe de charge » désigne le couple (puissance active moyenne P , puissance réactive moyenne Q) échantillonnée à une fréquence $F_e = 10$ Hz. Un exemple de signaux est illustré sur la figure (2.21). On constate que chaque courbe est un mélange de créneaux présentant des pics de très forte amplitude (en comparaison avec l'amplitude du signal) à chaque enclenchement. On constate également qu'il existe deux familles de compresseurs :

- ceux ayant une forte amplitude de l'ordre de 5000 W/VAr et dont la régulation est telle que le nombre de cycles de fonctionnement et de l'ordre d'une dizaine par heure ;
- ceux ayant une faible amplitude (cf. zoom) de l'ordre de 500 W/VAr et dont la régulation est plus rapide que dans le cas précédent.

La CdC générale de l'hypermarché est constituée principalement des trois usages suivants : le froid alimentaire, les fours de la boulangerie et l'éclairage. Un exemple de la CdC générale sur l'intervalle [00 h,01 h] est présenté sur la figure (2.22). On constate que les principaux événements de la CdC générale coïncident avec ceux des compresseurs. Mais certains événements générés par d'autres usages sont également présents et devront être pris en compte lors de la reconstruction du scénario de fonctionnement du froid alimentaire. Souvent, ces événements n'ont pas de composante réactive, ce qui facilite l'identification de la reconnaissance des événements générés par les compresseurs du froid alimentaire. Etant donné que l'objectif est de caractériser les compresseurs (amplitude et scénario de gestion), *a priori* il semble judicieux d'exploiter les plages horaires où cet usage est prépondrant à savoir $[21h : 00\,h[$ et $[00h : 5h]$. Ce choix sera consolidé par l'analyse de la CdC des fours pendant ces plages horaires.

« *A priori* à dire d'expert » sur le parc de compresseurs de l'installation ciblée

- Le fonctionnement du parc de compresseurs est théoriquement géré par un automate tel qu'on minimise la sollicitation du même compresseur (l'objectif est de répartir la charge, en énergie et en durée de fonctionnement, le plus uniformément possible entre les compresseurs du parc). Par ailleurs, la mise en marche et l'arrêt des compresseurs sont effectués selon la règle suivante : *le premier enclenché est le dernier à déclencher*. Cette règle est un *a priori* fort pour décoder la série des événements de compresseurs afin de reconstruire les signaux élémentaires.
- La durée minimale entre deux événements successifs $E_v(i)$ et $E_v(i+1)$ est supérieure à deux secondes. Cet *a priori* pourra être utile pour l'estimation de la série des événements utiles pour l'usage froid alimentaire à partir de la CdC générale.
- Le signal de puissance active (réactive) se stabilise (régime permanent atteint) au bout d'une durée maximale de s. Cet *a priori* peut servir à l'estimation de l'amplitude des signaux de compresseurs.

En résumé, qu'il s'agisse du secteur résidentiel ou du secteur tertiaire, on constate que :

- les signaux présentent un pic au démarrage ;
- ils ont un fonctionnement en tout ou rien et la gestion des appareils est régie par une règle précise de séquencement des enclenchements et des déclenchements ;
- les amplitudes moyennes sont quasiment identiques pour tous les cycles de fonctionnement ;
- les compresseurs sont faciles à détecter en l'absence d'autres usages.

Quant aux spécificités de chaque classe de signaux :

- dans le secteur résidentiel, les compresseurs ont une amplitude faible et une régulation en tout ou rien ;
- dans l'hypermarché, les compresseurs ont une forte amplitude et ont une régulation FIFO.

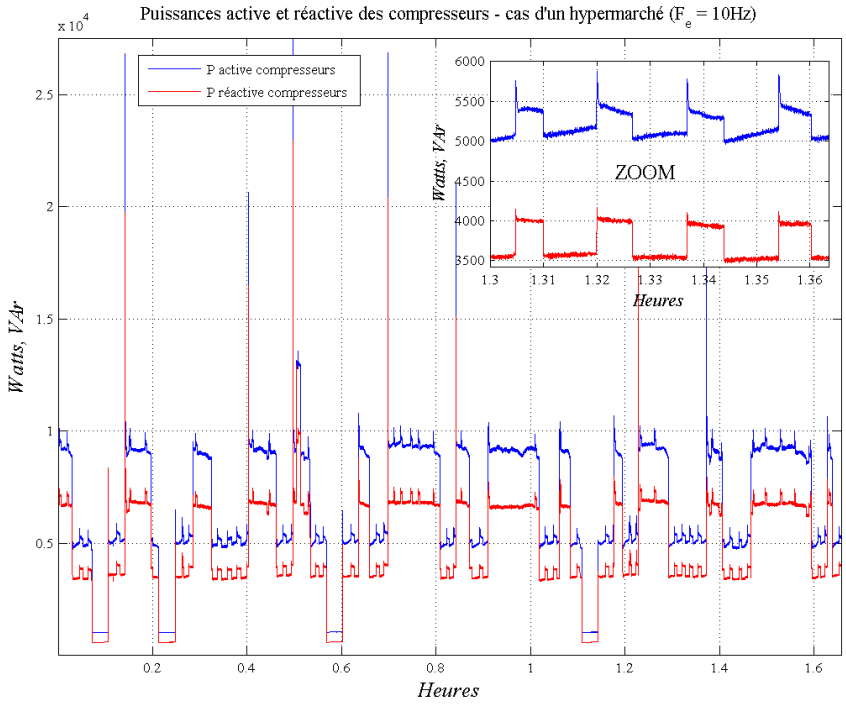


FIGURE 2.21: CdC du froid alimentaire d'un hypermarché

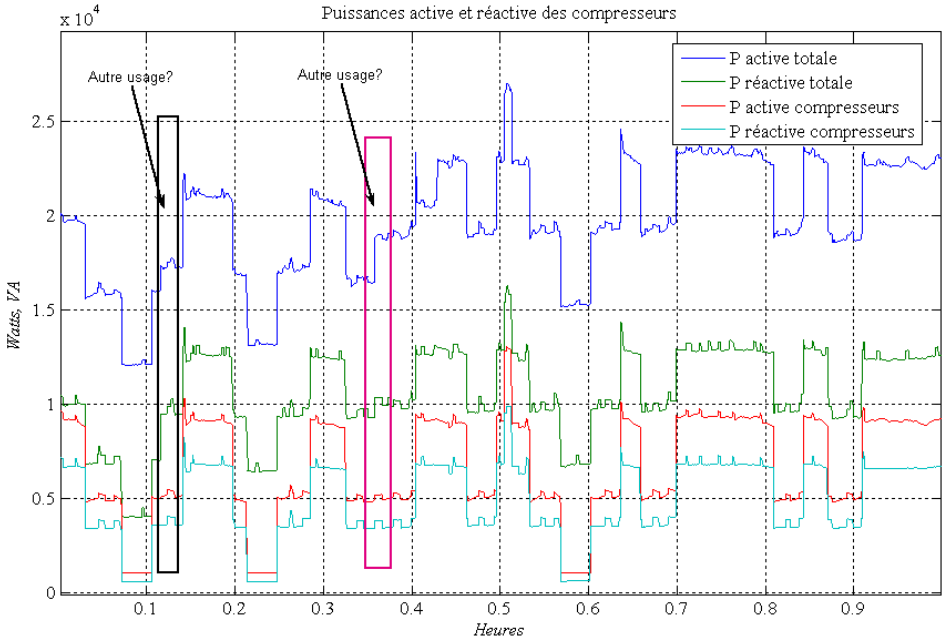


FIGURE 2.22: Zoom sur la CdC générale d'un hypermarché

2.2.5 Les « gros » appareils électroménagers

Cette classe d'usages regroupera les appareils suivants : le lave-linge (**LVL**), le lave-vaisselle (**LVV**) et le sèche-linge (**SL**). Chacun de ces appareils est constitué de plusieurs charges.

2.2.5.1 Présentation des appareils électriques ciblés

a) Lave-linge Cet appareil est constitué de trois charges : un tambour, la résistance de chauffage et une pompe.

b) Lave-vaisselle Les deux principales charges sont une résistance pour chauffer l'eau et une pompe pour l'aspersion et la vidange de l'eau.

c) Sèche-linge Cet appareil est constitué de plusieurs charges, dont un tambour, une pompe et une résistance pour chauffer l'air.

Présentation des données disponibles

Malheureusement, très peu de données représentatives de cette classe d'appareils ont pu être collectées pendant la thèse. Nous disposions uniquement d'un seul cas d'étude issu de la Maison Domotique. De plus, un seul essai avait été réalisé.

Des signaux collectés récemment (K_4- 2008/2009) dans le cadre du projet R&D STABILO seront exploités dans des travaux futurs pour une caractérisation plus fine de ces signaux.

Dans la suite nous présenterons à titre d'illustration des signaux issus de ces dernières campagnes de mesure. Mais, aucune analyse approfondie n'a été réellement menée sur cette classe de signaux.

2.2.5.2 Analyse de signaux

a) Lave-linge La figure (2.23) présente trois exemples de signaux de lave-linge issus de trois appareils différents. Le signal observé est un mélange de deux signaux élémentaires : des régulations rapides (de l'ordre d'une minute) et de faible amplitude (environ 100 W) et des créneaux d'amplitude élevée ($\geq$ 2000 W). Ce dernier signal correspond *a priori* à la puissance appelée par la résistance de chauffage de l'eau.

On note :

- ➢ que la première phase de chauffage est souvent la plus longue ;
- ➢ qu'il existe des cycles de plus courte durée qui correspondent à des relances de chauffage de l'eau.

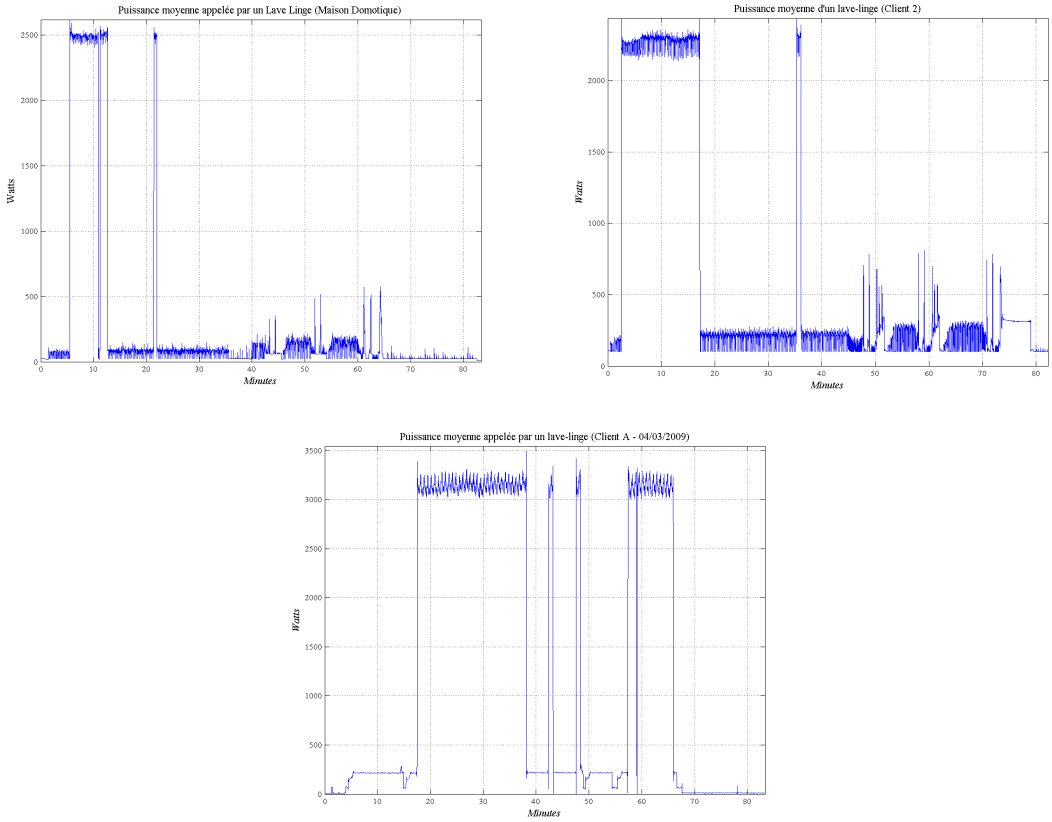


FIGURE 2.23: Puissance moyenne appelée par un lave-linge ($T_e = 1\ s$)

b) Lave-vaisselle Nous disposions d'un seul exemple de cette classe d'appareils. La puissance moyenne appelée par un lave-vaisselle ($T_e = 1$ s) est illustrée sur la figure (2.24). L'analyse de plusieurs cycles de fonctionnement montre :

- ➢ que le signal observé est un mélange de deux familles de signaux : un signal de faible amplitude (environ 200 W) correspond à la puissance appelée par le moteur et un signal de forte amplitude (≥ 2000 W) correspond à la puissance appelée par la résistance de chauffage de l'eau.
- ➢ que la puissance appelée par la résistance présente une régulation en tout ou rien, où la durée *on-off* Δt_{10} varie de quelques minutes à une demi-heure.

L'analyse de plusieurs cycles de fonctionnement n'a permis d'identifier aucune autre propriété particulière de ce signal.

c) Sèche-linge Nous disposions de la signature d'un seul sèche-linge (Maison Domotique) et d'un seul essai, le signal mesuré est présenté sur la figure (2.25).

- ➢ Sur l'intervalle $[0,\ 15\ \text{mn}]$: le signal observé est un mélange d'un créneau ($A_1 = 200$ W) et d'un créneau quasi-périodique ($A_2 = 100$ W). Il s'agirait (à dire de spécialiste) du mélange des puissances appelées respectivement par la résistance et par un moteur.
- ➢ Sur l'intervalle $[15\ \text{mn}, 57\ \text{mn}]$: le signal observé est un mélange de deux créneaux périodiques d'amplitude A_1 et A_2. Le caractère quasi-périodique de ces signaux est mis en évi-

dence par le spectre du signal (2.25). La pseudo-période est d'environ $T_0 = 80$ s. Rappelons qu'il s'agit de l'une des périodicités caractéristiques des convecteurs électriques.
  - Par conséquent, on peut identifier une indétermination entre un signal de convecteur et ce segment du signal associé au sèche-linge. Le sèche-linge sera *a priori* une source potentielle de fausse alarme lors de la détection du chauffage électrique.

➢ Sur l'intervalle $[57\text{ mn}, 64\text{ mn}]$: on observe un créneau quasi-périodique ($T_0 = 80$ s) de faible amplitude, qui correspondrait à la puissance appelée par la pompe.

Remarquons également que les amplitudes des signaux sont très proches de celles du lave-linge ou encore du lave-vaisselle.

Il est évident que l'on ne peut pas généraliser les observations tirées de cet essai. Mais les signaux disponibles en 2008 ont permis de confirmer ces observations, tel qu'illustrés par la figure (2.26) (a) où le signal global observé est la puissance appelée par l'ensemble des appareils dit gros électroménagers. Les deux zooms confirment les observations précédentes et permettent d'estimer les amplitudes maximales des signaux (sèche-linge : 2500 W ; lave-linge : 2000 W).

Le foisonnement d'un sèche-linge et d'un lave-linge est illustré par la figure (2.26) (b). En effet, le signal observé sur la **Zone 2** correspond à la signature d'une fin de cycle de fonctionnement. Le signal observé sur la **Zone 1** est un mélange de créneaux d'amplitudes de l'ordre de 2500 W et 2000W. Il s'agit donc bien d'un foisonnement des signaux du lave-linge et du sèche-linge.

En résumé Les puissances moyennes appelées par les appareils du groupe gros électroménagers échantillonnées à une fréquence de l'ordre du hertz ont les propriétés suivantes :

➢ *Les signaux observés sont eux-mêmes des signaux composites. Chaque mélange est constitué d'une composante résistive de forte amplitude et d'un signal de faible amplitude* (*souvent associé à un moteur*).

➢ *La puissance nominale de ces appareils n'est pas caractéristique de cette classe d'usage et ne permettrait pas de les discriminer.*

➢ *Le sèche-linge a une régulation quasi-périodique, identique à celle des convecteurs.*

➢ *Le lave-linge et le lave-vaisselle ont une régulation en tout ou rien.*

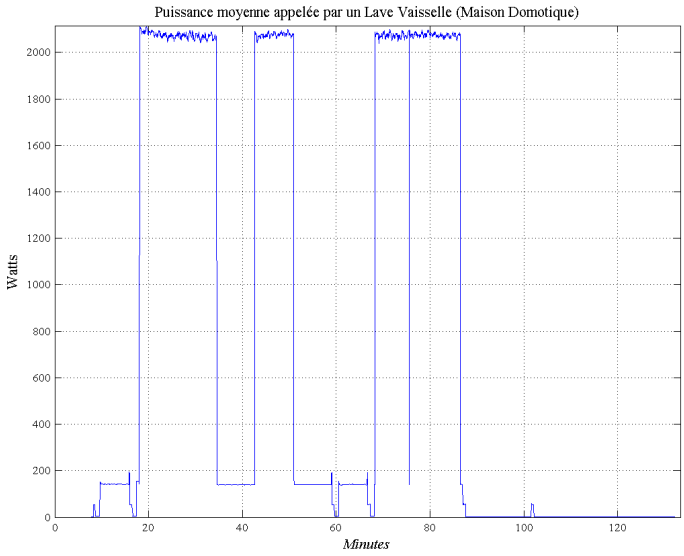


FIGURE 2.24: Puissance moyenne appelée par un lave-vaisselle (Maison Domotique -K_1,$T_e = 1\ s$)

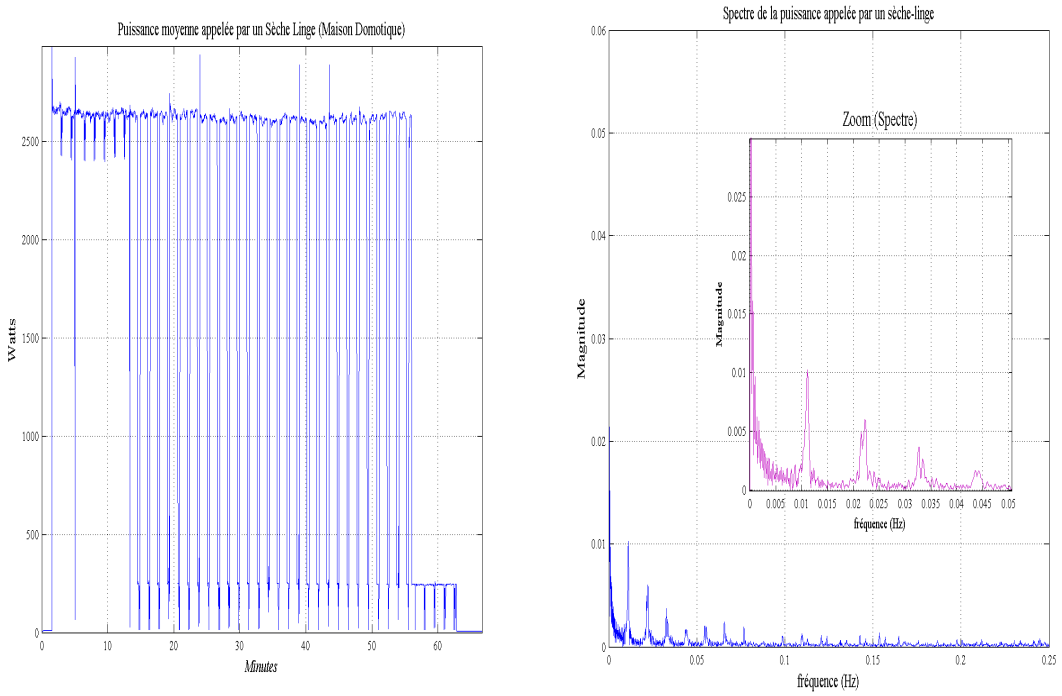
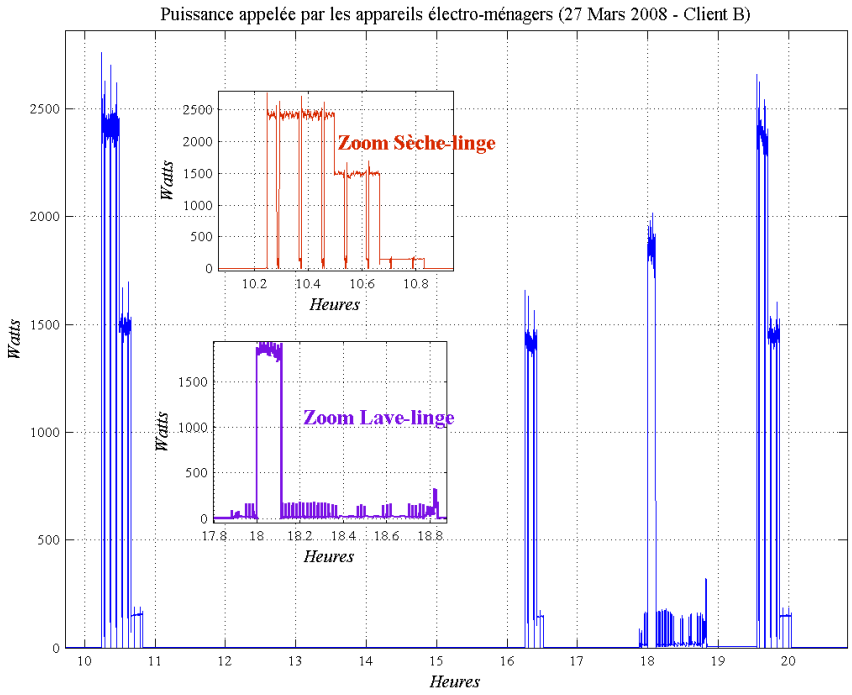


FIGURE 2.25: Puissance moyenne appelée par un sèche-linge ($T_e = 1$ s) et son spectre

a) Groupe des appareils électroménagers



b) Lave-linge

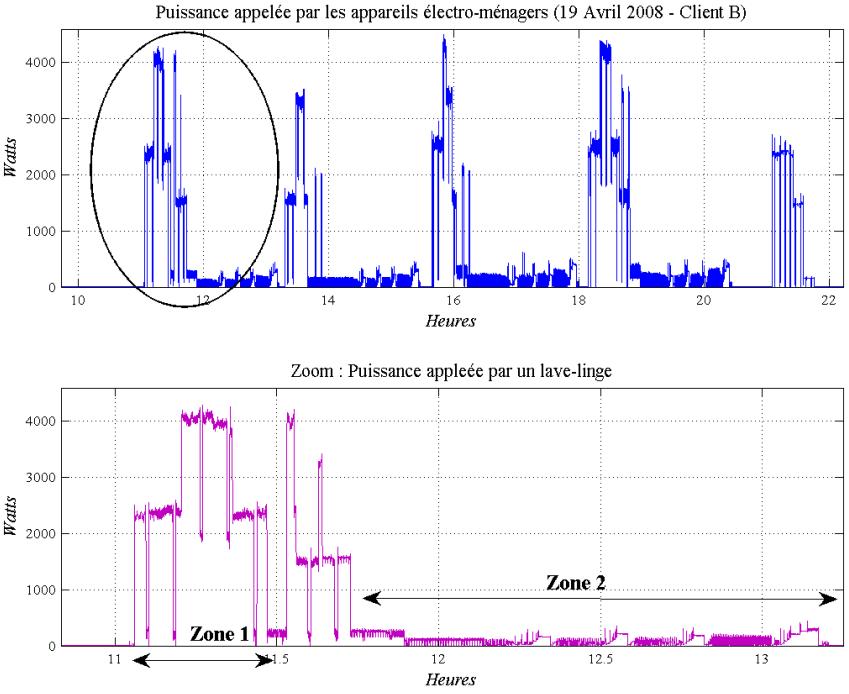


FIGURE 2.26: a) Puissance moyenne appelée par le groupe des gros életroménagers : (Client B -K_2,$T_e = 1s$)
b) Puissance moyenne appelée par un lave-linge (K_4,$T_e = 1$ s) : plusieurs cycles de fonctionnement (Client B)

2.2.6 Les appareils de cuisson

Les appareils de cuisson présentent un caractère à dominante résistive. Bien que ces appareils ne faisaient pas partie des usages ciblés, nous nous sommes intéressés à leur signature électrique afin d'anticiper leur impact sur la caractérisation des trois usages ciblés. La figure (2.27) illustre la puissance moyenne appelée par un micro-ondes, un four et une plaque de cuisson.

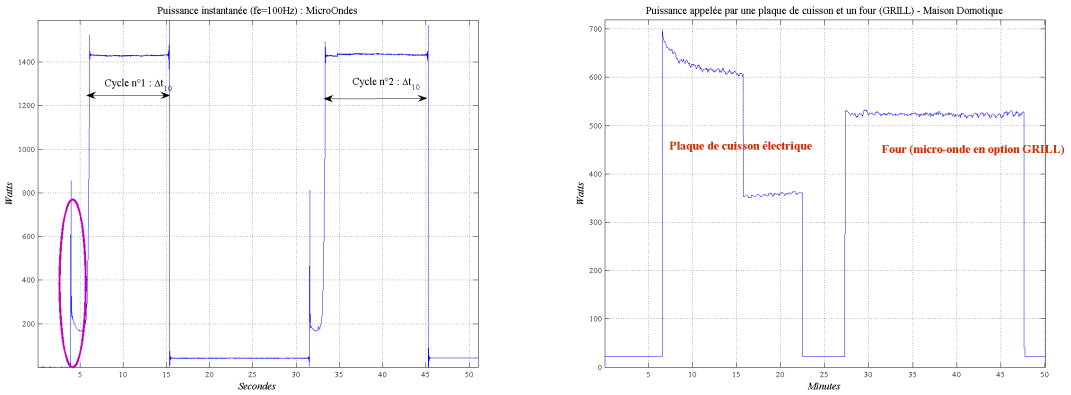


FIGURE 2.27: Puissance appelée par les appareils de cuisson ($T_e = 1$ s)

a) Le four

Le caractère purement résistif du four se traduit par une signature sous forme d'un créneau symétrique [3].

b) Le micro-ondes

La puissance appelée par un micro-ondes est un signal composite issu du mélange d'une composante purement résistive et d'une composante purement inductive (moteur assurant la rotation du plateau de l'appareil). Dans ce cas, le signal observé est constitué de deux cycles de fonctionnement. Remarquons que :

- ➢ La durée d'un cycle (τ_i^m) est de l'ordre de 10 s.
  - – c'est une durée proche de la durée de cycle que peut avoir un convecteur électrique.
- ➢ La durée entre deux cycles successifs (t_i^+) est d'environ 30 s. Ici, on considère uniquement les événements associés à la composante résistive.
  - – la puissance appelée par un micro-ondes serait-elle pseudo-périodique ?
  - – cette durée est du même ordre de grandeur que l'une des périodicités observées pour la classe des convecteurs électriques.

Compte tenu de l'importance de la caractérisation de l'usage chauffage, nous nous sommes alors penchés sur l'analyse des signaux de cette source potentielle de fausse-alarme pour le chauffage, plus précisément sur l'analyse des durées (t_i^+), (t_i^-) et $(t_i^\dagger)$.

3. il s'agit d'une symétrie entre l'enclenchement et le déclenchement de l'appareil

Les données exploitées pour cette analyse sont issues de 4 appareils différents et un total de 10 essais. Les essais correspondent à des utilisations différentes (contenu, consigne, mode d'utilisation, etc.).

L'histogramme (2.28) des durées entre événements successifs de même nature (de la composante résistive) met en évidence :

- ➢ *la présence d'une pseudo-période, qui est en plus commune à l'ensemble des 4 appareils considérés* ;
- ➢ *cette pseudo-période est très proche d'une période de régulation de convecteur* ;
- ➢ *la durée moyenne d'un cycle de fonctionnement est de l'ordre d'une minute.*
  - Cette indétermination possible entre un convecteur et un micro-ondes peut être levée si l'on peut exploiter le signal du micro-ondes appelé par la charge inductive.

c) La plaque de cuisson

Le signal observé met en évidence les propriétés suivantes, dont la généralisation nécessitera une base de signaux « plus conséquente » :

- ➢ *le caractère inductif (il s'agissait d'une plaque à induction) se traduit par une décroissance exponentielle au démarrage de l'appareil* ;
- ➢ *la variation de l'amplitude du signal en fonction de la consigne fixée (ici deux niveaux de consommation).*

2.2.7 Analyse d'exemples de CdC

2.2.7.1 Cartographie des principaux usages chez deux clients particuliers

Rappelons que la terminologie Courbe de Charge générale désigne la consommation totale d'une installation électrique. Dans cette partie, la grandeur électrique est la puissance moyenne (active ou apparente) appelée par l'installation.

A titre d'illustration et pour fournir des ordres de grandeur, la figure (2.29) présente l'énergie quotidienne consommée chez un client particulier pendant le mois de janvier, ainsi que l'énergie consommée par le chauffage et l'eau chaude sanitaire. Précisons que l'énergie consommée par ces usages est l'énergie de référence évaluée à partir de la puissance moyenne mesurée pour chacun des usages.

La figure (2.30) présente la part de chaque usage par rapport à la consommation globale. On distingue deux profils de consommation en présence ou non du client.

La consommation globale décroît notablement du 7 janvier au 11 janvier. En effet, le client était absent. Néanmoins, sa consommation n'est pas complètement nulle, elle est de l'ordre de 15 kWh par jour. En effet, les deux usages suivants continuent à fonctionner :

- ➢ l'ECS fonctionne (relances du chauffe-eau) pour compenser les déperditions thermiques ;
  - elle représente plus de 20% de la consommation totale (2.30)
- ➢ le froid alimentaire ;
- ➢ certains convecteurs ;
  - dont la consommation représente environ 7% de la consommation totale (2.30)
- ➢ certains appareils en mode de veille.

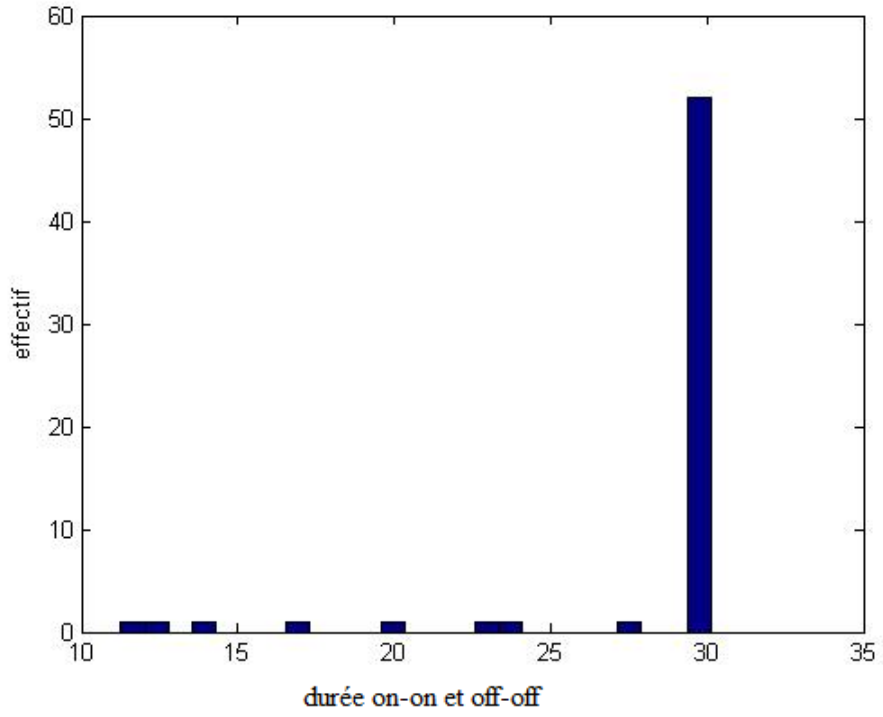


FIGURE 2.28: Histogramme des durées de l'usage micro-ondes (K_4)

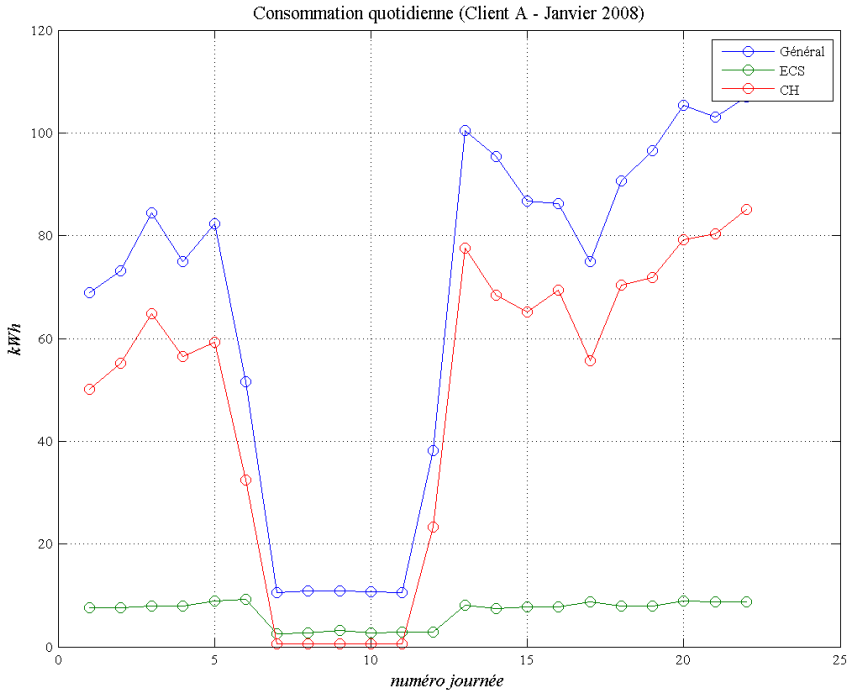


FIGURE 2.29: Consommation quotidienne du client A : janvier 2008

En présence du client, on constate que le chauffage représente souvent plus que 50% de sa consommation quotidienne, alors que l'ECS représente environ 10% de la consommation quotidienne.

La part de chaque usage dans la facture du client dépend du tarif choisi. Dans le cas d'un tarif de base, la cartographie de la facture par usage est identique à la cartographie de la consommation totale par usage.

Dans le cas d'un tarif Heure Creuse / Heure Pleine (HP/HC), la cartographie détaillée de la facture dépend de l'énergie consommée sur chacune des plages horaires.

La part du froid alimentaire dans la consommation quotidienne a été étudiée à partir des données du client B à partir de journées (2.31) de la période comprise entre janvier 2008 et avril 2008.

La part de cet usage dans la consommation totale est faible, mais l'énergie quotidienne consommée par cet usage n'est pas négligeable. En effet, le client a un chauffage « tout électrique » et deux ballons d'eau chaude sanitaire. L'énergie consommée par ces deux usages est très élevée, ce qui explique cette faible part de la consommation du froid alimentaire.

Cet exemple particulier nous amène à la question suivante : quelle est la grandeur utile et significative à fournir au client pour décrire la consommation de chaque usage ? On peut envisager par exemple de lui fournir :

- ➢ l'énergie consommée par usage ;
- ➢ la part (pourcentage) de l'energie consommée dans la consommation totale de chaque usage ;
- ➢ le coût de consommation par usage ? Ce point n'a pas été étudié dans le cadre de la thèse, mais mériterait d'être approfondi.

2.2.7.2 Propriétés de la CdC générale et représentations adaptées

Après avoir analysé séparément les signaux de chaque classe d'appareils, nous présenterons ici quelques exemples de CdC de clients particuliers afin de mettre en évidence leurs caractéristiques. Des modes d'analyse pour mettre en évidence certaines caractérisques des usages ciblés sont également présentés.

La CdC : un foisonnement de signaux non-stationnaires et localement stationnaires

La CdC (puissance moyenne) échantillonnée à une fréquence de l'ordre du hertz peut être vue comme la superposition de plusieurs séries d'événements. Chaque série d'événements est en réalité la suite d'enclenchements et de déclenchements d'un appareil.

Les événements de la CdC dépendent :

- ➢ du comportement du client, qui peut être régulier ou complètement aléatoire ;
  - exemple : heure d'utilisation des appareils de cuisson, de l'éclairage, de la télévision.
- ➢ des conditions d'utilisation de l'appareil (consignes, contenu) : aléatoire ;
  - exemple : micro-ondes, réfrigérateur, etc.
- ➢ du modèle de fonctionnement de l'appareil : déterministe ;
  - exemple : séquencement des opérations (tâches) d'un lave-vaisselle.
- ➢ d'autres contraintes liées au climat et aux caractéristiques du bâtiment par exemple.
  - exemple : chauffage électrique.

La figure (2.32) présente un exemple de CdC chez un client ayant uniquement deux convecteurs d'appoint. Précisons qu'il s'agit d'une CdC pour laquelle nous ne disposons d'aucun sous-comptage. On constate que lorsque le chauffage est en marche, le nombre d'événements

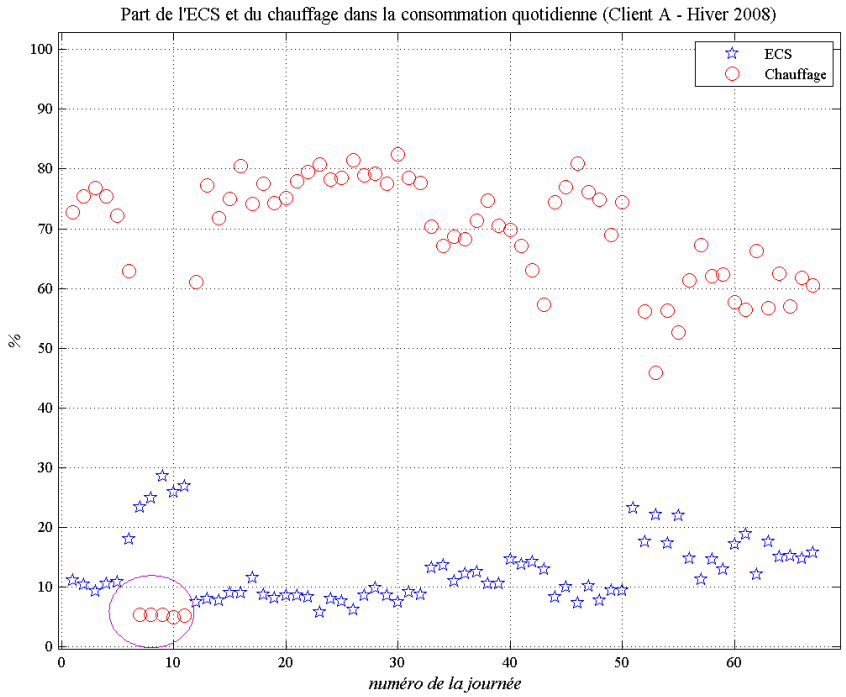


FIGURE 2.30: Part du chauffage et de l'ECS dans la consommation quotidienne chez le client A

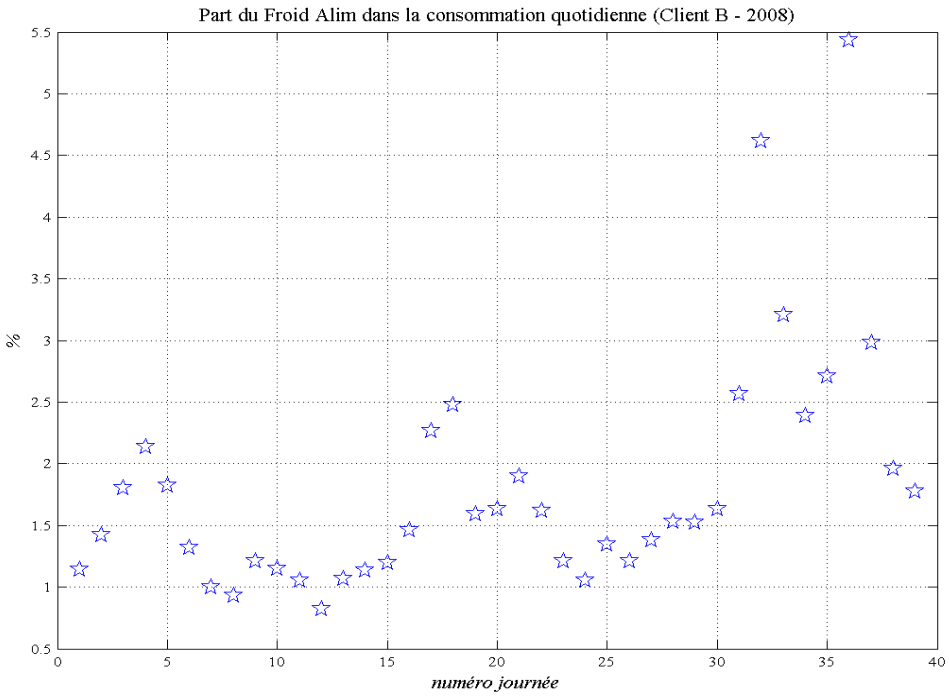


FIGURE 2.31: Part du froid alimentaire dans la consommation quotidienne chez le client B

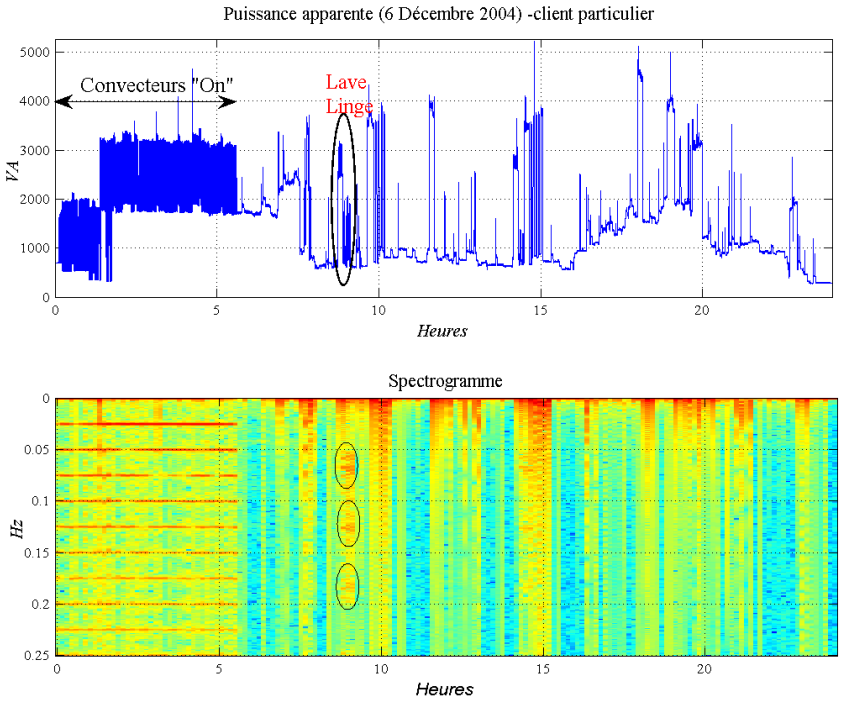


FIGURE 2.32: Exemple de CdC chez un client particulier

est élevé : les événements ne sont plus lisibles à l'œil nu. Ce qui n'est plus le cas quand le chauffage est éteint. Comme la puissance d'un convecteur est un signal pseudo-périodique de rapport cyclique lentement variable, la puissance du chauffage est un signal localement stationnaire. En absence de chauffage électrique, l'exemple présenté met en évidence le caractère non-stationnaire de la courbe de charge.

Un mode d'analyse adapté pour mettre en évidence la présence de chauffage et les périodicités est le spectrogramme (2.32). La présence de chauffage se traduit par la présence de raies horizontales équidistantes. Cet exemple met en évidence l'apport d'une représentation du type spectrogramme pour la mise en évidence de la présence du chauffage. La figure (2.32) met en évidence la présence de « motifs spectraux » (entourés avec des ellipses noires) plus étalés spectralement que les raies spectrales des convecteurs. Le support spectral de ces motifs a une intersection non nulle avec le support des raies des convecteurs. Le zoom sur le signal temporel présenté sur la figure (2.33) met en évidence la présence d'un lave-linge, dont les régulations pseudo-périodiques génèrent les motifs observés. Les deux zooms présentés sur la figure (2.33) mettent également en évidence le foisonnement des usages électriques. La même analyse est effectuée pour une CdC, où certains sous-comptages sont instrumentés. La plage horaire de fonctionnement d'un convecteur est entourée avec une ellipse noire sur la figure (2.34), dont la signature sur le spectrogramme est mise en évidence.

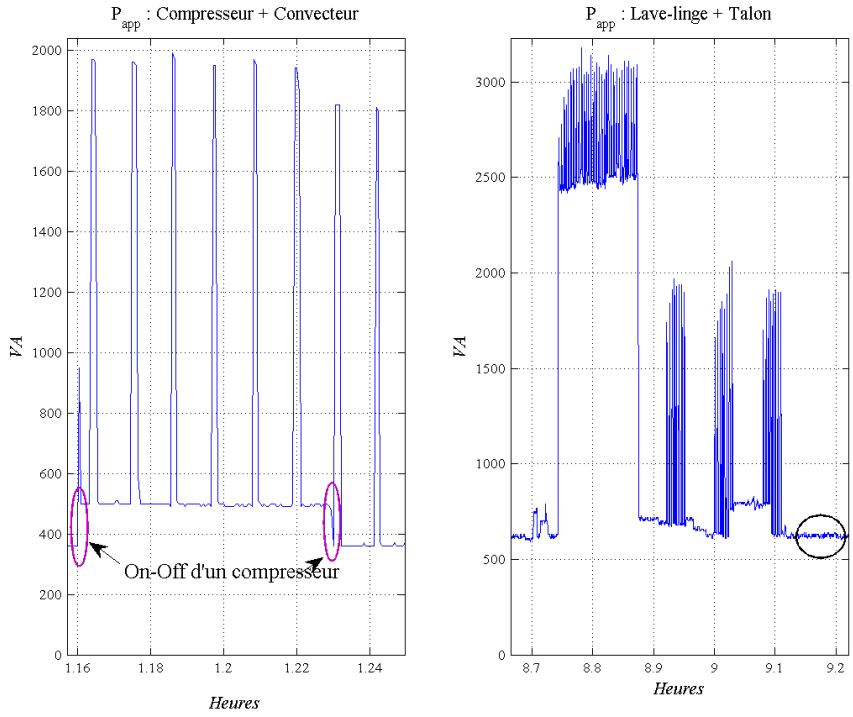


FIGURE 2.33: Zoom - foisonnement d'usages

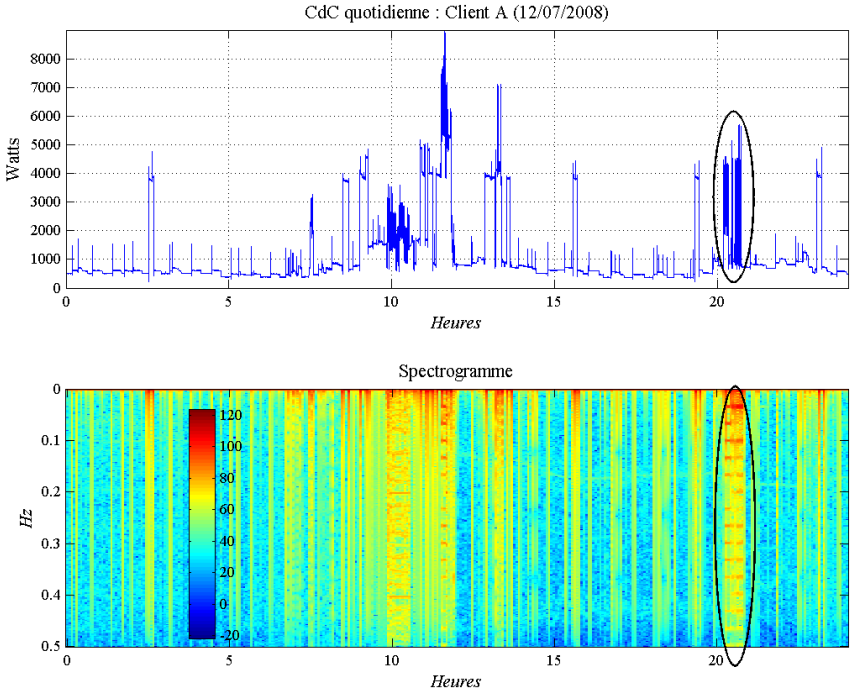


FIGURE 2.34: CdC quotidienne - 12 juillet 2008 (Client A)

Cette CdC ne contenant que peu de chauffage et par conséquent très peu d'événements peut sembler facile voire évidente à caractériser. Cependant certaines sources d'indétermination entre usages sont inévitables. Les signaux correspondant aux sous-comptages ECS, cuisson et lave-vaisselle, illustrés sur (2.35) mettent en évidence les difficultés suivantes :

- ➢ Le chauffe-eau et le lave vaisselle présentent des signaux très similaires, plus particulièrement ils ont ici la même amplitude.
- ➢ Le zoom sur la plage horaire de $[11\text{ h}, 14\text{ h}]$ montre que l'enclenchement du deuxième cycle de l'ECS coïncide avec le déclenchement de la résistance du lave-linge :
  - cette superposition donne lieu à deux événements manquants : un *on* de l'ECS et un *off* du lave-vaisselle. Il faudra donc pouvoir gérer les collisions d'événements.
- ➢ Les courbes des puissances moyennes des deux réfrigérateurs montrent que les trois compresseurs ont la même amplitude.
  - cette propriété aurait pour conséquence une forte indétermination sur l'appariement des événements même si l'on dispose de la somme des trois signaux uniquement.
  - l'amplitude des pics générés par le démarrage des compresseurs pourrait être un paramètre discriminant entre les deux appareils.
- ➢ La courbe de charge des autres appareils se caractérise par la présence d'une composante d'amplitude constante (500 W)
  - cette composante est nette sur la plage horaire $[00\text{ h}, 5\text{ h}]$;
  - il s'agit d'un exemple de talon de consommation ;
  - cet exemple met également en évidence l'importance de la part de la consommation des appareils compris dans le talon : ici l'amplitude est supérieure à celle d'un réfrigérateur et le talon est présent en permanence.

2.2.7.3 Analyse événementielle pour la reconnaissance du chauffage

Comme mentionné précédemment, la régulation rapide du convecteur en comparaison avec les autres appareils électriques ainsi que les durées de fonctionnement ont pour conséquence un nombre d'événements plus élevé quand le chauffage est en marche [4].

Une analyse des événements de la CdC totale peut se résumer au comptage du nombre d'événements sur une fenêtre d'analyse donnée (analogie avec la Transformée de Fourier Court Terme). On introduit une densité d'événements illustrée sur la figure (2.36).

La courbe représentant la densité des événements évaluée à partir de la CdC globale a des valeurs plus significatives sur les plages horaires où le chauffage est en marche. **Une transformée utile pour la détection du chauffage serait alors la densité d'événements.**

4. Exception : tous les convecteurs sont en régime de saturation en permanence

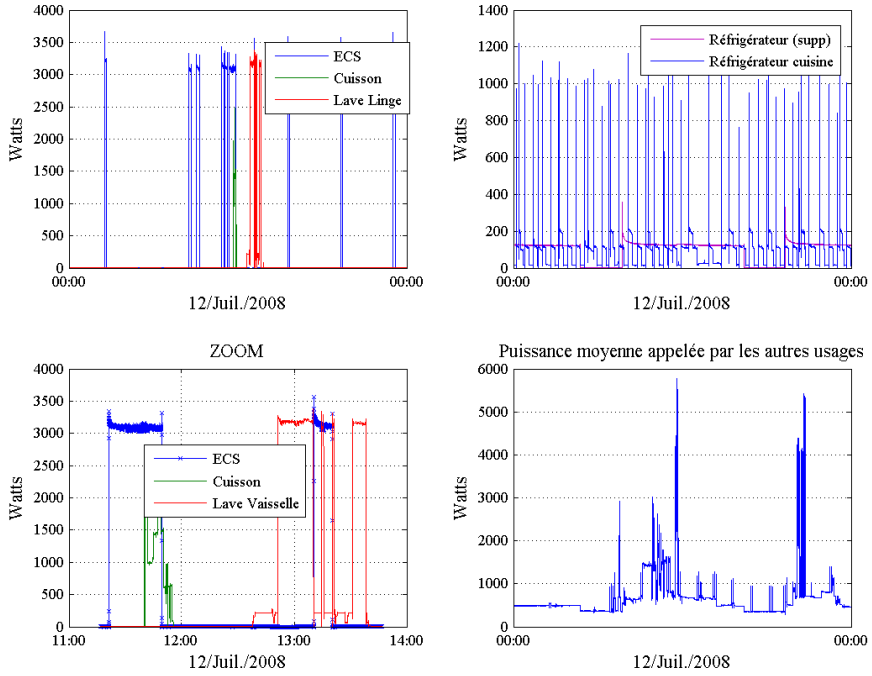


FIGURE 2.35: Composantes de la CdC quotidienne - 12 juillet 2008 (Client A)

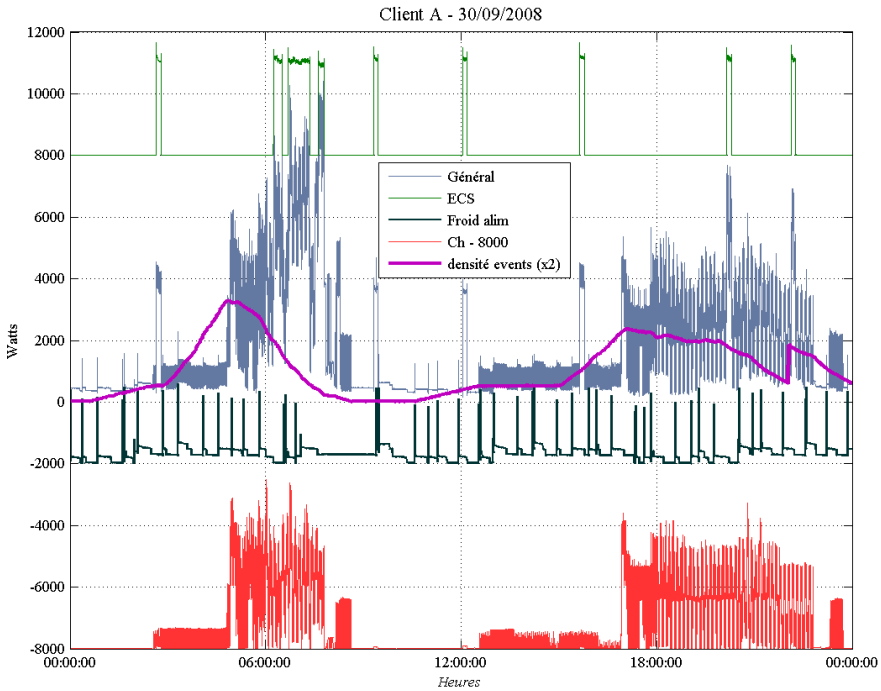


FIGURE 2.36: CdC quotidienne - 30 septembre 2008 (Client A)

2.2.7.4 Foisonnement du chauffage et du chauffe-eau

L'impact d'une forte consommation du chauffage sur la CdC est illustrée sur la figure (2.37). Si l'on observe la puissance active de du chauffe-eau, on constate que malgré le foisonnement des deux principaux usages :

- ➢ *la présence de l'ECS se traduit par des variations importantes, brusques et quasi-régulières de la puissance active générale ;*
- ➢ *la caractérisation de l'ECS à partir de la puissance totale serait a priori plus simple sur la plage horaire allant de minuit jusqu'à 5 h.*

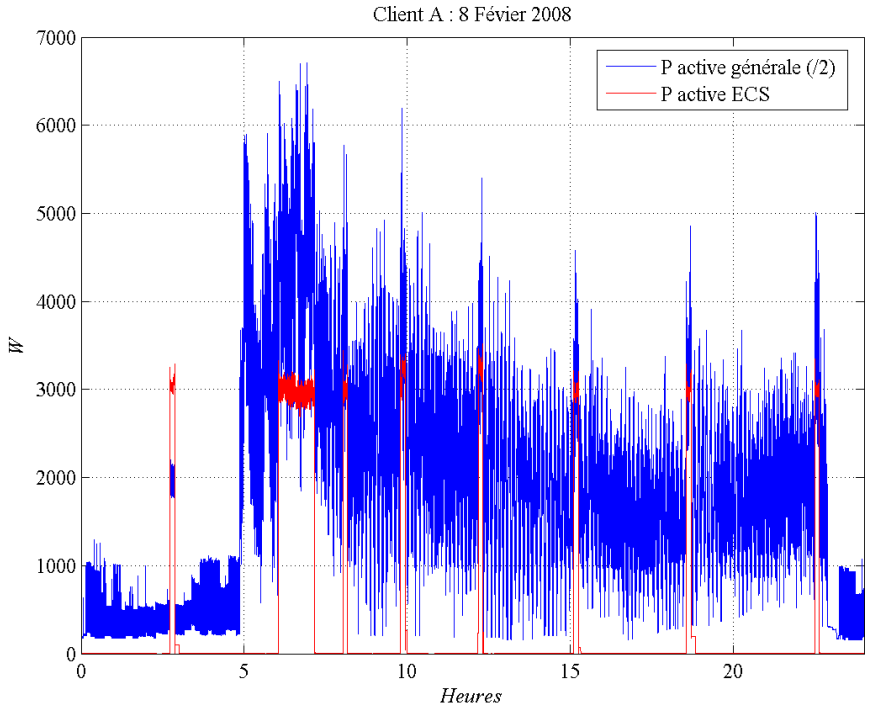


FIGURE 2.37: Exemple de CdC quotidienne générale + ECS (Client A)

Synthèse de l'analyse des signaux d'usages électriques

Bien qu'effectuée sur des données issues de trois maisons témoins uniquement, cette étude des signaux visait à étudier la reproductibilité des signaux des usages électriques, puis identifier les descripteurs discriminants entre usages et les représentations adaptées dans une optique de caractérisation des usages. L'analyse des signaux (puissance active ou puissance apparente moyenne) des appareils électriques met en évidence les paramètres discriminants (forme des signaux, type de régulation, amplitude, etc.) entre certains usages et permet d'identifier des similarités entre certains signaux (puissance moyenne d'un convecteur et celle d'un sèche-linge par exemple).

Nous proposons une description quantitative de certains usages à travers l'analyse de certains paramètres tels que les variations du rapport cyclique d'un convecteur, la péridocité des signaux de convecteurs, la symétrie entre les *on* et les *off* des convecteurs, l'asymétrie entre le *on* et le *off* d'un cycle de fonctionnement de l'ECS.

L'analyse des CdC générales montre qu'il est souvent difficile d'interpéter la CdC dans son domaine d'observation direct (le domaine temporel). Cependant, il existe pour chaque usage au moins un mode de représentation où la reconnaissance de l'usage (interprétation qualititative) est plus simple que dans une analyse directe. Par exemple, le spectrogramme et de la densité des événements de la CdC permettent de reconnaître « une signature » caractéristique des usages à régulation très rapide, en l'occurence le chauffage.

En résumé, une base de connaissances apprises est constituée grâce à l'analyse des signaux réels. Les informations extraites sont de différentes natures. Il s'agit :

- ➢ de descripteurs discriminants entre usages ;
- ➢ de domaines transformés où la séparation des usages est envisageable ;
- ➢ des similarités entre signaux et par conséquent des potentielles sources de confusion entre usages ;
- ➢ d'une description temporelle des signaux.

Orientations méthodologiques

L'analyse des signaux a permis d'extraire des connaissances complémentaires sur les usages de natures différentes : densité de probabilité d'un paramètre donné, modèle paramétrique des signaux ou encore des transformées permettant de discriminer certains usages. Ces informations nous permettent de tirer profit d'une part du cadre générique des approches inverses bayésiennes [Ia01] pour intégrer les connaissances *a priori* sur certaines grandeurs et de la méthodologie générale de séparation de sources mono-capteur fondée sur l'interprétation du mélange à partir d'une représentation dans un domaine transformé [BB03, BGB01]. Une présentation des deux approches et des principaux travaux identifiés dans la littérature sont fournis en annexe (E).

Les trois chapitres qui suivent sont respectivement dédiés à la segementation de la CdC, l'estimation des paramètres des usages et la décomposition de la CdC. Ils s'appuient sur les deux méthodologies citées précédemment et sur les connaissances apprises grâce à l'analyse des signaux. Dans chaque chapitre, nous présentons la méthode proposée, des critères d'évaluation des algorithmes développés et les premies résultats obtenus à partir des signaux réels.

Chapitre 3

Segmentation automatique de la courbe de charge

Introduction

Connaître le domaine de présence de certains usages est utile pour le développement de services tels que la détection d'anomalies (mauvaise utilisation d'un usage, le maintien à domicile de personnes âgées par exemple), l'optimisation des scénarios de gestion d"une installation donnée ainsi que l'analyse détaillée de l'énergie quotidienne consommée par usage (chapitre 4) . Dans ce chapitre, nous présentons une approche automatique de segmentation de la CdC résidentielle fondée sur la fusion de l'ensemble des détecteurs dédiés aux différents usages d'intérêt, ces derniers étant appliqués indépendamment les uns des autres à la CdC. Cette approche, dite approche « classe /autre » est déclinée dans le cas où les usages étudiés sont le chauffage, le chauffe-eau et le froid alimentaire. Nous présentons également les premiers résultats de l'évaluation des algorithmes de détection à partir de signaux réels issus de trois maisons témoins.

3.1 Méthodologie de segmentation de la CdC

3.1.1 Position du problème

La CdC résidentielle est la superposition des signaux des différents appareils électriques, un intervalle de temps donné peut donc être alloué à des composantes différentes. Par exemple, le domaine de présence du chauffage est souvent inclu dans le celui du froid alimentaire. Segmenter la puissance moyenne d'une installation électrique consiste à déterminer le domaine de présence (support temporel) de chaque usage (ou appareil) à partir de la CdC générale. Précisons que la définition du domaine de présence dépend de la nature de l'usage étudié. En effet, il peut s'agir :

- ➢ soit de l'union des intervalles de temps pendant lesquels la puissance appelé par l'usage est strictement positive (exemple : le chauffe-eau, sèche-linge, etc.) ;
- ➢ soit de l'union des intervalles de temps pendant lesquels l'usage est en marche, mais sa puissance peut s'annuler de par son mode de régulation (exemple : le chauffage, le froid alimentaire, etc.).

Aucune cohérence entre les domaines de définition des usages n'est requise étant donnée l'indépendance entre les usages qui contribuent à la CdC. Les classes d'usages présents ainsi que leur nombre (noté N) sont inconnus. Néanmoins des usages d'intérêt sont prédéfinis dans le contexte de la thèse : le chauffage par convecteurs électriques, le chauffe-eau (**ECS**) et le froid alimentaire.

3.1.2 Approche envisagée

Le problème générique est celui de la segmentation d'un signal non-stationnaire, les classes recherchées étant préalablement définies. Ce problème a fait l'objet de nombreux travaux qui s'appuient souvent sur l'utilisation de représentations temps-fréquence adaptées aux signaux étudiés [MD04]. Deux approches sont envisageables. La première approche consiste à considérer le signal observé dans un même et unique espace de représentation, la segmentation étant effectuée par exemple par segmentation morphologique [PM95] d'une représentation temps-fréquence ou en utilisant un modèle statistique du mélange dans le domaine transformé choisi [Vin95, Hor, DDT02]. Une application à la segmentation signaux sonores à partir d'une seule observation fondée sur un modèle statistique du mélange est présentée dans [PGB]. La deuxième approche consiste à identifier des domaines transformés dédiés aux différentes classes. Le problème initial est ainsi ramené à N problèmes de détection, les résultats des détecteurs constituant les entrées d'un module de décision.

Cette approche est dénommée approche « classe / autre ». Cette approche présente les deux avantages suivants : elle permet d'améliorer la discrimination d'un usage donné sans en détériorer les performances de discrimination des autres composantes et elle offre la possibilité d'intégrer des classes supplémentaires sans modifier la procédure existante.

La segmentation de la CdC présente deux particularités qui justifient le choix d'une méthodologie générale du type « classe / autre ». En effet, l'analyse des signaux des usages 2 a permis d'identifier des représentations (ou fonctions d'observation) permettant de discriminer certains usages. Par ailleurs, bien que certains usages soient choisis, l'approche proposée doit pouvoir s'enrichir progressivement par l'intégration d'autres usages.

Le schéma général décrivant l'approche dédiée à la segmentation de la CdC en N classes est présenté sur la figure (3.1), où chaque module de détection se résume par le test d'hypothèse suivant :

$$\begin{cases} \mathscr{H}_1 & : \text{l'usage ciblé est } \textbf{actif} \text{ (en marche)} \\ \mathscr{H}_0 & : \text{l'usage ciblé } \textbf{n'est pas actif} \text{ (appareil éteint ou inexistant dans l'installation)} \end{cases}$$

Dans ce chapitre, nous nous focalisons sur le chauffage et le chauffe-eau. La segmentation de la CdC est décomposée en deux modules : un détecteur du chauffage (chauffage / autre usage) et un détecteur du chauffe-eau (ECS / autre). La segmentation du signal, illustrée sur la figure (3.1), s'effectue en fusionnant les résultats de segmentation intermédiaires de sorte que le domaine temporel soit découpé en quatre domaines :

- ➢ un domaine temporel où le *chauffage et chauffe-eau* fonctionnent simultanément ;
- ➢ un domaine temporel où le *chauffage est actif* alors que que le *chauffe-eau ne l'est pas* ;
- ➢ un domaine temporel où le *chauffe-eau est actif* alors que que le *chauffage ne l'est pas* ;
- ➢ un domaine temporel *où le chauffage et le chauffe-eau* sont *tous les deux absents* (inactifs).

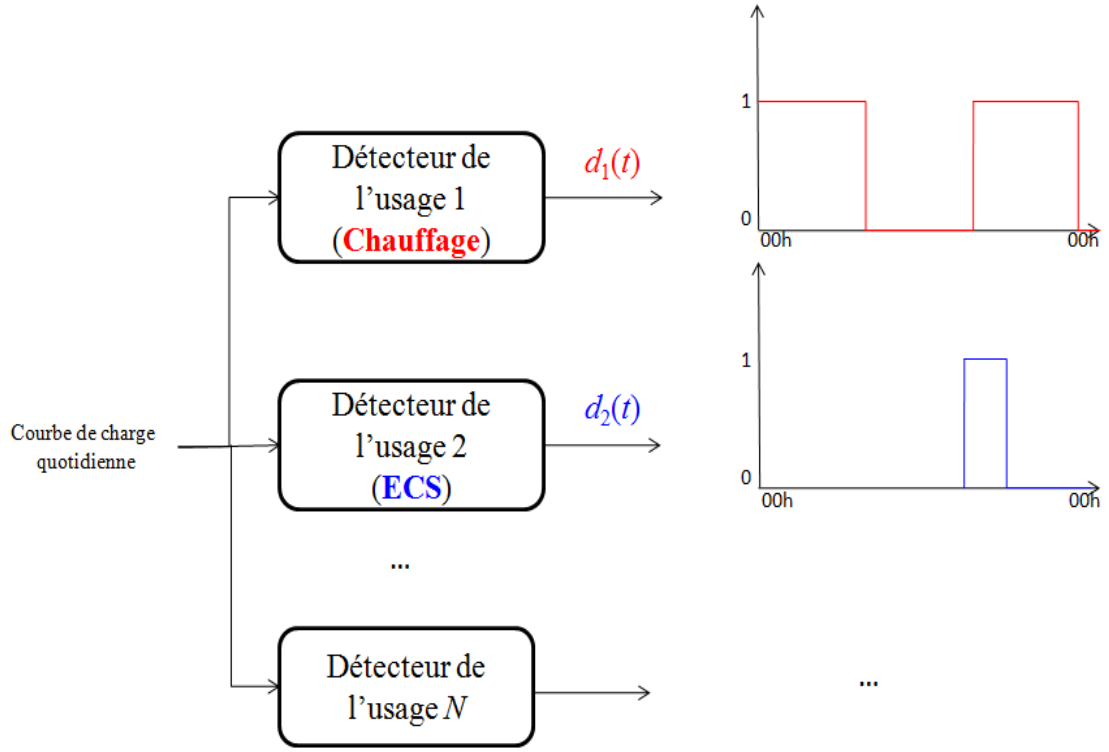


FIGURE 3.1: Schéma général de la segmentation de la CdC quotidienne résidentielle

Dans les sections suivantes, nous détaillons les détecteurs mis en oeuvre, les critères d'évaluation de chaque étape segmentation intermédiaire ainsi que les performances obtenus à partir des signaux réels.

3.2 Segmentation « chauffage / autre »

3.2.1 Domaines transformés adaptés

3.2.1.1 Le spectrogramme : un domaine de représentation adapté au chauffage

L'analyse des signaux élémentaires et des CdC générales a mis en évidence les deux propriétés suivantes du chauffage :

➢ le caractère quasi-périodique des signaux des convecteurs ;

➢ la régulation rapide du chauffage en comparaison avec les autres usages.

Compte tenu du caractère localement stationnaire[Fla93] de la CdC du chauffage, une repésentation temps-fréquence adaptée est la Transformée de Fourier Court Terme (TFCT) [Fla93], plus particulièrement son module : le *spectrogramme*. Dans le cas discret, chaque coefficient du spectrogramme est défini par l'équation suivante (3.1) :

$$\rho_y[k,\nu] = \left| \frac{1}{M} \sum_{m=1}^{M} h[m-k]\, y[m]\, e^{-2\pi j \frac{m\nu}{N_f}} \right|^2 \tag{3.1}$$

où y est le signal discret étudié, h est la fonction (fenêtre) d'analyse de taille N_f, M est la taille (nombre d'échantillons du signal analysé), k représente le temps et ν la fréquence réduite.

Dans cet espace de représentation, la présence de chauffage se traduit par la présence de raies horizontales équidistantes. On peut donc envisager de détecter le chauffage en segmentant sa représentation temps-fréquence. Le détecteur défini et mis en œuvre dans ce domaine est détaillé en (3.2.2).

3.2.1.2 La densité des événements

La régulation rapide des convecteurs (par rapport aux autres usages) se traduit par un nombre très élevé d'enclenchements et de déclenchements quand le chauffage est mis en marche. En effet, si un covecteur, un compresseur et un chauffe-eau sont en marche pendant un quart d'heure, l'ordre de grandeur du nombre maximum d'événements par appareil est respectivement 40, 5 et 2. Nous introduisons une grandeur, la densité des événements, qui sera utile à la détection du chauffage.

Définition

Soit $y(t)$ le signal observé , $y'(t)$ sa dérivée et t_{c0} la durée minimale de fonctionnement du chauffage (en pratique, t_{c0} est de l'ordre de 10 mn). Soit $g_s(t)$ une fonction d'analyse définie comme suit :

$$g_s(t, t_{c_0}) = \begin{cases} 1 \; si \; t \in \left[-\frac{1}{2}t_{c_0} \; : \; \frac{1}{2}t_{c_0}\right] \\ 0 \; ailleurs \end{cases} \tag{3.2}$$

La densité des événements s_y est définie ainsi :

$$s_y(t, t_{c_0}) = y'(t) \star g_s(t, t_{c0}) \tag{3.3}$$

où $\star$ désigne l'opérateur de convolution.

Le détecteur défini et mis en œuvre dans ce domaine est détaillé dans la section (3.2.3).

3.2.2 Détection du chauffage dans le domaine temps-fréquence

3.2.2.1 Une « nouvelle » méthode d'interprétation du spectrogramme

Un exemple de représentation temps-fréquence, ici le spectrogramme, d'une CdC générale en présence du chauffage est illustré sur la figure (3.2). Les raies horizontales équidistantes constituent la signature du chauffage sur le spectrogramme. Les motifs spectraux entourés avec des ellipses noires sur le spectrogramme correspondent à la signature du lave-linge dans ce domaine. On constate que ces derniers apparaissent à des fréquences communes avec celles du chauffage. En effet, la CdC d'un lave-linge présente une composante « régulations rapides » dont la période de régulation est quasiment égale à celle des convecteurs. Cette observation vaut également pour le sèche-linge qui présente une composante similaire (régulation de la résistance). Néanmoins,

la reconnaissance du chauffage est immédiate dans cet espace car les raies horizontales associées couvrent tout le domaine fréquentiel d'analyse, ce qui n'est pas le cas pour le lave-linge. Nous proposons alors de caractériser le chauffage à partir de cette représentation afin déduire le support temporel de cet usage.

De manière générale, l'interprétation d'une représentation temps-fréquence consiste en deux processus :

- ➢ segmentation de la représentation temps-fréquence ;
- ➢ classification des motifs spectraux détectés.

Plusieurs approches paramétriques et non-paramétriques, supervisées ou non-supervisées sont proposées dans la littérature [DDC94, Vin95, PM95]. Nous investiguons une nouvelle approche semi-paramétrique et non-supervisée de segmentation du spectrogramme illustrée sur la figure (3.3).

Nous détaillerons cette approche [GDL+09] dans le cas de la caractérisation d'un nombre inconnu de signaux « quasi-périodiques » localement stationnaires [Fla93] à partir d'une observation unique. La démarche proposée consiste en six étapes (détaillées dans les paragraphes qui suivent) :

- ➢ représenter la courbe de charge $y(t)$ dans le domaine temps-fréquence (par un spectrogramme par exemple ρ_y) ;
- ➢ segmenter le spectrogramme obtenu en utilisant un détecteur de non stationnarités dans l'espace temps-fréquence. Le détecteur mis en oeuvre est fondé sur la modélisation statistique de la TFCT [Hor] et permet d'obtenir une carte temps-fréquence constituée par l'ensemble des motifs spectraux pertinents. Ce détecteur est paramétré par la probabilité de fausse alarme (fixée) P_{fa}. Cette étape est détaillée en (3.2.2.2).
- ➢ binariser (seuillage) la carte temps-fréquence (L_{ga}). Cette est un prétraitement pour l'étape de classification des motifs détectés.
- ➢ extraire les motifs fréquents (M_{sp}) (dont nous donnerons une définition) par une méthode appelée *extraction des motifs fréquents* décrite en (3.2.2.3).
- ➢ classer les motifs fréquents pour séparer ceux du chauffage (M_{CH}) de ceux associés aux autres usages ;
- ➢ déterminer le domaine temporel de présence du chauffage ($d_1(t)$).

Ces différentes étapes (hormis la première étape) sont détaillées dans les paragraphes qui suivent.

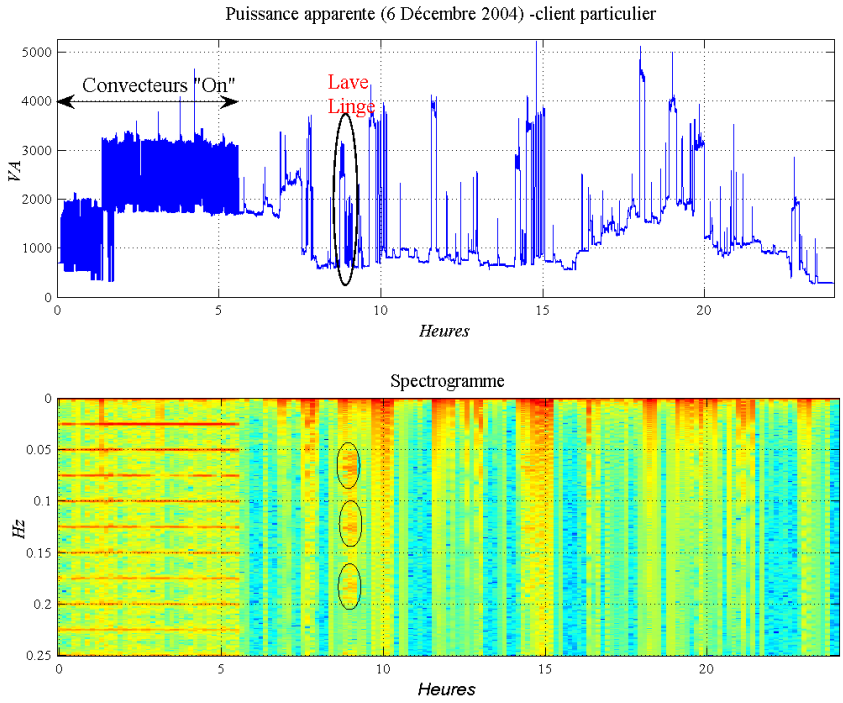


FIGURE 3.2: Exemple de spectrogramme d'une CdC générale ($T_e = 2$s, $N_f = 512$ échantillons)

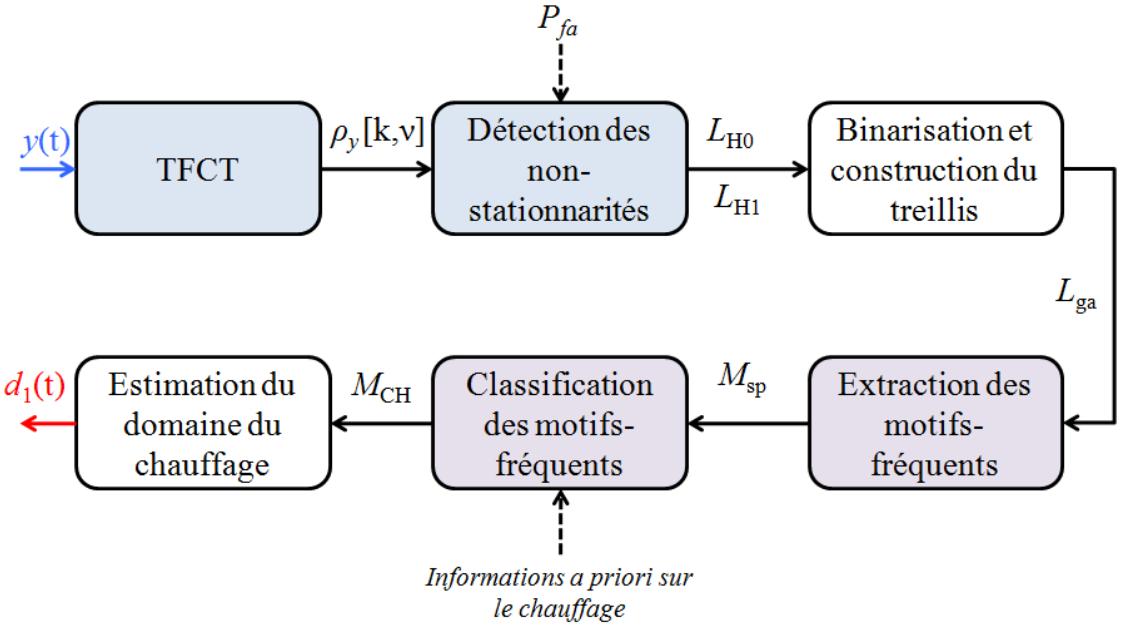


FIGURE 3.3: Schéma général de la détection du chauffage dans le domaine temps-fréquence

3.2.2.2 Segmentation du spectrogramme

Objectif L'objectif est d'obtenir une segmentation du spectrogramme $\rho_y[k,\nu]$ (où k représente le temps et ν la fréquence réduite) en deux régions : L_{H_0} est la région contenant les motifs spectraux « utiles » et L_{H_1} correspondant au bruit de fond de la RTF, ce qui revient à détecter les non-stationnarités temporelles et fréquentielles dans ce domaine. Le détecteur de ces non-stationnarités [MD04] s'appuie sur les variations du moment d'ordre 2 du signal dans le domaine temps-fréquence. Il est défini pour représentation temps-fréquence dont on sait calculer une estimation : un spectrogramme ou un corrélogramme glissant par exemple et en introduisant le minimum d'informations *a priori*.

Dans la suite, on présente le principe général du détecteur ainsi que quelques éléments sur le test d'hypothèse dans le domaine temps-fréquence, les principales étapes de l'algorithme et les résultats obtenus avec cette méthode.

Principe du détecteur L'étape de détetcion consiste à confronter les deux hypothèses suivantes :

H_0 : *Signal aléatoire stationnaire, de puissance moyenne finie, éventuellement augmenté d'un signal déterministe, sa densité de probabilité marginale p_0 est connue mais de paramètres inconnus.*

H_1 : *Signal stationnaire par morceaux de densité de probabilité inconnue.*

La statistique du détecteur est définie à partir d'un seul coefficient de la TFCT, ce qui permet de tenir compte de la localité temps-fréquence et détecter les non-stationnarités.

Ainsi, pour un signal y de M échantillons et pour une probabilité de fausse alarme P_{fa} fixée, le problème de détection se réécrit comme suit :

$$\text{A la fréquence réduite } \nu, \text{ si } \rho_y[k,\nu] \geq \eta_{P_{fa}}[\nu] \text{ alors : l'hypothèse } H_0 \text{ est retenue,}$$

où $\eta_{P_{fa}}[\nu]$ est le seuil de détection fixé pour une fréquence donnée en fonction de la probabilité de fausse alarme choisie. Ce seuil est défini par l'équation suivante (3.4) :

$$Pfa = \int_{\eta_{P_{fa}}[\nu]}^{\infty} p_0(u)\,du \tag{3.4}$$

Afin de déterminer ce seuil, il est nécessaire de définir le modèle statistique du spectrogramme, puis d'inverser l'équation (3.4) pour estimer le seuil.

Densité de probabilité des coefficients du spectrogramme Afin d'expliciter le choix de cette loi de probabilité des coefficients du spectrogramme, nous nous ramenons au test d'hypothèse simplifié suivant :

H_0^* : *Signal aléatoire stationnaire b, gaussien centré de variance σ^2, éventuellement augmenté d'un signal déterministe d.*

H_1^* : *rejet de H_0^*.*

Si de plus la fenêtre d'analyse (utilisée pour construire le spectrogramme) est rectangulaire, d'après le théorème central limite, chaque coefficient du spectrogramme du signal b suit une loi du χ_2 à l degrés de liberté. Par conséquent, chaque coefficient du spectrogramme du signal y suit une loi Gamma $\Gamma(l/2, \sigma^2/2, 0)$.

Si un signal déterministe vient s'additionner au bruit gaussien observé, alors les coefficients de la TFCT suivent des lois du χ_2 décentrés du même degré de liberté l et dont les paramètres de décentrage sont définis par la relation suivante :

$\delta = l\frac{\rho_d[k,\nu]}{\rho_b[k,\nu]}$ où $\rho_d[k,\nu]$ et $\rho_b[k,\nu]$ sont respectivement les cofficients du spectrogramme du signal d et du signal b.

Par conséquent, le spectrogramme peut être modélisé par un mélange de lois de χ_2 centrés et décentrés. Les paramètres des lois décentrés sont foncion du rapport signal à bruit.

Seuil de décision Etant donné la modélisation statistique du spectrogramme, proposée dans [MD04], le seuil de décision est obtenu par inversion de l'équation (3.4) et est déterminé en fonction de la moyenne des observations stationnaires de l'espace $L_{H0}[\nu]$. Pour plus de pécisions sur ce calcul, le lecteur pourra se référer à [MD04, GDL+09].

Algorithme de détection de non-stationnarités sur un spectrogramme Le test de détection dans le plan temps fréquence repose sur la connaissance de la moyenne des éléments de $L_{H0}[\nu]$ que l'on cherche. C'est pourquoi, la construction des deux espaces d'observation L_{H0} et L_{H1} est réalisée itérativement. Les différentes étapes de l'algorithme sont illustrées par la figure (3.4).

Pour chaque fréquence ν, $L_{H0}[\nu]$ est initialisé par les $p\%$ coefficients de plus faible intensité, le paramètre p étant un paramètre fixé par l'utilisateur, 80% dans notre cas. En effet, les amplitudes des coefficients du spectrogramme ont de plus faibles intensités sous H_0 (signal stationnaire) que sous H_1. A chaque itération, le seuil de détection $\eta_{P_{fa}}[\nu]$ est actualisé en fonction de la moyenne de l'ensemble des observations de $L_{H0}[\nu]$ à l'itération précédente et les deux espaces d'observations sont actualisés. Le critère d'arrêt est la stabilité des deux ensembles recherchés. Le résultat de cette procédure est une carte temps-fréquence qui peut être codée comme suit : les éléments de L_{H1} sont représentés par un code couleur, ceux de L_{H0} sont mis à zéro.

En résumé, le détecteur mis en œuvre est récursif et utilise uniquement la connaissance de la famille de p_0. Pour plus de développements sur ce détecteur et son adapatation à la CdC électrique, le lecteur peut se reporter à [GDL+09]. Le résultat obtenu est une carte temps-fréquence, que l'on peut assimiler à une image constituée de plusieurs motifs (les motifs spectraux), dont ceux générés par les signaux de convecteurs. Un exemple de résultat de l'application de ce détecteur est présenté sur la figure (3.5) : on y observe la région de bruit de fond (bleu, 0 dB) et les motifs spectraux utiles d'intensité strictement supérieure à 0 dB.

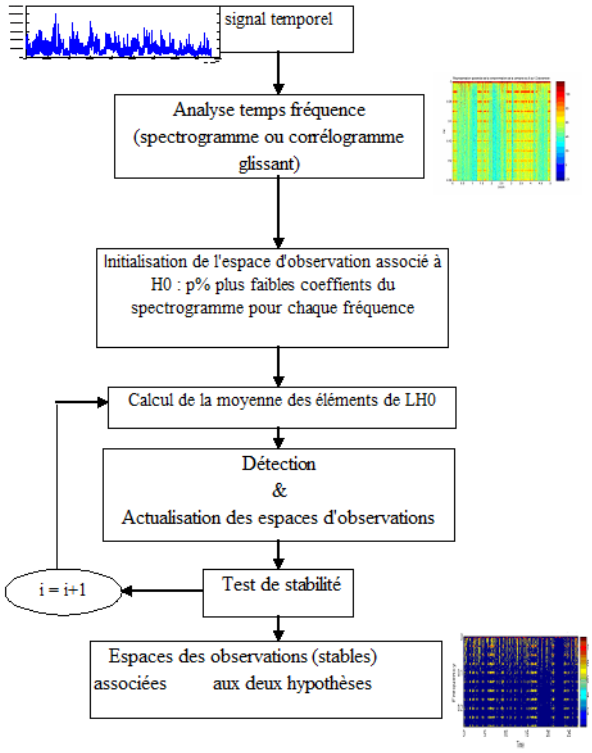


FIGURE 3.4: Schéma général de la détection du chauffage dans le domaine temps-fréquence

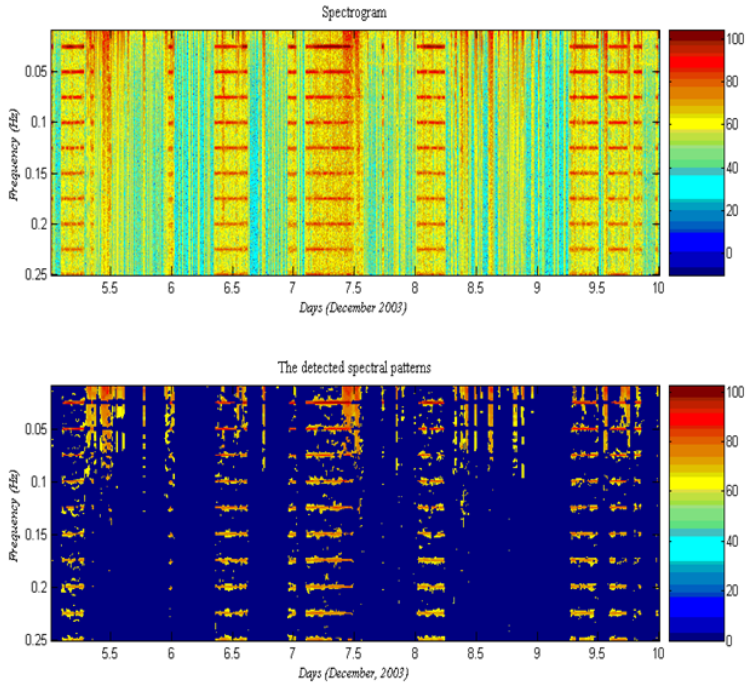


FIGURE 3.5: Segmentation du spectrogramme d'une CdC électrique

3.2.2.3 Extraction des motifs spectraux « pertinents »

Une fois la segmentation du spectrogramme effectuée, nous obtenons la réunion de deux familles de motifs spectraux : ceux générés par les convecteurs et ceux générés par les autres appareils.

Il faut extraire automatiquement les motifs spectraux pertinents pour notre problème, à savoir ceux des convecteurs. L'analyse des images obtenues montre que les signatures des convecteurs se différencient principalement :

➢ par leurs formes géométriques (quasi-rectangulaire) ;

➢ ainsi que par la fréquence d'apparition de chaque motif spectral.

Par ailleurs, nous observons que sur un intervalle de temps où le chauffage est enclenché, les motifs sont quasiment de même taille et sont équidistants. Ces *a priori* sont exploités pour définir une méthode automatique d'extraction de ces motifs. Nous avons utilisé une méthode issue du domaine du Data Mining : l'**extraction de motifs fréquents [CR03].**

La base des données traitées doit être binaire. Cette base est représentée par une matrice dont les lignes sont appelées : **objets** et les colonnes **attributs**. On obtient ainsi des groupes d'objets qui partagent les mêmes attributs, appelés motifs d'objets ou motifs d'attributs.

Remarque : Dans notre cas d'application, les objets (ou transactions) sont les indices temporels, les attributs sont les fréquences.

Définition *Un motif fréquent est un motif pour lequel le nombre d'objets contenant ce motif, appelé support, dépasse un certain seuil appelé fréquence de coupure* **[CR03].**

Les premiers travaux sur l'extraction de motifs fréquents dans une base de données remontent à 1996, date de la conception du premier algorithme APRIORI par Agrawal [Agr96].

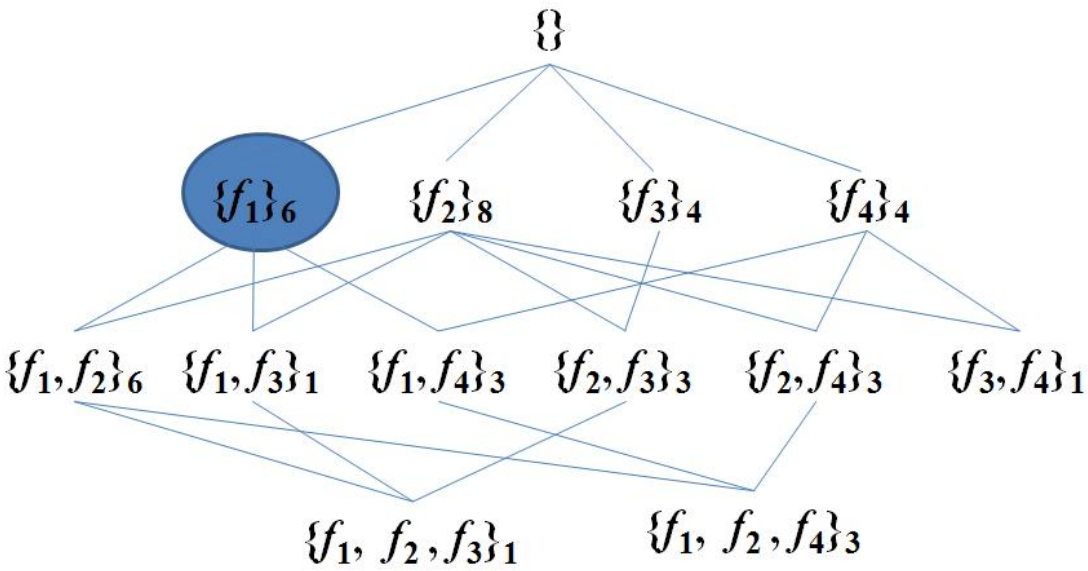
L'extraction des motifs fréquents consiste à chercher, dans une large base de données, les motifs d'attributs rencontrés « fréquemment ». Plus précisément, la recherche des motifs s'effectue sur une représentation condensée des données sous forme d'un treillis (de Galois).

Exemple illustratif Afin d'illustrer le principe de cette méthode, nous considérons un cas simlifié d'une matrice binaire présentée sur la figure (3.6), où les lignes correspondent aux indices des fenêtes d'analyse et les colonnes représentent les fréquences (observées (1) ou non (0)) pour une fenêtre d'analyse donnée.

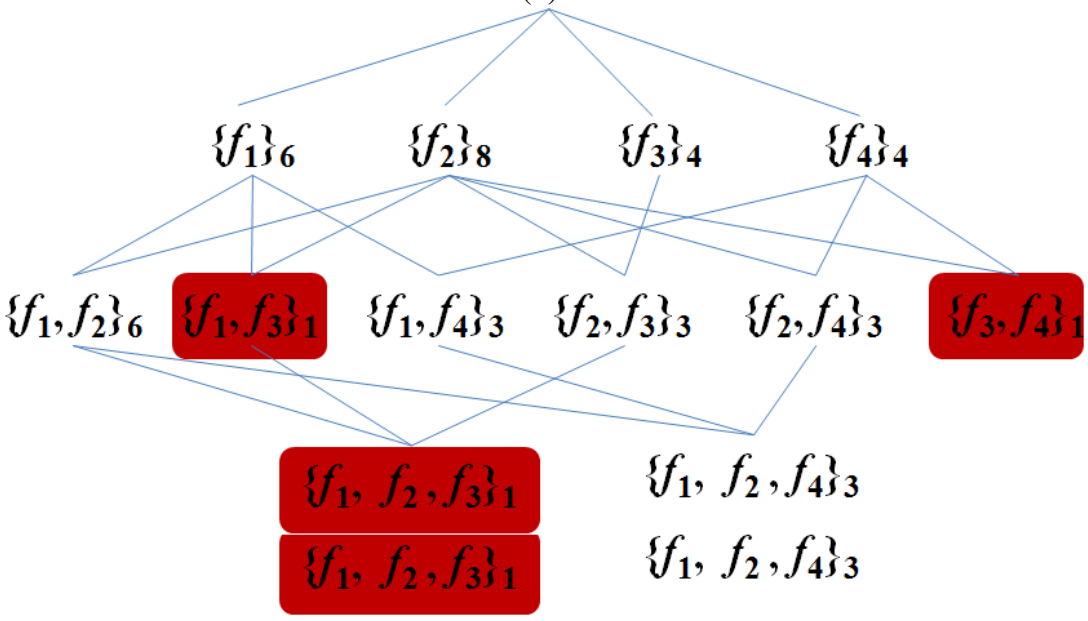
	0	N_f	$2N_f$	$3N_f$	$4N_f$	$5N_f$	$6N_f$	$7N_f$	$8N_f$
f_1=0	1	1	1	0	0	0	1	1	1
f_2=0.01	1	1	1	0	1	1	1	1	1
f_3=0.02	0	0	0	1	1	1	0	0	1
f_4=0.03	1	1	1	1	0	0	0	0	0

N_f=512

(a)



(b)



(c)

FIGURE 3.6: Illustration de l'extraction des motifs fréquents : **(a)** Matrice des données **(b)** Treillis **(c)** Motifs obtenus pour $f_c = 2$

Le treillis associé à cette matrice est présenté sur la figure (3.6). Il est obtenu de la manière suivante :

- ➢ On cherche les fréquences (isolées) présentes. Pour chaque fréquence, on compte le nombre d'apparition sur le domaine temporel d'étude (fenêtres d'analyse). Ainsi, nous constatons que la fréquence f_1 (entourée avec une ellipse bleu sur la figure (3.6) (b)) apparaît 6 fois par exemple sur le domaine temporel d'analyse.
- ➢ Une fois ces fréquences identifiées, on cherche les couples (triplet, quadruplet, etc) de fréquences qui apparaissent au même instant (indice de fenêtre temporelle) et on compte le nombre d'apparition de ces motifs. Dans le cas considéré, les fréquences f_1 et f_2 apparaissent 6 fois à la même fenêtre d'analyse. Il s'agit des apparitions aux indices 0, N_f, $2N_f$, $3N_f$, $6N_f$, $7N_f$ et $8N_f$.

Considérons maintenant une fréquence de coupure $f_c = 2$. L'application de cette fréquence de coupure au treillis implique la suppresion de toutes les associations dont le nombre d'occurence est infénrieur à la fréquence choisie. Les motifs supprimés sont entourés avec des rectangles rouges sur la figure 3.6 (c).

Cette méthodologie est appliqué à la carte temps-fréquence binarisée après détection des non-stationnarités.

3.2.2.4 Identification du support temporel du chauffage électrique

Les motifs fréquents sont extraits avec un logiciel développé au sein de l'université de Caen : *MvMiner*. L'extraction des motifs fréquents dans le cas de la CdC électrique est réalisée pour une fréquence de coupure définie empiriquement en fonction de la durée minimale de fonctionnement d'un convecteur Δt^H_{min} (de l'ordre de 10 mn) de la manière suivante (3.5) :

$$f_c = E\left(\frac{2\,\Delta t^H_{min}}{T_e\,N_f}\right) \tag{3.5}$$

où E désigne l'opérateur partie entière et N_f est la taille de la fenêtre d'analyse utilisée pour la construction du spectrogramme. En pratique, nous avons choisi $N_f = 512$.

Les motifs générés par le chauffage sont inclus dans l'ensemble des motifs fréquents extraits. Ils ont la particularité suivante : ils couvrent la totalité de l'axe fréquentiel : ce sont *les motifs les plus longs*. Un test sur la longueur des motifs fréquents permet de sélectionner ceux du chauffage. Pour obtenir le support temporel du chauffage, il suffit d'associer une fenêtre rectangulaire de taille N_f à chaque colonne de la carte des motifs extraits dès que celle-ci n'est pas partout nulle (c.à.d elle contient des pixels attribués au convecteurs).

Conséquence Par construction le détecteur proposé ne donne pas lieu à de fausses alarmes. En effet, les sources d'alarme tels que les signaux générés par le sèche-linge ou les appareils de cuisson sont des motifs fréquents, mais ne sont pas les plus longs.

La figure (3.7) présente la carte des motifs fréquences binarisée issue de la détection de non-stationnarités (haut de la figure) ainsi qu'une représentation des motifs associés aux convecteurs (bas de la figure). Les deux régions entourées avec des ellipsesvertes mettent en évidence la fragmentation des motifs extraits. Ce phénomène est lié en réalité au fait qu'après détection de non-stationnarités et binaraisation, les motifs ne sont plus compacts. Les motifs entourés par

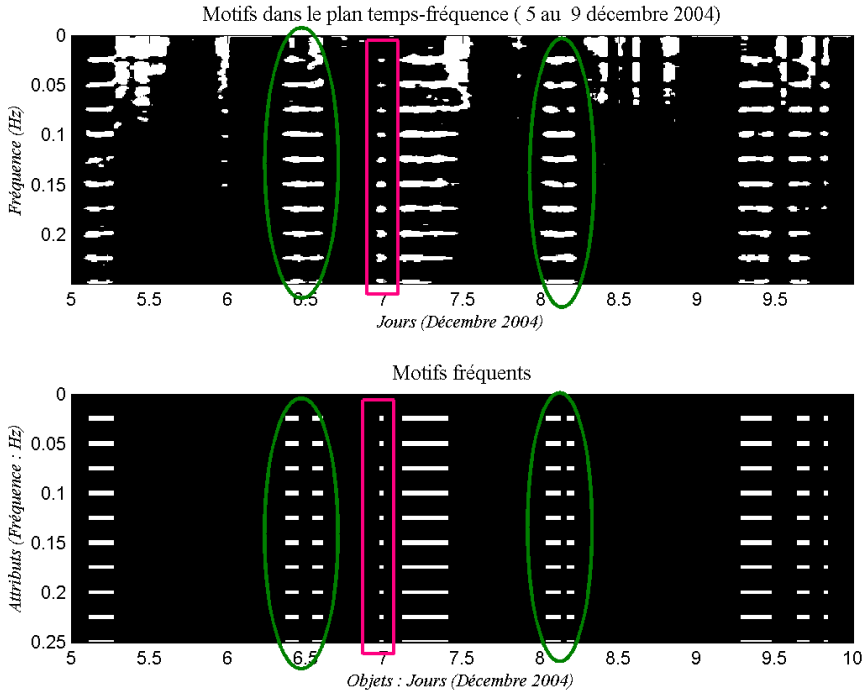


FIGURE 3.7: Extraction des motifs du chauffage électrique : CdC du 5 au 9 Décembre 2004 ($T_e = 2$s)

des rectangles rouges rendent compte de la sous-estimation du domaine associé au chauffage et par conséquence une sous-estimation de la durée quotidienne de fonctionnement de cet usage. Ce risque est d'autant plus important que la forme du motif spectral après binarisation diffère d'un rectangle, notamment à cause de l'apodisation du spectre.

Ces deux phénomènes donnent lieu naturellement à des zones de non-détection du chauffage, et *a priori* à une sous-estimation de l'énergie consommée.

Un critère d'évaluation des performances du détecteur consiste à quantifier la part de la durée de présence de chauffage non-détecté. Si $\hat{d}_1(t)$ est le détecteur de présence du chauffage (grandeur binaire) et $d_1(t)$ est le support temporel de référence du chauffage, on définit un taux de non-détection (3.6) :

$$\beta_{CH}^{glob} = \frac{\int_I \hat{d}_1(\tau)\, d\tau}{\int_I d_1(\tau)\, d\tau} \tag{3.6}$$

où I est le domaine d'analyse de la CdC (une journée).

Le détecteur proposé est testé sur une base de signaux réels d'un seul client (pour des raisons de disponibilité des données). Une étude de performance de ce détecteur sur les nouvelles données de la campagne de mesures est envisagée en perspective de cette étude.

La figure (3.8) présente la CdC de chauffage et le résultat du détecteur défini. Sur cette figure, la courbe en gris représente la CdC du chauffage obtenue par segmentation manuelle de la CdC générale. En effet, nous ne disposions pas du sous-comptage du chauffage. La courbe en rouge représente le détecteur obtenu $\hat{d}_1(t)$ multiplié par un facteur 6.2 kW pour mettre en évidence les performances de la détection. Les zones entourées avec des ellipses noires correspondent à des zones de chauffage non-détectées par l'algorithme.

Le taux de non-détection obtenu sur l'ensemble des CdCs quotidiennes du mois de décembre 2004 est d'environ 10%. Un post-traitement des résultats, notamment un filtrage morphologique pour pallier le problème de fragmentation des motifs, permettrait d'améliorer considérablement ce résultat.

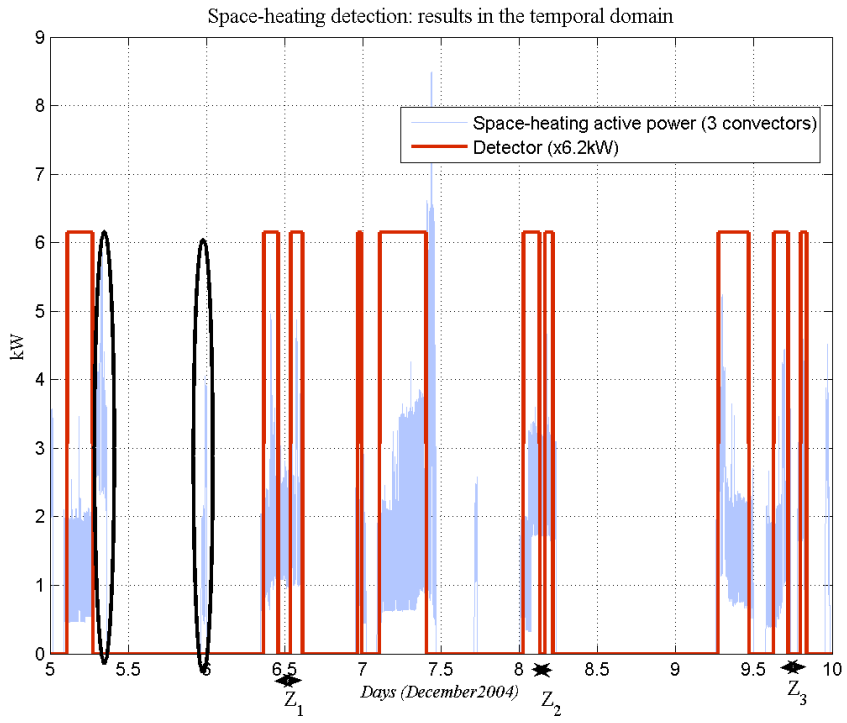


FIGURE 3.8: Détection du chauffage électrique à partir de l'extraction des motifs fréquents les plus longs

Remarquons aussi que ce résultat est global, et sert à évaluer les performances globales du détecteur. D'un point de vue offre de services énergétiques, il n'a de sens qu'en fonction de l'erreur d'estimation de l'énergie consommée qui serait induite par les zones de non-détection.

3.2.2.5 Contributions et discussions

Une nouvelle approche automatique pour interpréter une représentation temps-fréquence, plus précisément le spectrogramme, est proposée. La méthode développée est semi-paramétrique et non-supervisée. Elle repose sur la combinaison de deux méthodes existantes pour réaliser la procédure d'interprétation. Le processus de segmentation est effectué par l'intermédiaire d'un détecteur temps-fréquence fondée sur la modélisation statistique de la TFCT. Le processus de classification fait appel à la méthode d'extraction des motifs fréquents. L'utilisation conjointe de ces deux méthodes est nouvelle et permet de répondre au problème particulier de la segmentation du chauffage à partir de la CdC générale. Notons que ce détecteur ne nécessite aucune connaissance sur les appareils présents chez le client. De plus, la détection du chauffage ne dépend pas du nombre de convecteurs. Cette propriété est en réalité due au nombre limité des périodes de régulation (2 valeurs uniquement).

Les premiers résultats sur des données réelles sont encourageants au regard des critères souhaités par l'entreprise, à savoir une erreur relative maximale de $\pm 15\%$. Néanmoins, une étude de performance du détecteur et son évaluation conjointement avec l'estimation de l'énergie reste encore à réaliser.

Le principal inconvénient de cette méthode et qu'elle ne permet pas de détecter le chauffage quand tous les convecteurs sont en permanence en régime de saturation.

Les perspectives de cette approche portent sur le volet applicatif ainsi que le volet méthodologique. Dans le cas particulier de la CdC, il serait intéressant d'étudier l'applicabilité de cette approche à la caractérisation d'autres usages électriques.

En effet, il serait intéressant d'étudier de manière plus générique l'apport de l'utilisation conjointe des deux méthodes de segmentation et de classification pour l'interprétation d'une représentation temps-fréquence des signaux non-stationnaires.

3.2.3 Détection du chauffage par seuillage de la densité des événements

3.2.3.1 Connaissances *a priori*

Nous avons vu que la densité des événements constitue une transformée adaptée à la reconnaissance du chauffage à partir de la CdC générale. Cette propriété est liée à la régulation rapide des convecteurs. Il est alors plus pertinent de considérer la densité des événements de la composante dite « régulations rapides » $y_r(t)$ définies par la durée caractéristique T_r (cette durée est de l'ordre d'une minute). Ce signal est la superposition de deux signaux : les régulations rapides des convecteurs $y_r^{CH}(t)$ et les régulations rapides d'autres usages $y_r^{\overline{CH}}(t)$ (le sèche-linge, le lave-linge et certains appareils de cuisson). La composante régulations rapide s'écrit ainsi (3.7) :

$$y_r(t) = y_r^{CH}(t) + y_r^{\overline{CH}}(t) \tag{3.7}$$

L'analyse des signaux (chapitre 2) montre que les régulations des usages autre que le chauffage électrique vérifient l'une des deux propriétés suivantes :

➢ *L'amplitude des événements de la composante régulations (de ces usages) est plus faible que l'amplitude minimale d'un événement généré par un convecteurs.*

➢ *La durée de la phase de régulation rapide (de ces usages) est inférieure à la durée minimale de fonctionnement d'un convecteur.*

Nous exploitons ces connaissances apprises de l'analyse des données pour définir une méthode de segmentation du chauffage, que nous décrivons dans la sous-section suivante.

3.2.3.2 Approche proposée

L'approche proposée repose sur un test d'hypothèses sur une grandeur pertinente pour la détection du chauffage. Il s'agit de la densité des événements de la composante régulations rapides de la CdC analysée.

Le schéma général de la segmentation est illustré sur la figure (3.9). La méthode consiste en trois étapes :

➢ l'extraction des régulations rapides du signal observé ;
➢ l'évaluation de la densité des événements définissant ainsi une fonction de contraste pour le test d'hypothèses ;
➢ l'application d'une règle de décision (par seuillage).

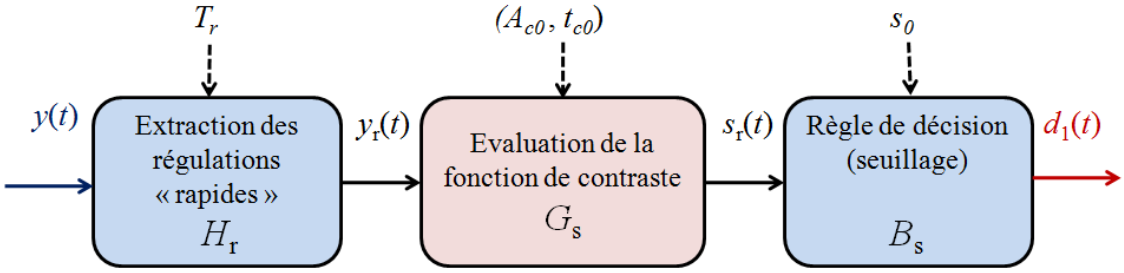


FIGURE 3.9: Schéma général de l'approche événementielle pour la détection du chauffage électrique

Ces trois modules de la procédure ainsi que le choix des paramètres sont décrits ci-après.

Extraction des régulations L'objectif est d'extraire le signal $y_r(t)$ sachant que $y(t) = y_r(t) + y_{au}(t)$ où $y_{au}(t)$ est la composante «autre que les régulations rapides» et $y_r(t) = y_r^{CH}(t) + y_r^{\overline{CH}}(t)$. Une solution fondée sur un filtre passe-haut n'est pas appropriée car elle donne lieu à une déformation des créneaux de la composantes des régulations rapides. Or, l'évaluation de la densité des événements nécessite de garder cette sutructure des créneaux du mélange.

Notre objectif est de construire un filtre h_r paramétré par une fenêtre rectangulaire (élément structurant) de largeur T_r vérifiant les trois conditions suivantes :

➢ les ruptures liées aux *on* et *off* sont conservées ;
➢ la durée T_r est choisie de manière empirique dans l'intervalle $[T_{01}, T_{02}]$ où T_{01} (valant 40 s) et T_{02} (valant 80 s) sont les périodes des caractéristiques des convecteurs ;
➢ les régulations de durée inférieure ou égale à la durée caractéristique T_r sont préservées.

La morphologie mathématique offre la possibilité de construire cette classe de filtre. Le filtre choisi, dénommé filtre Chapeau Haut de Forme, est le résidu entre l'identité et une ouverture dont l'élément structurant est fenêtre rectangulaire de largeur T_r, cette dernière consistant à évaluer les minima du signal à filtrer sur une fenêtre glissante de largeur T_r.

Pour une définition du filtrage morphologique et une illustration de l'ouverture, fermeture ou une combinaison des deux dans un ordre donné, le lecteur peut se reporter à l'annexe F et aux ouvrages de George Matheron et Jean Serra (*Morphologie Mathématiques et granulométries en place*, eisti.fr).

Densité des événements des régulations rapides La densité d'événements de la composante « régulations » $y_r(t)$ est calculée pour une valeur donnée t_{c0}, ce paramètre étant la durée minimale de fonctionnement d'un convecteur dont on souhaite garantir la détection. En pratique, nous choisissons $t_{c0} = 10$ mn.

Soit $\widetilde{y}_r(t) = y_r(t+T_e) - y_r(t)$ la série des accroissements de $y_r(t)$ (gradient de la composante régulations). Avant d'évaluer la densité des événements, nous filtrons les événements d'amplitude inférieure à l'amplitude minimale choisie A_{c0}, appelée amplitude critique. En pratique, A_{c0} vaut 350 W, valeur définie en fonction de l'amplitude minimale d'un signal de convecteur à savoir 450 W. Ainsi, les événements générés par les compresseurs et les appareils de veille par exemple sont filtrés.

La série des événements significatifs $\widetilde{\delta}_r(t, A_{c0})$ par rapport à l'amplitude critique A_{c0} est alors définie en (3.8) :

$$\widetilde{\delta}_r(t, A_{c0}) = \begin{cases} 1 \text{ si } \mid \widetilde{y}_r(t) \mid \geq A_{c0} \\ 0 \text{ sinon} \end{cases} \tag{3.8}$$

La densité des événements des régulations rapides d'amplitude significative est donnée par l'équation (3.9) :

$$s_r(t) = \mid \widetilde{\delta}_r(t) \mid \star g_s(t, t_{c0}) \tag{3.9}$$

Cette fonction est un compteur des événements *significatifs* sur une fenêtre de largeur t_{c0}.

Remarque

Nous avons vu que le paramètre t_{c0} est défini empiriquement par la durée minimale de fonctionnement d'un convecteur que l'on souhaite détecter. Une autre manière de définir ce paramètre consiste à fixer le nombre minimum de périodes N_{c0} nécessaire pour détecter un convecteur isolé en régime de régulation. La durée t_{c0} est alors approchée par la relation suivante : $t_{c0} = N_{c0} \times T_r$.

Règle de décision La règle de décision permettant de détecter le chauffage est en effet une règle de seuillage, qui permet de séparer Ω_s en deux régions de recouvrement non vide, correspondant respectivement à l'usage chauffage et aux usages autre que le chauffage.

La détection du chauffage est alors formulée ainsi (3.10) :

$$\begin{cases} H_1 : s_r(t) \geq s_0 \text{ alors le chauffage est en marche} \\ H_0 : \text{sinon le chauffage est éteint} \end{cases} \tag{3.10}$$

où s_0 est un seuil de détection défini qualitativement en se basant sur les connaissances apprises à partir de l'analyse des signaux.

Choix du seuil de détection On propose de choisir un seuil au plus égal au nombre d'événements (*on-off*) généré par un convecteur de période de régulation T_r fonctionnant sur une durée t_{c0} et au moins égal à la moitié de cette valeur. Une expression de ce seuil est explicitée par l'équation suivante :

$$s_0 = \alpha_0 \frac{t_{c0}}{T_r} \text{ où } \alpha_0 \in \left[\frac{1}{2}, 1\right] \tag{3.11}$$

où α_0 est un paramètre à ajuster. La valeur $\alpha_0 = 1$ correspond à appliquer un seuil égal à N_{c0}, ce qui ne permet pas de détecter des convecteurs en marche si leur durée de fonctionnement est inférieur à la durée limite choisie.

3.2.3.3 Illustrations

La figure (3.10) illustre l'extraction des régulations rapides à partir d'un exemple de CdC semi-synthétique [1]. Le signal étudié est un mélange des CdCs élémentaires des appareils suivants : 2 convecteurs fonctionnant sur la plage temporelle $[t_1, t_1'] \cup [t_2, t_2']$, un lave-linge, un sèche-linge, un réfrigérateur et un chauffe-eau. Idéalement, le signal $y_r(t)$ est quasiment nul en dehors du domaine de fonctionnement du chauffage. Or, on constate que ce signal peut prendre des valeurs significatives (de l'ordre de 2000 W) vers 12h. En effet, ces régulations son celles générées par le sèche-linge et le lave-linge (zoom sur la figure (3.11)).

La densité des événements significatifs ainsi que le seuil de décision sont présentés sur la figure (3.11). On constate que le détecteur mis en œuvre permet d'identifier le support temporel de l'usage chauffage avec une fausse alarme liée au fonctionnement du sèche-linge (ellispe rouge sur la figure 3.11).

1. Le mélange est simulé, les signaux du mélange sont réels.

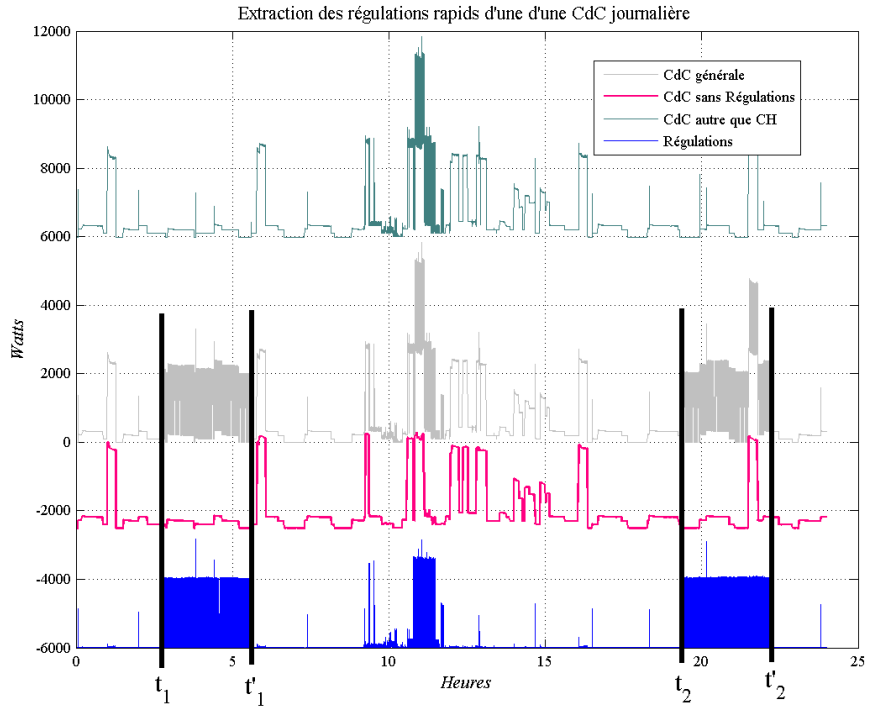


FIGURE 3.10: Extraction des régulations rapides à partir d'une CdC générale ($A_{c0} = 350$ W, $T_e = 1$ s, $t_{c0} = 1400$ s et $T_r = 80$ s)

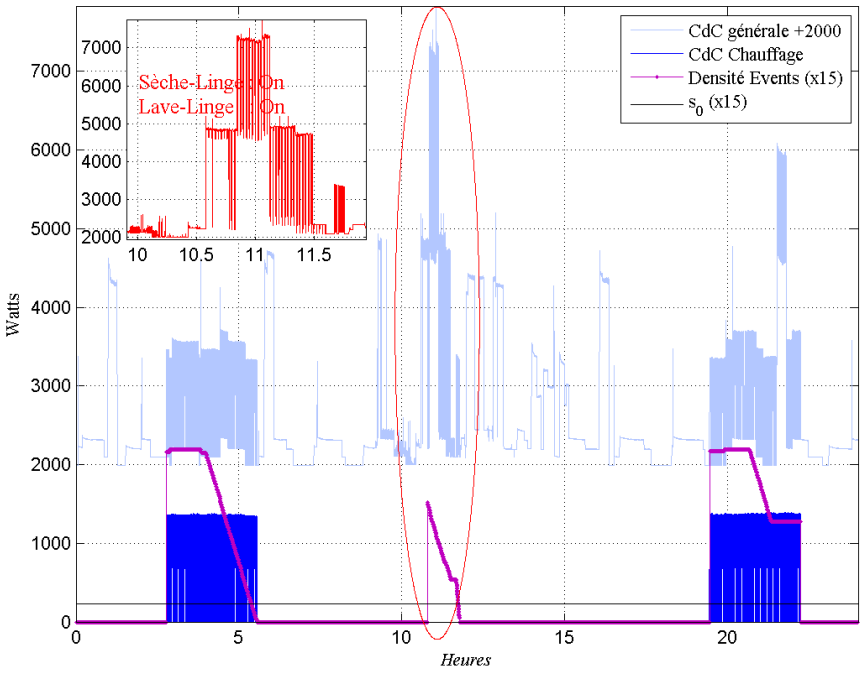


FIGURE 3.11: Illustration de la détection par seuillage de la densité des événements ($s_0 = 35$)

3.2.3.4 Amélioration du détecteur du chauffage

Les conséquences d'une fausse alarme, au sens de la détection du chauffage mais aussi de l'estimation de l'énergie consommée par cet usage, sont conséquents en cas d'absence de chauffage (pas de chauffage électrique ou hors période de chauffe) ou bien pour les clients ayant très peu de convecteurs. L'objectif est de réduire l'impact de cette source de fausse alarme.

Les signaux de convecteurs et du sèche-linge ont les mêmes propriétés fréquentielles. L'amplitude des signaux ne permet pas non plus de les discriminer. Mais, l'analyse des signaux de convecteurs a mis en évidence la propriété de symétrie des événements d'un convecteur, ce qui n'est pas le cas pour le sèche-linge. Rappelons que la CdC d'un sèche-linge est un foisonnement de la puissance de la charge résistive (source de régulations quasi-périodiques) et d'un moteur (faible amplitude, signaux piqués).

Afin de tirer profit de cette différence entre les deux classes de signaux, en vue de la détection du chauffage, nous introduisons un nouvel espace de représentation des événements issus d'une CdC donnée dénommé *l'espace* (P^+, P^-) : il s'agit de représenter les événements du type *off* (accroissements négatifs) en fonction des événements du type *on* (accroissements positifs). La série des événements est en effet le gradient de la CdC observée.

Comme illustré sur la figure (3.12), la signature d'un convecteur dans ce domaine est un nuage de points réparti autour de la première bissectrice. Une procédure de détection du sèche-linge et d'estimation de son support temporel a été mise en œuvre, et exploite les deux *a priori* suivants :

- ➢ asymétrie de la signature du sèche-linge dans le domaine (P^+, P^-) ;
- ➢ le minimum de la puissance nominale d'un sèche linge (environ 2000 W).

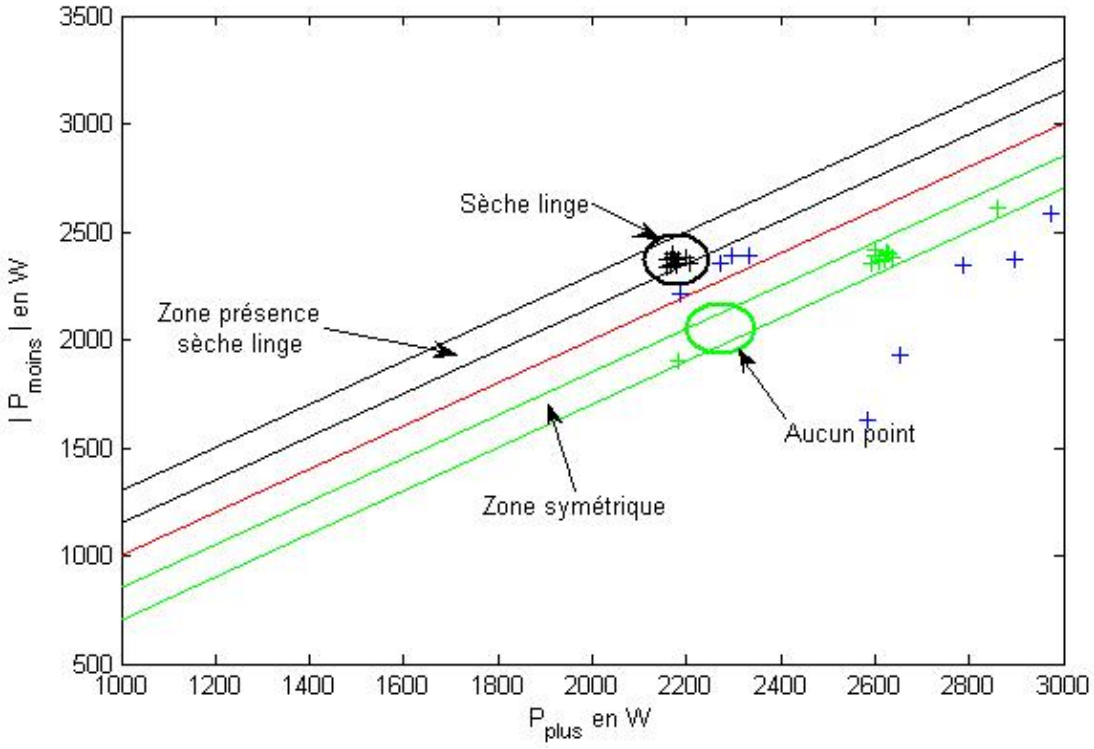


FIGURE 3.12: Signatures de convecteurs et d'un sèche-linge dans le domaine (P^+, P^-)

3.2.3.5 Contributions et discussions

Le chauffage par convecteurs électriques est le résultat du foisonnement d'un nombre inconnu de CdC élémentaires de convecteurs. Les caractéristiques des convecteurs (amplitude, période de régulation, évolution du rapport cyclique) sont inconnues. Afin de détecter la présence de cet usage, à partir d'une CdC générale quelconque, et d'identifier le support temporel de cet usage, nous proposons de détecter cet usage à partir de la fonction de densité des événements. Ce détecteur s'appuie sur la propriété fondamentale suivante : un convecteur a des régulations rapides définies par une durée de régulation critique et une amplitude minimale des régulations. La procédure de détection proposée présente les avantages suivants :

- ➢ elle ne nécessite aucune connaissance sur l'installation étudiée ;
  - – si le chauffage est en marche, il suffit qu'un des convecteurs présents ne soit pas en régime de saturation pour que la détection de l'usage soit garantie ;
  - – la méthode de détection proposée est paramétrée par rapport à une connaissance a priori de l'ordre de grandeur des régulations des convecteurs. Par conséquent, tant que le type d'appareil de chauffage présente des régulations et qu'on a un *a priori* sur la période maximale et sur la période minimale de régulation, l'algorithme développé pour des convecteurs peut être généralisé à d'autres systèmes de chauffage.
  - – par construction, le détecteur défini est robuste aux sources potentielles de fausses alarmes suivantes : les appareils de cuisson, les régulations du lave-linge ou du lave-vaisselle.

Par construction le détecteur de chauffage proposé donne lieu à une non-détection systématique si tous les convecteurs en fonctionnement sont soit en régime de saturation soit sur des durées très courtes par rapport à la durée minimale de détection choisie. Une fausse alarme systématique apparaît en présence de régulations du sèche-linge fonctionnant en absence de chauffage.

Une amélioration du détecteur à travers la détection du sèche-linge a également été proposée et sera utile lors de l'estimation de l'énergie quotidienne consommée.

3.3 Segmentation « ECS / autre »

L'analyse des signaux a permis de mettre en évidence les caractéristiques suivantes de la CdC de l'eau chaude sanitaire (pour une fréquence d'échantillonnage de l'ordre du Hertz) :

- ➢ Cet appareil a un fonctionnement en « tout ou rien » ;
  - – Cette propriété sera résumée par un modèle dit de fonctionnement « tout ou rien » Λ_{TR}
- ➢ Asymétrie de l'amplitude d'un *on* et d'un *off* du même cycle de fonctionnement.
  - – Cette propriété sera résumée par le modèle d'asymétrie des amplitudes noté Λ_{AA}. Au regard des données analysées, ce modèle se traduit par une loi normale dont les caractéristiques sont celles estimées empiriquement à partir des données.

Nous avons par ailleurs mis en évidence les propriétés temporelles de cet usage en étudiant les durées Δt_{ij} avec $i,\ j \in \{0,1\}$, que nous résumerons par le modèle global des durées $\Lambda_{Durées}$.

3.3.1 Principe

L'objectif est de détecter les événements de la CdC vérifiant les propriétés précédentes quelle que soit la CdC étudiée. Cet objectif peut être atteint en deux étapes :

➢ la première étape consiste à détecter des événements d'amplitude significative ;
➢ la seconde consiste à sélectionner uniquement les événements vérifiant les caractéristiques intrinsèques à l'ECS.

Or le chauffage électrique génère des événements de forte amplitude également. Une démarche séquentielle telle que proposée donnerait lieu à plusieurs événements détectés parmi lesquels il faudrait extraire ceux de l'ECS. Afin d'éviter une grande combinatoire et de réduire le coût calculatoire du détecteur, nous proposons d'appliquer la procédure de détection à la CdC ne contenant pas les régulations rapides générées par chauffage.

Le principe du détecteur proposé est résumé sur la figure (3.13), où :

➢ γ_0 et h_0 sont des paramètres nécessaires à l'étape de détection des événements à proprement parler ;
➢ $d_2(t)$ est le détecteur de l'ECS (vaut 1 quand un *on* est détecté, -1 quand un *off* est détecté et 0 ailleurs).

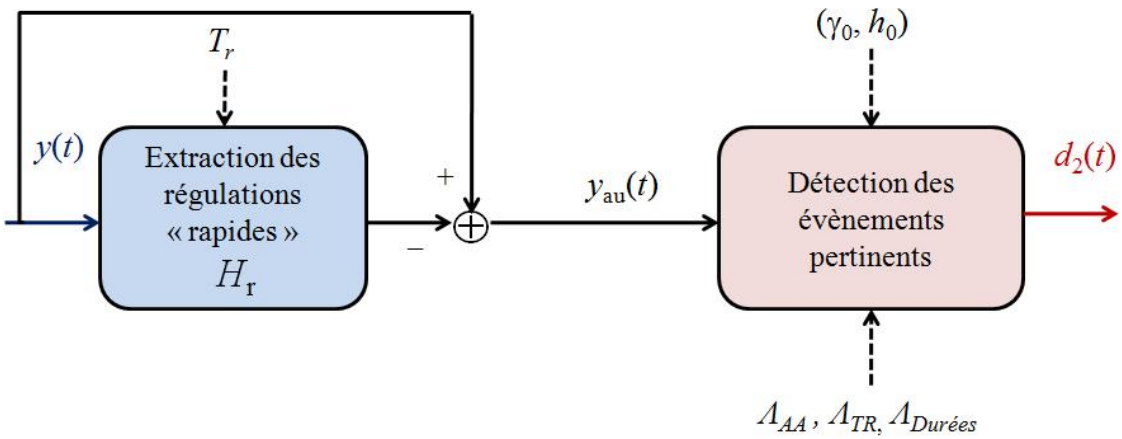


FIGURE 3.13: Schéma général de détection de l'ECS à partir d'une CdC générale ($T_e \sim 1$s)

La détection des événements associés à l'usage ECS est fondée sur l'utilisation d'un test de Page-Hinkley. Seuls les événements qui vérifient les propriétés spécifiques à l'ECS sont retenus. Ce module est détaillé dans le paragraphe suivant.

Illustration de l'extraction des « régulations »

Les figures (3.14, 3.15, 3.16) présentent trois exemple de CdCs générales, les signaux filtrés $y_{au}(t)$ et les puissances appelées par l'ECS mesurées dans chaque cas. On constate qu'en absence de chauffage électrique ou dans le cas où très peu de convecteurs sont en marche sur une durée limitée, le filtrage met en évidence la CdC de l'ECS. Le cas de la CdC où plusieurs convecteurs sont en marche montre que la décomposition du signal filtré est plus simple que celle du signal brut. Cependant, certaines sources de fausses alarmes persistent, par exemple quand des appareils électroménagers sont présents ou encore en présence d'une composante basse fréquence générée par le foisonnement de plusieurs convecteurs (zone entourée avec une ellipse noire). D'où l'utilité d'introduire les connaissances *a priori* sur les lois des durées et de l'amplitude (asymétrie entre l'enclechement et le déclenchement) du chauffe-eau afin de sélectionner les événements pertinents.

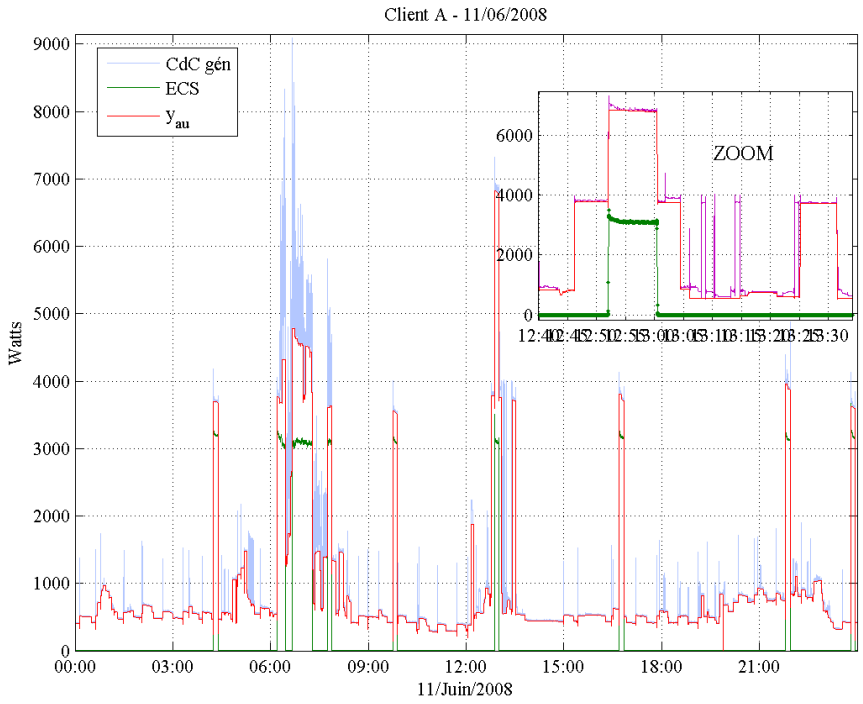


FIGURE 3.14: Illustration du filtrage de la CdC générale en absence de chauffage électrique

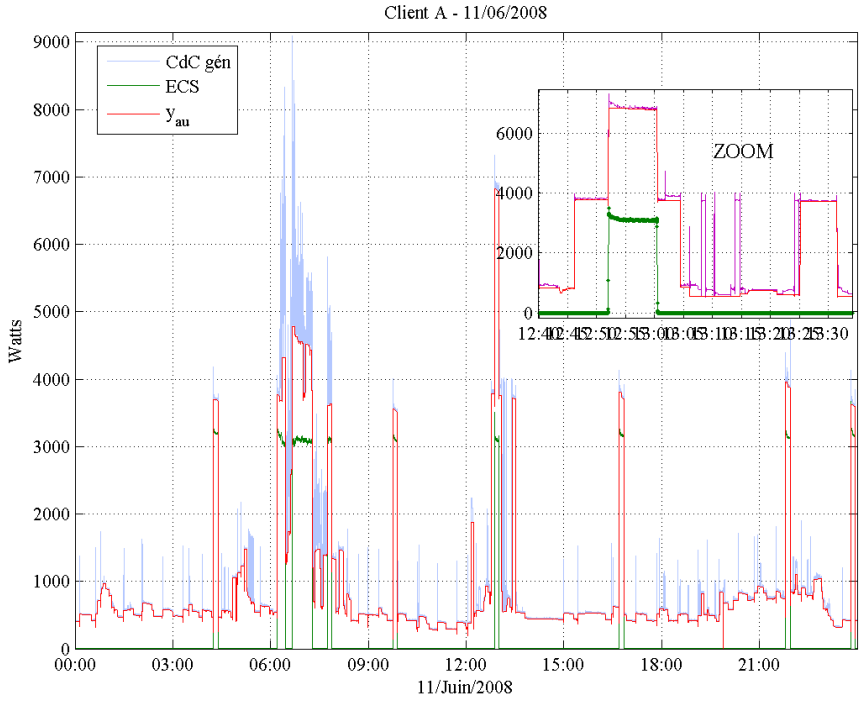


FIGURE 3.15: Illustration du filtrage de la CdC générale en présence chauffage électrique de faible consommation

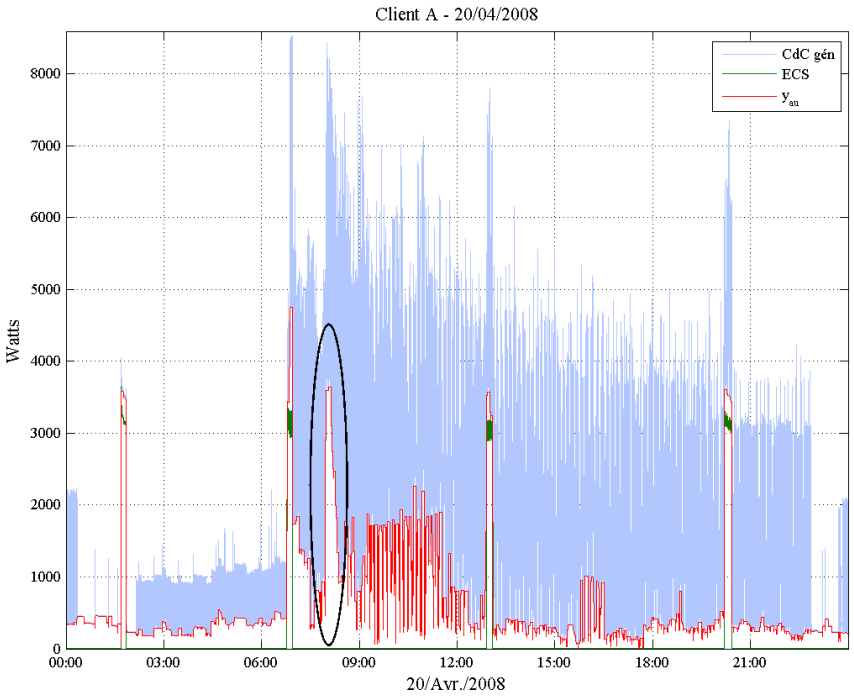


FIGURE 3.16: Illustration du filtrage de la CdC générale en chauffage en marche toute la journée (forte consommation)

3.3.2 Détection des événements

L'objectif est de détecter les événements (*on, off*) générés par l'ECS. Le détecteur proposé consiste à :

- ➢ détecter des sauts de la moyenne (détection des sauts positifs et des sauts négatifs) à partir du signal filtré $y_{au}(t)$;
- ➢ ne retenir que les événements vérifiant les propriétés respectant les modèles Λ_{AA} et $\Lambda_{Durées}$ caractéristiques de l'ECS.

3.3.2.1 Test de Page-Hinkley avec *a priori*

Différents algorithmes de détection de ruptures de la moyenne existent dans la littérature [BN93] et exploitent ou non la connaissance a priori de la variation d'amplitude à détecter, dont le test de Page-Hinkley. Ce test a pour finalité de détecter les instants des ruptures positives et négatives et d'estimer la moyenne avant et après la rupture. Nous l'avons adapté au cas d'application de l'ECS.

Soit A_2^{min} l'amplitude minimale du saut à détecter (de l'ordre de 1000 W en pratique), N la taille du signal observé et μ la moyenne à détecter.

Le détecteur consiste à mener deux tests d'hypothèses en parallèle afin de détecter un accroissement ou une décroissance.

On considère la variable intermédiaire γ définie par la relation suivante :

$$\gamma[k] = \mu[k] + \varepsilon[k] \text{ pour } k \in [1, N]$$

où ε est un bruit blanc gaussien centré de variance σ^2.

Si une rupture a lieu à l'instant k_0, alors $\begin{cases} \mu[k] = \mu_0 \text{ si } k \leq k_0 - 1 \\ \mu[k] = \mu_1 \text{ si } k \geq k_0 \end{cases}$.

La détection d'un saut négatif de la moyenne est obtenue par la procédure suivante :

$$\begin{cases} W_0 = 0 \\ W_n = \sum\limits_{1 \leq k \leq n} \left(\gamma[k] - \mu_0 + \frac{A_2^{min}}{2} \right) \\ M_n = \max\limits_{0 \leq k \leq n} (W_k) \\ \text{Une alarme a lieu dès que } M_n - W_n \geq \lambda \end{cases} \tag{3.12}$$

La détection d'un saut positif de la moyenne est obtenue par la procédure suivante :

$$\begin{cases} U_0 = 0 \\ U_n = \sum\limits_{1 \leq k \leq n} \left(\gamma[k] - \mu_0 + \frac{A_2^{min}}{2} \right) \\ m_n = \min\limits_{0 \leq k \leq n} (U_k) \\ \text{Une alarme a lieu dès que } U_n - m_n \geq \lambda \end{cases} \tag{3.13}$$

Le seuil de détection λ est estimé de manière itérative. Il est initialisé à la valeur $\lambda = 2h_0\sigma / A_2^{min}$. Les deux tests étant menés simultanément, la décision prise correspond au test qui vérifie le critère d'arrêt en premier. Les termes de sommes cumulées sont mis à jour après chaque rupture détectée.

3.3.2.2 Sélection des événements pertinents

Aux critères d'arrêt définis en (3.12) et (3.13), d'autres critères traduisant les contraintes que doit vérifier un *on* ou un *off* sont introduits afin de sélectionner les candidats à la liste d'événements de l'ECS :

➢ Δt_{11}, Δt_{00} et Δt_{10} désignent respectivement la durée entre deux *on* successifs, la durée entre deux *off* successifs et la durée entre un *on* et un *off* successifs.

En pratique, soit l l'étiquette de l'événement détecté, l vaut 0 si l'événement détecté est du type *off* et 1 s'il s'agit d'un *on*.

Mise à part la phase d'initialisation, $E_v(t_i^c, A_i^c)$ est l'événement détecté (avec le test de Page-Hinkley) et $E_v(t_j, A_j)$ est le dernier événement accepté. Prenons le cas particulier où l'événement détecté est tel que l vaut 0.

$\begin{cases} \text{Si } A_j A_i^c \geq 0 \text{ alors le nouvel évènement (rupture de la moyenne) n'est pas retenu.} \\ \text{sinon} \\ \quad \begin{cases} \text{mise à jour des durées en fonction de l'événement détecté} \\ \text{Si les durées suivent le modèle } \Lambda_{Durées} \text{ alors l'évènement détecté est accepté.} \\ \text{sinon, aucune alarme.} \end{cases} \end{cases}$

A l'issue de l'application du test de Page-Hinkley, dont les critères d'arrêt sont modifiés, on obtient une liste d'événements candidats à l'usage ECS. Le détecteur obtenu, dédié à l'ECS, sera appelé détecteur de Page-Hinkley avec contraintes du type ECS.

3.3.3 Détecteur séquentiel de l'ECS

Afin d'extraire les événements générés par l'ECS à partir d'une CdC globale donnée, nous proposons une démarche séquentielle qui s'appuie sur la mise en œuvre des deux étapes décrites dans le paragraphe précédent.

Plus précisément, l'agorithme mis en œuvre est composé des étapes suivantes :

1. Segmenter la CdC
2. Appliquer le détecteur d'événements pertinents (Page-Hinkley avec contraintes) séparément à un ensemble de segments pré-définis de la CdC, et dont l'obtention est détaillée ci-après.

Segmenter la CdC

Les résultats de l'application du détecteur intial proposé ont mis en évidence la grande sensiblité du détecteur à la présence du chauffage (par l'intermédiaire de la composante basse-fréquence) et à la présence d'un gestionnaire d'énergie. Afin d'améliorer la robustesse du détecteur dans de telles configurations, nous proposons de traiter séparément des zones dites « critiques » Z_c de la CdC et des zones « non-critiques » Z_{au}. Les zones critiques sont les zones à forte consommation de chauffage. La segmentation de la CdC est effectuée de la manière suivante :

➢ Déterminer les zones Z_c et Z_{au} par un opérateur de seuillage du signal $y_r(t)$ (composante régulations de la CdC étudiée) sous-échantillonné à la période 10 mn.

➢ Déterminer les intervalles sur lesquels le détecteur sera appliqué : intersection de la liste des zones Z_c et Z_{au} avec la liste des intervalles définis par défaut à savoir $[00\ h, 06\ h]$,$[05\ h, 10\ h]$, $[08\ h, 20\ h]$, $[18\ h, 22\ h]$ et $[20\ h, 00\ h[$ [2].

Appliquer le détecteur sur chaque intervalle d'intérêt

Soit I_m la zone étudiée et $y^m_{au}(t)$ le signal $y_{au}(t)$ tronqué sur cet intervalle.

Le détecteur fondé sur le test de Page-Hinkey avec contraintes est appliqué au signal $y^m_{au}(t)$ sur I_m, ainsi qu'au signal (retourné temporellement) $y^m_{au}(-t)$. Ce choix est justifié par les observations suivantes, qui sont liées au foisonnement des usages :

➢ certaines non détections peuvent être évitées grâce à ce test supplémentaire ;

➢ certaines fausses alarmes sont pénalisées dans un cas et pas dans l'autre.

On obtient ainsi deux détecteurs $d^r_2(t)$ et $d^g_2(t)$ sur chaque intervalle d'étude I_m .

2. Notons que les intervalles définis par défaut sont recouvrants. Ce choix permet de gérer les cycles de fonctionnement ayant lieu sur deux intervalles successifs.

Fusion des événements détectés

Sur chaque intervalle d'intérêt, le détecteur final de l'ECS $d_2(t)$ est obtenu ainsi :
- ➢ mise à *un* sur toute la période où les deux détecteurs élémentaires sont à un ;
- ➢ mise à *un* sur toute la période où un seul des deux détecteurs est à un sans recouvrement avec l'autre détecteur ;
- ➢ mise à *zéro* sur toute la période où un seul des deux détecteurs est à un, mais qui recouvre une autre zone où l'autre détecteur est à un ;
- ➢ mise à *zéro* sur toute la période où les deux détecteurs valent zéro.

Remarque

Nous avons choisi d'intégrer l'étape de sélection des événements dans la définition du détecteur, car l'objectif est d'estimer l'énergie consommée par cet usage sur son domaine de définition.

3.4 Évaluation des algorithmes mis en oeuvre

Nous présentons les critères d'évaluation des algorithmes développés dans les deux sections précédentes et les résultats obtenus à partir de signaux réels. La méthode de segmentation du chauffage à partir de l'interprétation du spectrogramme (développée en 2006) n'a pas été évaluée dans le cadre de la thèse car les premiers résultats obtenus (taux de non-détection moyen de 10% pour environ 3 semaines de données) n'a pas été jugé satisfaisant par l'entreprise. Des améliorations méthodologiques restent encore à effectuer sur cette méthode avant l'étape d'évaluation.

3.4.1 Performances de la segmentation chauffage / autres usages

Nous présentons ici les performances de l'algorithme de segmentation chauffage / autres usages fondée fondé sur un test d'hypothèses sur la densité des événements. Cet algorithme a été testé sur des données semi-synthétiques de la Maison Domotique ainsi que sur une centaine de cas de mélange collectés chez le client A.

3.4.1.1 Taux d'erreur de segmentation

Les critères de performances utilisés sont le taux de non-détection et le taux de fausse alarme. Le taux de non-détection est la durée totale de fonctionnement de chauffage non détecté relativement à la durée totale de fonctionnement du chauffage. Le taux de fausse alarme est la durée totale des zones attribuées au chauffage alors que le chauffage n'est pas en marche relativement à la durée totale de fonctionnement de cet usage.

Cas de CdC générales semi-synthétiques

Les résultats obtenus sur ces CdC sont représentés sur les figures 3.17 et 3.18, où nous présentons respectivement le taux de non-détection et de fausse alarme en fonction de la durée quotidienne effective de fonctionnement du chauffage (rangée par ordre croissant).

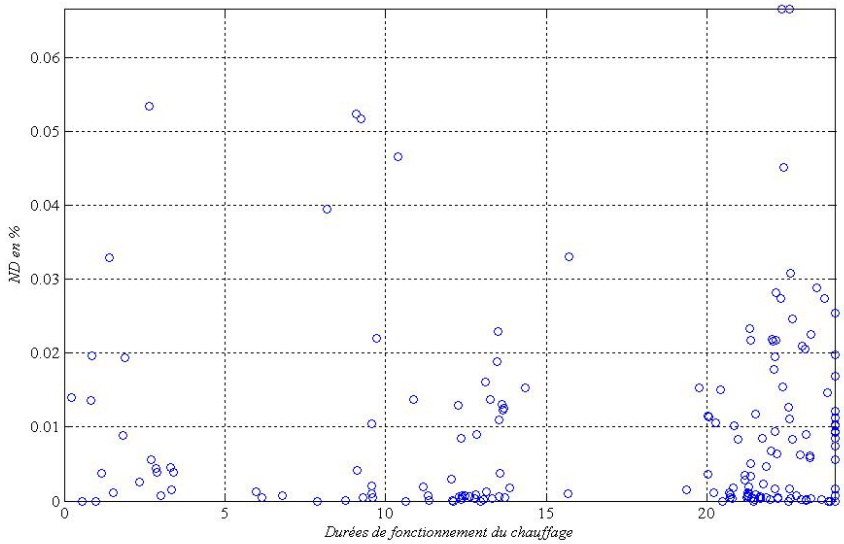


FIGURE 3.17: Taux de non-détection du chauffage (cas de la Maison Domotique)

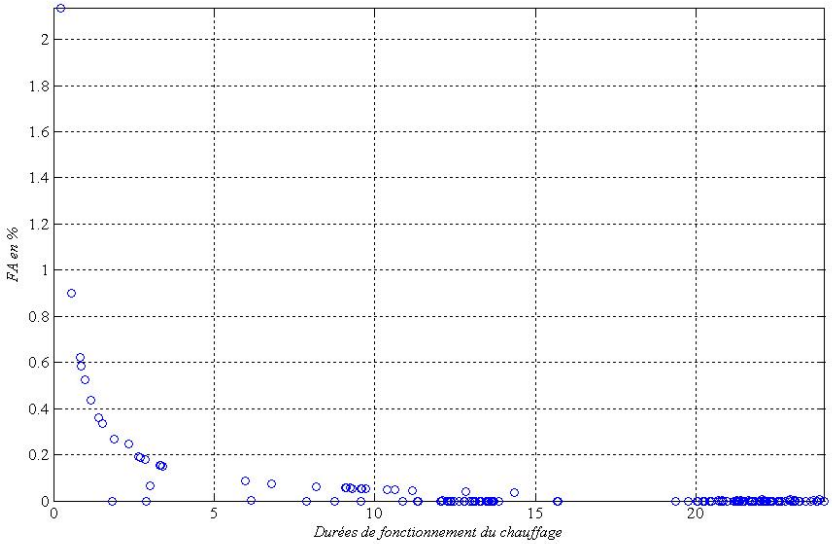


FIGURE 3.18: Taux de fausse-alarme du chauffage (cas de la Maison Domotique)

On constate que le taux de non-détection est quasi-nul, ce qui est prévisible par construction de la fonction d'observation considérée. Notons également qu'aucune corrélation entre le taux de non-détection et la durée effective de fonctionnement des convecteurs n'est identifiée.

Nous précisions que ce résultat est particulier au cas où les convecteurs en marche ne fonctionnent pas simultanément en régime saturé pendant une durée significative. Quant au taux de fausse-alarme, nous constatons qu'il est globalement inférieur à 2%. L'analyse du taux de fausse alarme (3.18) pour les journées testées met en évidence une corrélation entre ces deux grandeurs : le taux de fausse alarme décroît en fonction de la durée de fonctionnement du chauffage. Ce résultat est prévisible. En effet, la principale source de fausse alarme est le sèche-linge. A durée de fonctionnement égale du sèche-linge, le taux de fausse alarme décroît si la durée de fonctionnement du chauffage croît. La décroissance exponentielle est particulière aux cas testés, plus précisément aux profils de consommateurs simulés notamment à la fréquence d'utilisation du sèche-linge.

Cas de CdC générales réelles

Les résultats obtenus sur les données du client A (taux de fausse alarme et taux de non-détection) sont présentées sur la figure (3.19). Toutes les valeurs sont inférieures à 1% à l'exception du taux de non-détection obtenu pour la journée n°24. Après analyse des signaux, il s'est avéré qu'il y avait un problème d'acquisition des mesures lors cette journée. Nous retrouvons les mêmes ordres de grandeur pour les autres données.

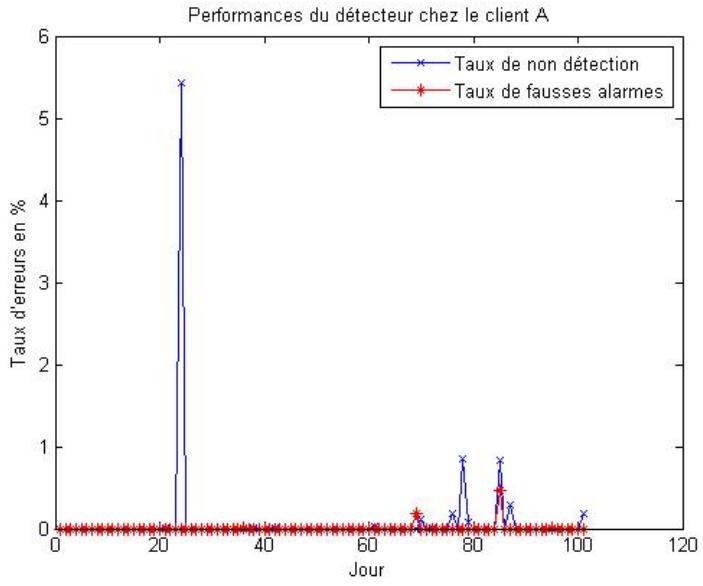


FIGURE 3.19: Taux de fausse alarme et taux de non-détection du détecteur de chauffage

En résumé, les taux d'erreur de segmentation chauffage / autres usages sont globalement inférieurs à 2%. Ces résultats sont très satisfaisants d'autant plus que les résultats de la segmentation servent à l'estimation de l'énergie consommée (chapitre 4).

3.4.2 Performances de la segmentation du chauffe-eau / autres usages

Quatre exemples de CdC résidentielles et les résultats de la détection de l'usage ECS (détecteur multiplié par 3300 W). La figure (3.20) présente la puissance moyenne appelée par l'installation d'un client résidentiel absent. Le signal mesuré est la superposition de deux usages : le réfrigérateur et le chauffe-eau. Le résultat de détection superposé à la puissance moyenne appelée par le chauffe-eau seul illustre les bonnes performances du détecteur dans ce cas. Les figures (3.21) et (3.22) illustrent la CdC en présence du chauffage électrique. Dans les deux cas, les résultats obtenus mettent en évidence deux fausses alarme liés à la présence d'appareils de cuisson respectivement vers 8h et vers midi. Le dernier cas, illustré sur la figure (3.23), met en évidence la robustesse du détecteur en présence d'une forte consommation du chauffage.

En résumé, les résultats de détection obtenus dans ces cas sont très satisfaisants malgré la présence de fausses alarmes :

- ➢ tous les cycles de fonctionnement ont été détectés ;
- ➢ les dates de début et de fin de chaque cycle sont déterminés avec une résolution inférieure à 1 mn.

La figure (3.24) illustre un exemple de non détection et un exemple de fausse alarme. Les deux cas de fausse-détection sont liés au foisonnement des convecteurs qui génère une composante basse-fréquence dans le premier cas, et donne lieu à une forte asymétrie entre l'enclenchement et le déclenchement détecté dans le second cas.

La détection est parfois nettement moins performante. L'exemple présenté sur la figure (3.25) correspond à un taux de non-détection très élevé lié au foisonnement important des usages.

L'erreur de détection du chauffe-eau sera détaillée dans le chapitre suivant (4.3) où l'on présente l'erreur d'estimation de l'énergie quotidienne consommée par le chauffe-eau, cette dernière étant proportionnelle à la durée de fonctionnement de l'usage.

Synthèse

Ce chapitre présente une méthodologie générale pour la segmentation de la courbe de charge générale. Elle est fondée sur une approche « classe / non-classe », où « non-classe » est un raccourci pour désigner les éléments en dehors de la classe recherchée. Cette approche est très utilisée dans le cadre d'une problèmatique similaire, celle de la segmentation voix / musique. Les méthodes et algorithmes développés dans le cas de la segmentation chauffage / non-chauffage et chauffe-eau / non-chauffe-eau sont détaillés. Les résultats de segmentation obtenus à partir de signaux réels sont présentés. Cette première évaluation est satisfaisante et met en évidence l'intérêt de cette approche. En effet, les performances de la segmentation du chauffage sont très satisfaisants (non-détection et fausse alarme). Celles de la segmentation du chauffe-eau le sont moins. L'approche adoptée nous permet d'envisager des améliorations de ce module sans détériorer le premier ainsi que d'introduire d'autres modules de segmentation. Une fois ces améliorations apportées, il faudrait définir plus précisément des critères d'évaluation du système de segmentation et mener l'étude de performance à une base de signaux plus significative.

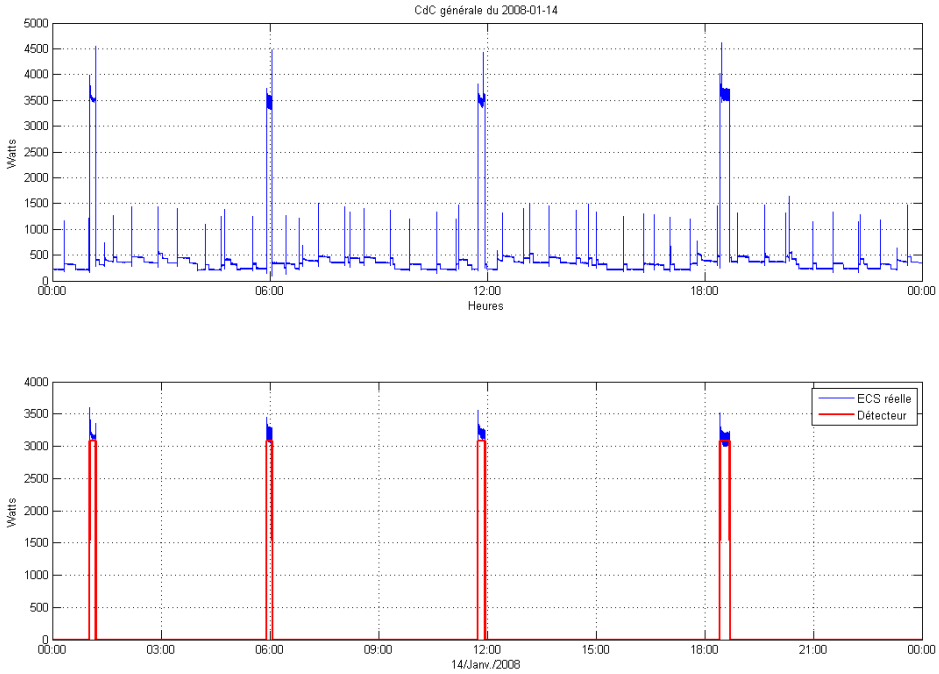


FIGURE 3.20: Détection de l'ECS : CdC en absence du client

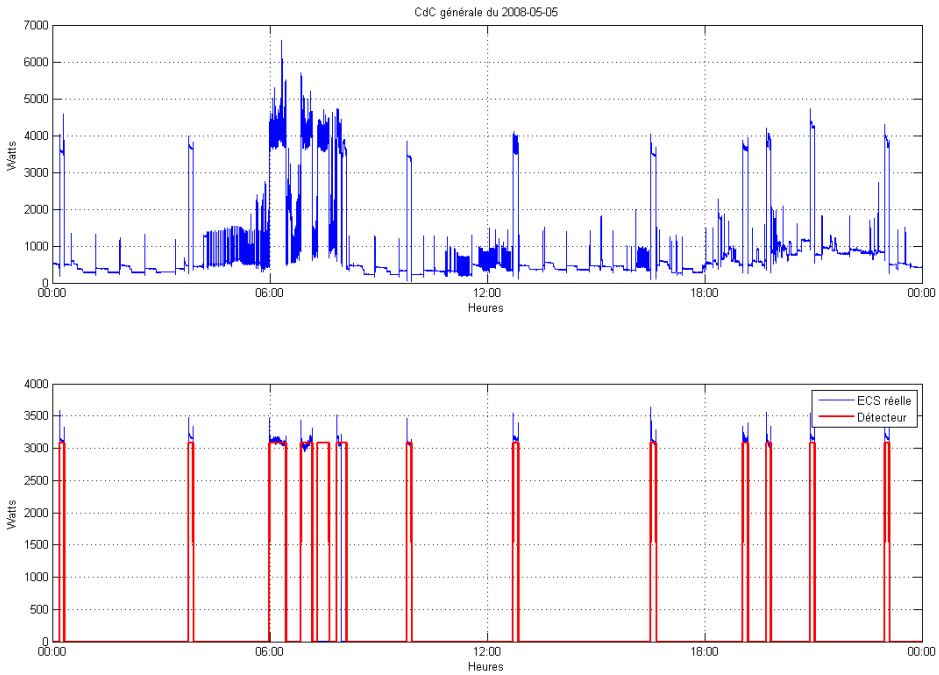


FIGURE 3.21: Détection de l'ECS : CdC avec très peu de chauffage

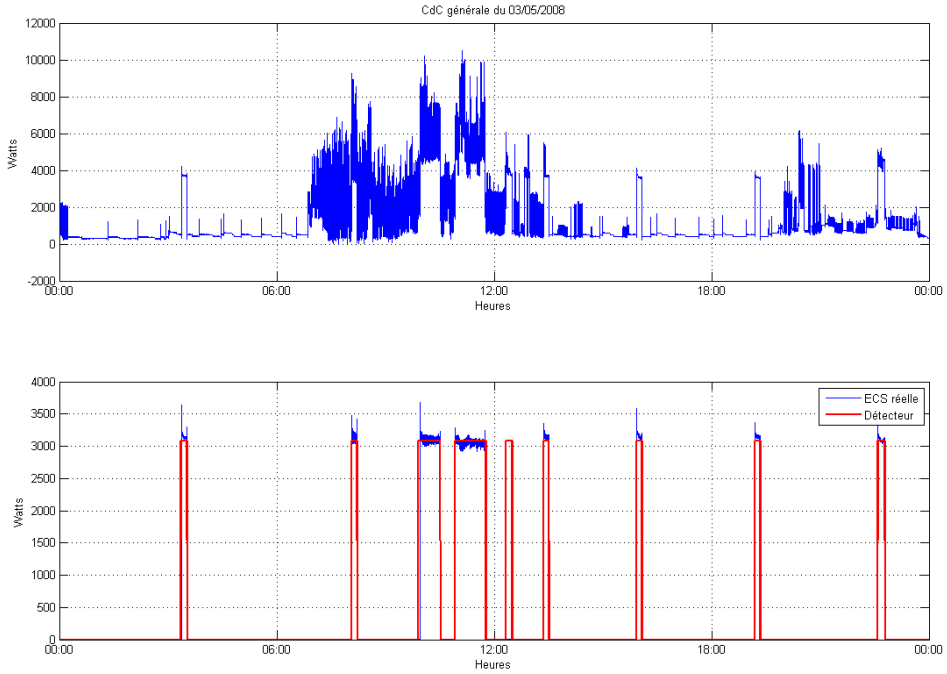


FIGURE 3.22: Détection de l'ECS : CdC avec une consommation moyenne du chauffage

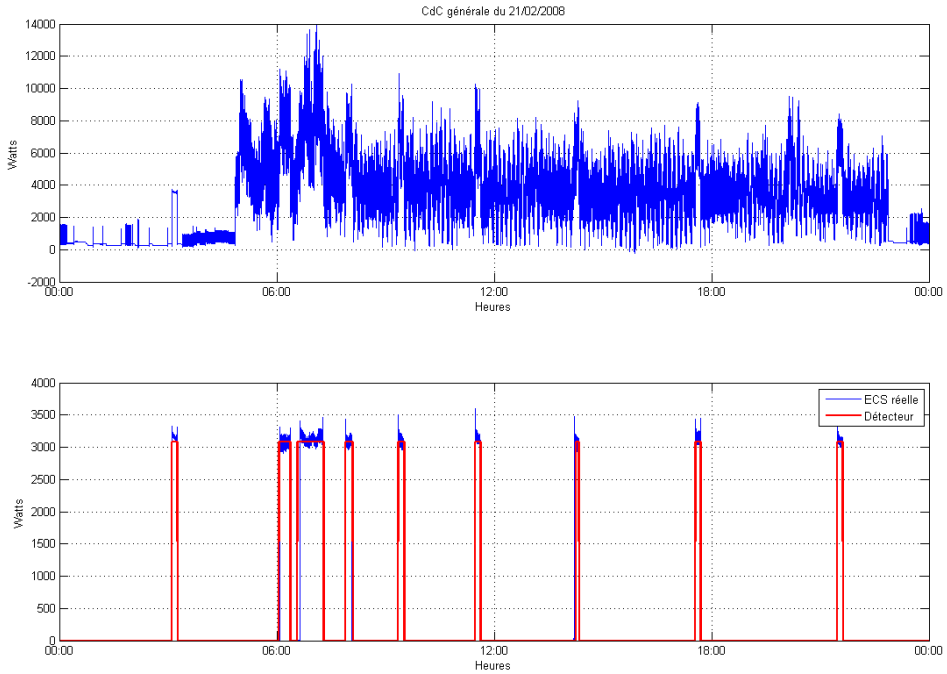


FIGURE 3.23: Détection de l'ECS : CdC avec une forte consommation du chauffage

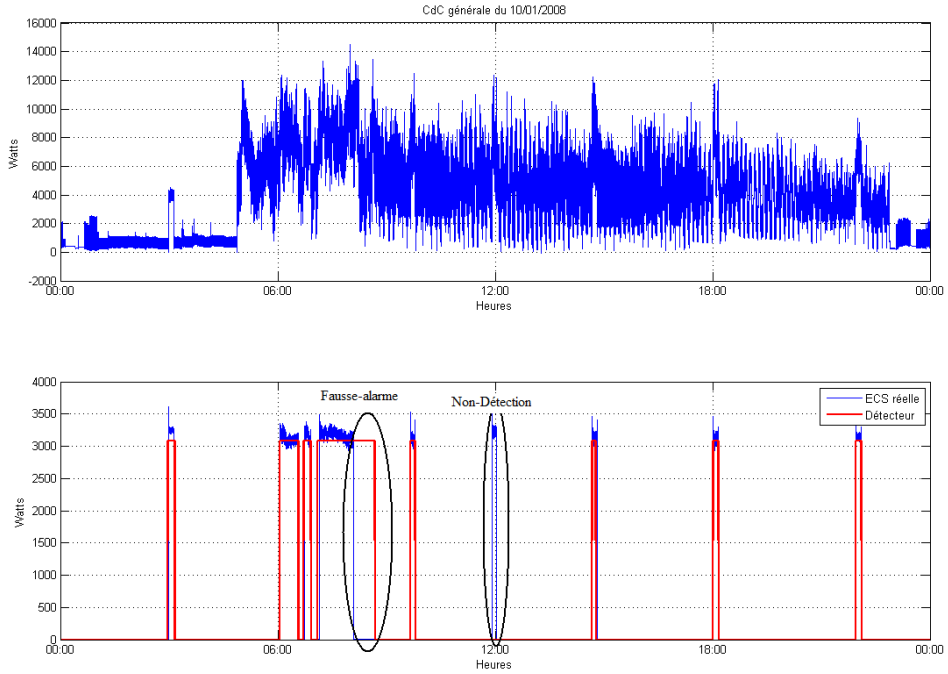


FIGURE 3.24: Illustration de non détection et fausse alarme du détecteur de l'ECS

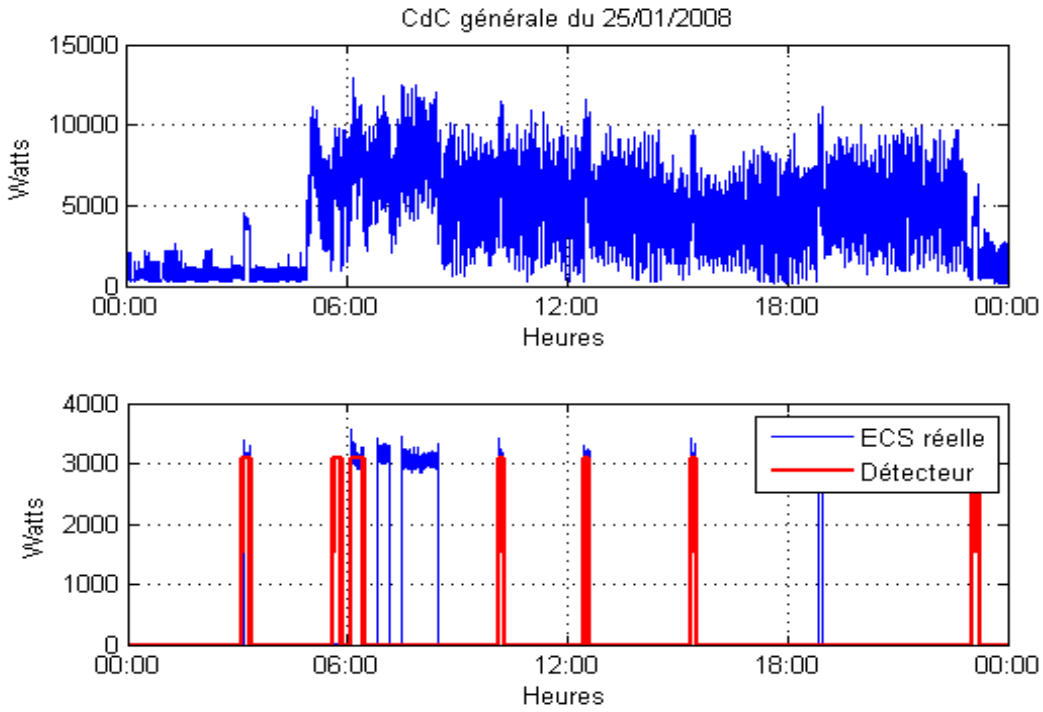


FIGURE 3.25: Détection de l'ECS : exemple de non-détection importante

Chapitre 4

Estimation des caractéristiques d'usages à partir de la courbe de charge

Introduction

Dans ce chapitre, nous présentons différentes méthodes pour estimer les caractéristiques des usages électriques. Dans certains cas, l'estimation des paramètres suppose que l'étape de la détection est préalablement effectuée (exemple : estimation de l'amplitude du signal du chauffe-eau). Dans d'autres cas, l'estimation des paramètres s'effectue dans le cadre de la reconstruction des signaux (exemple : estimation de l'amplitude des compresseurs d'un hypermarché).

4.1 Objectifs

Les paramètres des usages électriques sont utiles pour le développement d'une offre de services énergétiques. L'analyse détaillée de l'énergie totale consommée par usage constitue une offre énergétique à part entière. Connaître les puissances nominales est également une information fondamentale dans le cadre de l'audit énergétique des installations complexes dans le secteur tertiaire.

Nous nous intéressons ici à l'estimation de deux catégories de paramètres des usages dont la définition est fournie ci-après.

Les paramètres globaux Ce sont des paramètres macroscopiques décrivant un usage donné tels que le nombre d'appareils ($N_k \in \mathbb{N}$) de la même classe C_k, l'énergie consommée par usage $e \in \mathbb{R}^+$.

Les paramètres intrinsèques de l'appareil Il s'agit de grandeurs liées soit au modèle électrotechnique de l'appareil (exemple : puissance nominale de fonctionnement) soit au modèle de fonctionnement (exemple : caractéristiques de la régulation). Nous nous intéressons tout particulièrement à l'estimation de l'amplitude moyenne $a \in \mathbb{R}^+$ d'un usage de la classe C_k.

Dans la suite, nous exposons trois méthodes utilisées pour répondre à ce problème d'estimation, qui font intervenir respectivement des techniques de filtrage non-linéaire, un modèle paramétrique déterministe et un modèle stochastique du signal observé.

Remarque L'estimation du nombre de composantes est inhérent à toutes ces problématiques. Nous n'avons pas investigué les méthodes spécifiques à ce problème. Cependant, nous le traitons dans le cadre de l'estimation d'autres paramètres ou encore lors de la reconstruction des signaux du mélange.

4.2 Estimation de l'énergie consommée par le chauffage

Nous présentons ici une méthode d'estimation de l'énergie quotidienne consommée par le chauffage électrique à partir de la CdC générale.

4.2.1 Démarche

Nous avons vu en (3.2.3) que :
- ➢ la régulation des convecteurs est plus rapide que celle des autres appareils ;
- ➢ la CdC générale est la somme d'un signal $y_r(t)$ des « *régulations rapides* » et d'un signal ne contenant pas de régulations $y_{au}(t)$;
- ➢ La composante « régulations rapides » contient le signal associé aux régulations du chauffage $y_r^{CH}(t)$ et celui associé aux régulations des autres usages $y_r^{\overline{CH}}(t)$ (exemple : sèche-linge).

Nous proposons d'estimer l'énergie consommée par le chauffage en la décomposant en deux contributions : celle de $y_r(t)$ estimée par filtrage morphologique décrit en (3.2.3.2) et un terme de correction que nous détaillerons en (4.2.2).

4.2.2 Modèle de l'énergie consommée par le chauffage

Nous proposons un modèle paramétrique de l'énergie quotidienne consommée par le chauffage e_{CH} décrit par l'équation (4.1) :

$$e_{CH} = e_r + \alpha\, e_H + \beta\, e_B \tag{4.1}$$

où :
- ➢ e_r est l'énergie des régulations rapides ;
- ➢ e_H est l'énergie de l'enveloppe haute $x_H(t)$ de la CdC paramétrée par une durée T_H, la défintion de ce paramètre et l'estimation de l'énergie sont détaillés en (4.2.3.1) ;
- ➢ e_B est l'énergie de l'enveloppe basse $x_B(t)$ de la CdC paramétrée par une durée T_B , la défintion de ce paramètre et l'estimation de l'énergie sont détaillés en (4.2.3.1) ;
- ➢ α et β sont des constantes qui dépendent du nombre de convecteurs en marche et du profil du consommateur.

Les coefficients de pondération α et β sont définis par les deux relations suivantes (4.2) :

$$\begin{cases} \alpha = \int_{t_d}^{t_f} (a N_{co}(t) + b)\, \hat{d}_1(t)\, dt \\ \beta = \int_{t_d}^{t_f} (a' N_{co}(t) + b')\, \hat{d}_1(t)\, dt \end{cases} \tag{4.2}$$

où :

- $[t_d, t_f]$ est le domaine d'observation de la CdC $y(t)$;
- $\hat{d}_1(t)$ est le détecteur de chauffage ;
- $N_{co}(t)$ est le nombre de convecteurs en marche simultanément ($N_{co}(t) \in [0, 15]$) estimé dans (4.2.3.2).
- a, a', b et b' sont appelés les **«constantes du profil-consommateur».**

Ainsi, le modèle proposé pour estimer l'énergie consommée par le chauffage dépend des paramètres de six paramètres : les constantes du profil consommateur (estimés à partir de données d'apprentissage en 4.2.3.3) et les durées T_r, T_H et T_B (fixés en s'appuyant sur l'analyse des signaux des usages) nécessaires à l'estimation de la contribution des régulations rapides, de l'enveloppe haute et de l'enveloppe basse de la CdC à l'énergie totale consommée par les convecteurs.

Une justification qualitative de l'introduction des coefficients de pondération α et β est fournie en (4.2.3.1).

Remarque

Le modèle que nous avons proposé est le résultat d'une approche entièrement empirique, qui s'appuyait sur l'analyse de signaux de convecteurs. En effet, le manque de données pendant la première moitié de la thèse rendait la défintion d'un «vrai» modèle de régression de l'énergie du chauffage connaissant la CdC globale difficilement envisageable. Le parti pris était donc de définir un modèle et de l'évaluer et l'améliorer une fois les données étaient disponibles.

4.2.3 Estimation du modèle de l'énergie du chauffage

4.2.3.1 Estimation des énergies e_r, e_H et e_B

Les énergies e_r, e_H et e_B sont évaluées respectivement à partir des signaux $y_r(t)$, $x_H(t)$ et $x_B(t)$ tronqués sur le domaine de présence du chauffage noté $\hat{d}_1(t)$ obtenu par seuillage de la densité des événements significatifs des régulations rapides. Les expressions des différentes énergies intermédiaire sont décrites par le système d'équation (4.3) :

$$\begin{cases} e_r = \int_{t_d}^{t_f} y_r(t)\, \hat{d}_1(t)\, dt \\ e_H = \int_{t_d}^{t_f} x_H(t)\, \hat{d}_1(t)\, dt \\ e_B = \int_{t_d}^{t_f} x_B(t)\, \hat{d}_1(t)\, dt \end{cases} \tag{4.3}$$

Les signaux temporels $y_r(t)$, $x_H(t)$ et $x_B(t)$ sont extraits par filtrage de la CdC générale. Dans la sute, nous fournirons une défintion de chaque composante, les différents filtres mis en oeuvre ainsi que l'ordre de grandeur des paramètres des filtres. L'ajustement de ces paramètres sera exposé ultérieurement à partir de signaux d'apprentissage.

a) Estimation des « régulations rapides » Le signal de régulations rapides $y_r(t)$ dépend le paramètre T_r, le filtre mis en oeuvre est détaillé en 3.2.3.2. La durée T_r est de l'ordre d'une minute (choix fondé sur les connaissances *a priori* sur les signaux de régulation rapide).

b) Estimation de l'enveloppe haute

Définition L'*enveloppe haute* désigne (idéalement) le signal issu du foisonnement des usages autre que le chauffage électrique et que les compresseurs. Les composantes élémentaires de ce signal sont caractérisées par une régulation lente (à l'exception du sèche-linge) de durée minimale caractéristique T_H de l'ordre de 10 mn.

Filtrage morphologique mis en oeuvre L'analyse de CdCs contenant peu d'appareils montre que l'interprétation de ce signal est quasiment immédiate à partir des **maxima locaux** de la CdC sur un horizon temporel de durée T_H. Ceci est illustré sur la figure (4.1) dans le cas d'un mélange synthétique d'un chauffe-eau et d'un convecteur. Nous proposons d'extraire ce signal par une fermeture morphologique h_{T_H} dont l'élément structurant est une fenêtre de durée T_H : c'est un filtre qui extrait le signal défini par les maxima locaux évalués sur une fenêtre glissante de duréeT_H.

Estimation de l'enveloppe basse

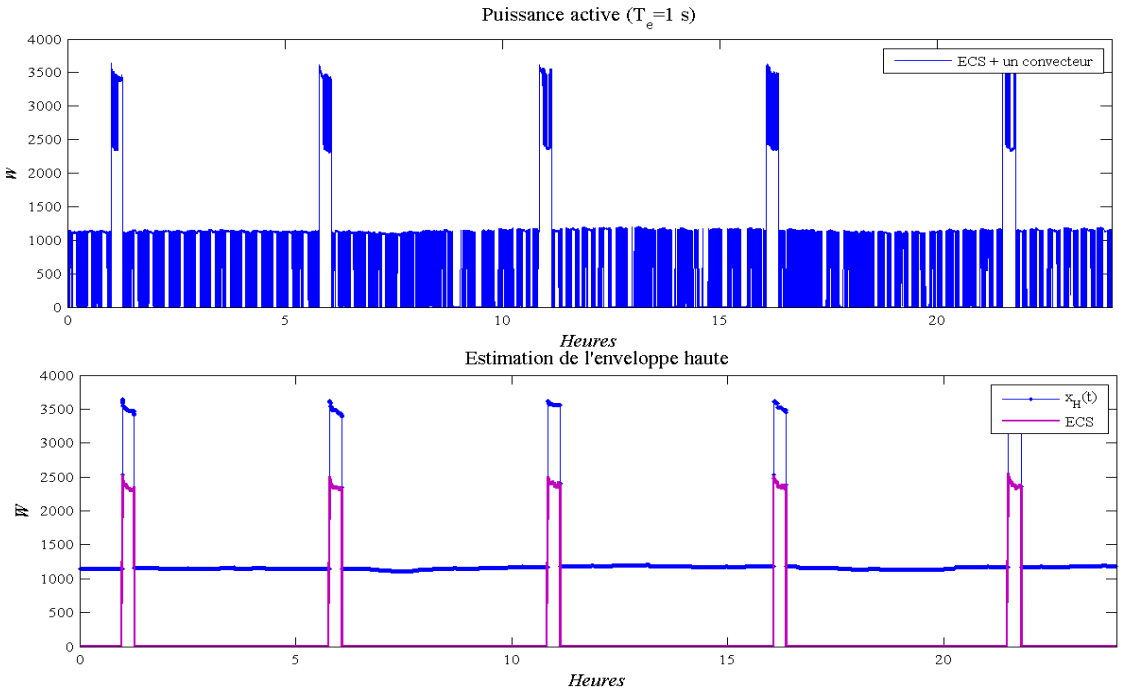


FIGURE 4.1: Filtrage : extraction de l'enveloppe haute de la CdC ($T_H = 700$ s)

Définition L'*enveloppe basse* désigne (idéalement) le signal issu du foisonnement des compresseurs et de la composante basse fréquence générée par le chauffage électrique. Les composantes élémentaires de ce signal sont caractérisés par une régulation lente de durée minimale caractéristique T_B très supérieure à la période de régulation des convecteurs (10 mn à 30 mn).

Filtre morphologique mis en oeuvre L'analyse de CdCs contenant peu d'appareils montre que l'interprétation de ce signal est quasiment immédiate à partir des **minima locaux** de la CdC sur un horizon temporel de durée T_B. Ceci est illustré sur la figure (4.2) dans le cas d'un mélange synthétique d'un convecteur et d'un réfrigérateur. Nous proposons d'extraire ce signal par une ouverture morphologique h_{T_B} dont l'élément structurant est une fenêtre de durée T_H : c'est un filtre qui extrait le signal défini par les minima locaux évalués sur une fenêtre glissante de duréeT_B.

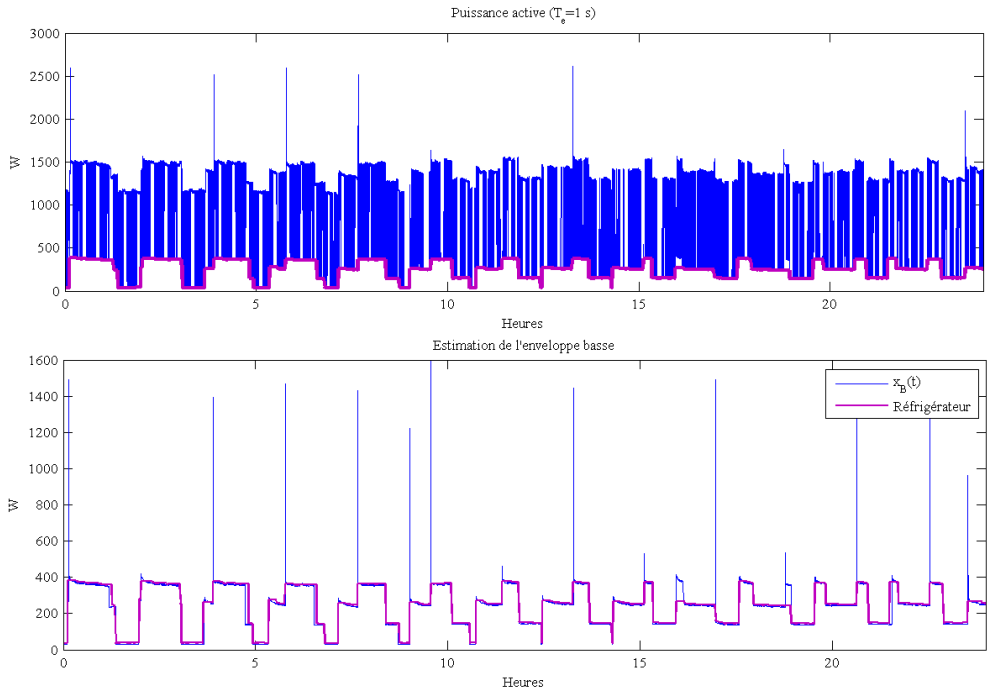


FIGURE 4.2: Filtrage : Illustration de l'extraction de l'enveloppe basse de la CdC ($T_B = 700$ s)

d) Illustrations Sur la figure 4.3, nous constatons que l'enveloppe haute est à une constante près égale à la puissance de l'ECS sur l'intervalle de temps [00h, 8h]. Une composante supplémentaire, liée au foisonnement des signaux de convecteurs, vient s'ajouter à ce signal sur l'intervalle [8h, 24h[. Par conséquent, il est difficile d'obtenir uniquement les usages autre que les convecteurs et les compresseurs dans l'enveloppe haute. L'enveloppe basse illustrée sur la figure 4.4 est le foisonnement d'un signal basse-fréquence généré par la superposition des signaux de convecteurs et du signal des compresseurs.

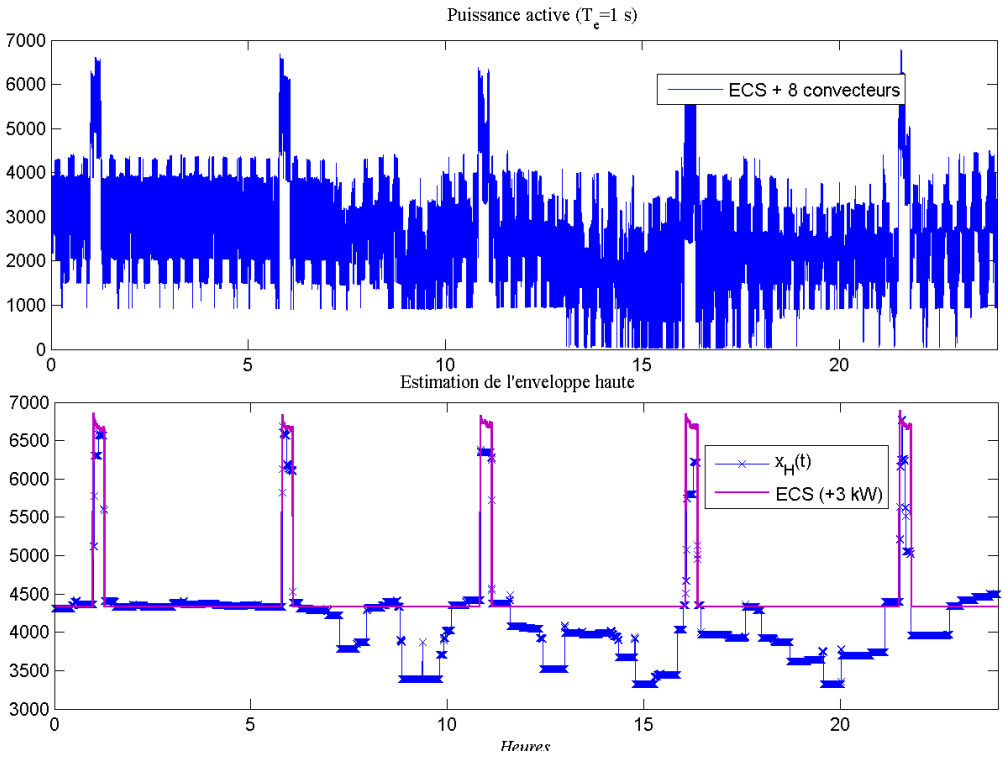


FIGURE 4.3: Filtrage : Illustration de l'extraction de l'enveloppe haute de la CdC ($T_H = 700$ s)

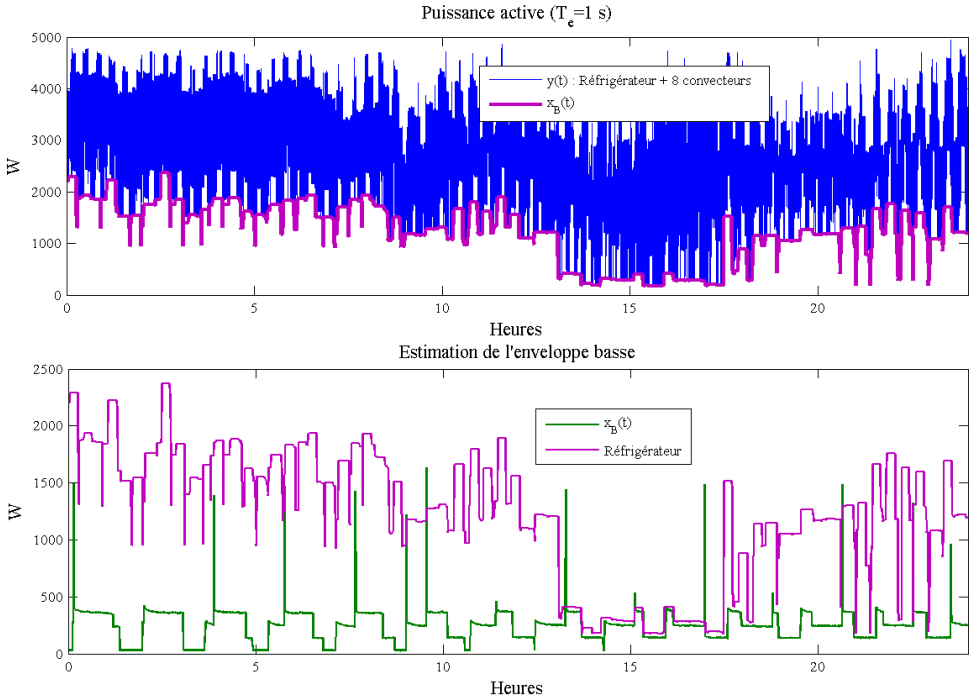


FIGURE 4.4: Filtrage : Illustration de l'extraction de l'enveloppe basse de la CdC ($T_B = 700$ s)

Ces deux exemples mettent en évidence l'impact de la CdC du chauffage sur l'extraction des enveloppes haute et basse et par conséquence l'estimation de la contribution des usages autre que le chauffage à l'énergie totale consommée. Cet effet est pris en compte dans le modèle décrit en (4.1) en pondérant l'énergie associée à chaque enveloppe par un coefficient qui dépend du nombre de convecteurs au cours du temps et du profil du consommateur.

4.2.3.2 Estimation du nombre de convecteurs $N_{co}(t)$

La méthode d'estimation du nombre de convecteurs consiste en deux étapes. La première vise à segmenter la CdC en zones à nombre de convecteurs actifs constant. La deuxime s'attache à estimer le nombre de convecteurs sur chaque zone.

Segmentation en zones à N_{co} constant La détermination des zones recherchées est obtenu par segmentation de la densité des événements $s_r(t)$ des régulations rapides dela CdC. En effet, la mise en marche ou l'arrêt d'un convecteur modifie la densité d'événements des régulations rapides. La segmentation de $s_r(t)$ est effectuée par l'algorithme CUSUM [BN93]. Un exemple de résultat est présenté sur la figure (4.5). Le seuil de détection de rupture est ajusté telle qu'une sur-segmentation du signal soi favorisée et ce afin d'améliorer les performances de l'estimation locale du nombre d'appareils (voir la figure 4.5).

Estimation de N_{co} Etant donné les intervalles à nombre de convecteurs (simultanéments en marche) constant, N_{co} est estimé à partir de la répartition (puissances cumulées par ordre croissant) de la valeur absolue des accroissements $\widetilde{y}_r(t)$ illustrée sur la figure (4.6). En effet, un convecteur en marche et ne fonctionnant pas en régime saturé génère une variation de la composante «régulations rapides» d'amplitude égale à celle de fonctionnement de l'appareil. Le résultat obtenu est une approximation du nombre minimum de composantes, car deux convecteurs d'amplitudes égales génèrent la même variation d'amplitude du signal $\widetilde{y}_r(t)$. Ainsi, le nombre de convecteur est estimé par le nombre de ruptures de la fonction de répartition de $\widetilde{y}_r(t)$, ces ruptures étant détectées par l'algorithme CUSUM. Sur l'exemple présenté sur la figure (4.6), cinq ruptures sont détectées, le nombre de convecteurs estimé est donc égal à cinq, le nombre effectif de convecteur étant égal à six.

Performances Bien que le nombre de convecteurs n'intervienne pas localement dans le modèle empirique, nous avons évalué les performances de son estimation locale. Les résultats obtenus à partir d'une base de CdCs semi-synthétiques (scénario de consommation simulés et appliquées à des signaux réels) sont présentés sur le tableau (4.1).

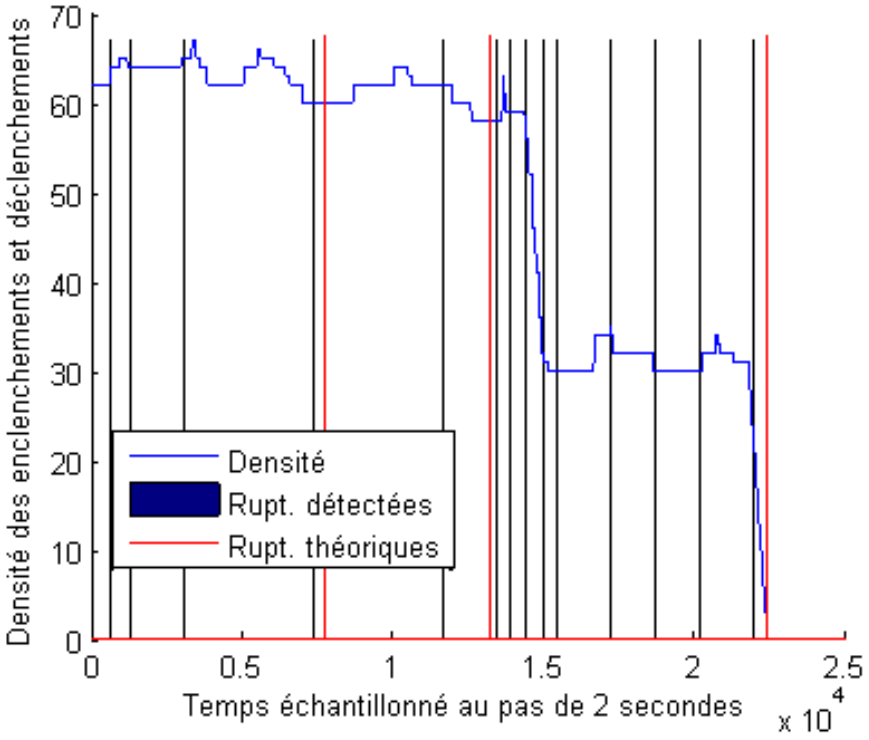


FIGURE 4.5: Segmentation de la densité des évévenements en zones à N_{co} constant ($t_{c0} = 700$ s)

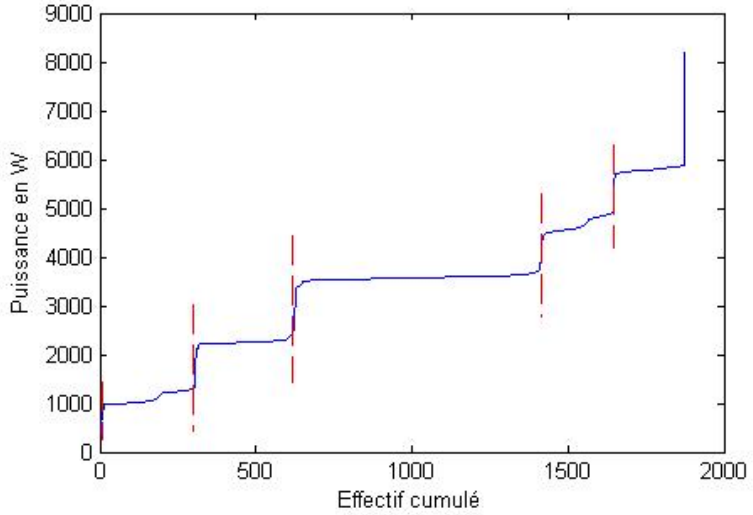


Figure 4.6: Détection de ruptures de la répartition de la composante « régulations »

N_{co}	Bonne estimation	Sous-estimation: -1	Sur-estimation : +1	Sous-estimation -2	Sur-estimation: > +2
1	48 %	10 %	29 %	9 %	3 %
2	36 %	7 %	35 %	19 %	2 %
3	30 %	20 %	42 %	5 %	3 %
> 3 et < 8	10 %	64 %	8 %	17 %	1 %
> 7 et < 12	5 %	27 %	63 %	4.5 %	0.05 %
> 11	1 %	14 %	82 %	3 %	0

TABLE 4.1: Performances de l'estimation du nombre de convecteurs à partir de la CdC générale

	a	b	a'	b'
profil 1	0.0049	-0.061	0.0345	0.8149
profil 2	0.0050	-0.0568	0.0552	0.7469
profil 3	0.0045	-0.0531	0.0526	0.7511
profil 4	0.0122	-0.1032	0.0131	0.8388
profil 5	0.0063	-0.0720	0.0256	0.8388
profil 6	0.0051	-0.0545	0.0434	0.7805

TABLE 4.2: Apprentissage des constantes du profil consommateur

Nous constatons que :
➢ si N_{co} est inférieure à 3, alors il est correctement estimé à $\pm$ 1 près ;
➢ si N_{co} est compris entre 3 et 8, est souvent sous-estimé de 1 ;
➢ si N_{co} est compris entre 7 et 12, est souvent sur-estimé de 1 ;
➢ si N_{co} est supérieur à 11, est souvent sur-estimé de 1.

4.2.3.3 Estimation des « constantes du profil-consommateur »

Ces paramètres sont estimés à partir d'une base de CdCs générales semi-synthétiques obtenue par simulation de cinq profils « type » de consommateurs et de plusieurs scénarios de consommation. Pour chaque profil, nous disposons de 100 CdCs quotidiennes.

Les paramètres recherchés sont estimés ainsi (4.4) :

$$(a, a', b, b') = \theta = \underset{\theta}{\operatorname{argmin}}(\mid \mathrm{e_{ch}} - \hat{\mathrm{e}}_{\mathrm{ch}}(\theta) \mid^{\mathbf{2}}) \tag{4.4}$$

où e_{ch} est et $\hat{e}_{ch}$ sont respectivement l'énergie effectivement consommée par le chauffage et l'énergie estimée de cet usage.

Les résultats obtenus sont résumés dans le tableau (4.2). Nous constatons que tous les coefficients sont positifs sauf le paramètre b. En effet, l'enveloppe haute contient la contribution des usages autre que le chauffage, dont il faut soustraire l'énergie consommée au modèle de l'énergie consommée par le chauffage, et une partie de la composante basse-fréquence du chauffage.

4.2.4 Procédure d'estimation de l'énergie du chauffage

Soit $y(t)$ la CdC quotidienne mesurée en sortie du compteur d'une installation donnée et $\hat{d}_1(t)$ le support temporel de présence du chauffage obtenu par seuillage de la densité des événements (détaillé dans la section 3.2.3). Les étapes d'estimation de l'énergie quotidienne consommée par le chauffage sont les suivantes :

1. Extraire les enveloppes haute $x_H(t)$ et basse $x_B(t)$ et le signal de régulations rapides $y_r(t)$ à partir de $y(t)$.
2. Sur le domaine de présence du chauffage, déterminer les zones à nombre de convecteurs constants.

3. Estimer le nombre de convecteurs sur chaque zone d'inétrêt identifiée en 2.
4. Estimer l'énergie quotidienne consommée par le chauffage définie par l'équation 4.1.

4.2.5 Evaluation de l'algorithme d'estimation de l'énergie consommée par chauffage

La base des signaux testés est constituée de 150 CdCs quotidiennes collectées chez deux clients résidentiels (une centaine de CdCs du client A et une cinquantaine du client B) pendant les six premiers mois de 2008.

Les résultats d'estimation de l'énergie consommée (client A et client B) sont présentés sur la figure (4.7), où l'énergie quotidienne totale consommée par tous les usages est présentée par la courbe bleue, l'énergie quotidienne réellement consommée par le chauffage est présentée par la courbe verte et l'estimée de l'énergie consommée par cet usage est présentée par la courbe rouge.

Précisons que nous ne disposons pas de la CdC chauffage de référence totale car trois convecteurs n'étaient pas instrumentés. Ces derniers correspondent à des convecteurs dont la consommation quotidienne est faible et dont le support temporel est réduit (exemple : sèche-serviette). La CdC de ces convecteurs peut être obtenue en deux étapes :

➢ Soustraie au mélange observée la puissance mesurée sur sept voies différentes (de l'appareil de mesure) ;
➢ extraire la composante régulation du signal obtenu [1], ce dernier approchant la puissance appelée par les trois convecteurs résiduels.

La comparaison de l'énergie de référence du chauffage à celle de tous les appareils met en évidence la part importante de cet usage dans la consommation quotidienne totale. Nous notons également que l'énergie estimée est très proche de(s) énergie(s) de référence pour les deux clients. Nous constatons que l'énergie consommée par le chauffage est souvent sur-estimée dans le cas du client B. L'analyse des signaux de ce client a mis en évidence des régulations rapides du chauffe-eau similaires à celles des convecteurs. Ces régulations rapides, qui sont révèlent un défaut de fonctionnement de l'appareil (non encore identifié), expliquent la sur-estimation du chauffage.

Pour évaluer l'algorithme, nous considérons l'erreur relative en points définie par l'équation (4.5) :

$$err = \frac{\hat{e}_{ch} - e_{ch}}{e_g} \tag{4.5}$$

où e_g est l'énergie quotidienne totale consommée, e_{ch} et $\hat{e}_{ch}$ étant respectivement l'énergie réelle et l'énergie estimée du chauffage. Les résultats obtenus à partir des signaux réels sont présentés sur la figure (4.8). Les erreurs d'estimation se situent globalement dans le domaine d'acceptabilité défini par EDF à $\pm 15\%$.

Nous retrouvons la tendance à la sur-estimation de l'énergie du chauffage chez le client B. Quant à l'erreur d'estimation observée chez le client A, elle présente deux comportements différents. Plus précisément, il y a globalement une tendance à sous-estimer le chauffage sur la première

1. Le filtre utilisé est le même exploité pour extraire $y_r(t)$.

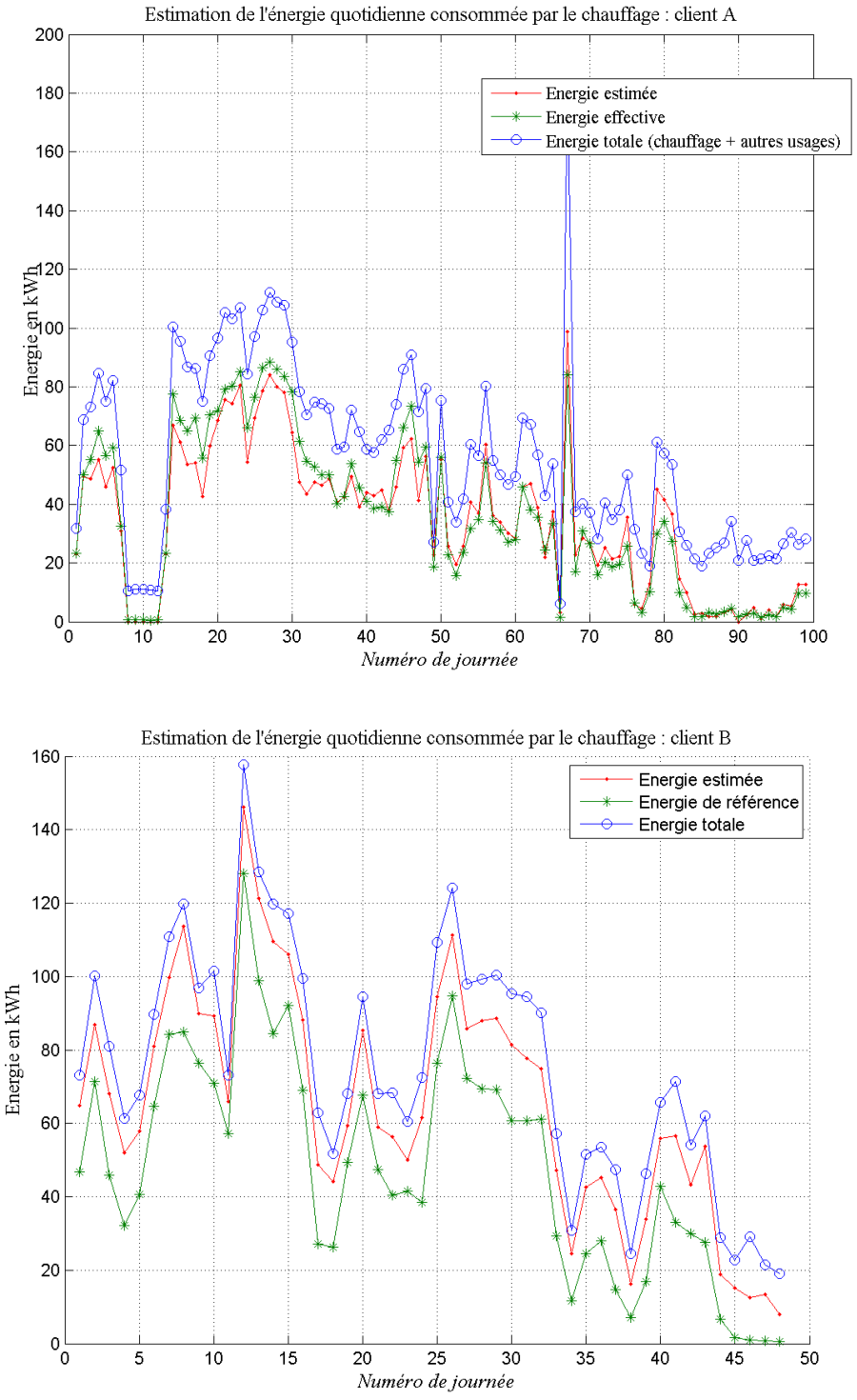


FIGURE 4.7: Estimation de l'énergie consommée par le chauffage (en kWh)

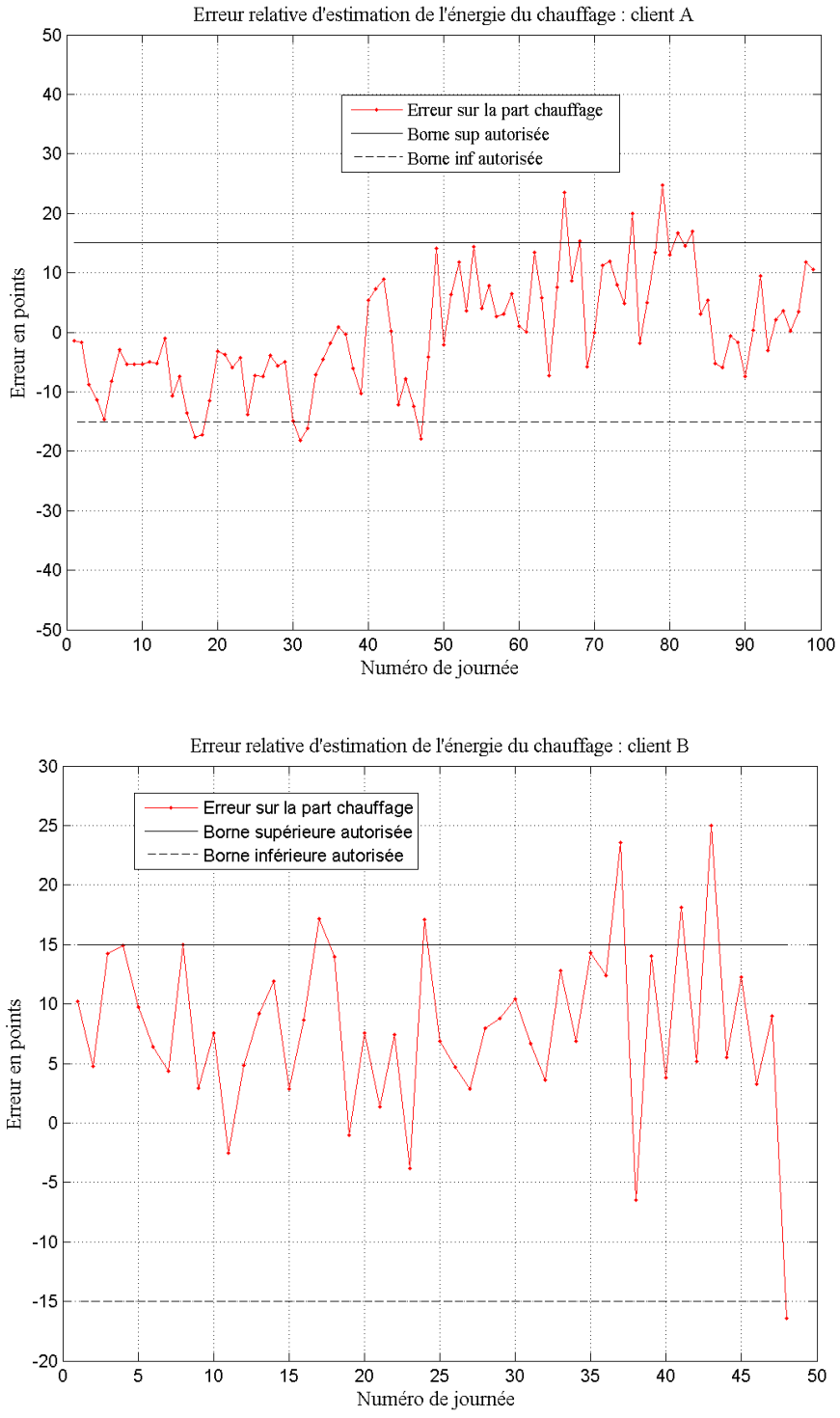


FIGURE 4.8: Erreur relative (en points) d'estimation de l'énergie consommée par le chauffage

cinquantaine de journées et plutôt une tendance à sur-estimer cet usage sur la deuxième moitié de la base de données. Ces deux comportements correspondent à un changement du niveau de consommation du chauffage : la figure (4.7) montre en effet que la consommation quotidienne de cet usage baisse d'un facteur 2 sur la deuxième moitié de la base de signaux-test, qui correspond à la période printanière et au début du mois de juin. Par conséquence, les variations observées de l'erreur d'estimation chez le client A nous amènement aux deux conclusions suivantes.

- ➢ L'algorithme a tendance à sous-estimer l'énergie consommée par le chauffage, quand la consommation de ce dernier est très élevée. Cette sous-estimation peut s'expliquer par la saturation de convecteurs et /ou la part importance de la composante basse fréquence du signal de chauffage généré par un foisonnement des convecteurs qui sont quasiement tous en marche.
- ➢ L'algorithme a tendance à sur-estimer l'énergie consommée par le chauffage pendant les périodes où la consommation de ce dernier baisse par rapport à la consommation en hiver. Cela peut s'expliquer par la présence du sèche-linge, dont la consommation n'est plus négligeable par rapport à celle réllement consommée par le chauffage, et par le choix des constantes du profil consommateur. En effe, il faudrait envisager d'appliquer des constantes différentes en fonction de la saison.

Ces deux conclusions peuvent alarmer le client. Mais il ne sera a priori pas lésé car l'erreur commise est quasiment symétrique entre les deux périodes. Cela se traduit par exemple par un histogramme global de l'erreur relative d'estimation de la part du chauffage dans la consommation quotidienne illustré sur la figure (4.9) quasi-symétrique autour de zéro.

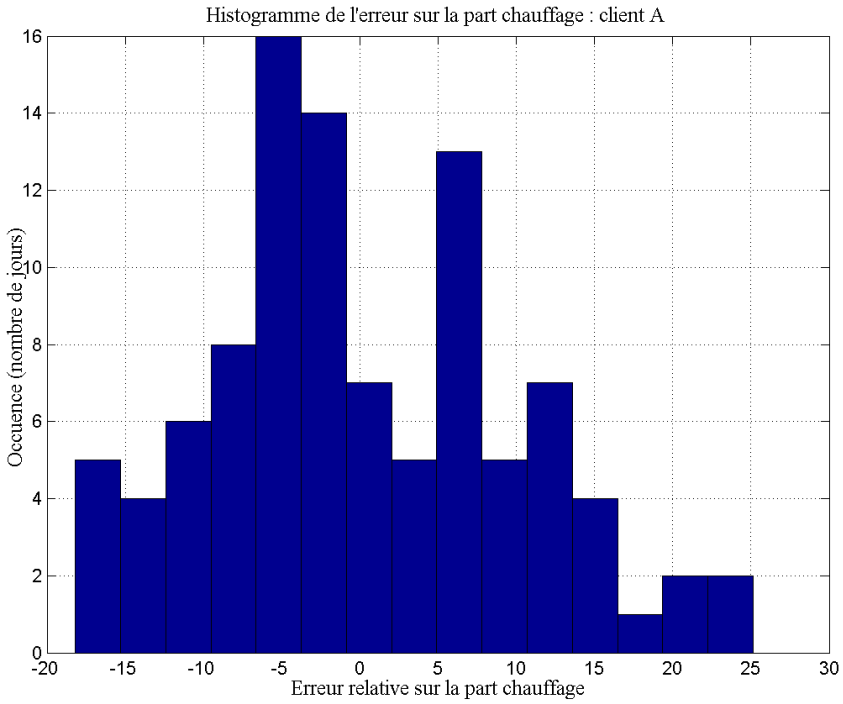


FIGURE 4.9: Histogramme de l'erreur relative sur l'estimation de la part du chauffage dans l'énergie consommée par jour

4.3 Estimation de l'énergie consommée par le chauffe-eau

4.3.1 Principe

L'approche adoptée pour estimer la consommation quotidienne de l'ECS est résumée sur la figure (4.10).

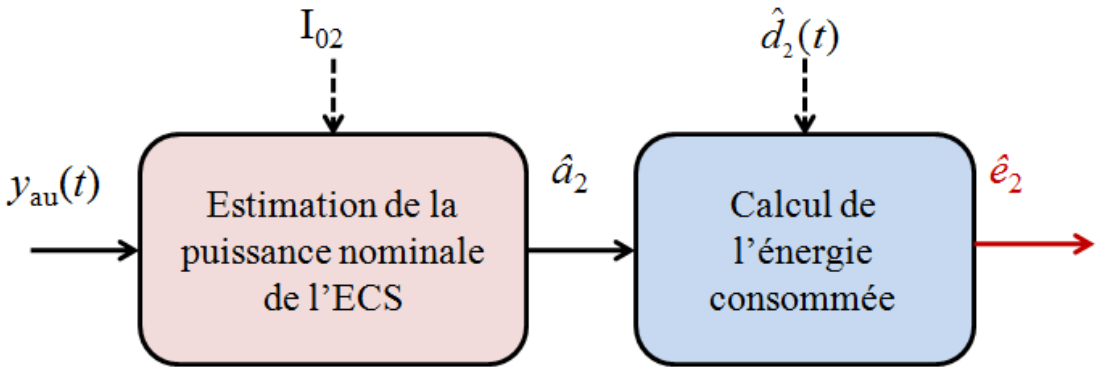


FIGURE 4.10: Estimation de l'énergie consommée par l'ECS

La puissance nominale est estimée [2] à partir du signal filtré $y_{au}(t)$ sur un intervalle de temps correspondant à une baisse d'activité et à une consommation réduite du chauffage. Nous avons retenu l'intervalle $I_{02} = [00\ h, 05\ h]$.

L'énergie quotidienne de l'usage étudié est alors donnée par : $\hat{e}_2 = \hat{a}_2 \int \hat{d}_2(t)\,dt$ où $\hat{d}_2(t)$ et $\hat{a}_2$ sont respectivement le détecteur de l'ECS et l'amplitude de fonctionnement estimée de cet usage.

4.3.2 Résultats et discussions

L'algorithme d'estimation de l'énergie quotidienne consommée par l'ECS est évaluée sur une base de signaux constituée d'environ 140 CdCs quotidiennes mesurées chez le client sur une période allant de janvier 2008 à juillet 2008. L'énergie consommée par le chauffe-eau et la part de cette consommation par rapport à l'énergie totale consommée sont représentées sur la figure (4.11). Nous constatons que l'énergie consommée présente une faible variabilité. Mais la part de cet usage dans l'énergie totale est très variable. En effet, elle dépend fortement de la présence ou non du chauffage dont la consommation est souvent très élevée.

L'histogramme de l'erreur d'estimation (définie par l'équation (4.5)) est présenté sur la figure (4.12). Nous remarquons que la moyenne de l'erreur d'estimation est quasi-nulle, ce qui nous rassure quant à l'absence d'un biais d'estimation. Tout comme pour le chauffage, l'erreur d'estimation de l'énergie quotidienne consommée par le chauffe-eau respecte globalement l'objectif fixé par EDF [CG08].

2. médiane des amplitudes des événements d'amplitude supérieure à l'amplitude minimale de l'ECS

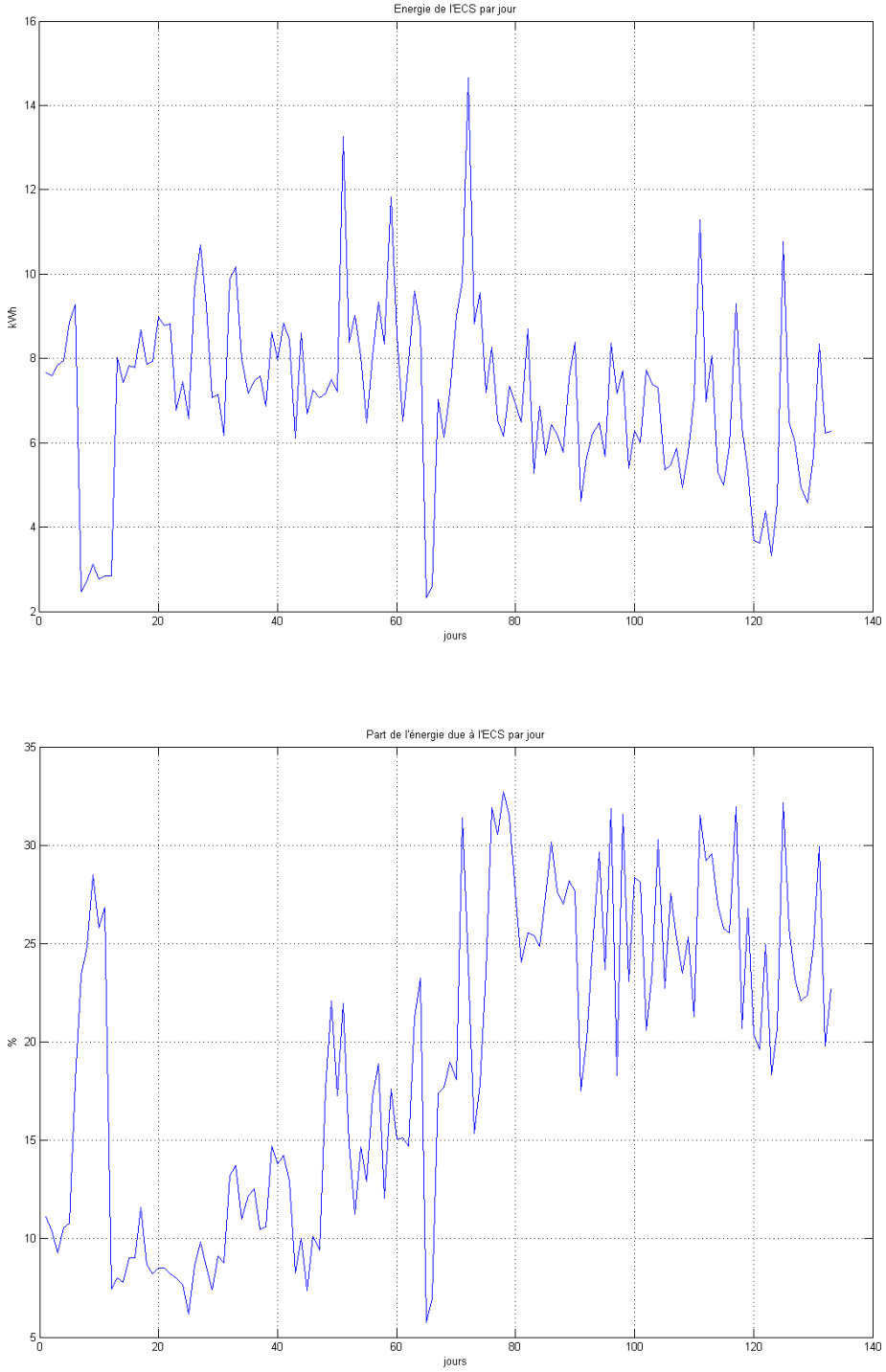


FIGURE 4.11: Energie consommée par le chauffe-eau

4.4 Identification des paramètres d'un modèle

4.4.1 Exemples de modèles d'amplitude

L'analyse de la puissance moyenne $x_2(t)$ appelée par un chauffe-eau particulier met en évidence une décroissance exponentielle de l'amplitude du signal mesuré durant un cycle de fonctionnement $[t_0, t_1]$ illustré sur la figure (2.13). Nous proposons alors le modèle suivant pour chaque cycle de fonctionnement du chauffe-eau (4.6) :

$$x_2(t) = c \exp\left(-\frac{t}{\tau}\right) + d \tag{4.6}$$

où c, d et τ sont respectivement la variation maximale de l'amplitude du signal sur $[t_0, t_1]$, la puissance nominale de fonctionnement du chauffe-eau et une constante de temps caractéristique de l'appareil et des conditions de fonctionnement.

4.4.2 Estimation du modèle de l'ECS

L'estimation des paramètres du modèle est effectuée à partir des signaux mesurés pour l'usage ECS seul.

Etant donnés les instants de démarrage et de fin de cycles du chauffe-eau (3.3), les paramètres du chauffe-eau sont estimés par minimisation d'un critère L_2 dont une illustration est fournie sur la figure (4.13). L'estimation des paramètres du modèle est un problème d'optimisation non-linéaire sans contraintes (résolu par la fonction *fminsearch.m* de Matlab).

4.4.3 Résultats

Les paramètres du modèle sont identifiés à partir de la CdC de l'ECS (Maison Domotique) sur une durée de 11 jours. Le nombre de cycles de fonctionnement vaut 51. Les résultats obtenus sont présentés sur la figure (4.14). Nous constatons une faible dispersion des valeurs estimées pour chaque paramètre, à l'exception de la puissance nominale où certaines s'écartent significativement de la moyenne des estimées. En effet, il s'agit d'une conséquence d'une détection imprécise du début et de la fin des cycles de fonctionnement.

Le modèle proposé est également évalué à partir de l'analyse de l'histogramme de l'erreur de modélisation (écart entre le modèle et le signal observé pour chaque cycle de fonctionnement) illustré sur la figure (4.15). Cet histogramme est de moyenne nulle et d'écart type 18. Ainsi, nous vérifions l'absence de biais d'estimation et notons que l'écart type du résidu est du même ordre de grandeur du bruit de mesure (variant entre 20 et 25 en fonction du site étudié).

Remarque Ces paramètres sont utiles pour :

- ➢ améliorer l'estimation de l'énergie consommée par le chauffe-eau ;
- ➢ aider le client à choisir le « meilleur » chauffe-eau (volume, puissance nominale) en fonction de ses besoins ;
- ➢ surveiller l'état de l'appareil à travers la constante de temps par exemple (un entartrage donnerait *a priori* lieu à une décroissance moins rapide du signal).

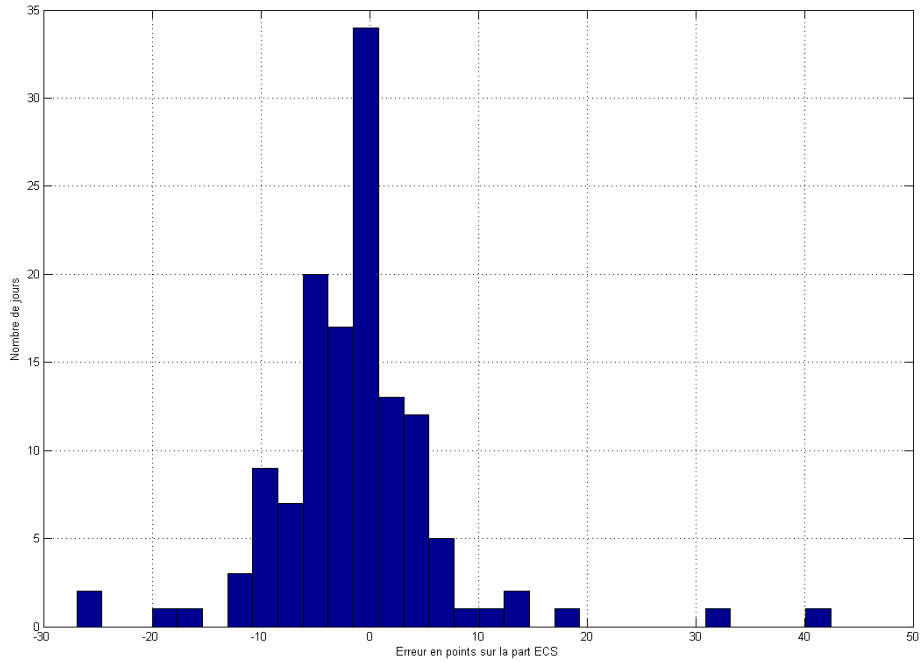


FIGURE 4.12: Erreur d'estimation de l'énergie consommée par le chauffe-eau

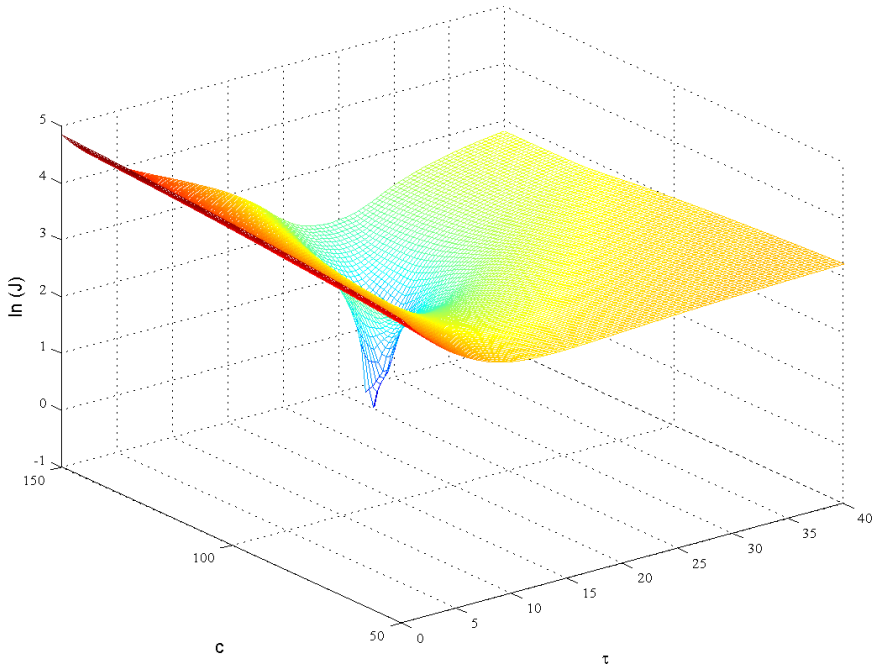


FIGURE 4.13: Estimation du modèle de l'ECS :illustration du critère (d =2380 W)

4.5 Estimation des paramètres dans un cadre probabiliste

4.5.1 Estimation par MV des amplitudes des compresseurs

Soit $(t_v, a_v(t))$ la série des événements de la CdC des compresseurs d'un hypermarché. Les histogrammes des amplitudes des événements détectés sont présentés sur la figure (4.16) obtenus à partir de la puissance active et de la puissance réactive.

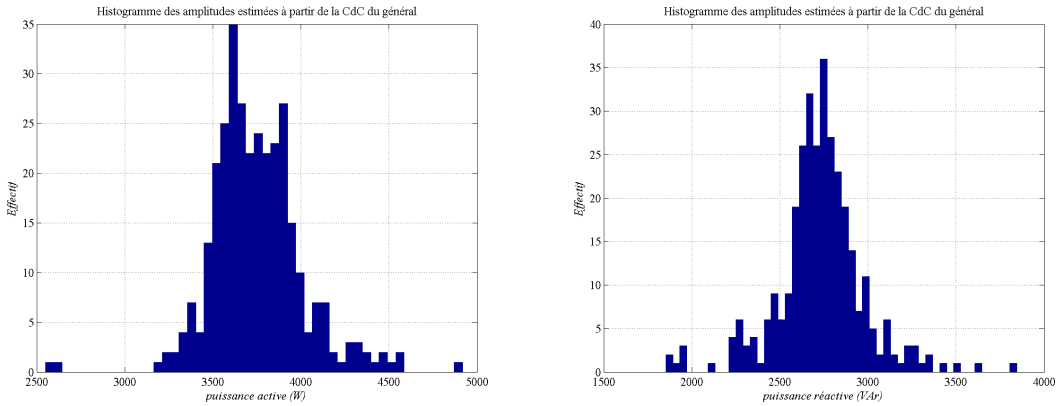


FIGURE 4.16: Estimation de l'amplitude moyenne des compresseurs à partir de la CdC générale d'un hypermarché

L'amplitude des événements est considérée comme la réalisation d'une variable aléatoire A_v gaussienne de moyenne d (c'est la puissance nominale de l'appareil) et de variance σ_v^2. En effet, mise à part la présence de pics de démarrage de compresseurs, l'amplitude du signal varie très peu durant un cycle de fonctionnement. Les pics des signaux sont filtrés par un filtre morphologique (ouverture puis fermeture). Nous détaillons la détection des événements et le filtrage dans le chapitre suivant (5.2.3). L'estimation de d au sens du maximum de vraisemblance est la moyenne des valeurs mesurées $a_v(t)$. Les résultats obtenus sont détaillés en (5.2.3). En effet, l'estimation de ce paramètre fait partie d'un algorithme de reconstruction des signaux de compresseurs d'un hypermarché.

4.5.2 Estimation MAP des caractéristiques du chauffage électrique

Dans le cas des convecteurs électriques, nous souhaitons connaître les amplitudes, les périodes de régulation ainsi que l'évolution du rapport cyclique au cours du temps. Nous disposons de connaissances *a priori* sur ces inconnues acquises grâce à l'analyse des signaux. Ces paramètres sont estimés par inférence bayésienne détaillée dans la section (5.1). La méthode exploite un modèle probabiliste des signaux de convecteurs.

Synthèse

L'estimation des caractéristiques (énergie, amplitude, etc) de certains usages est fondamentale pour la conception des services énergétiques. Différentes méthodes déterministes / probabilistes faisant intervenir ou non un modèle contribuent à la résolution de ce problème. Ces méthodes nécessitent souvent une étape de détection des événements ou la reconnaissance des signaux de la CdC. Dans ce chapitre, nous avons principalement présenté une méthode de cartographie de l'énergie quotidienne. Plus précisément, deux algorithmes dédiés à l'estimation de l'énergie consommée par le chauffage et celle consommée par le chauffe-eau sont présentés. Les permiers résultats obtenus à partir de signaux réels et semi-synthétiques sont prometteurs. Dans les deux cas, bien que les méthodes soient automatiques, elles nécessitent l'ajustement de plusieurs paramètres. Ce problème a été abordé de deux manières : certains paramètres sont fixés en s'appuyant sur des connaissances a priori sur les caractéristiques des signaux, d'autres sont ajustés en exploitant une base de signaux (d'apprentissage) où les configurations des usages sont connues (exemple : constantes du profil du consommateur dans le modèle de l'énergie consommée par chauffage). Maintenant que nous disposons d'une base de signaux de plus en plus riche, il reste à étudier en profondeur le choix des différents paramètres dans le cadre d'une étude de performances des algorithmes plus approfondie.

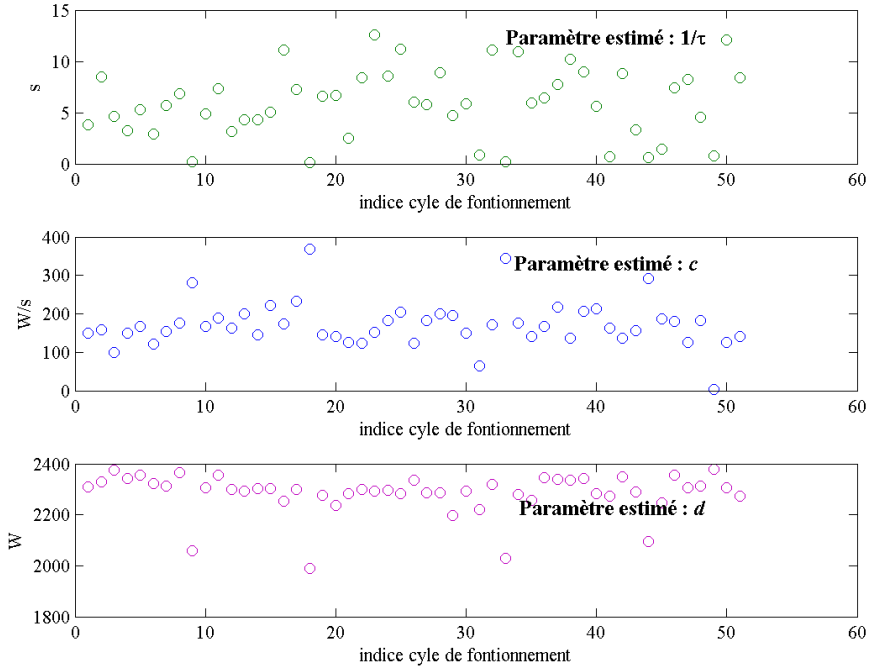


FIGURE 4.14: Estimation du modèle de l'ECS à partir des données réelles

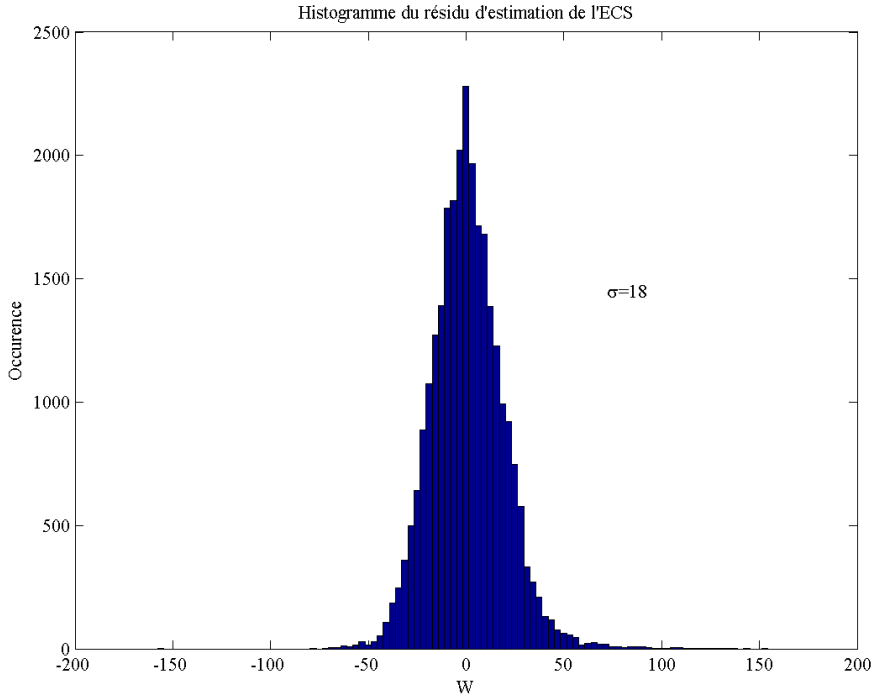


FIGURE 4.15: Histogramme du résidu d'identification du modèle

Chapitre 5

Séparation des sources de la courbe de charge générale

Introduction

L'analyse des signaux a mis en évidence l'impact du foisonnement des usages et de leur diversité sur la complexité de l'analyse de la CdC générale. Cependant, la caractérisation complète de tous les signaux d'une CdC globale peut être envisagée dans certains cas, par exemple :

- ➢ Le système étudié est constitué d'un seul usage ;
- ➢ Le système étudié est composé de deux usages que l'on peut décrire par des dictionnaires de formes temporelles parcimonieux entre eux, par exemple un mélange de la CdC du chauffage électrique et de la CdC du froid alimentaire.

En pratique, il peut s'agir :

- ➢ soit d'un sous-comptage ;
- ➢ soit d'une CdC générale, où l'on a quasiment un seul usage en marche, par exemple :
  - – une CdC d'un client particulier, quand le client est absent ;
  - – ou encore celui d'une CdC générale d'un hypermarché entre minuit et 5 heures, où les deux usages prépondérants sont les compresseurs (signaux intermittents) et l'éclairage (CdC quasi-constante sur la durée de fonctionnement).

Cette classe de signaux sera appelée CdC générale quasi-mono-usage. Les caractéristiques des signaux élémentaires de cette classe de signaux rend leur reconstruction envisageable.

L'estimation des composantes élémentaires de la CdC [1] s'inscrit dans le cadre du problème générique de séparation de sources non-stationnaires de nombre de composantes inconnu, à partir d'une seule observation. Ce problème générique a fait l'objet de nombreuses études depuis les années 2000, principalement au sein de la communauté de traitement de signaux sonores. Les méthodes proposées dans la littérature sont principalement des méthodes à base de dictionnaires. Elles s'appuient par exemple sur la parcimonie des sources dans un domaine transformé judicieusement choisi [BGB01, BB03, BMa06]. Dans le cas de la CdC électrique quasi-mono-usage, une telle démarche n'est pas envisageable au regard des propriétés des signaux élementaires : une démarche dédiée s'inspirant des travaux antérieurs est nécessaire. Ce choix est justifié par les deux particularités suivantes de ce signal :

1. Puissance active ou puissance réactive

➢ Des signaux associés au même usage (chauffage par exemple) peuvent se superposer localement ou sur l'ensemble du domaine d'observation.
➢ Ces signaux partagent la même morphologie.

Nous proposons une méthode de reconstruction des signaux qui intègre les connaissances *a priori* sur les signaux élémentaires. Un modèle stochastique du mélange est proposé et se fonde sur la description des signaux à partir d'un dictionnaire de formes temporelles pré-établies. Un échantillonneur du type MCMC est ensuite utilisé pour estimer la solution au sens du maximum *a posteriori*.

Cette démarche est détaillée dans le cas particulier de la CdC du chauffage où les sources du mélange sont des signaux localement stationnaires et quasi-périodiques. Mais, elle pourra être étendue à d'autres classes de CdC, tant que les signaux élémentaires peuvent être décrits à partir d'un dictionnaire de formes et pour lesquels des connaissances a priori sur les lois temporelles des événements ou des durées de cycles de fonctionnement sont disponibles.

Remarque

Le choix du chauffage est lié aux propriétés des signaux et aussi à la part importante de la consommation de cet usage. Estimer ses composantes élémentaires permettrait de développer des services énergétiques tel que le délestage par exemple.

5.1 Approche probabiliste de décomposition de la CdC du chauffage

5.1.1 Modèle stochastique

La CdC du chauffage est un mélange d'un nombre inconnu de créneaux quasi-périodiques $x_i(t)$:

$$y_{ch}(t) = \sum_{i=1}^{N_b} x_i(t) + \varepsilon_{ch}(t) \tag{5.1}$$

où $\varepsilon_{ch}(t)$ est un bruit gaussien centré et N_b est le nombre (inconnu) des convecteurs.

Un exemple de signaux élémentaires est présenté sur la figure (5.1).

Nous avons mis en évidence lors de l'analyse des données que ces signaux se caractérisent par :
➢ le caractère quasi-périodique de la puissance moyenne d'un convecteur ;
➢ la symétrie entre l'amplitude d'un *on* et celle d'un *off* successifs ;
➢ l'évolution au cours du temps du rapport cyclique.

Un modèle paramétrique de chaque signal de convecteur peut être défini par :
➢ des inconnues qui sont déterministes, à savoir la période de régulation T_{0i} et l'amplitude A_i ;
➢ des inconnues qui sont des variables aléatoires :
  – date de début t_{0i} ;
  – le nombre de cycles de fonctionnement N_{ci} ;
  – les valeurs du rapport cyclique correspondant respectivement représentées par $\rho_{T_{0i}} = [\rho_{T_{0i},1}, \rho_{T_{0i},2}, ..., \rho_{T_{0i},N_{ci}}]$.

Soit $[t_d, t_f]$ le support temporel du signal étudié, les domaines de définition des différents paramètres de chaque convecteur sont définis comme suit :

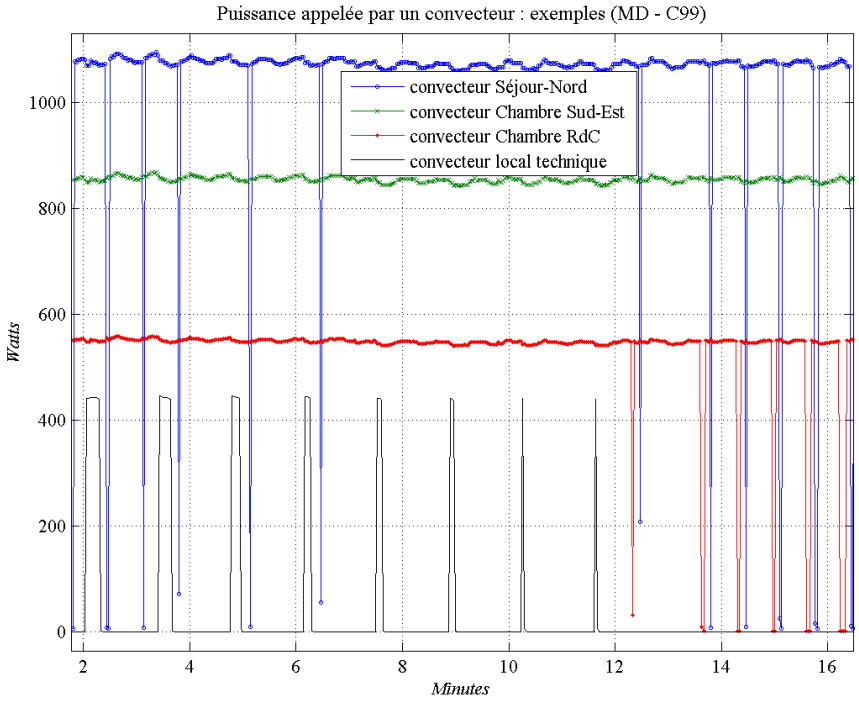


FIGURE 5.1: Signaux de puissance de convecteurs fonctionnant simultanément ($T_e = 2$ s)

$$\begin{cases} (1) \forall i \in [1, N_b],\ t_d \leq t_{0i} \leq t_f \\ (2)\ 0 \leq N_{ci} \leq N_{im} \\ (3)\ \forall k \in [1, N_{ci}],\ 0 \leq \rho_{T_{0i}}(k) \leq 1 \end{cases} \tag{5.2}$$

où N_{im} est le nombre de cycles maximum du *ième* convecteur sur l'intervalle $\left[t_d, t_f\right]$.
Soit $\theta_i = \left[t_{0i}, N_{ci}, \rho_{T_{0i}}\right]$ le vecteur des inconnues aléatoires décrivant un convecteur donné.
L'espace d'état (espace des paramètres) est de grande dimension. Disposer de lois *a priori* sur ces paramètres permettrait de faciliter la recherche de ces paramètres.

5.1.1.1 Lois *a priori*

Date de début En absence d'information sur la date de démarrage d'un convecteur donné, cette inconnue est décrite par une loi uniforme définie sur le domaine d"étude du signal :

$$\forall i \in [1, N_b],\ t_{0i} \sim U_{[t_d, t_f]}$$

Période de régulation Quant à la « période » du signal d'un convecteur donné, on peut la modéliser ainsi :
$T_{0i} = T_0 + \varepsilon_{T_0}$ où T_0 vaut 40 s ou 80 s, où ε_{T_0} est un bruit gaussien centré dont la variance, égale à 3, est calculée à partir de l'analyse des signaux (chapitre 2, figure 2.5).

Durées de « on » (rapport cyclique) L'analyse des variations du rapport cyclique de différents convecteurs a mis en évidence les variations lentes de cette grandeur. Un modèle fondé sur une transformation logistique sera alors utilisé pour décrire cette grandeur. Plus précisément, nous avons choisi de modéliser les durées de la phase « *on* » des créneaux d'un convecteur par l'intemédiaire d'une loi logistique, décrite par la relation suivante (5.3) :

$$\begin{cases} \rho_{T_{0i}}(k) = \frac{\exp(U_k)}{1-\exp(U_k)} \\ U_k = \frac{1}{2}(U_{k-1} + U_{k+1}) + \varepsilon_k \text{ où } \varepsilon_k \text{ est un bruit additif blanc gaussien centré.} \end{cases} \tag{5.3}$$

En supposant que que le différents paramètres sont indépendants, la loi *a priori* sur les paramètres $f(\Theta)$ est le produit des lois a priori des différents paramètres.

5.1.1.2 Vraisemblance des données

Le problème de décomposition de la CdC peut ainsi être ramené à un problème de minimisation d'énergie sur l'espace des paramètres. Il s'agit alors d'estimer les signaux élémentaires de telle sorte que l'ajustement des dates d'enclenchement, de l'amplitude et des rapports cycliques soit optimal. On propose alors de considérer un critère d'énergie constitué d'un terme d'attache aux données, décrit par l'équation (5.4)

$$J(\Theta) = \|y_{ch}(t) - y_m(t)\|^2 \tag{5.4}$$

où $y_m(t) = \sum_{i=1}^{Nb} x_i(t, T_{0i}, A_i, \theta_i)$ et Θ est le vecteur qui représente les inconnues du modèle.
La vraisemblance des données $f(y_{ch}/\Theta)$ est une loi normale (étant données les propriétés du bruit de modélisation.

5.1.1.3 Loi *a posteriori*

La loi *a posteriori* des paramètres notée $f(\Theta)$ est donnée par la relation suivante (5.5):

$$f(\Theta) \propto f(y_{ch}/\Theta)\, f(\Theta) \tag{5.5}$$

L'espace d'état est de dimension finie mais de grande cardinalité. Une approche possible pour estimer les paramètres consiste à utiliser un échantillonneur stochastique pour explorer l'espace d'état et un algorithme de recuit simulé. L'algorithme d'optimisation mis en œuvre est décrit dans la section suivante.

5.1.2 Estimation des paramètres

Hormis la grande taille de l'espace des paramètres recherchés, notons que le nombre de variables à estimer est inconnue et peut évoluer au cours du temps suite à la mise en marche ou à l'arrêt d'un convecteur par exemple. Les méthodes RJ-MCMC[KS] (méthodes de Monte Carlo par Chaînes de Markov à Sauts-Réversibles) offrent la possibilité de traiter cette classe de problèmes. Une introduction à ces approches est présentée en annexe (E). Nous pésentons ici les développéments dédiés à la courbe de charge uniquementà savoir les transitions définies pour

explorer l'espace d'état et la procédure finale (fondée sur l'algorithme du recuit simulé) pour l'estimation des paramètrs au sens du maximum *a posteriori*.

Dans le cas d'application étudié, nous considérons que le nombre total de convecteurs présents dans l'installation, mais pas nécessairement actifs, est connu. Les amplitudes des signaux associés sont connues.

5.1.2.1 Exploration de l'espace d'état de la CdC

L'exploration de l'espace des paramètres est assurée par un échantillonneur stochastique, ici l'échantillonneur de Metropolis-Hastings.

Transformations élémentaires Les transformations élémentaires considérées sont les suivantes :

Translation d'un signal de convecteur : Il s'agit de modifier uniquement la date de début. Cette perturbation est réalisée selon une loi uniforme ou selon un modèle autorégressif d'ordre 1.

Modification d'une valeur du vecteur rapport cyclique : Cette perturbation consiste à modifier une seule valeur du vecteur $\rho_{T_{0i}}(k)$ tirée selon une loi uniforme parmi les valeurs courantes. La modification consiste à remplacer la valeur courante en fonction des deux valeurs voisines en tenant compte de la relation (5.3).

Ajout/Suppression d'un créneau en début de séquence : Un créneau ajouté a une durée *on* égale à celle du premier créneau de la séquence de créneaux en cours (du convecteur étudié). Une mise à jour de la date de début est effectuée.

Dans le cas d'une suppression de créneau, le nombre de cycles de la séquence est mis à jour.

Ajout/Suppression d'un créneau en fin de séquence : Un créneau ajouté a une durée *on* égale à celle du dernier créneau de la séquence de créneaux en cours (du convecteur étudié).

En cas d'ajout ou de suppression, le nombre de cycles de la séquence est mis à jour.

Modification de la date de début et ajustement du rapport cyclique : Les valeurs des durées de « on » des créneaux de la séquence courante sont modifiées selon une transformation logistique.

Création des signaux de convecteurs initiaux : Soit $\left[t_d, t_f\right]$ l'intervalle d'intérêt.

La configuration initiale des convecteurs est générée de la manière suivante :

1. Initialiser le nombre de convecteurs N_b initial. Deux méthodes sont proposées

    a) ce paramètre est estimé en exploitant la fonction de répartition des accroissements du signal observé ;

b) N_b est généré aléatoirement selon une loi de probabilité. On peut choisir par exemple $\mathscr{U}([0, N_{bmax}])$ où N_{bmax} est le nombre maximum autorisé de convecteurs, ou encore un processus poissonien.

2. Initialiser l'amplitude de chacun des N_b convecteurs. Plusieurs procédés sont envisageables, dont :

   a) l'amplitude est générée selon une loi uniforme, dont le domaine est pré-défini à travers l'étude de la fonction de répartition des accroissements du signal observé.

   b) l'amplitude est générée selon une loi uniforme, dont le domaine est fonction des amplitudes minimale et maximale autorisées.

3. Générer les périodes de régulation des N_b convecteurs, où les deux périodes possibles (40 s et 80 s) sont équiprobables.

4. Pour chaque convecteur :

   a) initialiser la date de début : $t_{0i} \sim \mathscr{U}([t_d, t_f])$

   b) initialiser le nombre de cycles : $N_{ci} = (t_f - t_{0i})/T_{0i}$

   c) initialiser le vecteur rapport cyclique à la valeur 0.5.

5.1.2.2 « Chaînage » de signaux de convecteurs

Etant donné que la durée de fonctionnement d'un convecteur est souvent élevée, nous avons choisi de concevoir un algorithme séquentiel pour décomposer la CdC du chauffage. Plus précisément, la CdC observée est découpée en plusieurs fenêtres de durées égales à une constante T_{CH}. En pratique, $T_{CH} \sim 5$ mn $-$ 10 mn, et correspond à la durée minimale de fonctionnement d'un convecteur que l'on souhaite détecter et caractériser. La CdC observée est traitée séquentiellement sur les différentes fenêtres successives. Traiter le signal observé sur chaque domaine séparément n'est pas judicieux étant donné :

➢ que d'une part un signal de convecteur peut se prolonger sur plusieurs fenêtres successives ;

➢ et que d'autre part, un signal de convecteur présente une structure particulière (quasi-périodique) qui pourrait simplifier l'exploration de l'espace d'état.

D'où l'introduction d'une procédure de chaînage des signaux de convecteurs entre fenêtres successives.

Soient $W_k = [t_{d,k}, t_{f,k}]$ et $W_{k+1} = [t_{d,k+1}, t_{f,k+1}]$ deux fenêtres successives d'analyse du signal observé $y_{ch}(t)$. On se place dans le cas où l'algorithme de reconstruction des signaux de convecteurs appliqué sur la fenêtre W_k a convergé vers une solution non-strictement nulle.

L'objectif du chaînage est de définir les signaux de convecteurs initiaux sur la fenêtre W_{k+1} compte tenu de la solution obtenue sur la fenêtre W_k définie par l'équation suivante :

$$\hat{y}_m(t) = \sum_{i=1}^{\hat{Nb}} \hat{x}_i(t, \hat{T}_{0i}, \hat{A}_i, \hat{\theta}_i) \text{ où } t \in [t_{d,k}, t_{f,k}]$$

On considère les convecteurs $\hat{x}_i(t)$ tour à tour. Pour chaque signal élémentaire, trois cas de figure se présentent illustrés sur la figure (5.2).

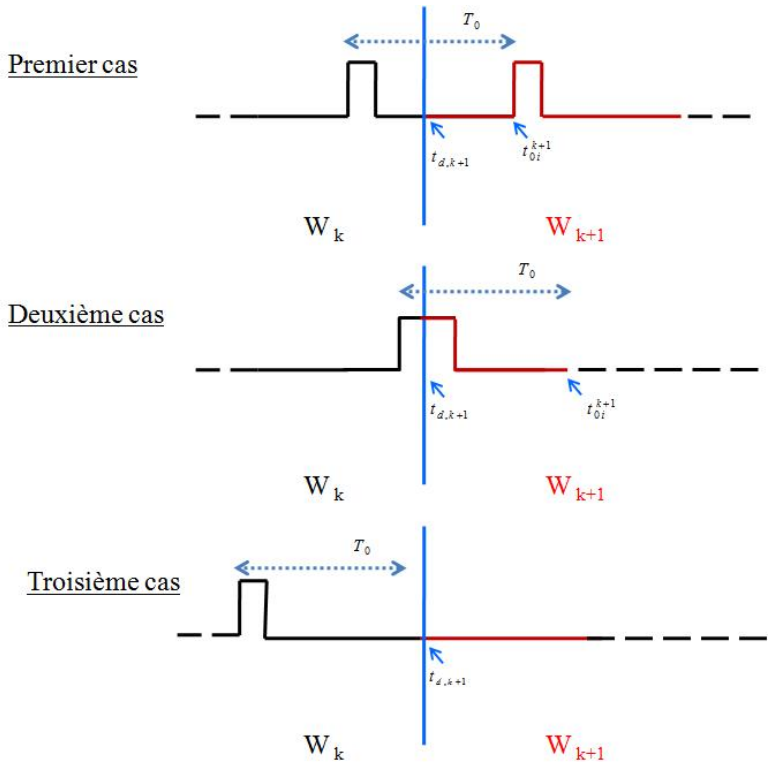


Figure 5.2: Configurations possible d'un convecteur sur deux fenêtres successives

Premier cas

Fin du dernier cycle du convecteur dans la fenêtre W_{k+1} et dernier «*on*» se terminant dans la fenêtre W_k

- ➢ initialisation de la date de début : $t_{0i}^{k+1} = t_{0i}^{kopt} + N_{ci}^{kopt} T_{0i}$;
- ➢ initialisation du nombre de cycles : $N_{ci}^{k+1} = \frac{t_{1,k+1} - t_{0i}^{k+1}}{T_{0i}}$;
- ➢ initialisation des durées de «*on*» : le premier «*on*» a la même valeur que le dernier «*on*» de la fenêtre précédente, les autres «*on*» sont générés soit selon un modèle autorégressif d'ordre 1 soit selon une transformation logistique des durées de «*on*» de la fenêtre précédente.

Deuxième cas

Fin du dernier cycle du convecteur dans la fenêtre W_{k+1} et dernier «*on*» se terminant dans la fenêtre W_{k+1}

- ➢ initialisalisation de la date de début dans la fenêtre courante et du nombre de cycles : identique au premier cas ;
- ➢ initialisation des durées de «*on*» : identique au cas précédent sauf pour le cycle se situant sur les deux fenêtres.

Troisième cas : continuité de fonctionnement avec une probabilité donnée

Fin du dernier cycle du convecteur dans la fenêtre W_k

- ➢ Le convecteur continue à fonctionner dans la fenêtre W_{k+1} avec une probabilité r prédéfinie.

- ➢ Si la continuité de fonctionnement est assurée, l'initialisation des différents paramètres est celle décrite dans le paragrpahe précédent.

5.1.3 Procédure de décomposition du chauffage

La procédure de décomposition de la puissance appelée par le chauffage $y_{ch}(t)$ se résume par l'algorithme final décrit ci-après :

Algorithm 5.1 Algorithme stochastique de décomposition du chauffage

1. Estimer le nombre de convecteurs de l'installation ainsi que les amplitudes correspondantes.
2. Identifier les zones de présence du chauffage, en utilisant par exemple la méthode de détection décrite en (3.2.3).
3. Découper chaque zone de présence du chauffage en fenêtres d'analyse de durée caractéristique T_{CH}.
4. Sur chaque fenêtre d'analyse, appliquer jusqu'à convergence l'algorithme du recuit simulé pour décomposer le signal observé. Les perturbations des configurations et les contraintes de chaînage de signaux sont appliquées aux convecteurs tour à tour.

Remarque : paramètres de la mise en œuvre du recuit simulé

Nous avons choisi de faire décroître la température selon un schéma de décroissance exponentielle. À chaque itération n, la température est définie ainsi :

$$\mathscr{T}_{\mathrm{n}} = N\sigma^2,\ \sigma^2{=}\sigma_0^2 \exp\left(-c\frac{n}{NITERM}\right)$$

où N est le nombre d'échantillons du signal observé, $NITERM$ est le nombre maximum d'itérations, c est un paramètre qui permet de régler la décroissance de la température et σ_0^2 est la variance de l'erreur initiale (critère quadratique).

La définition de la température tient compte :

- ➢ de l'évolution du critère au fur et à mesure des itérations, donc de l'énergie du système à travers l'erreur ;
- ➢ et de la dimension des données traitées, de sorte que l'on tienne compte de l'erreur moyenne par échantillon.

5.1.4 Exemples de résultats

Nous allons présenter les premiers résultats obtenus par cette méthode, afin d'illustrer les performances de l'agorithme proposé dans le cas de mélanges simples en présence ou non de régime saturé. Des tests complémentaires sur des mélanges plus complexes et sur des durées de fonctionnement plus conséquentes afin d'évaluer l'intérêt de cette méthode feront l'objet de travaux futurs.

Sauf précision contraire, la durée caractéristique d'une fenêtre d'analyse est $T_{CH} = 10$ mn.

5.1.4.1 Convecteurs en régime de régulation

La figure (5.3) présente le résultat de la reconstruction d'un mélange de deux convecteurs de périodes égales à 40 s et 80 s. Les amplitudes estimées sont respectivement 670 W et 1190 W. On constate que les événements du mélanges sont correctement identifiés, bien que la configuration initiale est très éloignée de la solution recherchée.

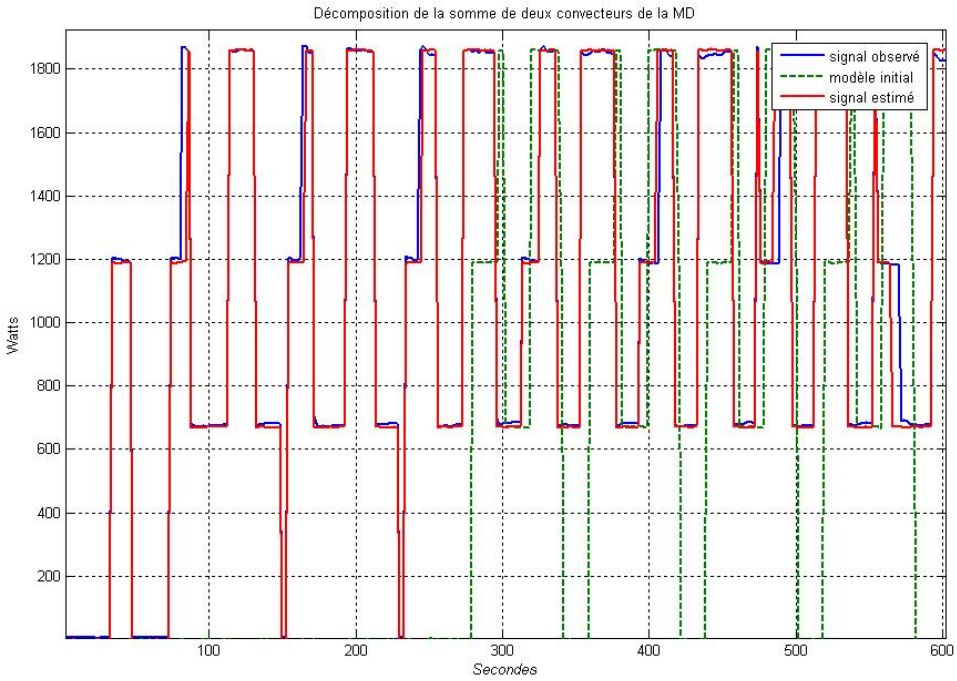


FIGURE 5.3: Décomposition d'un mélange de deux convecteurs - sans saturation

5.1.4.2 Convecteurs en régime de régulation et/ou de saturation

Afin d'étudier la capacité de l'algorithme à prendre en compte la saturation des convecteurs, nous l'avons d'abord testé sur différents signaux d'une durée de l'ordre de 30 mn présentant à la fois une phase de régulation et une phase de saturation. Un exemple de résultat est présenté sur la figure (5.4). Il s'agit d'un signal où le rapport cyclique évolue progressivement jusqu'à atteindre le régime de saturation. On constate que le signal de convecteur est reconstruit avec une bonne précision et que la phase de saturation est correctement identifiée.

Le résultat de reconstruction des composantes d'un mélange d'un convecteur en régime de régulation et d'un convecteur présentant une phase intermédiaire de saturation ainsi que les signaux de test sont présentés sur la figure (5.5). Globalement les paramètres du mélange sont correctement estimés.

Remarquons que ce résultat est obtenu en prenant en compte le chaînage des signaux sur les deux fenêtres d'une durée de 10 mn.

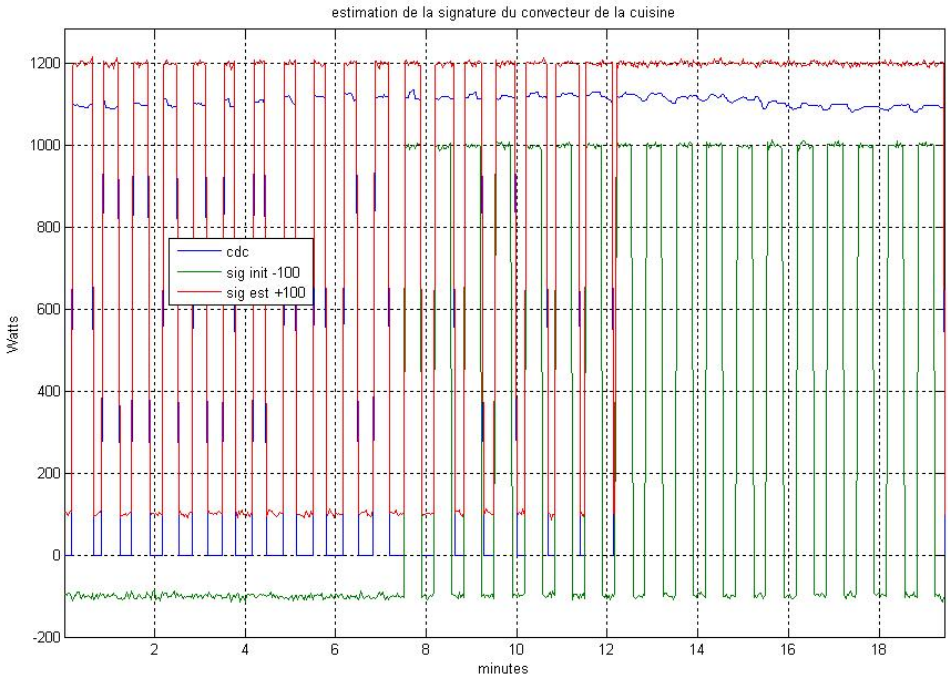


FIGURE 5.4: Décomposition d'un convecteur en présence d'une plage de saturation

Synthèse et discussion

Un modèle stochastique de la CdC du chauffage en vue de sa décompostion a été présenté. Ce modèle exploite les *a priori* sur cette classe de signaux, plus précisément le caractère cyclique des signaux et l'évolution lente du rapport cyclique au cours du temps.

Les paramètres du modèle évoluent dans un espace de dimension élevée. Le problème d'estimation des signaux élémentaires est ramené à la résolution d'un problème d'optimisation combinatoire. Une stratégie séquentielle mettant en œuvre un algorithme d'optimisation stochastique sur des fenêtres adjacentes a été mise en œuvre.

Les premiers tests réalisés pour démontrer la faisabilité de la méthode sur des cas particuliers pour étudier la robustesse de l'algorithme au phénomène de saturation des convecteurs ainsi que sur des mélanges simples sont satisfaisants.

Soulignons que la méthode de décomposition de courbe de charge du chauffage ne nécessite pas forcément une étape de détection des zones de mise en marche de cet usage. En effet, le modèle intègre ce cas à travers la valeur minimale nulle du rapport cyclique. Cependant l'utilisation d'un détecteur permet d'éviter l'exploration des zones où tous les convecteurs sont éteints.

Les éléments fournis dans cette section constituent le résultat d'une étude de faisabilité de cette classe de méthode dans le cas de la CdC du chauffage. Cette étude devra être complétée par des tests de validation sur une large base de signaux d'une part, et par une étude détaillée des transitions afin d'améliorer les propriétés de convergence de l'algorithme d'optimisation.

On peut également envisager d'enrichir le modèle en introduisant les modèles élémentaires d'autres usages, en particulier le froid alimentaire.

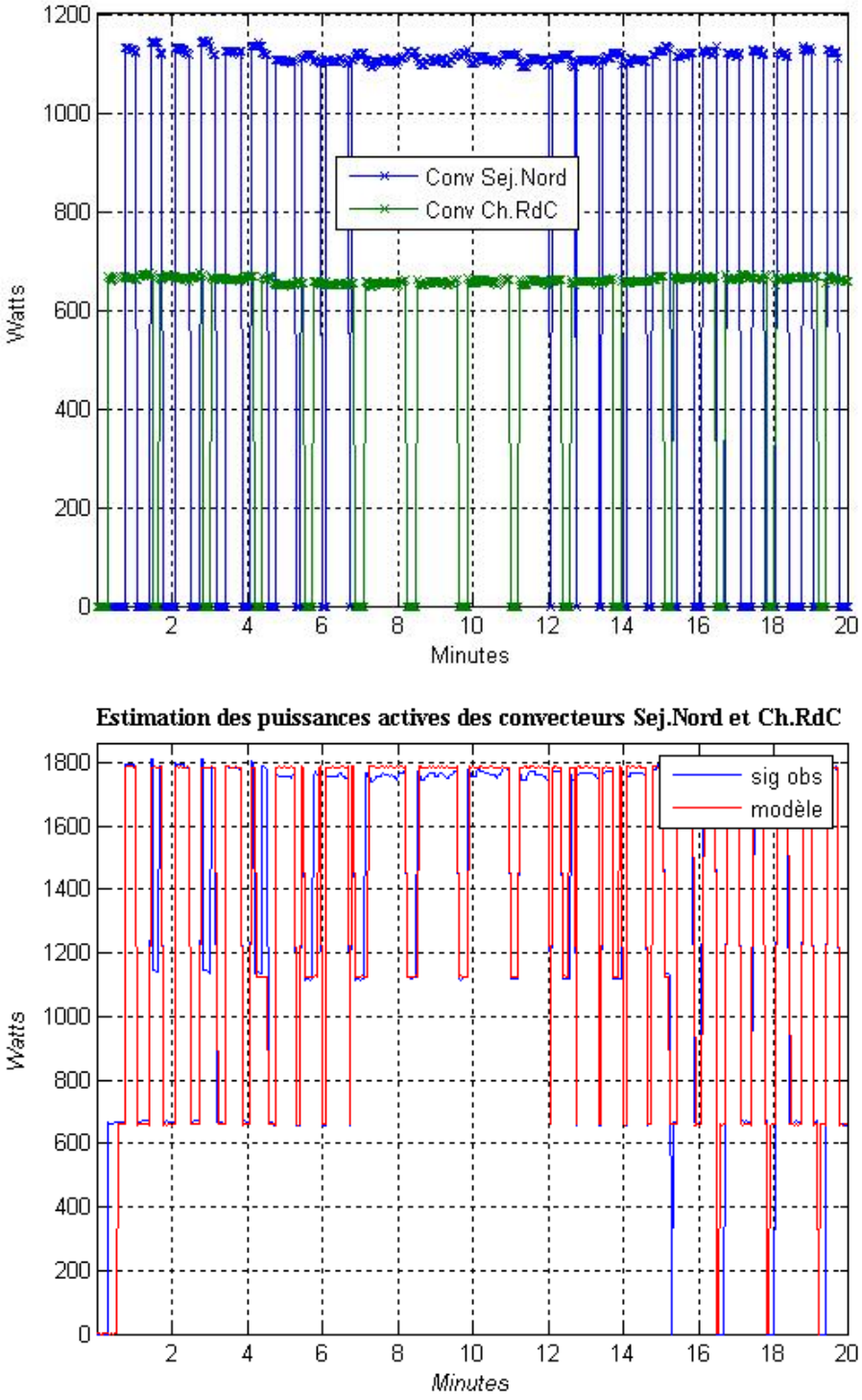


FIGURE 5.5: Décomposition d'un mélange de deux convecteurs en présence d'une zone de saturation

5.2 Approche du type « décodage » - application au froid alimentaire

La CdC du froid alimentaire est le foisonnement des puissances appelées par l'ensemble des compresseurs disponibles. Nous avons remarqué lors de l'analyse des CdC générales, l'existence d'intervalles de temps où le froid alimentaire est l'usage prépondérant. Dans le secteur résidentiel, on peut considérer les journées d'absence du client. Dans le cas d'un hypermarché, la période nocturne (ou pendant les jours fermés) semble être la plus appropriée pour caractériser à moindre coût et le plus finement possible les signaux des compresseurs.

Compte tenu des informations *a priori* sur les signaux analysés (forme, amplitude, scénario de fonctionnement des compresseurs, régulation d'un compresseur), une méthode de décomposition de la CdC du froid alimentaire est envisagée. Nous allons voir qu'une méthode dédiée par secteur est nécessaire, compte tenu des différences de fonctionnement dans chaque cas. Mais la méthodologie générale est commune aux deux cas d'application.

5.2.1 Principe

Les spécificités de l'usage froid alimentaire, dans le secteur résidentiel et dans le cas particulier d'un hypermarché, sont les suivantes :

- ➢ Modèle de fonctionnement d'un compresseur : « tout ou rien »
- ➢ Modèle de fonctionnement d'un ensemble de compresseurs d'une installation donnée : le séquencement des enclenchements et des déclenchements des différents compresseurs obéit à une régle pré-établie.
  - Dans le secteur résidentiel : Si un « *on* » et suivi par un « *off* », et que les amplitudes des événements vérifient le modèle Λ_A alors ces deux événements constituent un cycle de fonctionnement du même compresseur.
  - Dans le cas d'un hypermarché : le scénario de gestion du parc de compresseurs est défini de telle sorte que l'on minimise les sollicitations successives du même compresseur. Le cas d'application auquel nous nous sommes intéressé obéit à une règle de fonctionnement du type **FIFO** (le premier compresseur enclenché est le premier à déclencher).
- ➢ Forme du signal : les signaux observés présentent souvent des « pics » au démarrage du compresseur, la durée de cette phase est de l'ordre d'une seconde.

On constate qu'il y a un a priori fort sur le séquencement des événements d'un seul compresseur ou encore sur le séquencement des enclenchements et des déclenchements de l'ensemble des compresseurs présents dans l'installation étudiée. Nous proposons d'exploiter cet *a priori* pour décoder la série des événements de la CdC étudiée afin de reconstruire les signaux des compresseurs.

La démarche proposée est décrite sur la figure (5.6). Elle consiste à ramener le problème d'extraction des signaux élémentaires de la CdC à un problème de classification et d'appariement des événements à partir du signal de mélange $y(t)$. La démarche consiste en deux phases :

Une phase d'apprentissage (hors-ligne)

Cette phase permet d'apprendre les modèles Λ_A et Λ_S définissant respectivement la variation d'amplitude entre un « *on* » et le « *off* » correspondant et le séquencement des événement (règles

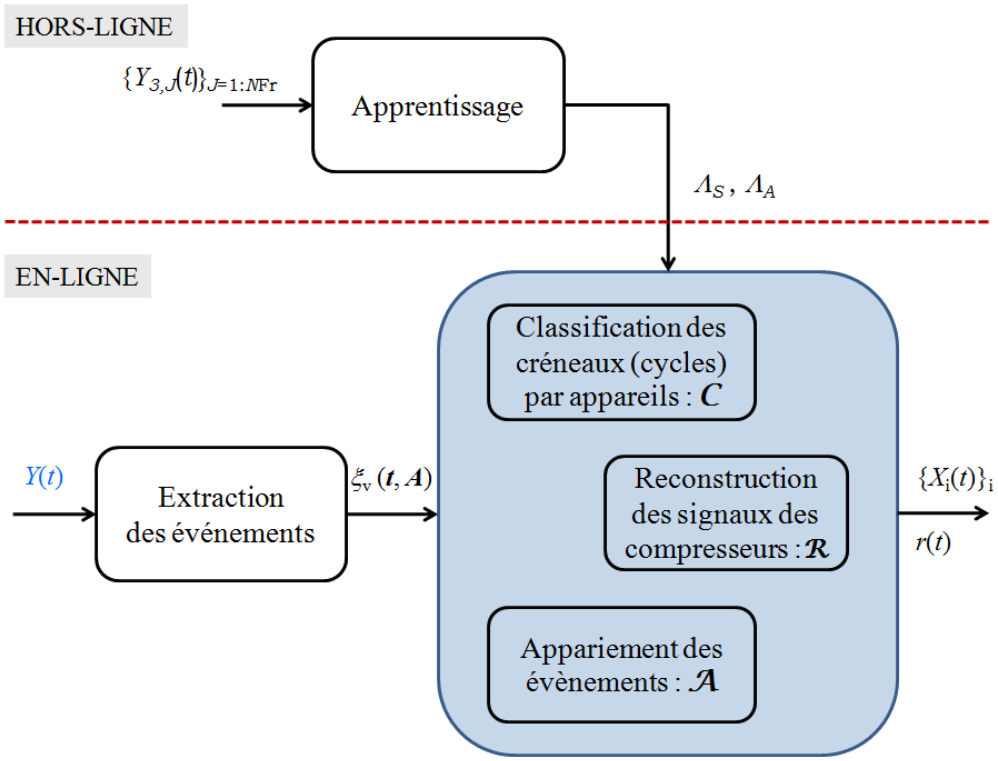


FIGURE 5.6: Schéma général de la méthode d'extraction des signaux de compresseurs

entre événements d'un compresseur et de plusieurs compresseurs). Les signaux d'entrée sont des exemples de signaux de NFr compresseurs, acquis pendant J jours.

Décomposition du mélange (en ligne)

Etant données les amplitudes des signaux de compresseurs et les règles qui régissent le séquencement des événements, l'objectif est de décomposer un signal de mélange $y(t)$. La première étape consiste à extraire la série des événements $\xi_v(t, A)$ *significatifs* à partir du signal observé. Cette étape est identique quelle que soit le cas d'application étudié, et consiste en deux étapes :

- ➢ un filtrage des « pics » : filtrage morphologique (ouverture et fermeture) dont la durée caractéristique est $T_{fm} = 3s$.
- ➢ détermination de la série des événements significatifs :
  - Soit $y'(t) = y(t + t_{v0}) - y(t)$, où t_{v0} est une fenêtre temporelle de calcul des accroissements du mélange (en pratique $t_{v0} = T_e$).
  - les événements significatifs (d'amplitude minimale en valeur absolue A_{3min}) sont obtenus ainsi :

$$\xi_v(t, A) = \left\{(t_i, A_i),\ tels\ que\ y'(t_i) = A_i\ et\ |A_i| \geq A_{3min}\right\}$$

Une fois cette série des événements obtenue, il reste à apparier chaque enclenchement au déclenchement correspondant (réalisé par l'opérateur $\mathscr{A}$), à classifier l'ensemble des événements par appareils (réalisé par l'opérateur $\mathscr{C}$) et de reconstruire les signaux de chaque compresseur (réalisé par l'opérateur $\mathscr{R}$). L'ordre de ces opérations est conditionné par la nature des signaux élémentaires, plus précisément par le modèle régissant l'ordre des événements Λ_S.

Nous montrerons que deux approches sont envisageables pour mettre en œuvre cette méthodologie : une approche conjointe où l'appariement et la reconnaissance des événements sont combinés, et une approche séquentielle où l'on appliquera les opérateurs décrits séparément selon un ordre précis.

5.2.2 Décomposition de la CdC du froid alimentaire dans le secteur résidentiel

Afin d'extraire les signaux de compresseurs, nous proposons une procédure itérative pour extraire les couples (t_i, A_i) vérifiant certaines propriétés (provenant d'informations *a priori*), les durées des cycles de fonctionnement étant la grandeur utilisée par la suite pour séparer les différents compresseurs.

5.2.2.1 Informations *a priori*

La puissance moyenne appelée par le froid alimentaire chez un client résidentiel est le foisonnement des puissances appelées par les différents compresseurs.

L'analyse des données du client considéré montrent que :

- ➢ les compresseurs n'ont pas nécessairement la même amplitude ;
- ➢ pour un compresseur donné, l'amplitude en régime établi est quasiment constante ;
- ➢ un compresseur fonctionne en « tout ou rien », ainsi un *on* est nécessairement suivi par un *off* ;
- ➢ les événements de la CdC obéissent globalement à une règle du type **LIFO** (Last In First Out). Ainsi, un événement du type *on* suivi par un événement du type *off* sont nécessairement générés par le même compresseur. Nous les appellerons « *on-off* » *isolés* ;
- ➢ les signaux se différencient par la durée des cycles de fonctionnement.

5.2.2.2 Prétraitements des signaux et extraction des événements pertinents

La figure (5.7) illustre l'étape de filtrage des pics des compresseurs. La détection des événements pertinents est obtenue par seuillage puis filtrage morphologique des accroissements du signal étudié.

5.2.2.3 Algorithme de décomposition de la CdC du froid alimentaire

Nous proposons une méthode qui consiste en deux étapes.

❶ Estimation des cycles de fonctionnement des compresseurs

Un algorithme itératif fondé sur la loi de séquencement des événements permet d'extraire et d'isoler les cycles de fonctionnement des compresseurs. L'architecture de l'algorithme mis en œuvre est présentée sur le schéma bloc (5.2).

A chaque itération, la série des événements significatifs est estimée puis les couples de « on-off » isolés sont extraits. Etant donnée la liste des événements extraits, une estimation de la puissance moyenne de la série courante des signaux est évaluée. Le signal estimé est soustrait à

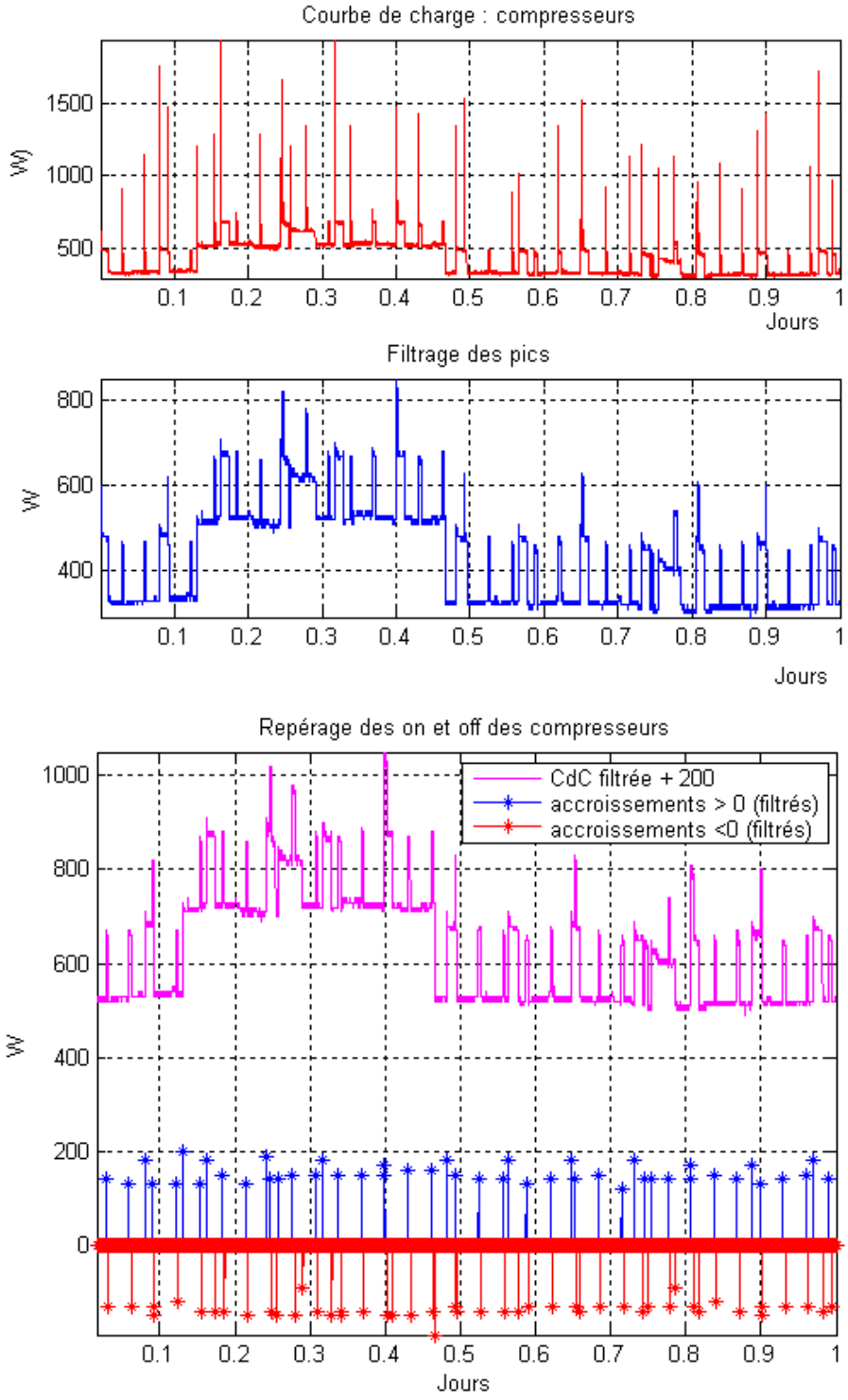
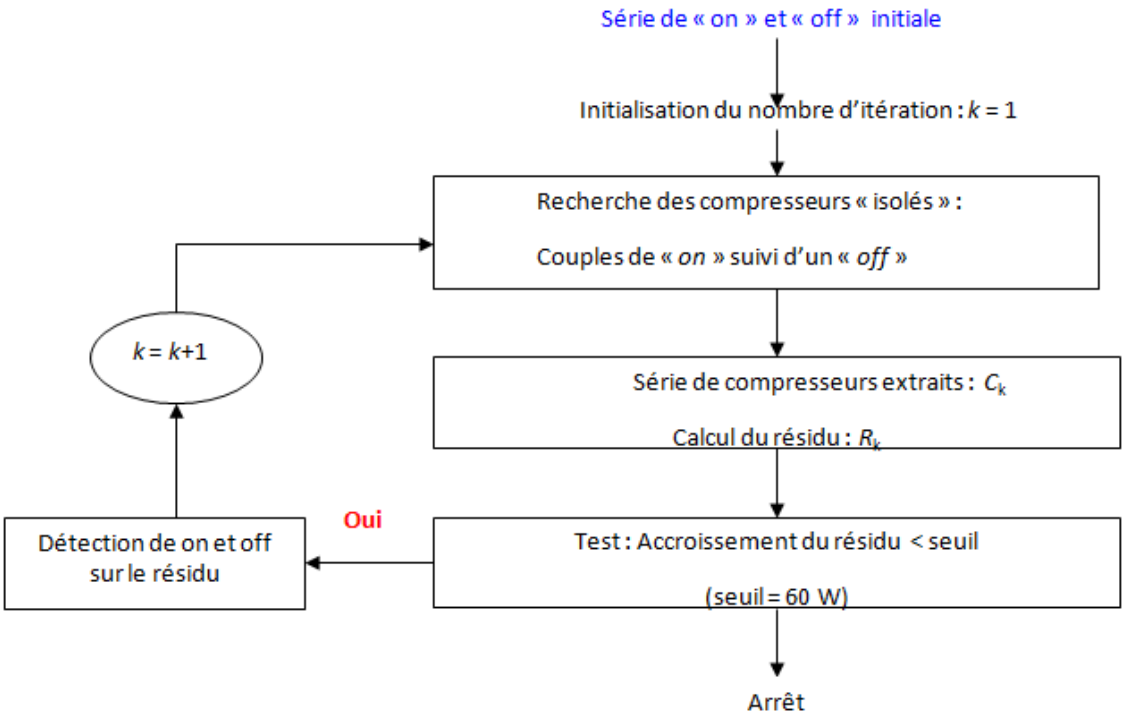


FIGURE 5.7: Extraction des événements pertinents des compresseurs à partir de la CdC du froid alimentaire ($T_e = 2$ s)

Algorithm 5.2 Algorithme itératif d'extraction des compresseurs



la CdC courante. Ces étapes sont appliquées itérativement tant que l'ensemble des événements significatifs n'est pas vide.

Les figures (5.8), (5.9) et (5.10) présentent uniquement **un zoom** sur le signal estimé et le résidu obtenu pour les 3 itérations de l'algorithme afin de mieux mettre en évidence la qualité du résultat obtenu.

Remarquons que les signaux extraits (cycles de fonctionnement) à chaque itération n'ont pas nécessairement la même amplitude et présentent des durées de fonctionnement variables. Il ne s'agit pas de signaux issus du même compresseur. Une étape de classification des signaux est alors nécessaire.

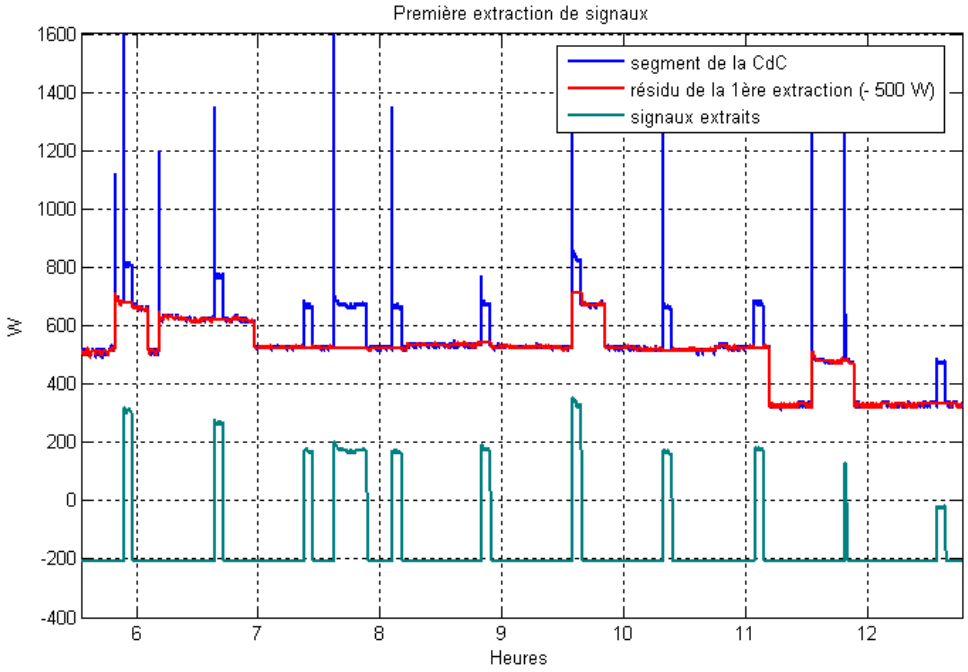


FIGURE 5.8: Zoom sur le signal estimé et sur le résidu- Première itération

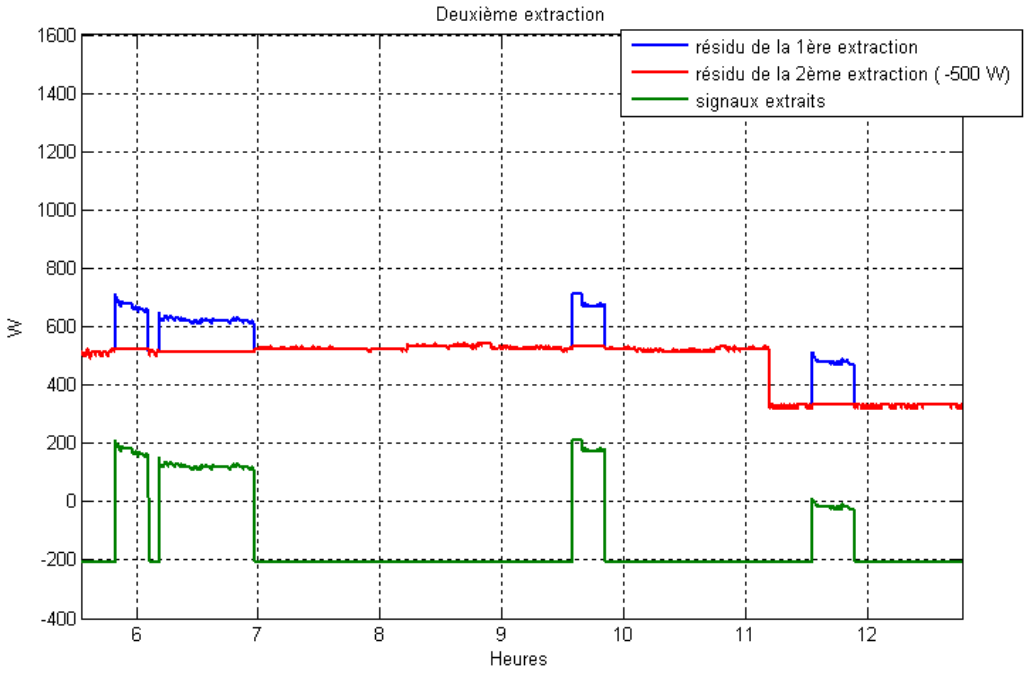


FIGURE 5.9: Zoom sur le signal estimé et sur le résidu- Deuxième itération

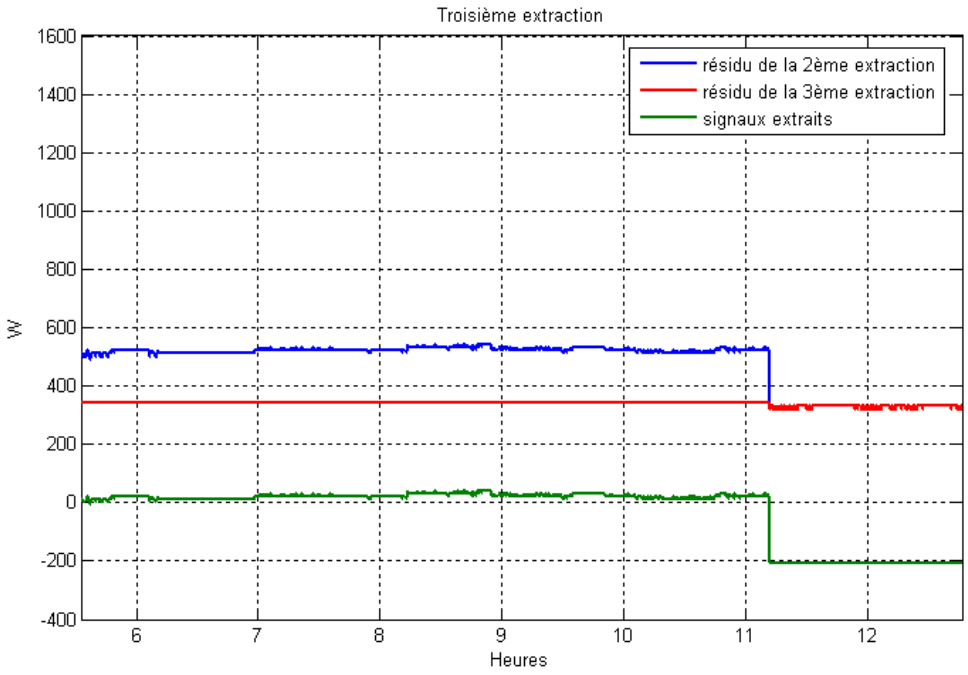


FIGURE 5.10: Zoom sur le signal estimé et sur le résidu - Troisième itération

❷ Classification des signaux

Les signaux sont classés selon la durée du cycle de fonctionnement. La figure (5.11) illustre les histogrammes de ces paramètres obtenus respectivement pour les signaux de la première et de la deuxième itération regroupés puis pour l'unique cycle extrait à la troisième itération. Nous constatons qu'il y a trois classes de signaux (au regard du critère choisi). Le compresseur dont la durée « *on-off* » est d'environ 8 h est celui d'une pompe à chaleur. Celui dont la durée caractéristique est de l'ordre de 45 mn est a priori le congélateur. Enfin, les signaux dont la durée « *on-off* » varie entre 15 mn et 20 mn sont générés par le compresseur du réfrigérateur.

Remarques

➢ L'analyse des résultats (extraction des signaux et classification) a été effectuée de manière aveugle. Des informations collectées auprès du client (sur l'installation et la fréquence d'utilisation des appareils) ont permis de valider cette approche.

➢ La méthode développée est automatique. Une extension à la CdC du froid alimentaire d'un hypermarché est envisageable, sous réserve de bien prendre en compte les *a priori* spécifiques à ce secteur.

5.2.3 Décomposition de la CdC du froid alimentaire d'un hypermarché

L'objectif de la caractérisation du froid alimentaire est ici la reconstruction du scénario. Ce scénario permet d'estimer le taux horaire de fonctionnement de chaque compresseur τ_H : c'est le ratio entre la durée de fonctionnement effectif d'un appareil D_{uni} (ici un compresseur) et la durée totale D_{tot} de fonctionnement de l'usage correspondant à l'appareil étudié (ici le froid alimentaire) décrit par l'équation (5.6).

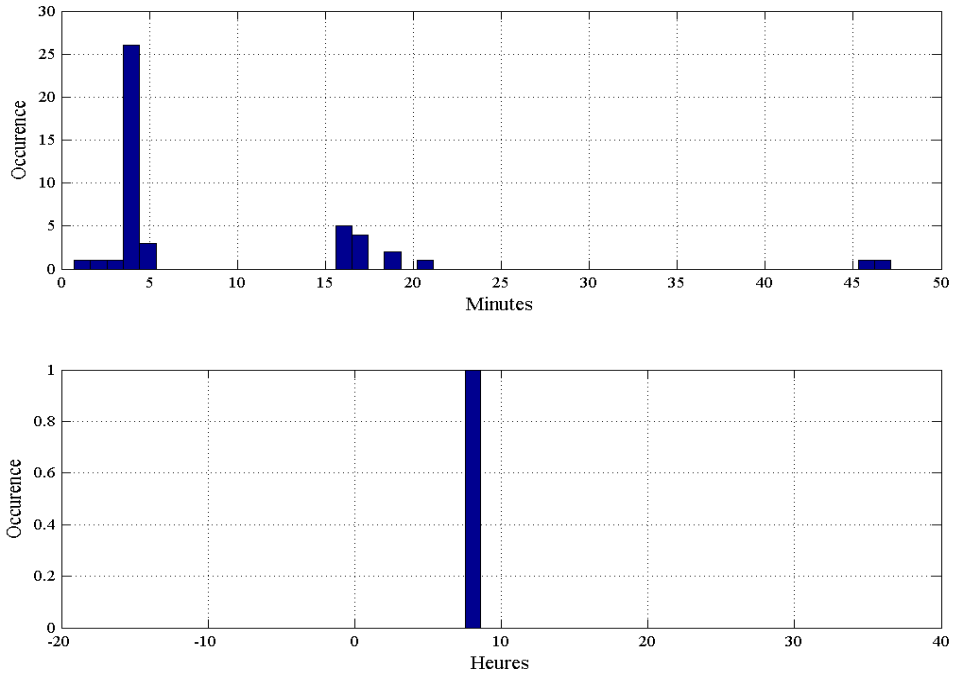


FIGURE 5.11: Durées des cycles de fonctionnement des signaux extraits

$$\tau_H = \frac{D_{uni}}{D_{tot}} \tag{5.6}$$

L'analyse de cette grandeur contribue à l'identification des sources potentielles d'optimisation de la gestion des appareils. Une répartition très hétérogène entre appareils est significative d'une mauvaise gestion. En effet, certains compresseurs sont très sollicités, ce qui augmente le risque de défaillance de ces appareils.

Nous nous intéressons également à l'estimation des caractéristiques dimensionnantes des compresseurs telles que les amplitudes moyennes des signaux en régime établi. Ces paramètres permettent d'estimer la puissance nominale de l'appareil.

Remarquons que l'estimation du nombre de compresseurs, des scénarios d'usage et des amplitudes moyennes permettent d'estimer les CdCs élémentaires des différents compresseurs.

5.2.3.1 Informations *a priori*

Les connaissances a priori fournies par un spécialiste de l'audit de ces installations sont les suivantes :

- ➢ Les compresseurs étudiés nominale supérieure à 1000 W et la même puissance nominale. Mais, l'installation contient également d'autres compresseurs de puissance nominale plus faible.
- ➢ Le séquencement des enclenchements et déclenchements des compresseurs obéit à une régle du type LIFO.

➢ Le froid alimentaire est l'usage prépondérant sur la plage horaire [00 h, 5 h] et [21 h, 00 h[, les principaux autres usages étant l'éclairage (mais de faible consommation la nuit) et la ventilation.

Remarque

Le chauffage et la climatisation n'étaient pas en marche pendant la durée des essais (3 jours).

5.2.3.2 Estimation des événements des compresseurs à partir de la CdC générale

Pour atteindre cet objectif, nous utilisons l'algorithme de détection des événements développé pour le secteur résidentiel (à la définition du seuil de détection (1000 W) près).

Cet algorithme est appliqué à la puissance active et à la puissance réactive afin de minimiser les erreurs de détection. Les événements retenus doivent vérifier la condition suivante : un événement n'est retenu que s'il est détecté simultanément (dans une fenêtre de tolérance d'une durée d'une seconde) sur la puissance active et sur la puissance réactive.

L'amplitude moyenne de fonctionnement (active ou réactive) est approchée par la moyenne de la variation du signal étudié autour des événements détectés. La fenêtre de calcul de la variation d'amplitude a une largeur de 30 s centrée autour de l'événement considéré.

Un exemple de résultat obtenu à partir de l'analyse de la CdC générale quotidienne tronquée sur la plage horaire [00 h, 5 h] est décrit sur la figure (5.12). On obtient une estimation des amplitudes et de l'incertitude (ici écart type de l'erreur). Les résultats obtenus ont été jugés satisfaisants par l'auditeur du site par rapport aux informations dont il dispose sur le parc étudié et des parcs similaires audités par ailleurs.

Cette estimation est d'autant plus intéressante pour l'auditeur du site qu'elle ne nécessite qu'une journée (voire une demi-journée) de données uniquement, ce qui est compatible avec les durées d'audition du site et la durée d'acquisition tolérée par le client.

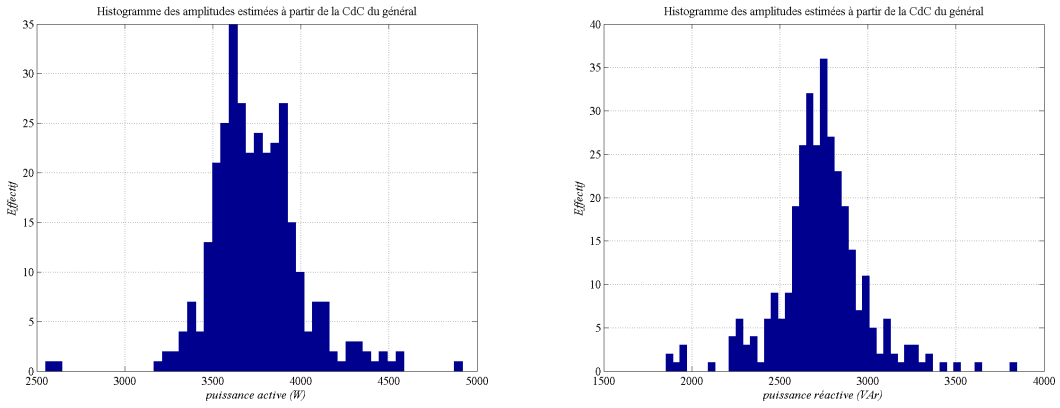


FIGURE 5.12: Estimation de l'amplitude moyenne des compresseurs à partir de la CdC générale d'un hypermarché

5.2.3.3 Reconstruction des scénarios de gestion des compresseurs

Rappelons que le nombre de compresseurs est inconnu. Nous proposons un algorithme séquentiel fondé sur la logique combinatoire et qui permet d'étiqueter les événements et d'estimer le nombre de compresseurs conjointement.

La figure (5.13) présente un exemple de résultat obtenu de manière aveugle sur la CdC générale tronquée sur le domaine d'étude.

Le nombre de compresseurs estimé est égal à 3. Le nombre effectif des composants étant inconnu, nous proposons de comparer l'estimation obtenue à partir de la CdC générale avec celle issue de l'application du même algorithme à la CdC du froid alimentaire seul observée sur une journée complète. Un zoom sur le scénario de référence sur le domaine d'étude de la CdC générale est présenté sur la figure (5.14). Nous constatons l'existence d'un quatrième compresseur qui s'enclenche vers 2 h.

Cette observation met en évidence les limites de l'algorithme du décodage des événements liés à la grande combinatoire : si le nombre de compresseurs est inconnu, la régle LIFO du séquencement des événements n'est pas toujours suffisante pour résoudre le problème combinatoire considéré.

Après cette expérimentation sur les données réelles, nous introduisons une contrainte supplémentaire (fournie par l'auditeur du site) : la solution recherchée est celle qui maximise le nombre de composants tout en respectant la règle LIFO de gestion des appareils.

Le nombre de compresseurs estimé avec l'algorithme amélioré appliqué à la CdC générale est égal à 4. Un autre critère d'évaluation des performances de l'algorithme consiste à comparer les taux horaires de fonctionnement évalués à partir les scénarios de référence (reconstruits à partir de la CdC du froid alimentaire seul) τ_H^{ref} et ceux évalués à partir des scénarios estimés à partir de la CdC générale τ_H^{est}. Précisons que la durée totale D_{tot} est calculée à partir des signaux de référence et que les grandeurs comparées sont celles évaluées sur le domaine temporel d'étude considéré ([00 h, 5 h]) .

Les résultats résumés sur le tableau (5.1) montrent :

- ➢ une bonne adéquation entre les valeurs estimées à partir la CdC générale et les valeurs de référence ;
- ➢ un taux de perte (durée de fonctionnement de compresseurs non détectée) très faible (2%).

	Compresseur n°1	**Compresseur n°2**	**Compresseur n°3**	**Compresseur n°4**
τ_H^{ref} (total : 100%)	27	27	27	17
τ_H^{est} (total : 98%)	28	26	27	19

TABLE 5.1: Résultats d'estimation du taux horaire de fonctionnement

Ces résultats sont très satisfaisants, mais une conclusion générale sur les performances de l'algorithme nécessite une phase d'évaluation sur une base de signaux significative.

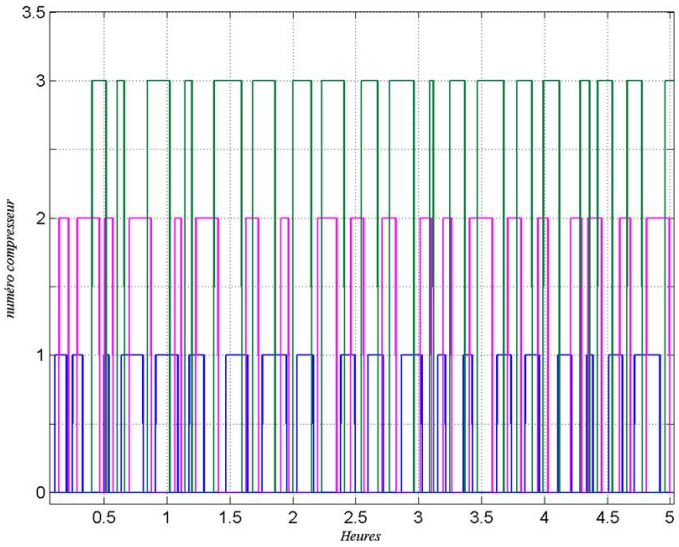


FIGURE 5.13: Reconstruction de scénario de fonctionnement d'un nombre inconnu de compresseurs à partir de la CdC générale d'un hypermarché

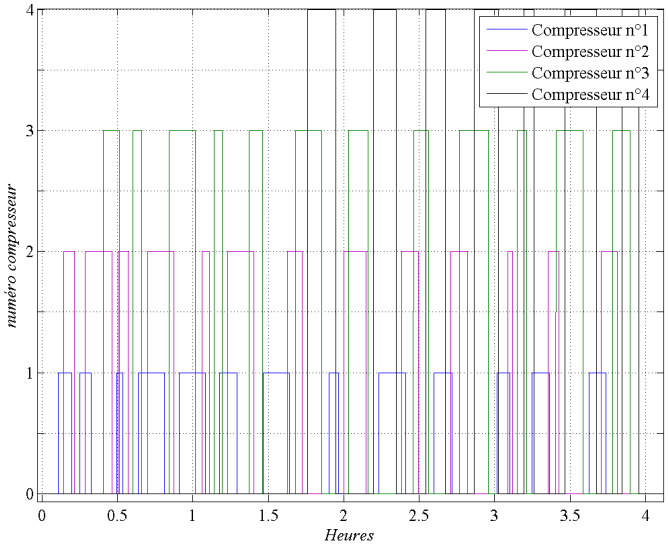


FIGURE 5.14: Reconstruction de scénario de fonctionnement d'un nombre inconnu de compresseurs à partir de la CdC du froid alimentaire seul

Conclusions sur les méthodes de décodage des événements

L'estimation des signaux élémentaires de la CdC du froid alimentaire est traitée sous l'angle du décodage de la série d'événements de la CdC. Des informations contextuelles sur la période où cet usage est prépondérant (période d'absence du client par exemple dans le secteur résidentiel, et la nuit dans le cas des hypermarchés) nous permettent de traiter ce problème d'estimation des sources à partir d'une CdC quasi-mono-usage. Dans les deux cas traités, le problème de reconstruction des signaux est décomposé en plusieurs étapes à savoir la détection des événements, la classification des événements par appareil et l'estimation de leur amplitude moyenne. Ces informations permettent de reconstruire les signaux élémentaires. Les algorithmes sont fondés sur des informations *a priori* sur le mode de fonctionnement (séquencement des mises en marche et d'arrêts des appareils) ainsi que sur les amplitudes des signaux. Les premiers résultats obtenus confirment l'intérêt de cette approche pour la caractérisation du froid alimentaire à partir de la CdC générale. Une évaluation sur une base significative de signaux reste à réaliser.

Synthèse

L'objectif de ce chapitre est d'explorer les possibilités et méthodes de décomposition de la CdC d'une installation donnée. Afin de s'affranchir de la grande variabilité des signaux d'une CdC quelconque, nous proposons de se restreindre à des cas particuliers, à savoir les CdC quasi-mono-usage (un usage prépondérant).

Dans ce cadre, une méthode probabiliste est proposée pour décomposer la CdC du chauffage dont les premiers résultats sont encourageants. En effet, l'algorithme développé permet de prendre en compte des événements simultanés de différents appareils ainsi que la saturation des convecteurs.

Deux algorithmes fondés sur l'extraction des événements de la CdC et la prise en compte des informations *a priori* sur le séquencement du fonctionnement des compresseurs permettent de décomposer :

- la CdC résidentielle des compresseurs seuls (en absence du client par exemple) ;
- et la CdC générale d'un hypermarché pendant la période de nuit.

Dans les deux cas, les premiers résultats sont satisfaisants. Mais une évaluation à partir d'une base de signaux plus représentatives reste à faire.

Conclusion générale

Dans ce manuscrit, nous considérons le problème de caractérisation aveugle de la courbe de charge (CdC) générale de manière entièrement non-intrusive. Trois sous-problématiques sont sous-jacentes à la *caractérisation*, à savoir la détection, la classification et l'estimation (estimation de paramètres ou extraction de signaux) (chapitre 1) . Cette problématique s'inscrit dans le cadre générique de l'extraction automatique du contenu informationnel des sources d'un mélange mono-valué à des fin de prise de décision (supervisée ou non-supervisée). La CdC est un signal non-stationnaire qui constitue une observation indirecte des objects recherchés (énergie quotidienne par usage, domaine de fonctionnement, nombre de composantes, état de santé des machines, etc.). La caractérisation de cette classe de signaux de mélange est traitée par exemple dans le cas de signaux sonores, où différentes approches sont proposées pour séparer les sources à partir d'une observation unique [BGB01, BB03, GL]. L'analyse de ces travaux (chapitre E) met en évidence l'importance de la connaissance complémentaire apportée ou encore le choix d'une représentation adaptée aux sources pour rendre le problème soluble. Un autre problème analogue est celui de l'extraction automatique d'objets à partir d'une image [Ort95], qui se situe dans le cadre générique des problèmes inverses. Les solutions proposées montrent l'intérêt de la simulation bayésienne pour cette classe de problèmes. L'analyse de notre besoin et de ces approches nous ont conduit à adopter une démarche qui se décline en deux phases. Nous verrons que l'étape d'évaluation des méthodes est intégrée à la deuxième phase.

a) Extraction de connaissances a priori à partir des signaux élémentaires Une analyse détaillée de signaux (chapitre 2) issus de différentes campagnes de mesure a permis d'extraire des propriétés des usages (la morphologie, le contenu fréquentiel, les lois d'évolution de paramètres, etc.). Cette phase a également permis d'identifier des représentations adaptées à la reconnaissance de certains usages, notamment le spectrogramme et la densité des événements pour le chauffage électrique.

b) Définition d'une méthodologie générale pour la caractérisation de la CdC La méthodolgie générale de la caractérisation de la CdC s'appuie sur la classification des estimées recherchées en cinq classes : les caractéristiques physiques des usages, l'énergie consommée par usage, le support temporel de chaque usage, la série des événements par usage et enfin les signaux temporels. Quant au choix de(s) approches théoriques, orienté par les spécificités du problème, nous avons également opté pour une décomposition du problème industriel en trois cas d'applications. Ce choix permet de sélectionner l'approche théorique la plus appropriée à chaque classes de signaux et de répondre au besoin industriel. Dans la suite, nous décrivons les

méthodes développées et performances obtenus pour chaque « sous-problème » en fonction des objectifs annoncés.

Caractérisation d'une CdC générale quelconque Nous nous intéressons tout particulièrement aux clients ayant du chauffage électrique par convecteurs. La diversité des CdCs et la complexité apporté par le foisonnement important des signaux en présence du chauffage nous ont amené à restreindre les objectifs de la caractérisation à la segmentation de la CdC et à l'estimation des paramètres des usages (caractéristiques physiques, énergie consommée par usage).

La segmentation automatique est traitée dans le chapitre 3. Une approche « classe / non-classe » est présentée et déclinée dans le cas particulier du secteur résidentiel. Nous avons développé deux méthodes de segmentation du chauffage. Elles sont respectivement fondées sur l'interprétation du spectrogramme et sur l'étude de la densité des événements. Un détecteur résultant d'une adaptation du test de Page-Hinkley (en ajoutant des informations *a priori*) permet de détecter les cycles de fonctionnement du chauffe-eau. L'étude de performances à partir de données réelles met en évidence la robustesse de la segmentation chauffage / autres usages. La détection du chauffe-eau reste à améliorer.

Le problème d'estimation de l'énergie consommée, de l'amplitude et du nombre de composants est considéré dans le chapitre 4. Les résultats obtenus avec les algorithmes de cartographie de l'énergie totale consommée à partir des CdCs réelles sont satisfaisants et répondent globalement aux attentes de l'entreprise. Néanmoins certaines améliorations sont nécessaires, notamment pour traiter le cas où les convecteurs fonctionnent en régime saturé pendant une durée significative (quelques heures).

Décomposition d'une CdC quasi-mono-usage Isoler ce cas d'application particulier nous permet d'envisager des méthodes de reconstruction des sources du mélange dans des conditions réelles. Ce qui est difficilement envisageable dans le premier cas d'application.

Deux méthodes de reconstruction de signaux sont présentées dans le chapitre 5. La première s'appuie sur un algorithme de décodage des événements de l'usage étudié et utilise un modèle d'amplitude des signaux. Une application à la CdC du froid alimentaire (résidentiel et tertiaire) met en évidence l'intérêt de cette méthode et la robustesse des algorithmes développés. La deuxième est une méthode générative, où les paramètres du modèle sont estimés par un algorithme du recuit simulé. Une application au chauffage électrique constitue une illustration de la faisabilité de la méthode, qu'il faudra compléter par des tests sur des signaux réels.

En résumé

Les contributions de cette thèse portent sur deux niveaux. Le premier niveau concerne les méthodes et modèles proposés par adapatation et amélioration des approches existantes. Le second niveau concerne l'aspect applicatif des solutions proposées, au sens les solutions que l'entreprise envisage d'industrialiser à terme.

Les contributions méthodologiques Les principales contributions méthodologiques sont les suivantes :

- ➢ Extraction de connaissances *a priori* sur les usages électriques et définition de domaines transformés dédiés aux usages éléctriques ;

- ➢ Nouvelle méthode d'interprétation d'une représentation temps-fréquence fondée sur l'utilisation conjointe d'un détecteur de non-stationnarités dans le plan temps-fréquence et d'une méthode d'extraction des motifs fréquents ;
- ➢ Une nouvelle approche de séparation de sources mono-capteur (sans apprentissage à base d'exemples et dans le domaine d'observation des signaux) fondée sur deux modèles stochastiques de la CdC et des événements de la CdC.

Les contributions « applicatives » D'un point de vue applicatif, certains algorithmes développés dans le cadre de la thèse contribuent à répondre au besoin industriel court et moyen termes. Il s'agit principalement de :

- ➢ la segmentation de la CdC résidentielle quotidienne ;
- ➢ la cartographie de l'énergie quotidienne par usages dans le secteur résidentiel ;
- ➢ la reconstruction des scénarios de fonctionnement des compresseurs dans les hypermarchés ;
- ➢ l'estimation des caractéristiques d'usages (nombre de convecteurs, nombre de compresseurs, taux horaire de fonctionnement des compresseurs d'un hypermarché).

Ces résultats constituent des composantes d'une offre énergétique telle qu'une analyse détaillée de la consommation d'un client particulier ou encore un appui à l'audit de sites tertiaires en vue de l'optimisation de la gestion des appareils ou de la détection d'anomalies.

Les algorithmes d'estimation de l'énergie quotidienne feront l'objet d'une évaluation à grande échelle en 2010.

Perspectives

Ce chapitre décrit les perspectives méthodologiques, algorithmiques et d'évaluation dans la suite des méthodes proposées dans le cadre de cette thèse. Nous présentons brièvement deux études initiées en pespectives de la thèse : la première s'intéresse à l'apport de la modélisation par processus stochastiques marqués pour la décomposition de la CdC, la deuxième s'intéresse à la caractérisation des usages électriques tel qu'envisagée dans notre thèse mais à partir de signaux électriques différents (intensité et tension de courant échantillonnées à quelques kHz).

Amélioration de la modélisation

Dans le cadre des méthodes du type filtrage, nous proposons d'améliorer le modèle de la consommation quotidienne du chauffage en adoptant une démarche du type régression de modèle. Rappelons que la définition du modèle décrit dans ce mémoire est fondée sur des constatations qualitatives par manque de données au moment où le modèle a été conçu. Nous proposons d'identifier la fonction reliant la CdC du chauffage, les signaux de régulation rapide et d'autres paramètres de la CdC générale en utilisant les données d'apprentissage où la CdC chauffage de référence est parfaitement connue. Rappelons que le modèle proposé dans le cadre de la thèse est empirique et avait initialement été proposé sans connaître les signaux du mélange.

Caractérisation des usages dans le domaine temps-fréquence

Nous avons présenté une méthode de segmentation de la CdC dans le domaine temps-fréquence en vue de la reconnaissance des usages. Le choix de la représentation est orienté par les propriétés des signaux à caractériser. Il serait intéressant d'étudier une généralisation de cette approche d'interprétation des usages à partir d'une ou de plusieurs représentations temps-fréquence. Nous avons investigué une méthode de détection et d'estimation des usages fondée sur la méthode des paquets d'ondelettes. Les premiers résultats montrent que cette approche est prometteuse. En effet, comme les usages étudiés se différencient par leurs constantes temporelles (la durée de cycle de fonctionnement), une interprétation des signaux fondée des représentations temps-échelle est a priori pertinente. Le choix des paquets d'ondelette offrent plus de possiblités en termes de détection et de classification du contenu de la CdC. Cette approche s'insprise d'une étude portant sur la segmentation et la classification des signaux EMG [Che06].

Evaluation des méthodes de caractérisation de la CdC

La problématique d'évaluation des méthodes n'a pas été traitée à part entière dans le cadre de la thèse. Cependant, des critères d'évaluation sont proposés pour chaque méthode. Ces critères ont été utilisés au mieux compte tenu de la qualité des données disponibles. De manière générale, il reste à évaluer les méthodes sur une base de données plus significative. Des données issues d'une campagne de mesure « grande échelle » (une cinquantaine de clients) seront disponibles au second semestre 2010 afin d'évaluer les algorithmes d'estimation de l'énergie consommée par le chauffage et celle consommée par l'ECS. Ces données permettront également d'évaluer les performances des deux détecteurs mis en œuvre pour les deux usages étudiés. Il reste également à évaluer les algorithmes d'estimation des signaux temporels des compresseurs à partir de CdC quasi-mono-usage. Les autres algorithmes nécessitent quelques améliorations et extensions avant d'envisager une telle démarche d'évaluation globale.

Vers une méthode de reconstruction des signaux d'une CdC quelconque

Cette section présente une méthode que nous proposons en perspective de cette étude pour décomposer une CdC quelconque. Il s'agit d'une étude de faisabilyé. Nous esquissons la démarche générale, la formalisation du problème et une illustration des premiers résultats obtenus.

Ici, nous considérons le problème de la séparation aveugle des signaux du mélange observé (CdC) sous l'angle de la classification des événements électriques de la CdC globale par classe d'appareils. L'analyse des signaux a mis en évidence des propriétés morphologiques et temporelles communes entre certains appareils n'appartenant pas nécessairement au même usage. Nous définissons alors de nouvelles classes de signaux du mélange en fonction de ces propriétés. La reconstruction de ces classes de signaux est décomposée en une étape de classification et une étape d'appariement des événements observés. Etant donnée la dimension de l'espace des paramètres recherchés et le caractère combinatoire du problème, nous formalisons le problème initial de séparation de sources dans le cadre de la simulation bayésienne. La méthode proposée est fondée sur une modélisation stochastique de la série des événements observés. La solution au sens du maximum *a posteriori* est obtenue en utilisant un échantillonneur du type MCMC, plus précisément l'échantillonneur de Metropolis-Hastings. Nous commençons ce chapitre par une présentation du problème de séparation des composantes de la CdC sous l'angle de la classification et de l'appariement conjoints des événements de classes de signaux que nous avons introduits.

1. Position du problème

1.1. Définition des classes d'événements

Soit $y(t)$ la puissance moyenne mesurée à une fréquence d'échantillonnage T_e chez un client résidentiel. Le problème de séparation aveugle des signaux du mélange consiste à estimer les sources $x_i(t)$ du mélange. L'analyse des signaux a permis d'identifier certains paramètres discriminants entre appareils tels que la symétrie ou non entre un enclenchement et le déclenche-

ment correspondant, les lois des durées temporelles ainsi que l'amplitude moyenne des signaux. Afin de prendre en compte ces connaissances, la représentation naturelle de la CdC est la série des événements électriques. Rappelons qu'un événement électrique traduit le passage d'un état stable à un autre état stable et se traduit par une variation de l'amplitude entre ces deux états (voir l'équation (1.1)).

Ainsi l'observation $y(t)$ est représentée par la série $(t_i, A_i)_i$ notée z_g qui est le foisonnement de 5 classes.

Les classes de signaux (événements) considérées sont les suivantes.

- ➢ **Classe** $\mathscr{C}_1$: cette classe correspond aux signaux « symétriques à régulation très rapide » (exemple : fours, bouilloire, signal de la résistance de chauffage de l'eau d'un lave-linge), où :
  - *« symétrique »* signifie qu'un événement du type *on* et l'événement du type *off* correspondant ont quasiment la même amplitude en valeur absolue. Une modélisation de cette caractérisation par une loi de probabilité sera présentée ultérieurement.
  - *« régulation très rapide »* signifie dans notre cas d'application que la constante de temps caractéristique de la régulation de l'appareil T_{rg} est de l'ordre d'une minute. Le domaine de définition précis de ce paramètre, dans notre cas d'application, est présenté dans la section suivante.
- ➢ **Classe** $\mathscr{C}_2$: cette classe correspond aux signaux « symétriques à régulation moyenne » (exemple : fours, bouilloire, signal de la résistance de chauffage de l'eau d'un lave-linge) où :
  - *« régulation moyenne »* signifie dans notre cas d'application que la constante de temps caractéristique de la régulation de l'appareil T_{rg} est de l'ordre de quelques minutes. Le domaine de définition précis de ce paramètre dans notre cas d'application est présenté dans la section suivante.
- ➢ **Classe** $\mathscr{C}_3$: cette classe correspond aux signaux « asymétriques à régulation lente et de forte amplitude » (exemple : ECS) où :
  - *« asymétrique »* signifie qu'un événement du type *on* et l'événement du type *off* correspondant ont des amplitudes significativement différentes en valeur absolue, relativement à un seuil prédéfini. Une modélisation de cette caractérisation par une loi de probabilité sera présentée ultérieurement.
  - *« régulation lentes »* signifie dans notre cas d'application que la constante de temps caractéristique de la régulation de l'appareil T_{rg} est au moins égale à une dizaine de minutes. Le domaine de définition précis de ce paramètre dans notre cas d'application est présenté dans la section suivante.
  - « *forte amplitude* » signifie que l'amplitude moyenne de chaque événement est supérieure (en valeur absolue) à un seuil prédéfini A_{3min} fixé à 1000 W dans le cas de la CdC.
- ➢ **Classe** $\mathscr{C}_4$: cette classe correspond aux signaux « asymétriques à régulation lente de faible amplitude » (exemple : compresseurs du froid alimentaire, pompe à chaleur), où :
  - « *faible amplitude* » signifie que l'amplitude moyenne de chaque événement est inférieure (en valeur absolue) à un seuil prédéfini A_{4min} fixé à 500 W dans le cas de la CdC.
- ➢ **Classe** $\mathscr{C}_5$: cette classe regroupe les événements de la CdC n'appartenant à aucune des classes précédentes.

Les séries d'événements (paires d'événements) correspondant à ces cinq classes sont respectivement notées z_k où $k \in \{1,2,3,4,5\}$.

	f_a	f_{11}	f_{00}	f_{10}
Classe $\mathscr{C}_1$	$f_a(.) = \frac{1}{2b} exp(-\frac{\lvert . \rvert}{b})$	$\mathscr{N}(0, s_{11})$	$\mathscr{N}(0, s_{11})$	$\mathscr{U}(D_1^{00})$
Classe $\mathscr{C}_2$	$\mathscr{N}(\mu_{a2}, \sigma_{a2})$	$\mathscr{N}(\mu_{11}, s_{11})$	$\mathscr{U}(D_2^{00})$	$\mathscr{U}(D_2^{10})$
Classe $\mathscr{C}_3$	$\mathscr{N}(\mu_{a3}, \sigma_{a3})$	$\mathscr{U}(D_3^{11})$	$\mathscr{U}(D_3^{00})$	$\mathscr{U}(D_3^{10})$
Classe $\mathscr{C}_4$	$\mathscr{N}(\mu_{a4}, \sigma_{a4})$	$\mathscr{U}(D_4^{11})$	$\mathscr{U}(D_4^{00})$	$\mathscr{U}(D_4^{10})$

TABLE 5.2: Lois a priori des paramètres des classes d'événements de la CdC

Notons d'ores et déjà que ce choix n'est pas complètement figé : la description des appareils de chaque classe, les caractéristiques et le nombre de classes seront complétés en fonction de l'analyse des signaux d'autres usages que ceux considérés dans cette thèse.

1.2. Informations *a priori*

Les paramètres décrivant les différentes classes sont :

- ➢ la variation d'amplitude entre deux événements successifs de nature différente (*on-off*) notée $(A_{10}^{(k)})$ pour la classe k . Ces variations sont modélisées par les lois de probabilité $f_a^{(k)}$ pour chaque classe étudiée. Cette loi est appelée *loi des amplitudes.*
- ➢ la durée entre deux *on* successifs, la durée entre deux *off* successifs et la durée entre un *on* et un *off* successifs : ces durées sont respectivement modélisées par les lois de probabilités $f_{11}^{(k)}$, $f_{00}^{(k)}$ et $f_{10}^{(k)}$. Ces lois de probabilités seront résumées par un notation commune $f_{df}^{(k)}$ où $d, f \in \{1, 2, 3\}$ en fonction du type de chacun des deux événements considérés.

Ainsi, deux événements (A_i, t_i) et (A_j, t_j) successifs de la classe $\mathscr{C}_k$ vérifient les deux relations suivantes 5.7 :

$$\begin{cases} \mid A_i - A_j \mid \sim f_{a,k} \\ \mid t_i - t_j \mid \sim f_{df}^{(k)} \end{cases} \tag{5.7}$$

Le choix des lois *a priori* de chaque classe est fondé sur l'analyse des signaux ou des connaissances sur le fonctionnement des usages.

A titre d'illustration, nous présentons sur la figure (5.15) les histogrammes de la variation d'amplitude observée entre deux événements successifs pour un convecteur et un ballon d'eau chaude sanitaire. L'histogramme obtenu est centré pour le convecteur et a une moyenne nulle traduisant ainsi la symétrie « *on-off* » des événements ; alors que celui obtenu pour l'ECS est décentré, ce qui traduit l'asymétrie des événements « *on-off* » pour cet usage. Nous considérons une lois double exponentielle pour le chauffage et une loi normale décentrée (de moyenne $\mu_2 = 150$ pour l'ECS). Ci-après les lois *a priori* des paramètres des classes d'événements considérées.

1.3. Lois *a priori* des variations d'amplitude d'un couple d'événements « *on-off* » successifs

Les lois considérées sont résumées dans le tableau (5.2).

1.4. Deux approches envisagées

L'objectif consiste à extraire les quatre séries d'événements utiles, ce qui revient à attribuer à chaque événement une étiquette (*label*) parmi les cinq valeurs possibles et à les apparier. Ce problème peut être vu sous l'angle d'un test d'hypothèses multiples. Une méthode a été envisagée dans ce cadre et a abouti à la mise en œuvre d'un algorithme séquentiel de classification des événements. Le problème posé est un problème d'optimisation combinatoire, que nous proposons d'aborder en utilisant une méthode du type MCMC. Ce choix permet d'intégrer les *a priori* sur les signaux et d'explorer plus efficacement l'espace des solutions possibles.

2. Formalisation dans le cadre de la simulation bayésienne

A chaque signal observé $y(t)$, on associe la série des événements (A_i, t_j) qui est la réunion des séries d'événements des classes considérées.

Les événements de chaque classe $\mathscr{C}_k$ sont modélisés de la manière suivante :

$$\begin{cases} |\, A_i - A_j \,| \sim F_k \\ |\, t_i - t_j \,| \sim G_k \end{cases} \tag{5.8}$$

où F_k et G_k sont respectivement la loi d'amplitude et la loi des durées associées à la classe considérée.

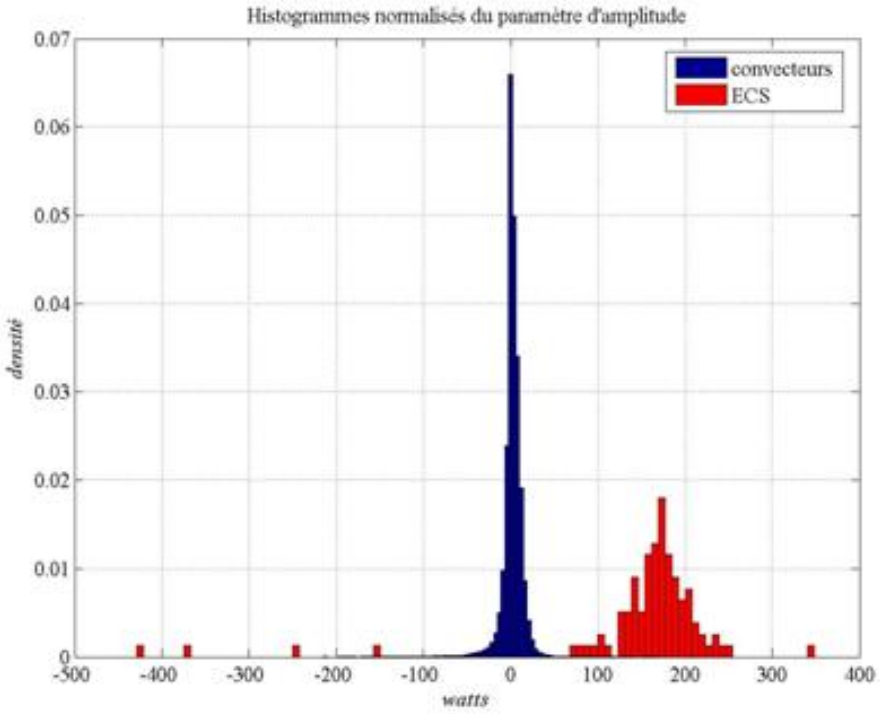


FIGURE 5.15: Histogramme des variations d'amplitude *on-off* d'un convecteur et de l'ECS

Soit Π_k la probabilité de la classe $\mathscr{C}_k$. Les séries d'événements recherchées sont celles qui expliquent au mieux, au sens du maximum a posteriori, la série observée. Il s'agit donc d'estimer les étiquettes (*labels*) k des événements observés en maximisant la loi *a posteriori* définie par l'équation suivante :

$$f(z) = \prod_{\text{paires } (i,j)} G_k(t_i - t_j) F_k(t_i - t_j)\, \Pi_k\, \Omega(z)$$

où :
- ➢ z est la série des événements (A_m, t_m, k_m) ;
- ➢ $\Omega(z)$ est la loi *a priori* (c'est le produit des lois a priori sur les différents paramètres du modèle).

3. Algorithme

L'algorithme mis en œuvre pour simuler la loi a posteriori est similaire à l'algorithme du recuit simulé présenté pour la décomposition de la CdC du chauffage décrit dans le chapitre (**??**).

Rappelons qu'une paire d'événements est un couple d'événements de même nature ou non. Nous utilisons cette terminologie lors de la description des perturbations (transitions) envisagées.

3.1. Définition des transitions

- ➢ Permuter deux paires d'événements : il s'agit d'intervertir les labels de deux paires d'événements tirés aléatoirement selon une loi uniforme.
- ➢ Supprimer une paire d'événement d'une classe ;
- ➢ Ajouter une paire d'événements tirée aléatoirement parmi ceux de la classe $\mathscr{C}_5$;
- ➢ Modifier le label d'un événement sur deux d'une classe donnée ;
- ➢ Supprimer tous les événements de la série d'une classe donnée (autre que $\mathscr{C}_5$) ;
- ➢ Ajouter une série d'événements tirée aléatoirement de $\mathscr{C}_5$ à une classe donnée.

3.2. Initialisation des séries

Chaque classe est initialisée par une série dont la taille est fonction de la probabilité d'apparition de la classe ($\Pi_k.N$ où N est le cardinal de l'ensemble des événements). Si N_k est le nombre initial de paires de la classe $\mathscr{C}_k$, on tire aléatoirement N_k instants t_{ik} dans la liste des « *on* ». Les événements du type « off » sont alors générés aléatoirement connaissant les dates de t_{ik} selon la loi de probabilité f_{10}.

4. Validation sur des cas simplifiés

L'algorithme proposé a été implémenté. Il a été validé sur deux cas-test simplifiés. Le premier cas-test est un mélange de deux convecteurs d'amplitude moyenne différente et de même période de régulation. Le second cas test est un mélange de deux signaux : la CdC du froid alimentaire et celle du ballon d'eau chaude. Les événements à séparer sont présentés sur la figure (5.16).

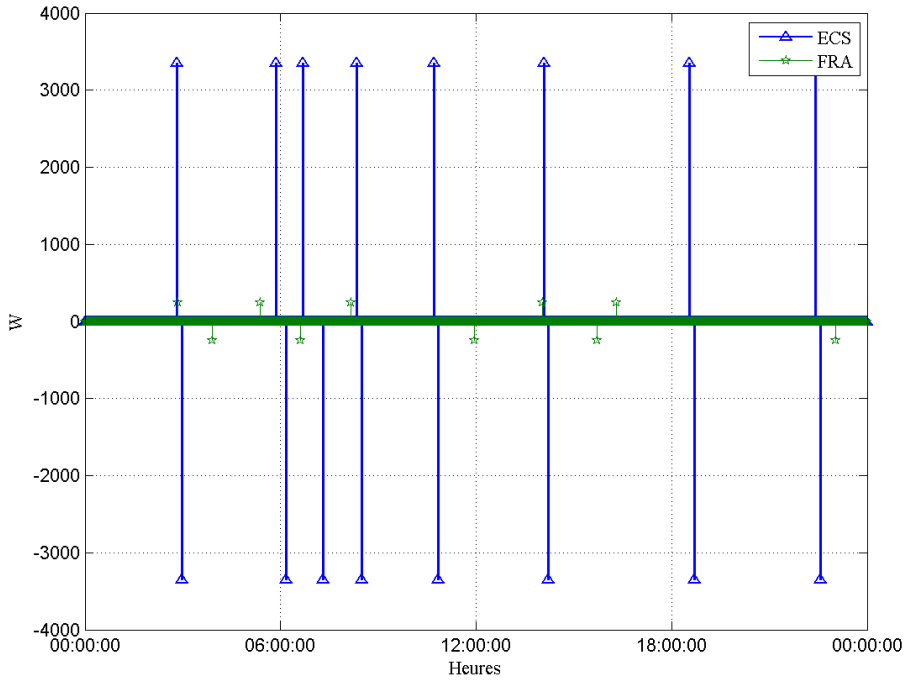


FIGURE 5.16: Cas test - mélange du froid alimentaire et de l'ECS

La série des événements du signal observé $y(t)$ échantillonné à 1 s est obtenue par seuillage de la série des différences finies du signal. Le seuil appliqué (en valeur absolue) est de 100 W. L'application de l'algorithme permet d'identifier parfaitement les événements dans les deux cas tests. Il s'agit bien de deux tests pour valider la démarche. Mais, il ne s'agit que de premiers résultats de preuve de faisabilité obtenus en fin de thèse. Des améliorations de la méthode, portant plus précisément sur le choix des noyaux de transition et sur la gestion des cas d'événements manquants ou simultanés, sont nécessaires avant d'envisager des tests sur des mélanges réels.

Remarque

La méthode générative fondée sur une modélisation paramétrique a été développée dans le cadre précis de la CdC quasi-mono-usage (chauffage par exemple). Sa généralisation à une CdC quelconque n'est pas évidente à cause de la grande diversité des signaux d'une part, et de la combinatoire du problème. Une extension de cette approche dans le cadre de la théorie des processus ponctuels marqués a été proposée par Julien Bect et al dans [GBLa09] et se poursuit dans le cadre d'une thèse démarrée en mars 2009 au département Signaux et Systèmes Électroniques de Supélec.

Approche fondée sur un « enrichissement » des données d'entrée

Nous avons considéré le problème de la caractérisation dans des conditions réelles avec l'infrastructure existante. D'autres approches peuvent être envisagées. Par analogie avec les approches

NIALM [Cox06, Lee03], il serait judicieux d'investiguer des méthodes de traitement du signal à partir de signaux échantillonnés à une fréquence plus élevée. L'analyse des algorithmes de la famille NIALM montre qu'il est envusageable de s'affranchir de la phase d'apprentissage en adoptant une démarche de traitement du signal telle que effectuée dans cette thèse.

Une autre approche potentielle consisterait à tirer profit des systèmes d'information disponibles, qui sont de plus en plus présents dans les maisons. Cette approche ainsi qu'une étude de l'apport des systèmes de mesure dits « multi-prises » ou *prises intelligentes* sont en cours d'investigation au sein d'EDF en collaboration avec plusieurs partenaires industriels et académiques.

Publications

Articles de conférence avec comité de relecture

M. El Guedri, C. Lajaunie, G. Fleury et G. D'Urso, *Modèle stochastique pour la séparation des composantes de la courbe de charge électrique*, XXIIème colloque **GRETSI**, Dijon, France 2009.

M. El Guedri, G. D'Urso, C. Lajaunie, G. Fleury et S. Parfouru, *Time-Frequency characterization for electric load Monitoring*, **EUSIPCO** (**EU**ropean **S**ignal and **I**mage **P**rocessing **CO**nference), Glasgow, Ecosse 2009.

M. El Guedri, J. Bect, C. Lajaunie, G. Fleury, R. Séraoui, A. Garcia et A. Girard, *RJMCMC Point-Process Sampler for Single Sensor Source Separation : an Application to Electric Load Monitoring*, **EUSIPCO** (**EU**ropean **S**ignal and **I**mage **P**rocessing **CO**nference), Glasgow, Ecosse 2009.

Brevet

B. Charbonnier, G. D'Urso, M. El Guedri, *Method and system for detecting and estimating the electric power consumption of a subscriber's facilities*, **European patent n°EP20080157409**, Octobre 2008.

ANNEXES

Annexe A

Description du compteur bleu électronique

En 2005, le nombre de compteurs de relevé de consommation électrique dans le secteur résidentiel s'élève à 35 millions, toutes catégories comprises. En 2007, 21 millions de compteurs sont électromécaniques et un million de compteurs électroniques sont installés par an. La figure (A.1)présente une image d'un CBE en précisant ses principales composantes.

Le compteur bleu électronique (CBE) est un objet communicant : différentes informations sont disponibles dans le compteur, certaines étant transmises via une sortie dédiée appelée *sortie télé-information* (télé-info).

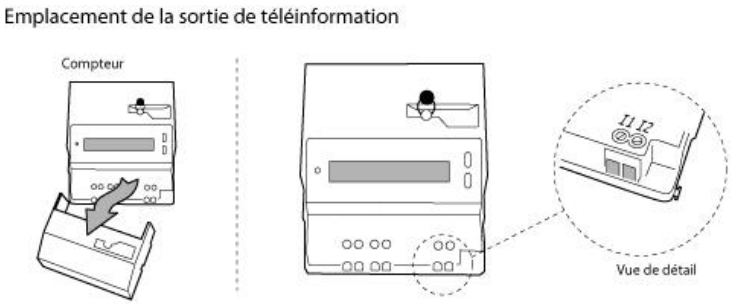


FIGURE A.1: Sortie Télé-info du CBE

En effet, les compteurs bleus électroniques (CBE) disposent de deux fonctions de communication : d'une part un accès « adressé » aux données de comptage par un bus de TELERELEVE au standard EURIDIS et d'autre part une sortie *téléinformation* (*Téléinfo*) spécifique EDF mettant en continu des données à disposition d'un équipement récepteur.

Sortie télé-information (client particulier)

Une transaction par le bus *TELERELEVE* se fait en trois ou quatre secondes à l'initiative d'un équipement de relève à distance qui a un accès partagé à *N* compteurs qu'il doit interroger l'un

après l'autre. Par conséquent, la fréquence de ces transactions est limitée par le nombre N de compteurs à relever. La sortie TELEINFO émet vers un équipement récepteur côté client en continu –s'il n'est pas interrompu par une TELERELEVE- et envoie un « train » de données nouvelles environ toutes les secondes (la fréquence exacte est liée à l'option tarifaire du client). Cette notion d'interruption en cas de TELERELEVE est intéressante car elle peut servir de synchronisation pour une action de vérification de cohérence entre les données de télérelève et les données de téléinfo.

La Téléinformation du client est réalisée par une liaison série numérique modulée qui diffuse en permanence des information contenues dans les mémoires du compteur. Cette liaison peut être utilisée pour connecter des systèmes tels que : un afficheur déporté, un terminal ou encore un gestionnaire d'énergie.

Principe

Les informations sont transmises cycliquement en série sur la ligne. Chaque donnée transmise est précédée d'une étiquette permettant de l'identifier. L'ensemble des données transmises dépend de la programmation du compteur. Les groupes d'informations inutiles, compte tenu du mode de fonctionnement programmé, ne sont pas émis.

Grandeurs délivrées par la sortie télé-info

Les grandeurs fournies par la sortie télé-info peuvent être classées en deux catégories :

- ➢ information sur le client : tarif, puissance souscrite, etc.
- ➢ La consommation du client : puissance appelée, Index, Intensité instantanée, etc. Ces grandeurs sont mises à jour chaque seconde, ce qui permet de suivre leur évolution au cours du temps.

La résolution des principales grandeurs qui nous intéressent est résumée dans le tableau (A.1).

Grandeur électrique	**Résolution**
Puissance active	10 W
Puissance apparente	10 VA
Intensité de courant	$\pm$ 0.5A

TABLE A.1: Pas de quantification des mesures électriques en sortie du CBE

Remarque

La seule grandeur disponible en sortie du CBE est la puissance apparente.

Annexe B

Définitions de quelques grandeurs électriques

L'objectif de cette annexe est de définir les grandeurs électriques d'intérêt ainsi que de présenter les propriétés physiques de certains composants (dipôles). Nous nous limiterons au cas du courant alternatif en régime monophasé.

Soit D un dipôle soumis à la tension instantanée $u(t)$ et traversé par l'intensité de courant instantanée $i(t)$. On distingue deux types de dipôles : les dipôles linéaires et les dipôles non-linéaires, en fonction de la relation reliant l'intensité de courant instantanée à la tension telle que l'illustre la figure (B.1) où :

$$\begin{cases} i_1(t) = i(\omega, t) \\ i_2(t) = i(\Omega, t) \quad \text{avec } \Omega = [\omega, 2\omega, 3\omega, ..]^t \end{cases}$$

La puissance instantanée du dipôle est définie par l'équation (B.1) :

$$P(t) = u(t)\, i(t) \tag{B.1}$$

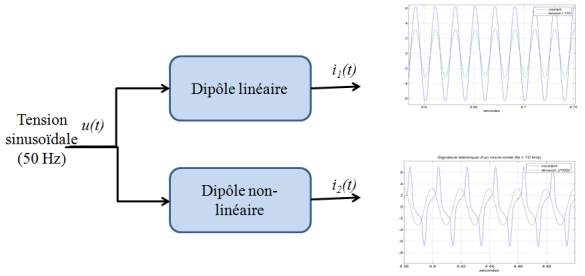


FIGURE B.1: Dipôle linéaire et dipôle non-linéaire

Cas d'un dipôle linéaire : Puissances en régime sinusoïdal de courant et de tension

Le dipôle D est caractérisé par son impédance complexe $Z = R + jX = Z(\cos\theta + j\sin\theta)$, où R et X sont respectivement la résistance et la réactance du dipôle, θ est le déphasage du courant par

rapport à la tension. En supposant que la tension est un signal purement sinusoïdal de période T_0 (fréquence $f_0 = \frac{1}{T_0} = 50\,\text{Hz}$), on définit les trois puissances suivantes :

➢ **La puissance active** est, par définition, la valeur moyenne de la puissance instantanée sur une période. Elle correspond à la puissance électrique consommée par la composante résistive du dipôle. Nous avons alors :

$$P(t) = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} v(\tau)\, i(\tau)\, d\tau = UI\cos\theta$$

où U et I sont les valeurs efficaces de tension et de courant.

➢ **La puissance réactive** correspond à la part associée à la réactance du dipôle, et est définie ainsi :

$$Q(t) = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} v\left(\tau - \frac{T_0}{4}\right) i(\tau)\, d\tau = UI\sin\theta$$

➢ **La puissance apparente** est le produit des valeurs efficaces de courant et de tension : $S = UI$ et vérifie la relation $S^2 = P^2 + Q^2$.

Rappel : Théorème de Boucherot

La puissance active (*réactive*) *dissipée dans un groupement de dipôles* (*fonctionnant en régime sinusoîdal*) *est égale à la somme des puissances actives* (*réactives*) *dissipées dans chacun des dipôles, quelle que soit leur association.*

Conséquence

A tout instant, la puissance active d'une installation électrique (par exemple une maison) est égale à la somme des puissances actives appelées par N appareils électriques :

$$P(t) = \sum_{m=1}^{N} P_m(t) \tag{B.2}$$

Cas d'un dipôle non-linéaire : Puissances en régime sinusoïdal de tension et en régime non-sinusoïdal de courant

La distribution de l'électricité se fait en régime sinusoïdal, si l'on néglige la pollution du réseau. Mais, la plupart des appareils électriques utilisés par les particuliers et les industriels appellent un courant périodique mais pas nécessairement sinusoïdal, à cause de la présence de plus en plus de convertisseurs de l'électronique de puissance dans ces appareils. Par conséquent, l'intensité de courant appelée par cette catégorie d'appareils contient des *courants harmoniques* en plus de la composante oscillant à la fréquence fondamentale (décomposition en série de Fourier). Tradionnellement, les sources d'harmoniques sont essentiellement attribuées au secteur industriel. Or des campagnes de mesure menées en France ont montré que le pic de pollution du réseau est atteint le dimanche soir, moment où les téléviseurs sont massivement allumés. Le tableau (B.1) présente les principaux générateurs d'harmoniques chez les industriels et chez les particuliers.

Secteur résidentiel	Lampes basse consommation	Tubes fluorescents	Téléviseurs, ordinateurs
Secteur industriel	Variateurs de vitesse (moteurs)	Fours à arc (métallurgie)	Soudeuses

TABLE B.1: Exemples de générateurs d'harmoniques

La définition de *puissance* en présence de courants harmoniques telle que cette puissance ait un sens physique est un défi qui n'est pas complètement résolu. Par analogie avec la définition de la puissance active (réactive) en régime sinusoïdal pour un dipôle linéaire, la puissance active et la puissance réactive associées à l'harmonique de rang k sont définies comme suit (B.3) :

$$\begin{cases} P_k(t) = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} v(k\tau)\, i(\tau)\, d\tau = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} V\cos\left(k\frac{2\pi}{T_0}\right) i(\tau)\, d\tau \\ Q_k(t) = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} v\left(k\tau - \frac{T_0}{4}\right) i(\tau)\, d\tau = \frac{1}{T_0}\int_{t-T_0/2}^{t+T_0/2} V\sin\left(k\frac{2\pi}{T_0}\right) i(\tau)\, d\tau \\ \text{pour } k = 1,\, 2,\, 3,\, \ldots \end{cases} \tag{B.3}$$

Annexe C

Sythèse des travaux antérieurs sur la caractérisation des usages électriques

Approches statistiques pour la décomposition non-intrusive de la CdC

Modèles de sources markoviennes cachées (HMMs)

Nous avons vu que Georges Hart a ramené le problème de la décomposition de la CdC électrique (puissance active et réactive en régime permanent) au décodage de la série des événements définie par l'équation (1.3). Cette phase de décodage a été réalisée en se basant sur une modélisation des appareils électriques par des machines à deux états $(on,\ off)$. Cette méthode est fondée sur les deux hypothèses suivantes :

H_1 Des usages distincts ont des signatures distinctes dans l'espace de représentation $(\Delta P, \Delta Q)$.

H_2 Un événement se traduit par une transition d'un niveau de consommation quasiment constant à un autre niveau de consommation quasiment constant.

Le décodage des événements a fait l'objet de différents algorithmes, principalement fondé sur la modélisation des usages par des machines d'état. Cette idée a été reprise dans le cadre des travaux de la thèse de Marc Bons [Bon96] et a abouti à une nouvelle modélisation de la CdC électrique dans le secteur résidentiel. Cette méthode consiste à modéliser la puissance mesurée en sortie du compteur d'électricité par des chaînes de Markov cachées [Rab89].

Cette méthode est basée sur le constat suivant : les CdCs des particuliers présentent une grande diversité liée à la diversité des installations électriques et aux différents modes de consommation (qui dépendent principalement des habitudes des clients et des tarifs choisis). Ces éléments ainsi que d'autres facteurs tels que la température et des événements particuliers (départ en vacances, vague de froid très importante, etc.) confèrent un caractère erratique aux données analysées, qui suggère une modélisation stochastique. On peut donc considérer qu'il existe des périodes de consommation *homogènes* et d'autres aléatoires.

Ainsi, nous disposons d'une grandeur observée qui est la puissance totale appelée, les classes des différentes courbes de charge élémentaires étant des variables cachées. Ce contexte sug-

gère l'utilisation d'un modèle statistique du type chaînes de Markov cachées (**H**idden **M**arkov **M**odels : **HMM**).

Généralités sur les HMM Les modèles à sources markoviennes cachées (HMM) constituent un outil de modélisation stochastique qui permet d'utiliser des distributions de probabilité et d'incorporer des information structurelles. Ces techniques ont connu un essor très important depuis les années 80 grâce à leur application dans le domaine de la reconnaissance de la parole.

Un modèle de chaîne de Markov cachée est défini par une séquence finie d'états, où chaque état est associé à une distribution de probabilité. Les transitions entre les différents états sont gouvernées par *des probabilités de transition*. Quand le système est dans un état, une observation est générée selon une *probabilité d'observation*. Les états ne sont pas observables, d'où le terme « cachés». Plus précisément, un modèle de chaînes de Markov cachées est un processus doublement stochastique. Les deux processus sont :

- le processus des états cachés noté $(q(t))_t$:
  - C'est un processus markovien. On suppose qu'à l'instant t, le système est dans l'état $q(t) \in \{S_1, ..., S_N\}$. Par exemple, dans le cas d'un appareil électrique, les états sont les états de marche et arrêt si on modélise un seul appareil à deux états, et sont les configurations possibles si on modélise toute l'installation électrique.
    Les changements d'état sont définis par des probabilités de transition : $a_{ij} = P(q(t+1) = j/q(t) = i)$. La matrice des probabilités de transition est notée A.
- le processus de génération des observations $O_t(t)$
  - On définit également l'alphabet des observations : $\{v_1, ..., v_M\}$, qui correspond aux niveaux (amplitudes) de consommation dans notre application. Les probabilités associées aux observations sont définies par : $b_j(k) = P(O_t = v_k \mid q(t) = S_j)$ qui est la probabilité d'observer le symbole v_k à l'instant t sachant que l'état courant est S_j.

Le modèle est donc défini par la donnée de la matrice d'observation, la matrice des probabilités de transition et le vecteur des probabilités initiales. Les trois problèmes fondamentaux des HMMs sont :

- l'apprentissage pour ajuster les paramètres du modèle ;
- l'évaluation de la vraisemblance des observations étant donné un modèle spécifié ;
- et le décodage qui consiste à déterminer la séquence d'état la plus vraisemblable (critère choisi).

La méthode proposée dans la thèse de Bons repose sur la modélisation des appareils électriques (par des machines d'états) ainsi que l'interaction entre appareils, pour prendre en compte des cas d'événements simultanés ou manquants. Cette modélisation globale permet de conforter le décodage par la prise en compte de la grande majorité des interactions entre appareils.

Cette approche a pour but d'augmenter la fiabilité du décodage des événements, en comparaison avec les résultats des travaux de Hart. Ici, le décodage a pour but d'attribuer la "meilleure" attribution (au sens de maximum de vraisemblance) aux appareils compte tenu du modèle global. **Le problème de décomposition de la CdC est ainsi ramené à l'estimation de la séquence d'états la plus vraisemblable, où un état de l'installation globale est constitué de micro-états associés aux différents appareils**. Cette phase est réalisée par un algorithme de Viterbi [Vit67].

Modélisation de l'installation électrique Le modèle global de l'installation électrique est l'agrégation des modèles de chacun des appareils. Les états possibles du modèle agrégé sont appelés macro-états, où un macro-état est le $n-uplet$ dont chaque élément est l'état d'un des différents appareils modélisés. Chaque appareil est modélisé par une machine à deux ou plusieurs états en fonction du nombre de composantes de l'appareil.

Par exemple, un convecteur électrique est modélisé par une machine à deux états (marche/arrêt), alors que la modélisation d'un lave-linge fait intervenir au moins trois composantes : une résistance (chauffage de l'eau), un moteur (tambour) et une pompe (vidange de l'eau). Chaque composante du lave-linge est modélisée par une machine à deux états, les différentes composantes étant ensuite interconnectées entre elles en tenant compte du fonctionnement de l'appareil, ce qui peut amener à multiplier les états.

Modélisation des durées L'étude des événements de la CdC effectuée dans l'étape de décodage est fondée sur la succession des événements (au cours du temps) et sur leurs puissances. Une telle méthode peut être évaluée en termes de pourcentage d'événements correctement identifiés. Ce critère est loin d'être satisfaisant étant donné d'une part que l'on s'intéresse à l'énergie consommée par l'usage et d'autre part à la grande dispersion entre usages au sens de leur densité d'événements. Par exemple, un ballon d'ECS génère une dizaine d'événements par jour, alors qu'un convecteur génère quasiment autant d'événements pour seulement une durée de fonctionnement de l'ordre de cinq minutes. Des erreurs égales de reconnaissance des événements ne donnent pas lieu à des erreurs similaires en énergie consommée.

Afin d'améliorer la phase de décodage, M. Bons propose dans sa thèse d'introduire des connaissances *a priori* sur les durées de fonctionnement par analogie avec l'utilisation des lois de durée en reconnaissance de la parole [Fer80, Vas95]. Deux grandeurs sont alors observées : la CdC globale (séquence des événements) et les durées évaluées.

En résumé La modélisation des appareils électriques par des machines à deux états est une méthode originale qui a permis de résoudre les problèmes liés à la superposition d'événements et les cas d'appareils de même puissance. Elle permet de prendre en compte des appareils ayant plusieurs composantes (charges) ; ce qui n'a pas été traité dans les travaux de Hart. De plus, cette méthode fondée sur une phase d'apprentissage et une phase de décodage, et prenant en compte les lois des durées de certains appareils, permet de prendre en compte les spécificités de comportement de la clientèle des particuliers. A l'issu de ce travail, le chauffe-eau, les convecteurs et le réfrigérateur sont souvent correctement estimés. Des appareils tels que le lave-linge sont souvent sous-estimés bien que correctement détectés. Le lave-vaisselle, le sèche-linge et les appareils de cuisson n'ont pas été pris en compte dans cette étude.

Application d'un algorithme génétique à la décomposition de la puissance active

Des travaux de recherche à l'Université de Paderborn [BV03, BV04] ont pour objectif de mettre en place un système de reconnaissance des usages électriques et d'estimation des énergies consommées :

➢ de manière non-intrusive ;

- ➢ en exploitant l'unique mesure disponible à la sortie du compteur d'électricité (puissance active totale échantillonnée à 1 s et sous-échantillonnée à 16 s) ;
- ➢ et sans information *a priori*.

Exploiter uniquement des mesures disponibles en sortie du compteur permettrait alors de réduire le coût du système final. Plus précisément, les données d'entrée exploitées sont : la puissance active échantillonnée à une période comprise entre 1 s et 16 s. Les usages ciblés par la méthode sont uniquement les usages fréquents ou récurrents à l'échelle d'une journée, tels que le chauffage, le froid alimentaire et l'eau chaude sanitaire. La méthode développée est fondée sur la modélisation de chaque usage par une machine d'états. Elle exploite plusieurs paramètres discriminants entre usages : leur amplitude, le nombre d'occurrences par unité de temps (densité des événements) ainsi que la durée d'un cycle de fonctionnement.

La démarche de séparation des usages est toujours celle proposée par Hart dans les années 80 : détecter les événements, les étiqueter et reconstituer les courbes de charge des usages. Mais cette approche utilise uniquement la puissance active, et ne peut donc bénéficier des apports du domaine de représentation $(\Delta P, \Delta Q)$. Ce manque d'information sur les propriétés physiques des appareils est compensé par les propriétés des signaux temporels notamment de la forme du signal, de la densité des événements et des durées des cycles de fonctionnement.

L'algorithme proposé vise à détecter les plages temporelles de fonctionnement des usages, et à estimer leurs énergies à l'échelle de la journée. Il est composé de quatre étapes :

- ➢ Détecter les événements, ceux ayant une faible amplitude étant négligés ;
- ➢ Classer les événements par la méthode de clustering flou :
  - – les événements ayant une structure similaire (amplitude, durée entre événements successifs…) sont regroupés dans un même cluster. Les combinaisons des différents clusters sont des candidats possibles aux différents appareils.
- ➢ Générer l'ensemble des candidats potentiels (machines à états) à partir des combinaisons possibles des clusters
  - – Cette étape est réalisée par un algorithme génétique, pour coder et déterminer les différents candidats.
- ➢ L'identification des appareils est obtenue en optimisant les séquences des machines d'états candidates.
  - – La méthode utilisée pour cette phase d'optimisation est un algorithme génétique.

La reconnaissance des usages électriques est effectuée en s'appuyant sur le résultat de l'algorithme précédent et sur une base de données de référence, contenant des mesures de la puissance active des usages.

En résumé

Cette méthode est assez récente (2003-2004). Elle tente de réduire le coût et l'intrusion de l'installateur d'un système NILM actuel, en exploitant les données disponibles, à savoir la puissance active seule. Cette simplification de la partie hardware se traduit par une grande complexité de l'algorithme de reconnaissance des usages, comme le prévoyait Georges Hart. Mais, celà est-il suffisant pour abandonner l'analyse de la puissance active en vue de la reconnaissance des usages, si on souhaite intégrer les résultats à une offre de services énergétiques à grande échelle ?

Soulignons que cette méthode est, à notre connaissance, l'unique méthode qui a tenté de décomposer la puissance active échantillonnée à $T_e = 1$ s de manière non-intrusive et sans intervention

sur le site, même si une phase d'apprentissage supervisé à partir de données préalablement collectées de différents sites est nécessaire.

Cependant, une approche similaire n'est pas envisageable dans le cadre de nos travaux pour deux raisons : la complexité des algorithmes est incompatible avec une offre énergétique à grande échelle, et le fait qu'elle n'est robuste que pour des cas de mélanges « simples » au sens où il y a très peu de convecteurs électriques voire pas du tout. L'analyse de la méthode montre que tenir compte de cette composante rendrait l'algorithme encore plus complexe et nécessiterait des informations supplémentaires.

Il est important de préciser que cette étude s'inscrit dans le cadre d'une forte volonté d'obtenir un produit qui sera commercialisé et intégré dans les maisons « intelligentes » *(smart-homes)* en exploitant les moyens de communication, en particulier Internet.

Modélisation stochastique de la CdC à base de dictionnaire de formes

Le problème à traiter consiste à décomposer la consommation annuelle d'un site donné (cas d'application : hypermarché ou hôtel) en ses principales composantes à savoir le chauffage, la climatisation, l'éclairage (et les fours dans le cas des hypermarchés). La CdC considérée est la puissance moyenne échantillonnée à $T_e = 10\,\text{mn}$ ainsi que la température extérieure et observée sur une année. Ces données sont traditionnellement analysées dans un espace de représentation appelé : **le graphe d'intensité**.

Il s'agit d'une représentation $(temps, temps)$, où la CdC est représentée sous forme d'une image qui se compose d'une juxtaposition de CdCs (quotidiennes ou hebdomadaires). Pour des CdC quotidiennes par exemple, l'axe des abscisses comporte des heures $0 \rightarrow 24\,h$ et celui des ordonnées des jours $0 \rightarrow 365$.

Dans le secteur tertiaire, la CdC à décomposer est caractérisée par une certaine régularité liée à la reproductibilité des usages. Par ailleurs, certains usages (chauffage/climatisation) ont une forte corrélation avec les conditions météorologiques. Ces connaissances « métiers » constituent des a priori pour la résolution du problème.

La démarche proposée

La démarche proposée dans [Laj02] est fondée sur le constat suivant : les usages étudiés ont des signatures (ici formes géométriques) différentes sur le graphe d'intensité $(heure, jours, semaine)$,

Christian Lajaunie propose alors un modèle probabiliste de la CdC à base d'un dictionnaire de formes géométriques. Ce modèle est décrit par l'équation C.1,

$$P(i,j,k) = \sum_u w_u \phi_u(i,j,k;\theta_u) \tag{C.1}$$

où

- i, j et k désignent respectivement la tranche horaire ($\{1,2,...,144\}$) de la journée, l'indice du jour de la semaine ($\{1,2,...,7\}$) et le numéro de la semaine ($\{1,2,...,53\}$) ;

➢ w_u est l'amplitude de l'usage u ;
➢ ϕ_u le modèle de l'usage défini par les paramètres θ_u .
L'espace d'état étant de grande dimension, un algorithme d'optimisation classique n'est pas envisageable. L'algorithme mis en œuvre est issu du domaine de la simulation bayésienne [DG01], il s'agit d'un algorithme de recuit simulé où un échantillonneur stochastique (Metropolis-Hastings [Has70]) permet d'explorer « au mieux » l'espace d'état et en conséquence de s'approcher au mieux du minimum global.

Modélisation stochastique des usages électriques des hypermarchés sur le graphe d'intensité

Les principaux usages d'un hypermarché sont : la base (représente le talon de consommation permanent), la climatisation, le chauffage, l'éclairage des aires de vente et les fours. L'analyse de la signature de ces composantes sur le graphe d'intensité à partir des données d'apprentissage a permis de définir *un dictionnaire de formes géométriques* caractéristiques des usages ciblés. Trois exemples de modèles définis sont présentés sur la figure (C.1).

La base correspond à une consommation permanente toute l'année et modélisée par un parallélépipède rectangle (C.1) paramétrée par la hauteur.

Le modèle *du chauffage* dépend de huit paramètres qui permettent de décrire le profil en fonction de l'heure de la journée (C.1). Ces paramètres sont présentés ci-après :
➢ T_0 est l'heure de mise en œuvre ;

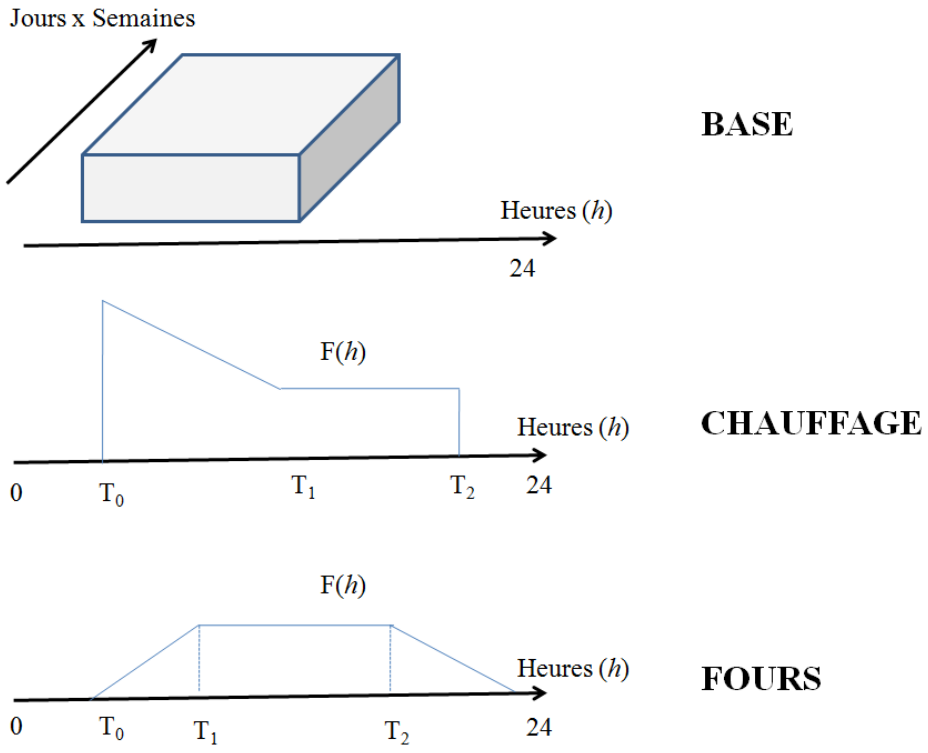


FIGURE C.1: Modèles des composantes de la CdC d'un hypermarché [Laj02]

- T_1 est l'heure de la fin de la phase de décroissance ;
- T_2 est l'heure de coupure ;
- P_3 est la fraction de la puissance initiale retrouvée après stabilisation
- S_4 et S_5 indiquent les semaines de fin et de début de la saison de chauffe ;
- P_6 : puissance minimale à la mise en œuvre et P_7 : incrément de la consommation par degré de température perdu

La puissance moyenne de cet usage s'écrit :

$$\begin{cases} P(h,\theta,s) = F(h)\,\Phi(\theta)\,G(s) \\ \Phi(\theta) = \left\{P_6 + P_7\left(\theta_{chauff} - \theta\right)\right\} 1_{\theta \leq \theta_{chauff}} \end{cases}$$

où

θ_{chauff} la température extérieure à partir de laquelle le chauffage s'enclenche, $G(s)$ une variable indicatrice de saison , Φ un coefficient d'amplitude qui dépend de la température et (h,θ,s)designent respectivement l'heure, le jour et la semaine du point observé.

Quant à la composante climatisation, elle est paramétrée par les six variables suivantes :

- T_0 est l'heure de mise en œuvre ;
- T_1 est l'heure de début de la phase constante ;
- T_2 est l'heure de début de la phase de décroissance ;
- T_3 est l'heure de coupure
- P_4 et P_5 sont respectivement la puissance d'attaque et la puissance par degré supplémentaire.

$$\begin{cases} W(h,\theta) = F(h)\,\Phi(\theta) \\ \Phi(\theta) = \{P_4 + P_5(\theta - \theta_{clim})\}\, 1_{\theta \geq \theta_{clim}} \end{cases}$$

avec θ_{clim} la température extérieure à partir de laquelle la climatisation s'enclenche.

Algorithme d'optimisation

Les paramètres du modèle sont estimés en minimisant itérativement l'erreur quadratique ou l'erreur absolue entre le modèle et le signal observé. L'algorithe d'optimisation mis en œuvre est le recuit simulé. A chaque itération, une transition (perturbation d'un ou de plusieurs paramètres) est proposée. L'erreur est évaluée, puis comparée à l'erreur précédente. Dans le cas d'une diminution du critère, la modification est acceptée. Dans le cas d'une augmentation de l'erreur, la transition est acceptée avec une probabilité appelée probabilité d'acceptation-rejet. Cette probabilité doit permettre de favoriser plus d'acceptation au début de l'algorithme afin de mieux explorer l'espace d'état.

Résultats et commentaires

Les résultats obtenus par cette méthode sont satisfaisants. L'algorithme est intégré à un logiciel dédié à l'offre de services énergétiques dans le secteur tertiaire.

La méthode dans [Laj02] exploite la signature des usages dans un domaine transformé, le graphe d'intensité. L'idée d'introduire un dictionnaire de formes géométriques pour les usages est innovante. Elle présente l'avantage d'offrir la possiblité de décrire paramétriquement les objets recherchés, et de pouvoir intégrer de nouvelles composantes. La solution obtenue est celle qui explique au mieux (au sens du critère défini) les observations.

Cette méthode est transférable à la CdC résidentielle (1 s) sous réserve qu'il existe un espace de représentation de signaux où une modélisation des usages est possible.

Une déclinaison de cette méthode pour la décomposition de la puissance moyenne du chauffage résidentiel est décrite dans le chapitre **??**.

Cependant, elle n'est pas compatible avec une offre énergétique d'analyse détaillée de la consomation par usage dans le secteur résidentiel à cause du coût de calcul. Mais, elle peut être utile pour d'autres services tels que l'optimisation de la gestion du chauffage électrique.

Méthodes de reconnaissance de forme

Des travaux sur la décomposition de la courbe de charge par usage à l'université de Concordia ont abouti à une méthode de reconnaissance de formes pour la détection des usages et l'estimation des énergies consommées [Far99, MZ00]. Les usages ciblés sont l'ECS, le chauffage et le froid alimentaire. L'approche proposée consiste à estimer les usages à partir de la puissance active globale (période d'échantillonnage $Te = 16\,s$) et nécessite une phase d'apprentissage pendant quelques jours (une semaine) à partir d'une base de signaux issus du site étudié. Cette contrainte confère à cette méthode un caractère faiblement intrusif. L'estimation des consommations élémentaires est obtenue en deux étapes :

- ➢ Extraction des caractéristiques des usages à partir des données obtenues pendant la phase de collecte de mesures sur site (mesures de l'intensité de courant totale et de tous les usages envisagés [1]).
- ➢ Analyse de la puissance active globale : détection des usages et estimation des énergies consommées par les appareils ciblés. L'algorithme mis en œuvre est fondé sur un ensemble de règles d'association et utilise des informations complémentaires, notamment l'amplitude des signaux. Le système final s'appelle RPRA (*Rule-based Pattern Recognition Algorithm*).

Ces travaux sont les premiers (à notre connaissance) qui se sont intéressés à la décomposition d'une observation unique (la puissance active). L'un des apports méthodologiques de ces travaux est la recherche de critères discriminants à partir des signaux d'usages. Malheureusement, leur extraction nécessite une intervention chez le client.

La méthode développée en 2004 à l'université de Padderborn (C) s'inscrit dans la suite de cette approche.

1.
   - Les puissances sont ensuite calculées en supposant que la tension est constante.

Les méthodes du type NIALM

Plusieurs méthodes non-intrusives d'analyse et de décomposition de la courbe de charge électrique ont été proposées durant les deux dernières décennies. Les travaux les plus aboutis dans le cadre de l'identification et de l'estimation des usages, on été menés au MIT. Rappelons que les études sur la courbe de charge aux Etats-Unis ont été initiées par Georges Hart au début des années 80. Elles ont abouti à un système de détection et d'estimation des usages électriques dans le secteur résidentiel [Har89, Har92]. Les méthodes mises en œuvre sont classées en trois familles en fonction de la grandeur électrique étudiée, plus précisément en fonction du régime de fonctionnement des signaux exploités. La première s'intéresse à l'analyse des variations des puissances active et réactive en régime permanent ($T_e = 1$ s). La seconde exploite les signaux de courant et de tension en régime transitoire (T_e 100 µs). La troisième famille d'algorithmes est hybride : la détection des événements est souvent effectuée à partir des signaux en régime permanent, la reconnaissance des usages exploite les signaux en régime transtoire.

Identification des usages en régime permanent

Une méthode « non-intrusive » de décomposition de la consommation globale par usage nécessite de définir les paramètres discriminants entre usages. La méthode de détection et d'estimation des usages en régime permanent se fonde sur les deux observations suivantes :

- Les usages électriques peuvent être répartis en plusieurs classes, compte tenu de leurs caractéristiques physiques, en particulier de leur caractère résistif, inductif ou capacitif que l'on peut étudier en utilisant conjointement la puissance active et la puissance réactive :
  Ce constat implique qu'on peut caractériser et reconnaître les signaux élémentaires dans l'espace $(\Delta P, \Delta Q)$: variations de la puissance active et variations de la puissance réactive.
- les appareils électriques se distinguent également par le niveau de la consommation (puissance nominale par exemple) :
  - Cette observation nous amène à détecter les enclenchements et les arrêts des usages en analysant les variations de la puissance active.

La méthode [Har89, Har92], initialement proposée par G. Hart est fondée sur les variations de la puissance active et les variations de la puissance réactive (fréquence d'échantillonnage $f_e = 1$ Hz). Un événement se traduit alors par le passage d'un état permanent [2] à un autre état permanent.

Données d'entrée

En pratique, les courants et les tensions des trois phases de l'installation sont mesurés. Les événements sont détectés à la fois sur les variations de la puissance active et celles de la puissance réactive.

2. Un état permanent est défini dès que la puissance appelée se stabilise (puissance appelée quasiment constante) sur un horizon temporel prédéfini.

Principe de la détection et identification des usages en régime permanent

La décomposition de la courbe de charge par usage est effectuée en cinq étapes [DK92, LLC+03] regroupées dans trois modules.

➢ **Détection des événements**
Il s'agit de détecter les variations de la puissance active et de la puissance réactive calculées par un pré-processeur à partir des mesures de courant et de tension. Cette détection est réalisée avec un détecteur de rupture de la moyenne (et de la variance).

➢ **Classification**
La classification a pour objectif d'attribuer chaque événement détecté à une classe parmi les usages ciblés. Elle s'appuie sur la représentation des événements dans le domaine $(\Delta P, \Delta Q)$. Les événements détectés sont représentés dans cet espace, on obtient alors des clusters. Une illustration (schéma d'explication seulement) est présentée sur la figure C.2.

Le classifieur réalise les tâches suivantes :

1. *les clusters de même amplitude et de signes opposés sont regroupés, et correspondent à la signature d'un appareil fonctionnant en « tout ou rien » (mode on/off). Remarquons qu'une classe d'usages est constituée d'un ou de plusieurs clusters.*
2. *les événements n'appartenant à aucun cluster sont soit associés à l'un des clusters existants, selon un critère de distance ou de maximum de vraisemblance, soit regroupés dans une famille de signatures inconnues.*
3. *chaque cluster est associé à un usage électrique, connaissant sa puissance nominale, et éventuellement d'autres information a priori. Ces informations sur les amplitudes des usages sont collectées lors d'une phase d'apprentissage. La série temporelle (ou un détecteur par usage) de chaque usage est alors reconstituée.*

La réalisation de cette phase de classification des événements nécessite de modéliser les classes associées aux usages et de définir le classifieur. Dans le cadre de l'étude en régime permanent, il est envisageable de modéliser chaque usage par un processus gaussien de moyenne et de matrice de covariance connues (phase d'apprentissage). Le problème de classification des événements est équivalent à un test d'hypothèse multiple. Deux classifieurs ont été appliqués. Le premier est basé sur la minimisation de la distance entre l'observation et le centre de chaque cluster. Le second prend en compte la distribution statistique de chaque classe (chaque classe étant modélisée par un processus gaussien). Un critère de maximum de vraisemblance est utilisé pour la reconnaissance des classes de la série d'événements. Pour plus de détails sur ces deux étapes, on peut se reporter au document de synthèse [MEG07]. Cette méthode est implémentée dans un produit appelé *Steady-State NILM*.

➢ **Estimation des énergies consommées**
Ce module estime l'énergie consommée par chaque usage, connaissant son amplitude et la série des événements qui lui est associée.

L'architecture du système NILM (Non-Intrusive Load Monitoring) le plus récent dans le cas du régime permanent est présentée sur la figure (C.3).

Avantages et limites de l'identification des usages domestiques en régime permanent

La méthode d'estimation des usages électriques, basée sur le régime permanent, a été intégrée dans certains produits (Non-Intrusive Appliances Load Monitoring : NIALM, Single-Point Electrical End-uses Disaggregation : SPEED) commercialisés par la société ENETICS [ene]. Cependant, les systèmes développés n'ont pas été commercialisés à grande échelle (dans le secteur résidentiel), car ils nécessitent l'intervention d'un personnel qualifié pour installer le système de mesure et collecter les données nécessaires à la phase d'apprentissage. De plus le système proposé est assez coûteux (**SPEED** : 1200 euros).

En résumé, la méthode de décomposition de la courbe de charge proposée présente l'avantage d'être simple à mettre en œuvre, et de permettre une estimation de la consommation globale par usage, sans instrumenter tous les appareils en exploitant les signatures électriques des appareils dans le domaine $(\Delta P, \Delta Q)$. La modélisation par processus gaussiens des différentes classes dans cet espace de représentation est simple, mais elle nécessite de déterminer préalablement les paramètres de chaque classe. Elle permet d'identifier des charges ayant des puissances différentes et/ou des caractéristiques physiques (résistif, inductif) différentes. Cette approche présente des limites dans le secteur tertiaire, le secteur industriel et de plus en plus le secteur résidentiel pour les raisons suivantes :

- ➢ *La présence d'un bruit généré par les appareils électroniques rend la définition des conditions de « steady-state » difficile, et par conséquent la série des changements d'états donc la série des événements associés aux différents usages ;*
- ➢ *Il y a de plus en plus de cas de on et/ou off superposés (à cause de l'augmentation du nombre d'appareils) ;*
- ➢ *La présence d'usages électriques ayant une puissance variable au cours du temps.*

De plus, elle ne tient pas compte des usages donnant lieu à des clusters confondus totalement ou partiellement dans le domaine d'étude. En effet, l'approche précédente est basée sur les hypothèses suivantes :

- ➢ H_1 **: Des usages distincts ont des signatures distinctes dans l'espace d'étude.**
  Ici, le terme « usage» désigne un seul appareil électrique. Ce n'est pas forcément le cas dans le secteur tertiaire (commercial) où il y a beaucoup plus d'usages que dans une maison. L'espace d'étude est alors ***perturbé*** *par la présence de plusieurs signatures, et devient de moins en moins discriminant au fur et à mesure que le nombre d'usages augmente. En fait, dans le tertiaire on trouve plusieurs usages (par exemple les moteurs) dont les clusters se superposent dans le plan* $(\Delta P, \Delta Q)$.

- ➢ H_2 **: Un événement se traduit par le passage d'un niveau de consommation quasiment constant à un autre (« steady-state »).**
  Cette hypothèse implique qu'il faut attendre la fin du régime transitoire, et que l'appareil atteigne le régime permanent, pour effectuer la mesure qui correspond à cet état. Dans le secteur industriel ou commercial, trouver une période d'échantillonnage convenable pour disposer de la puissance des usages en régime permanent semble difficile. Une période trop longue donne lieu à des non-détections, car des activations d'usages ne seront pas perçues par le détecteur. Une période trop faible peut fournir des mesures pendant le transitoire, ce qui donne lieu à des fausses alarmes. Certaines composantes du HVAC telles que des réfrigérateurs ou des ventilateurs peuvent nécessiter une durée qui varie de quelques secondes à

quelques minutes avant d'atteindre le régime permanent (vitesse finale de fonctionnement). Certaines machines à vitesse variable (Variable Speed Drives) peuvent ne jamais atteindre un état permanent, ce qui empêche de déterminer des états permanents et de disposer des mesures de puissance associées.

➢ H_3 **: Les Steady-State NILM effectuent la décomposition de CdC journalière off-line.** *Cette approche ne peut être suffisante dans le secteur industriel ou commercial, où l'on souhaite non seulement décomposer la courbe de charge globale par usage, mais où l'enjeu de surveillance des machines et leur diagnostic est essentiel. Afin de pallier les limites du « Steady-State NILM », une nouvelle approche a été proposée. Elle consiste à augmenter la fréquence d'échantillonnage, et à utiliser la signature des appareils en régime transitoire pour les identifier.*

En résumé, cette méthode d'estimation des usages électriques à partir du couple (puissance active, puissance réactive) échantillonnée à 1 Hz :

➢ nécessite une phase d'apprentissage sur le site concerné pour avoir les puissances des usages ;
➢ n'est pas adaptée aux secteurs tertiaire et industriel, où l'on a une plus grande diversité d'appareils,
➢ et présente des limites dans le secteur résidentiel dès que le nombre d'appareils (et par conséquent le nombre d'événements) devient assez élevé.

Afin de résoudre ces problèmes et d'étendre le domaine d'application de la méthode au secteur commercial, une nouvelle approche a été proposée. Elle exploite la signature électrique de chaque appareil en régime transitoire, en plus des signatures en régime permanent [Lee93].

Identification des usages en régime transitoire

Nous avons vu dans le paragraphe précédent que l'estimation des usages en régime permanent présente des limites au niveau de la détection (cas d'événements très proches dans le temps) et de la classification (cas d'usages ayant des signatures identiques dans le plan $(\Delta P, \Delta Q)$). Pour franchir ces limites, une analyse plus fine de la signature des appareils s'impose, et ce en augmentant la fréquence d'échantillonnage par exemple. En effet, le transitoire d'un appareil électrique est étroitement lié au modèle électrotechnique de l'appareil et à son utilisation. Par exemple, le transitoire associé à un moteur à induction et celui associé à une lampe sont distincts, parce qu'augmenter la vitesse du rotor est fondamentalement différent du chauffage du filament de la lampe (figure C.4).

La puissance appelée par la lampe présente un pic à l'enclenchement (puissance élevée nécessaire pour chauffer le filament), et qui se stabilise très rapidement (au bout de 0.2 s). Alors, que dans le cas du moteur, le régime transitoire est plus long (environ 1 s) pour assurer une augmentation progressive de la vitesse du rotor.

La méthode de reconnaissance des appareils est une méthode d'identification qui exploite les motifs des transitoires des différents appareils. Cette méthode est robuste, sous réserve que le bruit de mesure ne soit pas très important. Elle consiste en deux étapes : détecter les événements sur les signaux transitoires et étiqueter ces événements en leur associant un type d'appareil donné. Le système développé est dénommé « Transient-Non-Intrusive Load Monitoring ».

Détection des événements en régime transitoire (Transient Detector)

La détection des événements est effectuée en deux étapes. La première est réalisée par un pré-processeur. C'est une étape de prétraitement, qui permet de calculer les séries temporelles sur lesquelles on peut détecter les événements. La deuxième phase consiste à détecter les enclenchements et les arrêts des usages.

Première étape : calcul des puissances active et réactive (aux différentes harmoniques)
Le système SS-NILM détecte les événements à partir des séries temporelles constituées des variations de la puissance active et des variations de la puissance réactive échantillonnées à une fréquence de l'ordre d'un hertz. La détection des événements en régime transitoire fait intervenir les enveloppes spectrales, qui sont proportionnelles aux puissances harmoniques définies par l'équation (B.3). Plus précisément, il s'agit des coefficients de la série de Fourier de l'intensité de courant. En pratique, ces coefficients sont calculés à partir des séries temporelles discrètes de courant et de tension, échantillonnées à une fréquence $f_e = 8$ kHz. Ces coefficients renseignent sur l'évolution au cours du temps du déphasage entre le courant et la tension. De plus, ils renseignent également sur le contenu fréquentiel au cours du temps du signal de courant. L'analyse des puissances harmoniques permet de discriminer certains usages, qui ne peuvent être séparés à partir de la puissance active et de la puissance réactive, comme l'illustre l'exemple d'une lampe à incandescence fonctionnant simultanément avec un ordinateur (figure C.5). Cet exemple est extrait de [LLC+03].

Deuxième étape : détection des événements
Un événement se traduit souvent par une variation de l'amplitude du signal observé (enveloppe spectrale), et dans certains cas seule la variance du signal mesuré varie. Un test d'hypothèse binaire basé sur le ratio de vraisemblance généralisé permet de détecter les séries de (on, off) qui se traduisent uniquement par une rupture de la moyenne. Un détecteur basé sur un test d'hypothèse composite de rupture de la moyenne et de la variance a été proposé [Lee93, LMLS93, Lee03] pour détecter les événements liés à une rupture de la moyenne et/ou de la variance du signal analysé. Ce détecteur est détaillé dans [MEG07]. Deux améliorations ont été proposées pour accroître les performances du détecteur en présence de bruit, à savoir utiliser un filtre médian pour filtrer les oscillations et les pics d'appels de courant [NL96] ou appliquer le détecteur à différentes échelles temporelles [Don02].

Classification / Identification des événements en régime transitoire

Les méthodes existantes de reconnaissance des signatures des appareils électriques en régime transitoire sont fondées sur les deux conditions nécessaires suivantes : *l'unicité de la signature électrique d'un appareil (en particulier du motif observé en régime transitoire) et la reproductibilité de cette signature*.

La méthode de classification exploite une bibliothèque de signaux collectés sur site. Les principaux classifieurs proposés [Sha00, Lee93] sont basés sur la segmentation du signal transitoire observé en « zones caractéristiques » dites $v - sections$ illustrées sur la figure C.6.

L'identification d'un motif est réalisée en minimisant un critère de moindres carrées qui permet de prendre en compte un facteur de proportionnalité (et/ou d'offset) ou un déphasage entre le motif à identifier et son homologue dans la base de référence. Cette procédure est appliquée à toutes les zones caractéristiques du signal observé.

Avantages et limites de l'identification des usages en régime transitoire

L'analyse des signaux électriques en régime transitoire offre plus de possibilités de discrimination des usages par rapport à la méthode fondée sur l'étude des signaux en régime permanent, et permet de tenir compte du modèle physique des appareils. Mais certains appareils ont des signatures très proches en régime transitoire, au regard du classifieur mis en œuvre, et des amplitudes différentes en régime permanent.

Notons par ailleurs que le classifieur proposé n'exploite ni des descripteurs des signaux transitoires ni des modèles de ces signaux. Il se limite à une comparaison exhaustive du motif par rapport à la base de référence.

Identification des usages électriques :utilisation conjointe des signatures en régime permanent et en régime transitoire

Une méthode hybride tirant profit des deux méthodes précédentes a fait l'objet des travaux récents [Lee03] au LEES. Le principe de la méthode est résumé sur la figure (C.7). Notons la présence des deux classifieurs correspondant aux deux régimes (transitoire et permanent), ainsi qu'une procédure d'arbitrage entre les deux modules de classification pour résoudre les cas de contradiction entre les décisions obtenues par les deux méthodes.

Caractérisation des usages électriqes à partir de données autre que la CdC

Les méthodes qui seront présentées ci-après visent à estimer des paramètres de la CdC et des composantes élémentaires à partir de différentes données d'entrée de nature physique ou climatique par exemple, et éventuellement de données de consommation d'électricité. Ces méthodes sont réparties en trois classes en fonction des techniques utilisées : les méthodes dites « engineering approaches », les méthodes statistiques et les méthodes hybrides.

Approche empirique fondée sur les données contextuelles

Il s'agit d'une approche qui combine des connaissances *a priori* ou des hypothèses sur la fréquence d'utilisation des appareils électriques et sur les consignes de température pour certains usages, couplées avec un modèle thermodynamique du bâtiment (du local) étudié. Les principaux systèmes (autres que ceux développés à EDF) utilisant cette approche (jusqu'à 1999) sont :

- HOT2000 : ce système développé au CTEC-CANMET [CTE95] est principalement utilisé dans le secteur résidentiel pour l'estimation de la consommation du chauffage électrique et de l'eau-chaude sanitaire.
- DOE2.1E : ce système a été développé à l'université de Californie (Lawrence Berkley Laboratory) en 1991 [GL91]. Il s'agit d'un simulateur très adapté à la simulation de la consommation globale d'un bâtiment ou d'une agrégation de plusieurs logements ou de sites industriels.

Cette approche est intrusive, dans la mesure où elle nécessite un audit du bâtiment ; ce qui constitue un inconvénient majeur pour une offre de services à grande échelle dans le secteur résidentiel.

Les approches statistiques

Il s'agit des approches qui s'appuient sur la description de chaque usage par un modèle paramétré par des données économiques et climatiques, des caractéristiques physiques de l'usage, des information sur les occupants et sur le type de logement. Un historique des consommations est obtenu soit à partir de données issues d'intervention sur site (cas du secteur tertiaire) soit à partir des bases de données renseignant sur la consommation globale du client. Les trois principales méthodes mises en œuvre sont l'analyse conditionnelle de la demande, la régression multi-variée et la classification.

Remarque

Les approches basées sur une modélisation statistique des grandeurs électriques ne feront pas partie de cette section.

Analyse conditionnelle de la demande (Conditional Demand Analysis : CDA) [ASK84, LN04]

Cette méthode consiste à construire un modèle économétrique du logement du client particulier (données démographiques et économiques). Elle est principalement utilisée pour l'agrégation des consommations ou le profilage des consommateurs, et non pour l'estimation de la consommation par usage. Ce modèle a notamment été utilisé pour la décomposition de la courbe de charge dans le secteur résidentiel, principalement au Canada (Modèle d'estimation de la consommation d'électricité, étude sur le comportement énergétique des ménages québécois).

Remarquons qu'aucune mesure de nature électrique n'est exploitée pour estimer les paramètres recherchés.

Régression multivariée (Multivariate Regression Analysis)

Un modèle de régression en fonction de la température, de l'ensoleillement et d'autres facteurs climatiques est défini. Les paramètres du modèle sont estimés par minimisation d'un critère quadratique. Cette méthode est simple, mais les résultats d'estimation de la consommation par usage sont souvent médiocres. Cette approche est implémentée dans le logiciel PRISM [Fel86] développé à l'université de Princeton.

Classification (Objective Classification Analysis)

Cette méthode [Had93] combine l'analyse en composantes principales (ACP) et une méthode de clustering pour identifier les jours où la consommation est corrélée au climat, puis pour les classifier par jour type dépendant du climat. Elle a été initialement développée pour la prédiction des consommations, en particulier pour les systèmes **HVAC** (**H**eating **V**entilation **A**ir-**C**ondtionning) (pour les industriels et les commerces).

Des méthodes similaires exploitant des techniques de classification par arbre, des réseaux de neurones ont été développées au sein d'EDF pour répondre à des besoins tels que la prévision des consommations ou le profilage des consommateurs.

Les approches hybrides

Ce sont des méthodes qui combinent des techniques utilisées dans les deux approches précédentes afin d'améliorer l'estimation des usages. Les principaux modèles utilisés pour la décomposition de la courbe de charge, autres que ceux proposés par EDF, sont les suivants (C.8) :
➢ le modèle SAE [Tra92] (Statistically Adjusted Engineering Model) ;
➢ le modèle RECAP [Xen](Residential Energy Consumption Analysis Program) ;
➢ le modèle REEPS [Koo95](Residential End-Use Planning Systems) ;
➢ et le modèle EDA [Akb95](End-Use Disaggregation Algorithm).
Une approche hybride exploitant des connaissances *a priori* sur le modèle physique et le mode de fonctionnement de l'usage ciblé a également été développée au sein d'EDF-R&D [Boz00]. Plus précisément, il s'agit d'une méthode d'estimation des consommations électriques et de la puissance nominale d'une installation d'eau chaude sanitaire (ECS), asservie au tarif heure creuse (HC), à partir de la courbe de charge en sortie du CBE. La méthode mise en œuvre utilise essentiellement les caractéristiques électriques et des hypothèses sur le mode de fonctionnement de l'installation ciblée :
➢ l'heure d'enclenchement théorique (fixée par le tarif HP/HC [3]) ;
➢ l'enclenchement de l'ECS est le plus important au cours de la nuit.
Les résultats de cette étude sont :
➢ Estimation de la puissance nominale : erreur d'envrion 6%.
➢ Estimation de l'énergie consommée par l'ECS : erreur de $\pm 15\%$.
Au cours de cette étude, une classification des courbes de charge à partir du « courboscope » [Heb, Kop03], outil statistique développé à EDF-R&D, a également été réalisée. La classification, utilisant un algorithme de type cartes auto-organisatrices de Kohonen, a permis :
➢ d'obtenir une répartition en 3 classes principales ;
➢ et d'extraire des information susceptibles d'être utilisées pour améliorer les algorithmes d'estimation des consommations d'usages.
D'autres services ont été proposés, et exploitent les fonctions de communication du CBE. Un service particulier consiste à identifier les cycles d'enclenchements et de déclenchements (on, off) d'un ensemble de convecteurs. Cette fonction est réalisée en prenant en compte différents paramètres relatifs au bâtiment (déperditions thermiques, inertie thermique du local, etc) et des constantes de temps caractéristiques des convecteurs.

Modèle RECAP

La société XENERGY a mis au point au début des années 1990 le programme d'estimation RECAPtm destiné à estimer la consommation des appareils et équipements d'une habitation pendant une période donnée. Les données d'entrée de la méthode sont les suivantes :
➢ un questionnaire d'enquête détaillé qui fournit les caractéristiques de l'habitation et du ménage ainsi que la liste et le mode d'utilisation des appareils électriques (usages) ;

3. Heures Pleines/Heures Creuses

- ➢ des informations liées à la facturation (relevés estimés du compteur, éléments de facturation du client, etc.) ;
- ➢ les conditions météorologiques (température extérieure, etc) ;
- ➢ et d'autres informations sur les prix de vente des appareils électriques sur le marché.

La méthode consiste à estimer les consommations unitaires grâce à une modélisation statistique en fonction des caractéristiques du client.

- ➢ Découpage de la courbe de charge de chaque client
  C'est l'estimation par client des charges de climatisation (chauffage et rafraîchissement) non sensibles à la température en exploitant le modèle suivant $C_{ip} = \beta_1 NJ_{ip} + \beta_2 CH_{ip} + \beta_2 CL_{ip} + \varepsilon_{ip}$ pour un client i
  où

C_{ip} est la consommation d'énergie pour la période p ;

NJ_{ip} est le nombre de jours sur la période p ;

CH_{ip} est le nombre de degré-jours de chauffage pour la période p ;

CL_{ip} est le nombre de degré-jours de rafraîchissement pour la période p ;

β_1, β_2 et β_3 sont des coefficients du modèle et ε_{ip} est le bruit résiduel.

- ➢ Estimation d'un modèle d'utilisation par appareil et par habitation sur la base de données techniques et des modèles statistiques ;
- ➢ Phase de réconciliation des différents éléments ;
  - – Il s'agit de réconcilier les modèles d'utilisation, les estimations tirées de la facturation et les factures proprement dites. Cette réconciliation est réalisée par un algorithme qui cherche à maximiser ou minimiser une fonction économique sous la contrainte du recalage de la consommation totale estimée avec la consommation totale observée pour chaque client. Le résultat est une estimation en BTU pour une période donnée pour l'appareil (ou l'usage) l du client i, les BTU estimés peuvent être convertis en kWh.

Synthèse et analyse des méthodes présentées

Les méthodes présentées exploitent des données d'entrée de nature non-électrique et éventuellement la consommation en électricité d'un usage particulier ou de l'installation globale. Les trois approches sont résumées sur la figure (C.8). Soulignons que les finalités sont principalement la prédiction de la consommation ou l'estimation des usages électriques chez un client particulier ou tertiaire. Les méthodes d'estimation des usages nécessitent des connaissances *a priori* sur le client, ses usages ou encore l'audit du site ou l'instrumentation de certains usages. Les méthodes de prédiction sont semi-intrusives et utilisent principalement des informations contextuelles, alors que les méthodes d'estimation des usages sont intrusives et peu robustes. Ces méthodes ont l'avantage de prendre en compte des données qui constituent en effet des variables explicatives de la consommation totale d'un client donné ainsi que son comportement "énergétique". Bien qu'elles peuvent permettre de caractériser la CdC au sens de détection d'anomalies ou de prédiction de la CdC d'un site donné, elles ne peuvent être envisagées pour une décomposition suffisamment précise de la courbe de charge quotidienne d'un client particulier.

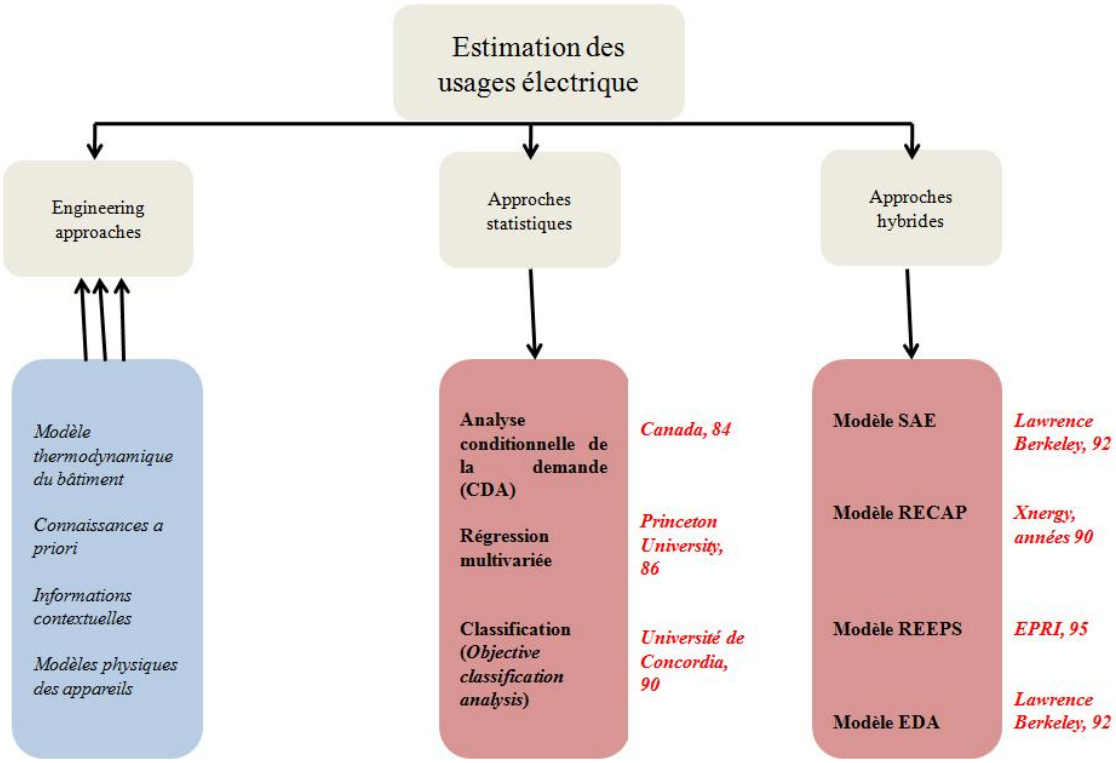


FIGURE C.8: Estimation des usages : première famille de méthodes

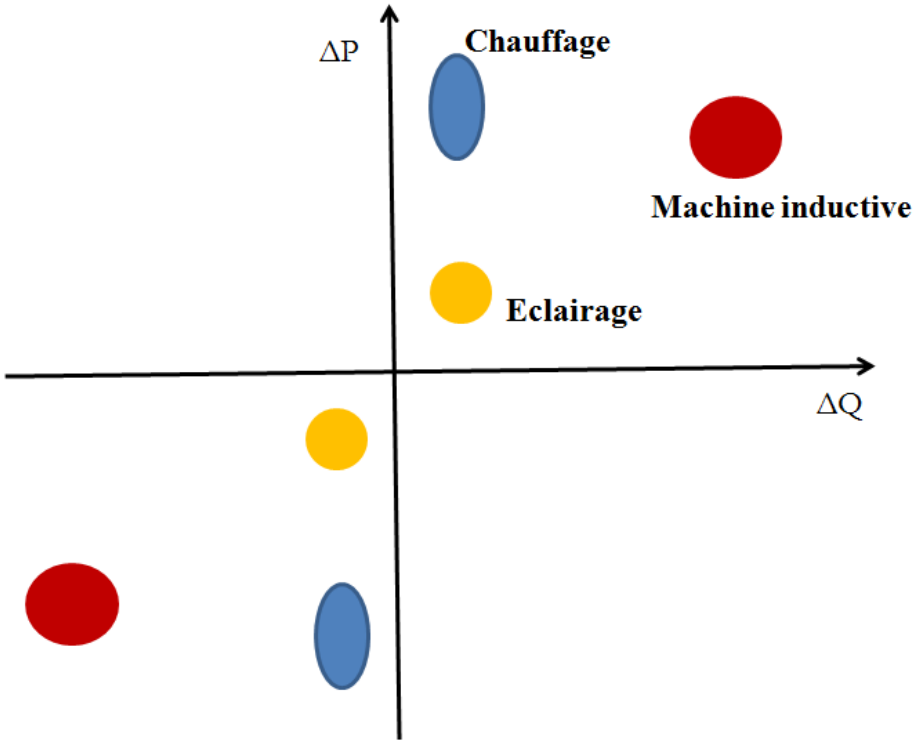


FIGURE C.2: Schéma de principe illustrant des exemples de signatures électriques dans le domaine $(\Delta P, \Delta Q)$

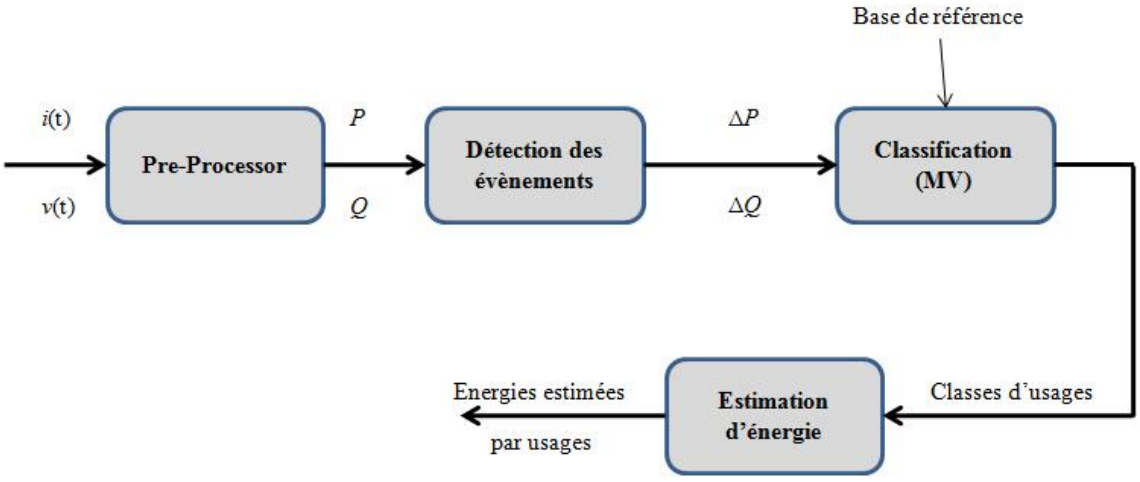


FIGURE C.3: Architecture du software du système NILM (régime permanent) de la société ENETICS

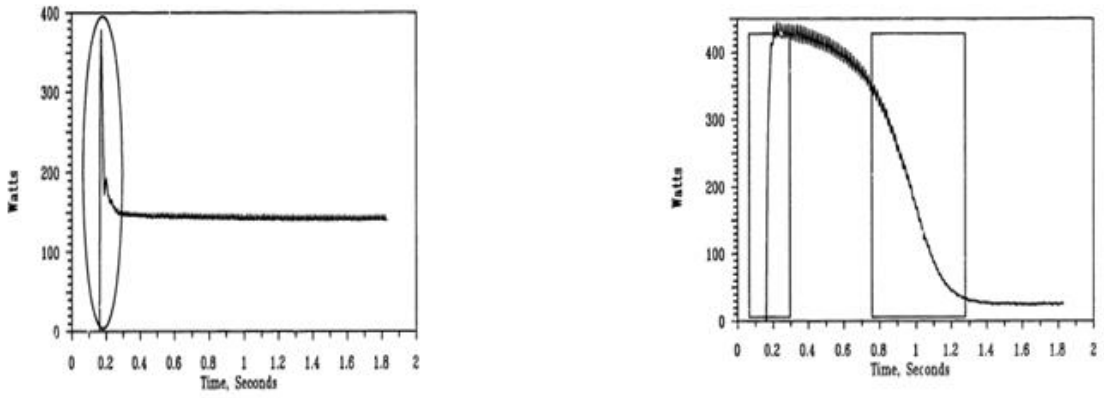


FIGURE C.4: Puissance active d'une lampe et d'un moteur (120 Hz) (Extrait de [Lee93])

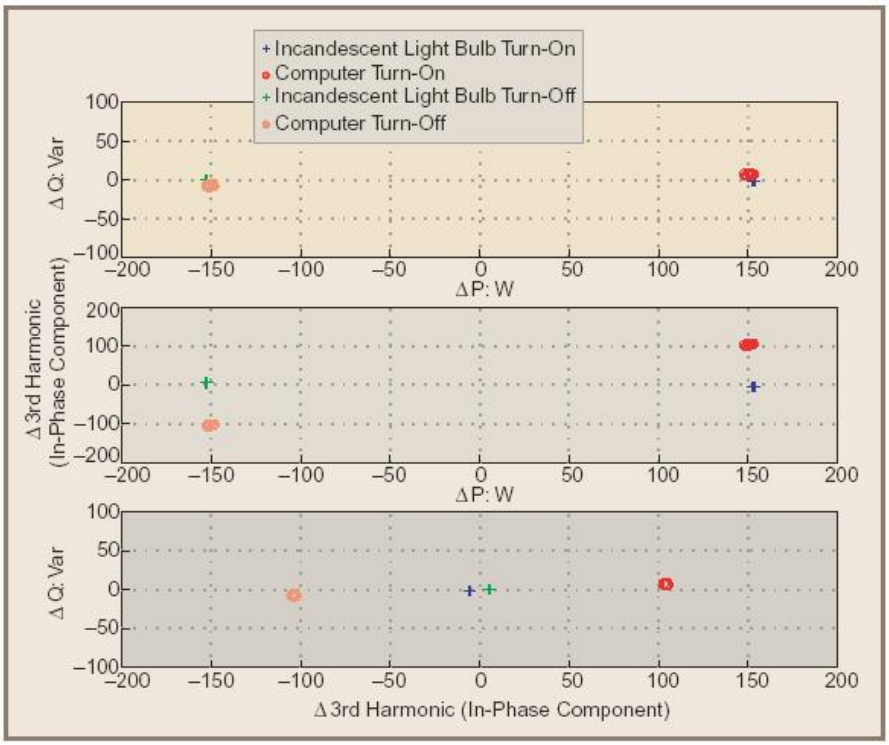


FIGURE C.5: Séparation d'usages basée sur les puissances harmoniques de rang 3 (Extrait de [LLC+03])

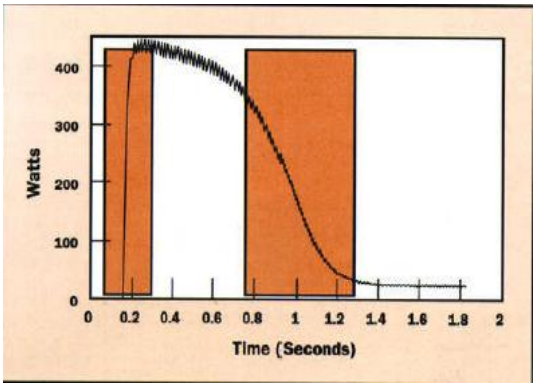


FIGURE C.6: ν – *sections* d'un moteur à induction (Extrait de [LMLS93])

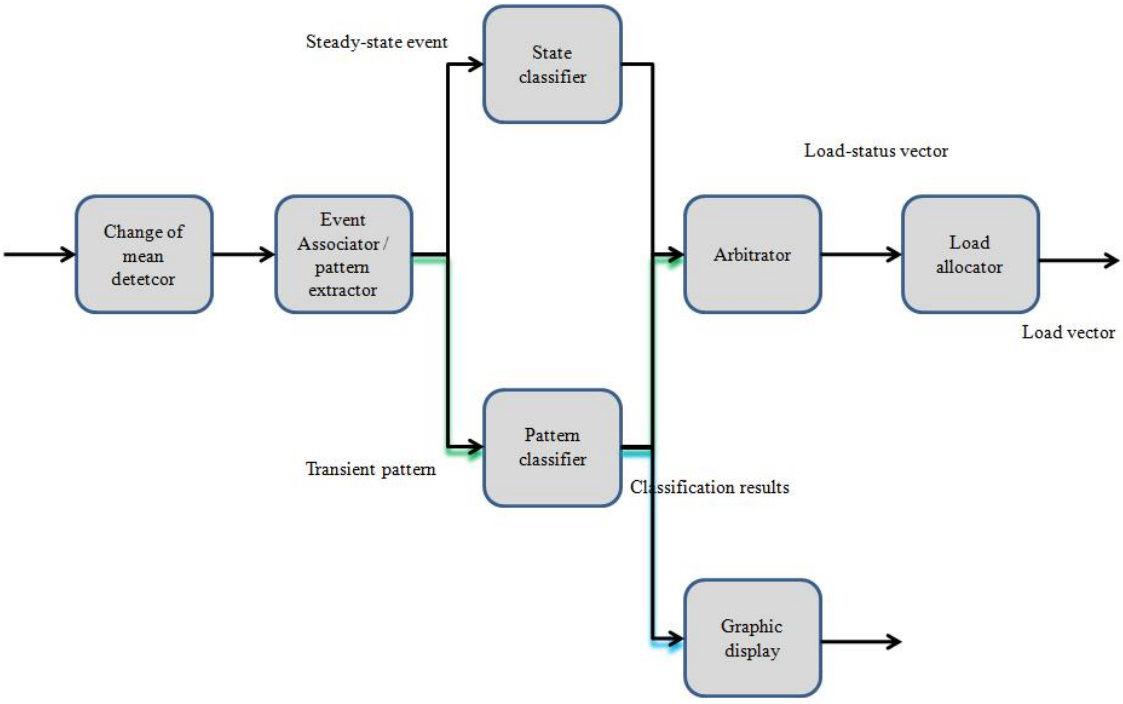


FIGURE C.7: Méthode hybride de décomposition de la CdC électrique

Annexe D

Thermostatically Controlled Appliances

Chauffe-eau

Il existe différents systèmes d'eau chaude sanitaire : à accumulation, solaire et thermodynamique. On s'intéresse en particulier aux chauffe-eau à accumulation. Le système ciblé est consituté d'un réservoir équipé d'une résistance électrique pour le chauffage de l'eau. Un thermostat permet de régler la témpérature de l'eau à l'intérieur du réservoir et assure une sécurité contre la sur-chauffe. La durée de chauffage de l'eau dépend en particulier du tarif du cient (tarif de base ou tarif Heures Creuses / Heures Pleines). Un schéma de principe d'un chauffe-eau est présenté sur la figure (D.1) :

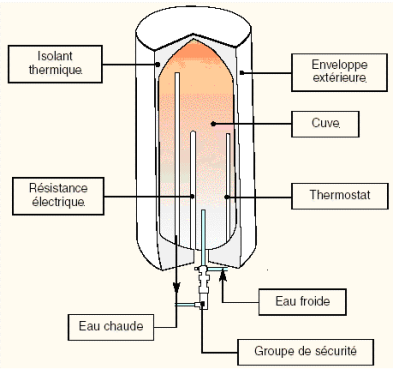


Figure D.1: Chauffe-eau à accumulation

Froid alimentaire

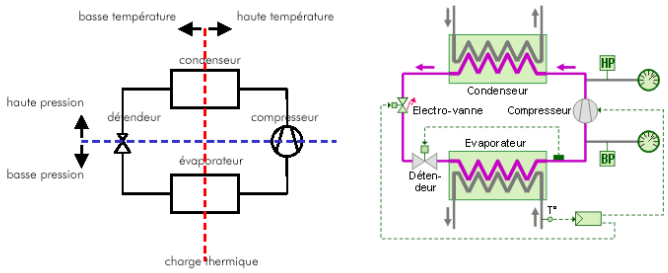


Figure D.2: Schéma de principe d'un système à compression

Annexe E

Approches pour la décomposition des signaux composites

Cette annexe présente les approches probabilistes et les approches de séparation de sources pour la décomposition d'un mélange unique plusieurs sources.

Approches probabilistes pour la décomposition de mélanges mono-valués

Nous avons vu que résoudre un problème inverse mal posé au sens de Hadamard [Had76] nécessite de compléter notre connaissance du problème et des inconnues. Rappelons qu'un problème inverse est dit bien posé au sens de Hadarmard si les trois propriétés suivantes sont vérifiées : existence d'une solution, unicité et continuité de la solution. Les connaissances complémentaires sont ensuite utilisées soit dans un cadre déterministe soit dans un cadre probabiliste. En effet, deux points de vue sont possibles [Ia01] :

- ➢ Les informations complémentaires permettent de restreindre l'espace de recherche des solutions et de contraindre la solution. C'est le point de vue des approches classiques dites approches régularisantes.
- ➢ Les information acquises sur les inconnues du problème sont utilisées pour rendre le problème bien posé dans un espace plus large, qui est un espace des distributions de probabilité. C'est le point de vue des approches probabilistes, dont les approches bayésiennes.

Dans ce chapitre, nous commençons par présenter le formalisme probabiliste pour les problèmes inverses. Nous présenterons ensuite les principaux estimateurs dans un cadre bayésien ou non.

Formalisme probabiliste pour les problèmes inverses

L'approche probabiliste pour l'inversion est fondée sur les principes suivants [KS] :

- ➢ Toutes les variables du problème sont modélisées par des variables aléatoires.
- ➢ L'aléa permet de décrire le «degré» d'information / de connaissance que nous avons sur les réalisations de ces variables.

- La connaissance que nous avons des variables est décrite par l'intermédiaire de distributions de probabilité. Cela signifie que l'espace des inconnues doit être spécifié de manière exhaustive dès le début.
- La solution du problème inverse est la distribution *a posteriori* du paramètre recherché dans le cas de l'inférence bayésienne.

Inférence statistique

L'inférence peut se résumer par la question : «Comment tirer les meilleures conclusions de l'information incomplète disponible» [DG01]. Cette approche s'applique autant au problème d'estimation qu'au problème de choix de modèle.

La première étape consiste à définir l'espace d'état, qui est l'espace des inconnues du problème, appelé parfois *espace des épreuves* quand il s'agit de l'espace des résultats possibles Ω et *espace des hypothèses* $\mathscr{H}$ s'il décrit les hypothèses que l'on souhaite vérifier.

Soit $y \in \mathbb{R}^m$ ($m = 1$ dans le cas de la CdC) le signal mesuré souvent entaché d'un bruit. Ce signal est une mesure indirecte de $x \in \mathbb{R}^n$(où l'on confond composantes du mélange et paramètres). Cette quantité est accessible par l'intermédiaire d'un modèle, qui s'écrit classiquement dans le cadre des problèmes inverses $y = f(x,e)$ où :

- $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ est le modèle reliant les observations aux inconnues ;
- $e \in \mathbb{R}^k$ est un vecteur qui contient le bruit (mesure, modélisation) et les paramètres mal connus [1].

On se limitera au cas de bruit additif.

Dans le cadre d'une approche probabiliste, le modèle de formation des données est décrit par l'équation (**??**). Par conséquent, les distributions de probabilité des variables aléatoires X, Y et E sont reliées.

$$Y = f(X) + E \tag{E.1}$$

Une étape essentielle dans une démarche d'inférence est la description de l'incertitude sur le bruit e par une densité de probabilité du bruit $e \longmapsto \Pi_{bruit}(e)$. On peut ensuite déduire la distribution directe des données ou vraisemblance des données ainsi :

$$\Pi(y \mid x) = \Pi_{bruit}(y - f(x) \mid x) \tag{E.2}$$

Cette fonction peut s'exprimer sous forme d'un modèle paramétrique. Les modèles les plus courants sont le modèle gaussien (MG), les modèles de mélanges de gaussiennes (MMG), les modèles de chaînes de Markov Cachées (MMC) [Rab89] et les processus ponctuels marqués (PPM) [Lie00].

Etant donnée cette seule distribution, une solution du problème inverse est celle du *maximum de vraisemblance* (MV) , où y prend la valeur observée y_{obs} et x est la variable.

$$\hat{x}_{MV} = \underset{x \in \mathscr{H}}{argmax}\ \Pi(y = y_{obs} \mid x) \tag{E.3}$$

1. pour lesquels on dispose de très peu ou pas du tout d'information

Approche bayésienne

Souvent, nous disposons d'informations supplémentaires sur X issues de différentes sources (contexte, données d'apprentissage, expertise, ...). Dans un cadre bayésien, on suppose que cette information peut être codée par une densité de probabilité $x \longmapsto \Pi_{pr}(x)$ appelée *densité a priori*. Cette densité traduit la connaissance *a priori* sur l'inconnue du problème.

Loi marginale

On considère que la probabilité conjointe de X et Y est définie après avoir étudié les données observées et les *a priori* sur les paramètres. La densité marginale de l'inconnue X doit vérifier la relation suivante :

$$\int_{\mathbb{R}^m} \Pi(x,y)\, dy = \Pi_{pr}(x) \tag{E.4}$$

Fonction de vraisemblance

Si x est une réalisation de X, la probabilité conditionnelle de Y étant donnée cette information définie par la relation (E.5) est appelée fonction de vraisemblance.

$$\Pi(y \mid x) = \frac{\Pi(x,y)}{\Pi_{pr}(x)} \tag{E.5}$$

Distribution *a posteriori*

Soit y_{obs} une observation (réalisation) de la variable aléatoire Y. La loi conditionnelle définie par la formule (E.6) est la distribution *a posteriori* de X.

$$\Pi(x \mid y_{obs}) = \frac{\Pi(x,y_{obs})}{\Pi(y_{obs})} \tag{E.6}$$

Cette distribution décrit la connaissance que nous avons sur X après une réalisation de Y.

Théorème de Bayes (Problèmes inverses)

Dans un cadre bayésien, le problème inverse est défini ainsi : «étant donné $Y = y_{obs}$, trouver la distribution *a posteriori* $\Pi(x \mid y_{obs})$ de la variable X», ce qui se résume par le théorème de Bayes décliné dans le cadre des problèmes inverses.

Théorème [KS]

Supposons que la variable aléatoire X a une densité de probabilité $\Pi_{pr}(x)$ et que les données se résument par y_{obs}, réalisation d'une variable aléatoire observable Y tel que $\Pi(y_{obs}) > 0$. La distribution a posteriori de X étant donné y_{obs} est définie ainsi :

$$\Pi_{post}(x) = \Pi(x \mid y_{obs}) = \frac{\Pi_{pr}(x)\,\Pi(y_{obs} \mid x)}{\Pi(y_{obs})} \tag{E.7}$$

Inférence bayésienne pour les problèmes inverses : démarche

Dans un cadre bayésien, un problème inverse peut être abordé en trois étapes.

1. *Définition des densités a priori* : Il s'agit de formaliser toute connaissance *a priori* sur X par une densité *a priori* Π_{pr} qui traduit judicieusement l'information disponible.
2. *Définition de la fonction de vraisemblance* : Trouver la fonction $\Pi(y \mid x)$ qui décrit au mieux la relation entre les inconnues et les observations.
3. *Estimation de la distribution a posteriori* : Déduire de 1) et 2) la densité *a posteriori* $\Pi_{post}(x)$ des paramètres. Ici l'enjeu est souvent la méthode d'échantillonnage utilisée pour approcher la densité de probabilité et en déduire un estimateur ponctuel (par exemple le maximum *a posteriori*).

Remarques

➢ Notons que le maximum de vraisemblance est un cas particulier du maximum *a posteriori* sous l'hypothèse d'uniformité de la densité *a priori*.
➢ Quelques définitions des estimateurs bayésiens sont fournies en annexe (**??**)..

Echantillonnage stochastique

Nous avons vu que résoudre un problème inverse mal posé dans un cadre bayésien conduit soit à un problème d'optimisation (maximum *a posteriori*) ou encore à un problème d'intégration (moyenne *a posteriori* par exemple) sur des espaces de grande dimension. Il faut alors disposer d'outils performant pour explorer l'espace des paramètres (espace d'état) et pour s'affranchir des difficultés calculatoires. L'échantillonnage stochastique (génération pseudo-aléatoire de réalisations d'une loi de probabilité) cons
titue un cadre approprié pour contourner ces problèmes.

Lorsque les quantités à générer sont de faible dimension, il existe des méthodes directes pour simuler différentes lois de probabilités. La plus classique consiste en un changement de variable à partir d'une loi dont l'échantillonnage est maîtrisé, telle que la loi uniforme.

Dans les cas plus complexes, une des techniques courante est l'échantillonnage par chaîne de Markov.

Méthodes de Monte Carlo par Chaînes de Markov

Intégration par méthode de Monte Carlo

Avant de présenter le principe de l'échantillonnage par les méthodes de Monte Carlo par Chaînes de Markov, nous rappelons ici le principe de l'intégration par la méthode de Monte Carlo.

Soit μ une mesure de probabilité sur $\mathbb{R}^n$ et g une fonction mesurable, intégrable sur $\mathbb{R}^n$ définie par rapport à la mesure de référence μ. L'objectif est de calculer l'intégrale de g. Une approche déterministe consiste à discrétiser le support de la fonction et à approcher l'intégrale recherchée comme suit :

$$\int_{\mathbb{R}^n} g(x)\,\mu(dx) \simeq \sum_{j=1}^{N} w_j\, g(x_j)$$

où $x_1, x_2, ..., x_N$ sont les points choisis sur le support de la fonction et $w_1, w_2, ..., w_N$ sont les poids associés.

La méthode de Monte Carlo consiste à générer les points support de manière aléatoire selon une certaine loi de probabilité. Les poids associés sont ensuite déterminés par la mesure μ.

Soit X une variable aléatoire dont la distribution est μ. Supposons que l'on dispose d'un générateur aléatoire permettant d'obtenir N réalisations $\{x_1, x_2, ..., x_N\}$ de X distribués selon la loi μ. L'intégrale recherchée peut alors être approchée de la manière suivante :

$$\int_{\mathbb{R}^n} g(x)\,\mu(dx) = E(X) \simeq \frac{1}{N}\sum_{j=1}^{N} g(x_j) \tag{E.8}$$

Principe des méthodes MCMC

Les méthodes de Monte Carlo par Chaînes de Markov constituent un outil pour générer des échantillons distribués suivant une loi cible.

L'échantillonnage par chaîne de Markov consiste à générer une suite $X^{(0)}, X^{(1)}, ...X^{(i)}, ...$ qui converge en loi vers la loi cible Π. Cette chaîne est générée de la manière suivante :

- ➢ $X^{(0)}$ est générée à partir d'une loi arbitraire ;
- ➢ les autres composantes sont obtenues de proche en proche à partir de transformations aléatoires judicieusement définies.

Chaque transformation est définie par une densité conditionnelle $K(x, x')$ appelée *noyau de transition*. Le noyau de transition doit permettre d'explorer l'espace d'état (ou encore la distribution cible). Pour cela certaines conditions (dont la justification est fournie en annexe) sont requises. En effet, la convergence en loi vers la loi cible dépend de la construction du noyau de transition, c'est-à-dire de la loi qui permet de générer $X^{(k+1)}$ connaissant $X^{(k)}$. Nous avons le résultat suivant, qui est un énoncé simplifié du théorème (3.21) de [Rob96]:

Soit $(X^{(k)})$ une chaîne de Markov homogène apériodique et irréductible, dont le noyau de transition admet la loi Π de densité μ pour loi invariante, alors $(X^{(k)})$ converge en loi vers μ.

L'homogénéité signifie que le noyau de transition n'évolue pas au cours des itérations. L'invariance signifie que si Π est la loi de $(X^{(k)})$, alors la loi de $(X^{(k+1)})$ est également Π. Ces conditions se trouvent souvent respectées par des choix intuitifs et simples.

En pratique, différentes procédures permettent de construire le noyau de transition qui a les «bonnes» propriétés souhaitées. L'une des procédures standards est l'algorithme de Metropolis Hastings.

Construction d'un noyau de transition : algorithme de type Metropolis-Hastings

De l'invariance de la chaîne définie par le noyau de transition K, découle la relation (E.9) appelée équation de balance :

$$\Pi(x)\,K(x,x') = \Pi(x')\,K(x',x) \tag{E.9}$$

L'algorithme de Metropolis-Hastings consiste à construire un noyau de transition vérifiant l'équation de balance, et dont le principe est de proposer une perturbation de la chaîne puis l'acceptation ou le refus de la proposition.

La mise en œuvre de cet algorithme est décrite par les étapes suivantes :

1. Initialisation : $k = 1$, choisir un état initial $\mathbb{R}^n$.
2. Générer $x' \in \mathbb{R}^n$ selon la densité intrumentale $q(x_k, x')$ et calculer le taux d'acceptation $\alpha(x_k, x') = \min\,(1, \frac{\Pi(x')\,q(x',x_k)}{\Pi(x_k)\,q(x_k,x')})$.
3. Générer $u \in [0,1]$ selon une loi uniforme.
4. Si $\alpha(x_k, x') \geq u$ alors $x_{k+1} = x'$. Sinon, $x_{k+1} = x_k$.

Remarques

➢ Si le noyau de transition est symétrique, alors $\alpha(x_k, x') = \min\,(1, \frac{\Pi(x')}{\Pi(x_k)})$. La marche aléatoire est un cas particulier pour lequel la propriété de symétrie est vérifiée.

➢ D'autres algorithmes du type Metropolis-Hastings existent. Geyer et Moller ont proposé un algorithme (GM) avec un noyau de type naissance et mort afin d'explorer des sous ensembles de l'espace d'état (naviguer dans différentes dimensions de l'espace d'état). Un autre algorithme, celui de Metropolis-Hastings-Green, autorise plusieurs perturbations assurant ainsi une meilleure mélangeance de la chaîne.

Optimisation

La recherche de la solution au sens du maximum a posteriori conduit à un problème d'optimisation, où le critère à optimiser présente souvent des minima locaux. Par ailleurs, dans les applications industrielles concernées, l'espace d'état est un espace de dimension finie et grande. Le recuit simulé [KG83] est un algorithme d'optimisation stochastique qui permet d'échapper au piège de minima locaux.

La structure de base est de simuler non plus une chaîne de Markov $(X^{(k)})$ homogène convergeant vers la loi cible Π, mais une chaîne non-homogène par la décroissance lente d'un paramètre de température au cours des itérations. Soit f la densité de la loi à maximiser. On introduit :

$$f_T(x) = \frac{1}{Z(T)}(f(x))^T$$

où T est un paramètre de température et $Z(T)$ une constante de normalisation. Soit $K_T(x,x')$ un noyau de transition permettant d'échantillonner f_T.

Les deux principales propriétés du recuit simulé sont les suivantes :

➢ Echantillonnage inhomogène : il s'agit de générer une suite aléatoire $(X^{(k)})_{k\in\mathbb{N}}$ suivant le noyau de transition $K_{T(k)}(x,x')$ où $(T(k))_{k\in\mathbb{N}}$ est une suite déterministe décroissante et tendant vers zéro. Lorsque la température initiale est suffisamment haute et si la décroissance est suffisamment lente, alors la loi de $(X^{(k)})$ est proche de $f_{T(k)}$ lorsque k est suffisamment grand.

➢ Lien entre la densité échantillonnée et le MAP : à température élevée, cette densité devient uniforme, à très basse température, elle se concentre autour du maximum *a posteriori*.

Le recuit simulé permet d'explorer plusieurs minima locaux en acceptant parfois des perturbations qui augmentent l'énergie du modèle.

Méthodes de séparation de sources mono-capteur pour la décomposition de mélanges mono-valués

Généralités

La séparation de sources consiste à estimer un ensemble de signaux dont on observe un ou plusieurs mélanges. L'exemple traditionnellement présenté dans la littérature est celui du *« cocktail party »* : plusieurs personnes parlent en même temps et les signaux sont captés par des micros.

La démarche de la séparation de sources peut être résumée de la manière suivante (E.1) :

➢ *définir une mesure de contraste (mesure de diversité) ;*

➢ *puis séparer les sources selon un critère d'optimalité choisi.*

La méthode la plus connue est l'analyse en composantes indépendantes (ACI) [Com90, HP00] fondée sur l'indépendance statistique entre les sources. Cette méthode a fait ses preuves pour de nombreuses applications en traitement du signal et des images. Nous nous limitons dans ce chapitre au cas de mélange linéaire instantané avec une seule observation (mesure), cadre dans lequel s'inscrit la problématique de séparation des composantes constitutives de la CdC à partir de la puissance active. Ce cas particulier est un problème difficile à résoudre que nous rencontrons dans différents domaines, dont trois exemples [PGB, Che06, Peu08] sont résumés dans le tableau (E.1).

Plusieurs travaux menés depuis les années 2000 ont permis de définir une méthodologie pour séparer les signaux dans le cas particulier d'une seule observation, qui est devenu un axe de recherche à part entière. Précisons d'ores et déjà que ce cas a principalement été traité pour les signaux sonores (parole, musique).

Paradigme de séparation de sources mono-capteur

Dans le cas d'un modèle de mélange linéaire instantané, le problème s'écrit comme suit : $y(t) = \sum_i a_i x_i(t) + b(t)$. Les premiers travaux sur le cas mono-capteur sont apparus dans le cadre de l'approche **CASA** (**C**omputational **A**uditory **S**cene **A**nalysis)[2]. Depuis les années 2000, plusieurs chercheurs se penchent sur ce problème principalement dans le cadre du traitement de la parole.

Le cadre général de l'analyse en composantes indépendantes ne permet pas de résoudre le cas mono-capteur, car la matrice de mélange dans ce cas est «complètement dégénérée». Une séparation aveugle devient inenvisageable. Une phase préliminaire d'apprentissage / d'entraînement sur des signaux séparés est incontournable afin d'apporter des informations complémentaires.

2. Cette approche vise à reproduire ou à approcher notre manière de percevoir les sons sous forme d'algorithmes. L'analogie réside dans le fait que l'oreille est capable de distinguer différents instruments de musique dans le cas d'un enregistrement unique polyphonique ou composite.

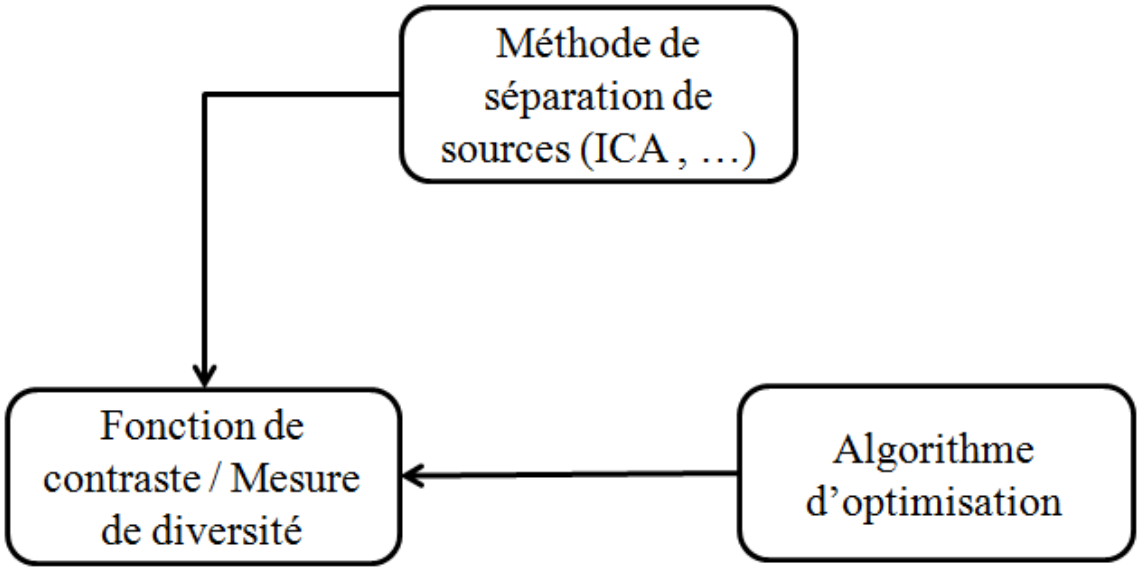


FIGURE E.1: Schéma de séparation de sources

Domaine	Mélange (un exemple)	Objectif	Approche utilisée
Traitement de la parole	Signal audio : - Plusieurs locuteurs - Musique et voix chantée	Séparation des signaux	Représentations parcimonieuses
Ingénierie biomédicale	Electromyogramme	Classification d'évènements	Classification supervisée & décomposition par paquets d'ondelettes
Contrôle non-destructif (ultrasons)	Ascan / Bscan : échos de la structure du composant mélangé aux échos générés par la présence de défauts	Extraction des échos générés par les défauts	Débruitage & paquets d'ondelettes

TABLE E.1: Exemples de problèmes de séparation de sources à partir d'un seul capteur

Comme précisé précédemment, la théorie bayésienne semble être un cadre propice pour résoudre ce type de problème. Nous montrerons que toutes les approches proposées dans la littérature s'inscrivent dans ce cadre générique, qui permet d'intégrer toutes les connaissances apprises au préalable. Le paradigme de séparation de sources à partir d'un seul capteur se résume ainsi en deux étapes :

- **l'apprentissage (hors-ligne)** : cette phase est réalisée séparément sur les différentes sources. Elle peut porter sur des modèles de sources ou encore sur leurs caractéristiques intrinsèques. D'un point de vue problème inverse, cette étape correspond à la phase de collecte d'information a priori et de formalisation de ces connaissances *a priori*.
- **l'estimation des sources (en ligne)** : les sources sont estimées en minimisant un critère dont la définition et la minimisation dépendent de la méthode de séparation à proprement parler.

La phase d'apprentissage sert souvent à choisir le domaine transformé dans lequel le signal composite sera analysé ainsi que la méthode d'estimation des sources. Deux cas de figure se présentent.

- Il existe un modèle des sources dans le domaine transformé : dans ce cas la séparation des sources est effectuée dans le domaine transformé en utilisant des méthodes à base de modèles statistiques.
- Les sources possèdent des propriétés permettant de les discriminer directement dans ce domaine telles que la parcimonie et la diversité morphologique. Les méthodes de décomposition adaptative de signaux sont alors utilisées.

  Dans les deux cas, une fonction de contraste est définie. La reconstruction des sources dépend de la nature de la transformée utilisée, plus précisément de son caractère inversible ou non.

La figure (E.2) résume ces deux options.

Les approches proposées peuvent être classifiées en fonction de deux critères : le choix du domaine transformé et l'utilisation ou non d'un modèle du signal composite dans ce domaine. Dans la suite, nous détaillerons uniquement deux classes de méthodes : les méthodes à base de modèles de mélange de lois et celles fondées sur la parcimonie ou la diversité morphologique. Ces méthodes ont fait leur preuves sur des données réelles tels que les signaux de parole ou encore des images.

Méthodes à base de modèles statistiques dans un domaine transformé

Généralités Le modèle décrit par l'équation (E.10) présente le problème dans le domaine « direct » d'observation des données. Afin de pallier le manque d'information sur les sources, certains chercheurs ont proposé de traiter le problème de séparation de sources mono-capteur dans un domaine transformé (par exemple la Tranformée de Fourier Court Terme) et d'aborder le problème dans un cadre bayésien.

$$y(t) = \sum_i x_i(t) + b(t). \tag{E.10}$$

Plus précisément, les signaux $x_i(t)$ sont des réalisations de différents processus stochastiques. Ces réalisations sont soumises à une transformée F qui peut être **redondante ou non**, mais souvent **linéaire**. Un modèle statistique est alors défini dans le domaine transformé en s'appuyant sur une base de signaux d'entraînement (apprentissage) qui guide le choix de la transformation

et du modèle approprié. L'estimation des sources réalisée en ligne consiste souvent à minimiser l'espérance d'une mesure de distorsion entre la source ciblée non-observée et son estimation.

Les méthodes s'inscrivant dans cette démarche (E) constituent les méthodes de séparation de sources monocapteur à base de modèles statistiques *a priori*.

Le choix de la transformée dépend des caractéristiques intrinsèques des signaux sources. Le choix du modèle s'en déduit naturellement.

Exemples de transformées La transformée permet de coder la source dans un domaine judicieusement choisi qui définira le choix du modèle des sources.

Rappelons qu'une proportion importante des chercheurs ayant travaillé sur ce sujet se sont intéressés au cas des signaux audio.

Les transformées utilisées sont souvent des représentations temps-fréquence / temps-échelle adaptées à cette classe de signaux.

Exemples de modèles Nous citons par exemple les modèles de mélange de gaussiennes (**GMM**) [BB03], et plus généralement des modèles de chaînes de Markov cachées (**HMM**) à loi conditionnelle gaussienne. Ces modèles sont couramment utilisés pour les signaux sonores. En effet, les GMM permettent de prendre en compte la diversité des motifs spectraux d'un son. L'utilisation des HMM est classique en reconnaissance de la parole.

Une extension des GMM a également été proposée par Laurent Benaroya [BB03] : il s'agit d'un modèle de mélange de gaussiennes à facteurs d'amplitude (**MMGA**).

Les auteurs [MD04, Hor] proposent un modèle de mélange de chi-deux dans le cadre de l'extraction de motifs à partir d'une représentation temps-fréquence**.**

Dans la suite, nous illustrons les méthodes basées sur des modèles statistiques *a priori* dans le cas des signaux sonores. Nous nous limiterons à l'étape d'estimation des sources à proprement parler.

Modèles statistiques *a priori* pour la séparation de sources à partir d'un seul capteur Comme mentionné précédemment, le problème de séparation de sources mono-capteur est souvent rencontré dans le cas des signaux sonores. Prenons l'expemple de séparation de deux sources à partir d'un seul enregistrement. Il peut s'agir par exemple de la voix chantée. Soit x le signal de mélange, x_1 la voix chantée, x_2 la musique et ε un bruit gaussien centré de variance σ_ε^2.

$$x = x_1 + x_2 + \varepsilon \tag{E.11}$$

Cas de sources stationnaires gaussiennes (Filtrage de Wiener) Supposons que les processus sources sont stationnaires gaussiens. Soient $\sigma_1^2(f)$ et $\sigma_2^2(f)$ les densités spectrales de puissance (DSP) des deux sources. Ici la transformée $\mathscr{F}$ est la transformée de Fourier. L'estimation des sources au sens du maximum *a posteriori* est donnée par (E.12) :

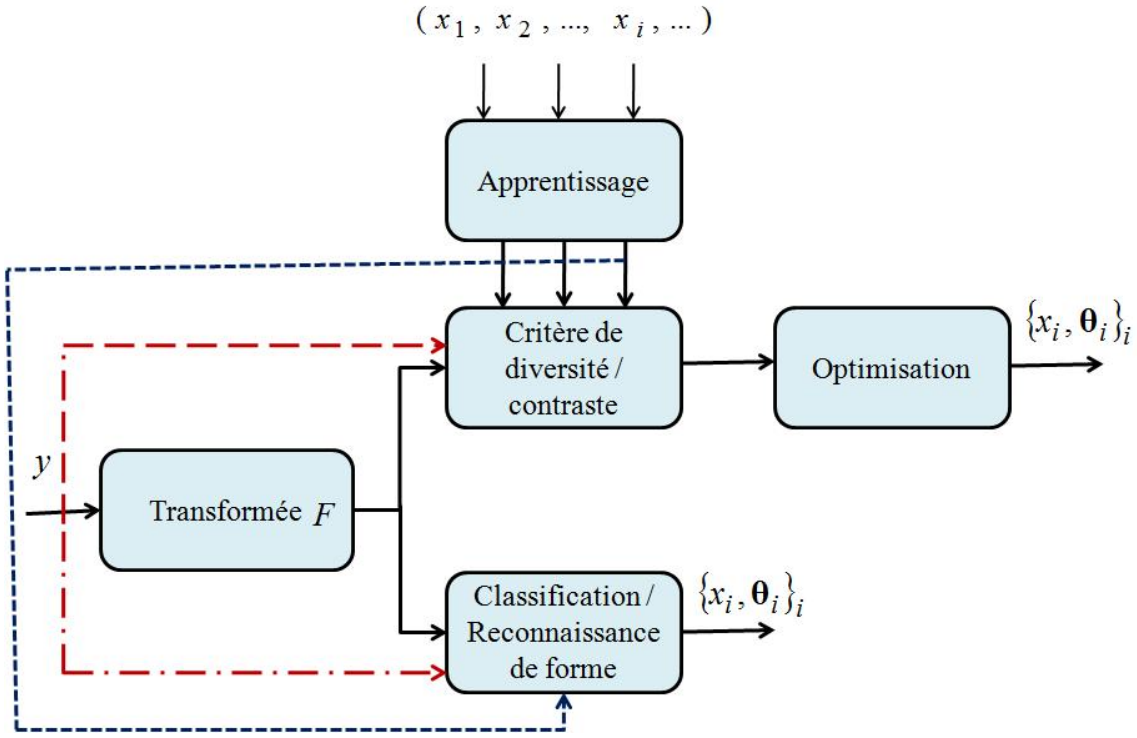


Figure E.2: Séparation des sources à partir d'un seul capteur dans un domaine transformé

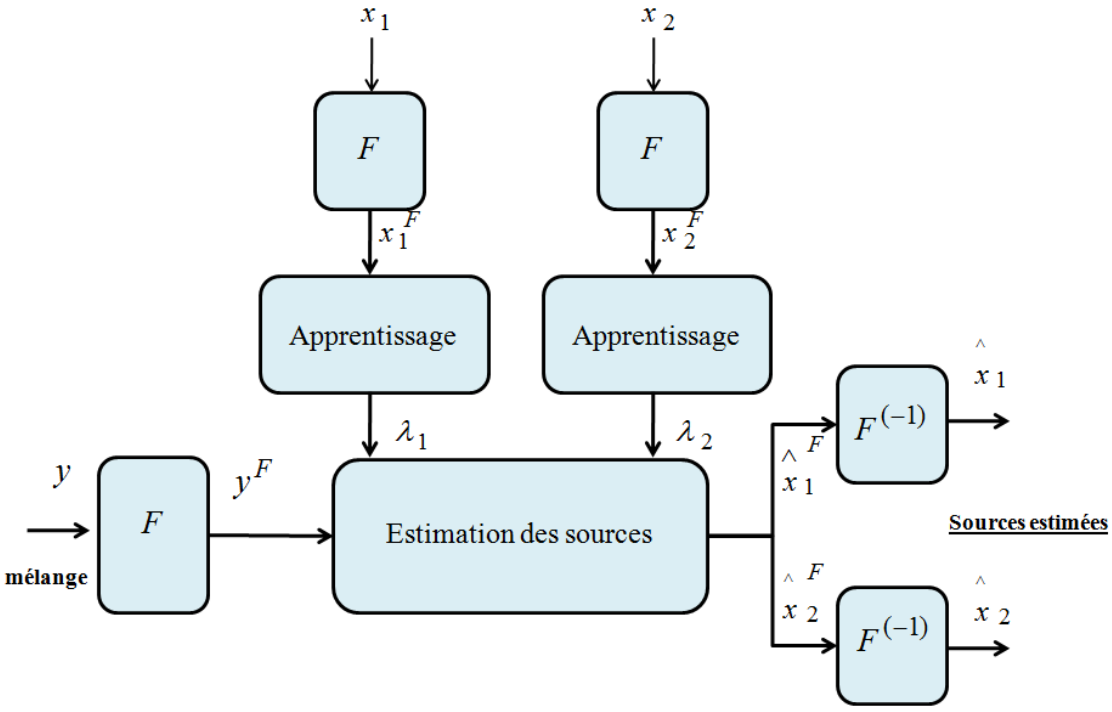


FIGURE E.3: Schéma général de séparation basée sur des modèles statistitiques *a priori*

$$\begin{cases} \widehat{\mathscr{F}_{x_1}(f)} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2+\sigma_\varepsilon^2}\mathscr{F}_x(f) \\ \widehat{\mathscr{F}_{x_2}(f)} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2+\sigma_\varepsilon^2}\mathscr{F}_x(f) \end{cases} . \tag{E.12}$$

La solution obtenue est celle du ***fitrage du Wiener***. Ce filtrage purement fréquentiel n'est pas approprié aux signaux réels souvent non-stationnaires tels que les signaux sonores.

Extension au modèle de mélange de gaussiennes (GMM) Le filtrage de Wiener ne permet pas de traiter le cas de signaux non-stationnaires. Il faudrait alors choisir une transformée appropriée et un modèle statistique permettant de prendre en compte l'évolution au cours du temps des propriétés spectrales du signal. Notons toutefois que la démarche reste la même que précédemment.

Prenons le cas de deux sources sonores. Il peut s'agir d'une voix chantée superposée à la musique ou la voix d'un homme et la voix d'une femme, ou encore deux instruments de musique. Ces signaux sont non-stationnaires, plus précisément stationnaires par morceaux au sens où Laurent Benaroya le propose dans [BB03]. Il est alors judicieux de s'intéresser plutôt à une représentation temps-fréquence, par exemple la Transformée de Fourier à Court Terme. Le problème de séparation est reformulé ainsi :

$$\mathscr{F}_x(t,f) = \mathscr{F}_{x_1}(t,f) + \mathscr{F}_{x_2}(t,f) + \mathscr{F}_\varepsilon(t,f). \tag{E.13}$$

Etant donné que chaque source est caractérisée par l'évolution de ses propriétés spectrales au cours du temps, Benaroya propose de considérer chaque trame du signal $\mathscr{F}_{x_i}(t,f)$ (pour $i=1,2$) à t fixé, comme la réalisation d'un processus stochastique. Par ailleurs, les trames de la même source sont des réalisations indépendantes.

Chaque trame du bruit est modélisée par un processus gaussien centré de variance σ_ε^2.

Les modèles a priori des sources dans le domaine temps-fréquence sont des modèles de mélange de gaussiennes centrées. Etant donné ce modèle et celui du bruit, Benaroya a établi des formules d'estimation des sources par leurs espérances conditionnelles respectives [BB03]. Une généralisation du filtrage de Wiener pour la séparation des sources sonores à partir d'un capteur est donc effectuée. D'autres extensions aux HMM ou encore aux GMMA ont été proposées par le même auteur. Pour plus de détails sur ces modèles, le lecteur peut se référer à [BB03, PGB].

Soulignons que la démarche est commune à toutes ces méthodes et est dédiée aux signaux sonores.

Synthèse

La séparation de sources à partir d'un seul capteur sans modèle des sources peut être envisagée dans un domaine transformé. Il s'agit en effet de choisir judicieusement la transformation et de définir un modèle statistique des sources dans ce domaine transformé commun aux sources à séparer.

Cette approche se décline en deux étapes. La première étape d'apprentissage permet de définir des modèles statistiques des sources étudiées dans un domaine transformé. La deuxième étape

consiste à estimer les sources par optimisation d'une fonction de contraste. Cette approche nécessite les conditions suivantes :

- ➢ *disposer d'une base d'exemples des signaux à estimer qui soit suffisamment significative pour définir des modèles robustes ;*
- ➢ *il existe un domaine transformé dans lequel les sources présentent une certaine diversité.*

Les développements méthodologiques réalisés dans le cadre de cette approche émanent principalement de la communauté du traitement de la parole. Les méthodes sont alors souvent dédiés aux signaux sonores.

Parcimonie et séparation de sources mono-capteur

Depuis les années 2000, de nombreux travaux de recherche portent sur l'utilisation de la diversité (spectrale, morphologique, temporelle) des sources en s'appuyant sur les représentations adaptatives des signaux pour la séparation des sources. Nous montrerons par la suite qu'à la différence de la première approche, cette dernière met en œuvre plusieurs transformées *adaptées* au sens d'un critère (à définir) aux sources à extraire. Une troisième approche, hybride, consiste à utiliser conjointement les deux méthodes.

Représentations adaptatives Une représentation adaptative consiste à décomposer le signal dans un dictionnaire redondant. Soit $x(t)$ un signal de longueur T et soit $\mathscr{D}$ un dictionnaire redondant composé de $N \geq T$ vecteurs g_i de dimension T. Si $\mathscr{D}$ est un système générateur de $\mathbb{R}^T$, il existe une infinité de représentations exactes de $x(t)$ dans $\mathscr{D}$ du type : $x(t) = \sum_{i \leq 1 \leq N} \alpha_i g_i(t)$.

Le principe de cette décomposition est de sélectionner, parmi toutes les décompositions possibles, la « meilleure » d'entre elles au sens d'un critère, par exemple l'économie de la représentation. Le critère le plus utilisé, dit critère de parcimonie, est la norme $\mathscr{L}_p$ des coefficients α_i avec $0 \leq p \leq 1$.

Trois grandes familles d'algorithmes permettent d'obtenir une décomposition optimale ou sous-optimale d'un signal dans un système redondant : l'approche par *Meilleure Base* (*Best Basis*), l'approche par *Poursuite de Base* (*Basis Pursuit*) et l'approche par *Poursuite Adaptative* (*Matching Pursuit*).

Parcimonie et séparation aveugle de sources à partir d'un seul capteur En s'appuyant sur le cadre théorique des représentations adaptatives, des chercheurs ont étendu ce cadre à la séparation de sources sonores à partir d'un seul capteur [Kje, BGB01] et à la décomposition d'images [SED04]. En effet, dans le cadre d'un mélange linéaire additif, le problème d'extraction des sources du signal composite est ramené à la recherche de bases adaptées à chacune des sources du mélange. La séparation à proprement parler est ainsi réalisée grâce à la décomposition parcimonieuse du mélange sur la réunion de bases adaptées connues (préalablement apprises).

Prenons le cas de deux sources sonores x_A et x_B de longueur T. Le signal observé est $x = x_A + x_B$.

L'approche proposée par les auteurs [DH99, ZP99] suppose que chaque source admet une représentation parcimonieuse :

$x_A = \alpha_A D_A$ et $x_B = \alpha_B D_B$ où α_A et α_B sont les coefficients de la représentation de chacune des sources dans sa base adaptée respectivement notée $\mathscr{D}_A$ et $\mathscr{D}_B$. Cette condtion de parcimonie se traduit par les deux propriétés suivantes :

1. *Tout signal de la même classe que x_A (respectivement x_B) admet une représentation «sparse» dans le dictionnaire $\mathscr{D}_A$ (respectivement $\mathscr{D}_B$) au sens où il est représenté par très peu de coefficients.*
2. *Le dictionnaire d'une classe de signaux n'est pas parcimonieux pour la deuxième classe au sens où la représentation d'un signal d'une classe dans le dictionnaire adaptée à la seconde n'est pas parcimonieuse.*

Le nouveau paradigme de séparation de sources mono-canal consiste alors à décomposer parcimonieusement le mélange x sur le dictionnaire redondant $\mathscr{D} = \mathscr{D}_A \cup \mathscr{D}_B$, ce qui revient à résoudre le problème d'optimisation sous contraintes suivant :

$$\begin{cases} \{\alpha_A^{opt}, \alpha_B^{opt}\} = \underset{\alpha_A, \alpha_B}{argmin} \|\alpha_A\|_{L_p} + \|\alpha_B\|_{L_p} \\ \text{sous contrainte que : } x_A = \alpha_A D_A \text{ et } x_B = \alpha_B D_B \end{cases} \tag{E.14}$$

Plusieurs algorithmes de décomposition sous-optimale ou optimale dans un dictionnaire redondant existent dont nous citons l'algorithme du Basis Poursuit, la méthode FOCUSS régularisée ou encore la méthode du seuillage alterné . Plus de précisions sur la méthode de décomposition parcimonieuse, appelée Analyse en Composantes Morphologiques (*MCA*) sont présentées dans l'article [SED04].

Notons que la qualité de la décomposition dépend étroitement de la parcimonie de chaque source. La problématique de construction des dictionnaires (apprentissage) est extrêmement sensible dans ce cadre. Elle nécessite en particulier de disposer d'une base d'exemple de signaux. Elle a bénéficié de l'émergence des décompositions temps-fréquence et temps-échelle redondantes pendant les deux dernières décennies.

Cette démarche a été appliquée en séparation de signaux sonores ainsi que pour la décomposition des images (séparation texture-contour par exemple) .

Approche hybride : parcimonie et modèles statistiques

La démarche basée sur des modèles statistiques peut être transposée dans le cadre des représentations parcimonieuses. Il s'agirait en effet de considérer un modèle probabiliste décrivant la distribution des coefficients de la représentation parcimonieuse de chaque source. Le modèle de mélange porte ainsi sur la distribution des coefficients dans un système redondant réunissant plusieurs bases adaptées aux sources présentes dans le mélange, ce qui n'est pas le cas de l'approche décrite en 3.2. La densité de probabilité des coefficients d'une classe de signaux donnée est choisie de sorte que la propriété de parcimonie soit respectée, telles que les densités leptokurtiques (de kurtosis positif) [Bob08]. Cette approche a été étudiée uniquement dans le cadre de séparation de sources multi-capteurs dans la thèse de Jerôme Bobin et a donné lieu à une nouvelle méthode appelée « Analyse en Composantes Mutuellement Parcimonieuses » (*ACMP*). Le fonctionnement de cette méthode est analogue à celui de l'ACI. Pour plus de précisions sur cette méthode, le lecteur peut se référer au chapitre 3 de la thèse de Jérôme Bobin.

Le cas mono-capteur ne peut pas être abordé exactement de cette manière, mais plutôt par une approche « problèmes inverses » dans un cadre bayésien.

Conclusion : Vers une démarche dédiée pour la caractérisation des signaux composites

Dans cette partie, le problème de caractérisation de signaux composites posé dans le cadre générique de la décision est décliné en trois problèmes : la détection, la classification et l'estimation. Afin de répondre à la question « caractériser les signaux d'un mélange donné à partir de très peu d'informations », nous avons montré qu'il est nécessaire de compléter notre connaissance du problème et ce à travers des données contextuelles, des données d'expertise ou encore des données d'apprentissage. Une phase de formalisation permet de rendre ces connaissances exploitables dans un système de caractérisation, qui se situe dans le cadre :

- ➢ *des problèmes inverses, plus précisément l'approche bayésienne pour les problèmes inverses* ;
- ➢ *ou de la séparation de sources à partir d'un seul capteur, où le mélange est étudié dans un domaine transformé adapté aux caractéristiques des sources à extraire ou à détecter.*

La démarche de la caracatérisation des signaux telle que envisagée dans notre cas se résume en trois phases.

Apprentissage

Nous utilisons ce terme ici pour désigner à la fois l'apprentissage d'un modèle à partir d'exemples (données d'entraînement) mais aussi la phase de définition des *a priori* sur les signaux ou sur leurs caractéristiques (morphologiques, spectrales).

Choix d'une mesure de diversité / de contraste

Il s'agit de définir une mesure qui permette de discriminer les composantes du mélange. Elle peut porter sur les signaux ou sur leurs représentations dans des domaines adaptés ou enocre sur certaines caractéristiques des signaux.

Optimisation d'un critère

Cette étape a pour objectif d'extraire les objets recherchés à partir du mélange au sens du critère défini.

Annexe F

Filtrage morphologique

Définitions

Morphologie mathématique

Cette théorie a été introduite par G. Matheron et J. Serra dans les années 1960. Elle permet d'analyser les structures spatiales des images et est fondée sur la théorie des treillis.

Filtre morphologique (G. Matheron, J. Serra)

Toute transformation croissante et idempotente sur un treillis définit un filtre morphologique. La croissance assure que la structure de base du treillis, c'est à dire la relation d'ordre, est conservée lors d'un filtrage morphologique. L'idempotence signifie que le signal filtré est transformé en un invariant pour l'opérateur en question.

Ouverture

L'ouverture d'un ensemble X par un élément structurant B est l'espace balayé par B quand B est entièrement inclus dans X. En traitement des images, par ouverture, les petites particules sont supprimés ; les contours des grosses particules sont lissés.

Fermeture

La fermeture d'un ensemble X correspond à l'ouverture de son complémentaire.

Illustrations

La figure illustre l'effet d'une ouverture (O) et d'une fermeture (F) appliquées séparément à un signal.
La figure illustre l'effet de l'application d'une ouverture puis d'une fermeture d'une part et l'application d'une fermeture puis d'une ouverture au même signal.

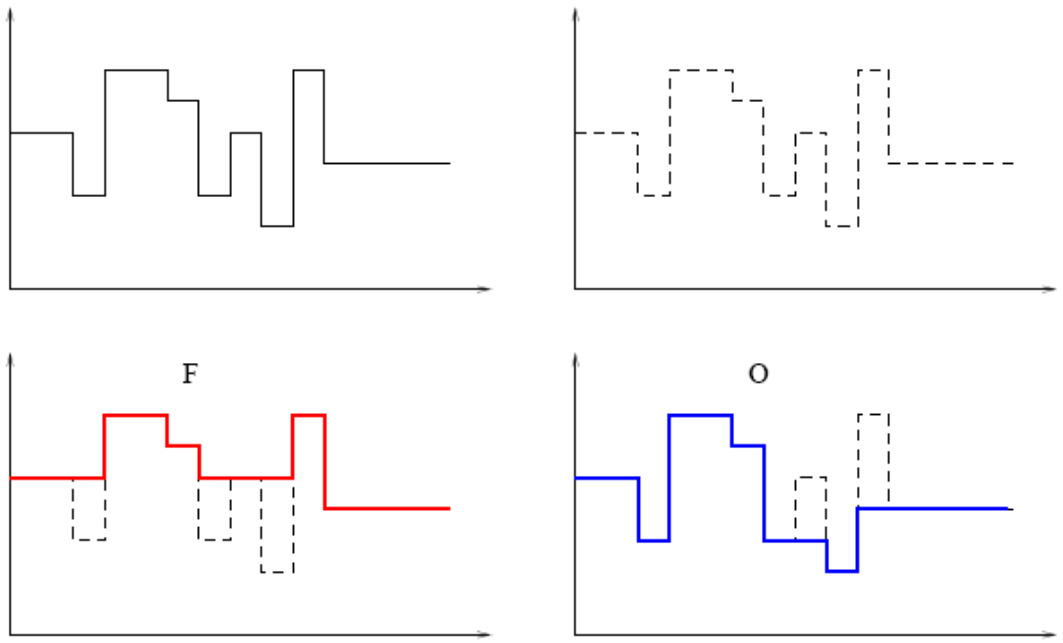


Figure F.1: Ouverture et fermeture morphologiques

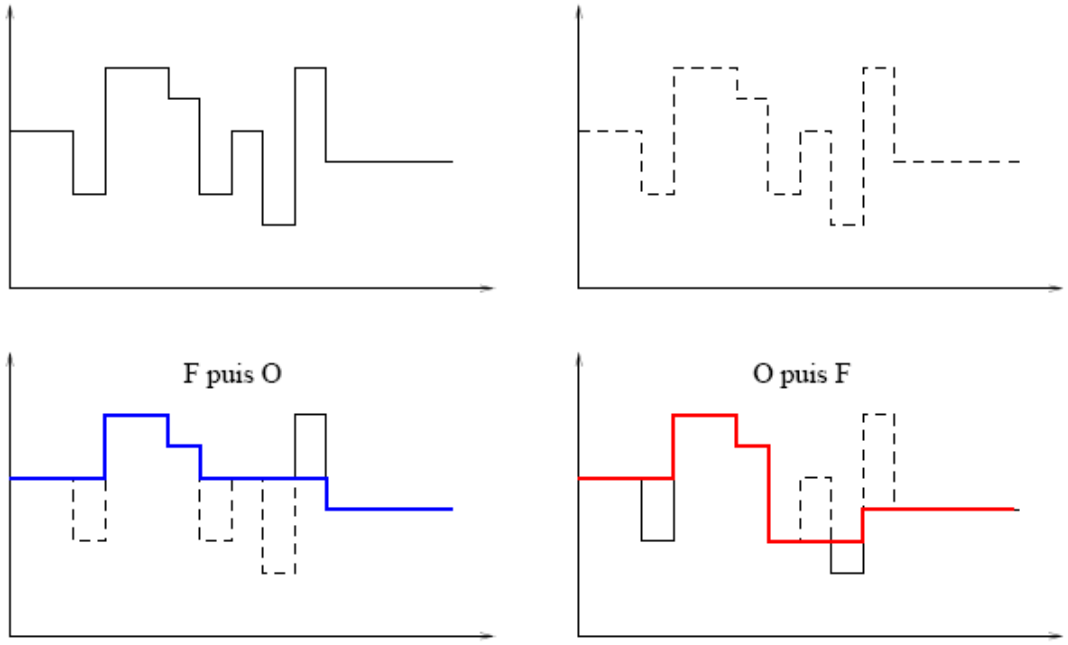


Figure F.2: Ouverture et fermeture morphologiques

Bibliographie

[Agr96] R. Agrawal. Fast discovenry of association rules. *Advances in Knowledge Discovery and Data Mining*, 1996. 3.2.2.3

[Akb95] H. Akbari. Validation of an algorithm to disaggregate whole-building hourly electrical load into end-uses. *Energy*, 20 :1291–130, dec 1995. C

[ASK84] D.J. Aigner, C. Sorooshian, and P. Kerwin. Conditional demand analysis for estimating residential end-use load profiles. *Journal of Energy*, 26 :81–97, 1984. C

[BB03] L. Benaroya and F. Bimbot. Wiener based source separation with hmm/gmm using a single sensor. *Proceedings of 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Japan*, to add :957–961, 2003. 2.2.7.4, 5, 5.2.3.3, E, E, E

[BC93] O. Bouliou and B. Chazottes. Logiciel d'indentification des signatures des appareils électro-domestiques - bilan des travaux. Technical Report HR22/2953, Electricité de France-Recherche et Développement, juil 1993. 1.3.1

[BGB01] L. Benaroya, R. Gribonval, and F. Bimbot. Représentations parcimonieuses pour la séparation de sources à partir d'un seul capteur. *GRETSI*, 2001. 2.2.7.4, 5, 5.2.3.3, E

[BMa06] J. Bobin, Y. Moudden, and al. Morphological Diversity and Source Separation. *IEEE Transactions on Signal Processing*, 13 :409–412, 2006. 5

[BN93] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes : Theory and Applications*. Prentice Hall Information and System Sciences Series, 1993. 3.3.2.1, 4.2.3.2

[Bob08] J. Bobin. *Diversité Morphologique et Analyse de Données Multivaluées*. PhD thesis, Laboratoire CEA-IRFU/SEDISAP, octobre 2008. E

[Bon96] M. Bons. *Modèles à sources markoviennes cachées appliqués à l'analyse non-intrusive des principaux usages domestiques de l'électricité*. PhD thesis, Université de Rennes I, 1996. (document), 1.3.1, 1.3.3, C

[Boz00] L. Bozzi. Modélisations statistiques des courbes de charge 10min et développement de téléservices en immeubles collectifs. Technical report, Electricité de France-Recherche et Développement, sep 2000. 1.3.1, C

[BV03] M. Baranski and J. Voss. Non-Intrusive Load Appliance Load Monitoring based on an Optical Sensor. *IEEE Power Tech Conference*, jun 2003. (document), 1.3.1, C

[BV04] M. Baranski and J. Voss. Genetic Algorithm for Pattern Detection in NIALM Systems. *IEEE International Conference on Systems Man and Cybernetics*, jul 2004. (document), 1.3.1, C

[CG08] B. Charbonnier and M. El Guedri. Détection et estimation de l'eau chaude sanitaire à partir de la courbe de charge générale au pas 1 seconde : améliorations, évaluation et perspectives. Technical Report H-P1B-2008-037749, Electricité de France-Recherche et Développement, Février 2008. 4.3.2

[Che06] M. Chendeb. *Détection et classification de signaux non-stationnaires par utilisation des ondelettes : Application aux signaux électromyographiques utérins*. PhD thesis, Université de Technologie deTroyes - France, 2006. 5.2.3.3, E

[Com90] P. Comon. Analyse en composantes indépendantes et identification aveugle. *Traitement du Signal, N°5, Spécial Non linéaire non gaussien*, 17 :435–450, 1990. E

[Cox06] R. Cox. *Minimally Intrusive Strategies for Fault Detection and Energy Monitoring*. PhD thesis, Massachusetts Institute of Technology, Septembre 2006. (document), 5.2.3.3

[CR03] B. Crémilleux and F. Rioult. Optimisation d'extraction de motifs : une nouvelle méthode fondée sur la transposition des données. *Cap*, 2003. 3.2.2.3, 3.2.2.3

[CRI] CRIEPI. http ://criepi.denken.or.jp/en/index.html. (document)

[CTE95] CANMET CTEC. HOT2000 version 7-10 User's Md, Centre de la Technologie de l'Energie de CANMET-Varennes. Ressources Naturelles Ottawa - Canada, 1995. C

[DDC94] I. Vincent M. Davy, C. Doncarli, , and E. Le Carpentier. Nonstationary signals classification using time-frequency distributions. *Proceedings of the International Synopsium on TFTS*, pages 233–236, 1994. 3.2.2.1

[DDT02] M. Davy, C. Doncarli, , and J.Y Tourneret. Classfication of chirp signals using hierarchical bayesian learning and MCMC methods. *IEEE Transactions on Signal Processing*, 50 :377–388, fev 2002. 3.1.2

[DG01] G. Demoment and Y. Goussard. *Approches bayésiennes pour les problèmes inverses, chapitre : Inversion dans un cadre probabiliste*. Lavoisier, 2001. C, E

[DH99] D.L. Donoho and X. Huo. Uncertainty Principles and Ideal Atomic Decomposition. June 1999. E

[DK92] S. Drenker and A. Kader. Non-intrusive appliance load monitoring. *Proceedings IEEE Computer Applications in Power*, 80 :1870–1891, dec 1992. C

[Don02] Dong Luo and Leslie K. Norford and Steven.R. Shaw and Steven. B. Leeb. Monitoring HVAC Equipment from a Centralized Location- Methods and Field Test Results. *ASHRAE Transactions*, 108, 2002. 1.3.1, C

[DR09] G. Darnis-Rebec. *Etude d'une fonction d'identification de la consommation d'électricité par usage intégrable dans le compteur bleu électronique*. PhD thesis, Université du Sud Toulon-Var (Laboratoire SNC), Janvier 2009. (document)

[DWEG05] G. Durso, P. Wagner, and M. El-Guedri. Apport du Traitement du Signal à la détection des usages électriques à partir de la courbe de charge générale. Technical Report H-P1B-2005-02556, Electricité de France-Recherche et Développement, dec 2005. 2.1.1

[ene] www.enetics.com. C

[Far99] L. Farinaccio. *The Disaggregation of Whole-House Electric Load into the Major End-Uses Using a Rule-Based Pattern Recognition Algorithm*. PhD thesis, Université de Concordia, 1999. (document), 1.3.3, C

[Fel86] M.F. Fels. PRISM : An introduction. *Energy and Buildings*, 9 :5–18, 1986. C

[Fer80] J.D. Fergusson. Variable duration models for speech. *Proc. Synopsium on the Applications of HMMs to text Speech - Princeton New Jersey*, pages 143–179, 1980. C

[Fla93] P. Flandrin. *Temps-Fréquence*. Hermès edition, 1993. 3.2.1.1, 3.2.2.1

[GBLa09] M. El Guedri, J. Bect, C. Lajaunie, and al. RJ-MCMC sampler for One Sensor Source Separation : an Application to Electric Load Monitoring. *European Signal Processing Conference 2009 (EUSIPCO) : Glasgow-Scotland*, 2009. 5.2.3.3

[GDL+09] M. El Guedri, G. D'Urso, C. Lajaunie, G. Fleury, and S. Parfouru. Time Frequency Characterization for Electric Load Monitoring. *European Signal Processing Conference 2009 (EUSIPCO) : Glasgow-Scotland*, 2009. 3.2.2.1, 3.2.2.2, 3.2.2.2

[GL] R. Gribonval and S. Lesage. A survey on Sparse Component Analysis for source separation : principles, perspectives and new challenges. *14th European Synopsium on Artificial Neural Networks (ESANN*, pages 323–330. 5.2.3.3

[GL91] Simulation Research Group and Lawrence Berkley Laboratory. DOE-2 BAsics LBL-29140. Univsersity of Berkley, 1991. C

[Had76] J. Hadamard. *Sur les problèmes aux dérivées partielle et leur signification physique*. Editions MIR, 1976. E

[Had93] D.L. Hadley. Daily variations in wac system electrical energy consumption in response to different weather conditions. *Energy and Buildings,*, 19 :235–247, 1993. C

[Har89] G.W. Hart. Energy Monitoring and Computerized Surveillance via Utility Power Flows. *IEEE Technologie and Society Magazine*, 6 :12–16, 1989. (document), 1, 1.3.1, 1.3.1, C, C

[Har92] G.W. Hart. Non-intrusive appliance load monitoring. *Proceedings of the IEEE...*, 80 :1870–1891, dec 1992. 1.3.1, 1.3.1, C, C

[Har94] G.W. Hart. Automatic construction of finite state load behavior models. *Proceedings of the 4th International Synopsium on Distribution Automation and Demand-Side Management*, jan 1994. 1.3.1

[Has70] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57 :97, 1970. C

[Heb] G. Hebrail. Practical data mining in a large utility compagny. Technical report, Electricité de France-Recherche et Développement. (document), C

[Hor] C. Hory. *Mixtures of Chi2 for the interpretation of a time frequence representation*. PhD thesis. 3.1.2, 3.2.2.1, E

[HP00] A. Hyvärinen and P. Pajunen. Independant component analysis : Algorithms and applications. *Neural Networks*, 13 :411–430, 2000. E

[Ia01] J. Idier and al. *Approches bayésiennes pour les problèmes inverses*. Lavoisier, 2001. 2.2.7.4, E

[KG83] S. Kirkpatrick and C.D. Gelatt. Optimizing by simulated annealing. *Sciences, series 220*, pages 671–680, 1983. E

[Kje] E. Kjertsi. *Frame Based Signal Representation and Compression*. PhD thesis. E

[Koo95] J.G. Koomey. Residen tial Sector End-Use Forecasting with EPRI-REEPS : Summary Input Assumptions and Results. Technical Report LBL-34044, Energy and Environment Division - Lawrence Berkeley National Laboratory, 1995. C

[Kop03] P. Kopf. Technical opportunities for developing services about remote power metering. Technical report, Electricité de France - Recherche et Développement, dec 2003. C

[KS] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer (Applied Mathematical Sciens, vol. 160). 5.1.2, E, E

[Laj02] C. Lajaunie. Estimation stochastique de la consommation d'usages pour les hypermarchés. Technical report, Ecole des Mines de Paris (Armines), 2002. (document), 1.3.1, 1.3.3, 1.9, 1.3.3, C, C.1, C

[LB06] V. Le-Bars. Spécifications pour compteurs bleu électronique triphasés : palier 2007. Technical Report HR-43/06/013/A, Electricité de France - Recherche et Développement (MIRE-Clamart), dec 2006. 1.1.4

[Lee93] Steven B. Leeb. *Conjoint Pattern Recognition Approach to Non-Intrusive Load Monitoring*. PhD thesis, Massachusetts Institute of Technology, feb 1993. (document), 1.3.1, C, C, C, C.4

[Lee03] Kwangduck Douglas Lee. *Electric System for Non-Intrusive Power Minotor*. PhD thesis, Massachusetts Institute of Technology, jun 2003. 1.3.1, 5.2.3.3, C, C

[Lie00] M.N.M. Lieshout. *Point Processes and their applications*. Imperial College Press, 2000. E

[LLC+03] C. Laughman, K. Lee, R. Cox, S. Shaw, L. Norford, and P. Armstrong. Power signature analysis. *IEEE Power and Energy Magazine*, pages 55–63, mars 2003. (document), C, C, C.5

[LMLS93] Steven B. Leeb, James L. Kirtley Michael, S. Levan, and Joseph P. Sweeney. Development and Validation of a Transient Event Detector. *AMP Journal of Technology*, 3 :69–74, nov 1993. (document), 1.3.1, C, C.6

[LN04] B.M. Larsen and R. Nesbakken. Household electricity end-use consumption : results from econometric and engineering models. *Energy Economics*, 26 :179–200, 2004. C

[MD04] N. Martin and C. Doncarli. *Décision Temps-Fréquence*. Hermès, 2004. 3.1.2, 3.2.2.2, 3.2.2.2, E

[MEG07] G. D'Urso M. El-Guedri. Caractérisation des composantes constitutives de la courbe de charge électrique- état de l'art. Technical Report HP-P1B, Electricité de France - Recherche et Développement (Chatou), juil 2007. C, C

[MZ00] M.L. Marceau and R. Zmeureanu. Non-Intrusive Load Disaggregation computer program to estimate the energy consumption of major end-uses in residential buildings. *Energy Conversion and Management*, 41 :1389–1403, 2000. C

[Naj09] H. Najmeddine. *Developpement de méthodes de classification et d'identification des appareils électriques du secteur résidentiel*. PhD thesis, Université Blaise-Pascal Clermont-Ferrand (LASMEA), 2009. (document)

[NL96] L.K. Norford and S.B. Leeb. Non-intrusive electrical load monitoring in commercial buildings based on steady state and transient load-detection algorithms. *Energy and Buildings*, 24 :51–64, 1996. C

[Ort95] M. Ortner. *Processus Ponctuels Marqués pour l'Extraction Automatique de Caricatures de bâtiments à partir de Modèles Numériques d'Elevation*. PhD thesis, Univeersité de Nice Sophia-Antipolis, 1995. 5.2.3.3

[Peu08] P. Peureux. Caractérisation de défauts plans par contrôle ultrasonore et traitement de signal adapté. *Journée COFREND*, to add :to add, 2008. E

[PGB] A. Ozerovand P. Philippe, R. Gribonval, and F. Bimbot. Choix et adaptation de modèles statistiques pour la séparation de voix chantée à partir d'un seul microphone. *to add*, to add :to add. 3.1.2, E, E

[PM95] V. Pierson and N. Martin. Watershed segmentation of time-frequency images. *Proceedings of the IEEE Workshop on Non-linear Signal and Image Processing*, 1995. 3.1.2, 3.2.2.1

[Rab89] Lawrence R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(22), feb 1989. C, E

[Rob96] C. Robert. Méthodes de monte carlo par châines de markov. *Economica*, 1996. E

[SED04] J.L. Starck, M. Elad, and D. Donoho. Redundant multiscale transformes and their applications for morphological component analysis. *Advances in Imaging and Electron Physics*, 132 :to add, 2004. E, E

[Sha00] S. Shaw. *System Identification Techniques and Modeling for Non-Intrusive Load Monitoring*. PhD thesis, Massachusetts Institute of Technology, 2000. C

[Tra92] K.E. Train. An assessment of the accuracy of statiscally adjusted engineering models on end-use load curve. *Energy*, 7 :713–723, 1992. C

[Vas95] S.V. Vaseghi. State duration modelling in hidden Markov models. *Signal Processing*, 41 :31–41, 1995. C

[Vin95] P. Vincent. *Classification de signaux non-stationnaires*. PhD thesis, Ecole centrale de Nantes, 1995. 3.1.2, 3.2.2.1

[Vit67] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum algorithm. *IEEE Transactions on Theory and Information*, IT-13 :260–269, Apr 1967. C

[Wag05] P. Wagner. Estimation de la consommation des usages à partir de la courbe de charge d'un client particulier - état de l'art. Technical Report HP-P1B/05/011A, Electricité de France - Recherche et Développement (Chatou), dec 2005. 1.1.1

[Xen] Xenergy. RECAP, Residential Energy Audit. www.xenergv.com/nwweb.nsf. C

[ZP99] M. Zibulevsky and B.A. Pearlmutter. Blind Source Separation by Sparse Decomposition. Technical Report CS99-1, University of Mexico, 1999. E