



HAL
open science

Détection et caractérisation de signaux transitoires : application à la surveillance de courbes de charge

Matthieu Sanquer

► **To cite this version:**

Matthieu Sanquer. Détection et caractérisation de signaux transitoires : application à la surveillance de courbes de charge. Electronique. Université de Grenoble, 2013. Français. NNT : 2013GRENT095 . tel-01561742

HAL Id: tel-01561742

<https://theses.hal.science/tel-01561742>

Submitted on 13 Jul 2017

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 7 août 2006

Présentée par

Matthieu Sanquer

Thèse dirigée par **Nadine Martin**

préparée au sein **Gipsa-lab**

et de **École Doctorale Électronique, Électrotechnique, Automatique & Traitement du Signal**

Détection et caractérisation de signaux transitoires

Application à la surveillance de courbes de charge

Thèse soutenue publiquement le **15 mars 2013**,
devant le jury composé de :

M. Olivier Gaudoin

Professeur à Grenoble-INP, Président

M. Cédric Richard

Professeur à l'Université de Nice Sophia-Antipolis, Rapporteur

M. Jean-Yves Tourneret

Professeur à l'Institut National Polytechnique de Toulouse, Rapporteur

M. Pierre Borgnat

Chargé de recherche au CNRS, Examineur

Mme Nadine Martin

Directeur de recherche au CNRS, Directeur de thèse

M. Florent Chatelain

Maitre de conférence à Grenoble-INP, Co-encadrant



Avant-propos

Ce travail a été réalisé dans le cadre d'un contrat de thèse CIFRE entre le laboratoire de recherche Gipsa-Lab à Grenoble et le département STEP (Simulation et Traitement de l'information pour l'Exploitation des systèmes de Production) de EDF R&D.

Je tiens à remercier l'ensemble des personnes ayant participé à l'élaboration de ce manuscrit ainsi que les personnes m'ayant soutenu et encouragé pendant ces années de doctorat.

Je remercie ma directrice de thèse, Nadine Martin pour m'avoir guidé et conseillé tout au long de cette thèse et pour m'avoir aidé à prendre du recul sur mes travaux. Je remercie mon co-encadrant, Florent Chatelain pour avoir donné son aide, consacré son temps sans compter et grandement amélioré la rigueur de certains chapitres de ce manuscrit. Je les remercie encore chaleureusement tout les deux pour m'avoir soutenu et supporté dans les moments difficiles et pour avoir permis la finalisation de ce manuscrit.

Je remercie Mabrouka El-Guedri pour m'avoir permis de travailler sur ce sujet de recherche riche et innovant qu'est l'analyse de courbes de charge. Je remercie Bruno Charbonnier pour sa relecture attentive de ce manuscrit et pour m'avoir renseigné sur le contexte industriel de ces travaux.

Je remercie Jean-Yves Tourneret et Cédric Richard d'avoir accepté d'être les rapporteurs de ce travail, pour leur lecture éclairante de ce manuscrit et la pertinence de leurs commentaires.

Je remercie Olivier Gaudoin d'avoir accepté de présider le jury de cette thèse et Pierre Borgnat pour avoir accepté dans faire partie et d'assister à la présentation de ce travail. Je les remercie tout deux de l'intérêt qu'ils ont porté à ces travaux.

Je remercie collectivement l'équipe SAI-GA du Gipsa-lab et l'équipe PIB de EDF R&D pour m'avoir accueilli pendant ces années de thèse. Je remercie les doctorants du Gipsa-lab et en particulier Fabien, Cyrille, Bastien, Lionel et Cedric pour les parties de coinche entre midi et deux ainsi que Jonathan, Gailene, Wei, Robin, Aude et Vincent pour les soirées et les moments passés ensemble.

Je remercie mes amis et ma famille pour leur soutien au quotidien.

Table des matières

Avant-propos	3
Notations	9
Introduction	11
Problématique	11
Organisation du manuscrit	12
Plan de lecture	12
1 Suivi de courbes de charges	15
1.1 Problématique et enjeux du suivi de courbe de charge	15
1.1.1 Contexte	15
1.1.2 Objectifs	17
1.1.3 Motivations	17
1.1.4 Protection de la vie privée	18
1.2 Généralités sur l'analyse de courbes de charge	18
1.2.1 Courant et tension	19
1.2.2 Puissance	19
1.2.3 Régime permanent et régime transitoire	20
1.3 État de l'art des méthodes de suivi de courbe de charge	20
1.3.1 Signature électrique des usages	20
1.3.2 Classification générale des usages électriques	21
1.3.3 Analyse macroscopique	22
1.3.4 Analyse microscopique	26
1.3.5 Base de données de courbes de charges	29
1.3.6 Applications industrielles commercialisées	29
1.3.7 Détection d'événements et décomposition de courbes de charges	30
1.4 Conclusions	31
2 Présentation des signaux électro-domestiques de la base de données SISED	33
2.1 Présentation générale de la base de données SISED	33
2.1.1 Conditions d'acquisition	33
2.1.2 Sélection des cas d'études	34
2.2 Régime transitoire des usages électriques	34
2.2.1 Constitution d'une base de données de signaux transitoires	35
2.2.2 Analyse générale des signaux transitoires de mise sous tension	35
2.2.3 Puissance active en régime transitoire	36
2.2.4 Analyse individuelle de chacune des classes d'usages	39
2.3 La puissance active pour la caractérisation des courbes de charge	46
2.4 Description de la puissance active par ajustement de modèle paramétrique	47
2.5 Conclusions	47

3	Modèle de régression à transitions régulières	49
3.1	Problématique de l'ajustement de courbes	49
3.1.1	Modèles de régression par morceaux	49
3.1.2	Modèles de régression à transitions régulières	50
3.2	Formulation générale du modèle	51
3.2.1	Sous-modèles	51
3.2.2	Fonctions de transition	52
3.2.3	Contrainte de convexité sur les poids des sous-modèles	53
3.2.4	Exemple simulé	53
3.3	Propriétés du modèle	54
3.3.1	Régularité de la fonction de régression	54
3.3.2	Formulation matricielle du modèle	56
3.3.3	Fonction de vraisemblance du modèle	56
3.4	Modélisation bayésienne hiérarchique	56
3.4.1	Densité de probabilité <i>a priori</i> des paramètres	57
3.4.2	Densité de probabilité <i>a priori</i> des hyperparamètres	61
3.4.3	Marginalisation de la densité conjointe <i>a posteriori</i>	63
3.5	Estimation par méthode MCMC à sauts réversibles	65
3.5.1	Vue d'ensemble de l'algorithme RJMCMC	66
3.5.2	Expression générale du rapport de Green	67
3.5.3	Mouvement de naissance/mort d'une transition	67
3.5.4	Mouvement de fusion/division d'une transition	69
3.5.5	Mouvement de mise à jour des transitions	70
3.5.6	Mouvement de mise à jour de l'ordre d'un sous-modèle	71
3.5.7	Mise à jour de δ^2	71
3.6	Évaluation des performances du modèle pour la représentation de signaux tests	72
3.6.1	Conditions de simulation	72
3.6.2	Signaux tests réguliers	73
3.6.3	Signaux tests non réguliers	74
3.7	Conclusions	77
4	Segmentation des signaux transitoires électro-domestiques	81
4.1	Spécification du modèle à transitions régulières pour la modélisation de signaux transitoires électriques	81
4.1.1	Objectifs	81
4.1.2	Sous-modèles	82
4.1.3	Paramétrisation des transitions : la fonction exponentielle étendue	82
4.1.4	Modèle bayésien hiérarchique pour l'estimation des signaux transitoires	87
4.1.5	Algorithme RJMCMC pour l'estimation des paramètres	89
4.2	Résultats de modélisation des signaux transitoires de la base de données SISED	91
4.3	Conclusions	92
5	Apprentissage de transitoires électro-domestiques	97
5.1	Apprentissage statistique : enjeux et méthodes	97
5.1.1	Définitions	97
5.1.2	Enjeux	98
5.1.3	Choix d'une méthode d'apprentissage	98
5.2	Apprentissage bayésien	99

5.2.1	Apprentissage bayésien standard	99
5.2.2	Modèle bayésien hiérarchique pour l'apprentissage	101
5.2.3	Densité <i>a priori</i> des vecteurs de caractéristiques	103
5.2.4	Loi <i>a posteriori</i> marginalisée	108
5.2.5	Loi <i>a posteriori</i> marginale	112
5.2.6	Algorithme RJMCMC pour l'apprentissage des signaux de la base de données	112
5.2.7	Résultats : apprentissage des signaux transitoires	114
5.2.8	Estimation de la classe d'un signal	117
5.2.9	Résultats : estimation de la classe des signaux tests	126
5.2.10	Validation croisée	127
5.3	Conclusions	131
Conclusions et perspectives		133
	Contributions	133
	Perspectives	134
	Publications	135
	Actes de conférences	135
	Article de revue avec comité de lecture	135
A	Annexes	137
A.1	Définition et paramétrisation des fonctions spéciales et des lois de probabilité standards	137
	A.1.1 Fonctions spéciales	137
	A.1.2 Lois de probabilité standards	137
A.2	Absence d'harmoniques pairs dans la forme d'onde du courant	138
A.3	Preuve de la proposition 4.1.1	139
A.4	Preuve de la proposition 4.1.2	140
A.5	Marginalisation des hyperparamètres de l'apprentissage bayésien hiérarchique	142
Bibliographie		147

Notations

\mathbb{R}	Ensemble des nombres réels
\mathbb{R}_+	Ensemble des nombres réels positifs
\mathbb{N}	Ensemble des entiers naturels
$\mathbb{E}(\cdot)$	Fonction partie entière
$\text{tr}(\cdot)$	Trace d'une matrice
$\text{card}(\cdot)$	Cardinal d'un ensemble
$\mathbb{I}_{\mathcal{E}}(\cdot)$	Fonction indicatrice de l'ensemble \mathcal{E}
$\Gamma(\cdot, \cdot)$	Fonction gamma (cf annexe A.1)
$B(\cdot, \cdot)$	Fonction bêta (cf annexe A.1)
$\mathcal{N}(\cdot, \cdot)$	Loi normale (cf annexe A.1)
$\mathcal{G}(\cdot, \cdot)$	Loi gamma (cf annexe A.1)
$\mathcal{IG}(\cdot, \cdot)$	Loi inverse-gamma (cf annexe A.1)
$\mathcal{B}(\cdot, \cdot)$	Loi bêta (cf annexe A.1)
\mathbf{x}, \mathbf{x}_i	Signal observé
$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_i$	Signal de bruit résiduel
$\boldsymbol{\theta}, \boldsymbol{\theta}_i$	Vecteur des paramètres du modèle
Θ	Ensemble de définition du vecteur des paramètres
$\boldsymbol{\phi}$	Vecteur des hyperparamètres
Φ	Ensemble de définition des hyperparamètres
σ^2, σ_i^2	Variance du bruit résiduel
K, K_i	Nombre de ruptures du modèle
$\boldsymbol{\tau}, \boldsymbol{\tau}_i$	Paramètres de localisation des fonctions de transitions
$\boldsymbol{\lambda}, \boldsymbol{\lambda}_i$	Paramètres d'échelle des fonctions de transitions
$\boldsymbol{\alpha}, \boldsymbol{\alpha}_i$	Paramètres de forme des fonctions de transitions
\mathbf{c}, \mathbf{c}_i	Paramètres de type des fonctions de transitions
$\boldsymbol{\eta}, \boldsymbol{\eta}_i$	Paramètres des fonctions de transitions
\mathbf{Q}	Ordres des sous-modèles
$\boldsymbol{\beta}, \boldsymbol{\beta}_i$	Paramètres des sous-modèles
$Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}), Z, Z_i$	Matrice de régression du modèle

Introduction

Peut-on identifier un appareil électrique à partir de l'énergie électrique qu'il consomme ? C'est essentiellement cette question qui guide les travaux de recherche présentés dans cette thèse. Historiquement les compteurs électriques ne fournissent que la quantité d'énergie totale consommée. Cette unique donnée ne permet qu'une identification sommaire des appareils électriques. Cependant les récents développements en matière de compteurs électriques « intelligents » permettent de disposer en temps réel de la consommation électrique d'un foyer, d'un bâtiment ou d'une entreprise. Ces compteurs permettent donc de disposer du profil temporel de la consommation, qui constitue un outil puissant pour l'analyse de l'électricité consommée. La problématique de cette thèse s'inscrit dans le cadre général du suivi de courbes de charges, c'est à dire l'analyse de la consommation électrique globale d'une installation afin d'obtenir des informations sur les différentes charges autonomes qui la compose.

Dans le cadre de cette thèse, nous nous intéressons plus particulièrement aux applications dans le secteur résidentiel. Dans ce cas, chaque charge individuelle correspond à un appareil électrodomestique : radiateur, réfrigérateur, chauffe-eau, ... Par ailleurs, les informations que l'on cherche à obtenir peuvent être de différentes natures : type d'appareil, consommation individuelle, présence d'un défaut de fonctionnement ... Les méthodes de suivi de courbe de charges s'attachent à obtenir ces informations de manière aveugle ou non-intrusive, c'est-à-dire en n'utilisant qu'une mesure globale de la consommation électrique. Dans le cadre d'une application résidentielle il est crucial de développer ce type d'approche car instrumenter chacun des appareils est irréalisable pour des raisons autant pratiques qu'économiques.

La motivation principale au développement de telles méthodes est d'accroître notre maîtrise de la consommation électrique. Disposer d'une analyse détaillée de sa consommation électrique constitue une étape préliminaire essentielle à une maîtrise voire une baisse de l'électricité consommée. De plus l'évolution actuelle du réseau électrique vers un pilotage décentralisé (« *smart grid* ») motive la recherche d'informations précises sur la consommation électrique afin d'optimiser à une échelle locale la distribution d'énergie.

Dans le cadre de cette thèse, nous nous intéresserons plus particulièrement à l'analyse des signaux transitoires générés par la mise sous-tension des appareils électrodomestiques. En effet ces transitoires sont caractéristiques de l'appareil mis en route car ils dépendent étroitement de la fonction physique remplie par l'appareil. Ainsi les caractéristiques de la consommation électrique d'un appareil pendant le court instant qui suit sa mise sous tension peuvent être utilisées à l'identification de cet appareil.

Problématique

La problématique industrielle qui guide cette thèse est l'identification des appareils électrodomestiques à partir des signaux transitoires qu'ils génèrent. Cette problématique générale peut se décomposer en deux questions sous-jacentes :

1. Comment caractériser les signaux transitoires générés par la mise sous tension des appareils électrodomestiques ? En particulier il faudra identifier la nature de ces signaux transitoires, c'est-à-dire identifier leurs caractéristiques communes ainsi que leur caractéristiques individuelles. Il s'agit également de préciser les informations physiques *a priori* dont on dispose et de les exploiter pour extraire une signature discriminante des signaux transitoires.

2. Comment différencier les appareils à partir des caractéristiques identifiées de leur régime transitoire ? En d'autres termes, comment utiliser la signature des signaux transitoires afin de détecter automatiquement la nature de l'appareil auquel elle correspond.

Organisation du manuscrit

Le présent manuscrit est organisé en cinq chapitres principaux.

1. Le premier chapitre, intitulé « Suivi de courbes de charges », définit plus précisément la problématique industrielle ainsi que les enjeux industriels et économiques associés. En s'appuyant sur la description de quelques notions générales de l'analyse de courbes de charges, nous établissons un état de l'art de la problématique industrielle.
2. Le deuxième chapitre, « Présentation des signaux électro-domestiques de la base de données SISED », fournit une analyse précise des signaux transitoires de la base de données SISED d'Electricité de France (EDF) qui constituent les données de tests qui serviront à valider les différentes méthodes mise en place par la suite.
3. Le troisième chapitre, « Modèle de régression à transitions régulières », définit le modèle de régression développé pour la représentation des signaux transitoires électro-domestiques. Ce chapitre s'affranchit du cadre applicatif de la thèse en définissant le modèle dans un cadre méthodologique général. Ainsi, la méthode définie est testée sur des signaux standards et comparée à d'autres méthodes de l'état de l'art en matière de régression et de lissage de courbes.
4. Le quatrième chapitre, « Segmentation des signaux transitoires électro-domestiques », présente l'application du modèle de régression à transitions régulières aux signaux de la base de données SISED d'EDF. Certains éléments du modèle sont donc spécifiés par rapport au cadre applicatif de la thèse.
5. Le cinquième chapitre, « Apprentissage de transitoires électro-domestiques », s'attache à définir une méthode de classification des signaux étudiés à partir des paramètres du modèle extraits de chaque signal.

Enfin, le manuscrit se clôt par la description des conclusions, contributions et perspectives de ce travail.

Plan de lecture

Ce manuscrit étant partagé entre aspects applicatifs et aspects méthodologiques, nous proposons au lecteur deux axes de lecture différents en fonction des problématiques qui l'intéressent.

- Le lecteur intéressé par l'application de suivi de courbes de charges pourra lire le chapitre 1 pour une description détaillée de la problématique industrielle puis le chapitre 2 pour se familiariser avec les signaux étudiés. La lecture du paragraphe 3.1.2 du troisième chapitre suffit à une compréhension générale du modèle de régression étudié. Le chapitre 4 permet de se rendre compte des choix effectués pour adapter le modèle de régression à l'application traitée. En particulier les paragraphes 4.1.2 et 4.1.3 exposent les éléments du modèle spécifiques à l'application traitée. Enfin, le chapitre 5, paragraphe 5.2.9, expose les résultats obtenus en terme de classification des signaux transitoires étudiés.
- Le lecteur intéressé par les aspects méthodologiques de la thèse pourra se consacrer à la lecture du chapitre 3, pour ce qui concerne le modèle de régression développé, et à la lecture du

chapitre 5 pour l'utilisation des paramètres de ce modèle afin de classifier automatiquement les signaux transitoires observés.

Suivi de courbes de charges

1.1 Problématique et enjeux du suivi de courbe de charge

1.1.1 Contexte

Durant ces dernières décennies, la consommation en énergie de la France n'a cessé d'augmenter. Une part de plus en plus importante de la consommation énergétique totale est dévolue à l'électricité au détriment du pétrole et du charbon. Cette tendance se vérifie dans tous les secteurs et plus particulièrement dans le secteur résidentiel.

Plus récemment, l'ouverture progressive du marché de l'énergie à la concurrence a induit de profonds changements pour ses acteurs. D'une part, la décentralisation des moyens de production a renforcé la nécessité de disposer de nouveaux outils de gestion des réseaux. D'autre part, l'ouverture à la concurrence a conduit les fournisseurs à proposer de nouveaux services pour se démarquer les uns des autres.

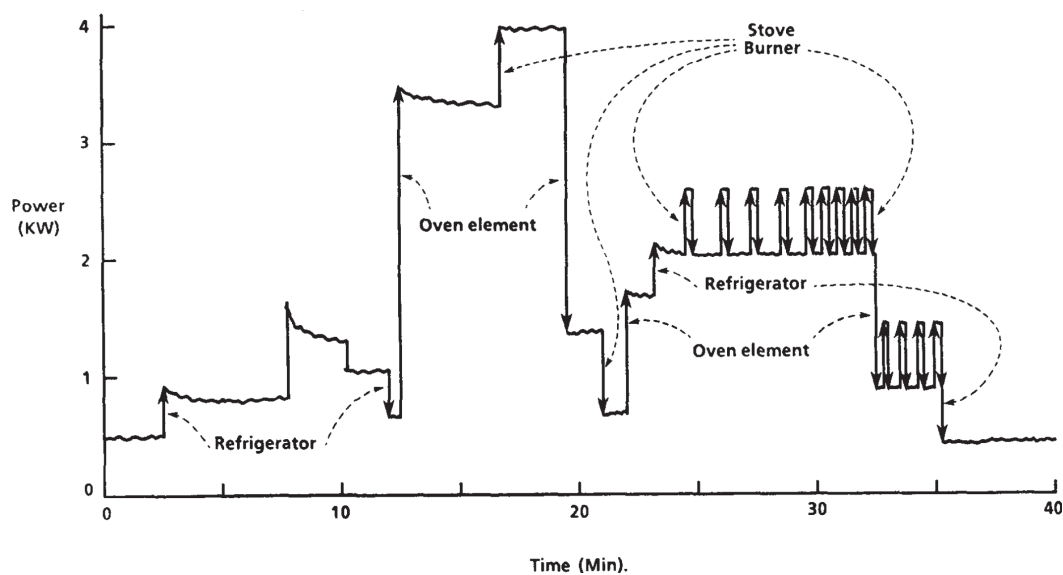
Simultanément, la sensibilité du public aux questions environnementales s'est considérablement accrue. Dans ce cadre, les énergies renouvelables (éolien, solaire) se sont développées et de nouveaux besoins en matière d'intégration de ces énergies au réseau existant ont émergé. Naturellement, la maîtrise, voire la baisse, de la consommation est devenue un enjeu majeur, engendrant le besoin de nouveaux outils en matière d'adéquation entre production et consommation d'une part, en matière d'information des consommateurs d'autre part.

Réseaux intelligents

Cette évolution profonde des enjeux énergétiques a accéléré l'émergence de la thématique des réseaux intelligents (« *smart grids* » en anglais). Sous ce terme générique se dessinerait la prochaine génération de réseaux électriques visant à répondre au mieux aux problématiques associées à la production et la distribution d'électricité. Il n'existe pas de définition unique pour un tel réseau mais les enjeux évoqués sont généralement la fiabilité, la flexibilité, l'efficacité, l'équilibre entre puissance produite et consommée, le lissage voire la baisse de la consommation totale ou encore l'intégration de sources d'énergies renouvelables intermittentes. Le principal défi technologique à relever est le passage de la structure actuelle pyramidale à une forme plus décentralisée du réseau de distribution pour répondre à ces nouveaux enjeux.

Le développement d'un tel système de distribution implique d'interconnecter, d'une part un ensemble de moyens de production, d'autre part un ensemble de consommateurs. Chacun des éléments de ce système possédant ses caractéristiques propres (quantité d'énergie consommée/produite, production/demande contrôlée ou non, prévisible ou non) auxquelles il faut s'adapter au mieux.

La gestion d'un tel système nécessite la mise en place d'un réseau de capteurs remontant en temps réel ces multiples informations quand elles sont disponibles.



source [Hart 1992]

FIGURE 1.1: Exemple de courbe de charge. Les usages qui produisent chacun des événements de la courbe sont annotés.

Compteurs communicants

Le compteur intelligent ou communicant (« *smart meter* » en anglais) s'est imposé naturellement ces dernières années comme un élément essentiel de ce bouleversement énergétique. En France, le déploiement de compteurs intelligents au domicile des consommateurs est prévu dans le cadre du projet "Linky" qui devrait équiper en compteurs communicants la plupart des foyers Français à l'horizon 2020¹. La fonction première de ces compteurs est la relève en temps réel et à distance de la puissance électrique consommée afin de piloter un réseau de nouvelle génération.

Le relevé de ces données ouvrirait également de toutes nouvelles perspectives dans le domaine du suivi de courbe de charge (« *load monitoring* » en anglais).

Suivi de courbe de charge

Le terme "courbe de charge" désigne l'évolution dans le temps de l'ensemble des grandeurs électriques représentatives de la consommation d'une installation. Les grandeurs directement mesurées sont généralement le courant et la tension. D'autres grandeurs électriques comme les puissances actives ou réactives peuvent en être déduites. Un exemple de courbe de charge est représenté sur la figure 1.1. L'évolution temporelle de la puissance électrique est représentée. On distingue nettement l'influence de chaque appareil.

La courbe de charge générale ou agrégée (« *Agregated load* » en anglais) désigne la courbe de charge disponible au point d'entrée d'une installation électrique, typiquement au niveau du compteur électrique pour les applications dans le secteur résidentiel. Elle est la somme des courbes de charge individuelles de chaque usage présent dans l'installation observée. 'Usage' est un terme

1. ERDF, <http://www.erdfdistribution.fr/Linky>

générique désignant un ou plusieurs appareils électriques de même nature (chauffage électrique, réfrigérateur, éclairage).

Une courbe de charge, qu'elle soit agrégée ou individuelle, véhicule de nombreuses informations sur la composition, les caractéristiques et l'évolution de l'installation électrique considérées.

1.1.2 Objectifs

L'objectif général des méthodes de suivi de courbe de charge est d'extraire les diverses informations véhiculées par la consommation électrique d'une installation. Cet objectif général peut être décliné en fonction des informations recherchées qui peuvent être plus ou moins précises et de différentes natures :

- *identification de la présence ou de l'absence d'un usage*, c'est-à-dire l'estimation de la probabilité de présence d'un usage donné,
- *caractérisation d'un usage*, c'est-à-dire l'estimation de grandeurs intrinsèques des appareils (puissance nominale par exemple) ou l'estimation du nombre de composantes d'un usage (nombre de radiateurs électriques dans une maison par exemple),
- *décomposition ou désagrégation de la courbe de charge*, c'est-à-dire l'estimation des courbes de charge individuelles à partir de la courbe de charge agrégée, permettant par exemple d'estimer la consommation individuelle de chaque usage,
- *détection et éventuellement caractérisation d'un défaut ou du comportement anormal d'un appareil*, comme par exemple la surconsommation d'un réfrigérateur.

Le point le plus couramment étudié est la décomposition de la courbe de charge générale (« *Disaggregation Load* » en anglais), car il permet de traiter la plupart des autres problèmes (estimation de la puissance nominale des appareils, détection de défauts) *a posteriori* en utilisant les courbes de charge individuelles estimées.

1.1.3 Motivations

Les motifs du développement de méthodes de suivi de courbe de charge sont multiples et différents en fonction du point de vue adopté. Nous différencierons par la suite les intérêts du consommateur de ceux du fournisseur ou du producteur.

Motivations pour le consommateur

Les bénéfices des méthodes de suivi de courbe de charge pour le consommateur portent essentiellement sur la maîtrise de sa consommation électrique. En effet il peut ainsi disposer, en plus du relevé de sa consommation totale, d'une estimation de la consommation individuelle de chaque usage. Or la plupart des utilisateurs ont une vision assez floue des détails de leur consommation énergétique. L'accès à des informations plus précises permettrait de mieux gérer la consommation en adaptant leur comportement afin d'économiser sur la dépense totale. De nombreuses études montrent que fournir ce type d'information aux consommateurs (facture détaillée, consommation électrique affichée en temps réel) permet de réduire leur consommation électrique. Selon certaines études [Darby 2006, Neenan 2009], cette réduction pourrait atteindre 15% de la consommation totale. Cette économie d'énergie peut naturellement être valorisée d'un point de vue écologique, mais également d'un point de vue économique, d'autant plus dans un contexte de hausse des prix de l'énergie.

Un autre aspect intéressant le consommateur est la surveillance systématique et à distance de son installation électrique, lui permettant d'être averti automatiquement d'une défaillance ou encore

de la mise en route ou de l'arrêt inopiné d'un appareil à son domicile. On peut également imaginer des applications en matière de pilotage à distance de l'installation électrique.

Motivation pour le producteur

Les informations détaillées obtenues sur la consommation des usages servent au producteur afin d'identifier le profil énergétique exact de chaque consommateur et de pouvoir ainsi prédire de façon plus fiable sa consommation. L'intérêt est de pouvoir prédire la consommation globale de tous les usagers d'affiner la compréhension des variations de la consommation en période de pointe. Ces informations sont utiles au producteurs pour améliorer sa gestion des moyens de production en temps réel.

En particulier, si l'on dispose depuis longtemps de la consommation totale et d'outils de prédiction à l'échelle de la France, on ne dispose pas de tels outils à des échelles plus locales. Or de tels outils sont nécessaires au développement de réseaux décentralisés de nouvelle génération. Les méthodes de suivi de courbe de charge permettraient la mise en place de tels moyens.

Motivation pour le fournisseur/commercialisateur

Pour le fournisseur, l'intérêt de ces méthodes réside dans le fait de fournir à ses clients des services supplémentaires. Ces services peuvent être sous forme d'une facture détaillée usage par usage, ou encore sous forme d'audit énergétique alimenté par les informations obtenues par surveillance. Dans le prolongement de ces services, l'identification du profil de l'utilisateur permettrait de lui proposer des moyens ciblés de maîtrise de l'énergie.

1.1.4 Protection de la vie privée

Les nombreuses applications potentielles du suivi de courbe de charge peuvent soulever des craintes en matière de protection de la vie privée. En effet, la courbe de charge d'un foyer recèle de nombreuses informations sur les usages et habitudes de ses occupants. Il est ainsi relativement facile de détecter la présence ou l'absence d'occupants dans une maison, l'heure à laquelle ils se lèvent, et même le programme télévisé ou le film qu'ils regardent [Carluccio 2011] !

Du fait de ces risques, une courbe de charge est une information sensible et il est préférable que les données qu'elle contient soient protégées ou du moins qu'on ne puisse faire le lien entre une courbe de charge et l'identité des usagers. En 2010, un communiqué de la CNIL (Commission Nationale de l'Informatique et des Libertés) recommandait de ne stocker, en la matière, que les données nécessaires à l'application visée et d'adopter les mesures de sécurité qui s'imposent pour protéger ses données².

Enfin notons qu'il existe déjà des études scientifiques proposant des méthodes permettant de "brouiller" une courbe de charge pour neutraliser le bon fonctionnement des méthodes de surveillance afin de protéger l'utilisateur d'intrusions non désirées [Laughman 2003, Varodayan 2011].

1.2 Généralités sur l'analyse de courbes de charge

Le domaine du suivi de courbe de charge fait appel à quelques notions élémentaires d'électronique de puissance et de réseaux électriques. Dans ce paragraphe nous rappelons ces notions de bases ainsi que quelques hypothèses fondamentales pour les travaux qui suivent.

2. CNIL, 5 août 2010, <http://www.cnil.fr/en-savoir-plus/fiches-pratiques/fiche/article/les-compteurs-electriques-intelligents-en-questions/>

1.2.1 Courant et tension

On considère la distribution de l'énergie électrique à une installation domestique. La source d'énergie est le réseau électrique de distribution. La charge est constituée de l'ensemble des appareils sur le réseau domestique. Ces appareils sont branchés en parallèle.

On fait l'hypothèse que le réseau électrique de distribution est un générateur parfait de tension alternative. C'est-à-dire que quelle que soit la charge, la tension au niveau du compteur électrique (et donc de tous les appareils puisqu'ils sont branchés en parallèle) est une sinusoïde d'amplitude et de fréquence constante. La fréquence fondamentale F_f est connue (respectivement $F_f = 50$ Hz en Europe, Asie et Afrique et $F_f = 60$ Hz aux États-Unis). L'amplitude efficace de cette sinusoïde est $V_e = 230V$. On note donc :

$$v(t) = \sqrt{2}V_e \cos(2\pi F_f t + \theta_0). \quad (1.1)$$

Cette hypothèse est théoriquement fautive puisque un générateur parfait de tension ne peut exister. En particulier, l'enclenchement d'appareils sur le réseaux génère de faibles perturbations de la tension théorique $v(t)$. Il existe d'ailleurs des travaux visant à utiliser ces perturbations pour caractériser les appareils branchés [Cox 2006, Patel 2007]. Cependant ces perturbations sont suffisamment faibles pour être négligées dans le cadre de nos travaux. On supposera que seul le courant est caractéristique de la charge en présence, la tension étant fixée par le réseau de distribution.

Le courant appelé par la charge n'est pas nécessairement sinusoïdal. En effet de nombreux appareils domestiques comportent des charges non linéaires (tout appareil doté d'une alimentation à découpage par exemple). Le courant appelé est alors simplement périodique de fréquence fondamentale F_f puisque que la tension est elle-même périodique. La représentation en série de Fourier est adaptée à la modélisation du courant appelé $i(t)$, on notera :

$$i(t) = \sum_{n=1}^{\infty} \sqrt{2}I_n \cos(2\pi F_f n t + \theta_n), \quad (1.2)$$

où I_n et θ_n sont les coefficients de Fourier de $i(t)$. On suppose que la valeur moyenne du courant sur une période est nulle.

1.2.2 Puissance

La puissance active est la puissance moyenne consommée sur une période

$$P = \frac{1}{T} \int_0^T v(t)i(t)dt, \quad (1.3)$$

où $T = \frac{1}{F_f}$ est la période des signaux de tension et de courant.

On remarque que seule la composante fondamentale du courant participe à la puissance active. En effet, en prenant en compte les expressions de $i(t)$ et $v(t)$ obtenues précédemment, on obtient :

$$P = V_e I_1 \cos \phi, \quad (1.4)$$

avec $\phi = \theta_0 - \theta_1$.

Dans le cas où la charge est purement résistive, le courant et la tension sont parfaitement en phase ($\cos(\phi) = 1$). Alors la puissance instantanée $v(t)i(t)$ est positive à tout instant t , c'est-à-dire que l'énergie ne circule que de la source vers la charge. Cependant dans le cas général, la charge comporte des éléments inductifs ou capacitifs ($\cos(\phi) < 1$). Alors il existe des instants t où la

puissance instantanée $v(t)i(t)$ est négative, c'est-à-dire que la charge renvoie temporairement de l'énergie vers la source. Pour caractériser ce phénomène on définit la puissance réactive :

$$Q = V_e I_1 \sin \phi. \quad (1.5)$$

On a remarqué que les harmoniques du courant ne participent pas en moyenne à la transmission de puissance de la source vers la charge. Cependant ces grandeurs dépendent de la nature de la charge et peuvent être utiles à sa caractérisation dans le cadre du suivi de courbe de charge. On définit les puissances actives P_n et réactives Q_n harmoniques du courant :

$$P_n = V_e I_n \cos(\theta_0 - \theta_n), \quad (1.6)$$

$$Q_n = V_e I_n \sin(\theta_0 - \theta_n), \quad (1.7)$$

pour tout $n > 1$.

1.2.3 Régime permanent et régime transitoire

La relation entre la tension et le courant appelé par la charge est régie par une équation différentielle dont l'expression exacte dépend des composants de la charge et de leur organisation. La solution de cette équation différentielle est la somme de deux parties, une solution libre et une solution forcée. La première correspond à la réponse de la charge à une perturbation ponctuelle, la seconde à la réponse de la charge à une contrainte extérieure continue.

En électricité, ces deux parties de la solution correspondent respectivement au régime transitoire et au régime permanent. Lors d'une mise en fonctionnement, ou lors d'un changement de mode de fonctionnement, la charge est soumise à une perturbation instantanée ce qui génère un régime transitoire et la tension du réseau lui est appliquée ce qui génère le régime permanent. Le régime transitoire disparaît avec un temps caractéristique qui dépend de la charge (généralement quelques secondes). Le régime permanent dure jusqu'à ce que l'appareil soit à nouveau perturbé (changement de mode de fonctionnement ou arrêt de l'appareil).

1.3 État de l'art des méthodes de suivi de courbe de charge

Les premiers travaux en matière de surveillance de courbe de charge ont été entrepris au M.I.T dans les années 1980 par Georges W. Hart [Hart 1992]. Ces travaux ont connu de nombreux développements et extensions par la suite. Le nombre croissant de publications sur le sujet ces dernières années témoigne de l'intérêt de la communauté scientifique pour cette thématique. En particulier, un article de synthèse paru en 2011 résume une partie des travaux existants dans ce domaine [Zeifman 2011] (voir également les actes de conférences suivants en complément [Jiang 2011, Najmeddine 2008, Du 2010]).

1.3.1 Signature électrique des usages

L'un des enjeux scientifiques principaux du suivi de courbe de charge est l'identification de la signature électrique des usages. On entend ici par "signature" le comportement électrique de l'usage tel qu'il peut être mesuré au niveau du compteur (ou d'un autre point d'intérêt considéré). Dans l'état de l'art, les méthodes développées exploitent des signatures de natures très variables. On se référera à [Hart 1992, Liang 2010] où des inventaires exhaustifs de ces signatures sont réalisés.

Signature macroscopique

On appelle "signature macroscopique" toute signature extraite à une fréquence d'échantillonnage faible par rapport à la fréquence fondamentale des signaux de courant et tension F_f . Cela correspond typiquement à des fréquences de l'ordre de 1 Hz. À de telles échelles, on observe essentiellement les changements globaux de la consommation dus à des changements d'états des appareils (événements ON/OFF, changement de mode de fonctionnement, etc.). On peut alors utiliser la valeur des sauts pour identifier les appareils, ou encore essayer d'identifier des cycles de fonctionnement répétés au cours du temps comme le cycle compression/détente d'un réfrigérateur (voir un exemple sur la fig. 1.2). La puissance moyenne consommée ainsi que la durée de fonctionnement typique des usages sont les exemples les plus courants de signatures macroscopiques.

Signature microscopique

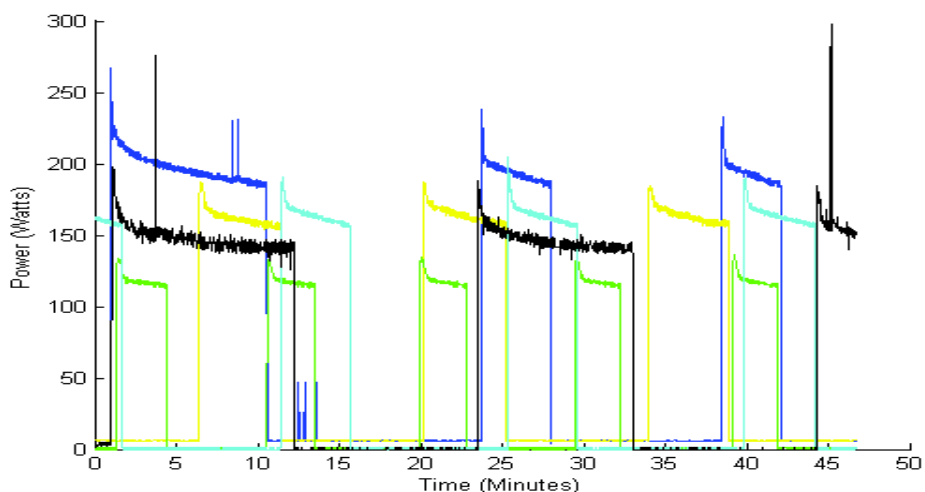
On appelle "signature microscopique" toute signature extraite à une fréquence d'échantillonnage élevée par rapport à F_f . Cela peut correspondre à des fréquences allant de quelques kHz à 1 MHz. À ces échelles, des informations telles que la forme du régime transitoire lors des changements de consommation sont maintenant observables. De plus comme on dispose de suffisamment de points par cycle, il devient possible d'utiliser la forme des ondes de courant, d'admittance ou de puissance instantanée pour caractériser des appareils que ce soit dans le domaine temporel ou fréquentiel.

1.3.2 Classification générale des usages électriques

Le groupe de travail du M.I.T propose une classification des usages électriques en fonction du profil de la courbe de charge individuelle de chacun des appareils. Celle-ci permet de rendre compte de la haute variabilité que peut présenter l'allure des courbes de charge individuelles de chaque usage. Cette classification est la suivante :

1. Les usages à consommation permanente : cette catégorie regroupe les usages consommant une puissance constante, 24 heures sur 24 et 7 jours sur 7. Une alarme incendie fait typiquement partie de cette catégorie.
2. Les usages ON/OFF : cette catégorie est constituée d'appareils possédant deux états seulement. Un état OFF où ils ne consomment pas, un état ON où ils consomment une puissance approximativement constante. De nombreux appareils courant dans le secteur résidentiel appartiennent à cette catégorie (bouilloire, ampoule électrique, grille pain).
3. Les usages à nombre d'états fini : cette catégorie inclut les usages passant par plusieurs états distincts où ils consomment une puissance constante. Typiquement, une machine à laver ou un sèche-linge font partie de cette catégorie.
4. Les usages à nombre d'états infini : cette catégorie inclut les usages dont la puissance consommée peut varier continûment sur une certaine plage de valeur, en général en fonction d'un réglage utilisateur. Cela inclut, par exemple, les aspirateurs à puissance variable.

Il est important de noter que cette classification est faite du point de vue de la consommation électrique de l'appareil et non des modes de fonctionnement qu'ils offrent à l'utilisateur, même si ces deux points de vue coïncident éventuellement. Par exemple, un réfrigérateur, allumé en permanence, semble, du point de vue de l'utilisateur appartenir à la première catégorie. Cependant fonctionnant selon une alternance de phase de compression (ou une puissance constante est appelée) et de phase de détente (ou aucune puissance n'est nécessaire) du fluide réfrigérant, cet usage appartient à la seconde catégorie.



source : blog O. Parson <http://op106phd.blogspot.fr/>

FIGURE 1.2: Exemple de courbe de charge décomposée à échelle macroscopique tirée de la base de données REDD [Kolter 2012]. Chaque couleur de courbe correspond à la courbe de charge individuelle d'un réfrigérateur.

Les méthodes développées jusqu'à présent se concentrent sur l'étude des usages appartenant à la deuxième et à la troisième catégorie. En effet les usages de la première catégorie ne se manifestent au niveau de la courbe de charge que par un niveau de consommation constant et permanent. À moins de disposer d'informations *a priori*, ces appareils sont indétectables. Tout au plus, on peut déterminer leur absence si la courbe de charge présente des plages de consommation nulle. Quant aux usages de la quatrième catégorie, leur comportement complexe et généralement non reproductible, rend leur détection/caractérisation extrêmement difficile, si bien que leur cas est rarement traité dans la littérature scientifique.

1.3.3 Analyse macroscopique

Dans les travaux pionniers sur les courbes de charge, ce sont les signatures macroscopiques des usages qui sont exploitées pour la décomposition de courbe de charge. En effet ces signatures sont plus faciles d'accès puisque des appareils standards (voltmètre, ampèremètre, wattmètre) permettent de réaliser des mesures à ces fréquences. De plus, comme on peut le constater sur la figure 1.2, il existe, au sein d'une même famille d'appareils (ici les réfrigérateurs) des caractéristiques communes au niveau macroscopique : les courbes associées à chaque appareil présentent des cycles de fonctionnement dont les formes sont similaires, même si l'amplitude et la fréquence de ces cycles varient selon les réfrigérateurs. Tout l'enjeu de l'analyse macroscopique est donc de capturer ces caractéristiques communes en dépit de leur variabilité d'un appareil à l'autre.

Identification de changements des puissances active et réactive

La méthode développée par Hart [Hart 1992] analyse conjointement deux grandeurs, la puissance active et la puissance réactive, à une fréquence d'échantillonnage de 1 Hz. Sa méthode

consiste en cinq étapes :

1. un algorithme type "détection de contours" extrait les changements de régime à partir de la puissance active,
2. un algorithme de classification non-supervisée (« *clustering* ») regroupe ces changements de régime dans un espace à deux dimensions, la puissance active P et la puissance réactive Q ,
3. les groupes symétriques par rapport à l'origine de l'espace P/Q sont associés, ils correspondent aux événements ON et OFF d'un même usage,
4. cette étape, dite de résolution d'anomalie, permet d'associer les groupes restants, soit en les fusionnant avec des groupes déjà associés, soit en étant associés entre eux.
5. l'étape finale consiste à donner un nom à chaque association de groupe d'événements en utilisant des données d'apprentissage.

Cet algorithme permet essentiellement de caractériser les appareils de type ON/OFF. Pour pouvoir caractériser des appareils à états multiples, la méthode proposée dans [Hart 1992] est de construire des machines à états finis émettant un message (c-à-d un changement dans le plan P/Q) à chaque changement d'état et de tenter de décoder la séquence émise. Des extensions de l'algorithme de Viterbi [Viterbi 1967] pour cette tâche spécifique ont été proposées dans [Bouloutas 1991, Hart 1993], même si ces travaux semblent en être restés à un stade théorique.

L'idée d'utiliser les changements macroscopiques de la puissance consommée a ensuite été reprise dans de nombreux travaux. Dans [Powers 1991] ou [Farinaccio 1999], cette idée est utilisée en conjonction de l'exploitation de règles de décision propres aux usages pour reconnaître les appareils. Ces méthodes donnent de bons résultats mais nécessitent une période d'apprentissage longue (de l'ordre d'une semaine). De plus les règles sont construites autour des spécificités de quelques usages et semblent donc difficiles à généraliser à n'importe quel scénario.

Dans [Marceau 2000], l'auteur propose d'utiliser des connaissances *a priori* sur les durées moyenne et maximale de fonctionnement de chaque usage en combinaison avec un algorithme de détection de changements dans la puissance active ce qui permet de diminuer le taux d'erreur de l'algorithme de classification.

Enfin citons les travaux de Baranski [Baranski 2004a, Baranski 2004b, Baranski 2003], qui propose un algorithme identifiant les changements les plus fréquents puis construisant des machines à états finis hypothétiques à partir de ces changements. Cet algorithme présente des points communs avec celui du M.I.T, mais traite l'ensemble du problème de manière globale, ce qui permet de détecter conjointement les événements ON et OFF d'un même appareil, ou d'ignorer les changements peu fréquents pour se focaliser sur les contributions les plus importantes à la courbe de charge totale.

Utilisation des harmoniques d'ordre supérieur

Une des limites principales des méthodes fondées seulement sur les puissances active et réactive est que lorsque le nombre d'usages en présence est important, il est fréquent que les signatures de plusieurs usages se superposent dans le plan P/Q . L'utilisation des puissances harmoniques d'ordre supérieur pour caractériser les usages permet de surmonter ce problème [McLaughlin 2011]. Cela consiste à généraliser les méthodes précédentes aux puissances harmoniques (P_n, Q_n) d'ordre $n > 1$. La généralisation ne présente pas de difficulté, on passe simplement d'un espace des signatures de dimension 2 à un espace de dimension $2n$.

Dans la plupart des cas, seuls les harmoniques d'ordre impair présentent une puissance significative. Cela est attribuable à la propriété de symétrie de la réponse des usages électriques par rapport

au signe de la tension appliquée (cf Annexe A.2). Ainsi on peut caractériser chaque usage par le vecteur des puissances actives et réactives associées aux harmoniques d'ordre impairs. Le vecteur de caractéristiques obtenu peut par la suite être utilisé pour entraîner un algorithme d'apprentissage statistique standard : réseau de neurones [Srinivasan 2006] séparateur vaste marge (SVM) [Onoda 2000] ou machine à noyaux [Murata 2001]. Ces méthodes requièrent cependant d'avoir identifié au préalable les événements ON/OFF de la courbe de charge.

Une méthode s'affranchissant de cette étape préliminaire est proposée dans [Yi-xin 2008]. Elle consiste à modéliser la courbe de charge comme un mélange linéaire des signatures de chaque usage, puis à utiliser un algorithme d'optimisation, (typiquement des moindres carrés non-négatifs « *non-negative least square* ») pour identifier les coefficients du mélange. On peut également rapprocher de ces derniers travaux la méthode proposée dans [El Guedri 2009]. Quoique n'utilisant que la puissance active pour caractériser la courbe de charge, une méthode globale d'optimisation (algorithme de chaîne de Markov par Monte-Carlo) est proposée permettant de s'affranchir de la détection individuelle de chaque événement de la courbe de charge.

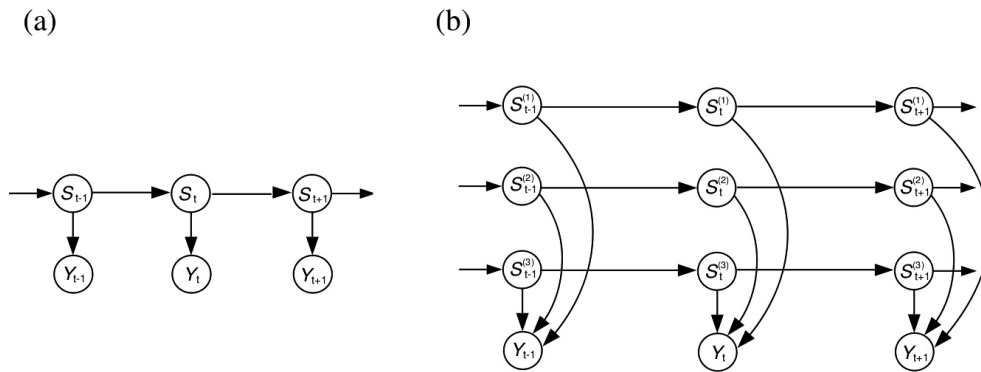
Modélisation par chaînes de Markov à états cachés

Dernièrement, plusieurs auteurs se sont intéressés à la modélisation de courbes de charge par des chaînes de Markov cachées [Parson 2011, Parson 2012, Kim 2011, Durand 2003]. Les chaînes de Markov cachées sont des modèles probabilistes fréquemment utilisés en analyse des séries temporelles, par exemple en reconnaissance de la parole [Rabiner 1989]. Un modèle de Markov caché est un modèle de Markov dont les états ne sont pas directement observables. A la place, chaque état est caractérisé par une distribution de probabilité qui modélise la fonction observée par rapport à chaque état. La figure 1.3(a) propose une représentation graphique d'un modèle de Markov caché. L'inférence des paramètres du modèle lors de la phase d'apprentissage peut être réalisée de manière classique par un algorithme de type EM (« *Expectation-Minimization* ») plus connu dans ce cas sous le nom d'algorithme de Baum-Welch. Étant donné cet apprentissage, le problème de la détection de la configuration des états la plus probable est alors classiquement résolu à l'aide d'algorithmes de programmation dynamique de type Viterbi [Viterbi 1967] ou de manière plus générale « *forward-backward* » [Rabiner 1989].

Dans le cadre de la modélisation de courbe de charge, l'idée consiste à expliquer la courbe de charge agrégée comme la partie observable d'une chaîne de Markov. Dans [Durand 2003], les états cachés correspondent à des plages d'activités différentes au cours de la journée (nuit, préparation du repas, lessive, etc.). Cependant il est difficile d'interpréter précisément ces états cachés et plus encore de les relier directement aux plages de fonctionnement individuelles des usages.

La prise en compte de plusieurs causes indépendantes (les usages) pour expliquer la variable observée s'avère difficile dans le cadre des chaînes de Markov cachées standard. Une solution consisterait à définir un vecteur d'état dont chaque composante correspondrait à un usage. Cependant la complexité de l'algorithme d'inférence des paramètres croît exponentiellement avec la taille de ce vecteur ce qui rend cette solution trop coûteuse algorithmiquement. Dans [Kim 2011, Kolter 2012], les auteurs préconisent l'utilisation de chaînes de Markov cachées factorielles [Ghahramani 1997]. Cette extension du modèle de Markov caché standard, permet de prendre en compte explicitement l'hypothèse d'indépendance entre les séquences d'états cachés. La figure 1.3(b) propose une représentation graphique d'un tel modèle. Ainsi chaque variable d'état dépend uniquement de son propre passé et non pas des autres variables d'états. Un algorithme d'inférence efficace peut alors être obtenu en approximant l'étape E (« *Expectation* ») soit par un échantillonnage de Gibbs soit par une approximation variationnelle [Ghahramani 1997].

Pour la modélisation de courbe de charge, chaque séquence d'état caché représente l'évolution



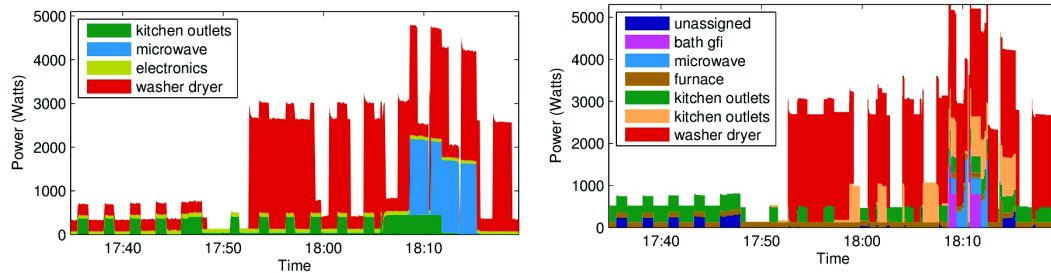
source [Ghahramani 1997]

FIGURE 1.3: Représentation graphique de modèles de chaînes de Markov cachées : chaque nœud du graphe correspond à une variable, les variables d'états sont notées s et les variables observées y , les arêtes représentent des dépendances conditionnelles entre les variables - (a) chaîne de Markov cachée standard (b) chaîne de Markov cachée factorielle munie de 3 processus cachés

de l'état (ON, OFF, etc.) d'un unique appareil. Restaurer les séquences d'états cachés revient à estimer directement les plages de fonctionnement de chaque usage. La figure 1.4 présente un exemple de décomposition de courbe de charge par une chaîne de Markov cachée factorielle.

Il existe plusieurs limites à ce modèle de Markov caché factoriel. D'une part le temps écoulé dans chaque état pour chaque usage est nécessairement modélisé par une loi géométrique (cela découle directement de la propriété de Markov). Cela implique que les courtes durées d'activation sont favorisées dans la modélisation des usages : une durée de 1 seconde sera plus probable qu'une durée de 1 minute, ce qui n'est pas le cas en réalité. D'autre part, les chaînes des états cachés sont nécessairement indépendantes alors qu'il existe parfois des dépendances entre les plages de fonctionnement des usages (le lecteur DVD sera probablement allumé en même temps que la télévision par exemple). Enfin ce modèle ne prend pas en compte des facteurs additionnels éventuellement disponibles, comme l'heure de la journée ou la température externe. Les auteurs de [Kim 2011] proposent une extension du modèle original de [Ghahramani 1997] permettant de modéliser explicitement la distribution du temps écoulé dans chaque état et de rajouter des dépendances conditionnelles entre les variables d'états et des facteurs externes en se basant respectivement sur les travaux de [Russell 1985] et [Brand 1997].

Notons que ces méthodes nécessitent une phase d'apprentissage pour être mises en oeuvre. Dans [Parson 2011, Parson 2012], l'auteur propose d'utiliser des connaissances *a priori* sous la forme de modèles génériques d'usages, pour pouvoir éviter cette phase d'apprentissage et obtenir des méthodes de décomposition de courbe de charge non-supervisées. Ces connaissances *a priori* concernent typiquement la durée de fonctionnement et la puissance consommée de chaque usage. Ces connaissances restent cependant souvent trop générales afin d'étiqueter précisément la nature des usages détectés.



(a) Courbes de charge réelles. La courbe agrégée correspond à l'enveloppe supérieure du graphe.

(b) Courbes de charge estimées

source [Kolter 2012]

FIGURE 1.4: Exemple de décomposition de courbe de charge par un modèle de Markov caché

Limites des approches macroscopiques

Au cours des derniers paragraphes, un état de l'art des méthodes de suivi de courbes de charge utilisant les signatures macroscopiques des usages a été présenté. Il est important de noter qu'il existe des limites inhérentes à ce type de méthodes. En effet, puisque les courbes de charge utilisées sont collectées à de faibles fréquences, le risque que deux usages changent d'état simultanément (au même instant d'échantillonnage) est important. De plus les grandeurs accessibles à ces fréquences sont essentiellement les puissances active et réactive du fondamental ainsi qu'éventuellement des puissances d'ordre supérieur. À ces fréquences, ces grandeurs s'avèrent parfois insuffisantes pour distinguer deux usages différents.

1.3.4 Analyse microscopique

Pour surmonter les difficultés inhérentes à l'analyse macroscopique, de nombreux auteurs se sont penchés sur l'utilisation de caractéristiques microscopiques souvent en complément de l'analyse macroscopique.

L'analyse microscopique des signaux permet de récolter des informations relatives au régime transitoire des usages observés. Ces informations sont particulièrement intéressantes du point de vue de l'analyse de la courbe de charge car les caractéristiques du régime transitoire sont directement reliées à la tâche physique exécutée par la charge [Laughman 2003]. Ainsi le régime transitoire d'une lampe à filament et celui d'un réfrigérateur sont extrêmement différents car la tâche consistant à chauffer le filament est fondamentalement différente de la mise en route du compresseur. La figure 1.5 représente le régime transitoire de 5 modèles différents de réfrigérateur. On peut constater qu'il existe indéniablement des caractéristiques communes à ces signaux, même si l'amplitude ou la forme exacte du signal varient d'un réfrigérateur à l'autre. L'intérêt de l'analyse microscopique sera donc de pouvoir étiqueter chaque changement d'état identifié sur la courbe de charge agrégée avec le type d'appareil.

L'analyse du régime transitoire dans son intégralité est une tâche complexe. En effet, contrairement au régime permanent, le régime transitoire ne se réduit pas à quelques caractéristiques simples (changement dans le plan puissance active/réactive). Les méthodes proposées dans la littérature pour résoudre ce point différent beaucoup. Certains auteurs proposent de réduire le régime transitoire à une ou deux caractéristiques élémentaires, d'autres proposent d'apprendre des portions entières de

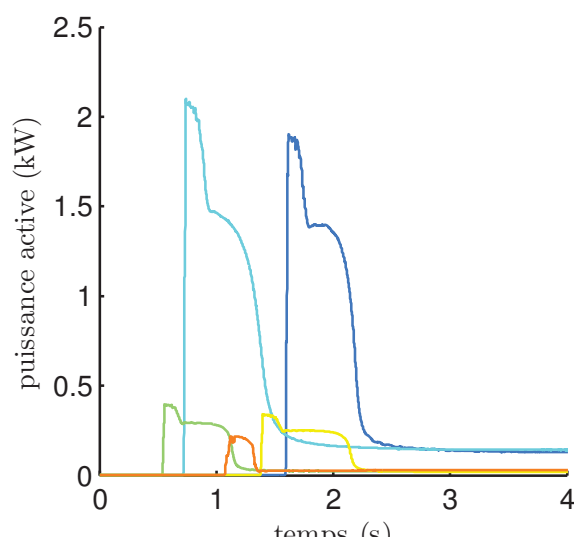


FIGURE 1.5: Signature microscopique : évolution de la puissance active en fonction du temps en régime transitoire. Chaque couleur de courbe correspond à un réfrigérateur différent

transitoires.

Utilisation de caractéristiques élémentaires

Dans [Sultanem 1991], l'auteur propose d'utiliser simplement la longueur du régime transitoire combinée à des caractéristiques macroscopiques pour caractériser le régime transitoire.

Alternativement, dans [Chang 2008, Chang 2010], les auteurs proposent de caractériser le régime transitoire par l'énergie consommée lors de la mise sous tension d'un usage (« *turn-on transient energy* ») :

$$W_t = \int_{t_s}^{t_s + \delta t} v(t)i(t)dt, \quad (1.8)$$

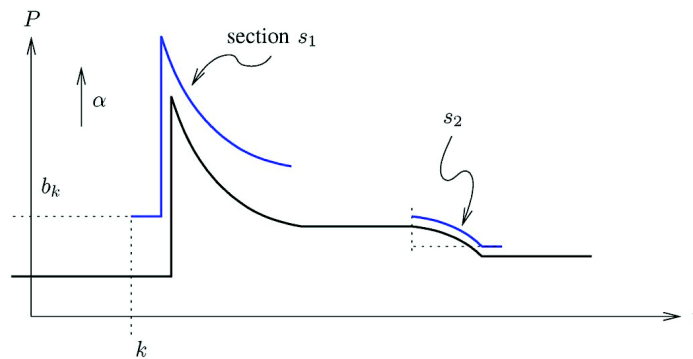
où chaque instant de départ t_s est testé et la longueur du régime transitoire δt est calculée durant une phase d'apprentissage. L'énergie W_t , la puissance active P et la puissance réactive Q en régime permanent sont ensuite utilisées pour entraîner un réseau de neurones à la reconnaissance d'appareils.

Régression de la forme du transitoire

Plutôt que de se limiter à l'extraction de quelques caractéristiques du régime transitoire, certains auteurs ont cherché à modéliser entièrement la forme des variations de la puissance active en régime transitoire.

Dans [Cole 1998], les auteurs proposent de modéliser la puissance active en régime transitoire par un changement abrupt « *edge* », une pente douce « *slope* » suivi du régime permanent constant. Ce modèle de régression, quoique très simple, permet de reconnaître certaines classes d'appareils.

Dans [Leeb 1995, Norford 1996, Shaw 2008], cette modélisation repose sur les sections du signal transitoire qui présentent des variations significatives de la puissance active, appelées « *v-*



source [Shaw 2008]

FIGURE 1.6: Modélisation du régime transitoire par la méthode des « *v-sections* ». La puissance active P est tracée en fonction du temps t . Deux « *v-sections* » sont notées s_1 et s_2 . Les paramètres k et b_k représentent respectivement la position et l'amplitude initiale des « *v-sections* ». α est le paramètre global de gain d'amplitude.

sections ». Ces « *v-sections* » sont dans un premier temps apprises pour chaque classe de transitoire sur une base d'exemples. Puis, durant la phase de reconnaissance, l'algorithme cherche à identifier des séquences de « *v-sections* » dans le signal observé. Pour cela, la position k et l'amplitude initiale b_k de chaque « *v-section* » s_k sont estimées conjointement à un paramètre de gain d'amplitude α qui s'applique à toutes les « *v-sections* » simultanément. Cette estimation est réalisée par un algorithme des moindres carrés. La figure 1.6 schématise le fonctionnement de cette méthode. Cette approche a été appliquée avec succès aussi bien à la surveillance de bâtiments industriels et commerciaux [Leeb 1995, Norford 1996] qu'à l'identification d'usages électriques dans une automobile [Shaw 2008].

Notons enfin l'approche de [Cox 2006] qui consiste à appliquer la modélisation précédente en « *v-sections* » au signal de tension au lieu de la puissance active. En effet, l'enveloppe du signal de tension présente d'infimes variations ($\pm 0.01V$) similaires aux variations du courant. L'intérêt de la méthode est de ne nécessiter qu'une seule mesure (tension) au lieu de deux (courant et tension).

Utilisation du bruit très haute fréquence généré par les usages

Une solution originale au problème de la caractérisation des usages électriques a été apportée dans [Patel 2007]. Elle consiste à exploiter la signature très haute fréquence générée par la mise sous tension ou la mise hors tension d'usages de type ON/OFF sur la tension mesurée.

L'approche des auteurs est de mesurer la tension à une fréquence de 1000 kHz et d'extraire des séquences de 2048 échantillons (soit environ 2ms). Une transformée de Fourier est réalisée sur chaque fenêtre d'observation, permettant d'obtenir un vecteur de caractéristiques. Le calcul de la distance euclidienne entre chaque vecteur permet d'identifier le début et la fin du régime transitoire qui correspondent aux instants où ce vecteur de caractéristiques varie. Un vecteur de caractéristiques moyen est ensuite calculé sur la durée du régime transitoire.

Le vecteur de caractéristiques moyen ainsi obtenu est finalement utilisé pour entraîner un classifieur de type SVM. Les taux de classification obtenus varient de 85% à 90% sur divers usages

dans le secteur résidentiel. Cependant les signatures obtenues par une telle approche sont fortement dépendantes de l'appareil appris et de la configuration du réseau électrique [Gupta 2010], ce qui limite grandement la généralité de la méthode.

Limites des approches microscopiques

Nous avons relevé dans les derniers paragraphes différentes méthodes d'analyse microscopique. Ces méthodes permettent de lever certaines ambiguïtés de l'analyse macroscopique en permettant l'accès à de nombreuses caractéristiques potentielles liés au régime transitoire. Cependant ces méthodes nécessitent des fréquences d'échantillonnage élevées ce qui implique une acquisition plus coûteuse des signaux et surtout le stockage d'un grand nombre de données.

1.3.5 Base de données de courbes de charges

L'intérêt accru de la communauté scientifique à la thématique des courbes de charge a donné lieu à la création de plusieurs bases de données scientifiques. Ces bases de données sont constituées de relevés de courbes de charge globales ainsi que des courbes de charge individuelles étiquetées qui constituent ainsi la vérité terrain.

Des bases de données privées ont été créées par des organismes de recherche et/ou des industriels (tels que EDF - Électricité de France, cf chapitre 2) afin de tester les méthodes et algorithmes développés en interne. A notre connaissance, il existe deux bases de données publiques à l'été 2012 :

- REDD : « *the Reference Energy Dissagregation Data set* »³ [Kolter 2012]
Cette base de données contient les relevés de la courbe de charge agrégée et de la vérité terrain pour 6 foyers en basse fréquence (1 Hz) et pour 2 foyers en haute fréquence (15 kHz). La durée des mesures varie de quelques jours à quelques mois.
- BLUED : « *Building-Level Fully labeled Electricity Disaggregation dataset* »⁴ [Anderson 2012]
Cette base de données contient les relevés de la courbe de charge agrégée et de la vérité terrain pour 16 foyers en haute fréquence (12 kHz). La durée des mesures est d'environ une semaine.

Il est important de préciser que puisque ces bases de données ont été publiées simultanément à la rédaction de ce manuscrit durant l'été 2012, nous n'avons pas encore pu en faire usage afin de tester nos méthodes.

1.3.6 Applications industrielles commercialisées

Si la thématique de suivi de courbes de charge est un problème de recherche actuel dans le milieu académique, le monde industriel s'est déjà intéressé de près à ces thématiques. Ces dernières années, un nombre conséquent de produits industriels ayant pour but d'analyser la consommation électrique des usagers a vu le jour. Évidemment, ces produits sont protégés par des brevets et il est donc difficile d'obtenir des informations techniques précises sur leur fonctionnement. A notre connaissance, voici la liste des sociétés commercialisant un produit permettant la décomposition par usages de la courbe de charge générale :

- Watteco (enPowerMe) <http://www.watteco.com/>
- Onzo <http://onzo.com/>
- Sentec (Coracle)⁵ <http://www.sentec.co.uk/>

3. À consulter à l'adresse suivante <http://redd.csail.mit.edu/>

4. À consulter à l'adresse suivante <http://nilm.cmubi.org/>

5. Start-up issue de l'université de Cambridge

- Navetas⁶ <http://www.navetas.com/>
- LoadIQ <http://www.loadiq.com/>
- Bidgely <http://bidgely.com/>
- Emme <http://www.getemme.com/>
- Plotwatt <https://plotwatt.com/>
- Enetics (SPEED - Single Point End-use Energy Disaggregation) <http://www.enetics.com/>
- Legrand (Energy Consumption Monitoring System) <http://www.enviro-friendly.com/home-energy-monitoring-system.shtml>
- Intel (WEST - Wireless Energy Sensing Technology) http://download.intel.com/newsroom/kits/research/2011/pdfs/Research@Intel-2011_DemoFactSheet.pdf

Ces produits utilisent essentiellement des technologies reposant sur la signature macroscopique des usages. Les mesures prises sont généralement les puissances active et réactive. A notre connaissance, seul le système de Navetas, utilisant des mesures échantillonnées à quelques kHz, est susceptible d'exploiter les signatures microscopiques.

EDF R&D a été amené à évaluer les performances de certains de ces produits. Même si certains usages, & plus particulièrement le froid alimentaire, sont identifiés avec de bonnes performances, les résultats obtenus sont globalement décevants et parfois très éloignés de la vérité terrain. De plus certains produits sont calibrés sur les spécificités d'un pays : le chauffage électrique, par exemple, est un mode de chauffage privilégié en France par rapport au Royaume-Uni. De fait, certains produits tendent à ignorer complètement cet usage alors qu'il représente une part importante de la consommation.

Finalement, notons que la fiabilité de l'identification des usages est un critère très important pour ce type de technologie. Le produit peut être complètement décrédibilisé aux yeux du client s'il fournit une information contraire à la réalité. En conséquence, il vaut mieux fournir une information imprécise (« usage non classifié ») qu'une information fausse.

1.3.7 Détection d'événements et décomposition de courbes de charges

Nous avons articulé notre état de l'art par rapport aux caractéristiques utilisées par les auteurs pour distinguer les différents usages présents dans une courbe de charge. Au regard des différentes méthodes examinées, il convient de noter qu'il existe deux approches fondamentalement différentes au suivi de courbes de charge.

La première approche, séquentielle, vise à détecter les événements d'intérêt dans la courbe de charge (ON/OFF, changement de régime), puis à caractériser et classifier ces événements pour ensuite en tirer les informations d'intérêt (type de l'usage en cas d'événement ON, consommation de l'usage détecté après son déclenchement,...). C'est cette approche qui a été utilisée dans les travaux pionniers [Hart 1992] et dans les nombreux travaux qui en découlent. C'est également cette approche qui domine les travaux d'analyse microscopique.

La deuxième approche est globale, c'est-à-dire que l'intégralité de la courbe de charge est décomposée en courbes individuelles. On peut alors expliquer à chaque instant quelle est la consommation associée à un usage particulier et quelle est sa part dans la consommation globale. Ce sont les résultats que délivrent typiquement les méthodes reposant sur des modèles de Markov à états cachés dont un exemple est présenté Fig. 1.4, mais également d'autres méthodes et modèles comme ceux proposés dans [Yi-xin 2008, El Guedri 2009].

6. Start-up issue de l'université d'Oxford

Dans le cadre de cette deuxième méthode, les signatures des usages choisies pour l'analyse doivent nécessairement vérifier un critère de superposition. En effet tous ces modèles reposent sur une hypothèse de mélange linéaire des sources associées à chaque usage. C'est-à-dire que la signature observée lorsque plusieurs usages sont présents doit être la somme des signatures individuelles de chaque usage. Il faut remarquer que ce critère de superposition est bien vérifié pour les puissances actives et réactives, ainsi que pour les puissances harmoniques. Ces grandeurs sont les plus couramment surveillées dans l'analyse des courbes de charge. Il existe cependant des contre-exemples. Ainsi l'enveloppe du signal de courant ne satisfait pas au critère de superposition. Notons enfin que cette hypothèse, même si elle est toujours souhaitable, n'est pas autant cruciale dans le cadre d'une analyse microscopique. On peut en effet relâcher un peu cette hypothèse en supposant que, d'une part, eu égard aux temps caractéristiques des transitoires surveillés (souvent inférieurs à la seconde), la probabilité d'une superposition de transitoires reste négligeable et d'autre part les grandeurs surveillées pour chaque événement, ou transitoire, ne dépendent pas des régimes permanents des autres usages.

1.4 Conclusions

Nous avons évoqué dans ce chapitre la problématique et les enjeux de la décomposition de courbe de charge. Nous avons dressé un état de l'art des méthodes existantes dans ce domaine en les distinguant en deux catégories : les approches macroscopiques qui consistent à identifier les grandeurs en régime permanent et les cycles de fonctionnement globaux des usages, et, les approches microscopique qui utilisent les spécificités du régime transitoire pour identifier des classes d'usages. Les applications industrielles existantes utilisent essentiellement des approches macroscopiques et souffrent donc de défauts quant à l'identification précise des usages. C'est pourquoi l'approche microscopique apparaît aujourd'hui comme un complément indispensable à l'analyse macroscopique. C'est justement tout l'enjeu de cette thèse à travers le développement de méthodes d'analyse microscopiques pour l'identification des usages.

Présentation des signaux électro-domestiques de la base de données SISED

Dans le cadre de cette thèse, nous disposons d'une base de données de courbes de charge réelles, fournie par EDF R&D, nommée SISED (Signatures Individuelles de Systèmes Électro-Domestiques). L'objet de ce chapitre est la description et l'analyse des signaux de cette base de données qui serviront à analyser les performances des méthodes développées dans les chapitres suivants.

2.1 Présentation générale de la base de données SISED

La base de données SISED est constituée des courbes de charges individuelles de différents usages. Pour chaque appareil référencé nous disposons d'une ou plusieurs mesures de la courbe de charge effectuées aux bornes de l'appareil en fonctionnement. Les appareils sont de modèle, de marque et d'ancienneté variables. De plus, d'une mesure à l'autre, les conditions de fonctionnement et les éventuels réglages de l'appareil peuvent changer.

La courbe de charge agrégée n'est pas mesurée. Étant réalisées directement aux bornes de l'appareil plutôt qu'au niveau du compteur électrique, les mesures ne tiennent pas compte de l'influence du réseau électrique domestique (pertes en ligne, défauts de conducteur).

Cette base de données n'a donc pas vocation à fournir un outil d'évaluation pour les méthodes de décomposition de la courbe de charge. L'objectif principal est de disposer d'exemples réels de courbes de charge individuelles pour chaque catégorie d'appareil afin de pouvoir étudier leur signature électrique.

La base de données SISED recense 39 clients correspondant aux personnes sollicitées pour la campagne de mesure. Sur l'ensemble de la base de données, 24 usages différents sont identifiés. Pour chaque client, seule une partie des usages a été mesurée. La durée de chaque mesure varie en fonction de la durée typique de fonctionnement des appareils, allant de quelques minutes (bouilloire, aspirateur), à plusieurs heures (réfrigérateur, machine à laver).

Pour chacune de ces mesures, nous disposons, en plus des mesures de courant et tension, du questionnaire rempli par le client comportant notamment, le type d'appareil, sa marque, sa puissance nominale, sa date d'achat approximative ainsi qu'un court descriptif des éventuelles manipulations effectuées durant l'acquisition (variation de puissance d'un aspirateur par exemple).

2.1.1 Conditions d'acquisition

L'acquisition de ces données a été faite par un prototype d'appareil de mesure permettant de mesurer de façon synchrone la tension et le courant appelés au niveau de la prise électrique d'un

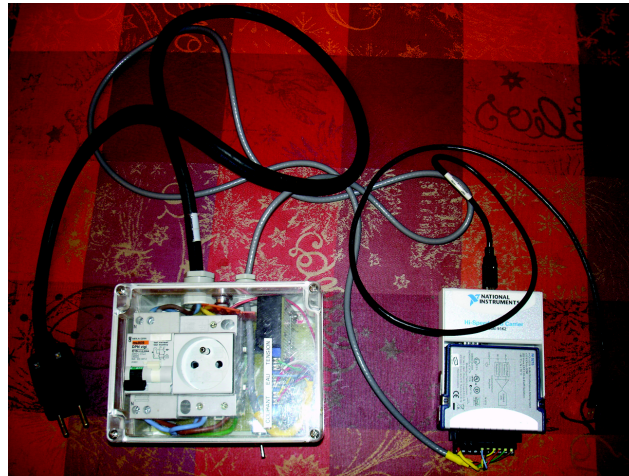


FIGURE 2.1: Photographie du prototype d'appareil de mesure : (gauche) boîtier de mesure et transmission de puissance ; (droite) bus d'acquisition des données

usage. L'appareil de mesure est branché sur le secteur tandis que l'usage électrique à mesurer est branché directement sur l'appareil de mesure. L'appareil de mesure transmet alors les données acquises à un ordinateur qui enregistre la valeur des échantillons dans un fichier texte. Le prototype ayant connu deux versions, la fréquence d'échantillonnage est soit de 5kHz soit de 10kHz pour les mesures plus récentes. Dans les deux cas, la fréquence d'échantillonnage est suffisante pour accéder à la signature microscopique des usages.

2.1.2 Sélection des cas d'études

Dans le cadre de cette thèse, nous ne traitons pas l'intégralité des usages présents dans la base de données SISED. En effet la problématique générale (la détection de n'importe quel usage) est trop complexe pour être appréhendée dans son ensemble. Nous abordons donc le problème de manière progressive, en sélectionnant quelques usages dont la détection, modélisation et classification nous constituent déjà un défi important, et pour lesquelles nous disposons d'un nombre d'exemples suffisant. Bien que nous ne validions notre méthodologie que sur un sous-ensemble des usages, la démarche présentée dans cette thèse est applicable aux autres types d'usages. Par la suite, nous nous concentrerons donc sur l'étude des 5 usages suivants : aspirateur, réfrigérateur, four micro-ondes, lampe halogène et bouilloire.

2.2 Régime transitoire des usages électriques

Dans le cadre de cette thèse, nous nous sommes intéressés à l'étude des régimes transitoires des usages électriques. Plus particulièrement, nous nous intéressons spécifiquement aux signaux transitoires générés par l'enclenchement (la mise sous tension) de ces usages. Nous avons déjà spécifié, lors de l'état de l'art des méthodes de surveillance de courbe de charge, que ces signaux transitoires sont caractéristiques du type d'appareil qui les génèrent (étant liés à la tâche physique sous-jacente réalisée : produire du froid, chauffer un filament ...). Nous comptons donc identifier le type d'appareil mis en fonctionnement à partir du signal transitoire qu'il génère.

Usage	Nombre d'exemplaires
Aspirateur	16
Bouilloire	12
Lampe Halogène	5
Four micro-ondes	35
Réfrigérateur	36
Total	104

TABLE 2.1: Répartition des différentes classe de signaux transitoires dans la base de données SISED

2.2.1 Constitution d'une base de données de signaux transitoires

Chacune des mesures de la base de données SISED est susceptible de contenir plusieurs signaux transitoires correspondant à l'enclenchement d'un appareil. De plus la majorité des échantillons mesurés correspond au régime permanent. Pour pouvoir travailler plus simplement sur les régimes transitoires qui nous intéressent, il est nécessaire d'isoler la ou les parties de chacune des mesures qui correspondent à l'enclenchement d'un appareil. Cette tâche a été effectuée manuellement dans un premier temps afin de pouvoir développer par la suite des méthodes de détection et d'analyse automatiques de ces signaux. Cette tâche, quoique répétitive, ne présente pas de difficulté particulière étant donné que chaque mesure correspond à un unique appareil et que le nombre de fois où l'appareil est mis sous tension est connu via les informations fournies par le client. De plus, la segmentation n'a pas besoin d'être précise, il s'agit simplement d'isoler grossièrement les parties d'intérêt du signal mesuré.

Au terme de cette phase d'extraction manuelle, nous avons isolé une centaine de signaux transitoires, répartis entre les cinq usages considérés. Le tableau 2.1 donne la répartition exacte des signaux entre chaque classe.

Nous allons procéder à l'analyse des signaux transitoires ainsi obtenus. Nous commencerons par identifier les caractéristiques communes à tous les signaux. Ensuite nous analyserons chaque classe d'appareil individuellement.

2.2.2 Analyse générale des signaux transitoires de mise sous tension

La figure 2.3 présente deux exemples de signaux transitoires tirés de la base de données SISED, respectivement un signal d'aspirateur et de réfrigérateur. Les mesures de courant et de tension sont présentées pour chaque appareil et un agrandissement d'une partie du signal permet d'observer la forme d'onde du courant et de la tension en régime permanent.

Dans un premier temps, on peut vérifier expérimentalement l'hypothèse formulée au chapitre 1, c'est-à-dire que la tension appelée ne dépend pas, ou de manière très marginale, de la charge et se présente comme une sinusoïde de fréquence fondamentale 50Hz et d'amplitude efficace 230V.

L'instant auquel l'appareil est mis sous tension est facilement identifiable : il correspond systématiquement à une augmentation instantanée de l'amplitude du courant consommé. Coïncidant avec l'instant de mise sous tension, on peut parfois constater la présence de faibles perturbations du signal de tension par rapport à la sinusoïde nominale (cf figure 2.2). Étant donné que ces perturbations sont occasionnelles et de très faible amplitude (2% de la tension crête sur la figure 2.2), nous les négligerons dans la suite de ces travaux.

Après la mise sous tension, le courant est un signal périodique de fréquence fondamentale 50Hz

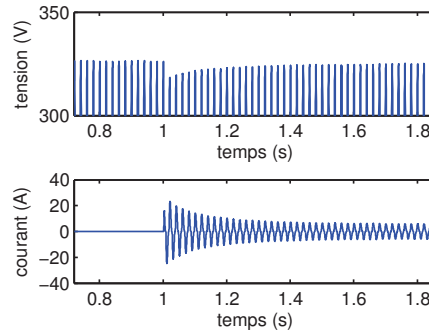


FIGURE 2.2: Exemple de perturbation du signal de tension à l'enclenchement d'un aspirateur : tension (haut) courant (bas)

dont l'amplitude varie pendant quelques secondes. Au terme de cette durée, l'amplitude se stabilise à une valeur constante définissant le régime permanent.

Pour illustrer les différentes phases d'évolution du signal et analyser plus précisément le régime transitoire, nous avons réalisé le spectrogramme des deux exemples de transitoires. Le spectrogramme $\gamma(t, f)$ d'un signal $x(t)$ est défini comme le carré de l'amplitude de la transformée de Fourier court-terme du signal :

$$\gamma(t, f) = \left| \int_{-\infty}^{+\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau \right|^2. \quad (2.1)$$

Pour les figures présentées dans ce chapitre, $w(t)$ est une fenêtre d'analyse de Hamming de la largeur temporelle 0.1s et le spectrogramme est représenté en décibels ($\omega_{dB}(t, f) = 10 \log_{10} \omega(t, f)$).

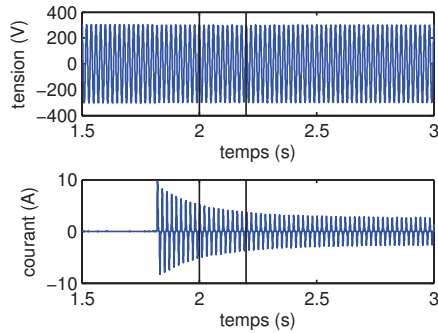
Sur les figures 2.3(e) et 2.3(f) l'instant de mise sous tension de l'appareil est bien visible. En effet la discontinuité du signal de courant qui le caractérise dans le domaine temporel est identifié dans le domaine temps fréquence par l'apparition rapide de nombreuses composantes fréquentielles. On distingue ensuite la présence d'un spectre de raies régulièrement espacées, caractéristique d'un signal périodique. Cependant ces composantes sont d'amplitude et de phase variables, certaines s'évanouissant rapidement avec le temps. Le signal n'est donc pas périodique au sens strict. On peut également constater l'absence (ou la très faible amplitude) des harmoniques de rang pairs tel que prévu par la théorie (cf Annexe A.2).

2.2.3 Puissance active en régime transitoire

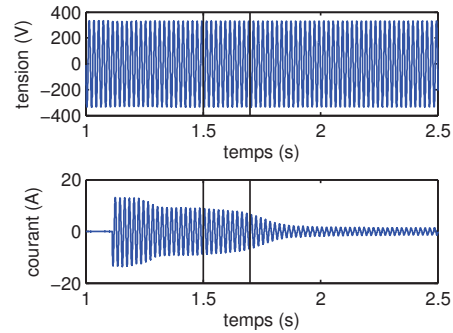
Dans le cas où le système n'est pas en régime permanent, la puissance active n'est pas constante et s'exprime à chaque instant t comme la moyenne suivante :

$$P(t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} v(\tau) i(\tau) d\tau, \quad (2.2)$$

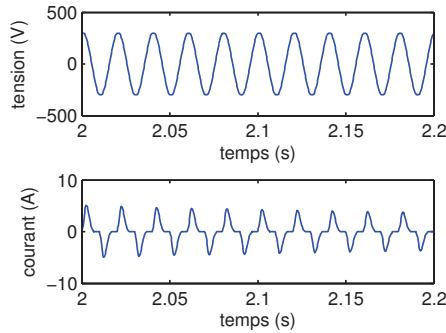
avec $T = \frac{1}{F_f}$ la période fondamentale des signaux de courant et tension. On peut développer cette expression de la puissance en considérant le développement en série de Fourier de la restriction des



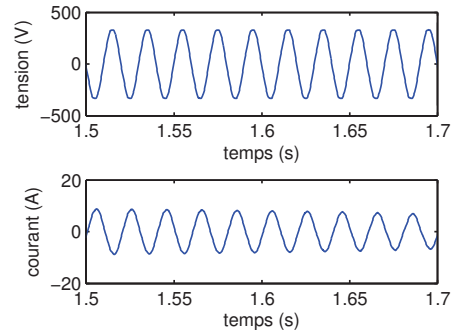
(a) Transitoire d'aspirateur : tension (haut) et courant (bas)



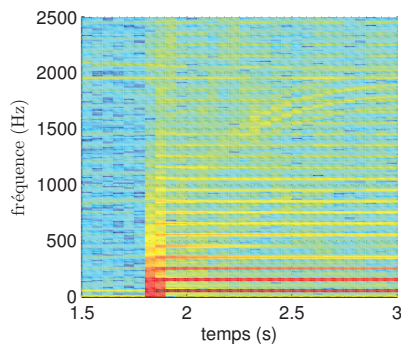
(b) Transitoire de réfrigérateur : tension (haut) et courant (bas)



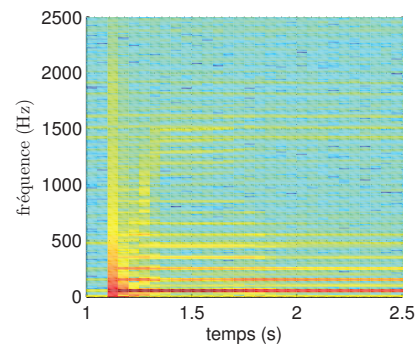
(c) Forme d'onde du signal d'aspirateur, agrandissement de la zone marquée sur la figure 2.3(a) : tension (haut) courant (bas)



(d) Forme d'onde du signal de réfrigérateur, agrandissement de la zone marquée sur la figure 2.3(b) : tension (haut) courant (bas)



(e) Spectrogramme du signal transitoire de courant de l'aspirateur (fenêtre d'analyse de Hamming de largeur temporelle 0.1s (500 échantillons))



(f) Spectrogramme du signal transitoire de courant du réfrigérateur (fenêtre d'analyse de Hamming de largeur temporelle 0.1s (500 échantillons))

FIGURE 2.3: Tension, courant et spectrogramme des signaux exemples

signaux de courant et tension à l'intervalle $[t - \frac{T}{2}, t + \frac{T}{2}[$, c-à-d, pour tout $\tau \in [t - \frac{T}{2}, t + \frac{T}{2}[$:

$$i(\tau) = \sum_{n \in \mathbb{Z}} I_n(t) e^{2\pi j n F_f \tau}, \quad (2.3)$$

$$v(\tau) = \sum_{n \in \mathbb{Z}} V_n(t) e^{2\pi j n F_f \tau}, \quad (2.4)$$

où j est le nombre complexe tel que $j^2 = -1$, et $I_n(t)$ et $V_n(t)$ sont les coefficients complexes du développement en série de Fourier :

$$I_n(t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} i(\tau) e^{-2\pi j n F_f \tau} d\tau,$$

$$V_n(t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} v(\tau) e^{-2\pi j n F_f \tau} d\tau.$$

Notons que l'on peut, de façon équivalente, calculer ces coefficients $I_n(t)$ et $V_n(t)$ en réalisant la transformée de Fourier court-terme munie d'une fenêtre d'analyse rectangulaire de largeur T des signaux de courant et tension. Le coefficient $I_n(t)$ (resp. $V_n(t)$) correspond alors à l'évolution de la composante de fréquence nF_f de la transformée de Fourier court-terme du courant (resp. de la tension). Dans le cas du courant, les lignes caractéristiques que l'on aperçoit sur les spectrogrammes (figures 2.3(e) et 2.3(f)) pour les multiples de la fréquence fondamentale, correspondent donc aux valeurs de $|I_n(t)|^2$.

En injectant les expressions (2.3) et (2.4) dans le calcul de la puissance (2.2) et en exploitant le fait que les fonctions $e^{2\pi j n F_f t}$ forment une base orthonormée de l'espace des fonctions périodiques de période T , on obtient directement que :

$$P(t) = \sum_{n \in \mathbb{Z}} I_n(t) V_n^*(t),$$

où $V_n^*(t)$ est le conjugué de $V_n(t)$. En utilisant l'hypothèse que la tension est simplement sinusoïdale, on déduit que seuls les coefficients $V_{\pm 1}(t)$ correspondants au fondamental ont une valeur non nulle. C'est à dire, quel que soit $n \neq \pm 1$, $V_n(t) = 0$. De plus cette valeur est indépendante du temps : $V_1(t) = V_1$. Alors, la puissance active se résume à

$$P(t) = 2\Re(I_1(t)V_1^*), \quad (2.5)$$

où $\Re(\cdot)$ désigne la partie réelle. On constate donc que seule la composante fondamentale du signal de courant participe à la puissance active. La puissance réactive et les puissances actives et réactives des harmoniques du courant peuvent se calculer de la même manière. On obtient alors l'équivalent de la formule (2.5) pour la puissance réactive :

$$Q(t) = 2\Im(I_1(t)V_1^*),$$

où $\Im(\cdot)$ désigne la partie imaginaire, et quel que soit $n > 1$:

$$P_n(t) = 2\Re(I_n(t)V_n^*),$$

$$Q_n(t) = 2\Im(I_n(t)V_n^*),$$

ce qui est équivalent aux équations (1.6) et (1.7) en régime permanent.

En pratique, les signaux de courant et tension sont échantillonnés à une fréquence F_e donnée ($F_e = 5\text{kHz}$ ou $F_e = 10\text{kHz}$ selon les cas). On note $\{v[k]\}_{k=1,\dots,n}$, $\{i[k]\}_{k=1,\dots,n}$ les échantillons mesurés de la tension et du courant, où n est le nombre de mesures réalisées. Pour calculer la puissance active, on peut alors utiliser au moins deux méthodes.

La première consiste à développer une approximation discrète de l'équation (2.2) :

$$P[k] = \frac{1}{L} \sum_{l=k-\frac{L}{2}}^{k+\frac{L}{2}} v[l]i[l], \quad (2.6)$$

où $L = \lfloor \frac{F_e}{F_f} \rfloor$, $\lfloor \cdot \rfloor$ désigne la partie entière.

La seconde consiste à développer une approximation discrète de l'équation (2.5) qui tient compte explicitement de l'hypothèse d'une tension sinusoïdale :

$$\begin{aligned} I_1[k] &= \frac{1}{L} \sum_{l=k-\frac{L}{2}}^{k+\frac{L}{2}} i[l] e^{-2\pi j \frac{F_f}{F_e} l}, \\ V_1[k] &= \frac{1}{L} \sum_{l=k-\frac{L}{2}}^{k+\frac{L}{2}} v[l] e^{-2\pi j \frac{F_f}{F_e} l}, \\ P[k] &= 2\Re(I_1[k]V_1[k]^*). \end{aligned} \quad (2.7)$$

Notons que nos signaux vérifiant bien l'hypothèse d'une tension sinusoïdale, les deux méthodes proposées ((2.6) ou (2.7)) donnent des résultats très similaires.

La puissance active obtenue, que ce soit avec l'une ou l'autre des méthodes d'approximation, est échantillonnée à la même fréquence F_e que les signaux de courant et tension. En pratique cependant, il n'est pas nécessaire de travailler à une telle fréquence. La puissance active étant issue d'un moyennage sur chaque période du signal, l'information qu'elle contient est redondante puisque les fenêtres d'analyse se recouvrent entre elles. Nous choisissons donc de sous-échantillonner les signaux de puissance active pour travailler à une fréquence d'échantillonnage de 50Hz quelle que soit la fréquence d'origine des signaux F_e . Cette fréquence d'échantillonnage permet d'éliminer toute redondance du signal de la puissance active. Il reste possible de choisir une fréquence supérieure à cette valeur, ce qui implique alors de traiter des signaux de plus grande taille.

La figure 2.4 présente la puissance active d'un signal transitoire de réfrigérateur calculée avec la première méthode (équation (2.6)) ou la deuxième méthode (équation (2.7)). Pour chaque méthode, le sous-échantillonnage de la puissance active à une fréquence de 50Hz est représenté.

Notons bien que, même si cette fréquence d'échantillonnage nous situe à la limite de l'analyse microscopique telle que définie au premier chapitre 1.3.1, les grandeurs utilisées pour calculer la puissance active doivent bien être échantillonnées à des fréquences supérieures. La puissance active nous permet simplement de résumer l'évolution de ces grandeurs.

2.2.4 Analyse individuelle de chacune des classes d'usages

Nous avons identifié la plupart des caractéristiques communes aux signaux de la base de données SISED. Chacune des cinq classes d'appareils étudiées présente également des spécificités. Ces spécificités interviennent essentiellement en deux endroits : au niveau de l'amplitude du signal de courant d'une part, et au niveau de la forme d'onde des signaux de courant i.e, de l'amplitude relative des différentes composantes harmoniques.

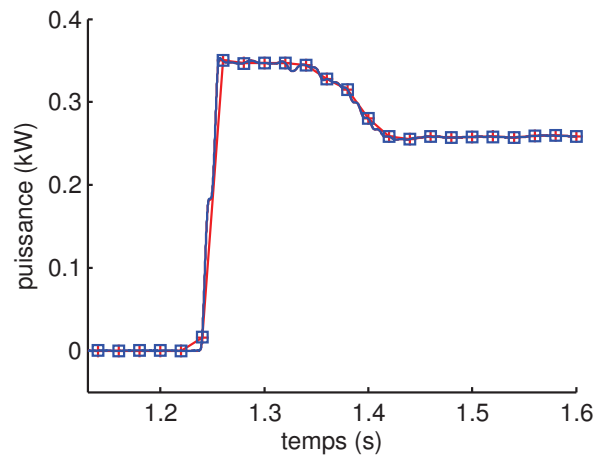


FIGURE 2.4: Portion d'un transitoire de réfrigérateur : ligne bleue $-$, puissance active (2.6) $Fe = 5000\text{Hz}$; carré bleu \square , puissance active (2.6) $Fe = 50\text{Hz}$; ligne rouge $-$, puissance active (2.7) $Fe = 5000\text{Hz}$; étoile rouge $+$, puissance active (2.7) $Fe = 50\text{Hz}$;

Cependant, l'enveloppe du signal de courant ne vérifie pas le critère de superposition (cf 1.3.7). Ainsi, plutôt que d'utiliser cette enveloppe, nous caractériserons les variations d'amplitude du signal de courant en utilisant la puissance active. La puissance active à l'avantage de vérifier le critère de superposition (cf Théorème de Boucherot). De plus, la tension étant supposée sinusoïdale, la puissance active capture essentiellement les variations du signal de tension.

Aspirateur

Quelques exemples de signaux d'aspirateur (puissance active et forme d'onde) sont présentés sur la figure 2.5. La puissance active de ces signaux présente, à partir de l'appel de courant initial, une allure décroissante jusqu'au retour à une valeur constante atteinte en régime permanent. À quelques rares exceptions près, seule la composante fondamentale et la troisième harmonique (150Hz) présentent une amplitude significative. D'un point de vue technologique, la puissance consommée par un aspirateur est entièrement dédiée à l'alimentation du moteur électrique permettant de mettre en rotation le ventilateur destiné à l'aspiration. C'est donc essentiellement la signature électrique de ce moteur que l'on observe.

Bouilloire

Les signaux de bouilloire sont les plus simples à décrire et interpréter parmi les classes de signaux traités. Ils se caractérisent par l'absence totale de variations dans l'amplitude et la phase, ainsi que par l'absence totale d'harmoniques. Plus simplement, le régime permanent est atteint instantanément dès l'appel de courant initial. Ce comportement est tout à fait conforme à ce qu'on peut attendre si l'on examine le fonctionnement physique de ce type d'appareil. La consommation électrique d'une bouilloire est attribuable essentiellement à l'élément chauffant, or celui-ci est équivalent à une simple résistance électrique qui dissipe l'énergie reçue par effet Joule. Le courant traversé par la bouilloire est donc régi par la loi d'Ohm dès lors que l'appareil est sous tension, ce qui produit un courant proportionnel à la tension. La figure 2.6 présente deux signaux transitoires de bouilloire tirés de la base de données.

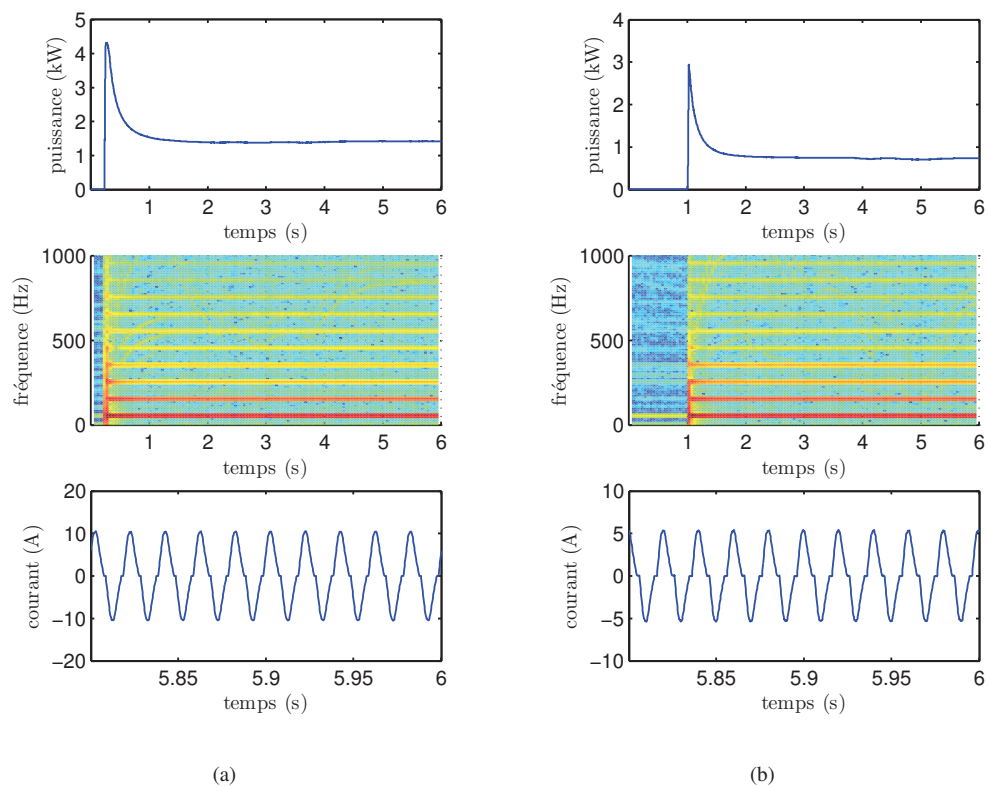


FIGURE 2.5: Exemples de signaux d'aspirateur : puissance active (haut), spectrogramme (fenêtre de Hamming de largeur $0.1s$ soit 500 échantillons) (milieu), forme d'onde du courant (agrandissement d'une partie du signal) (bas)

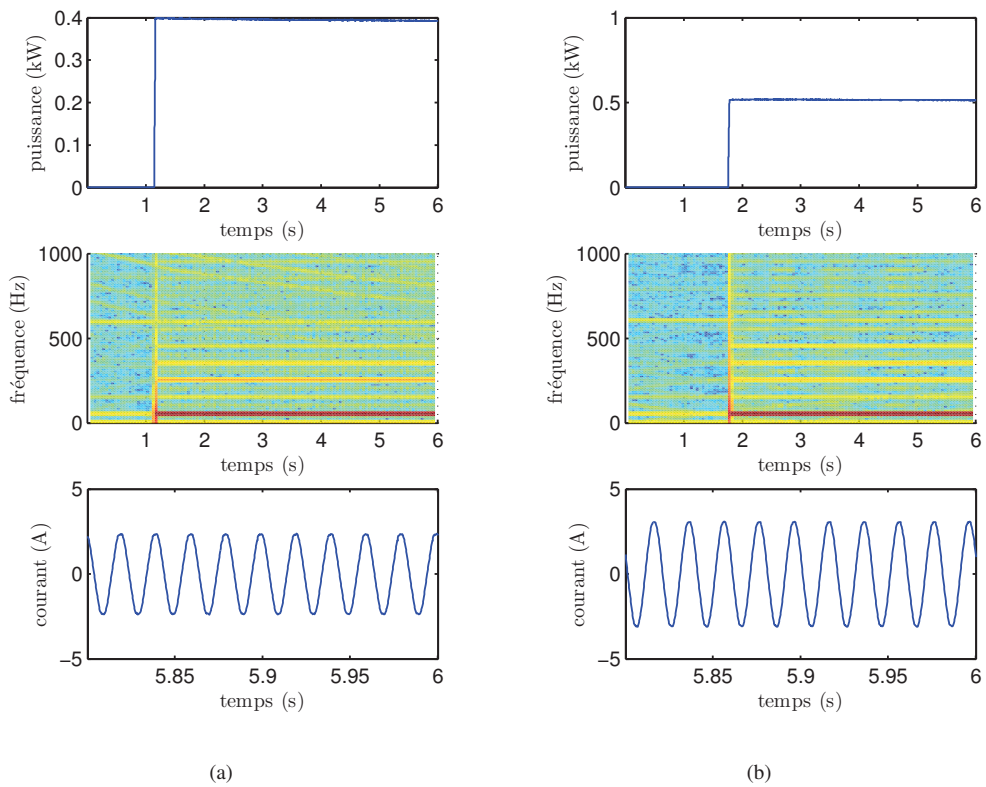


FIGURE 2.6: Exemples de signaux de bouilloire : puissance active (haut), spectrogramme (fenêtre de Hamming de largeur $0.1s$ soit 500 échantillons) (milieu), forme d'onde du courant (agrandissement d'une partie du signal) (bas)

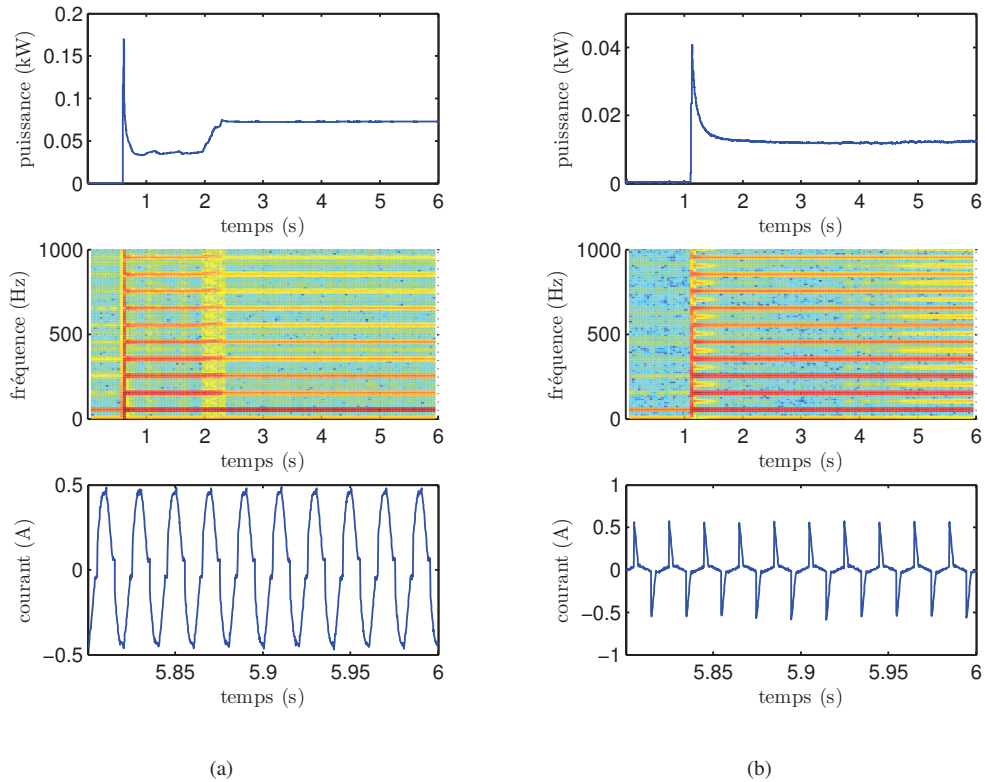


FIGURE 2.7: Exemples de signaux de lampe halogène : puissance active (haut), spectrogramme (fenêtre de Hamming de largeur $0.1s$ soit 500 échantillons) (milieu), forme d'onde du courant (agrandissement d'une partie du signal) (bas)

Lampe Halogène

L'évolution de la puissance active des transitoires de lampe halogène présente une allure similaire aux signaux d'aspirateur : une décroissance jusqu'au retour à une valeur constante lorsque le régime permanent est atteint. Cependant les deux classes de signaux se différencient au niveau de la forme d'onde. Là où les signaux d'aspirateurs ne présentaient que quelques harmoniques d'amplitude significative, les signaux de lampe halogène présentent une amplitude significative sur toutes les composantes observables. La forme d'onde des signaux d'halogène est donc beaucoup plus complexe que celle des signaux d'aspirateurs. La figure 2.7 présente deux exemples de signaux transitoires de lampe halogène. La forme d'onde particulière des signaux de lampe halogène est caractéristique de la présence d'un hacheur. C'est en fait le variateur d'intensité généralement associé à ce type de lampe qui produit cette signature caractéristique. Un hacheur est un interrupteur commandé électroniquement permettant de moduler l'intensité de la lampe en tronquant une partie du cycle de tension. Le courant traversant la lampe est alors distordu de manière non-linéaire ce qui explique la présence de nombreux harmoniques.

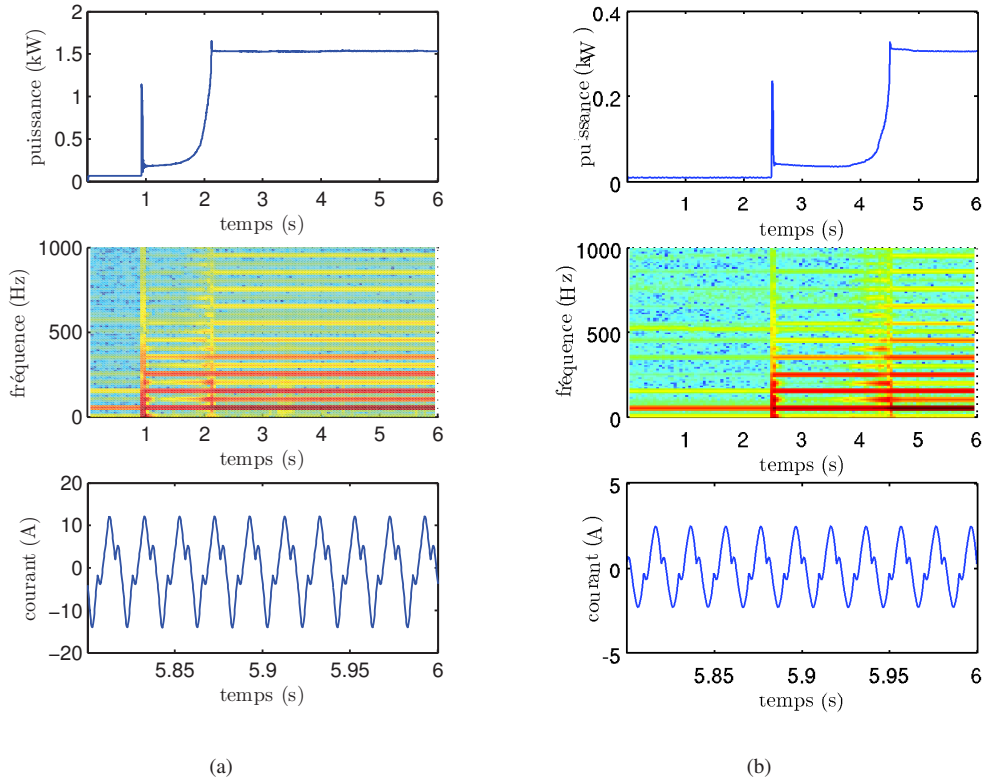


FIGURE 2.8: Exemples de signaux de four micro-ondes : puissance active (haut), spectrogramme (fenêtre de Hamming de largeur $0.1s$ soit 500 échantillons) (milieu), forme d'onde du courant (agrandissement d'une partie du signal) (bas)

Four micro-ondes

Les signaux transitoires de four micro-ondes sont plus complexes que ceux des autres classes. En effet, la puissance active observée se décompose en deux parties distinctes. Tout d'abord après l'appel de courant initial, on observe une décroissance exponentielle similaire à celle des aspirateur ou des lampes halogènes mais de durée caractéristique plus courte, de l'ordre du dixième de seconde. L'amplitude du signal se stabilise ensuite pendant une courte période à une valeur constante. La seconde partie du signal présente une amplitude d'allure exponentiellement croissante jusqu'à saturer à une valeur constante plus élevée que la première. Pour les deux parties du signal, l'amplitude des 5 ou 6 premiers harmoniques est significative. La figure 2.8 présente deux exemples de signaux de micro-ondes. Notons, de plus, que les signaux de micro-ondes présentent une signature macroscopique cyclique. C'est-à-dire que la signature représentée sur la figure 2.8, se répète à intervalles réguliers pendant la durée de fonctionnement de l'appareil. Pour cet usage, il est plus difficile de faire le lien entre la signature électrique et le fonctionnement technologique de l'appareil. En effet, un micro-ondes est un appareil complexe comportant de nombreuses sources de consommation électrique : le magnétron responsable de la cuisson, le moteur actionnant le plateau tournant, la lampe éclairant l'intérieur, le système de ventilation, le transformateur, etc.

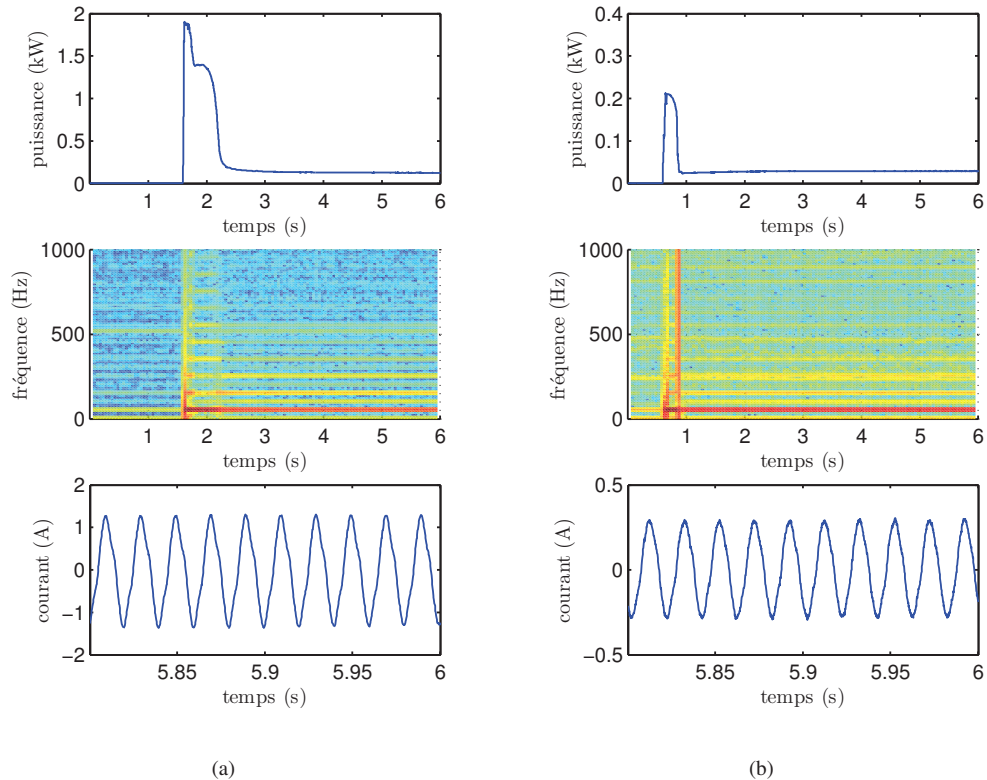


FIGURE 2.9: Exemples de signaux de réfrigérateur : puissance active (haut), spectrogramme (fenêtre de Hamming de largeur 0.1s soit 500 échantillons) (milieu), forme d'onde du courant (agrandissement d'une partie du signal) (bas)

Réfrigérateur

L'amplitude des signaux de réfrigérateurs présente généralement deux décroissances successives avant l'établissement du régime permanent. Contrairement aux variations observées pour les autres classes de signaux, ces décroissances ne présentent aucun caractère exponentiel. Leur allure est plutôt de type sigmoïdale, faisant apparaître un point d'inflexion. Notons qu'en fonction des appareils, il peut y avoir deux décroissances de ce type bien séparées par un palier d'amplitude constante ou n'y avoir qu'une seule décroissance avant le régime permanent. La figure 2.9 présente deux exemples de signaux transitoires de réfrigérateur. Comme pour les micro-ondes, la signature macroscopique des réfrigérateurs est cyclique. On peut d'ailleurs se reporter à l'exemple du chapitre 1, figure 1.2 pour illustrer la périodicité de la consommation électrique. Dans un réfrigérateur, c'est le compresseur, lui-même actionné par un moteur, qui est principalement responsable de la consommation électrique observée. Durant la phase de détente du fluide réfrigérant, la consommation électrique est nulle.

En s'appuyant en particulier sur l'analyse des signaux de micro-ondes, on peut constater qu'il est très difficile pour certains usages complexes de relier précisément la consommation électrique au fonctionnement électrotechnique de l'appareil. De plus, lorsque on en vient à l'analyse du régime

transitoire, la plupart des modèles classiques (pour les moteurs en particulier) ne sont plus valables. C'est pourquoi, dans cette thèse, nous utilisons une approche « boîte noire » pour la modélisation de la signature des usages. Nous ne cherchons pas à appréhender le fonctionnement technologique des appareils, mais nous modélisons le comportement observé expérimentalement pour chaque usage.

2.3 La puissance active pour la caractérisation des courbes de charge

Au regard de l'analyse que nous venons de conduire concernant les signaux transitoires de la base de données SISED, il convient de retenir un descripteur permettant de résumer les signaux de cette base en s'affranchissant des caractéristiques communes à toutes les classes tout en conservant les spécificités de chacune. En effet il semble peu judicieux de travailler à partir des grandeurs directement mesurées, à savoir le courant et la tension. Nous disposons de plusieurs connaissances *a priori* sur ces signaux qu'il s'agirait d'exploiter pour simplifier les signaux à analyser. De plus nous souhaitons conserver la propriété de superposition des grandeurs surveillées (voir paragraphe 1.3.7).

Comme nous l'avons vu, les signaux transitoires de chaque classe se distinguent essentiellement par l'évolution du signal de la puissance active consommée et par la forme d'onde du courant appelé. Par conséquent, l'évolution des puissances harmoniques actives et réactives (ou de façon équivalente les parties réelles et imaginaires des composantes associées aux harmoniques du 50Hz de la transformée de Fourier court-terme du signal de courant) est un descripteur pertinent afin de caractériser un transitoire et plus généralement un type d'usage. Ces profils harmoniques caractérisent en effet simultanément l'évolution de l'amplitude du signal de courant et le contenu harmonique de ce signal (c'est à dire la forme d'onde du courant).

Dans le cadre de cette thèse, nous nous limiterons à l'utilisation de la puissance active du signal qui est une grandeur liée uniquement aux composantes fondamentales des signaux de courant et de tension.

Remarquons qu'il est possible que cette seule caractéristique s'avère insuffisante afin de distinguer certaines classes de signaux. On peut, par exemple, constater visuellement que les profils de la puissance active des signaux d'aspirateur et de lampe halogène sont proches alors que leurs formes d'ondes sont bien distinctes. Cependant, nous allons voir par la suite (cf chapitre 4) que l'évolution de la puissance active est suffisamment discriminante pour séparer efficacement les classes de signaux étudiés. Notons de plus que la méthode développée pour l'analyse d'un signal monodimensionnel (la puissance active) peut être généralisée à l'analyse d'un signal multidimensionnel (les puissances actives et réactives harmoniques du courant) afin d'en améliorer les performances. Même si la méthode développée dans le cadre de ce travail se limite à l'heure actuelle à l'identification d'un nombre restreint d'usages en exploitant la seule puissance active, le principe sous-jacent est général. La méthode peut donc être étendue à un cadre plus opérationnel où l'on cherche à identifier un grand nombre d'usages possibles en utilisant toutes les puissances des harmoniques du courant.

2.4 Description de la puissance active par ajustement de modèle paramétrique

Nous avons identifié l'évolution de la puissance active en régime transitoire comme une grandeur pertinente pour l'identification des classes d'usages de la base de données SISED. Il reste donc à définir la manière dont nous allons utiliser cette caractéristique pour procéder à l'identification automatique des transitoires.

Une des difficultés liée à l'exploitation de cette puissance active est que l'on doit considérer non pas une grandeur scalaire, mais un signal temporel évoluant dans le temps. En conséquence, les méthodes d'apprentissage statistique standards utilisées pour la classification de grandeurs scalaires ou vectorielles (comme les puissances harmoniques en régime permanent par exemple) ne sont pas directement applicables à notre problème.

Le premier problème rencontré en vue d'un apprentissage permettant une classification des usages est lié à la taille des signaux. Le nombre d'échantillons qui composent les signaux de puissance active est trop grand pour être exploité efficacement. De plus, la durée de ces transitoires, et donc le nombre d'échantillons, est fortement variable selon les usages. Il convient donc d'extraire des caractéristiques pertinentes afin de synthétiser l'information utile contenue dans les profils temporels de la puissance active.

La démarche retenue dans ce travail consiste, préalablement à la classification de chaque signal temporel, à ajuster l'évolution de la puissance active à l'aide d'un modèle de régression paramétrique. En effet, l'ajustement des paramètres de ce modèle permet d'extraire une série de valeurs scalaires caractérisant l'évolution temporelle du signal de manière très synthétique. Les valeurs des paramètres ainsi inférées sont utilisées par la suite pour réaliser l'apprentissage statistique des signaux transitoires de chaque usage de la base de données SISED.

2.5 Conclusions

Dans ce chapitre, nous avons procédé à une analyse générale des signaux de la base de données SISED. Nous avons identifié l'évolution de la puissance active comme une caractéristique prometteuse pour la classification des signaux transitoires. Cependant, cette courbe de puissance retenue est une donnée trop volumineuse et présentant trop de redondances pour être exploitée directement comme entrée d'un algorithme de classification. Dans ce cas, l'information portée par l'évolution temporelle de la puissance active peut être résumée l'aide d'un modèle paramétrique d'ajustement de courbes. C'est cette démarche de classification reposant sur un modèle paramétrique de régression de la puissance active que nous choisissons d'utiliser. Le chapitre suivant est donc consacré à l'élaboration de ce modèle.

Modèle de régression à transitions régulières

Le chapitre précédent a souligné l'intérêt de disposer d'un modèle paramétrique afin de représenter l'évolution temporelle de la puissance active des signaux transitoires de la base de données SISED. Les valeurs des paramètres pourront en effet être exploitées afin de caractériser ces transitoires, et de discriminer les différents types d'usage. Dans ce chapitre, nous allons définir un modèle de régression par morceaux à transitions régulières adapté à la modélisation de ces signaux. Notons que le modèle proposé est très général et peut se transposer à d'autres applications. Nous l'exposons donc dans ce chapitre en conservant un formalisme le plus général possible. L'évaluation des performances du modèle est donc réalisée pour des problèmes standards de lissage de courbes. L'application et la validation de ce modèle aux signaux de la base de données sera l'objet du chapitre 4.

3.1 Problématique de l'ajustement de courbes

L'ajustement de courbe est une thématique centrale en traitement du signal. La problématique générale consiste à estimer une fonction qui représente au mieux un signal. Il existe plusieurs méthodes pour construire une telle fonction. D'une part les approches paramétriques consistent à choisir la fonction de régression au sein d'une famille de fonctions paramétrées par un vecteur de dimension finie. On citera, par exemple, les méthodes de régression polynomiale ou les méthodes fondées sur des modèles autorégressifs. D'autre part les approches non-paramétriques consistent à décomposer la fonction de régression sur une famille de fonctions de base. Un exemple de régression non-paramétrique largement utilisée dans la littérature est la régression par fonction splines [Unser 1999].

Dans le cadre de cette thèse, nous nous concentrerons sur les approches paramétriques. Ces dernières ont l'avantage de fournir directement des paramètres caractéristiques du signal régressé. Or nous avons besoin explicitement de ces paramètres pour réaliser la classification de nos signaux. Plus spécifiquement, nous nous sommes tournés vers les modèles de régression par morceaux.

3.1.1 Modèles de régression par morceaux

La régression par morceaux consiste à diviser le signal en segments temporels et à modéliser le signal par une fonction de régression indépendante sur chaque segment. La fonction de régression sur chaque segment est généralement linéaire par rapport aux paramètres. Ce type de modélisation s'avère utile dans de nombreux cas d'application et a fait l'objet de nombreuses publications [Basseville 1993, Gustafsson 2000]. L'utilisation de régression par morceaux est particulièrement pertinente dans le cas où le signal observé présente des changements de comportement abrupts entre des zones où son comportement est quasi-stationnaire.

L'une des principales difficultés quant à l'usage de ce type de modèle est l'évaluation du nombre de segments nécessaires à la représentation d'un signal donné. L'estimation de cette grandeur résulte

nécessairement d'un compromis entre la précision du modèle et le nombre de paramètres. En effet, augmenter le nombre de segments conduit inévitablement à une diminution de l'erreur entre le signal observé et le modèle. En contrepartie, le nombre de paramètres du modèle croît avec le nombre de segments.

Il existe plusieurs outils dans la littérature scientifique permettant la résolution de ce type de compromis. Les plus connus sont les critères d'Akaike [Akaike 1974] et de Schwarz [Schwarz 1978]. La prise en compte de ces critères permet de réaliser un compromis entre la minimisation de l'erreur quadratique moyenne entre le signal et son modèle d'une part et la minimisation du nombre de paramètres du modèle d'autre part. La différence entre les deux critères repose sur l'importance relative donnée à chaque terme. Notons également la méthode proposée dans [Lavielle 2001], où l'auteur propose de régler automatiquement le paramètre qui règle ce compromis, en le fixant à la valeur donnant une solution la plus stable possible.

Une alternative intéressante à ce type de critère est l'utilisation de méthodes d'estimation bayésienne [Robert 2006, Wasserman 2000]. Dans ce cadre, le nombre de segments est estimé exactement comme n'importe quel paramètre du modèle. A ce titre, une densité de probabilité *a priori* lui est attribué. C'est cette densité *a priori*, ainsi que celles attribuées aux autres paramètres du modèle qui vont conjointement régir le compromis entre l'erreur de modélisation et la taille du modèle. Ce type de méthode a déjà été utilisée avec de bons résultats dans [Andrieu 1999], et en particulier pour les modèles de régression par morceaux dans [Punskaya 2002].

Les signaux que nous souhaitons modéliser comportent de vastes portions où le signal est stationnaire. L'utilisation d'un modèle de régression par morceaux où la fonction de régression sur chaque morceau est simplement constante est un bon point de départ. Cependant, nous souhaitons également modéliser la forme des variations d'un régime à l'autre, ces variations étant caractéristiques à chaque classe de signal. Les modèles de régression par morceaux dans leur forme la plus simple ne permettent pas de modéliser ces transitions par autre chose qu'un changement abrupt. Nous allons donc nous tourner vers un modèle qui étend les caractéristiques des modèles de régression par morceaux.

3.1.2 Modèles de régression à transitions régulières

L'idée sous-jacente aux modèles de régression à transitions régulières est, au sein d'un modèle par morceaux, de modéliser les transitions d'un segment à l'autre par une fonction paramétrique plutôt que par une transition abrupte. Ce type de modèle est bien connu en économétrie dans le cas particulier du modèle auto-régressif par morceaux appelé STAR (« *Smooth Transition Autoregressive* ») [Lubrano 2000, van Dijk 2002]. Soient $m_1(t)$ et $m_2(t)$ deux modèles auto-régressifs. Soit $\pi(t)$ une fonction de transition croissante de 0 à 1 qui modélise la transition d'un régime au régime suivant. Alors on peut exprimer $m(t)$ le modèle autorégressif à transition régulière par :

$$m(t) = m_1(t)(1 - \pi(t)) + m_2(t)\pi(t). \quad (3.1)$$

Cette expression représente un modèle à deux segments. On peut noter que si l'on remplace $\pi(t)$ par une fonction échelon, on se ramène à un modèle auto-régressif par morceaux.

La généralisation de cette expression à un modèle doté d'un nombre quelconque de segments n'est pas évidente. D'une part, obtenir une expression similaire à (3.1) pour un nombre plus élevé de segments n'est pas immédiat. D'autre part, comme pour les modèles de régression par morceaux, à partir du moment où le nombre de segments n'est pas fixé, un compromis entre l'erreur de modélisation et le nombre de paramètres du modèle doit être pris en compte. Dans la littérature économétrique, plusieurs possibilités sont proposées. Dans [van Dijk 2002], les auteurs proposent d'étendre

le modèle (3.1) à un nombre quelconque de segments en imbriquant des modèles à deux segments. Plus précisément, on remplace $m_1(t)$ et $m_2(t)$ dans l'équation (3.1) par des modèles auto-régressifs à transitions régulières, obtenant ainsi un modèle à quatre segments. On peut répéter l'imbrication autant de fois que l'on souhaite pour atteindre un nombre de segments donné. Pour décider de la taille du modèle, les auteurs proposent, à chaque étape de l'imbrication, de décider si le nouveau niveau de modélisation est nécessaire par un test d'hypothèses (voir aussi [Lundbergh 2003]). Une procédure similaire de sélection de modèle est proposée dans [Medeiros 2005], à la différence que le modèle est formulé comme la sortie d'un réseau de neurones dont le nombre d'unités cachées détermine le nombre de segments du modèle. Enfin, dans [Maringer 2008], plusieurs procédures d'estimation (en particulier les méthodes de recuit simulé [Kirkpatrick 1983] ou d'évolution différentielle [Storn 1995]) sont proposées pour ce problème.

3.2 Formulation générale du modèle

Le signal observé est noté $x(t)$. On modélise ce signal par la somme de $m(t)$, la fonction de régression d'un modèle à transitions régulières, et d'un signal de bruit $\epsilon(t)$:

$$x(t) = m(t) + \epsilon(t). \quad (3.2)$$

Soit K le nombre de points de rupture du modèle. On note $\tau = [\tau_1, \tau_2, \dots, \tau_K]$ la séquence des K points de rupture entre les segments. Par convention, $[\tau_0, \tau_{K+1}]$ désigne la fenêtre temporelle sur laquelle le signal $x(t)$ est observé, et $L \equiv \tau_{K+1} - \tau_0$ correspond à la longueur du domaine d'observation. On suppose que la configuration τ est ordonnée telle que $\tau_0 < \tau_1 < \tau_2 < \dots < \tau_K < \tau_{K+1}$.

A chaque point de rupture τ_k est maintenant associée une fonction de transition $\pi_{\eta_k}(t)$, paramétrée par le vecteur η_k . Par convention, on associe également $\pi_{\eta_0}(t)$ et $\pi_{\eta_{K+1}}(t)$ à τ_0 et τ_{K+1} respectivement, tels que $\pi_{\eta_0}(t) = \pi_{\eta_{K+1}}(t) = H(t)$ où $H(t)$ est la fonction échelon de Heaviside.

Pour tout $k = 1, \dots, K+1$, on définit $m_k(t)$ le modèle de régression du signal sur le segment $[\tau_{k-1}, \tau_k]$. Nous appellerons par la suite les modèles $m_k(t)$ "sous-modèles", dans le sens où ils constituent les parties d'un modèle plus général $m(t)$. La fonction $m(t)$ s'exprime alors en un point quelconque $t \in [\tau_0, \tau_{K+1}]$ par :

$$m(t) = \sum_{k=1}^{K+1} p_k(t) m_k(t), \quad (3.3)$$

où $p_k(t)$ est le poids associé au modèle $m_k(t)$ à l'instant t :

$$p_k(t) = \pi_{\eta_{k-1}}(t - \tau_{k-1}) - \pi_{\eta_k}(t - \tau_k) \quad \forall k = 1, \dots, K+1. \quad (3.4)$$

Le modèle à transitions régulières $m(t)$ se présente donc comme une somme pondérée des sous-modèles, dont les poids sont pilotés par les fonctions de transitions.

3.2.1 Sous-modèles

Les sous-modèles déterminent le comportement de la fonction de régression sur chaque segment. Choisir un sous-modèle plutôt qu'un autre est essentiellement fonction de l'application visée. Par exemple, si le modèle décrit le comportement d'un système, alors chaque segment correspond à un régime établi du système, et chaque transition correspond à un changement de contraintes ou

de stimuli extérieur. Les sous-modèles doivent dans ce cas décrire le fonctionnement du système en régime permanent.

Dans cette thèse, nous supposons simplement que les sous-modèles s'expriment linéairement en fonction de leurs paramètres. En conséquence, on écrira :

$$m_k(t) = \sum_{q=0}^{Q_k} \beta_k^{(q)} \zeta^{(q)}(t) \quad \forall k = 1, \dots, K + 1, \quad (3.5)$$

où Q_k est l'ordre du sous-modèle associé à $[\tau_{k-1}, \tau_k]$, $\beta_k^{(q)}$ est le coefficient de régression d'ordre q du modèle k et $\zeta^{(q)}(t)$ sont les variables explicatives du modèle.

Cette classe de sous-modèles regroupe une large partie des modèles de régression existants dans la littérature, on citera en particulier les modèles polynomiaux d'une part ($\zeta^{(q)}(t) = t^q$) et les modèles auto-régressifs d'autre part ($\zeta^{(q)}(t) = x(t - q)$). Éventuellement des variables explicatives externes peuvent venir s'ajouter à la base de régression des sous-modèles. Même si le cadre de notre modèle de régression est très général, nous nous concentrons plus particulièrement dans la suite de ce mémoire à l'étude des sous-modèles polynomiaux.

3.2.2 Fonctions de transition

Comme pour les sous-modèles, de nombreux choix sont possibles en fonction de l'application voulue. De nombreuses familles de fonctions paramétriques peuvent définir les fonctions de transition. La seule réelle contrainte est de choisir des fonctions strictement croissantes entre 0 et 1 de sorte que les fonction $p_k(t)$ définissent bien des poids associés aux sous-modèles.

Dans un premier temps, on peut remarquer que si l'on associe des fonctions de transition abruptes à chaque point de rupture $\pi_{\eta_k}(t) = H(t) \quad \forall k = 1, \dots, K + 1$, alors les poids associés à chaque modèle deviennent les fonctions indicatrices de chaque segment :

$$p_k(t) = \begin{cases} 1 & \text{si } t \in [\tau_{k-1}, \tau_k], \\ 0 & \text{sinon.} \end{cases} = \mathbb{I}_{[\tau_{k-1}, \tau_k]}(t) \quad (3.6)$$

Le modèle $m_k(t)$ revient alors à une fonction de régression par morceaux. Dans ce cas le vecteur des paramètres des fonctions de transition η_k est l'ensemble vide.

Le choix le plus couramment fait dans la littérature scientifique pour définir des fonctions de transition est la fonction sigmoïde :

$$\pi_{\eta_k}(t) = \frac{1}{1 + \exp\left(-\frac{t}{\lambda_k}\right)}. \quad (3.7)$$

Auquel cas, le vecteur des paramètres η_k ne contient qu'un paramètre d'échelle $\eta_k = \{\lambda_k\}$. Cependant la méthodologie proposée ici peut s'appliquer à bien d'autres types de fonctions de transition. On peut même imaginer disposer de plusieurs familles de fonction différentes (sigmoïde, exponentielle, échelon) et laisser le modèle ajuster la bonne famille de fonction à chaque transition par l'intermédiaire d'une « étiquette ». Par exemple, si f_θ et g_θ définissent deux familles de fonction paramétrées par θ et le paramètre $e \in \{1, 2\}$ est une « étiquette », alors une famille de fonctions de transition peut être définie par :

$$\pi_{\eta_k}(t) = \begin{cases} f_{\theta_k} & \text{si } e_k = 1, \\ g_{\theta_k} & \text{si } e_k = 2, \end{cases} \quad (3.8)$$

avec $\eta_k = \theta_k \cup e_k$.

Le choix de la fonction de transition est finalement entièrement dépendant de l'application visée. Pour un modèle de lissage, un choix classique de fonction de transition sigmoïde est parfaitement approprié. Le modèle de régression résultant offre un bon compromis entre régularité de la fonction de régression et capacité à s'ajuster aux observations même si le signal observé présente des points singuliers. Dans le cas où la fonction de régression modélise un processus physique sous-jacent, un soin tout particulier doit être accordé au choix de la fonction de transition. Par exemple, pour l'analyse de signaux transitoires électriques, des fonctions de transition exponentielles peuvent modéliser la décroissance du courant dans un condensateur.

3.2.3 Contrainte de convexité sur les poids des sous-modèles

Nous avons exprimé notre modèle comme une somme pondérée des sous-modèles dont les poids évoluent avec le temps. Il est important que le modèle global $m(t)$ soit une combinaison convexe des sous-modèles. C'est-à-dire que les poids $p_k(t)$ vérifient la contrainte de convexité suivante :

$$\begin{cases} \sum_{k=1}^{K+1} p_k(t) = 1, \\ p_k(t) > 0 \quad \forall k = 1, \dots, K+1. \end{cases} \quad \forall t \in [\tau_0, \tau_{K+1}] \quad (3.9)$$

Cette contrainte n'est pas complètement indispensable, cependant elle améliore l'interprétabilité du modèle global $m(t)$. Attribuer un poids négatif à un sous-modèle peut ne pas avoir de sens pour certaines applications.

Étant donnée la manière dont nous avons construit les poids $p_k(t)$, la contrainte de convexité (3.9) se traduit par la contrainte suivante sur les fonctions de transitions :

$$\begin{cases} \sum_{k=1}^{K+1} \pi_{\eta_{k-1}}(t - \tau_{k-1}) - \pi_{\eta_k}(t - \tau_k) = 1, \\ \pi_{\eta_{k-1}}(t - \tau_{k-1}) - \pi_{\eta_k}(t - \tau_k) > 0 \quad \forall k = 1, \dots, K+1. \end{cases} \quad (3.10)$$

La première partie de la contrainte est nécessairement vérifiée puisque on montre facilement que

$$\sum_{k=1}^{K+1} \pi_{\eta_{k-1}}(t - \tau_{k-1}) - \pi_{\eta_k}(t - \tau_k) = \pi_{\eta_0}(t - \tau_0) - \pi_{\eta_{K+1}}(t - \tau_{K+1})$$

par simplifications de proche en proche des termes de la somme sur k . Or, par convention

$$\pi_{\eta_0}(t - \tau_0) - \pi_{\eta_{K+1}}(t - \tau_{K+1}) = H(t - \tau_0) - H(t - \tau_{K+1}) = 1 \quad \forall t \in [\tau_0, \tau_{K+1}]$$

La deuxième partie de la contrainte se traduit par contre par une contrainte sur les fonctions de transition :

$$\pi_{\eta_{k-1}}(t - \tau_{k-1}) > \pi_{\eta_k}(t - \tau_k), \quad \forall t \in [\tau_0, \tau_{K+1}], \quad \forall k = 1, \dots, K+1. \quad (3.11)$$

Cette contrainte n'étant pas automatiquement respectée, nous devons en tenir compte par la suite.

3.2.4 Exemple simulé

Nous allons montrer sur un exemple le comportement de ce modèle pour un jeu arbitraire de paramètres. Pour cet exemple, nous prendrons des fonctions de transition sigmoïdales et des sous-modèles polynomiaux. On fixe *a priori* le nombre de ruptures $K = 2$. On choisit une fenêtre

d'observation unitaire ($\tau_0 = 0$ et $\tau_{K+1} = 1$) et on choisit les points de rupture suivants : $\tau_1 = 0.1$, $\tau_2 = 0.5$. Les paramètres d'échelle de chaque fonction de transition valent respectivement $\lambda_1 = 0.02$ et $\lambda_2 = 0.05$. La figure 3.1(a) présente les fonctions de transition $\pi_k(t)$ ainsi définies. La figure 3.1(b) présente les poids $p_k(t)$ calculé à partir de ces fonction de transitions selon l'équation (3.4).

On définit les sous-modèles polynomiaux suivants ;

$$\begin{aligned} m_1(t) &= 1, \\ m_2(t) &= 2 + 5t - 8t^2, \\ m_3(t) &= -2 + 2t + t^2, \end{aligned}$$

Ceci correspond aux valeurs suivantes des paramètres : $Q_1 = 0$, $\beta_1^{(0)} = 1$, $Q_2 = 2$, $\beta_2^{(0)} = 2$, $\beta_2^{(1)} = 5$, $\beta_2^{(3)} = -8$ et enfin $Q_3 = 2$, $\beta_3^{(0)} = -2$, $\beta_3^{(1)} = 2$, $\beta_3^{(3)} = 1$. Ces sous-modèles sont représenté sur la figure 3.1(c). Sur cette même figure, est représenté le modèle à transitions régulières $m(t)$ résultant des fonctions de transitions et des sous-modèles précédemment définis.

3.3 Propriétés du modèle

3.3.1 Régularité de la fonction de régression

Il est intéressant, en particulier pour les applications de ce modèle en lissage de courbe, d'étudier la régularité de la fonction de régression. Cette régularité dépend entièrement de la régularité des sous-modèles et des fonctions de transitions.

Proposition 3.3.1. *Pour tout $n \in \mathbb{N}$, si $m_k \in \mathcal{C}^n(\mathbb{R})$ pour tout segment $k = 1, \dots, K + 1$ et $\pi_{\eta_l} \in \mathcal{C}^n(\mathbb{R})$ pour toute transition $l = 1, \dots, K$ alors*

$$m \in \mathcal{C}^n(\mathbb{R}),$$

où $\mathcal{C}^n(\mathbb{R})$ est l'espace fonctionnel des fonctions de $\mathbb{R} \rightarrow \mathbb{R}$, continues, dérivables n fois et dont la k -ième dérivée est continue pour tout $1 \leq k \leq n$.

Démonstration. Cette propriété du modèle m découle directement du fait que d'une part $\mathcal{C}^k(\mathbb{R})$ est un espace vectoriel et d'autre part m est une combinaison linéaire des m_k et des π_{η_k} via les équations (3.3) et (3.4). ■

Un corollaire immédiat est que lorsque les sous modèles sont polynomiaux ($m_k \in \mathcal{C}^\infty(\mathbb{R})$) et les fonction de transitions sont sigmoïdales ($\pi_{\eta_k} \in \mathcal{C}^\infty(\mathbb{R})$) alors la fonction de régression résultante m est $\mathcal{C}^\infty(\mathbb{R})$. Notons de plus, que même si la fonction de régression est $\mathcal{C}^\infty(\mathbb{R})$, sa dérivée peut atteindre une valeur arbitrairement grande au niveau des transitions si le paramètre d'échelle λ de la fonction sigmoïde (3.7) tend vers l'infini.

Cette capacité à produire des modèles à la fois très réguliers ($\mathcal{C}^\infty(\mathbb{R})$ par exemple) et de pouvoir s'ajuster à des données variant localement de façon très abrupte est une caractéristique très intéressante pour les applications de lissage de courbes. Au regard de cette propriété, les modèles à transitions régulières se comparent avantageusement à d'autres méthodes de régression classiques. Les fonctions splines par exemple produisent des fonctions de régression régulières mais nécessitent d'accumuler plusieurs nœuds en un même point pour produire des discontinuités. A l'inverse, les fonctions de régression par morceaux peuvent s'ajuster aux discontinuités présentes dans les données. Elles sont cependant seulement régulières par morceaux ce qui n'est pas optimal dans un contexte de lissage.

Les modèles à transitions régulières offrent donc une alternative à ces méthodes du point de vue du compromis régularité/adaptabilité.

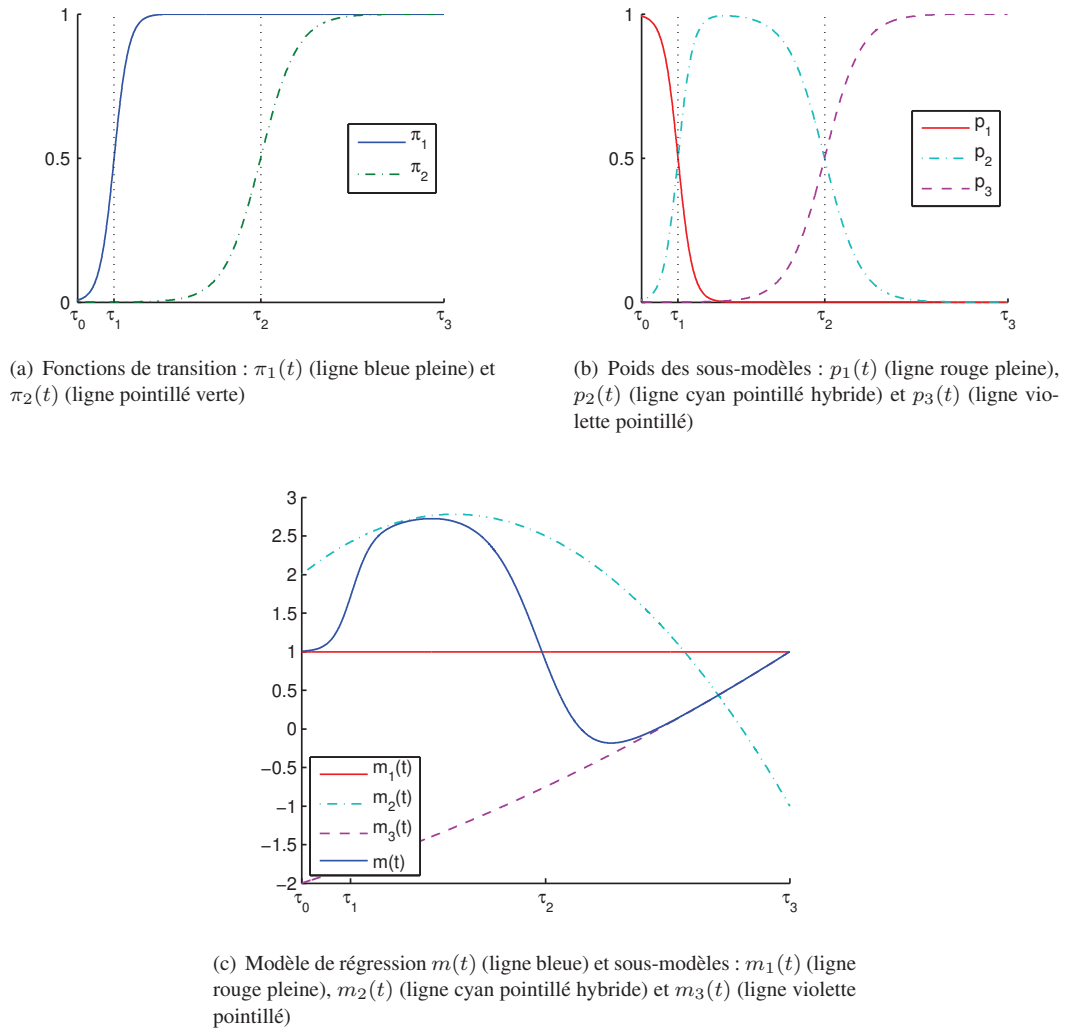


FIGURE 3.1: Exemple simulé

3.3.2 Formulation matricielle du modèle

Jusqu'ici nous avons raisonné sur des grandeurs continues. Dans la pratique, seuls n échantillons du signal observé $x(t)$ sont disponibles. On notera le vecteur de ces échantillons $\mathbf{x} = [x(t_1), \dots, x(t_n)]^T$, où $\mathbf{t} = [t_1, \dots, t_n]^T$ est le vecteur des instants d'échantillonnage, avec par convention $t_1 \equiv \tau_0$ et $t_n \equiv \tau_{K+1}$. Soit $\boldsymbol{\epsilon} = [\epsilon(t_1), \dots, \epsilon(t_n)]^T$ le vecteur des échantillons du bruit résiduel.

On définit par $\boldsymbol{\beta}_k = [\beta_k^{(0)}, \dots, \beta_k^{(Q_k)}]^T$ le vecteur des paramètres de régression du sous-modèle k . On désigne par $\boldsymbol{\beta} = [\beta_1, \dots, \beta_{K+1}]$ la concaténation des vecteurs $\boldsymbol{\beta}_k$ associés à chacun des sous-modèles. On adoptera également les notations suivantes :

- le vecteur $\boldsymbol{\tau} = [\tau_1, \dots, \tau_K]^T$ correspond au vecteur de localisation des points de ruptures entre les segments,
 - le vecteur $\boldsymbol{\eta} = [\eta_1, \dots, \eta_K]^T$ est le vecteur des paramètres des fonctions de transition,
 - le vecteur $\mathbf{Q} = [Q_1, \dots, Q_{K+1}]^T$ est le vecteur des ordres de chacun des sous-modèles.
- Le vecteur du signal observé peut alors s'écrire :

$$\mathbf{x} = Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}) \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.12)$$

où $Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q})$ est une matrice construite à partir des variables explicatives $\zeta^{(q)}(t)$ et des poids des sous-modèles $p_k(t)$ aux instants d'échantillonnage, qui dépend donc des paramètres $\boldsymbol{\tau}, \boldsymbol{\eta}$ et \mathbf{Q} :

$$Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}) = \begin{bmatrix} p_1(t_1)\zeta^{(0)}(t_1) & \dots & p_1(t_1)\zeta^{(Q_1)}(t_1) & \dots & p_{K+1}(t_1)\zeta^{(0)}(t_1) & \dots & p_{K+1}(t_1)\zeta^{(Q_k)}(t_1) \\ \vdots & & \vdots & & \vdots & & \vdots \\ p_1(t_n)\zeta^{(0)}(t_n) & \dots & p_1(t_n)\zeta^{(Q_1)}(t_n) & \dots & p_{K+1}(t_n)\zeta^{(0)}(t_n) & \dots & p_{K+1}(t_n)\zeta^{(Q_k)}(t_n) \end{bmatrix}. \quad (3.13)$$

L'équation (3.12) nous permet de souligner la dépendance linéaire existante entre les paramètres de régression des sous-modèles $\beta_k^{(q)}$ et le modèle à transitions régulières $m(t)$. En effet le modèle $m(t)$ est une combinaison linéaire des sous-modèles (3.3) et les sous-modèles sont, par hypothèse, linéaires par rapport à leur paramètres $\beta_k^{(q)}$, voir (3.5).

3.3.3 Fonction de vraisemblance du modèle

Muni de l'expression matricielle (3.12), il est facile d'exprimer la vraisemblance du modèle pour un jeu de paramètres donné. On supposera que le bruit résiduel $\boldsymbol{\epsilon}$ est un bruit blanc gaussien centré de variance σ^2 . Alors la vraisemblance du modèle est :

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \|\mathbf{x} - Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q})\boldsymbol{\beta}\|^2}, \quad (3.14)$$

où $\boldsymbol{\theta} = \{\boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, K, \sigma^2, \mathbf{Q}\}$ est le vecteur de l'ensemble des paramètres du modèle.

3.4 Modélisation bayésienne hiérarchique

Dans la partie précédente, nous avons défini un modèle de régression à transitions régulières. Ce modèle est caractérisé par un ensemble de paramètres $\boldsymbol{\theta}$. Le problème posé est donc l'estimation de ces paramètres à partir d'un signal observé \mathbf{x} donné. Comme nous l'avons évoqué dans le premier paragraphe 3.1 de ce chapitre, il existe une difficulté principale à l'estimation de ce type de modèle par morceaux : le nombre de paramètres du modèle dépend du nombre de segments. Ainsi, à moins

que le nombre de segments soit fixé *a priori*, nous sommes confrontés à un problème d'estimation et de sélection de modèle conjoints. En outre, si l'ordre des sous-modèles n'est pas connu, le problème de sélection de modèle est double : estimer le nombre de ruptures du modèle K et estimer l'ordre de chacun des sous-modèles.

Dans le cadre de cette thèse, nous proposons une méthode d'estimation bayésienne pour l'estimation du modèle. Suivant [Andrieu 1999, Punsakaya 2002], une méthode bayésienne permet conjointement d'estimer les paramètres du modèle et de sélectionner la taille adéquate du modèle. L'avantage de cette méthode est qu'elle permet d'éviter la recherche exhaustive de la valeur des paramètres et du nombre de ruptures dans un espace de très grande dimension.

Dans ce cadre, on associe à chacun des paramètres du modèle, y compris le paramètre déterminant le nombre de ruptures K et l'ordre des sous-modèles \mathbf{Q} , une densité de probabilité *a priori* traduisant le degré de croyance que nous attribuons à chaque valeur du paramètre. L'ensemble de ses densités *a priori* va conjointement réaliser un compromis entre la minimisation de la taille du modèle (qui dépend du nombre de ruptures K et des ordres des sous-modèles \mathbf{Q}) et la maximisation de la vraisemblance du modèle.

Dans cette partie, nous décrivons le cadre bayésien utilisé pour estimer les paramètres du modèle à transitions régulières.

L'estimation bayésienne repose sur l'expression de la densité de probabilité *a posteriori* des paramètres du modèle θ . La densité *a posteriori* contient toute l'information nécessaire pour procéder à l'estimation de θ . On exprime cette densité *a posteriori* par l'intermédiaire de la loi de Bayes :

$$p(\theta|\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|\theta)p(\theta)}{\int \mathcal{L}(\mathbf{x}|\theta)p(\theta)d\theta}, \quad (3.15)$$

où $\mathcal{L}(\mathbf{x}|\theta)$ est la vraisemblance du modèle, définie par (3.14), $p(\theta)$ est la densité *a priori* de θ et $p(\mathbf{x}) \equiv \int \mathcal{L}(\mathbf{x}|\theta)p(\theta)d\theta$ est un coefficient de normalisation, parfois appelée vraisemblance marginale. Pour exprimer la densité *a posteriori* des paramètres, il faut donc spécifier la forme de leur densité *a priori*.

3.4.1 Densité de probabilité *a priori* des paramètres

Nous supposons la plupart des paramètres indépendants *a priori*. Plus précisément, la densité *a priori* des paramètres s'exprime comme :

$$p(\theta) = p(K, \boldsymbol{\tau})p(\boldsymbol{\beta}, \sigma^2) \prod_k^K p(\eta_k) \prod_k^{K+1} p(Q_k). \quad (3.16)$$

Ceci signifie que les positions $\boldsymbol{\tau}$ et le nombre de ruptures K ne sont pas *a priori* indépendants. De même pour les coefficients de régression des sous-modèles $\boldsymbol{\beta}$ et la variance du bruit σ^2 .

Les choix de chacun des termes de la densité *a priori* sont détaillés ci-dessous. Notons que, autant que possible, les densités *a priori* sont choisies parmi les familles de lois conjuguées par rapport à la vraisemblance.

Densité *a priori* du nombre de ruptures et de leur positions

Nous choisissons de modéliser la position des ruptures par un processus ponctuel de Poisson [Daley 2002] homogène. Il faut noter que $\boldsymbol{\tau}$ a été définie comme une configuration ordonnée. Ici il s'agit bien sûr de la configuration non ordonnée (c-à-d indépendante de l'ordre entre les positions τ_i)

qui est modélisée par un processus ponctuel. Par abus de notation, nous noterons également τ cette configuration dans l'expression des densités. Par conséquent, le nombre de ruptures K suit une loi de Poisson, et la densité *a priori* jointe peut s'exprimer pour tout $K \in \mathbb{N}$ et $\tau_1, \dots, \tau_K \in [\tau_0, \tau_{K+1}]$ comme :

$$p(\tau, K | \mu_K) = \left(\frac{\mu_K}{L}\right)^K \frac{e^{-\mu_K}}{K!}, \quad (3.17)$$

où le paramètre μ_K correspond au nombre moyen de ruptures, alors que $L \equiv \tau_{K+1} - \tau_0$ désigne la longueur du signal. En marginalisant par rapport aux positions τ , on remarque en effet que :

$$\begin{aligned} p(K | \mu_K) &= \int_{\tau_0}^{\tau_{K+1}} \dots \int_{\tau_0}^{\tau_{K+1}} \left(\frac{\mu_K}{L}\right)^K \frac{e^{-\mu_K}}{K!} d\tau_1 \dots d\tau_K = \left(\frac{\mu_K}{L}\right)^K \frac{e^{-\mu_K}}{K!} L^K, \\ &= \frac{\mu_K^K e^{-\mu_K}}{K!}, \end{aligned}$$

i.e. que K est distribué *a priori* selon une loi de Poisson de moyenne μ_K . De plus, on retrouve directement l'expression des densités *a priori* des positions τ conditionnellement au nombre de ruptures :

$$p(\tau | K) = \frac{p(\tau, K | \mu_K)}{p(K | \mu_K)} = \frac{1}{L^K}, \quad (3.18)$$

pour tout $\tau_1, \dots, \tau_K \in [\tau_0, \tau_{K+1}]$. Ces densités conditionnelles témoignent des propriétés des processus de Poisson homogènes : conditionnellement à leur nombre, les positions des ruptures sont indépendantes et distribuées uniformément.

Une alternative possible à ce choix de densité *a priori* est de choisir une loi binomiale pour modéliser le nombre de ruptures K , comme dans [Punskaya 2002, Dobigeon 2007a], pour les modèles de régression par morceaux :

$$p(K) = p^K (1-p)^{n-K}, \quad K \in \{0, \dots, n\}. \quad (3.19)$$

En effet, dans le cas de la régression par morceaux, le modèle de régression aux instants observés t n'est pas modifié lorsque la position d'une rupture évolue entre deux instants voisins de t . Par conséquent, on modélise classiquement la position des ruptures en introduisant une séquence de variables de Bernoulli, associées aux instants observés t , qui prennent la valeur 1 avec une probabilité p lorsque une rupture est présente à l'instant d'échantillonnage. Le nombre de ruptures suit alors la loi binomiale 3.19. Dans le cas du modèle à transitions régulières, modifier la position d'une fonction de transition, même seulement entre deux instants voisins de t , modifie la valeur de la fonction de régression aux instants t suffisamment proches de la transition. Ce phénomène est particulièrement présent lorsque la fonction de transition est relativement abrupte. En effet, la dérivée de la fonction est alors particulièrement élevée et toute modification, même minime, de la position de la transition entraîne un changement important des valeurs de la fonction de transition aux instants d'échantillonnages proches. Les paramètres τ_k doivent donc bien être définis de manière continue sur $[\tau_0, \tau_{K+1}]$.

C'est ce fait qui nous a conduit à choisir la loi de Poisson plutôt qu'une loi binomiale pour modéliser le nombre de ruptures. La loi binomiale découle naturellement de l'hypothèse selon laquelle la position des ruptures est déterminée par un processus de Bernoulli, donc discret. On peut, en pratique, supposer que la loi du nombre de ruptures est binomiale tout en laissant la position des ruptures évoluer de façon continue. Cependant, il est plus naturel de modéliser la position des ruptures comme un processus de Poisson, et donc la loi du nombre de ruptures par une loi de Poisson, lorsque la position des ruptures évolue de façon continue.

Densité *a priori* des ordres des sous-modèles

La loi *a priori* des ordres des sous-modèles \mathbf{Q} est, comme pour le nombre de ruptures K , choisie parmi la famille des lois de Poisson.

$$p(Q_k|\mu_Q) = \frac{\mu_Q^{Q_k} e^{-\mu_Q}}{Q_k!} \quad \forall k = 1, \dots, K + 1, \quad (3.20)$$

où μ_Q est la valeur moyenne *a priori* de l'ordre des sous-modèles. En fixant μ_Q à une valeur faible, ce type de loi favorise alors les valeurs faibles des Q_k et permet de générer des représentations parcimonieuses du signal observé.

Densité *a priori* des paramètres des sous-modèles

Pour modéliser les paramètres des sous-modèles β et la variance du bruit σ^2 , nous choisissons des familles de loi conjuguées par rapport à la vraisemblance du modèle. Ainsi, une loi *a priori* de type g-prior [Zellner 1986] est choisie pour β , soit :

$$\beta|\delta^2, \sigma^2, \tau, \eta, \mathbf{Q} \sim \mathcal{N}\left(\mathbf{0}, \delta^2 \sigma^2 \left[Z(\tau, \eta, \mathbf{Q})^T Z(\tau, \eta, \mathbf{Q}) \right]^{-1}\right), \quad (3.21)$$

où $\delta^2 \in \mathbb{R}_+$ est un paramètre dont nous verrons par la suite qu'il est homogène à un rapport signal-à-bruit.

Notons qu'un choix alternatif plus classique de distribution *a priori* consisterait à celui de la régression *ridge* où

$$\beta|\delta^2, \sigma^2 \sim \mathcal{N}\left(\mathbf{0}, \delta^2 \sigma^2 I_{\sum_{k=1}^{K+1} Q_k}\right),$$

où $I_{\sum_{k=1}^{K+1} Q_k}$ est la matrice identité de rang $\sum_{k=1}^{K+1} Q_k$. Ce choix qui correspond à effectuer une régularisation au sens de Tikhonov [Tikhonov 1977] permettrait d'obtenir une estimation plus robuste lorsque la matrice $Z(\tau, \eta, \mathbf{Q})$ est mal conditionnée. Cependant, par construction de la matrice $Z(\tau, \eta, \mathbf{Q})$, ce cas de figure n'est pas critique dans notre modèle où la contrainte de convexité (3.10) et la forme des fonctions de transition utilisées imposent que les colonnes de $Z(\tau, \eta, \mathbf{Q})$ ne peuvent pas être liées. A l'inverse, le choix spécifique du g-prior (3.21) utilisé par la suite présente plusieurs avantages pour notre modèle, exposés ci-après.

Proposition 3.4.1. *La densité de probabilité (3.21) engendre un modèle (c-à-d la fonction de régression $m(t)$) dont la loi a posteriori est invariante par rapport à toute transformation linéaire de la base de régression choisie pour les sous-modèles.*

Démonstration. Nous avons besoin pour cette preuve de l'expression de la densité marginale *a posteriori* que l'on obtiendra plus tard dans ce manuscrit (voir les équations (3.33) et (3.34)). On admettra pour le moment que :

$$\beta|\theta/\{\beta\}, \mathbf{x} \sim \mathcal{N}(\mu_\beta, \Sigma_\beta), \quad (3.22)$$

avec $\mu_\beta = \frac{\delta^2}{1+\delta^2} [Z^T Z]^{-1} Z^T \mathbf{x}$, et $\Sigma_\beta = \frac{\delta^2}{1+\delta^2} [Z^T Z]^{-1}$.

Soit A une matrice carrée inversible définissant un changement de base. On note

$$Z'(\tau, \eta, \mathbf{Q}) = Z(\tau, \eta, \mathbf{Q})A, \quad (3.23)$$

la matrice de régression associée à cette nouvelle base. Alors la moyenne de β' évaluée dans la nouvelle base est :

$$\begin{aligned}\mu_{\beta'} &= \frac{\delta^2}{1 + \delta^2} [Z'^T Z']^{-1} Z'^T \mathbf{x}, \\ &= \frac{\delta^2}{1 + \delta^2} [A^T Z^T Z A]^{-1} A^T Z^T \mathbf{x}, \\ &= \frac{\delta^2}{1 + \delta^2} A^{-1} [Z^T Z]^{-1} Z^T \mathbf{x}, \\ &= A^{-1} \mu_{\beta},\end{aligned}\tag{3.24}$$

et la matrice de covariance de β' est :

$$\begin{aligned}\Sigma_{\beta'} &= \frac{\delta^2}{1 + \delta^2} (Z'^T Z')^{-1} = \frac{\delta^2}{1 + \delta^2} A^{-1} (Z^T Z)^{-1} A^{-T}, \\ &= A^{-1} \Sigma_{\beta} A^{-T}.\end{aligned}\tag{3.25}$$

Ces deux équations (3.24) et (3.25) montrent que la distribution *a posteriori* de $A\beta'$ et β sont des lois normales dont les moyennes et matrices de covariance sont identiques. Ces distributions sont donc identiques, et il en est de même pour la distribution de $m' = Z'\beta'$ et de $m = Z\beta$, car $Z' = ZA$. L'estimée de la fonction de régression est donc invariante par rapport à un changement linéaire de la base de régression des sous-modèles. ■

En particulier, si l'on choisit par exemple des sous-modèles polynomiaux, l'inférence réalisée sera indépendante de la base utilisée pour la régression des polynômes (polynômes d'Hermite, de Legendre, choix arbitraire de l'origine des temps, etc).

Proposition 3.4.2. *Le paramètre δ^2 associé au g-prior peut être interprété comme le rapport signal à bruit a priori entre le modèle $m(t)$ et le bruit résiduel $\epsilon(t)$, c-à-d que δ^2 peut s'exprimer a priori comme :*

$$\delta^2 = \frac{E[\mathbf{m}^T \mathbf{m}]}{E[\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}]} = \frac{E[\mathbf{m}^T \mathbf{m}]}{n\sigma^2},\tag{3.26}$$

où $\mathbf{m} = Z(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q})\boldsymbol{\beta}$ est le vecteur de la fonction de régression.

Démonstration. Cette propriété découle directement de l'expression du g-prior (3.21) :

$$\begin{aligned}E[\mathbf{m}^T \mathbf{m}] &= E[\text{tr}(\mathbf{m}^T \mathbf{m})] = E\left[\text{tr}\left(\boldsymbol{\beta}^T Z^T Z \boldsymbol{\beta}\right)\right] \\ &= \text{tr}\left(Z^T Z E\left[\boldsymbol{\beta} \boldsymbol{\beta}^T\right]\right) = \text{tr}\left(Z^T Z \sigma^2 \delta^2 (Z^T Z)^{-1}\right) \\ &= \text{tr}\left(\sigma^2 \delta^2 I_n\right) = n\sigma^2 \delta^2.\end{aligned}$$

■

Densité a priori de la variance du bruit

Pour la variance du bruit résiduel, nous choisissons une loi vague non-informative de Jeffreys [Jeffreys 1946] :

$$p(\sigma^2) \propto \frac{1}{\sigma^2} \mathbb{I}_{\mathbb{R}_+}(\sigma^2).\tag{3.27}$$

Ce choix de loi est standard pour la variance d'une loi normale, et assure une invariance de la distribution *a posteriori* par rapport à toute reparamétrisation du type $s = g(\sigma^2)$. Bien que la loi soit impropre, la densité *a posteriori* qui en résulte sera dans notre modèle bien définie.

Densité *a priori* des paramètres des fonctions de transition

La définition de la loi *a priori* des paramètres des fonctions de transition $f(\eta_k)$ dépend bien entendu de la forme choisie pour ces fonctions. Cependant quelque soit la famille des fonctions de transition, il n'est pas possible de choisir une loi conjuguée pour ses paramètres par rapport à la vraisemblance du modèle. En effet le modèle dépend des paramètres des fonctions de transition de manière non-linéaire et non-standard via la matrice $Z(\tau, \eta, \mathbf{Q})$ dans (3.14).

Si les fonctions de transition sont choisies sigmoïdales, il faut définir une loi *a priori* sur les paramètres d'échelle λ_k . Nous utilisons une loi *a priori* inverse-gamma telle que :

$$\lambda_k | (\nu_k, \rho_k) \sim \mathcal{IG}(\nu_k, \rho_k) \quad \forall k = 1, \dots, K, \quad (3.28)$$

où $\nu_k \in \mathbb{R}_+$ et $\rho_k \in \mathbb{R}_+$ sont respectivement le paramètre de forme et le paramètre d'échelle dans la paramétrisation standard de la loi inverse-gamma.

3.4.2 Densité de probabilité *a priori* des hyperparamètres

Nous avons défini des lois *a priori* pour l'ensemble des paramètres du modèle. La forme de ces lois *a priori* dépendent pour certaines d'autres paramètres. Nous appellerons les paramètres de ces lois *a priori* « hyperparamètres » pour les différencier des paramètres du modèle θ . Le réglage de ces hyperparamètres est important puisqu'ils influent sur les valeurs estimés finales des paramètres. Il existe plusieurs stratégies permettant de fixer ces hyperparamètres. Dans un cadre bayésien classique, ces paramètres sont fixés en fonction de connaissances *a priori* extérieures au modèle. Lorsqu'on ne dispose pas de telle connaissances, il existe d'autres approches, comme les approches bayésiennes empiriques [Carlin 1997], qui utilisent les données pour estimer les hyperparamètres. Dans la continuité de [Punskaya 2002, Dobigeon 2007a], nous utilisons ici une approche bayésienne hiérarchique. Dans ce cadre, les hyperparamètres sont considérés comme des variables aléatoires et sont estimés en fonction des données au sein d'un second niveau d'estimation bayésienne. Cette approche a l'avantage de conférer une grande versatilité au modèle et de fournir une estimation robuste des paramètres dans le sens où aucun réglage de ces paramètres n'est nécessaire pour ajuster le modèle à un signal donné.

Les hyperparamètres que nous devons évaluer sont $\{\mu_K, \mu_Q, \delta^2, \{\nu_k, \rho_k\}_{k=1, \dots, K}\}$. Nous allons fixer $\nu_k = 1$ pour tout $k \in [1, K]$. Une raison pour laquelle nous fixons ces hyperparamètres est qu'il n'est pas possible de choisir une loi conjuguée conjointement pour ν_k et ρ_k qui sont respectivement les paramètres de forme et d'échelle classiques de la loi inverse-gamma. En fixant $\nu_k = 1$, il est possible de choisir une loi *a priori* conjuguée pour ρ_k . D'autre part nous fixons $\mu_Q = 2$.

Nous notons ϕ le vecteur des hyperparamètres qui ne sont pas fixés de manière déterministe, c-à-d $\phi = \{\mu_K, \delta^2, \{\rho_k\}_{k=1, \dots, K}\}$. Ces paramètres sont supposés aléatoires avec une loi *a priori* connue. Ces loi *a priori* sont choisies non-informatives ou au moins considérées suffisamment vagues car nous ne disposons d'aucune information *a priori* sur les valeurs probables de ces hyperparamètres.

Pour μ_K nous choisissons une loi *a priori* impropre non informative de Jeffrey :

$$p(\mu_K) \propto \frac{1}{\sqrt{\mu_K}} \mathbb{I}_{\mathbb{R}_+}(\mu_K). \quad (3.29)$$

Pour δ^2 une loi *a priori* conjuguée inverse-gamma est choisie :

$$\delta^2 \sim \mathcal{IG}(\Xi, \Delta). \quad (3.30)$$

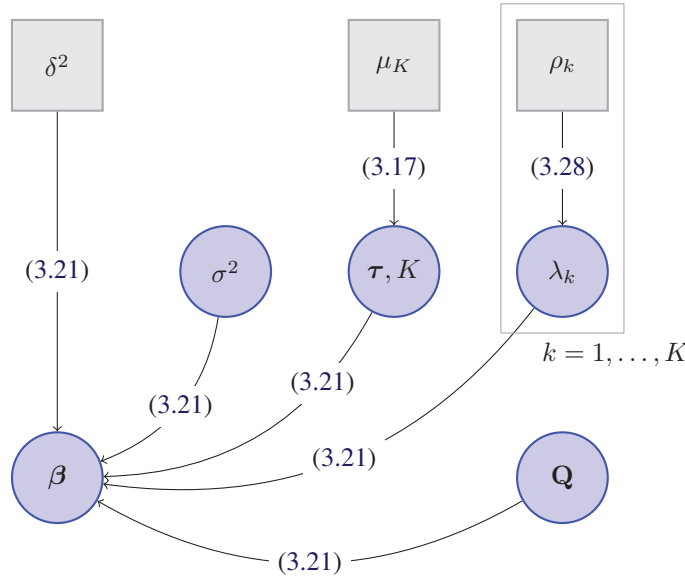


FIGURE 3.2: Graphe des paramètres du modèle : paramètres (cercles bleus), hyperparamètres (carrés gris). Chaque arc correspond à une dépendance conditionnelle entre les deux variables. L'étiquette des arcs correspond à l'équation qui souligne cette dépendance.

Les paramètres Ξ et Δ sont fixés de façon à ce que la loi soit suffisamment vague, $\Xi = 1$ et $\Delta = 100$.

Enfin pour les paramètres ρ_k une loi conjuguée gamma est définie :

$$\rho_k \sim \mathcal{G}(\chi, \psi) \quad \forall k = 1, \dots, K. \quad (3.31)$$

Là encore, χ et ψ sont choisis de manière à ce que la loi *a priori* soit vague, $\chi = 1$ et $\psi = 100$.

Le schéma 3.2 résume les relations entre les différents paramètres et hyperparamètres de la modélisation bayésienne. Les hyperparamètres fixés sont omis par souci de simplification. Les paramètres sont représentés dans des cercles bleus et les hyperparamètres sont représentés dans des carrés gris. Chaque arc du graphe représente une dépendance conditionnelle entre les deux variables connectées. La référence de l'équation soulignant cette dépendance est représentée sur l'arc.

Compromis nombre de ruptures / ordre des sous-modèles Nous avons choisi de fixer l'ordre moyen des sous modèles à la valeur $\mu_Q = 2$, mais de laisser le nombre moyen de ruptures μ_K variable. Ces deux lois *a priori* permettent de gérer le compromis qui existe entre le nombre de ruptures K et les ordres de chacun des sous-modèles $\{Q_k\}_{k=1, \dots, K+1}$. L'intérêt de ce compromis peut être mis en valeur eu égard à deux cas extrêmes dégénérés : Dans un cas, on peut représenter le signal observé de manière exacte par des sous-modèles très simples $Q_k = 0$ en multipliant le nombre de ruptures $K \rightarrow n - 1$ où n est le nombre de points observés. Dans l'autre cas, on ne représente le signal qu'avec un seul segment $K = 0$ mais on augmente l'ordre du sous-modèle jusqu'à obtenir une erreur de modélisation arbitrairement petite. Évidemment aucune de ces situations n'est souhaitée. Le choix des densités *a priori* permet de régler ce compromis. Dans cette thèse, nous avons choisi de favoriser les sous-modèles d'ordre faible en fixant l'ordre moyen à une valeur relativement faible $\mu_Q = 2$ tout en pénalisant peu les grandes valeurs du nombre de ruptures K . Ce

choix est pertinent en particulier pour des applications de lissage de courbes puisque les polynômes d'ordre élevé présentent des inconvénients notables (phénomène de Runge). Pour d'autres applications, des choix différents de loi *a priori* sont tout à fait valables pour K et $\{Q_k\}_{k=1,\dots,K+1}$ afin de régler ce compromis différemment.

3.4.3 Marginalisation de la densité conjointe *a posteriori*

La densité conjointe *a posteriori* des paramètres θ et des hyperparamètres ϕ du modèle bayésien hiérarchique s'exprime à un facteur multiplicatif près comme :

$$p(\theta, \phi | \mathbf{x}) \propto \mathcal{L}(\mathbf{x} | \theta) p(\theta | \phi) p(\phi). \quad (3.32)$$

Elle est composée du produit de la vraisemblance, des densités *a priori* des paramètres et des densités *a priori* des hyperparamètres.

Pour simplifier l'estimation du modèle, nous marginalisons la densité conjointe (3.32) par rapport aux paramètres/hyperparamètres (θ, ϕ) pour lesquels cette opération est réalisable analytiquement. Ceci revient à considérer ces paramètres comme des paramètres de nuisance, et à s'en affranchir dans la distribution *a posteriori* marginalisée. Remarquons que le choix de densités *a priori* conjuguées (tout au moins conditionnellement) s'avère crucial pour permettre l'expression analytique de cette loi marginale dans notre problème d'estimation.

Intégration de μ_K

L'hyperparamètre μ_K n'intervient dans la densité *a posteriori* conjointe (3.32) que par l'intermédiaire de la densité *a priori* de K (3.17) et sa propre loi *a priori* (3.29). La marginalisation de μ_K revient donc au calcul de la densité marginale *a priori* de K :

$$\begin{aligned} p(K) &= \int p(K | \mu_K) p(\mu_K) d\mu_K \\ &= \int \frac{\mu_K^{K-\frac{1}{2}} e^{-\mu_K}}{K!} \mathbb{I}_{\mathbb{R}_+}(\mu_K) d\mu_K \\ &= \frac{\Gamma(K + \frac{1}{2})}{K!} = \frac{\Gamma(K + \frac{1}{2})}{\Gamma(K + 1)}, \end{aligned}$$

pour tout $K \geq 0$. Cette expression montre que la loi *a priori* de K est une loi impropre pour laquelle les grandes valeurs de K ne sont que faiblement pénalisées. Finalement, après marginalisation par rapport à l'hyperparamètre μ_K , la loi jointe *a priori* du nombre de ruptures et de leur positions devient d'après l'expression précédente et l'expression (3.18)

$$p(\boldsymbol{\tau}, K) = p(\boldsymbol{\tau} | K) p(K) = \left(\frac{1}{L}\right)^K \frac{\Gamma(K + \frac{1}{2})}{\Gamma(K + 1)},$$

pour tout $K \geq 0$ et pour tout $\tau_1, \dots, \tau_K \in [\tau_0, \tau_{K+1}]$.

Intégration de $\{\rho_k\}_{k=1,\dots,K-1}$

Pour tout $k \in [1, K]$, l'hyperparamètre ρ_k n'intervient que via la loi *a priori* de λ_k (3.28) et sa propre loi *a priori* (3.31) dans la densité *a posteriori* conjointe (3.32). La marginalisation de ρ_k

revient au calcul de la densité marginale *a priori* de λ_k :

$$\begin{aligned} p(\lambda_k) &= \int p(\lambda_k | \nu_k, \rho_k) p(\rho_k) d\rho_k \\ &= \frac{1}{\lambda_k^{\nu_k+1} \Gamma(\nu_k) \Gamma(\chi) \psi^\chi} \int \rho_k^{\nu_k+\chi-1} e^{-\rho_k(\frac{1}{\lambda_k} + \frac{1}{\psi})} \mathbb{I}_{\mathbb{R}^+}(\lambda_k) d\rho_k \\ &= \frac{\Gamma(\nu_k + \chi)}{\Gamma(\nu_k) \Gamma(\chi)} \frac{\psi^{\nu_k+\chi}}{\chi^\psi} \frac{\lambda_k^{\chi-1}}{(\lambda_k + \psi)^{\nu_k+\chi}} \mathbb{I}_{\mathbb{R}^+}(\lambda_k). \end{aligned}$$

Intégration de β et σ^2

Les paramètres β et σ^2 interviennent dans la loi *a posteriori* conjointe (3.32) via la vraisemblance (3.14) et leur loi *a priori* respective (3.21),(3.27). L'intégration de β et σ^2 revient donc à :

$$\begin{aligned} & \int \int \mathcal{L}(\mathbf{x}|\theta) p(\beta|\delta^2, \sigma^2, \boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}) p(\sigma^2) d\beta d\sigma^2 \\ & \propto \int \int \frac{|Z^T Z|^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2} \|\mathbf{x} - Z\beta\|^2} e^{-\frac{1}{2\sigma^2 \delta^2} \beta^T Z^T Z \beta}}{(\sigma^2)^{\frac{n}{2}+1} (\sigma^2 \delta^2)^{\frac{1}{2} \sum_{k=1}^{K+1} (Q_k+1)}} \mathbb{I}_{\mathbb{R}^+}(\sigma^2) d\beta d\sigma^2 \\ & \propto \int \frac{e^{-\frac{1}{2\sigma^2} (\mathbf{x}^T \mathbf{x} - \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta)}}{(\sigma^2)^{\frac{n}{2}+1} (1 + \delta^2)^{\frac{1}{2} \sum_{k=1}^{K+1} (Q_k+1)}} \mathbb{I}_{\mathbb{R}^+}(\sigma^2) d\sigma^2 \\ & \propto \frac{(\delta^2 + 1)^{-\frac{1}{2} \sum_{k=1}^{K+1} (Q_k+1)}}{(\mathbf{x}^T \mathbf{x} - \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta)^{\frac{n}{2}}}, \end{aligned}$$

où μ_β et Σ_β sont les paramètres de la distribution *a posteriori* conditionnelle de β qui est une loi normale de moyenne μ_β et de matrice de covariance Σ_β tels que :

$$\mu_\beta = \frac{\delta^2}{1 + \delta^2} [Z^T Z]^{-1} Z^T \mathbf{x}, \quad (3.33)$$

$$\Sigma_\beta = \frac{\delta^2}{1 + \delta^2} (Z^T Z)^{-1}. \quad (3.34)$$

Finalement des expressions obtenues précédemment découle la densité marginale *a posteriori* suivante :

$$\boxed{p(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}, \delta^2 | \mathbf{x}) \propto p(K, \boldsymbol{\tau}) \left(\prod_{k=1}^{K+1} p(Q_k) \right) \left(\prod_{k=1}^K p(\lambda_k) \right) \frac{(\delta^2 + 1)^{-\frac{1}{2} \sum_{k=1}^{K+1} (Q_k+1)}}{(\mathbf{x}^T \mathbf{x} - \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta)^{\frac{n}{2}}} p(\delta^2)} \quad (3.35)$$

A ce stade, il n'est plus possible ni de marginaliser les paramètres et hyperparamètres restants $(\boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}, \delta^2)$, ni d'obtenir analytiquement l'expression de leurs estimateurs bayésiens classiques. En effet, leur influence sur la densité marginale *a posteriori* (3.35) s'avère non standard et trop complexe.

De manière très classique, nous proposons donc l'utilisation d'un algorithme d'échantillonnage MCMC (« *Markov-Chain Monte-Carlo* ») pour obtenir la distribution empirique de la densité marginale *a posteriori* (3.35).

3.5 Estimation par méthode MCMC à sauts réversibles

Les méthodes d'échantillonnage MCMC [Robert 2004] sont un moyen standard d'obtenir des échantillons asymptotiquement distribués selon une loi de probabilité, connue seulement à une constante multiplicative près, comme c'est le cas pour la densité marginale *a posteriori* de notre modèle (3.35). La particularité de notre modèle est que le vecteur des paramètres de la loi $(\tau, \eta, \mathbf{Q}, \delta^2)$ possède une dimension variable. Pour un nombre de transitions $K \in \mathbb{N}^*$ fixé, on a

$$(\delta^2, \tau, \eta, \mathbf{Q}) \in \mathbb{R}_+ \times \mathbb{R}^{k-1} \times \Upsilon^{k-1} \times \mathbb{R}^k, \quad (3.36)$$

où Υ est l'espace de définition des paramètres de la fonction de transition η_k (par exemple $\Upsilon = \mathbb{R}_+$ pour la fonction sigmoïde paramétrée par un paramètre d'échelle λ_k). Dans le cas général où K n'est pas fixé, l'espace de définition de ce vecteur de paramètres \mathcal{E} est la réunion de tous les espaces précédents :

$$\mathcal{E} = \mathbb{R}_+ \times \bigcup_{k \in \mathbb{N}^*} (\mathbb{R}^{k-1} \times \Upsilon^{k-1} \times \mathbb{R}^k). \quad (3.37)$$

La méthodologie proposée dans [Green 1995] permet de développer des algorithmes MCMC spécifiquement conçus pour échantillonner ce type d'espace trans-dimensionnel, appelés algorithmes MCMC à sauts réversibles (« *Reversible-Jump Markov-Chain Monte-Carlo* », RJMCMC). Nous avons choisi dans cette thèse de développer un algorithme de ce type. Il existe cependant d'autres méthodes pour gérer ce type de problèmes.

Citons par exemple les travaux de [Fearnhead 2005] qui proposent une méthode d'échantillonnage « parfait » (« *Perfect Sampling* », c'est-à-dire que l'échantillonnage est réalisé directement selon la loi cible) pour simuler la densité *a posteriori* d'un modèle de régression par morceaux. Malheureusement ces travaux semblent difficilement généralisables au cas des modèles à transitions régulières car ils nécessitent une hypothèse d'indépendance entre les fonctions de régression sur chaque segment, ce qui n'est pas le cas du modèle à transitions régulières.

Une autre méthode, alternative aux algorithmes RJMCMC, est l'utilisation du modèle Bernoulli-Gaussien [Lavielle 1998, Dobigeon 2007a, Dobigeon 2007b]. Dans ce cas, un vecteur de variables indicatrices binaires de la même taille que le signal observé est défini, chaque variable prenant la valeur 1 si une transition est présente à l'instant correspondant, la valeur 0 sinon. L'avantage considérable de cette approche est que la dimension de l'espace à échantillonner est alors fixée : on inclut en fait l'espace original \mathcal{E} de dimension variable dans un espace plus grand mais de dimension fixe. On peut alors appliquer des méthodes MCMC classiques et simples à mettre en œuvre, tel que l'algorithme de Gibbs [Gelfand 1990], pour échantillonner la loi voulue. Nous avons dans un premier temps évalué les performances de cette méthode pour le modèle. Elle présente cependant plusieurs inconvénients :

- D'une part, il faut alors ajouter, non pas une variable indicatrice par échantillon du signal \mathbf{x} comme pour les modèles de régression par morceaux, mais une variable indicatrice par échantillon et par paramètre de la fonction de transition. De plus la position des transitions évoluant de manière continue et non discrète, il nous faut encore ajouter une variable par échantillon pour modéliser la position d'une transition entre deux instants d'échantillonnage. Le nombre de variables additionnelles est donc considérable, ce qui conduit à un algorithme d'échantillonnage particulièrement coûteux d'un point de vue calculatoire.
- D'autre part, nous nous sommes heurtés à des problèmes majeurs de mélange de l'algorithme, c'est-à-dire des difficultés à explorer certaines zones de l'espace échantillonné. La raison majeure est que l'algorithme de Gibbs ne propose que des changements élémentaires

du vecteur de paramètre pour explorer l'espace, ce qui s'avère insuffisant pour explorer les modes de notre modèle à transitions régulières. Notons que des méthodes, développées afin de s'affranchir de ce problème, ont été récemment publiées [Kail 2012].

3.5.1 Vue d'ensemble de l'algorithme RJMCMC

L'algorithme RJMCMC développé pour l'estimation de nos signaux consiste en une boucle principale répétée un nombre d'itérations N_{iter} donné. On définit $(\boldsymbol{\tau}^{(i)}, \boldsymbol{\eta}^{(i)}, \mathbf{Q}^{(i)}, \delta^{2(i)})$ la valeur des paramètres à l'itération i . A chaque itération un mouvement, c'est-à-dire une perturbation de la valeur courante des paramètres $(\boldsymbol{\tau}^{(i)}, \boldsymbol{\eta}^{(i)}, \mathbf{Q}^{(i)}, \delta^{2(i)})$, est proposé. Les valeurs des paramètres proposés lors du mouvement sont notés $(\tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\eta}}, \tilde{\mathbf{Q}}, \tilde{\delta}^2)$. Ce mouvement est ensuite accepté ou rejeté selon une certaine probabilité qui dépend du modèle, du mouvement et des valeurs courante et proposée des paramètres. En cas d'acceptation, la valeur proposée devient la valeur courante, sinon la valeur courante est conservée pour la prochaine itération. Ci-dessous l'algorithme est présenté sous forme de pseudo-code. Les différents mouvements sont présentés individuellement dans la suite de ce chapitre. Les probabilités $p_{\text{naiss}}^{(i)}$, $p_{\text{mort}}^{(i)}$, $p_{\text{divi}}^{(i)}$, $p_{\text{fusion}}^{(i)}$ et $p_{\text{maj}}^{(i)}$ sont respectivement les probabilités qu'un mouvement de naissance, de mort, de division, de fusion ou de mise à jour des paramètres des fonctions de transition soit tenté à l'itération i . La valeur complémentaire $p_{\text{majQ}}^{(i)} = 1 - p_{\text{naiss}}^{(i)} + p_{\text{mort}}^{(i)} + p_{\text{divi}}^{(i)} + p_{\text{fusion}}^{(i)} + p_{\text{maj}}^{(i)}$ est donc la probabilité qu'un mouvement de mise à jour de l'ordre des sous-modèles soit tenté à l'itération i . Par la suite, nous considérerons que ces probabilités sont toutes égales entre elles.

```

Initialisation :  $K^{(0)} = 0$ ,  $\boldsymbol{\tau}^{(0)} = \boldsymbol{\lambda}^{(0)} = \emptyset$ ,  $\delta^{2(0)} = \text{var}(\boldsymbol{x})$ ,  $Q_1 = 0$  ;
for  $i = 1 : N_{\text{iter}}$  do
  Échantillonner  $u \sim \mathcal{U}_{[0 \ 1]}$  ;
  if  $u < p_{\text{naiss}}^{(i)}$  then
    | Exécuter un mouvement de naissance d'une transition (voir 3.5.3);
  else if  $u < p_{\text{naiss}}^{(i)} + p_{\text{mort}}^{(i)}$  then
    | Exécuter un mouvement de mort d'une transition (voir 3.5.3);
  else if  $u < p_{\text{naiss}}^{(i)} + p_{\text{mort}}^{(i)} + p_{\text{divi}}^{(i)}$  then
    | Exécuter un mouvement de division d'une transition (voir 3.5.4);
  else if  $u < p_{\text{naiss}}^{(i)} + p_{\text{mort}}^{(i)} + p_{\text{divi}}^{(i)} + p_{\text{fusion}}^{(i)}$  then
    | Exécuter un mouvement de fusion de deux transitions (voir 3.5.4);
  else if  $u < p_{\text{naiss}}^{(i)} + p_{\text{mort}}^{(i)} + p_{\text{divi}}^{(i)} + p_{\text{fusion}}^{(i)} + p_{\text{maj}}^{(i)}$  then
    | Exécuter un mouvement de mise à jour d'une transition (voir 3.5.5);
  else
    | Exécuter un mouvement de mise à jour de l'ordre d'un sous-modèle (voir 3.5.6);
  end
  Mettre à jour  $\delta^2$  (voir 3.5.7) ;
end

```

Algorithme 1: Algorithme RJMCMC pour l'estimation des paramètres du modèle à transitions régulières

3.5.2 Expression générale du rapport de Green

Dans les travaux de Green [Green 1995], l'auteur définit une méthode pour construire la probabilité qu'un mouvement soit accepté. L'auteur montre que cette méthode garantit la réversibilité de la chaîne ainsi générée se qui assure la convergence de la chaîne vers la loi cible $\pi(x)$. Pour construire la probabilité d'acceptation de passer d'un état x à un état \tilde{x} , on définit dans un premier temps deux variables auxiliaires, notées u et \tilde{u} telle que la transformation $(x, u) \rightarrow (\tilde{x}, \tilde{u})$ soit un difféomorphisme. Le rapport de Green s'exprime alors, pour chaque mouvement m du noyau de proposition conformément à [Green 1995] :

$$r_m = \frac{\pi(\tilde{x})p_m(\tilde{x})\tilde{q}_m(\tilde{u})}{\pi(x)p_m(x)q_m(u)} \left| \frac{\partial(\tilde{x}, \tilde{u})}{\partial(x, u)} \right| \quad (3.38)$$

où $\pi(x)$ est la loi cible, $p_m(x)$ est la probabilité de tenter le mouvement m depuis l'état x , $q_m(u)$ est la loi instrumentale qui permet de générer la variable auxiliaire u et $\left| \frac{\partial(\tilde{x}, \tilde{u})}{\partial(x, u)} \right|$ est le Jacobien de la transformation $(x, u) \rightarrow (\tilde{x}, \tilde{u})$. La probabilité que le mouvement m soit accepté est alors $\min(1, r_m)$. Notons que cette probabilité fait intervenir conjointement des grandeurs liées au mouvement m et au mouvement inverse pour garantir la réversibilité de la chaîne.

Dans cette thèse, pour chaque mouvement, les variables auxiliaires u, \tilde{u} ne dépendent pas de la configuration des paramètres x, \tilde{x} . De fait, le Jacobien de la transformation $(x, u) \rightarrow (\tilde{x}, \tilde{u})$ sera toujours égal à l'unité. De fait l'approche proposée est proche de la méthodologie proposée par les auteurs de [Geyer 1994]. La loi cible dans ce chapitre est la loi marginale *a posteriori* définie par (3.35). Sous ces hypothèses, on peut réécrire le rapport de Green comme :

$$r_m = \frac{p(\tilde{\theta}|\mathbf{x})p_m(\tilde{\theta})\tilde{q}_m(\tilde{u})}{p(\theta|\mathbf{x})p_m(\theta)q_m(u)} \quad (3.39)$$

3.5.3 Mouvement de naissance/mort d'une transition

Un mouvement de naissance consiste à échantillonner une nouvelle transition et à l'ajouter à la configuration courante des paramètres. La position de la nouvelle transition $\tilde{\tau}$, les paramètres de la fonction de transition associée $\tilde{\eta}$ et l'ordre du sous-modèle ajouté \tilde{Q} sont tirés suivant la loi instrumentale q_0 ,

$$\tilde{\tau}, \tilde{\eta}, \tilde{Q} \sim q_0(\tau, \eta, Q) = q_0(\tau)q_0(\eta)q_0(Q). \quad (3.40)$$

Les différentes composantes de cette loi instrumentale sont exposées ci-après.

Proposition de la position des ruptures On tire la position $\tilde{\tau}$ des nouvelles transitions selon

$$q_0(\tau) = g(\tau), \quad (3.41)$$

où $g(\tau)$ est une fonction d'intensité proportionnelle à la valeur absolue du gradient de x . Pour obtenir g , on calcule \bar{x} en filtrant les observations \mathbf{x} à l'aide d'une fenêtre gaussienne d'écart-type s (par la suite, $s = 4$ échantillons). Puis une approximation discrète du gradient de $x(t)$ est calculée :

$$\nabla x(t_i) = \frac{\bar{x}(t_{i+1}) - \bar{x}(t_{i-1})}{2t_e},$$

où t_e est la fréquence d'échantillonnage du signal. Enfin la fonction d'intensité g est définie comme :

$$g(t) = \frac{\sum_{i=1}^n |\nabla x(t_i)| \mathbb{I}_{[t_i - \frac{t_e}{2}, t_i + \frac{t_e}{2}]}(t)}{C}.$$

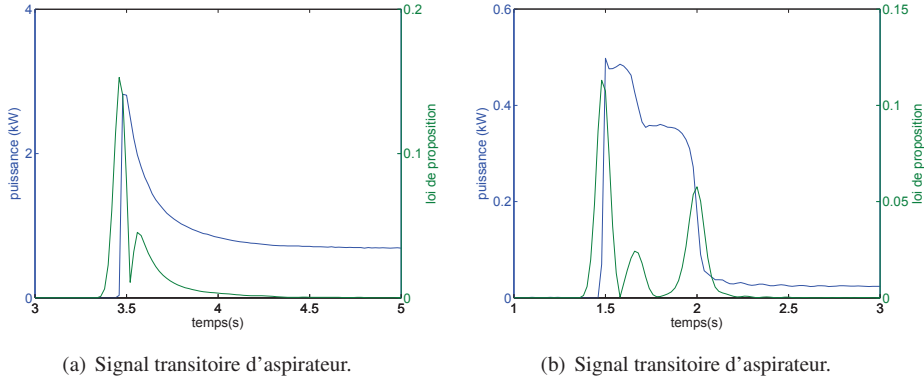


FIGURE 3.3: Exemple de loi instrumentale basée sur le gradient pour la position des ruptures. Signal réel (bleu) Densité de probabilité de la loi instrumentale (vert).

Où C est une constante de normalisation qui vaut :

$$\begin{aligned}
 C &= \int \sum_{i=1}^n \left(|\nabla x(t_i)| \mathbb{I}_{[t_i - \frac{t_e}{2}, t_i + \frac{t_e}{2}]}(t) \right) dt = \sum_{i=1}^n \left(|\nabla x(t_i)| \int \mathbb{I}_{[t_i - \frac{t_e}{2}, t_i + \frac{t_e}{2}]}(t) dt \right) \\
 &= \sum_{i=1}^n (|\nabla x(t_i)| t_e) \\
 &= t_e \sum_{i=1}^n (|\nabla x(t_i)|).
 \end{aligned}$$

Cette loi instrumentale permet de proposer la position des ruptures avec une plus grande probabilité aux instants où le gradient prends des valeurs importantes (positives ou négatives). Cela correspond à l'hypothèse que les transitions devraient se trouver à proximité des instants où le signal varie le plus. Cette stratégie améliore significativement l'efficacité de l'algorithme par rapport à une loi instrumentale uniforme sur $[\tau_0, \tau_K]$ par exemple. La figure 3.3 montre la densité de probabilité de cette loi instrumentale, pour deux exemples de signaux transitoires tirés de la base de donnée SISED, un signal d'aspirateur 3.3(a), et un signal de réfrigérateur 3.3(b).

Nous avons choisi ici de modéliser g par une fonction étagée changeant de valeur en chaque instant d'échantillonnage. Il serait également possible de réaliser un échantillonnage non uniforme de cette densité. Cela pourrait permettre d'améliorer la rapidité et la précision avec laquelle les valeurs $\tilde{\tau}$ sont tirées.

Proposition du paramètre d'échelle des ruptures Le paramètre d'échelle $\tilde{\lambda}$ des nouvelles fonctions transitions est tiré selon une loi exponentielle :

$$q_0(\lambda) = \frac{1}{\lambda_m} e^{-\frac{\lambda}{\lambda_m}} \mathbb{I}_{\mathbb{R}^+}(\lambda),$$

où $\lambda_m = 2$. Cette loi instrumentale favorise l'apparition de transitions abruptes tout en pouvant générer des valeurs arbitrairement larges.

Proposition de l'ordre du sous-modèle L'ordre du nouveau sous-modèle \tilde{Q} est tiré selon sa loi *a priori* (3.20),

$$q_0(Q) = p(Q) = \frac{\mu_Q^{Q_k} e^{-\mu_Q}}{Q_k!}. \quad (3.42)$$

Ratio d'acceptation du mouvement de naissance/mort La loi instrumentale $q_0(\tau, \eta, Q)$ pour les mouvements de naissance d'une rupture est donc assez générique, permettant d'échantillonner la totalité de l'espace des paramètres. La nouvelle transition générée selon la loi instrumentale est alors ajoutée à la configuration courante des transitions $\tilde{\theta} = \theta^{(i)} \cup \{(\tilde{\tau}, \tilde{\eta}, \tilde{Q})\}$. La variable auxiliaire u correspond donc à $(\tilde{\tau}, \tilde{\eta}, \tilde{Q})$ et la variable auxiliaire $\tilde{u} = \emptyset$. Le taux d'acceptation de Metropolis-Hastings-Green s'exprime donc

$$r_{\text{naiss}} = \frac{p(\tilde{\theta}|\mathbf{x}) p_{\text{mort}}(\tilde{\theta})}{p(\theta|\mathbf{x}) p_{\text{naiss}}(\theta)} \frac{1}{q_0(\tilde{\tau}, \tilde{\eta}, \tilde{Q})}. \quad (3.43)$$

Le mouvement inverse de la naissance d'une transition est le mouvement de mort d'une transition qui consiste à sélectionner au hasard une transition parmi celles existantes $k \sim \mathcal{U}_{[1, K^{(i)} - 1]}$ et à l'en retirer $\tilde{\theta} = \{\theta^{(i)} \setminus (\tau_k, \eta_k, Q_k)\}$. Le taux d'acceptation de Metropolis-Hastings-Green s'exprime de manière symétrique à celui de naissance :

$$r_{\text{mort}} = \frac{p(\tilde{\theta}|\mathbf{x}) p_{\text{naiss}}(\tilde{\theta})}{p(\theta|\mathbf{x}) p_{\text{mort}}(\theta)} q_0(\tau_k, \eta_k, Q_k). \quad (3.44)$$

Puisque la transition supprimée est sélectionnée au hasard, on a $p_{\text{mort}}(\theta) = \frac{1}{K}$.

Les paramètres ne dépendant pas de la transition ajoutée ou supprimée sont inchangés lors d'un mouvement de naissance ou mort. Les taux d'acceptations se simplifient donc :

$$r_{\text{naiss}} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_{\beta}^{(i)T} \Sigma_{\beta}^{(i)-1} \mu_{\beta}^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_{\beta}^T \tilde{\Sigma}_{\beta}^{-1} \tilde{\mu}_{\beta}} \right)^{\frac{n}{2}} \frac{p(\tilde{\lambda}) p(\tilde{Q})}{q_0(\tilde{\tau}, \tilde{\lambda}, \tilde{Q})} \frac{K^{(i)} + \frac{1}{2} (1 + \delta^2)^{-\frac{1}{2}(\tilde{Q}+1)}}{K^{(i)} + 1} \frac{1}{(K^{(i)} + 1)L} \quad (3.45)$$

$$r_{\text{mort}} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_{\beta}^{(i)T} \Sigma_{\beta}^{(i)-1} \mu_{\beta}^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_{\beta}^T \tilde{\Sigma}_{\beta}^{-1} \tilde{\mu}_{\beta}} \right)^{\frac{n}{2}} \frac{q_0(\tau_k, \lambda_k, Q_k)}{p(\lambda_k) p(Q_k)} \frac{K^{(i)}}{K^{(i)} - \frac{1}{2} (1 + \delta^2)^{-\frac{1}{2}(Q_k+1)}} \frac{K^{(i)} L}{1} \quad (3.46)$$

où $\tilde{\mu}_{\beta}$ et $\tilde{\Sigma}_{\beta}$ sont respectivement la moyenne et la matrice de covariance de la loi *a posteriori* de β conditionnellement à la configuration proposée $\tilde{\theta}$.

3.5.4 Mouvement de fusion/division d'une transition

Le mouvement de division consiste à sélectionner une transition aléatoirement dans la configuration courante $k \sim \mathcal{U}_{[1, K]}$ puis à tirer deux nouvelles transitions selon la loi instrumentale q_0 tronquée aux valeurs de τ comprises entre τ_{k-1} et τ_{k+1} :

$$\begin{aligned} \tilde{\tau}_1, \tilde{\eta}_1, \tilde{Q}_1 &\sim q_0(\tau, \eta, Q) \mathbb{I}_{[\tau_{k-1}, \tau_{k+1}]}(\tau), \\ \tilde{\tau}_2, \tilde{\eta}_2, \tilde{Q}_2 &\sim q_0(\tau, \eta, Q) \mathbb{I}_{[\tau_{k-1}, \tau_{k+1}]}(\tau). \end{aligned}$$

La transition sélectionnée est alors remplacée dans la nouvelle configuration par les deux transitions tirées $\tilde{\theta} = \{\theta^{(i)} \cup (\tilde{\tau}_1, \tilde{\eta}_1, \tilde{Q}_1) \cup (\tilde{\tau}_2, \tilde{\eta}_2, \tilde{Q}_2) \setminus \{(\tau_k, \eta_k, Q_k)\}$. Le mouvement de division correspond

donc à réaliser conjointement un mouvement de mort et deux mouvements de naissance. On peut donc déduire le taux d'acceptation de Metropolis-Hastings-Green du mouvement de division des taux d'acceptation de naissance (3.45) et de mort (3.46) :

$$r_{\text{division}} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_{\beta}^{(i)T} \Sigma_{\beta}^{(i)-1} \mu_{\beta}^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_{\beta}^T \tilde{\Sigma}_{\beta}^{-1} \tilde{\mu}_{\beta}} \right)^{\frac{n}{2}} \frac{p(\tilde{\lambda}_1) p(\tilde{Q}_1) p(\tilde{\lambda}_2) p(\tilde{Q}_2) q_0(\tau_k, \lambda_k, Q_k)}{q_0(\tilde{\tau}_1, \tilde{\lambda}_1, \tilde{Q}_1) q_0(\tilde{\tau}_2, \tilde{\lambda}_2, \tilde{Q}_2) p(\lambda_k) p(Q_k)} \\ \times \frac{K^{(i)} + \frac{1}{2} (1 + \delta^2)^{-\frac{1}{2}(\tilde{Q}+1)}}{K^{(i)} + 1} \frac{K^{(i)} L}{K^{(i)} L}. \quad (3.47)$$

Un mouvement de fusion consiste à sélectionner aléatoirement une paire de transitions voisines. C'est-à-dire que l'on tire $k \sim \mathcal{U}_{[1, K-1]}$ et on sélectionne les transitions $\{\tau_k, \tau_{k+1}\}$. On tire une nouvelle transition selon la loi instrumentale q_0 tronquée aux valeurs de τ comprises entre τ_{k-1} et τ_{k+2} :

$$\tilde{\tau}, \tilde{\eta}, \tilde{Q} \sim q_0(\tau, \eta, Q) \mathbb{I}_{[\tau_{k-1}, \tau_{k+2}]}(\tau). \quad (3.48)$$

Les deux transitions sélectionnées sont retirées de la configuration des ruptures tandis que la nouvelle transition y est ajoutée $\tilde{\theta} = \{\theta^{(i)} \cup (\tilde{\tau}, \tilde{\eta}, \tilde{Q}) \setminus \{(\tau_k, \eta_k, Q_k) \cup (\tau_{k+1}, \eta_{k+1}, Q_{k+1})\}\}$. Le mouvement de fusion correspond donc à réaliser conjointement deux mouvements de mort et un mouvement de naissance. Le taux d'acceptation de Metropolis-Hastings-Green est alors :

$$r_{\text{fusion}} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_{\beta}^{(i)T} \Sigma_{\beta}^{(i)-1} \mu_{\beta}^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_{\beta}^T \tilde{\Sigma}_{\beta}^{-1} \tilde{\mu}_{\beta}} \right)^{\frac{n}{2}} \frac{q_0(\tau_k, \lambda_k, Q_k) q_0(\tau_{k+1}, \lambda_{k+1}, Q_{k+1}) p(\tilde{\lambda}) p(\tilde{Q}_k)}{p(\lambda_k) p(Q_k) p(\lambda_{k+1}) p(Q_{k+1}) q_0(\tilde{\tau}, \tilde{\lambda}_k, \tilde{Q}_k)} \\ \times \frac{K^{(i)}}{K^{(i)} - \frac{1}{2} (1 + \delta^2)^{-\frac{1}{2}(Q_k+1)}} \frac{(K^{(i)} - 1) L}{K^{(i)} L}. \quad (3.49)$$

3.5.5 Mouvement de mise à jour des transitions

Ce mouvement consiste à sélectionner une transition aléatoirement dans la configuration courante $k \sim \mathcal{U}_{[1, K]}$ et à proposer une actualisation de (τ_k, η_k) . Les valeurs proposées sont tirées selon la loi instrumentale $q_1(\tau, \eta) = q_1(\tau) q_1(\eta)$. La loi instrumentale se divise donc en deux parties.

Proposition de la position des ruptures La nouvelle position de la rupture $\tilde{\tau}$ est échantillonnée selon la loi instrumentale

$$q_1(\tau) = w q_0(\tau) + \frac{1-w}{\sqrt{2\pi\sigma_{\tau}^2}} \exp\left\{-\frac{1}{2\sigma_{\tau}^2} \left[\tau - \tau^{(i-1)}\right]^2\right\}, \quad (3.50)$$

où $w = 0.2$, $\sigma_{\tau}^2 = 0.1$ et $\tau^{(i-1)}$ est la valeur courante de τ . Cette loi instrumentale est un mélange de la loi basée sur le gradient $q_0(\tau)$ utilisée pour les naissance de transitions et d'une marche aléatoire gaussienne autour de la valeur courante de τ .

Proposition du paramètre d'échelle Le nouveau paramètre d'échelle $\tilde{\lambda}$ est échantillonné selon la loi instrumentale

$$q_1(\lambda) = w q_0(\lambda) + \frac{1-w}{\sqrt{2\pi\sigma_{\lambda}^2}} e^{-\frac{1}{2\sigma_{\lambda}^2} (\lambda - \lambda^{(i-1)})^2} \mathbb{I}_{\mathbb{R}^+}(\lambda),$$

où $w = 0.2$, $\sigma_\lambda^2 = 0.1$ et $\lambda^{(i-1)}$ est la valeur courante de λ . Là encore, la loi instrumentale $q_1(\lambda)$ mélange la loi $q_0(\lambda)$ utilisée pour les mouvements de naissance de transitions et une marche aléatoire gaussienne autour de la valeur courante de λ .

Ce type de loi instrumentale ($q_1(\tau)$ et $q_1(\lambda)$) offre un bon compromis entre la capacité de q_0 à explorer l'intégralité de l'espace des paramètres et la marche aléatoire qui permet d'explorer les modes locaux de la distribution *a posteriori*.

La nouvelle transition est ajoutée tandis que l'ancienne est retirée de la configuration courante $\tilde{\theta} = \{\{\theta^{(i)} \cup (\tilde{\tau}, \tilde{\eta}, \tilde{Q})\} \setminus \{(\tau_k, \eta_k, Q_k)\}\}$.

Le taux d'acceptation de Metropolis-Hastings est alors :

$$r_{\text{maj}} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_\beta^{(i)T} \Sigma_\beta^{(i)-1} \mu_\beta^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_\beta^T \tilde{\Sigma}_\beta^{-1} \tilde{\mu}_\beta} \right)^{\frac{n}{2}} \frac{p(\tilde{\lambda}) q_1(\tau^{(i)}, \eta^{(i)})}{p(\lambda_k^{(i)}) q_1(\tilde{\tau}, \tilde{\eta})}. \quad (3.51)$$

3.5.6 Mouvement de mise à jour de l'ordre d'un sous-modèle

Ce mouvement consiste à choisir aléatoirement l'un des sous-modèles $k \sim \mathcal{U}_{[1, K+1]}$ et à proposer une nouvelle valeur \tilde{Q} de l'ordre du sous-modèle Q_k selon la loi instrumentale $q(Q)$:

$$\tilde{Q} \sim q(Q) = \begin{cases} 1/2 & \text{if } Q = Q^{(i-1)} - 1, \\ 1/2 & \text{if } Q = Q^{(i-1)} + 1, \\ 0 & \text{sinon.} \end{cases}$$

L'ordre du sous-modèle est donc soit augmenté soit diminué de 1 avec une probabilité égale. Évidemment, dans le cas où Q_k est nul, seul le mouvement $Q = Q^{(i-1)} + 1$ est proposé. Le taux d'acceptation de Metropolis-Hastings-Green est alors :

$$r_{\text{maj}Q} = \left(\frac{\mathbf{x}^T \mathbf{x} - \mu_\beta^{(i)T} \Sigma_\beta^{(i)-1} \mu_\beta^{(i)}}{\mathbf{x}^T \mathbf{x} - \tilde{\mu}_\beta^T \tilde{\Sigma}_\beta^{-1} \tilde{\mu}_\beta} \right)^{\frac{n}{2}} \times \frac{p(\tilde{Q}) q(Q_k)}{p(Q_k) q(\tilde{Q})} (1 + \delta^2)^{-\frac{1}{2}(Q - Q_k)}. \quad (3.52)$$

3.5.7 Mise à jour de δ^2

La densité conditionnelle *a posteriori* de δ^2 obtenue à partir de la densité marginale *a posteriori* (3.35) n'est pas standard. Cependant la distribution conditionnelle de δ^2 , étant donnée σ^2 et β , déduite de la densité conjointe *a posteriori* (3.32) est parfaitement connue et facile à simuler. Une stratégie d'échantillonnage de Gibbs avec augmentation (voir à ce propos [DYK, Park 2009]) des données est donc utilisée pour échantillonner conjointement δ^2 , σ^2 et β à partir de la densité conjointe *a posteriori* (3.32) :

1. La variance du bruit σ^2 est échantillonnée selon sa densité conditionnelle :

$$\sigma^2 | (\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}, \delta^2) \sim \mathcal{IG} \left(\frac{n}{2}, \frac{1}{2} \left(\mathbf{x}^T \mathbf{x} - \mu_\beta^T \Sigma_\beta^{-1} \mu_\beta \right) \right), \quad (3.53)$$

2. Les paramètres des sous-modèles β sont échantillonnés selon :

$$\beta | (\mathbf{x}, \boldsymbol{\tau}, \boldsymbol{\eta}, \mathbf{Q}, \sigma^2, \delta^2) \sim \mathcal{N} \left(\frac{\delta^2}{1 + \delta^2} [Z^T Z]^{-1} Z^T \mathbf{x}, \sigma^2 \frac{\delta^2}{1 + \delta^2} (Z^T Z)^{-1} \right), \quad (3.54)$$

3. Finalement, le paramètre δ^2 est échantillonné selon :

$$\delta^2 | (\boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma^2, \mathbf{Q}) \sim \mathcal{IG} \left(\Xi + \frac{1}{2} \sum_{k=1}^K (Q_k + 1), \Delta + \frac{1}{2\sigma^2} \boldsymbol{\beta}^T \mathbf{Z}^T \mathbf{Z} \boldsymbol{\beta} \right) \quad (3.55)$$

Les paramètres étant échantillonnés selon leur loi conditionnelle *a posteriori*, la loi cible et la loi instrumentale sont les mêmes, et le taux d'acceptation de ce mouvement est donc toujours égal à un.

3.6 Évaluation des performances du modèle pour la représentation de signaux tests

Afin de valider l'algorithme d'estimation et de comparer les performances du modèle à transition régulières à d'autres algorithmes de la littérature, nous avons réalisé plusieurs simulations sur des signaux synthétiques. Ces signaux se divisent en deux classes : les signaux réguliers d'une part, les signaux non-réguliers d'autre part.

3.6.1 Conditions de simulation

Pour tous les signaux tests, les fonctions de transitions sont sigmoïdales et les sous-modèles sont polynomiaux. L'algorithme RJMCMC 1 est utilisé pour générer une chaîne de Markov d'échantillons des paramètres $\theta^{(1)}, \dots, \theta^{(N_{\text{iter}})}$ distribués asymptotiquement selon la loi marginale *a posteriori* (3.35). Pour toutes ces simulations, nous fixons $N_{\text{iter}} = 5 \times 10^5$. De plus, les $N_{\text{chauffe}} = 10^5$ premiers échantillons correspondant à la période de chauffe (« *burn-in* ») ne sont pas utilisés pour estimer la fonction de régression.

Nous estimons la fonction de régression $\hat{f}(t)$ pour chaque signal $f(t)$ en effectuant un moyennage point par point des fonctions de régression $f_{\theta^{(i)}}(t)$ générées à partir des paramètres $\theta^{(i)}$ de la chaîne :

$$\hat{f}(t) = \frac{1}{N_{\text{iter}} - N_{\text{chauffe}}} \sum_{i=N_{\text{chauffe}}+1}^{N_{\text{iter}}} f_{\theta^{(i)}}(t). \quad (3.56)$$

Dans certains cas, nous présentons également l'estimateur au sens du maximum *a posteriori* (MAP) de la fonction de régression défini par :

$$\hat{f}_{\text{map}}(t) = f_{\theta^{(i_{\text{map}})}}(t), \quad (3.57)$$

avec i_{map} défini par :

$$i_{\text{map}} = \arg \max_i f(\theta^{(i)} | \mathbf{x}). \quad (3.58)$$

L'estimateur MAP (3.57) de la fonction de régression sera généralement moins satisfaisant au sens de l'erreur quadratique moyenne que l'estimateur bayésien (3.56), il est donc moins approprié pour des applications de lissage de courbe. Cependant, l'estimateur MAP à l'avantage de correspondre effectivement à un modèle à transition régulière, dans le sens où cet estimateur est déterminé par une unique valeur des paramètres $\theta_{\text{map}} = \theta^{(i_{\text{map}})}$ alors que l'estimateur (3.56) nécessite de connaître toute la chaîne de valeur $\theta^{(i)}$ pour être construit. L'estimateur MAP est donc plus approprié à des applications de réduction de dimension.

Pour mesurer la qualité de ces estimateurs et la parcimonie du modèle associé, nous utilisons deux indicateurs :

1. L'erreur quadratique moyenne entre la fonction $f(t)$ (non-bruitée) et son estimée $\hat{f}(t)$:

$$\text{eqm} = \frac{1}{n} \sum_{i=1}^n [\hat{f}(t_i) - f(t_i)]^2. \quad (3.59)$$

2. Le nombre de paramètres du modèle associé à l'estimateur MAP. C'est à dire la valeur

$$\begin{aligned} N_p &= \text{card}(\boldsymbol{\beta}^{(i_{\text{map}})}) + \text{card}(\boldsymbol{\tau}^{(i_{\text{map}})}) + \text{card}(\boldsymbol{\lambda}^{(i_{\text{map}})}), \\ &= \sum_{k=1}^{K^{(i_{\text{map}})+1}} (Q_k^{(i_{\text{map}})} + 1) + 2K^{(i_{\text{map}})}, \end{aligned} \quad (3.60)$$

3.6.2 Signaux tests réguliers

Nous utilisons, pour valider la méthode sur des signaux réguliers, deux signaux tests f_1 et f_2 issus de [Denison 1998] et qui s'expriment respectivement comme :

$$f_1(t) = t + 2 \exp(-16t^2) \quad t \in [-2, 2], \quad (3.61)$$

$$f_2(t) = \sin(2t) + 2 \exp(-16t^2) \quad t \in [-2, 2]. \quad (3.62)$$

Conformément à [Denison 1998], ces fonctions sont ré-échelonnées sur un intervalle unité $t \in [0, 1]$, et sont évaluées selon une grille d'échantillonnage uniforme de $n = 200$ points. Un bruit blanc gaussien de variance σ^2 est ajouté à chaque fonction $f_1(t)$ et $f_2(t)$ de sorte que le rapport signal à bruit soit égal à 3 (C'est-à-dire $\sigma = 0.4$ pour la fonction $f_1(t)$ et $\sigma = 0.3$ pour la fonction $f_2(t)$).

La figure 3.4 présente pour la première fonction test $f_1(t)$, le tracé de la fonction synthétique, sur la Fig. 3.4(a), un exemple de réalisation de la fonction bruitée, sur la Fig. 3.4(c), et la fonction de régression estimée $\hat{f}(t)$ par l'estimateur point par point (3.56) superposée aux estimateurs MAP de la position des transitions et de l'ordre des polynômes de chaque sous-modèle, sur la Fig. 3.4(e). Les figures 3.4(b), 3.4(d) et 3.4(f) présentent les mêmes tracés pour la deuxième fonction test $f_2(t)$. On peut apprécier visuellement la qualité de la reconstruction malgré la présence de quelques artefacts attribuables au bruit. Notons bien que la courbe tracée est estimée par moyennage point par point, alors que la position des ruptures correspond à l'estimateur MAP. La figure 3.5 permet, pour la fonction f_1 , de comparer la fonction de régression estimée $\hat{f}_{\text{map}}(t)$ par l'estimateur MAP avec des fonctions de transitions sigmoïdales, sur la Fig. 3.5(a) et la fonction de régression estimée $\hat{f}_{\text{map}}(t)$ par l'estimateur MAP avec des fonctions de transitions échelons, sur la Fig. 3.5(b). Les figures 3.5(b) et 3.5(d) présentent les mêmes tracés pour la fonction $f_2(t)$. Dans chacun des cas, l'erreur quadratique moyenne, dénotée eqm, est reportée dans la légende. La fonction de transition échelon réduit le modèle à transition régulière à un modèle de régression par morceaux classique (voir (3.6)). Ces simulations permettent de mettre en valeur l'intérêt du modèle à transition régulière par rapport à un modèle de régression par morceaux classique. En effet dans les deux cas, et plus particulièrement pour la première fonction, la fonction de régression à transition régulière reproduit mieux le comportement de la fonction test se qui se traduit par une erreur quadratique plus importante pour le modèle de régression par morceaux. Cette constatation est moins valable pour l'estimateur point par point car il produit un lissage statistique de la courbe, même si le modèle sous-jacent n'est pas régulier.

Nous avons appliqué l'algorithme RJMCMC 1 à ces deux fonctions synthétique bruitées pour 30 réalisations différentes du bruit. Les moyennes des résultats obtenus en terme d'erreur quadratique

fonction \ méthode	\widehat{N}_p [MTR]	\widehat{N}_p [Denison 1998]	eqm [MTR]	eqm [Denison 1998]
f_1	9	12	0.0078	0.0097
f_2	11	15	0.0065	0.0087

TABLE 3.1: Erreur quadratique moyenne et nombre de paramètres pour les deux fonctions tests régulières. Ces indicateurs sont donnés pour le modèle à transition régulière [MTR] et pour le modèle polynômial par morceaux présenté dans [Denison 1998]

moyenne et en terme de nombres de paramètres \widehat{N}_p sont reportées dans la table 3.1. Les résultats obtenus avec le modèle à transition régulière (MTR) sont comparés avec ceux obtenus dans [Denison 1998]. Dans cette référence, le modèle utilisé est un modèle polynômial par morceaux, où l'ordre des polynômes est fixé à 2 et des contraintes de continuité de la fonction de régression sont imposées aux niveau des transitions. Le modèle à transition régulière obtient de meilleurs résultats pour les deux fonctions tests, aussi bien en terme de qualité de la reconstruction caractérisée par l'erreur quadratique moyenne notée eqm, qu'en terme de parcimonie de la représentation caractérisée par le nombre de paramètres N_p .

Ces résultats sont donc très satisfaisants quant à la capacité du modèle à transitions régulières à représenter des fonctions régulières. Nous nous intéressons donc maintenant à la modélisation de fonctions non-régulières.

3.6.3 Signaux tests non réguliers

Pour valider le modèle sur des signaux tests non-réguliers nous utilisons quatre fonctions définies dans [Donoho 1994] : « *blocks* », « *bumps* », « *doppler* » et « *heavisine* ». Leur expression analytique avant remise à échelle est :

- Pour la fonction « *blocks* »,

$$f_{\text{blocks}} = \sum_{j=1}^{11} h_j H(t - \tau_j),$$

$$\text{avec } \begin{cases} \tau_j &= \{0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81\}, \\ h_j &= \{4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2\}. \end{cases}$$

- Pour la fonction « *bumps* »,

$$f_{\text{bumps}} = \sum_{j=1}^{11} h'_j \left(1 + \left| \frac{t - \tau_j}{w_j} \right| \right)^{-4},$$

$$\text{avec } \begin{cases} w_j &= \{0.005, 0.005, 0.006, 0.01, 0.01, 0.03, 0.01, 0.01, 0.005, 0.008, 0.005\}, \\ h'_j &= \{4, 5, 3, 4, 5, 4.2, 2.1, 4.3, 3.1, 5.1, 4.2\} \end{cases}$$

- Pour la fonction « *heavisine* »

$$f_{\text{heavisine}} = 4 \sin(4\pi t) - \text{sgn}(t - 0.3) - \text{sgn}(0.72 - t).$$

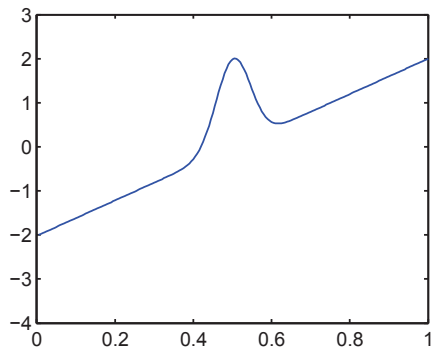
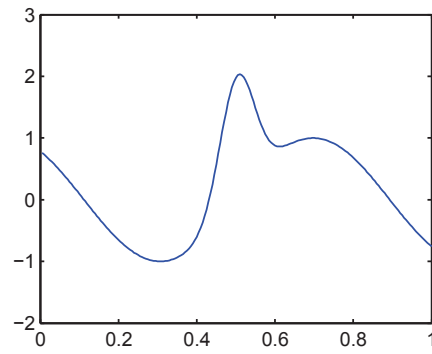
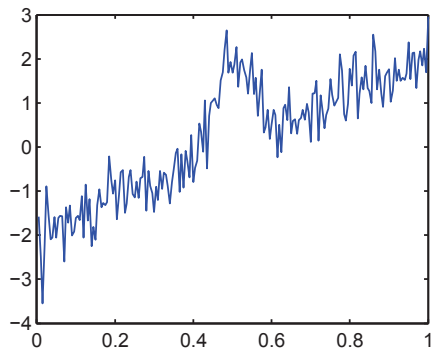
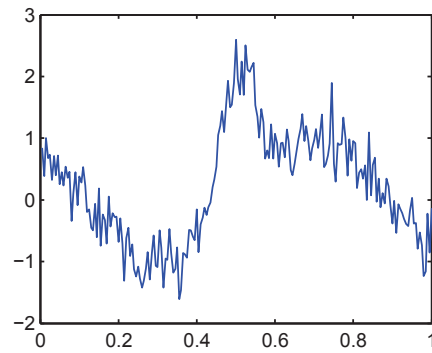
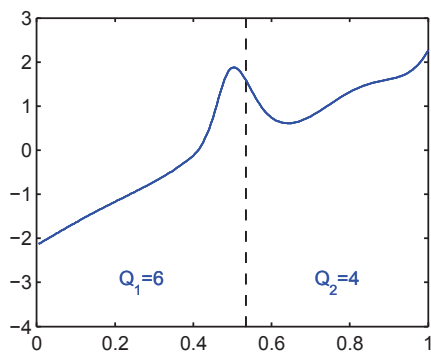
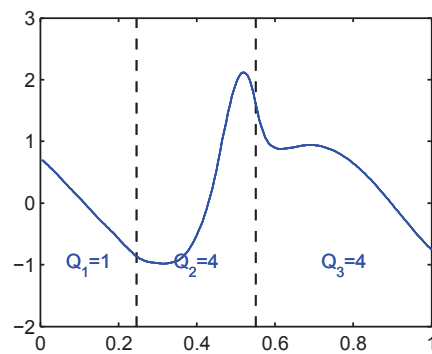
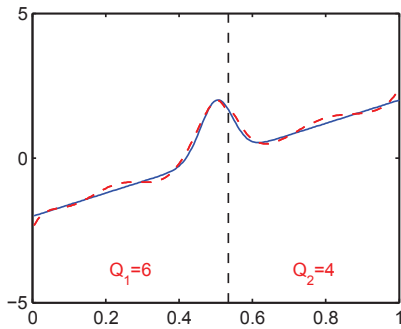
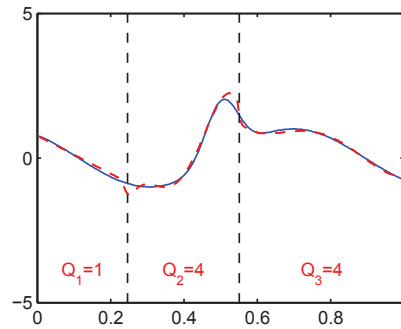
(a) fonction synthétique régulière $f_1(t)$ (b) fonction synthétique régulière $f_2(t)$ (c) fonction $f_1(t)$ bruitée(d) fonction $f_2(t)$ bruitée(e) fonction $\hat{f}_1(t)$ estimée point par point, position des transitions τ (ligne pointillée verticale) et ordre des sous-modèles Q estimés au sens du MAP(f) fonction $\hat{f}_2(t)$ estimée point par point, position des transitions τ (ligne pointillée verticale) et ordre des sous-modèles Q estimés au sens du MAP

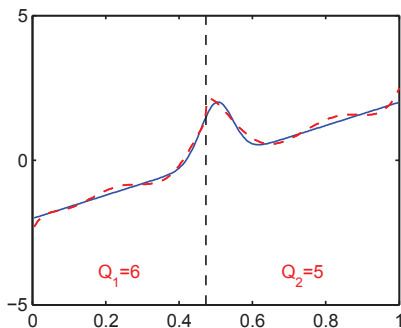
FIGURE 3.4: Estimation de fonctions tests régulières.



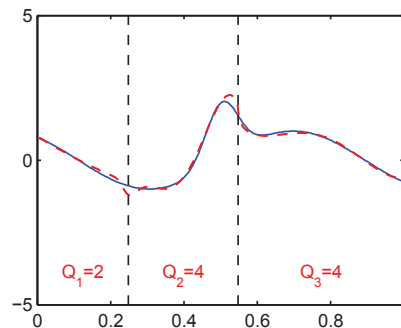
(a) Signal f_1 : fonction réelle (bleu) - fonction estimée au sens du MAP pour des fonctions de transitions sigmoïdales (pointillés rouges). Position des transitions τ (pointillés verticaux) et ordre des sous-modèles \mathbf{Q} . eqm = 1.303



(b) Signal f_2 : fonction réelle (bleu) - fonction estimée au sens du MAP pour des fonctions de transitions sigmoïdales (pointillés rouges). Position des transitions τ (pointillés verticaux) et ordre des sous-modèles \mathbf{Q} . eqm = 1.470



(c) Signal f_1 : fonction réelle (bleu) - fonction estimée au sens du MAP pour des fonctions de transitions échelon (pointillés rouges). Position des transitions τ (pointillés verticaux) et ordre des sous-modèles \mathbf{Q} . eqm = 4.075



(d) Signal f_2 : fonction réelle (bleu) - fonction estimée au sens du MAP pour des fonctions de transitions échelon (pointillés rouges). Position des transitions τ (pointillés verticaux) et ordre des sous-modèles \mathbf{Q} . eqm = 1.704

Comparaison entre le modèle à transition régulière et le modèle de régression par morceaux.

FIGURE 3.5

– Pour la fonction « *doppler* »,

$$f_{\text{doppler}} = (t(1-t))^{\frac{1}{2}} \sin\left(2\pi \frac{1+\epsilon}{t+\epsilon}\right),$$

avec $\epsilon = 0.05$.

Conformément à [Donoho 1994], ces signaux sont échantillonnés avec $n = 2048$ points et subissent un changement d'échelle de manière à conserver un rapport signal à bruit constant égal à 7 pour tous les signaux.

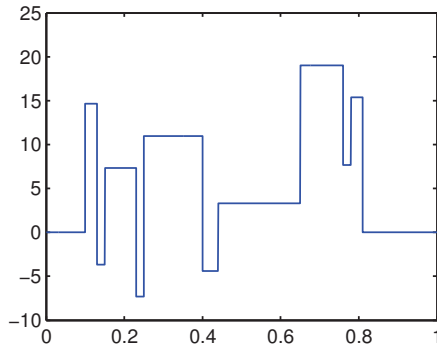
La figure 3.6(a) représente la fonction « *blocks* » après changement d'échelle. La figure 3.6(c) représente la fonction « *blocks* » additionnée d'un bruit blanc gaussien. La figure 3.6(e) représente l'estimation point par point de la fonction « *blocks* », ainsi que les positions des transitions estimées au sens du MAP. Les mêmes tracés sont réalisées respectivement pour la fonction « *bumps* » sur les figures 3.6(b), 3.6(d) et 3.6(f), pour la fonction « *doppler* » sur les figures 3.7(a), 3.7(c) et 3.7(e) et pour la fonction « *heavisine* » sur les figures 3.7(b), 3.7(d) et 3.7(f).

Comme pour les signaux tests réguliers, nous avons répété l'estimation de chacune des fonctions tests sur 30 réalisations différentes du bruit. Les résultats obtenus en terme d'erreur quadratique moyenne, moyennés sur les 30 réalisations sont reportés dans la table 3.2. Pour comparaison, nous avons reporté dans cette table les résultats obtenus sur ces même signaux dans d'autres travaux. Dans [Denison 1998], le modèle de régression est polynômial par morceaux avec un ordre fixé à 2 et des contraintes de continuité au niveau des transitions. Dans [Punskaya 2002], le modèle de régression est polynomial par morceaux également, mais sans contraintes et les ordres des polynômes sont estimés. Dans [Baladan. 2005], des splines cubiques sont utilisées conjointement à un terme de pénalisation variable en fonction du temps, permettant de limiter les phénomènes de sur-apprentissage des données.

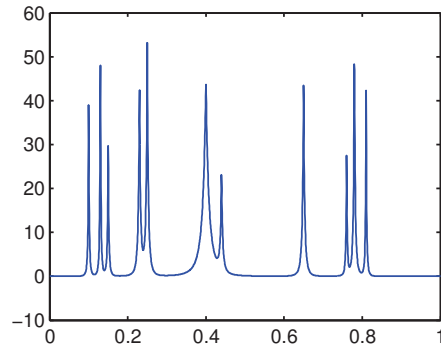
Pour deux des signaux tests, « *blocks* » et « *heavisine* », le modèle à transitions régulières offre de très bonnes performances en atteignant une erreur quadratique plus faible que toutes les autres méthodes. Pour le signal « *bumps* » les performances du modèle sont plus mitigées. Il s'avère que la valeur élevée de l'erreur quadratique obtenue pour ce signal est due au fait que les maxima de la fonction estimée sont légèrement tronqués. Ce phénomène découle directement de la propriété de régularité de la fonction de régression (voir la proposition 3.3.1). Les maxima du signal « *bumps* » sont en effet des points singuliers de la fonction. Or, muni de fonctions de transition sigmoïdales, le modèle ne peut pas prendre en compte de manière exacte de telles singularités. Notons que malgré cette incapacité inhérente à reproduire exactement la fonction « *bumps* », le reste du signal, incluant le voisinage des points singuliers, est bien reconstitué par le modèle à transitions régulières. Finalement, pour le signal « *doppler* », le modèle à transitions régulières atteint des performances dans la moyenne des autres méthodes. Le début du signal en particulier, correspondant aux hautes fréquences, n'est pas du tout reconstitué par le modèle à transitions régulières. Cela est attribuable, non pas au modèle lui-même, mais au cadre bayésien servant à l'estimation : le début du signal est considéré comme du bruit, ce qui est la conséquence des loi *a priori* sur les paramètres du modèle.

3.7 Conclusions

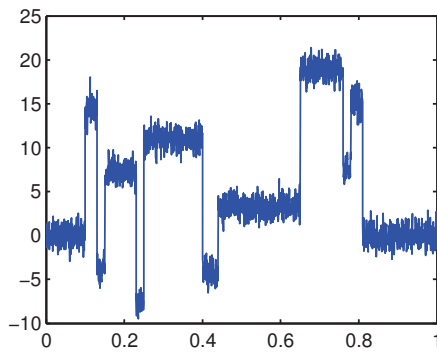
Dans ce chapitre, nous avons défini un modèle de régression à transitions régulières. Ce type de modèle généralise les modèles de régression par morceaux en modélisant les changements de sous-modèles par une fonction de transition paramétrique. Nous avons défini un cadre bayésien



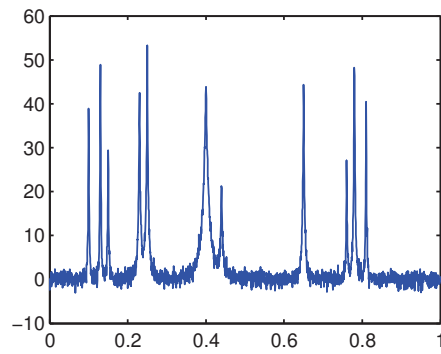
(a) fonction synthétique « blocks »



(b) fonction synthétique « bumps »



(c) fonction « blocks » bruitée



(d) fonction « bumps » bruitée

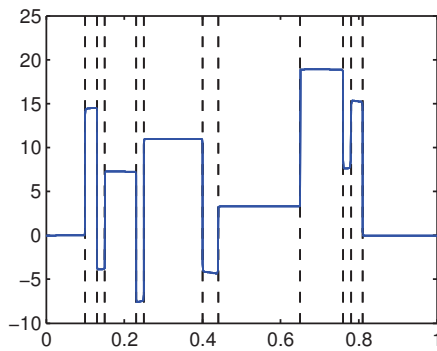
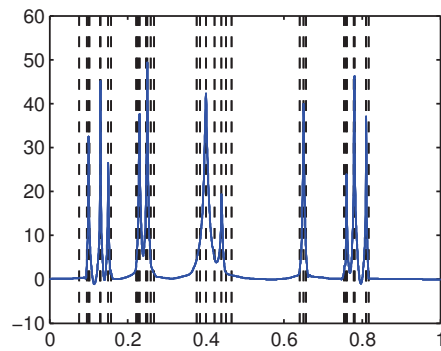
(e) fonction « blocks » estimée point par point, position des transitions τ au sens du MAP (pointillés)(f) fonction « bumps » estimée point par point, position des transitions τ au sens du MAP (pointillés)

FIGURE 3.6: Estimation de fonctions tests non-régulières.

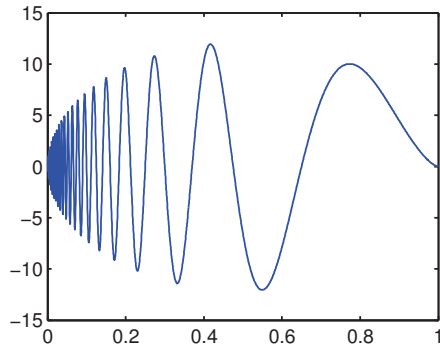
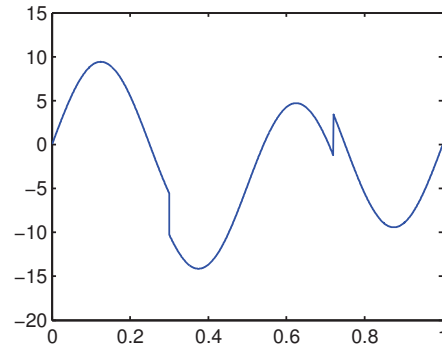
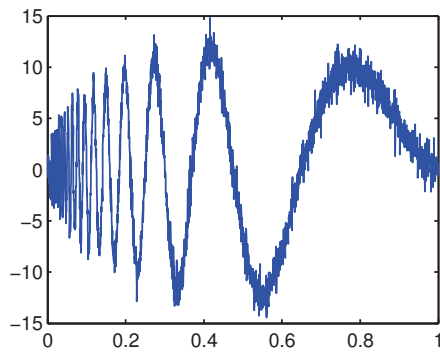
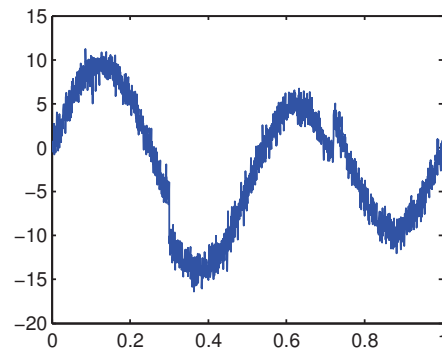
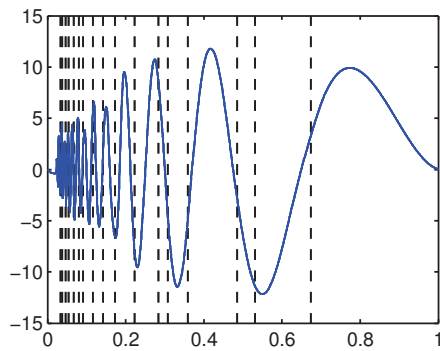
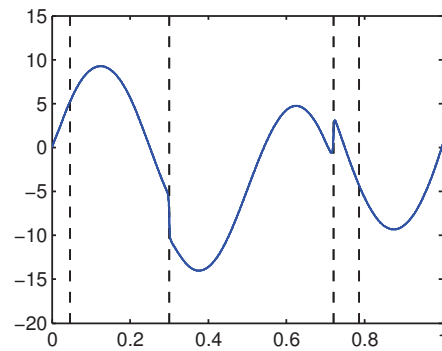
(a) fonction synthétique « *doppler* »(b) fonction synthétique « *heavisine* »(c) fonction « *doppler* » bruitée(d) fonction « *heavisine* » bruitée(e) fonction « *doppler* » estimée point par point, position des transitions τ au sens du MAP (pointillés)(f) fonction « *heavisine* » estimée point par point, position des transitions τ au sens du MAP (pointillés)

FIGURE 3.7: Estimation de fonctions tests non-régulières.

fonction \ méthode	[MTR]	[Denison 1998]	[Punskaya 2002]	[Baladan. 2005]
blocks	0.012	0.17	0.021	0.137
bumps	0.247	0.167	0.121	0.098
doppler	0.119	0.135	0.098	0.024
heavisine	0.014	0.033	0.019	0.028

TABLE 3.2: Erreur quadratique moyenne pour les différentes fonctions tests non-régulières. L'erreur est indiquée pour le modèle à transition régulières [MTR] ainsi que pour différents modèles de référence.

hiérarchique permettant d'estimer conjointement les paramètres de ce modèle et le nombre adéquat de paramètres pour représenter un signal donné. Un algorithme RJMCMC a été proposé pour échantillonner la loi *a posteriori* de ce modèle bayésien. Finalement, nous avons montré que cette méthodologie pouvait se positionner favorablement par rapport à quelques méthodes standard pour des problèmes de lissage de courbe. En effet, le modèle à transitions régulières présente de bonnes propriétés de régularité tout en gardant la possibilité de modéliser de façon satisfaisante les points singuliers d'un signal observé.

Nous avons également pu constater que ce modèle avait de bonnes propriétés de parcimonie, propriété qui nous intéresse le plus pour l'application de modélisation de signaux transitoires électriques. Le prochain chapitre nous conduit donc à exploiter ce modèle pour les signaux de la base de données SISED.

Segmentation des signaux transitoires électro-domestiques

Dans le précédent chapitre, nous avons présenté les outils méthodologiques nécessaires à l'estimation d'un modèle à transitions régulières. Naturellement nous souhaitons utiliser ce modèle pour la représentation des signaux transitoires électriques de la base de données SISED. Dans ce chapitre, nous précisons les sous-modèles et les fonctions de transition adaptées à la modélisation de ces signaux. Nous présentons enfin les résultats obtenus en appliquant ce modèle aux signaux de la base de données.

4.1 Spécification du modèle à transitions régulières pour la modélisation de signaux transitoires électriques

4.1.1 Objectifs

Avant d'exposer les modifications apportées à la méthode présentée au chapitre 3, il est important de préciser dans quel but ces modifications sont apportées et pourquoi nous n'utilisons pas directement le modèle présenté au chapitre précédent. Pour notre modèle de régression, toute la difficulté liée à la classification repose sur l'extraction de paramètres pertinents afin de caractériser convenablement chaque classe de signaux. Utiliser une approche trop générique peut ne pas être approprié. Ainsi, les approches étudiées dans le chapitre précédent sont destinées à obtenir une bonne approximation de la fonction de régression. Le risque est alors de sur-ajuster et donc de sur-apprendre les observations en modélisant, pour une classe donnée, des détails trop peu ou pas significatifs. Dans un contexte de classification, l'objectif doit se limiter à l'extraction des paramètres pertinents.

C'est pourquoi, d'une part, nous choisissons avec précaution les deux composantes essentielles du modèle à transitions régulières (la famille des sous-modèles et la famille des fonctions de régression) afin de capturer les caractéristiques essentielles des signaux observés. D'autre part, nous apportons une modification à l'estimation du paramètre δ^2 introduit au paragraphe 4.1.4 qui modélise le RSB *a priori* des signaux. Cette modification vise à améliorer la parcimonie de la représentation de nos signaux, c'est-à-dire à favoriser des représentations qui font intervenir peu de paramètres, quitte à dégrader légèrement la qualité de l'ajustement des données observées. Cette parcimonie vise à éviter le sur-ajustement, i.e., à garantir que les paramètres estimés seront les plus significatifs de notre modèle. Dans le cas contraire, i.e., si un trop grand nombre de paramètres devait représenter chaque signal, alors une large partie des paramètres estimés ne serait probablement pas caractéristiques de la classe considérée ; ils caractériseraient uniquement une réalisation bien particulière d'un signal de cette classe. Cet écueil correspond au phénomène de "sur-apprentissage" des données.

4.1.2 Sous-modèles

Dans le cadre de la modélisation des signaux transitoires de la base de données SISED, nous choisissons d'utiliser des sous-modèles constants :

$$m_k(t) = \beta_k, \quad \forall t \in [\tau_0, \tau_{K+1}]. \quad (4.1)$$

où β_k est la puissance active associée au segment $[\tau_{k-1}, \tau_k]$.

Ce choix s'explique par la présence de nombreuses plages temporelles où la puissance active est constante dans nos signaux. Ces zones « paliers » correspondent aux régimes permanents mais également à des phases intermédiaires comme ceux apparaissant sur les signaux de réfrigérateur micro-ondes.

Notons que, contrairement au cadre plus général exposé au chapitre 3, ces sous-modèles sont, par nature, d'ordre fixe. De fait, les ordres des sous-modèles Q n'auront pas à être estimés dans la suite de ce chapitre. L'intérêt de ce choix *a priori* est d'exploiter au mieux les propriétés physiques de ces signaux transitoires, et donc de se concentrer uniquement sur les paramètres significatifs du modèle.

4.1.3 Paramétrisation des transitions : la fonction exponentielle étendue

Introduction

Le choix d'une fonction de transition adaptée à notre application de modélisation de signaux transitoires est très important. En effet, choisir une fonction de transition représentant mal nos signaux conduirait à une sur-segmentation des signaux, c'est-à-dire à l'accumulation de plusieurs fonctions de transitions pour représenter une seule transition du signal observé. De plus une fonction de transition appropriée nous permettrait d'obtenir une segmentation plus représentative des phénomènes physiques sous-jacents.

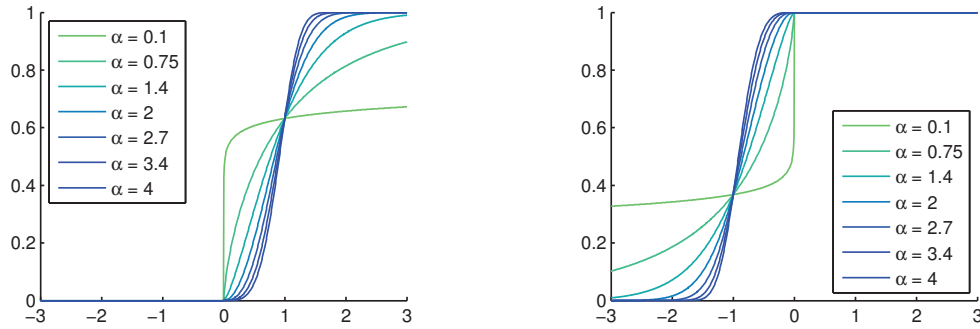
Comme nous l'avons vu, dans la littérature scientifique, les fonctions de transitions sont couramment choisies parmi la famille des fonctions sigmoïdes. Ce type de fonctions pourrait également être approprié à la modélisation de certains de nos signaux transitoires, en particulier ceux générés par les réfrigérateurs voir Fig. 2.9. Cependant, cette adéquation ne se généralise pas à tous les signaux de la base de données SISED. Les signaux d'aspirateur, Fig. 2.5, présentent plutôt des décroissances qui semblent exponentielles.

La fonction exponentielle est par ailleurs une candidate tout à fait valable à la représentation de nos fonctions de transition pour les transitoires électriques :

$$\pi_\lambda(t) = \begin{cases} 1 - \exp(-\frac{t}{\lambda}) & \text{si } t > 0, \\ 0 & \text{sinon.} \end{cases} \quad (4.2)$$

En dehors de coller visuellement à bon nombre de nos signaux, les relaxations exponentielles apparaissent fréquemment en tant que solution théorique de circuits électriques simples (circuits de type RLC, résistance, inductance, condensateur).

Cependant, deux raisons nous poussent à rechercher une famille de fonctions plus appropriée. D'une part, ce type de fonction asymétrique ne convient à la représentation de transitions comme celles des signaux de réfrigérateur. D'autre part, nos premières expérimentations ont montré que les relaxations exponentielles simples ne suffisent pas à représenter de façon satisfaisante les signaux transitoires d'aspirateur.



(a) Fonction exponentielle étendue, $\lambda = 1$, influence du paramètre de forme α .

(b) Fonction exponentielle étendue complémentaire, $\lambda = 1$, influence du paramètre de forme α .

FIGURE 4.1: Fonction de transition pour la modélisation des signaux transitoires de la base de données SISED.

Définition

Nous nous tournons donc vers une forme généralisée de la fonction exponentielle :

$$\pi_{\lambda,\alpha}(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^\alpha\right] & \text{si } t > 0 \end{cases} \quad (4.3)$$

Cette famille de fonction dépend donc de 2 paramètres : un paramètre d'échelle λ et un paramètre de forme α . La figure 4.1(a) montre le tracé de cette fonction pour $\lambda = 1$ et différentes valeur de α . On peut constater que cette famille de fonction permet de représenter une large variétés de formes : pour $\alpha = 1$, on retrouve la fonction exponentielle définie précédemment (4.2), pour $\alpha = 2$ on obtient une forme relativement proche de la fonction sigmoïdale (3.7).

Motivations et interprétation

Il existe plusieurs arguments théoriques renforçant ce choix. Ce type de fonction a été utilisé la première fois dans [Kohlrausch 1854] pour modéliser la décharge d'un condensateur pour une problématique qui n'était donc pas si éloignée de la notre. Par la suite, ce type de fonction a largement été utilisé pour la modélisation de différents phénomènes physiques de relaxation [Phillips 1996].

Notons que cette fonction peut être interprétée comme une somme intégrale de relaxation exponentielles « simples » :

$$e^{-t^\alpha} = \int_0^{+\infty} f_\alpha(u) e^{-\frac{t}{u}} du, \quad (4.4)$$

où $f_\alpha(u)$ est une distribution sur les temps de relaxation u qui peut être calculée pour tout $0 < \alpha < 1$ [Pollard 1946, Anderssen 2004]. Cette dernière interprétation est intéressante dans le sens où les relaxations exponentielles simples constituent la réponse de certains circuits de base des systèmes électriques : charge et décharge d'un condensateur dans une résistance (circuit RC), charge et décharge d'une bobine dans une résistance (circuit RL). La fonction exponentielle généralisée s'interpréterait donc comme la réponse d'un système électrique complexe constitué de la combinaison de plusieurs système simples.

Cette interprétation physique reste cependant sujette à caution car elle n'est pas valable pour $\alpha > 1$. De manière plus générale, notre objectif n'est pas de modéliser le fonctionnement physique des systèmes électriques étudiés, mais juste de définir une fonction de transition la plus adaptée.

Fonction de transition complémentaire

La fonction exponentielle généralisée permet de représenter la plupart des formes observées sur les signaux de la base de données SISED. Cependant, elle ne permet pas directement de représenter les formes observées sur certains signaux de micro-ondes par exemple sur la figure 2.8. Le premier appel de courant se caractérise par une décroissance exponentielle mais le deuxième serait plutôt une croissance exponentielle jusqu'à une valeur de saturation soit atteinte. Pour représenter ce type de transition, nous proposons d'utiliser la fonction complémentaire de la fonction exponentielle généralisée (4.5). La fonction exponentielle généralisée complémentaire s'écrit :

$$\pi_{\lambda,\alpha}(t) = \begin{cases} \exp\left[-\left(\frac{t}{\lambda}\right)^\alpha\right] & \text{si } t < 0, \\ 1 & \text{si } t > 0. \end{cases} \quad (4.5)$$

La figure 4.1(b) montre le tracé de cette fonction pour $\lambda = 1$ et différentes valeurs de α .

Pour pouvoir utiliser simultanément les deux versions de la fonction exponentielle généralisée définies précédemment, nous définissons un paramètre binaire $c \in \{0, 1\}$ et la fonction de transition $\pi_{\lambda,\alpha,c}^{(0)}(t)$ définie par :

$$\pi_{\lambda,\alpha,c}^{(0)}(t) = \begin{cases} 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^\alpha\right] & \text{si } t > 0, c = 0, \\ \exp\left[-\left(\frac{-t}{\lambda}\right)^\alpha\right] & \text{si } t < 0, c = 1, \\ 0 & \text{si } t < 0, c = 0, \\ 1 & \text{si } t > 0, c = 1. \end{cases} \quad (4.6)$$

Cette fonction de transition est celle qui sera utilisée par la suite dans cette thèse pour la modélisation des signaux transitoires électriques de la base de données SISED.

Propriétés

Nous examinons ci-après quelques-unes des propriétés de la fonction $\pi_{\lambda,\alpha,c}^{(0)}(t)$ (4.6). Ces propriétés nous seront utiles par la suite pour ajuster les paramètres de cette fonction.

Proposition 4.1.1. *L'argument du maximum du gradient de la fonction de transition $T_{\pi^{(0)}}(\alpha, \lambda)$ est donné par*

$$T_{\pi^{(0)}}(\alpha, \lambda) = \underset{t}{\operatorname{argmax}} \frac{d\pi_{\lambda,\alpha,c}^{(0)}}{dt}(t) = \begin{cases} 0 & \text{si } \alpha < 1, \\ \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 0, \\ -\lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 1. \end{cases} \quad (4.7)$$

Dans le cas où $\alpha > 1$, ce maximum correspond à un point d'inflexion de la fonction.

Démonstration. cf annexe A.3. ■

Proposition 4.1.2. *Soit $A_{\pi^{(0)}}(\lambda, \alpha)$, la somme de l'aire sous la courbe avant l'instant $t = T_{\pi^{(0)}}(\alpha, \lambda)$ et l'aire au dessus de la courbe après l'instant $t = T_{\pi^{(0)}}(\alpha, \lambda)$, définie par :*

$$A_{\pi^{(0)}}(\lambda, \alpha) = \int_{-\infty}^{T_{\pi^{(0)}}(\alpha,\lambda)} \left[\pi_{\lambda,\alpha,c}^{(0)}(t)\right] dt + \int_{T_{\pi^{(0)}}(\alpha,\lambda)}^{+\infty} \left[1 - \pi_{\lambda,\alpha,c}^{(0)}(t)\right] dt. \quad (4.8)$$

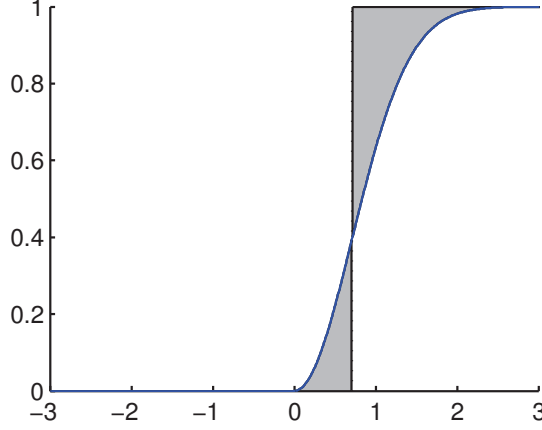


FIGURE 4.2: Fonction de transition $\pi_{\lambda, \alpha, c}^{(0)}(t)$ pour $\lambda = 1$, $\alpha = 2$ et $c = 0$, la surface grisée représente l'aire $A_{\pi^{(0)}}(\lambda, \alpha)$.

L'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ est proportionnelle à λ et vaut

$$A_{\pi^{(0)}}(\lambda, \alpha) = \begin{cases} \lambda \left[\frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) \right] & \text{si } \alpha < 0 \\ \frac{\lambda}{\alpha} \left[\frac{1}{\alpha} \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} + \Gamma\left(\frac{1}{\alpha}\right) - 2\gamma\left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha}\right) \right] & \text{si } \alpha > 0, \end{cases} \quad (4.9)$$

où Γ est la fonction gamma telle que $\Gamma : \begin{cases} \mathbb{R}_+ & \longrightarrow & \mathbb{R}_+ \\ (z) & \longmapsto & \int_0^{+\infty} t^{z-1} e^{-t} dt \end{cases}$ et γ est la fonction gamma

incomplète telle que $\gamma : \begin{cases} \mathbb{R}_+ \times \mathbb{R}_+ & \longrightarrow & \mathbb{R}_+ \\ (z, x) & \longmapsto & \int_0^x t^{z-1} e^{-t} dt \end{cases}$.

On note $L_{\pi^{(0)}}(\alpha) = \frac{1}{\lambda} A_{\pi^{(0)}}(\lambda, \alpha)$ le facteur de proportionnalité entre $A_{\pi^{(0)}}(\lambda, \alpha)$ et λ , qui ne dépend donc que de α .

Démonstration. cf annexe A.4. ■

La figure 4.2 permet de visualiser l'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ pour $\lambda = 1$, $\alpha = 2$ et $c = 0$.

Reparamétrisation de la fonction exponentielle étendue

Pour faciliter son utilisation dans nos algorithmes nous allons procéder à trois changements de paramètres par rapport à la fonction $\pi_{\lambda, \alpha, c}^{(0)}$. L'objectif de ces changements de paramètres est de pouvoir relier directement les paramètres de la fonction à des caractéristiques géométriques de sa courbe, ainsi que de mieux découpler les relations entre ces paramètres.

D'une part, nous souhaitons que la position des fonctions de transitions τ_k coïncide avec les maxima du gradient des poids $p_k(t)$, c'est-à-dire que le maximum du gradient des fonctions de transitions soit atteint pour $t = 0$. La raison est à la fois de faciliter l'interprétation des paramètres τ_k et de nous fournir un moyen simple pour obtenir des valeurs judicieuses de τ_k , ce qui nous servira dans la suite de ces travaux.

Nous exprimons donc la fonction de transition $\pi_{\lambda, \alpha, c}^{(1)}$ définie par :

$$\pi_{\lambda, \alpha, c}^{(1)}(t) = \pi_{\lambda, \alpha, c}^{(0)}(t - T_{\pi^{(0)}}(\alpha, \lambda)), \quad (4.10)$$

qui vérifie bien la propriété $T_{\pi^{(1)}}(\alpha, \lambda) = 0$.

De la même façon, nous souhaitons faire coïncider le paramètre d'échelle λ avec l'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ définie par (4.8). Cette propriété est à rapprocher de la propriété du temps caractéristique d'une décroissance exponentielle simple qui correspond à l'aire sous la courbe. Dans notre cas, on souhaite pouvoir relier le paramètre d'échelle λ à la durée d'une transition, c'est-à-dire le temps nécessaire pour passer d'un palier de puissance constante à un autre. Or on se rend compte facilement à l'observation de la figure 4.1 que, malgré le fait que le paramètre λ soit constant, certaines courbes croissent beaucoup plus rapidement que d'autres de 0 à 1.

On définit donc la fonction de transition $\pi_{\lambda, \alpha, c}^{(2)}(t)$ telle que :

$$\pi_{\lambda, \alpha, c}^{(2)}(t) = \pi_{\frac{\lambda}{L_{\pi^{(0)}}(\alpha)}, \alpha, c}^{(1)}(t), \quad (4.11)$$

qui vérifie bien $A_{\pi^{(2)}}(\lambda, \alpha) = \lambda$.

Enfin, l'un des inconvénients de la paramétrisation initiale de la fonction exponentielle généralisée est que la forme de la fonction n'évolue pas régulièrement avec α . En effet pour $0 < \alpha < 1$ la fonction évolue très vite de la fonction échelon vers une relaxation exponentielle. Puis pour $\alpha > 1$ la fonction converge très rapidement vers une fonction échelon localisée en $t = \lambda$ qui correspond au cas limite $\alpha \rightarrow +\infty$. Ce comportement n'est pas forcément gênant, cependant, pour faciliter la manipulation de cette fonction de transition nous proposons d'utiliser la fonction $\pi_{\lambda, \alpha, \tau, c}^{(3)}$ telle que :

$$\pi_{\lambda, \alpha, c}^{(3)} = \pi_{\lambda, \ln(\alpha), c}^{(2)}. \quad (4.12)$$

Ce changement de paramètre $\alpha \rightarrow \ln(\alpha)$ fournit ainsi une évolution de la forme de la fonction de transition plus régulière en fonction de α . Par la suite dans ce manuscrit, nous désignerons $\pi_{\lambda, \alpha, c}^{(3)}(t)$ par $\pi_{\lambda, \alpha, c}(t) = \pi_{\eta}(t)$, avec $\eta = \{\lambda, \alpha, c\}$, afin de simplifier les notations.

Fonction de transition reparamétrée

Finalement, la forme finale de la fonction de transition après reparamétrisation est :

$$\pi_{\eta}(t) = \begin{cases} 1 - \exp \left[- \left(\frac{t - T_{\pi^{(0)}}(\alpha, \lambda)}{\lambda / L_{\pi^{(0)}}(\alpha)} \right)^{\ln(\alpha)} \right] & \text{si } t > T_{\pi^{(0)}}(\alpha, \lambda), c = 0 \\ \exp \left[- \left(\frac{-t + T_{\pi^{(0)}}(\alpha, \lambda)}{\lambda / L_{\pi^{(0)}}(\alpha)} \right)^{\ln(\alpha)} \right] & \text{si } t < T_{\pi^{(0)}}(\alpha, \lambda), c = 1 \\ 0 & \text{si } t < T_{\pi^{(0)}}(\alpha, \lambda), c = 0, \\ 1 & \text{si } t > T_{\pi^{(0)}}(\alpha, \lambda), c = 1, \end{cases} \quad (4.13)$$

avec $\eta = \{\lambda, \alpha, c\}$ et :

$$T_{\pi^{(0)}}(\alpha, \lambda) = \operatorname{argmax}_t \frac{d\pi_{\lambda, \alpha, c}^{(0)}}{dt}(t) = \begin{cases} 0 & \text{si } \alpha < 1 \\ \lambda \left(\frac{\alpha-1}{\alpha} \right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 0 \\ -\lambda \left(\frac{\alpha-1}{\alpha} \right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 1 \end{cases} \quad (4.14)$$

et :

$$L_{\pi^{(0)}}(\alpha) = \begin{cases} \frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) & \text{si } \alpha < 0 \\ \frac{1}{\alpha} \left[\frac{1}{\alpha} \left(\frac{\alpha-1}{\alpha} \right)^{\frac{1}{\alpha}} + \Gamma\left(\frac{1}{\alpha}\right) - 2\gamma \left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha} \right) \right] & \text{si } \alpha > 0 \end{cases}. \quad (4.15)$$

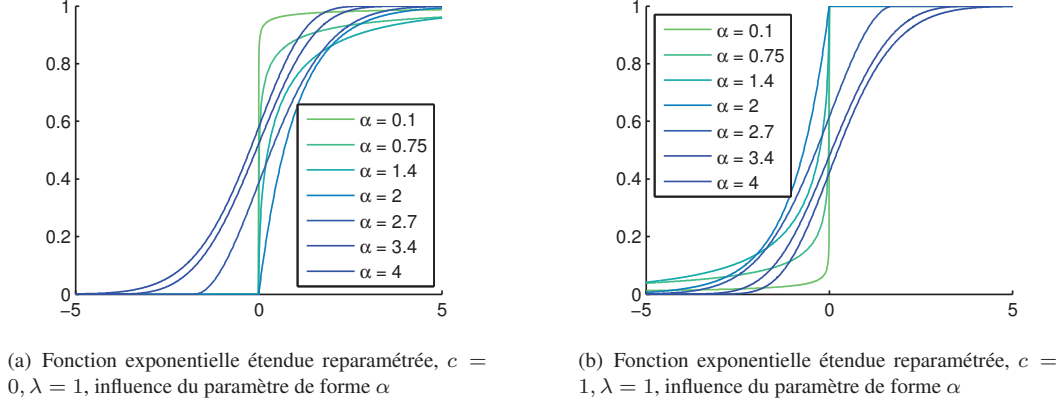


FIGURE 4.3: Fonction de transition pour la modélisation des signaux transitoires de la base de données SISED.

La figure 4.3 présente l'évolution de la fonction de transition $\pi_{\lambda, \alpha, \tau, c}^{(3)}(t)$ pour $\lambda = 0$ et $c = 0$ (figure 4.3(a)) et $c = 1$ (figure 4.3(b)). Chaque tracé correspond à une valeur différente de α . On peut constater que les contraintes imposées (point d'inflexion en $t = 0$ et aire sous la courbe constante pour λ constant) sont bien respectés pour la fonction de transition ainsi définie.

4.1.4 Modèle bayésien hiérarchique pour l'estimation des signaux transitoires

Le cadre bayésien hiérarchique exprimé de manière générale dans la partie 3.4 s'applique quasiment directement à la modélisation des signaux transitoires de la base de données SISED. Évidemment, comme l'ordre des sous-modèles n'a pas à être estimé ici, la loi *a priori* de \mathbf{Q} disparaît du problème. D'autre part, les fonctions de transition étant ici plus complexes que de simples sigmoïdes, il convient de définir $p(\eta) = p(\lambda, \alpha, c)$ la loi *a priori* des paramètres des fonctions de transitions. Nous choisissons une densité *a priori* indépendante pour chaque paramètre :

$$p(\eta) = p(\lambda)p(\alpha)p(c). \quad (4.16)$$

Chacune de ces lois *a priori* est choisie suffisamment vague, reflétant ainsi le fait que nous ne disposons pas d'information *a priori* sur la distribution de ces paramètres.

Loi *a priori* de λ

La loi *a priori* du paramètre d'échelle des fonctions exponentielles étendues est choisie identique à celle des paramètres d'échelle des fonctions de transitions sigmoïdales, c-à-d :

$$\lambda_k | (\nu_k, \rho_k) \sim \mathcal{IG}(\nu_k, \rho_k) \quad \forall k \in [1, K], \quad (4.17)$$

où l'hyperparamètre de forme est fixé $\nu_k = 1$ et l'hyperparamètre d'échelle $\rho_k \in \mathbb{R}_+$ suit une loi *a priori* telle que :

$$\rho_k \sim \mathcal{G}(\chi, \psi) \quad \forall k \in [1, K]. \quad (4.18)$$

Avec χ et ψ choisis de manière à ce que la loi *a priori* soit vague ($\chi = 1$ et $\psi = 100$). On peut marginaliser l'hyperparamètre ρ_k pour en déduire la loi marginale *a priori* de λ_k :

$$p(\lambda_k) = \frac{\Gamma(\nu_k + \chi)}{\Gamma(\nu_k)\Gamma(\chi)} \frac{\psi^{\nu_k + \chi}}{\chi^\psi} \frac{\lambda_k^{\chi-1}}{(\lambda_k + \psi)^{\nu_k + \chi}} \mathbb{I}_{\mathbb{R}^+}(\lambda_k) \quad \forall k \in [1, K]. \quad (4.19)$$

Loi *a priori* de α

La loi *a priori* du paramètre de forme α est choisie normale et centrée :

$$\alpha \sim \mathcal{N}(0, \sigma_\alpha^2), \quad (4.20)$$

où σ_α^2 est distribué selon une loi conjuguée inverse-gamma suffisamment vague :

$$\sigma_\alpha^2 \sim \mathcal{IG}(\chi, \psi). \quad (4.21)$$

On peut intégrer l'hyperparamètre σ_α^2 pour obtenir la densité marginale *a priori* de α :

$$p(\alpha) = \frac{\Gamma(\chi + \frac{1}{2})\psi^\chi}{\Gamma(\chi)\sqrt{2\pi}(\frac{\alpha^2}{2} + \psi)^{\chi + \frac{1}{2}}}. \quad (4.22)$$

Loi *a priori* de c

La loi *a priori* du paramètre étiquette c est choisie uniforme, telle que

$$p(c) = \frac{1}{2} \mathcal{U}_{\{0,1\}}(c). \quad (4.23)$$

Estimation de δ^2

Afin de permettre au lecteur de mieux appréhender les modifications apportées à l'estimation de δ^2 , nous présentons ici un résultat intermédiaire, obtenu sur un signal transitoire de lampe halogène en estimant δ^2 tel qu'exposé au chapitre 3. La figure 4.4 montre la puissance active d'un signal transitoire de lampe halogène et son estimation au sens du MAP ainsi que la position estimée au sens du MAP des transitions.

Le nombre de transitions estimé est très élevé ($\hat{K}_{\text{map}} = 14$). En effet chaque détail de la puissance active est pris en compte par le modèle par une transition. Bien que le signal observé soit correctement estimé par le modèle, ce résultat ne répond pas de manière satisfaisante à notre problème. Nous souhaitons obtenir une représentation beaucoup plus parcimonieuse des signaux en ne caractérisant que les variations les plus significatives de la puissance. Dans le cas présent, caractériser l'appel initial de puissance (entre 0.1 – 0.15s) et la variation finale de la puissance vers le régime permanent (entre 0.35 – 0.4s).

Cette sur-segmentation s'explique par le fait qu'il existe un écart entre le modèle que l'on souhaite obtenir et le signal observé, qui se matérialise par un bruit de modélisation non stationnaire. Le modèle de régression s'adapte donc à ces écarts en augmentant le nombre de transitions. Ce phénomène est exacerbé par le fait que, par ailleurs, le signal observé est très peu bruité (voir par exemple pendant le régime permanent, pour $t > 0.4$ sur la Fig. 4.4).

Pour résoudre ce problème, il existe plusieurs possibilités. Une solution pourrait être d'abandonner l'hypothèse d'un bruit stationnaire pour choisir une hypothèse plus réaliste. Par exemple, en utilisant un modèle de régression hétéroscédastique, ou encore en modélisant le bruit par un

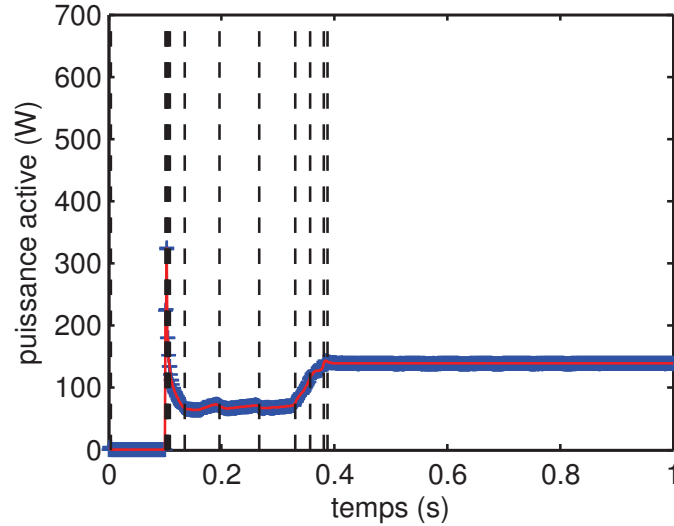


FIGURE 4.4: Modélisation d'un transitoire de lampe halogène : signal réel (bleu) signal estimé MAP (rouge) position estimées MAP des transitions (pointillés)

processus gaussien [Rasmussen 2006] ou un autre processus non-stationnaire. Cependant, ce serait sacrifier la simplicité de notre hypothèse, ce qui ne nous semble pas souhaitable dans le cadre de l'estimation d'un modèle déjà complexe.

Pour obtenir une représentation parcimonieuse des signaux observés, nous proposons de fixer une valeur *a priori* au paramètre δ^2 , plutôt que de l'estimer. En effet comme établi par la proposition 3.4.2, ce paramètre est homogène au rapport signal à bruit *a priori* entre la fonction de régression $m(t)$ et le bruit résiduel $\epsilon(t)$. Par conséquent, fixer ce paramètre à une valeur déterministe permet de déterminer la précision de la reconstruction recherchée. Lorsque δ^2 est estimé pour les signaux de la base de données SISED, le paramètre tend à prendre des valeurs très importantes ($\delta^2 = 4.4e+04$ soit environ 46dB pour la simulation présentée sur la figure 4.4). Cela concorde avec le fait que les signaux observés sont très peu bruités. En fixant maintenant δ^2 à une valeur inférieure à la puissance des variations peu significatives, nous forçons le modèle à sur-évaluer la variance du bruit et donc à incorporer ces variations parasites représentées sur la figure 4.4 dans le bruit résiduel.

4.1.5 Algorithme RJMCMC pour l'estimation des paramètres

L'algorithme RJMCMC présenté dans le cas général au chapitre 3.5 ne nécessite que quelques ajustements mineurs pour être appliqué aux signaux de la base de données SISED en tenant compte des modifications apportées au cadre bayésien.

Les mouvements du noyau de proposition sont les mêmes, à l'exception du mouvement de mise à jour des ordres des sous-modèles qui n'est plus nécessaire dans ce cas. Pour le reste, seules les lois de proposition d'une rupture (i.e., sa position τ et les paramètres η de la transition) doivent être redéfinies pour s'adapter aux fonctions de transition définies par (4.13). Ces nouvelles lois de propositions sont maintenant notées $q_0^*(\tau, \eta)$ (proposition d'une nouvelle rupture) et $q_1^*(\tau, \eta)$ (modification d'une rupture).

Loi de proposition $q_0^*(\tau, \eta)$

La loi de proposition $q_0^*(\tau, \eta)$ utilisée pour les mouvements de naissance/mort et de fusion/division des ruptures, est définie par :

$$q_0^*(\tau, \eta) = q_0^*(\tau)q_0^*(\lambda|\tau)q_0^*(\alpha)q_0^*(c), \quad (4.24)$$

c'est à dire que chaque paramètre est proposé de manière indépendante aux autres à l'exception de λ qui dépend de τ . La loi de proposition de τ reste inchangée : $q_0^*(\tau) = q_0(\tau)$ définie à l'équation (3.41).

Proposition du paramètre d'échelle λ Pour construire une loi de proposition pour le paramètre λ , nous nous basons sur la propriété que l'aire A_π , définie à la proposition 4.1.2, est égale à λ pour la fonction de transition reparamétrée (4.13). Supposons que $\tau \sim q_0^*(\tau)$ soit la k -ième transition du modèle, soit $\tau = \tau_k$. Soit

$$\tau^{(-)} = \frac{\tau_k + \tau_{k-1}}{2}, \quad \text{et} \quad \tau^{(+)} = \frac{\tau_k + \tau_{k+1}}{2},$$

respectivement le centre du segment $[\tau_{k-1}, \tau_k]$ et le centre de $[\tau_k, \tau_{k+1}]$. On normalise $x(t)$ entre ces deux points de sorte à produire une estimation de la fonction de transition :

$$\hat{\pi}_k(t) = \frac{x(t) - x(\tau^{(-)})}{x(\tau^{(+)}) - x(\tau^{(-)})} \in [0, 1], \quad (4.25)$$

pour tout $t \in [\tau^{(-)}, \tau^{(+)})$, où $\hat{\pi}_k(0) = 0$ et $\hat{\pi}_k(1) = 1$. On estime alors λ en approximant l'aire définie en 4.1.2 :

$$\hat{\lambda} = \sum_{t \in [\tau^{(-)}, \tau_k]} \hat{\pi}_k(t) + \sum_{t \in [\tau_k, \tau^{(+)})} 1 - \hat{\pi}_k(t). \quad (4.26)$$

Finalement on propose λ selon une loi gamma de variance $\sigma_\lambda^2 = 0.1$ et de moyenne $\hat{\lambda}$:

$$\lambda \sim \mathcal{G} \left(\frac{\hat{\lambda}^2}{\sigma_\lambda^2}, \frac{\sigma_\lambda^2}{\hat{\lambda}} \right). \quad (4.27)$$

Proposition du paramètre de forme α Le paramètre de forme α est tiré selon une loi normale :

$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad (4.28)$$

de moyenne $\mu_\alpha = 0$ et de variance $\sigma_\alpha^2 = 1$.

Proposition de l'étiquette c L'étiquette c est tirée uniformément sur $\{0, 1\}$ selon sa loi *a priori* :

$$q_0^*(c) = p(c). \quad (4.29)$$

Loi de proposition $q_1^*(\tau, \eta)$

A l'instar de $q_1(\tau, \eta)$ définie en (3.50), $q_1^*(\tau, \eta)$ est définie comme une loi de mélange entre la loi de proposition $q_0^*(\tau, \eta)$ et une marche aléatoire gaussienne autour des valeurs courantes des

paramètres τ, η . Les équations présentées au paragraphe 3.5.5 s'appliquent donc directement en remplaçant $q_0(\tau, \eta)$ par $q_0^*(\tau, \eta)$:

$$q_1^*(\tau, \eta) = q_1^*(\tau)q_1^*(\lambda)q_1^*(\alpha)q_1^*(c),$$

avec :

$$\begin{aligned} q_1^*(\tau) &= wq_0^*(\tau) + \frac{1-w}{\sqrt{2\pi\sigma_\tau^2}} e^{-\frac{1}{2\sigma_\tau^2}(\tau-\tau^{(i-1)})^2}, \\ q_1^*(\lambda) &= wq_0^*(\lambda) + \frac{1-w}{\sqrt{2\pi\sigma_\lambda^2}} e^{-\frac{1}{2\sigma_\lambda^2}(\lambda-\lambda^{(i-1)})^2} \mathbb{I}_{\mathbb{R}_+}(\lambda), \\ q_1^*(\alpha) &= wq_0^*(\alpha) + \frac{1-w}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{1}{2\sigma_\alpha^2}(\alpha-\alpha^{(i-1)})^2}, \end{aligned}$$

où $w = 0.2$ et $\sigma_\tau^2 = \sigma_\lambda^2 = \sigma_\alpha^2 = 0.1$. Enfin on définit la loi de proposition du paramètre c :

$$q_1^*(c) = wq_0^*(c) + (1-w)\mathbb{I}_{c=c^{(i-1)}}(c).$$

4.2 Résultats de modélisation des signaux transitoires de la base de données SISED

L'algorithme défini au chapitre 3, complété des modifications présentées dans le présent chapitre, est utilisé pour estimer les paramètres du modèle à transitions régulières pour chacun des signaux transitoires de la base de données SISED. Le paramètre δ^2 est fixé à $15dB$ soit $\delta^2 = 31.6$ environ. Nous simulons $N_{\text{iter}} = 1e4$ vecteurs des paramètres $(\theta^{(i)})_{i=1, \dots, N_{\text{iter}}}$. Les $N_{\text{chauffe}} = 2e3$ premiers vecteurs, correspondant à la période de chauffe, ne sont pas utilisés pour l'analyse. Nous estimons alors les paramètres du modèle au sens du maximum *a posteriori* marginal (MMAP). C'est-à-dire que nous calculons dans un premier temps le mode du nombre de sous-modèles \widehat{K} , en prenant la valeur la plus fréquente de K parmi les échantillons générés $(\theta^{(i)})_{i=1, \dots, N_{\text{iter}}}$. Puis, dans un second temps, on estime les paramètres du modèle $\widehat{\theta}$ au sens du MMAP :

$$\widehat{\theta} = \theta^{(\widehat{i})}, \tag{4.30}$$

avec \widehat{i} défini par :

$$\widehat{i} = \begin{cases} \arg \max & f(\theta^{(i)} | \mathbf{x}). \\ i = N_{\text{chauffe}} + 1, \dots, N_{\text{iter}} \\ K^{(i)} = \widehat{K} \end{cases} \tag{4.31}$$

La fonction de régression est obtenue à partir de l'estimation des paramètres $\widehat{\theta} = \{\widehat{\tau}, \widehat{\eta}, \widehat{\sigma^2}\}$:

$$\widehat{\mathbf{m}} = Z(\widehat{\tau}, \widehat{\eta})\widehat{\beta}, \tag{4.32}$$

où $\widehat{\beta}$ est la moyenne de la loi a posteriori conditionnelle de β (3.33) :

$$\widehat{\beta} = \frac{\delta^2}{1 + \delta^2} [Z^T Z]^{-1} Z^T \mathbf{x}. \tag{4.33}$$

Usage	N_e	$\widehat{\text{RSB}}_{\text{dB}}$ (dB)	$\widehat{\sigma}_\epsilon$ (W)	\widehat{K}
Aspirateur	16	39.2	18.5	2.8
Bouilloire	12	52.5	3.73	2
Lampe halogene	5	36.5	3.97	3
Réfrigérateur	36	30.1	17.9	3.8
Four micro-ondes	35	32.3	57.4	3.2

TABLE 4.1: Pour chaque usage : nombre d'exemples dans la base de données SISED (N_e). Rapport signal à bruit en décibels entre la fonction de régression estimée et le bruit résiduel ($\widehat{\text{RSB}}_{\text{dB}}$). Ecart-type du bruit $\widehat{\sigma}$ en Watts. Nombre estimé de sous-modèles (\widehat{K}). Toutes les indicateurs sont moyennés sur l'ensemble de la base de données.

Notons que l'estimé obtenu par moyennage point par point de la fonction de régression (3.56), n'est pas pertinent si l'on cherche à extraire des paramètres significatifs pour chaque signal. Nous ne l'utiliserons donc pas ici.

La figure 4.5 présente la puissance active de différents transitoires d'aspirateur (en bleu) et la fonction de régression estimée pour chacun d'eux (en rouge). La position des transitions $\widehat{\tau}$ est représentée par des lignes pointillées verticales. Les figures 4.6, 4.7, 4.8 et 4.9 présentent des résultats analogues pour, respectivement, les transitoires de bouilloire, de lampe halogène, de réfrigérateur et de four micro-ondes.

Pour mesurer la qualité de la modélisation de chaque transitoire, on calcule le rapport signal à bruit entre la fonction de régression et le bruit résiduel :

$$\widehat{\text{RSB}} = \frac{E[\widehat{\mathbf{m}}^T \widehat{\mathbf{m}}]}{E[(\mathbf{x} - \widehat{\mathbf{m}})^T (\mathbf{x} - \widehat{\mathbf{m}})]},$$

que l'on exprime en décibels :

$$\widehat{\text{RSB}}_{\text{dB}} = 10 \log_{10}(\widehat{\text{RSB}}).$$

ainsi que l'écart-type du bruit résiduel :

$$\widehat{\sigma}_\epsilon = \sqrt{E[(\mathbf{x} - \widehat{\mathbf{m}})^T (\mathbf{x} - \widehat{\mathbf{m}})]}.$$

Le tableau 4.1 présente le rapport signal à bruit en décibels $\widehat{\text{RSB}}_{\text{dB}}$, l'écart type du bruit $\widehat{\sigma}_\epsilon$ et le nombre de sous-modèle \widehat{K} , moyennés sur chaque classe de signaux étudiée, ainsi que le nombre d'exemples au sein de chaque classe N_e .

4.3 Conclusions

Les résultats de la modélisation des signaux transitoires de la base de données SISED s'avèrent très satisfaisants, aussi bien en terme de qualité de la représentation des signaux qu'en terme de parcimonie. Le modèle à transitions régulières, muni de sous-modèles constants et de fonctions de transitions exponentielles étendues, fournit une très bonne approximation des classes de signaux étudiés. L'approche consistant à fixer la valeur du rapport signal à bruit δ^2 *a priori* dans notre modèle bayésien permet de réaliser le compromis précision/parcimonie pour la reconstruction, le nombre de ruptures K s'ajustant afin de ne représenter que les variations principales des signaux observés. Ces

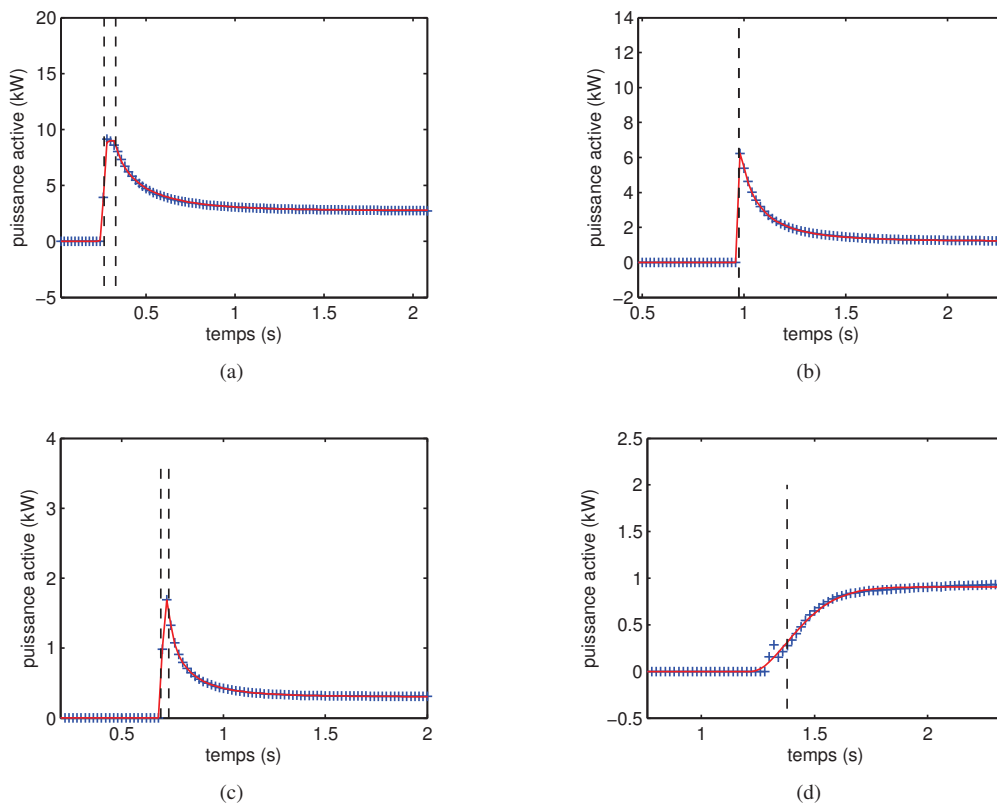


FIGURE 4.5: Transitoires d'aspirateurs : puissance active mesurée (points bleus) - estimateur MMAP du modèle à transition régulière (trait rouge) - positions estimées MMAP de ruptures (pointillés verticaux).

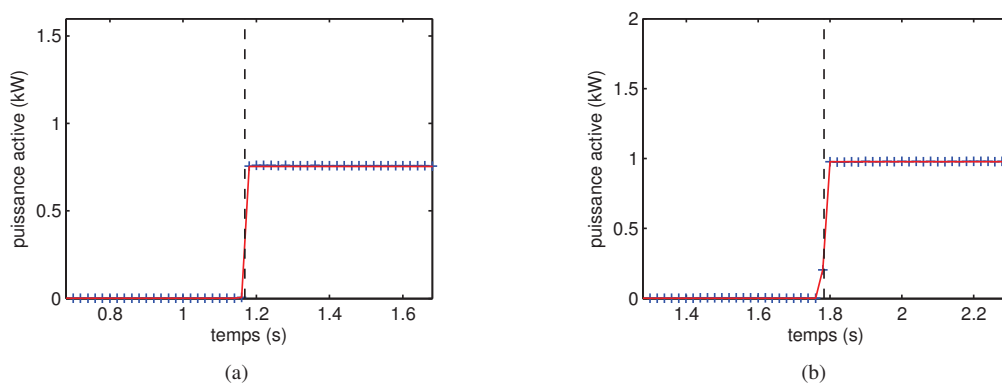


FIGURE 4.6: Transitoires de bouilloires : puissance active mesurée (points bleus) - estimateur MMAP du modèle à transition régulière (trait rouge) - positions estimées MMAP de ruptures (pointillés verticaux).

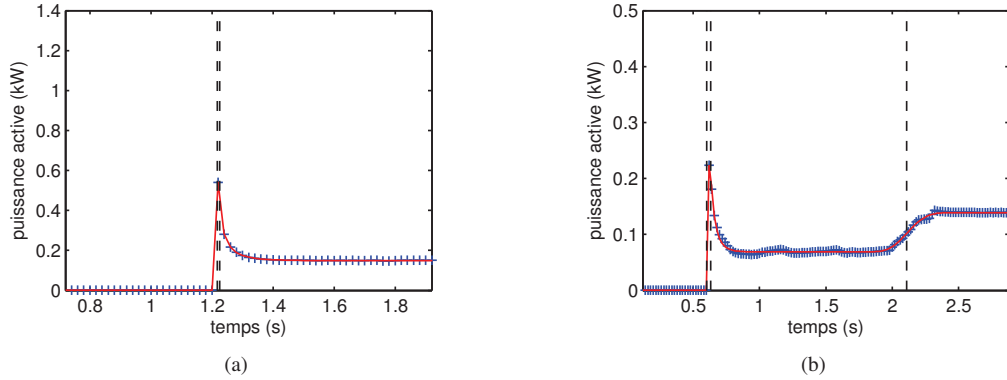


FIGURE 4.7: Transitoires de lampes halogènes : puissance active mesurée (points bleus) - estimateur MMAP du modèle à transition régulière (trait rouge) - positions estimées MMAP de ruptures (pointillés verticaux).

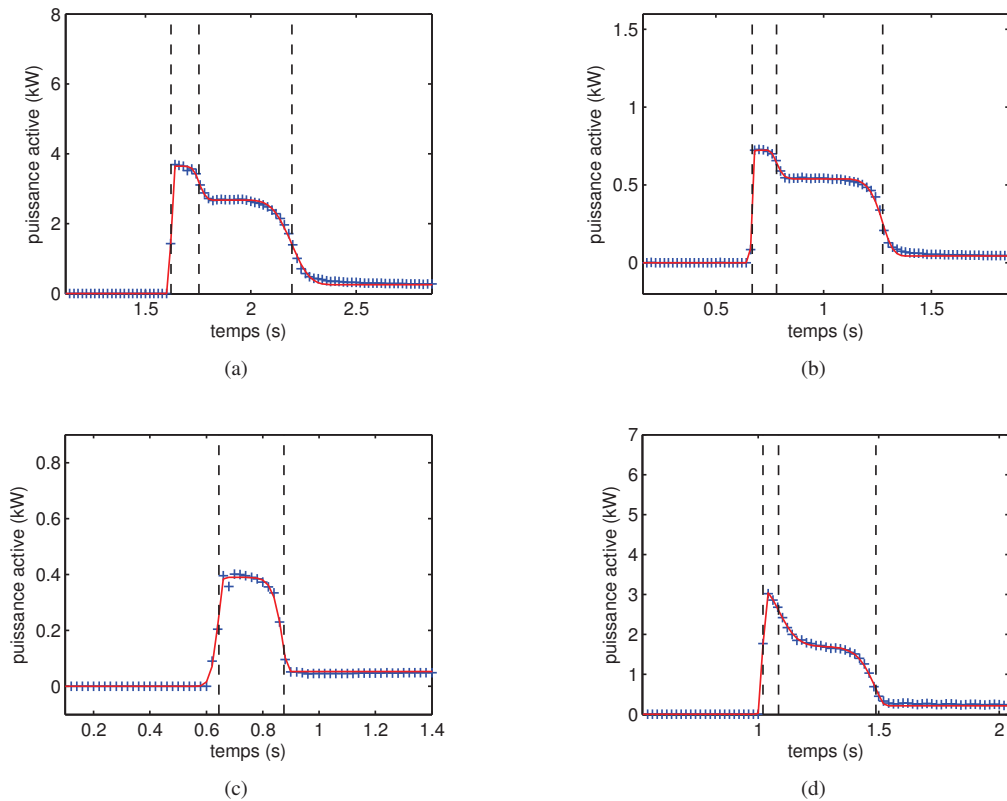


FIGURE 4.8: Transitoires de réfrigérateurs : puissance active mesurée (points bleus) - estimateur MMAP du modèle à transition régulière (trait rouge) - positions estimées MMAP de ruptures (pointillés verticaux).

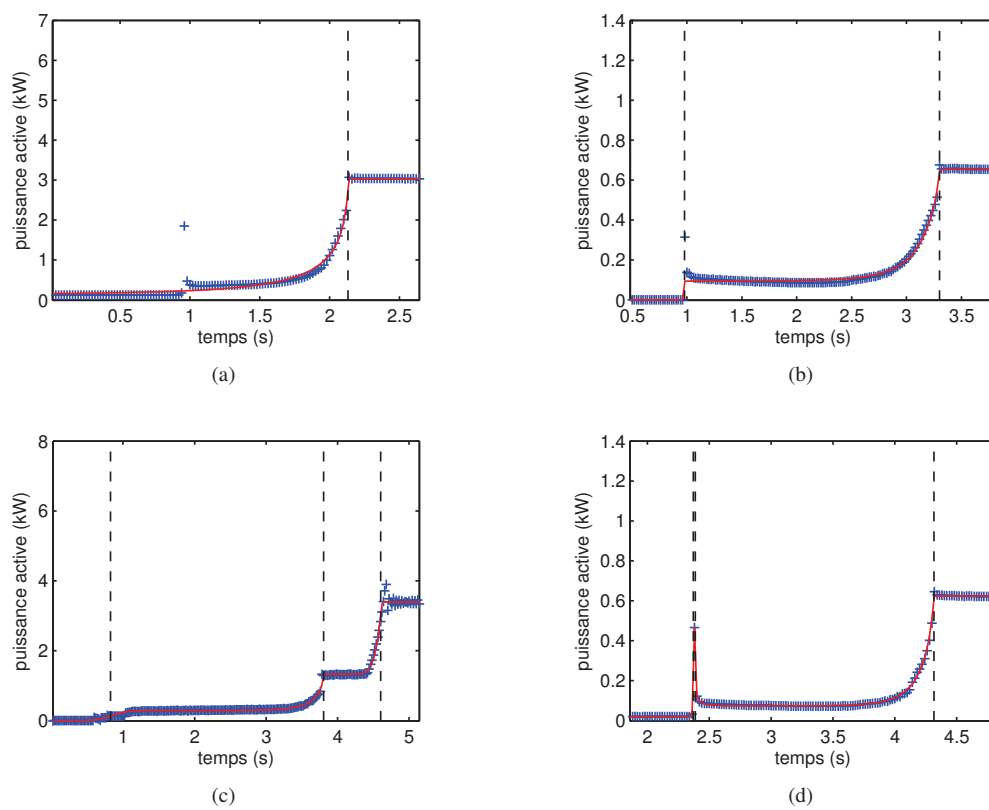


FIGURE 4.9: Transitoires de fours micro-ondes : puissance active mesurée (points bleus) - estimateur MMAP du modèle à transition régulière (trait rouge) - positions estimées MMAP de ruptures (pointillés verticaux).

résultats, même s'ils sont prometteurs, doivent maintenant être analysés plus globalement, classe par classe, afin de chercher à caractériser les types d'usage. Cela nécessite d'exploiter judicieusement les distributions *a posteriori* des paramètres estimés pour apprendre le comportement de chaque classe de signaux. Cette phase d'apprentissage fait l'objet du chapitre suivant.

Apprentissage de transitoires électro-domestiques

Lors des chapitres précédents, nous avons présenté un modèle de régression (chapitre 3) et nous avons constaté son efficacité pour modéliser les signaux de la base de données SISED (chapitre 4). Dans ce chapitre, nous mettons en œuvre une méthode d'apprentissage statistique afin, dans un premier temps, d'apprendre les caractéristiques des signaux au sein de chaque classe d'usage, et, dans un second temps, de déterminer automatiquement la classe de signaux extérieurs à cette base de données. Les classes de signaux que l'on cherche à caractériser sont les 5 classes d'usages sélectionnées dans la partie 2.1.2, c'est à dire la classe des signaux d'aspirateur, de réfrigérateur, de bouilloire, de lampe halogène et de four micro-ondes. La méthodologie proposée s'appuie naturellement sur l'estimation des paramètres du modèle de régression à transitions régulières.

5.1 Apprentissage statistique : enjeux et méthodes

5.1.1 Définitions

L'apprentissage statistique est le champ d'étude qui consiste à apprendre une fonction de prédiction à partir de données empiriques observées. Les données d'entrée appartiennent à l'espace des entrées, noté \mathcal{X} . Cet espace peut prendre des formes très diverses en fonction des applications (e.g., un ensemble de documents ou un ensemble de signaux que l'on cherche à classifier, les symptômes d'un patient que l'on souhaite diagnostiquer, etc.). La fonction de prédiction prend ses valeurs dans l'espace des sorties, noté \mathcal{Y} , qui constitue la grandeur d'intérêt (e.g., la classe du document ou du signal d'entrée, le diagnostic du patient). On supposera, dans le cadre de cette thèse, que \mathcal{Y} est un espace discret, que l'on notera \mathcal{C} , auquel cas on parle de classification (par opposition à la régression lorsque \mathcal{Y} est un espace continu).

Pour construire la fonction de prédiction, les méthodes d'apprentissage se fondent sur une base de données d'exemples, c'est-à-dire un ensemble de données d'entrées appartenant à \mathcal{X} . Ces méthodes d'apprentissage statistique se développent alors en deux phases :

1. Une phase d'apprentissage, consistant à identifier la structure des données observées, c'est-à-dire à construire la fonction de prédiction. On distingue ici l'apprentissage supervisé, où la classe (la valeur de la fonction de prédiction) de chaque donnée observée est connue, et l'apprentissage non-supervisé, où les classes doivent être identifiées simultanément à la fonction de prédiction, à partir des données observées.
2. Une phase de prédiction, consistant à déterminer la classe de nouvelles données d'entrées, c'est-à-dire à calculer la fonction de prédiction pour des points de \mathcal{X} qui n'appartiennent pas à l'ensemble d'apprentissage.

Il en résulte un algorithme permettant de généraliser le comportement analysé sur les données observées à de nouvelles données.

Notons que pour une majorité des problèmes concrets de classification, la fonction de prédiction n'est pas construite à partir des données d'entrées brutes, mais une phase d'extraction de caractéristiques est préalablement réalisée. Cette phase correspond à une transformation des données d'entrées afin d'extraire les caractéristiques significatives pour la classification. Chaque élément de \mathcal{X} est donc associé à un élément de l'espace des caractéristiques Θ . La fonction de prédiction est alors construite à partir des caractéristiques des données d'entrées.

5.1.2 Enjeux

Il y a deux enjeux majeurs, sous-jacents à cette méthodologie. D'une part la construction de la fonction de prédiction et son calcul pour toute nouvelle donnée présentée constitue un problème algorithmique pratique. D'autre part, les méthodes d'apprentissage statistique soulèvent un problème théorique : comment contrôler la capacité d'une méthode à généraliser le comportement appris sur quelques exemples à d'autres données. Plus précisément, une méthode d'apprentissage doit nécessairement répondre à un compromis entre deux phénomènes, le sous-apprentissage et le sur-apprentissage. Le sous-apprentissage se manifeste lorsque la fonction de prédiction ne prédit pas correctement la classe des données de la base d'apprentissage. Cela se produit dans le cas où la fonction de prédiction n'est pas suffisamment flexible pour s'adapter aux données d'entraînement. À l'opposé, le sur-apprentissage se manifeste par une prédiction correcte pour les données d'apprentissage, mais un taux d'erreur élevé pour la prédiction de nouvelles données. Cela se produit lorsque la fonction de prédiction est suffisamment complexe pour s'adapter exagérément aux données d'apprentissage qui peuvent être bruitées ou présenter une vision partielle de la réalité. Le compromis entre sous-apprentissage et sur-apprentissage est donc équivalent à un compromis entre capacité d'ajustement et complexité de la fonction de prédiction, qui est classique dans le domaine de la sélection de modèle. Pour une définition plus précise des notions présentées ci-dessus, le lecteur peut se rapporter à la théorie de Vapnik–Chervonenkis [Vapnik 2000].

5.1.3 Choix d'une méthode d'apprentissage

Dans la littérature, il existe de nombreuses méthodes pour la construction de la fonction de prédiction (voir par exemple [Bishop 2006, Hastie 2009] pour une vue d'ensemble de ces méthodes). On citera en particulier, les méthodes de partitionnement de données, tel que l'algorithme des k -moyennes qui permet de réaliser un apprentissage non-supervisé d'un ensemble de données en les regroupant par « paquets » homogènes par rapport à une distance donnée. Pour l'apprentissage supervisé, on peut considérer une méthode très simple : attribuer à toute nouvelle donnée la classe la plus fréquente parmi ses k plus proches voisins parmi les données d'apprentissage (algorithme KNN pour « *k-nearest-neighbor* »). Les méthodes basées sur les réseaux de neurones ont connues un important succès depuis les années 80. Enfin, depuis les années 90, les séparateurs à vastes marges (SVM, voir [Steinwart 2008]) ont fait l'objet de nombreux développements. Ces méthodes combinent deux éléments. D'une part, le coup du noyau (« *kernel-trick* ») permet, en transformant les données d'entrées, de remplacer un problème non-linéaire par un problème linéaire. D'autre part, les SVM recherchent un classifieur sous la contrainte de marge maximale pour limiter les problèmes de sur-apprentissage.

Cependant, ces méthodes ne peuvent être appliquées directement au problème de classification utilisant un modèle à transitions régulières. En effet, une difficulté spécifique à la démarche proposée est que le nombre de paramètres estimé est variable d'un signal à l'autre. Le nombre de caractéristiques utilisé pour classifier chaque signal est donc variable. Plus précisément, l'espace des caractéristiques est la réunion de plusieurs sous-espaces de dimension différentes (cf. équation

(3.37)). Or les méthodes standard de classification, ne permettent pas de manipuler directement ce type d'espaces.

Les méthodes citées jusqu'à présent appartiennent à la catégorie des méthodes discriminatives, dans le sens où elles visent à calculer directement la règle de classification $p(c|\boldsymbol{\theta})$, $\forall c \in \mathcal{C}, \forall \boldsymbol{\theta} \in \Theta$. Par opposition, les méthodes génératives visent à modéliser la densité de probabilité jointe $p(c, \boldsymbol{\theta})$, $\forall c \in \mathcal{C}, \forall \boldsymbol{\theta} \in \Theta$. La modélisation de cette loi jointe permet ensuite de calculer la règle de classification en utilisant la loi de Bayes. Bien que considérées comme sous-optimales par rapport aux méthodes discriminatives, les méthodes génératives proposent une modélisation globale du problème et permettent d'introduire facilement des connaissances *a priori* sur la structure des données d'entrée dans la règle de classification.

Nous proposons donc l'utilisation d'une méthode générative pour la classification, basée sur une approche bayésienne, dans la continuité méthodologique de la démarche adoptée pour l'estimation des paramètres du modèles. En plus de fournir un cadre théorique pour manipuler les éléments de l'espace des caractéristiques, nous pourrons alors combiner l'étape d'extraction des caractéristiques avec l'étape d'apprentissage au sein d'une unique structure bayésienne hiérarchique.

Dans un premier temps, ce chapitre décrit schématiquement la mise en œuvre d'un classifieur bayésien standard. Puis nous proposons de réaliser conjointement l'estimation des paramètres de chaque exemple de la base de données et l'apprentissage des paramètres par le moyen d'une approche bayésienne hiérarchique. Enfin, nous exposons les résultats obtenus pour l'apprentissage et la classification des signaux de la base de données SISED.

5.2 Apprentissage bayésien

5.2.1 Apprentissage bayésien standard

Notations

Dans le cadre de l'application de classification des signaux de la base SISED, l'espace des entrées, \mathcal{X} , correspond à l'espace des séries temporelles des échantillons de la puissance active d'un signal transitoire. Notons $\mathcal{X}_{\text{BDD}} \subset \mathcal{X}$ l'ensemble des signaux de la base de données d'apprentissage. Chaque élément $\boldsymbol{x} \in \mathcal{X}_{\text{BDD}}$ correspond donc à la série temporelle des échantillons de la puissance active d'un signal transitoire. Pour chaque signal $\boldsymbol{x} \in \mathcal{X}_{\text{BDD}}$ on dispose d'une étiquette de la classe du signal $c(\boldsymbol{x}) \in \mathcal{C}$ où \mathcal{C} est l'ensemble des étiquettes possibles $\mathcal{C} = \{1, \dots, N_{\text{class}}\}$ et N_{class} est le nombre de classes ($N_{\text{class}} = 5$ dans cette thèse). Enfin, notons $\mathcal{X}_{\text{BDD}_c}$ l'ensemble des signaux d'apprentissage appartenant à la classe c , $\mathcal{X}_{\text{BDD}_c} = \{\boldsymbol{x} \in \mathcal{X}_{\text{BDD}} : c(\boldsymbol{x}) = c\}$, $\forall c \in \mathcal{C}$.

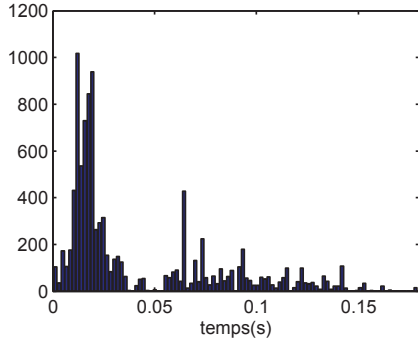
Dans un cadre bayésien, le problème de la classification supervisée d'un signal quelconque $\boldsymbol{x} \in \mathcal{X}$, se résume à calculer la densité *a posteriori* de l'étiquette du nouveau signal $c = c(\boldsymbol{x})$ connaissant la base de donnée étiquetée $p(c|\boldsymbol{x}, \mathcal{X}_{\text{BDD}})$. Si l'on choisit une fonction de coût, on peut alors déterminer l'estimateur de c . Par exemple, pour une fonction de coût 0 – 1, l'estimateur de c sera :

$$\hat{c} = \arg \max_{c \in \mathcal{C}} p(c|\boldsymbol{x}, \mathcal{X}_{\text{BDD}}). \quad (5.1)$$

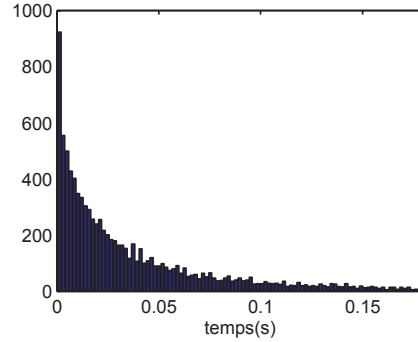
Par ailleurs, la densité *a posteriori* $p(c|\boldsymbol{x}, \mathcal{X}_{\text{BDD}})$ s'exprime d'après la loi de Bayes

$$p(c|\boldsymbol{x}, \mathcal{X}_{\text{BDD}}) \propto p(\boldsymbol{x}|c, \mathcal{X}_{\text{BDD}})p(c), \quad (5.2)$$

où $p(c)$ désigne l'*a priori* dont on dispose sur la classe c des signaux.



(a) Apprentissage bayésien standard : Distribution du paramètre d'échelle λ_2 pour la classe « Lampe Halogène, conditionnellement à $K = 3$. »



(b) Apprentissage bayésien hiérarchique : Distribution du paramètre d'échelle λ_2 pour la classe « Lampe Halogène, conditionnellement à $K = 3$. »

FIGURE 5.1: Exemple de régularisation de la distribution apprise des caractéristiques.

Extraction de caractéristiques

Pour les raisons discutées au paragraphe 2.4, nous ne souhaitons pas utiliser directement les échantillons du signal temporel \mathbf{x} pour identifier chaque classe de signaux. Nous cherchons à obtenir à partir des signaux étudiés un ensemble de caractéristiques θ pertinentes pour la classification. Cette étape permet de réduire la dimension des objets à classifier (les signaux temporels \mathbf{x}) pour en extraire l'information la plus pertinente possible. Ces caractéristiques θ correspondent aux paramètres du modèle à transitions régulières présenté précédemment. Dans un cadre bayésien, il existe plusieurs approches pour mettre en relation l'extraction de caractéristiques et l'apprentissage.

Dans une approche de classification standard, l'étape d'extraction de caractéristiques est réalisée lors d'une étape préliminaire à l'apprentissage, de façon entièrement indépendante. Chaque signal \mathbf{x} se voit associé un vecteur de caractéristiques θ le représentant et l'apprentissage est réalisé à partir de ces caractéristiques. Il suffit alors d'appliquer l'apprentissage bayésien présenté ci-dessus en remplaçant les signaux d'apprentissage et les signaux test par leur vecteur de caractéristiques. Il faut alors construire la densité de probabilité $p(\theta|c, \mathcal{X}_{\text{BDD}})$ pour chaque classe. Cette première approche est adaptée si l'étape d'extraction de caractéristique est directe, c'est-à-dire si l'estimation des caractéristiques d'un signal quelconque ne pose aucun problème.

Dans le cas présent, l'estimation des caractéristiques présente à elle seule une difficulté, que nous avons résolue à l'aide d'une méthode bayésienne dédiée (cf chapitre 3). Il serait envisageable dans ce cas de recourir à un estimateur bayésien standard des caractéristiques, par exemple l'estimateur du maximum *a posteriori* $\hat{\theta}_{\text{map}}$, et de l'utiliser comme entrée du classifieur en lieu et place des signaux \mathbf{x} . Cependant cette approche s'avère peu satisfaisante car l'ensemble de la distribution *a posteriori* $f(\theta|\mathbf{x})$ des paramètres θ est alors réduite à la seule valeur ponctuelle d'un estimateur.

Afin d'exploiter toute la distribution *a posteriori* des paramètres, une solution alternative consisterait à combiner les distributions *a posteriori* des paramètres estimés pour chacun des signaux d'entraînement d'une même classe afin de former une unique distribution représentative de la classe. Cette approche est relativement simple à mettre en œuvre, puisqu'il suffit, par exemple, de considérer que la loi $p(\theta|c, \mathcal{X}_{\text{BDD}_c})$ est constituée du mélange de chacune des lois *a posteriori* $p(\theta|\mathbf{x})$ pour $\mathbf{x} \in \mathcal{X}_{\text{BDD}_c}$.

A ce stade, le problème rencontré est que la loi ainsi obtenue ne se généralise pas forcément aux

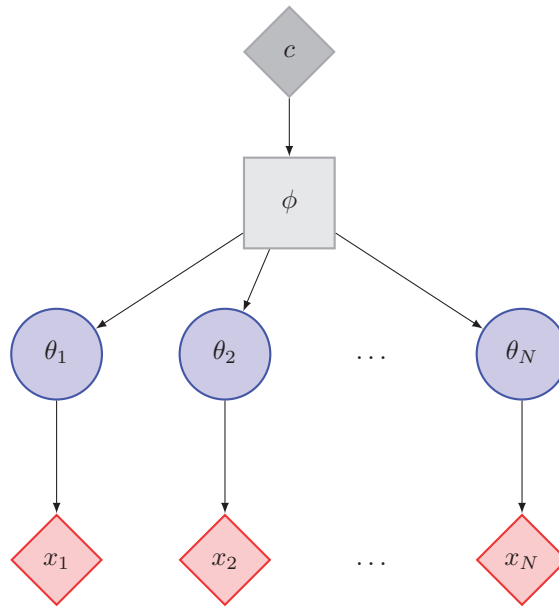


FIGURE 5.2: Graphe orienté acyclique des dépendances conditionnelles entre les observations x_i , les paramètres θ_i et les hyperparamètres ϕ et c pour le modèle d'apprentissage bayésien hiérarchique.

signaux n'appartenant pas aux signaux d'apprentissage. Par exemple, dans le cas des signaux transitoires électriques, la figure 5.1(a) représente la loi obtenue en mélangeant la loi *a posteriori* du paramètre d'échelle de la deuxième transition de chacun des exemples de la classe « Lampe Halogène ». Cette loi, peu régulière, présente deux modes principaux. Cela est dû au petit nombre d'exemples dont nous disposons pour la construire. Naturellement, en augmentant le nombre d'exemples, les « vides » de cette loi devraient probablement se combler et la loi apprise se régulariserait. Cependant, en l'absence de données supplémentaires, il faut forcer la régularisation de la loi apprise afin d'assurer la capacité de la méthode à généraliser le comportement appris sur un nombre réduit d'exemples.

Dans le cadre de méthodes bayésiennes standard, cette loi peut-être régularisée *a posteriori* en réalisant un lissage de la distribution des caractéristiques des exemples de la base de données. Dans cette thèse nous proposons d'utiliser une méthode d'apprentissage bayésien hiérarchique qui permet de contrôler *a priori* la forme de la distribution apprises des caractéristiques et de réaliser conjointement l'estimation des paramètres θ et le calcul de la loi $p(\theta|c, \mathcal{X}_{\text{BDD}})$. Ce type de méthode à l'avantage sur les méthodes classiques de permettre d'aboutir directement à des distributions régulières pour les caractéristiques classifiées (cf la Fig. 5.1(b)).

5.2.2 Modèle bayésien hiérarchique pour l'apprentissage

L'apprentissage bayésien hiérarchique [Davy 2002] consiste à combiner l'étape d'extraction des caractéristiques des signaux tests (i.e., l'estimation des paramètres du modèle à transitions régulières pour chaque signal) avec l'étape d'apprentissage (i.e., l'estimation de la distribution des caractéristiques au sein d'une classe). Ces deux étapes, habituellement traitées de manière séquentielle, sont réalisées conjointement au sein d'un modèle bayésien hiérarchique. Supposons que la

classe c contient N_c signaux d'apprentissage $N_c = \text{card}(\mathcal{X}_{\text{BDD}c})$, que l'on numérote arbitrairement de 1 à N_c . Pour chaque signal \mathbf{x}_i tel que $i = 1, \dots, N_c$, on considère le modèle génératif suivant :

$$\mathbf{x}_i \sim p(\mathbf{x}_i | \boldsymbol{\theta}_i), \quad (5.3)$$

$$\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta}_i | \phi), \quad (5.4)$$

$$\phi \sim p(\phi | c). \quad (5.5)$$

Les dépendances conditionnelles engendrées par les équations ci-dessus se traduisent par le graphe orienté acyclique de la figure 5.2. Ce modèle hiérarchique présente deux niveaux d'abstraction principaux. D'une part, chaque signal \mathbf{x}_i est contrôlé par un vecteur de caractéristiques $\boldsymbol{\theta}_i$ qui lui est propre. Cela permet au modèle de prendre en compte la variabilité des signaux au sein d'une classe : bien que les signaux d'une même classe possèdent des caractéristiques semblables, ils ne sont pas strictement identiques. D'autre part, la distribution $f(\cdot | \phi)$ de chacune de ces caractéristiques $\boldsymbol{\theta}_i$ obtenue pour chacun des signaux \mathbf{x}_i est la même pour tous les signaux $\mathbf{x}_1, \dots, \mathbf{x}_{N_c}$ d'une même classe c . C'est ce second niveau d'abstraction qui permet au modèle de regrouper les signaux au sein d'une même classe en considérant les caractéristiques suffisamment proches pour être modélisées selon la même loi. Ces deux niveaux d'abstraction correspondent en fait à des sur-hypothèses (en anglais « *overhypothesis* ») au sens où l'entend l'auteur de [Kemp 2007]. Chaque niveau du modèle bayésien hiérarchique proposé correspond donc à une étape standard de l'apprentissage supervisé :

1. estimer $\boldsymbol{\theta}_i$ à partir du signal \mathbf{x}_i et des hyperparamètres ϕ correspond à l'extraction du vecteur de caractéristiques,
2. estimer la loi de ϕ à partir de $\{\boldsymbol{\theta}_i\}_{i=1, \dots, N_c}$, étant donné une classe c , correspond à l'apprentissage de la distribution des caractéristiques au sein d'une classe,

ces étapes étant maintenant réalisées conjointement.

Pour pouvoir par la suite obtenir une règle de classification pour de nouveaux signaux, nous devons donc, pour chaque classe, estimer la loi des paramètres ϕ à partir des signaux de la base d'apprentissage $\mathcal{X}_{\text{BDD}c} = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_c}\}$. D'après la loi de Bayes, la loi *a posteriori* des hyperparamètres ϕ vaut :

$$p(\phi | \mathbf{x}_1, \dots, \mathbf{x}_{N_c}, c) \propto p(\mathbf{x}_1, \dots, \mathbf{x}_{N_c} | \phi) p(\phi | c), \quad (5.6)$$

Les signaux \mathbf{x}_i étant supposés indépendants conditionnellement à ϕ , il vient :

$$p(\phi | \mathbf{x}_1, \dots, \mathbf{x}_{N_c}, c) \propto p(\phi | c) \prod_{i=1}^{N_c} p(\mathbf{x}_i | \phi). \quad (5.7)$$

$$\propto p(\phi | c) \prod_{i=1}^{N_c} \int p(\mathbf{x}_i, \boldsymbol{\theta}_i | \phi) d\boldsymbol{\theta}_i. \quad (5.8)$$

$$\propto p(\phi | c) \prod_{i=1}^{N_c} \int p(\mathbf{x}_i | \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | \phi) d\boldsymbol{\theta}_i. \quad (5.9)$$

En pratique, la marginalisation par rapport aux vecteurs de caractéristiques $\boldsymbol{\theta}_i$ de chaque signal est rarement réalisable analytiquement. C'est pourquoi on s'intéresse à la distribution jointe des hyperparamètres ϕ et des vecteurs de caractéristiques $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{N_c}$:

$$p(\phi, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{N_c} | \mathbf{x}_1, \dots, \mathbf{x}_{N_c}, c) \propto p(\phi | c) \prod_{i=1}^{N_c} p(\mathbf{x}_i | \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | \phi). \quad (5.10)$$

Cette dernière équation est centrale puisqu'elle permet d'estimer conjointement le vecteur de caractéristiques de chaque signal et le vecteur des hyperparamètres qui détermine la distribution de ces vecteurs au sein d'une classe. Notons que le modèle hiérarchique proposé peut être identifié à un modèle à variables latentes (cf [Bishop 1999]), où on cherche à inférer la classe, à partir des signaux observés, les caractéristiques θ étant des variables latentes.

Jusqu'ici nous sommes restés dans un cadre général sans préciser les densités de probabilités (5.3), (5.4) et (5.5) qui déterminent le modèle hiérarchique. Dans le cadre de cette thèse, la densité de probabilité (5.3) correspond simplement à la vraisemblance du modèle à transitions régulières (3.14), telle qu'exposée au chapitre 3. La suite de ce chapitre est donc consacrée à préciser la densité de probabilité *a priori* des vecteurs de caractéristiques (5.4) et des hyperparamètres (5.5).

5.2.3 Densité *a priori* des vecteurs de caractéristiques

Comme nous l'avons vu, la densité de probabilité (5.4) spécifie, pour le modèle hiérarchique présenté, la distribution des vecteurs de caractéristiques pour une classe donnée. C'est donc cette densité de probabilité qui est garante du compromis entre sous-apprentissage et sur-apprentissage des signaux de la base de données. Le choix d'une densité de probabilité trop rigide (avec trop peu de paramètres) empêcherait le modèle de capturer la variabilité entre deux signaux de la même classe. Inversement, une densité de probabilité trop flexible (avec trop de degrés de liberté) provoquerait certainement des problèmes de sur-apprentissage.

Classification et modèles à dimension variables : hypothèses sur la loi *a priori*

La difficulté principale, liée à l'utilisation du modèle à transitions régulières, est que l'espace des caractéristiques est la réunion de sous-espaces de différentes dimensions (voir (3.37)). On rappelle que le jeu des paramètres associé au signal x_i s'exprime comme $\theta_i = (\tau_i, \eta_i, \beta_i, K_i, \sigma_i^2)$ pour le modèle à transitions régulières. τ_i est le vecteur contenant les positions des transitions, β_i est le vecteur des coefficients des sous-modèles, σ_i^2 est la variance du bruit et $\eta_i = (\lambda_i, \alpha_i, c_i)$ est le jeu des paramètres des fonctions de transitions présentées précédemment (voir paragraphe 4.1.3) et associées à chacune des transitions. λ_i est le vecteur des paramètres d'échelle, α_i le vecteur des paramètres de forme et c_i le vecteur des variables indicatrices de la fonction exponentielle généralisée ou de la fonction complémentaire. Il est important de remarquer que la dimension des vecteurs τ_i , η_i et β_i est une fonction du nombre de ruptures K_i .

Afin d'appréhender plus facilement les lois de ces paramètres qui prennent leurs valeurs dans des espaces de dimensions variables, les hypothèses suivantes sont supposées satisfaites. Conditionnellement au nombre de ruptures $K_i = k$:

- les lois *a priori* des vecteurs de paramètres $(\tau_i, \eta_i, \beta_i)$ sont gouvernées par des jeux d'hyperparamètres distincts notés $\phi_k \subset \phi : (\tau_i, \eta_i, \beta_i) | (K_i = k) \sim p(\cdot | \phi_k)$ pour tout $k \geq 0$,
 - les vecteurs de paramètres $((\tau_i, \eta_i, \beta_i) | \phi_k)_{k \geq 0}$ sont mutuellement *a priori* indépendants.
- Par conséquent, en supposant de plus que les paramètres K_i et σ_i^2 sont *a priori* indépendants sachant ϕ , la densité de probabilité *a priori* jointe s'exprime comme

$$p(\theta_i | \phi) = p(\sigma_i^2 | \phi) \sum_{k=0}^{+\infty} p(K_i = k | \phi) p(\tau_i, \eta_i, \beta_i | \phi_k, \sigma_i^2). \quad (5.11)$$

Le choix fait dans cette thèse de différencier les hyperparamètres associés à chaque valeur du nombre de ruptures K présente l'avantage de la simplicité en subdivisant les signaux d'apprentissage en sous-ensembles associés aux différentes valeurs de K . L'inconvénient est qu'en subdivisant

ainsi la base de données, le risque existe d'appauvrir l'information disponible dans une base de donnée déjà réduite. D'autre part, le nombre d'hyperparamètres permettant de représenter la distribution des caractéristiques des signaux au sein d'une classe, est alors conséquent.

Faire un choix différent pour la structure de la densité de probabilité $p(\theta_i|\phi)$ soulèverait immédiatement un problème d'identification des fonctions de transitions entre deux modèles ne possédant pas le même nombre de sous-modèles. Par exemple, admettons qu'un premier signal x_1 soit modélisé par une fonction de transition ($K_1 = 1$) alors qu'un second signal x_2 est modélisé par deux fonctions de transition ($K_2 = 2$). Il est possible que les deux signaux soient très similaires, mais que l'une des deux fonctions de transition du signal x_2 corresponde à un niveau de détail supplémentaire par rapport à la représentation du signal x_1 . Dans ce cas, l'autre fonction de transition de x_2 doit posséder des caractéristiques similaires à la fonction de transition de x_1 , et il pourrait être utile de considérer ces corrélations dans le cadre de l'apprentissage de ces signaux. Cependant, le problème consistant à identifier la fonction de transition du signal x_2 correspondant à la fonction de transition de x_1 n'est pas trivial, même si des travaux récents [Roodaki 2012] visent à interpréter les échantillons générés par un algorithme RJMCMC afin d'identifier ce type de corrélations dans des espaces trans-dimensionnels. Dans cette thèse nous ne tiendrons pas compte de ces corrélations possibles entre les représentations de dimension différentes.

Nous détaillons ci-après la distribution des jeux de paramètres qui caractérisent les signaux au sein d'une classe. Notons que le choix de chacune de ces lois est guidé par les mêmes principes généraux. D'une part, nous choisissons aussi souvent que possible des lois conjuguées par rapport à la vraisemblance du modèle. Mais contrairement aux chapitres précédents où les lois *a priori* étaient choisies afin d'être si possible non-informatives et de faire intervenir peu d'hyperparamètres, il peut être intéressant d'introduire des hyperparamètres supplémentaires. Ces hyperparamètres supplémentaires peuvent en effet capturer directement les différentes caractéristiques des lois des paramètres selon chaque classe. Néanmoins, les lois *a priori* choisies doivent également rester unimodales et, à l'inverse, ne pas dépendre de trop d'hyperparamètres. Il s'agit de limiter le phénomène de sur-apprentissage, notamment au vu de la taille réduite de la base de données SISED.

Loi du nombre de transitions K_i

Pour la loi du nombre de transitions, nous choisissons une loi de Poisson, paramétrée par sa moyenne μ_K :

$$p(K_i|\mu_K) = e^{-\mu_K} \frac{\mu_K^{K_i}}{K_i!}, \quad (5.12)$$

pour tout $K_i \in \mathbb{N}$. Ce choix est similaire à celui effectué au chapitre 3 pour l'estimation des paramètres d'un seul transitoire, voir paragraphe 3.4.1.

Loi de la variance du bruit σ_i^2

La variance du bruit σ_i^2 est supposée distribuée selon une loi inverse-gamma, de paramètre d'échelle ρ_σ :

$$\sigma_i^2|\rho_\sigma \sim \mathcal{IG}(1, \rho_\sigma). \quad (5.13)$$

Cette loi est marginalement conjuguée par rapport à la vraisemblance du modèle, et permet d'introduire un nouvel hyperparamètre ρ_σ par rapport au modèle (3.27) présenté lors du chapitre 3. Ce degré de liberté supplémentaire vise à capturer les principales tendances de la variance au sein d'une classe donnée.

Loi des paramètres des sous-modèles β_i

Dans le cadre de l'apprentissage, nous supposons que le premier sous-modèle est une fonction constante connue. Cela correspond à l'hypothèse de régime permanent de la consommation électrique avant l'appel de courant initial. Nous avons pu vérifier la validité de cette hypothèse lors de la phase de modélisation des signaux de la base de données SISED. De plus, on peut supposer, sans perte de généralité, que cette consommation initiale est nulle et est donc modélisée par la fonction nulle. Les paramètres des sous-modèles β_i sont alors supposés distribués selon une loi de type g-prior identique à la loi *a priori* utilisée pour l'estimation des paramètres du modèle (cf paragraphe 3.4.1) :

$$\beta_i | \sigma_i^2, Z_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \delta^2 [(Z_i)^T Z_i]^{-1}), \quad (5.14)$$

où δ^2 est un hyperparamètre et où Z_i^* est la matrice de régression du signal i , i.e., la matrice $Z(\tau_i, \eta_i)$, dont la première colonne a été retirée en accord avec l'hypothèse que le premier sous-modèle est nul en tout point.

Les avantages de ce choix pour la loi *a priori* des paramètres des sous-modèles β sont les mêmes que ceux exposés en 3.4.1. D'une part, l'estimation des paramètres est alors invariante par rapport à un changement de base de régression des sous-modèles (cf proposition 3.4.1). D'autre part, l'hyperparamètre δ^2 est alors identifiable à un rapport signal à bruit (cf proposition 3.4.2), ce qui nous permet de fixer *a priori* le niveau de reconstruction attendu de la modélisation à transition régulière (cf paragraphe 4.1.4). Par la suite, δ^2 est fixé à un niveau de 15dB, soit $\delta^2 = 31.6$.

Loi de la position τ_i des ruptures

La loi *a priori* de la position des ruptures ne porte en fait pas directement sur ces grandeurs. En effet, la position absolue des ruptures n'est pas significative, puisque la fenêtre d'observation des signaux de la base de données est arbitraire. La position de la première rupture représente en fait l'instant d'enclenchement de l'usage et n'est donc pas utile pour la classification de l'usage. On la modélise selon une loi uniforme sur une fenêtre de recherche donnée. Les grandeurs d'intérêt pour la classification sont donc plutôt la longueur des intervalles entre les transitions $d\tau_{k,i} = \tau_{k+1,i} - \tau_{k,i}$, pour tout $1 \leq k \leq K_i$, où la configuration des ruptures est supposée ordonnée : $\tau_{1,i} < \tau_{2,i} \dots < \tau_{K_i,i}$. Les longueurs de ces intervalles sont supposées *a priori* indépendantes et distribuées selon les lois exponentielles suivantes :

$$d\tau_{k,i} | \rho_{\tau_{k,K_i}} \sim \mathcal{G}(\nu_{\tau_{k,K_i}}, \rho_{\tau_{k,K_i}}) \quad \forall k = 1, \dots, K_i - 1. \quad (5.15)$$

où $\nu_{\tau_{k,K_i}}$ est un paramètre de forme fixé à $\nu_{\tau_{k,K_i}} = 1$ (i.e., une loi exponentielle) et $\rho_{\tau_{k,K_i}}$ est alors le paramètre d'échelle modélisant la moyenne du k -ième intervalle $d\tau_{k,i}$. Il est intéressant de souligner les similitudes avec le processus de Poisson homogène (3.17) choisi comme *a priori* au chapitre 3. En effet, une propriété classique dans le cas du processus de Poisson homogène est que les longueurs inter-occurrences sont indépendantes et distribuées selon des lois exponentielles. Cependant, ces lois exponentielles sont toutes identiques avec un paramètre qui dépend de l'intensité du processus. Avec ce nouvel *a priori*, un paramètre $\rho_{\tau_{k,K_i}}$ spécifique à chaque configuration ayant un nombre de segments K_i et à chaque segment k , pour $1 \leq k \leq K_i$, de cette configuration, est introduit afin de caractériser ces lois.

On note

$$\rho_{\tau_{K_i}} = (\rho_{\tau_{1,K_i}}, \dots, \rho_{\tau_{K_i-1,K_i}})$$

l'ensemble de ces paramètres d'échelle pour une valeur $K = K_i$. La loi *a priori* du vecteur τ est alors

$$p(\tau_i | \rho_{\tau_{K_i}}) = \frac{1}{L} \mathbb{I}_{[t_0^i, t_0^i + L]}(\tau_{1,i}) \prod_{k=1}^{K_i-1} \frac{1}{\rho_{\tau_{k,K_i}}} \exp\left(-\frac{\tau_{k+1,i} - \tau_{k,i}}{\rho_{\tau_{k,K_i}}}\right) \mathbb{I}_{\mathbb{R}_+}(\tau_{k+1,i} - \tau_{k,i}), \quad (5.16)$$

où $[t_0^i, t_0^i + L]$ désigne l'intervalle des valeurs possibles pour la première rupture $\tau_{1,i}$, la longueur L étant un paramètre déterministe fixé à la moitié de la fenêtre d'observation $L = \frac{t_0^i + t_n^i}{2}$. Notons que les hyperparamètres $\rho_{\tau_{K_i}}$ sont donc liés à la moyenne des longueurs des intervalles entre les positions des transitions.

Loi des paramètres d'échelle λ_i des fonctions de transition

Les paramètres d'échelle des fonctions de transition sont distribués selon une loi gamma au sein de chaque classe :

$$\lambda_{k,i} | \rho_{\lambda_{k,K_i}} \sim \mathcal{G}\left(\nu_{\lambda}, \rho_{\lambda_{k,K_i}}\right) \quad \forall k = 1, \dots, K_i, \quad (5.17)$$

où $\nu_{\lambda_{k,K_i}}$ est un paramètre de forme fixé à $\nu_{\lambda_{k,K_i}} = 1$ et $\rho_{\lambda_{k,K_i}}$ est un paramètre d'échelle lié à la valeur moyenne du paramètre λ_i . On note

$$\rho_{\lambda_{K_i}} = (\rho_{\lambda_{1,K_i}}, \dots, \rho_{\lambda_{K_i,K_i}}) \quad (5.18)$$

l'ensemble de ses paramètres d'échelle pour une valeur donnée de K_i .

Loi des paramètres de forme α_i des fonctions de transition

Les paramètres de forme des fonctions de transition sont distribués selon une loi normale de moyenne $\mu_{\alpha_{k,K_i}}$ et de variance $\sigma_{\alpha_{k,K_i}}^2$:

$$\alpha_k | \mu_{\alpha_{k,K_i}}, \sigma_{\alpha_{k,K_i}}^2 \sim \mathcal{N}\left(\mu_{\alpha_{k,K_i}}, \sigma_{\alpha_{k,K_i}}^2\right) \quad \forall k = 1, \dots, K_i. \quad (5.19)$$

On note $\mu_{\alpha_{K_i}} = (\mu_{\alpha_{1,K_i}}, \dots, \mu_{\alpha_{K_i,K_i}})$ et $\sigma_{\alpha_{K_i}}^2 = (\sigma_{\alpha_{1,K_i}}^2, \dots, \sigma_{\alpha_{K_i,K_i}}^2)$, respectivement les hyperparamètres de moyenne et de variance des lois de probabilité des paramètres de forme des fonctions de transition, pour une valeur donnée du nombre de ruptures K_i .

Loi du type c des fonctions de transition

Le type de chaque fonction de transition $c = \{c_1, \dots, c_{K_i}\}$ est supposé distribué selon une loi de Bernoulli de paramètre p_{k,K_i} :

$$\begin{cases} p(c_k = 0 | p_{k,K_i}, K_i) &= p_{k,K_i}, \\ p(c_k = 1 | p_{k,K_i}, K_i) &= 1 - p_{k,K_i} \end{cases} \quad (5.20)$$

pour tout $1 \leq k \leq K_i$. On note $p_{c_{K_i}} = (p_{c_{1,K_i}}, \dots, p_{c_{K_i,K_i}})$ le vecteur des paramètres associés.

Indépendance conditionnelle des lois *a priori* des paramètres θ_i

Finalement les différents jeux paramètres présentés ci-avant sont supposés être *a priori* mutuellement indépendants conditionnellement à ϕ et donc à la classe c considérée (voir le graphe de dépendances décrit fig. 5.2). C'est l'hypothèse classique de la classification bayésienne naïve. Bien que simplificatrice, elle s'avère optimale dans de nombreux cas [Domingos 1997]. De plus elle permet de limiter le nombre d'hyperparamètres, puisque les dépendances entre les caractéristiques n'ont pas à être apprises.

De plus, pour une classe c donnée, les seuls paramètres dont les lois *a priori* font apparaître des dépendances sont le vecteur des positions des ruptures τ (ce sont en fait la longueur des intervalles entre les transitions qui sont indépendantes), et le vecteur des paramètres des sous-modèles β (car le g-prior induit une corrélation entre les éléments de β).

Loi *a priori* des hyperparamètres

Le vecteur des hyperparamètres ϕ s'écrit finalement

$$\phi = (\mu_K, \rho_\sigma, \phi_0, \dots, \phi_k, \dots), \quad (5.21)$$

où $\phi_k = (\rho_{\tau_k}, \mu_{\beta_k}, \rho_{\lambda_k}, \mu_{\alpha_k}, \sigma_{\alpha_k}^2, p_{c_k})$ est le jeu des hyperparamètres associés à une configuration de $K = k$ ruptures. Il faut remarquer que ce vecteur ϕ prend ces valeurs dans un espace de dimension infinie. C'est en effet une conséquence de notre modélisation bayésienne non-paramétrique : la configuration des transitions est modélisée comme la réalisation d'un processus à valeurs dans l'espace exponentiel de toutes les configurations finies de transitions, qui est de dimension infinie.

Le dernier niveau (5.5) du modèle hiérarchique reste maintenant à spécifier. Cela correspond à déterminer la loi *a priori* de tous les hyperparamètres ϕ conditionnellement à la classe c .

Pour les hyperparamètres μ_K et ρ_σ , nous choisissons des loi *a priori* non informatives impropres de Jeffrey :

$$p(\mu_K | c) \propto \frac{1}{\sqrt{\mu_K}} \mathbb{I}_{\mathbb{R}_+}(\mu_K), \quad (5.22)$$

$$p(\rho_\sigma | c) \propto \frac{1}{\rho_\sigma} \mathbb{I}_{\mathbb{R}_+}(\rho_\sigma). \quad (5.23)$$

Pour les hyperparamètres $\phi_k = \{\rho_{\tau_k}, \mu_{\beta_k}, \rho_{\lambda_k}, \mu_{\alpha_k}, \sigma_{\alpha_k}^2, p_{c_k}\}$, nous choisissons des lois *a priori* conjuguées par rapport à la loi conditionnelle des paramètres $p(\theta | \phi)$. Ceci conduit à introduire des lois inverse-gamma pour les paramètres d'échelle, ou conditionnellement normales pour

les paramètres μ_{α_K} :

$$p(\rho_{\tau_K} | c) = \prod_{l=1}^{K-1} \Delta \frac{e^{-\frac{\Delta}{\rho_{\tau_l, K}}}}{\rho_{\tau_l, K}^2} \mathbb{I}_{\mathbb{R}_+}(\rho_{\tau_l, K}), \quad \forall K > 1, \quad (5.24)$$

$$p(\rho_{\lambda_K} | c) = \prod_{l=1}^K \Delta \frac{e^{-\frac{\Delta}{\rho_{\lambda_l, K}}}}{\rho_{\lambda_l, K}^2} \mathbb{I}_{\mathbb{R}_+}(\rho_{\lambda_l, K}), \quad \forall K > 0, \quad (5.25)$$

$$p(\mu_{\alpha_K}, \sigma_{\alpha_K}^2 | c) = \prod_{l=1}^K \Delta \frac{e^{-\frac{\Delta}{\sigma_{\alpha_l, K}^2}}}{(\sigma_{\alpha_l, K}^2)^2} \frac{e^{-\frac{\mu_{\alpha_l, K}^2}{2\Psi\sigma_{\alpha_l, K}^2}}}{(2\pi\Psi\sigma_{\alpha_l, K}^2)^{\frac{1}{2}}} \mathbb{I}_{\mathbb{R} \times \mathbb{R}_+}(\mu_{\alpha_l, K}, \sigma_{\alpha_l, K}^2), \quad \forall K > 0, \quad (5.26)$$

$$p(p_{c_K} | c) = \prod_{l=1}^K \mathbb{I}_{[0,1]}(c_{l,K}) \quad \forall K > 0, \quad (5.27)$$

avec Δ et Ψ deux paramètres déterministes, suffisamment grands pour que les lois *a priori* des hyperparamètres soient vagues. Par la suite, on fixe $\Delta = 100$ et $\Psi = 100$. Il est important de noter que les lois *a priori* des hyperparamètres ϕ_K ne sont pas impropres. En effet, ces hyperparamètres définissent la loi des paramètres du modèle pour une valeur donnée du nombre de sous-modèles k . Pour une configuration donnée des paramètres des signaux de la base d'apprentissage $(\theta_1, \dots, \theta_{N_c})$, les valeurs de k ne sont évidemment pas toutes représentées. En conséquence, il arrive qu'aucune observation ne vienne appuyer l'estimation de ces hyperparamètres. Dans cette situation, la loi *a posteriori* des hyperparamètres doit tout de même être définie, ce qui n'est pas le cas si leur loi *a priori* est impropre. Au contraire, pour les hyperparamètres μ_K et ρ_σ , qui sont communs à toutes les configurations de transition, nous pouvons choisir des lois *a priori* impropres car leur estimation implique les paramètres de tous les signaux de la base d'apprentissage et s'appuie donc sur autant d'observations.

Les relations entre les différentes variables du modèle hiérarchique proposé sont finalement synthétisées et décrites par le graphe orienté acyclique sur la figure 5.3.

5.2.4 Loi *a posteriori* marginalisée

La loi *a posteriori* qui résulte de cette modélisation dépend d'un grand nombre de paramètres et s'avère trop complexe pour être étudiée analytiquement. De manière classique, les paramètres sont donc échantillonnés à l'aide de méthodes MCMC. Afin d'obtenir un échantillonneur plus efficace pour simuler la configuration des transitions, il est intéressant de réduire la dimension de l'espace des paramètres. Ceci peut être fait en marginalisant par rapport à certains paramètres et hyperparamètres, ce qui permet de définir un échantillonneur de Gibbs partiellement marginalisé.

Lorsque c'est possible, nous intégrons donc les paramètres θ et les hyperparamètres ϕ pour marginaliser la densité de probabilité jointe *a posteriori* (5.10). Il faut noter que même si ces paramètres sont marginalisés, ils pourront toujours être simulés selon leur loi *a posteriori* conditionnelle. Les détails des calculs qui suivent sont présentés dans l'annexe A.5.

Intégration de l'hyperparamètre μ_K

L'hyperparamètre μ_K influence la loi de probabilité jointe par l'intermédiaire des lois *a priori* des paramètres K_i (5.12) et de sa propre loi *a priori* (5.22). A partir de ces deux équations, on

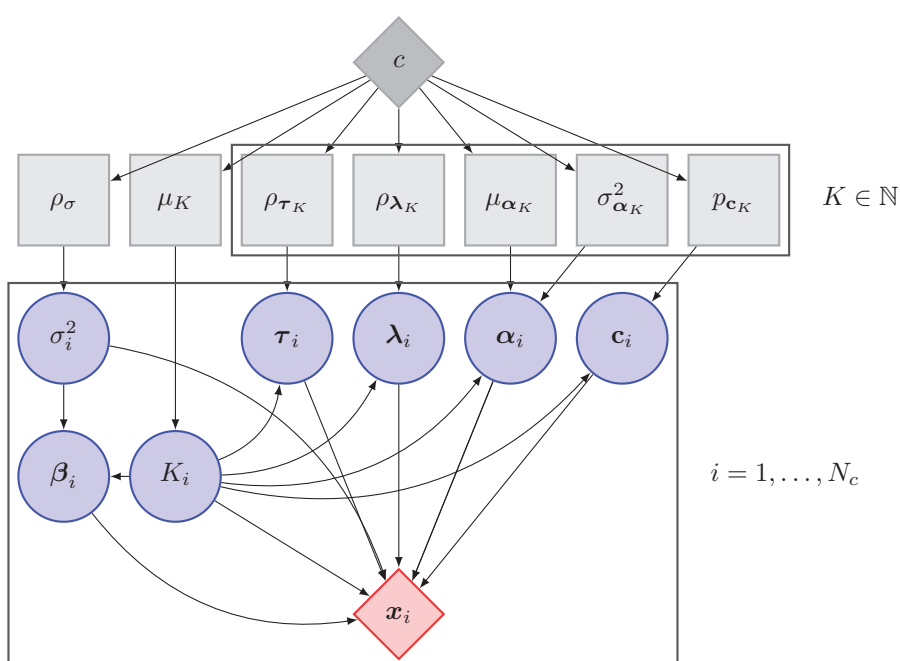


FIGURE 5.3: Graphe orienté acyclique des dépendances conditionnelles du modèle d'apprentissage bayésien hiérarchique. Les nœuds du graphe correspondent aux paramètres (ronds bleus), aux hyperparamètres (rectangles gris), aux signaux d'apprentissage (losange rouge) et à la classe étudiée (losange gris). Chaque arc correspond à une dépendance conditionnelle entre les deux variables. Les rectangles correspondent à une duplication des variables qu'ils contiennent.

exprime la loi marginale *a priori* des K_i :

$$\begin{aligned} p(K_1, \dots, K_{N_c}) &= \int p(\mu_K) \prod_{i=1}^{N_c} p(K_i | \mu_K) d\mu_K \\ &\propto \Gamma\left(\frac{1}{2} + \sum_{i=1}^{N_c} K_i\right) N_c^{-\left(\frac{1}{2} + \sum_{i=1}^{N_c} K_i\right)} \prod_{i=1}^{N_c} \frac{1}{K_i!} \end{aligned} \quad (5.28)$$

Intégration des paramètres des sous-modèles β et de la variance du bruit σ^2

Les paramètres β_i et σ_i^2 interviennent dans la loi *a posteriori* conjointe (5.10) via la vraisemblance du modèle à transitions régulières (3.14) et leur loi *a priori* respective (5.14), (5.13). L'intégration de β_i et σ_i^2 est donc, pour chaque signal $i = 1, \dots, N_c$:

$$\begin{aligned} &\int \int \mathcal{L}(\mathbf{x}_i | \boldsymbol{\theta}_i) p(\beta_i | \sigma_i^2, \boldsymbol{\tau}_i, \boldsymbol{\eta}_i) p(\sigma_i^2 | \rho_\sigma) d\beta_i d\sigma_i^2 \\ &\propto \frac{(\delta^2 + 1)^{-\frac{K_i}{2}}}{\left(\rho_\sigma + \frac{1}{2} (\mathbf{x}_i^T \mathbf{x}_i - \mu_{\beta_i}^T \Sigma_{\beta_i}^{-1} \mu_{\beta_i})\right)^{\frac{n_i}{2} + 1}}, \end{aligned}$$

où μ_{β_i} et Σ_{β_i} sont les paramètres de la distribution *a posteriori* conditionnelle de β_i qui est une loi normale de moyenne μ_{β_i} et de matrice de covariance $\sigma_i^2 \Sigma_{\beta_i}$ tels que :

$$\mu_{\beta_i} = \frac{\delta^2}{1 + \delta^2} [Z_i^T Z_i]^{-1} Z_i^T \mathbf{x}_i, \quad (5.29)$$

$$\Sigma_{\beta_i} = \frac{\delta^2}{1 + \delta^2} (Z_i^T Z_i)^{-1}. \quad (5.30)$$

Intégration des hyperparamètres ρ_{τ_K}

Ces hyperparamètres interviennent dans la loi *a priori* de la position des ruptures $\boldsymbol{\tau}$ (5.16) ainsi que dans leur propre loi *a priori* (5.24). Comme nous l'avons précisé précédemment, la forme particulière de la densité de probabilité des paramètres *a priori* (5.11) implique que les signaux dont le nombre de ruptures K ne sont pas identiques possèdent des paramètres $\boldsymbol{\tau}, \boldsymbol{\eta}, \beta$ indépendants. On peut en fait former des sous-classes de signaux en fonction de la valeur de K . On note ainsi

$$I_c^K = \{i = 1, \dots, N_c : \mathbf{x}_i \in \mathcal{X}_{\text{BDD}_c}, K_i = K\}, \quad (5.31)$$

l'ensemble des indices des signaux \mathbf{x}_i d'une classe c donnée modélisés à l'aide d'un nombre $K_i = K$ ruptures. Le nombre total de ces signaux est noté $N_c^K = \text{card}(I_c^K)$. On calcule donc la loi marginale *a priori* de la position des ruptures sur chaque sous-ensemble I_c^K :

$$\begin{aligned} p[(\boldsymbol{\tau}_i)_{i \in I_c^K}] &= \int p(\rho_{\tau_K}) \prod_{i \in I_c^K} p(\boldsymbol{\tau}_i | \rho_{\tau_K}) d\rho_{\tau_K}, \\ &= \prod_{k=1}^{K-1} \frac{\Delta \Gamma(1 + N_c^K)}{\left(\Delta + \sum_{i \in I_c^K} (\tau_{k+1,i} - \tau_{k,i})\right)^{1 + N_c^K}}, \quad \forall K > 1. \end{aligned} \quad (5.32)$$

Intégration des hyperparamètres ρ_{λ_K}

L'intégration des hyperparamètres ρ_{λ_K} est similaire à l'intégration des hyperparamètres ρ_{τ_K} puisque les loi *a priori* des paramètres d'échelle et des paramètres de position des transitions sont de même nature. A partir des équations (5.17) et (5.25), on calcule la loi marginale *a priori* des paramètres d'échelle des ruptures :

$$\begin{aligned} p[(\lambda_i)_{i \in I_c^K}] &= \int p(\rho_{\lambda_K}) \prod_{i \in I_c^K} p(\lambda_i | \rho_{\lambda_K}) d\rho_{\lambda_K}, \\ &= \prod_{k=1}^K \frac{\Delta \Gamma(1 + N_c^K)}{(\Delta + \sum_{i \in I_c^K} \lambda_{k,i})^{1+N_c^K}}, \quad \forall K > 0. \end{aligned} \quad (5.33)$$

Intégration des hyperparamètres $\mu_{\alpha_K}, \sigma_{\alpha}^2$

Ici encore, les hyperparamètres μ_{α_K} et σ_{α}^2 interviennent dans la loi *a priori* des paramètres de forme des transitions (5.19), ainsi que dans leur propre loi *a priori* (5.26). L'intégration de ces hyperparamètres correspond donc au calcul de la densité marginale *a priori* des paramètres α :

$$\begin{aligned} p[(\alpha_i)_{i \in I_c^K}] &= \int \int p(\mu_{\alpha_K}, \sigma_{\alpha_K}^2) \prod_{i \in I_c^K} p(\alpha_i | \mu_{\alpha_K}, \sigma_{\alpha_K}^2) d\mu_{\alpha_K} d\sigma_{\alpha_K}^2, \\ &= \prod_{k=1}^K \frac{\Delta \Gamma(\frac{N_c^K}{2} + 1)}{\Psi^{\frac{1}{2}}(N_c^K + \frac{1}{\Psi})^{\frac{1}{2}} (2\pi)^{\frac{N_c^K}{2}}} \left[\Delta + \frac{1}{2} \left(\sum_{i \in I_c^K} \alpha_{k,i}^2 - \left(\frac{1}{N_c^K + \frac{1}{\Psi}} \sum_{i \in I_c^K} \alpha_{k,i} \right)^2 \right) \right]^{-\frac{N_c^K}{2} - 1}, \end{aligned} \quad (5.34)$$

pour tout $K > 0$.

Intégration des hyperparamètres p_{c_K}

Finalement, on peut déduire de (5.20) et (5.27) la loi marginale *a priori* des paramètres c :

$$\begin{aligned} p[(c_i)_{i \in I_c^K}] &= \int p(p_{c_K}) \prod_{i \in I_c^K} p(c_i | p_{c_K}) dp_{c_K}, \\ &= \prod_{k=1}^K B \left(1 + \sum_{i \in I_c^K} c_{k,i}, 1 + N_c^K - \sum_{i \in I_c^K} c_{k,i} \right), \quad \forall K > 0. \end{aligned} \quad (5.35)$$

où $B(\cdot, \cdot)$ représente la fonction Bêta.

5.2.5 Loi *a posteriori* marginale

Après intégration des différents paramètres et hyperparamètres, on peut exprimer la loi marginale *a posteriori* qui en résulte :

$$p \left[(K_i, \boldsymbol{\tau}_i, \boldsymbol{\eta}_i)_{i=1, \dots, N}, \rho_\sigma, (\mu_{\beta_K})_{K \in \mathbb{N}} \mid \mathcal{X}_{\text{BDD}_c}, c \right] \propto \frac{1}{\rho_\sigma} \mathbb{I}_{\mathbb{R}_+}(\rho_\sigma) \sum_{K \in \mathbb{N}} p \left[(K_i)_{i \in \mathbb{I}_c^K} \right] p \left[(\boldsymbol{\eta}_i)_{i \in \mathbb{I}_c^K} \right] p \left[(\boldsymbol{\tau}_i)_{i \in \mathbb{I}_c^K} \right] \\ \times \prod_{i \in \mathbb{I}_c^K} \left\{ \frac{(\delta^2 + 1)^{-\frac{K_i+1}{2}}}{\left(\rho_\sigma + \frac{1}{2} (\mathbf{x}_i^T \mathbf{x}_i - \mu_{\beta_i}^T \Sigma_{\beta_i}^{-1} \mu_{\beta_i}) \right)^{\frac{n_i}{2} + 1}} \right\}. \quad (5.36)$$

Il n'est pas possible d'intégrer analytiquement les paramètres et les hyperparamètres restants. Par conséquent, nous adoptons une démarche identique à celle du chapitre 3, consistant à développer un algorithme RJMCMC pour échantillonner la loi marginale *a posteriori* (5.36).

5.2.6 Algorithme RJMCMC pour l'apprentissage des signaux de la base de données

L'algorithme RJMCMC utilisé pour échantillonner la loi (5.36) reprend en grande partie les mouvements de l'algorithme développés pour l'estimation des paramètres du modèle (cf Algorithme 1, p. 66). En fait, à chaque itération de l'algorithme, une modification des paramètres est proposée successivement pour chaque signal $(\mathbf{x}_i)_{i=1, \dots, N_c}$ de la classe étudiée. Le noyau de proposition pour modifier les paramètres d'un signal pris individuellement est exactement le même que les mouvements proposés dans l'algorithme 1 (mouvement de naissance/mort, mouvement de fusion/division, mise à jour des paramètres d'une transition). Par contre, le calcul du taux d'acceptation est évidemment réalisé à partir de la loi marginale *a posteriori* du modèle hiérarchique (5.36).

Après avoir proposé un mouvement pour chacun des signaux $(\mathbf{x}_i)_{i=1, \dots, N_c}$, les hyperparamètres du modèle sont mis à jour. Pour cela, nous utilisons une stratégie d'échantillonnage de Gibbs identique à celle exposée pour l'estimation de l'hyperparamètre δ^2 dans le cadre de l'évaluation des paramètres du modèle à transitions régulières (cf 3.5.7). C'est-à-dire, que nous échantillonons les hyperparamètres selon leur loi conditionnelle, dérivée de la loi jointe *a posteriori* (5.10). Notons que nous échantillonons de cette manière, aussi bien l'hyperparamètre ρ_σ que nous n'avons pas pu intégrer analytiquement, que le reste des hyperparamètres. En effet, ces derniers nous sont utiles, soit, dans le cas des $(\sigma_i^2)_{i=1, \dots, N_c}$ car ils interviennent dans la stratégie d'échantillonnage de Gibbs de ρ_σ , soit, simplement pour caractériser la loi des signaux de la base de données.

L'algorithme est présentée ci-dessous (Algorithme 2) sous forme de pseudo-code et les mouvements de mise à jour des hyperparamètres sont détaillés ci-après. Comme pour l'algorithme 1, les probabilités de tenter chacun des mouvements sont identiques ($p_{\text{naiss}}^{(n)} = p_{\text{mort}}^{(n)} = p_{\text{divi}}^{(n)} = p_{\text{fusion}}^{(n)} = p_{\text{maj}}^{(n)}$), quelque soit l'itération n considérée).

Mise à jour des hyperparamètres Une stratégie de Gibbs est donc utilisée pour mettre à jour les paramètres $(\beta_i)_{i=1, \dots, N_c}$, $(\sigma_i^2)_{i=1, \dots, N_c}$ et pour échantillonner les hyperparamètres ϕ .

1. Les paramètres $(\sigma_i^2)_{i=1, \dots, N_c}$ sont échantillonnés selon leur loi conditionnelle :

$$\sigma_i^2 \mid \mathbf{x}_i, \boldsymbol{\eta}_i, \rho_\sigma, \mu_{\beta_K} \sim \mathcal{IG} \left[\frac{n_i}{2} + 1, \rho_\sigma + \frac{1}{2} (\mathbf{x}_i^T \mathbf{x}_i - \mu_{\beta_i}^T \Sigma_{\beta_i}^{-1} \mu_{\beta_i}) \right], \quad \forall i = 1, \dots, N_c. \quad (5.37)$$

```

Initialisation  $\rho_\sigma = 1, K_i = 0, \tau_i = \eta_i = \emptyset,$ 
for  $n = 1 : N_{\text{iter}}$  do
  for  $i = 1 : N_c$  do
    Échantillonner  $u \sim \mathcal{U}_{[0,1]}$ 
    if  $u < p_{\text{naiss}}^{(n)}$  then
      | Exécuter un mouvement de naissance d'une transition pour le signal  $x_i$  (voir
      | 3.5.3)
    else if  $u < p_{\text{naiss}}^{(n)} + p_{\text{mort}}^{(n)}$  then
      | Exécuter un mouvement de mort d'une transition pour le signal  $x_i$  (voir 3.5.3)
    else if  $u < p_{\text{naiss}}^{(n)} + p_{\text{mort}}^{(n)} + p_{\text{divi}}^{(n)}$  then
      | Exécuter un mouvement de division d'une transition pour le signal  $x_i$  (voir 3.5.4)
    else if  $u < p_{\text{naiss}}^{(n)} + p_{\text{mort}}^{(n)} + p_{\text{divi}}^{(n)} + p_{\text{fusion}}^{(n)}$  then
      | Exécuter un mouvement de fusion de deux transitions pour le signal  $x_i$  (voir
      | 3.5.4)
    else
      | Exécuter un mouvement de mise à jour d'une transition pour le signal  $x_i$  (voir
      | 3.5.5)
    end
  end
  Mettre à jour  $\rho_\sigma$  (voir 5.2.6)
  Échantillonner les hyperparamètres  $\phi_k$  (voir 5.2.6)
end

```

Algorithme 2: Algorithme RJMCMC pour l'apprentissage des signaux de la base de données

2. Les paramètres $(\beta_i)_{i=1,\dots,N_c}$ sont échantillonnés selon leur loi conditionnelle :

$$\beta_i | \mathbf{x}_i, \sigma_i^2, \tau_i, \boldsymbol{\eta}_i \sim \mathcal{N}(\mu_{\beta_i}, \sigma_i \Sigma_{\beta_i}), \quad \forall i = 1, \dots, N_c. \quad (5.38)$$

3. L'hyperparamètre ρ_σ est échantillonné selon sa loi conditionnelle :

$$\rho_\sigma | (\sigma_i)_{i=1,\dots,N_c} \sim \mathcal{G}\left(N_c, \left[\sum_{i=1}^{N_c} \frac{1}{\sigma_i^2}\right]^{-1}\right). \quad (5.39)$$

4. Les hyperparamètres ρ_{τ_K} sont échantillonnés selon leur loi conditionnelle :

$$\rho_{\tau_{k,K}} | (\tau_{k,i})_{i \in \mathbb{I}_c^K} \sim \mathcal{IG}\left(1 + N_c^K, \Delta + \sum_{i \in \mathbb{I}_c^K} \frac{1}{\tau_{k+1,i} - \tau_{k,i}}\right), \quad \forall k = 1, \dots, K-1, \quad \forall K \in \mathbb{N} \quad (5.40)$$

5. Les hyperparamètres ρ_{λ_K} sont échantillonnés selon leur loi conditionnelle :

$$\rho_{\lambda_{k,K}} | (\lambda_{k,i})_{i \in \mathbb{I}_c^K} \sim \mathcal{IG}\left(1 + N_c^K, \Delta + \sum_{i \in \mathbb{I}_c^K} \frac{1}{\lambda_{k,i}}\right), \quad \forall k = 1, \dots, K, \quad \forall K \in \mathbb{N} \quad (5.41)$$

6. Les hyperparamètres $\mu_{\alpha_K}, \sigma_{\alpha_K}^2$ sont échantillonnés selon leur loi conditionnelle après complétion des données :

$$\sigma_{\alpha_{k,K}}^2 | (\alpha_{k,i})_{i \in \mathbb{I}_c^K} \sim \mathcal{IG}\left(\frac{N_c^K}{2} + 1, \Delta + \frac{1}{2} \left(\sum_{i \in \mathbb{I}_c^K} \alpha_{k,i}^2 - \frac{1}{N_c^K + 1} \left(\sum_{i \in \mathbb{I}_c^K} \alpha_{k,i} \right)^2 \right)\right), \quad (5.42)$$

$$\mu_{\alpha_{k,K}} | \sigma_{\alpha_{k,K}}^2, (\alpha_{k,i})_{i \in \mathbb{I}_c^K} \sim \mathcal{N}\left(\frac{1}{N_c^K + \frac{1}{\Psi}} \sum_{i \in \mathbb{I}_c^K} \alpha_{k,i}, \frac{\sigma_{\alpha_{k,K}}^2}{N_c^K + \frac{1}{\Psi}}\right), \quad (5.43)$$

$$\forall k = 1, \dots, K, \quad \forall K \in \mathbb{N}.$$

7. Les hyperparamètres p_{c_K} sont échantillonnés selon leur loi conditionnelle :

$$p_{c_{k,K}} | (c_{k,i})_{i \in \mathbb{I}_c^K} \sim \mathcal{B}\left(1 + \sum_{i \in \mathbb{I}_c^K} c_{k,i}, 1 + N_c^K - \sum_{i \in \mathbb{I}_c^K} c_{k,i}\right), \quad (5.44)$$

où $\mathcal{B}(\cdot, \cdot)$ désigne la loi Bêta avec la paramétrisation standard.

5.2.7 Résultats : apprentissage des signaux transitoires

Pour chacune des 5 classes de signaux c présentes dans la base de données, nous avons divisé en deux parties égales l'ensemble des signaux, afin de former un ensemble d'apprentissage et un ensemble de test. Nous avons simulé $N_{\text{iter}} = 1e6$ itérations de l'algorithme 2 afin de générer autant d'échantillons des hyperparamètres (et également des paramètres de chaque signal). Les $N_{\text{chauffe}} = 1e5$ premiers échantillons correspondant à la phase de chauffe sont mis de côté. L'hyperparamètre δ^2 représentant le rapport signal sur bruit *a priori* est fixé à $15dB$ soit $\delta^2 = 31.6$.

Apprentissage des hyperparamètres ϕ

Pour chaque classe c , la distribution des paramètres de position des ruptures et des paramètres des fonctions de transitions est apprise au travers de la loi *a posteriori* des hyperparamètres ϕ obtenue sur les signaux tests de la classe c :

$$p_{\text{app}}(\phi | \mathcal{X}_{\text{BDD}}, c) = p(\phi | \mathcal{X}_{\text{BDD}}, c).$$

Par la suite, on omet pour des raisons de brièveté la dépendance par rapport à la base de données \mathcal{X}_{BDD} dans les notations des lois apprises :

$$p_{\text{app}}(\phi | c) \equiv p_{\text{app}}(\phi | \mathcal{X}_{\text{BDD}}, c).$$

Afin d'illustrer les résultats obtenus, nous représentons ci-après les densités *a posteriori* empiriques des hyperparamètres obtenues après échantillonnage sur les figures 5.4, 5.5, 5.6, 5.7, et 5.8. Pour chaque classe, nous y représentons ainsi les densités *a posteriori* marginales du nombre de transitions K et de l'hyperparamètre ρ_σ (la moyenne du paramètre de variance du bruit au sein d'une classe). De plus, étant donné la valeur de K présentant la plus forte probabilité marginale, nous présentons la distribution *a posteriori* des hyperparamètres ϕ_K , c'est-à-dire :

- μ_{α_K} : la moyenne des paramètres de forme des fonctions de transitions,
- $\sigma_{\alpha_K}^2$: la variance des paramètres de forme des fonctions de transitions,
- ρ_{λ_K} : la moyenne des paramètres d'échelle des fonctions de transitions,
- p_{c_K} : la probabilité des paramètres déterminant le type des fonctions de transitions.

Notons que ces hyperparamètres sont définis pour chaque valeur possible de K . Nous ne les représentons que pour la valeur de K qui maximise sa loi *a posteriori* marginale, i.e., la valeur de l'estimateur du maximum *a posteriori* marginal, par souci de clarté.

Apprentissage du nombre des ruptures K

Le modèle hiérarchique proposé modélise la distribution du nombre de ruptures au sein de chaque classe par une loi de Poisson de paramètre μ_K . Nous pourrions apprendre la distribution du nombre de ruptures par l'intermédiaire de la distribution de cet hyperparamètre μ_K , comme nous le faisons pour les paramètres de position des ruptures et les paramètres des fonctions de transitions. Cependant cette approche impliquerait que la loi du nombre de ruptures soit positive pour tous les nombres entiers positifs. En pratique, lors de l'échantillonnage des paramètres de chaque signal, toutes les valeurs de K ne sont bien évidemment pas représentées. Or, à chaque valeur de K est associée un sous-ensemble des hyperparamètres ϕ_K . Ainsi, aucune information n'est apportée par l'apprentissage quant à la valeur de la plupart des sous-ensembles d'hyperparamètres ϕ_K , qui correspondent à des valeurs de K non représentées dans la base de données.

Plutôt que d'apprendre la loi de K par le biais des hyperparamètres introduits dans le modèle bayésien hiérarchique comme ceci est fait pour les autres paramètres du modèle, nous choisissons donc d'apprendre la loi de K grâce à sa loi empirique obtenue après échantillonnage. La loi du nombre de ruptures K au sein de chaque classe est donc modélisée par une loi dont le support fini correspond aux valeurs de K échantillonnées par l'algorithme 2, et dont les masses de probabilité sont les fréquences d'apparition de cette valeur pendant l'échantillonnage :

$$p_{\text{app}}(K | c) = p_{\text{emp}}(K | \mathcal{X}_{\text{BDD}c}, c) \equiv \frac{\text{card} \{n = N_{\text{chauffe}} + 1, \dots, N_{\text{iter}} : K^{(n)} = K\}}{N_{\text{iter}} - N_{\text{chauffe}}}.$$

Apprentissage des paramètres β

Le modèle hiérarchique proposé ne permet pas d'évaluer la distribution des paramètres des sous-modèles β_i via un ou plusieurs hyperparamètres spécifiques comme pour les autres caractéristiques des signaux. En effet, nous avons choisi de conserver pour les β une loi *a priori* semblable à celle proposée pour l'estimation des paramètres aux chapitres précédents.

La raison de ce choix est que les paramètres des sous-modèles jouent, via leur distribution *a priori*, un rôle de première importance pour la régularisation du problème d'estimation des paramètres et du nombre de ruptures K_i . Nous avons vu (cf 4.1.4), que le fait de fixer *a priori* l'hyperparamètre δ^2 permettait de fixer *a priori* la qualité de reconstruction attendue du modèle, et donc de régler l'estimation du nombre de sous-modèles K_i en évitant la sur-segmentation des signaux observés. Puisque cette structure permet de gérer simplement le compromis précision/régularisation, nous la conservons pour la loi apprise.

Cependant, il peut être intéressant d'extraire un minimum d'information sur les sous-modèles associés aux différents segments des signaux. Plutôt que de la prendre en compte à travers la distribution des paramètres β , il est possible d'intégrer cette information dans la loi des positions τ des transitions. En effet, dans le cadre du modèle à transitions régulières, les fonctions de transitions sont toujours croissantes de 0 à 1, et ce sont les paramètres des sous-modèles β_i qui déterminent le sens dans lequel varie le signal aux instants de transition. Nous utilisons le sens de ces changements pour apprendre et caractériser les classes de signaux. Pour déterminer le sens de ces changements, nous exploitons le signe du gradient du vecteur des observations \mathbf{x} aux instants de ruptures. Ainsi, le gradient des observations ∇x est approché par :

$$\nabla x(t_j) = \frac{\bar{x}(t_{j+1}) - \bar{x}(t_{j-1})}{2t_e},$$

en chaque instant d'échantillonnage $(t_j)_{j=1,\dots,n}$ où t_e est la période d'échantillonnage du signal. Puis le signe du gradient $S_{k,i}$ associé à la transition τ_k du signal \mathbf{x}_i de la base d'apprentissage est calculé :

$$S_{k,i} = \text{sign}[\nabla x_i(\tau_k)]. \quad (5.45)$$

Pour chaque classe et pour chaque valeur du nombre de ruptures K_i , nous calculons $S_{k,i}$ pour chaque échantillon des paramètres $\theta^{(n)}$ généré par l'algorithme 2. Si, pour un signal donné, un nombre de ruptures donné et une rupture donnée, la grandeur $S_{k,i}$ est toujours égal à 1 (respectivement, toujours égal à -1), alors nous apprenons une contrainte déterministe sur le sens de ce changement de régime. Lors de la classification des signaux tests, tout mouvement proposé amenant à violer l'une de ces contraintes sera systématiquement refusé.

Ceci conduit à rajouter une contrainte dure dans la loi apprise pour chaque classe c :

$$h_{\text{app}}(\tau|c) = \begin{cases} 1 & \text{si toutes les contraintes identifiées sont respectées,} \\ 0 & \text{sinon.} \end{cases} \quad (5.46)$$

Loi apprise sur la base de données

La loi des paramètres et des hyperparamètres apprise sur la base de données peut finalement s'exprimer comme

$$p_{\text{app}}(\theta, \phi|c) \propto p(\theta_{\tilde{K}}|\phi, K)p_{\text{app}}(K|c)h_{\text{app}}(\tau|c)p_{\text{app}}(\phi|c), \quad (5.47)$$

où $\theta = (\tau, \eta, \beta, \sigma^2, K)$ est le vecteur des paramètres alors que $\theta_{\tilde{K}} = (\tau, \eta, \beta, \sigma^2)$ est le vecteur des paramètres hormis K , et où $p(\theta_{\tilde{K}}|\phi, K)$ dérive de la loi *a priori* du modèle de régression à transitions régulières détaillée au paragraphe 5.2.3.

Par conséquent, suite à l'apprentissage effectué, la loi *a posteriori* de tout signal test x s'exprime comme :

$$p_{\text{app}}(\theta, \phi|x, c) \propto p(x|\theta)p_{\text{app}}(\theta, \phi|c), \quad (5.48)$$

où $p(x|\theta)$ est la vraisemblance du modèle de régression à transitions régulières (3.14).

Simulation de signaux synthétiques

A ce stade de la méthode, il est difficile de juger de la qualité des distributions apprises sur chacune des classes. En effet, seules des simulations sur l'ensemble de signaux test nous permettrons de valider la méthode de classification proposée. Cependant, nous pouvons facilement simuler des signaux synthétiques à partir de la distribution des hyperparamètres apprise sur chacune des classes. Il suffit pour cela, de sélectionner aléatoirement l'un des échantillons des hyperparamètres ϕ et d'échantillonner une valeur des paramètres θ selon la loi de densité $p(\theta|\phi)$. Comme c'est expliqué dans le paragraphe consacré à l'apprentissage des paramètres β , cette méthode ne permet pas de simuler de manière précise les paramètres des sous-modèles β_i . Pour pouvoir cependant représenter les signaux simulés, nous leur attribuons arbitrairement une valeur pour les paramètres des sous-modèle, en tirant aléatoirement l'une des valeurs de β_i échantillonnée sur les signaux tests.

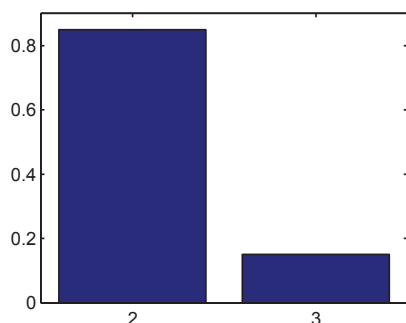
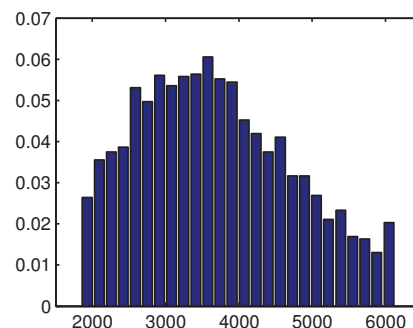
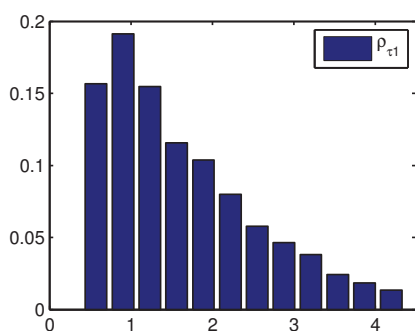
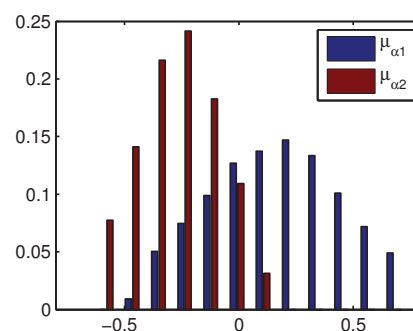
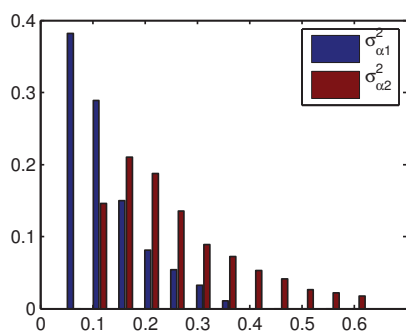
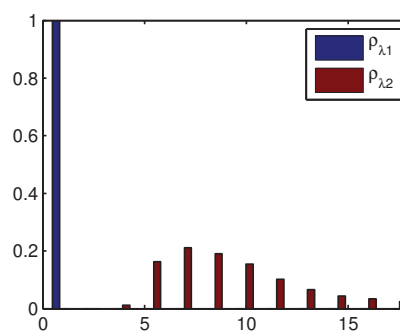
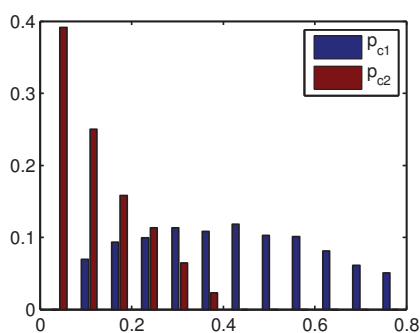
Pour chaque classe, nous simulons ainsi des exemples de signaux, en tirant uniformément l'instant de la première transition τ_1 . Nous traçons les signaux synthétiques ainsi obtenus, figures 5.9(a), 5.9(b), 5.9(c), 5.9(d), 5.9(e). Chaque couleur correspond à une réalisation différente des paramètres. On peut constater sur ces figures, que les signaux simulés sont, pour la plupart proches des signaux de la base d'apprentissage. Certains signaux peuvent cependant sembler assez éloignés de la réalité (par exemple le signal de four micro-ondes bleu). Mais ce type de réalisation est tout à fait cohérent, considérant que l'apprentissage comporte de nombreuses approximations. En particulier, beaucoup de paramètres sont, au moins conditionnellement, indépendants dans notre modèle d'apprentissage.

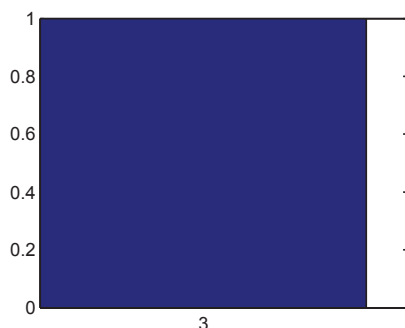
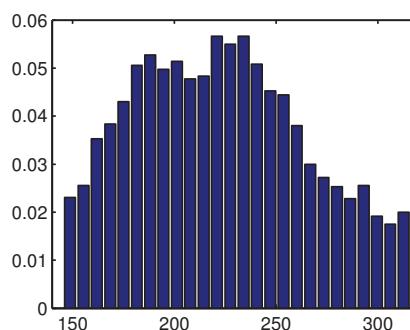
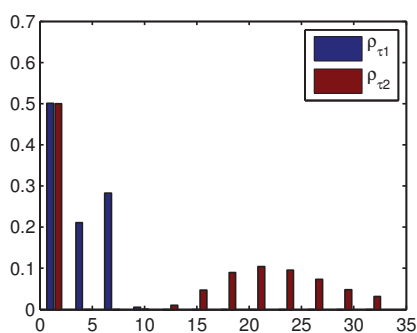
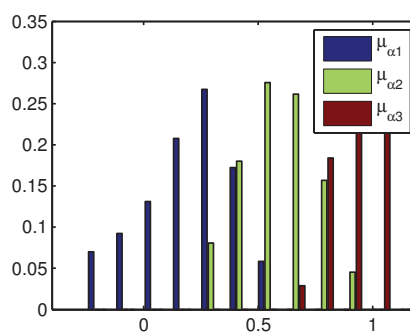
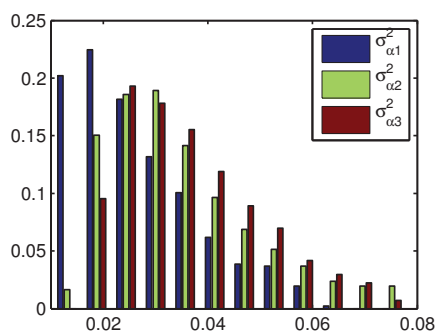
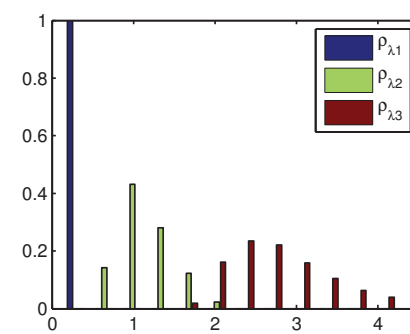
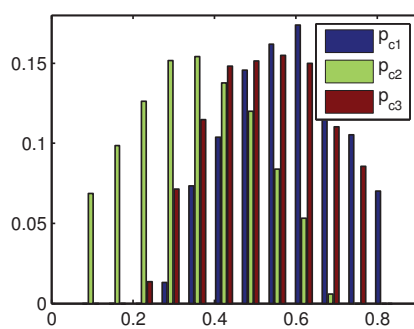
5.2.8 Estimation de la classe d'un signal

La méthode d'apprentissage présentée dans ce chapitre nous permet d'estimer conjointement les caractéristiques θ de chaque signal de la base de donnée, et un ensemble d'hyperparamètres ϕ qui définissent la distribution de ces caractéristiques au sein d'une classe. Pour tester cette méthode dans le cadre du problème de la classification des signaux de la base SISED, il faut maintenant établir un moyen d'attribuer une classe à un signal quelconque x qui n'appartient pas à la base de données d'apprentissage.

Échantillonnage conjoint de la classe et des paramètres

L'idée suivie dans cette thèse afin d'estimer la classe c d'un signal x consiste à formuler le problème de sélection de modèle à partir de la loi *a posteriori* marginale $p(c|x)$. Ceci revient à considérer que le paramètre c qui modélise la classe d'un signal x est maintenant une des variables de la loi *a posteriori*. D'un point de vue algorithmique, il s'agit donc de réaliser un échantillonneur MCMC à la fois sur l'espace des paramètres θ et des hyperparamètres ϕ , mais également sur l'espace des classes de signaux $\mathcal{C} = \{1, \dots, C\}$.

(a) Distribution *a posteriori* de K (b) Distribution *a posteriori* de ρ_σ (c) Distribution *a posteriori* de $\rho_{\tau_K} | K = 2$ (d) Distribution *a posteriori* de $\mu_{\alpha_K} | K = 2$ (e) Distribution *a posteriori* de $\sigma_{\alpha_K}^2 | K = 2$ (f) Distribution *a posteriori* de $\rho_{\lambda_K} | K = 2$ (g) Distribution *a posteriori* de $p_{c_K} | K = 2$ FIGURE 5.4: Distribution *a posteriori* des hyperparamètres de la classe Aspirateur conditionnellement à $K = \hat{K}_{\text{map}} = 2$.

(a) Distribution *a posteriori* de K (b) Distribution *a posteriori* de ρ_σ (c) Distribution *a posteriori* de $\rho_{\tau_K} | K = 3$ (d) Distribution *a posteriori* de $\mu_{\alpha_K} | K = 3$ (e) Distribution *a posteriori* de $\sigma_{\alpha_K}^2 | K = 3$ (f) Distribution *a posteriori* de $\rho_{\lambda_K} | K = 3$ (g) Distribution *a posteriori* de $p_{c_K} | K = 3$ FIGURE 5.5: Distribution *a posteriori* des hyperparamètres de la classe Refrigerateur conditionnellement à $K = \hat{K}_{\text{map}} = 3$.

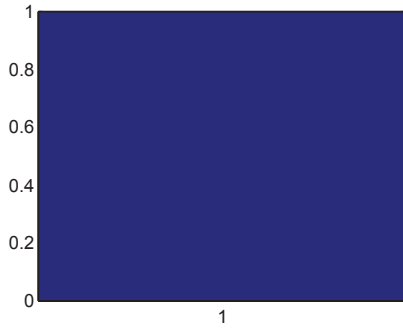
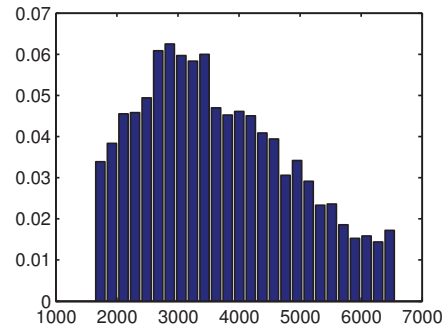
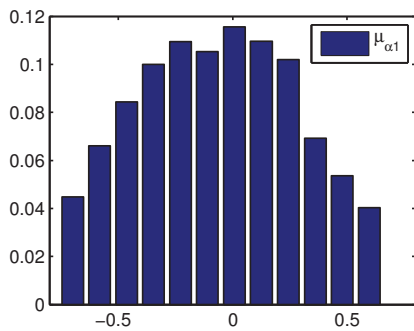
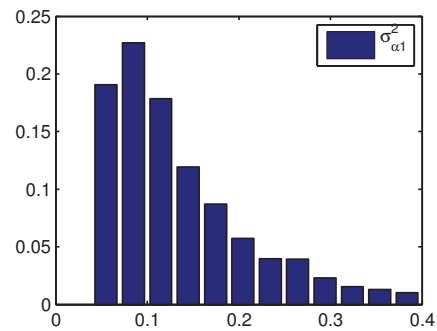
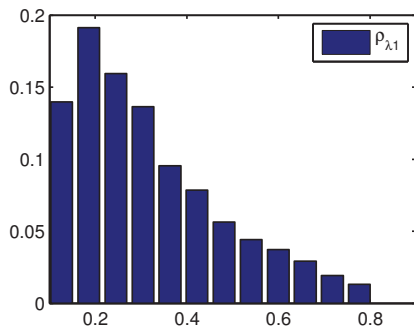
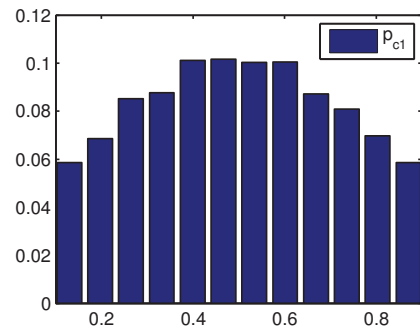
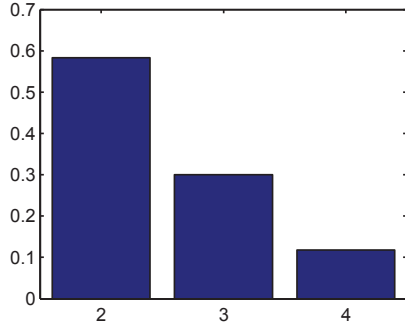
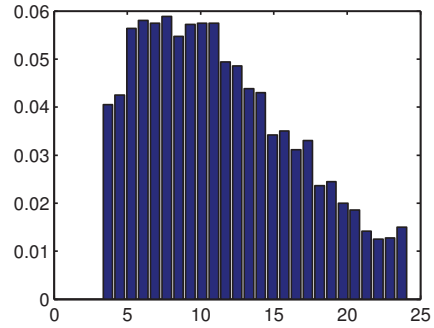
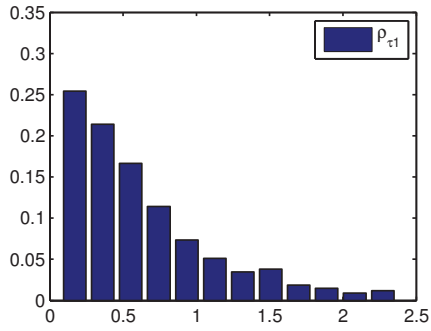
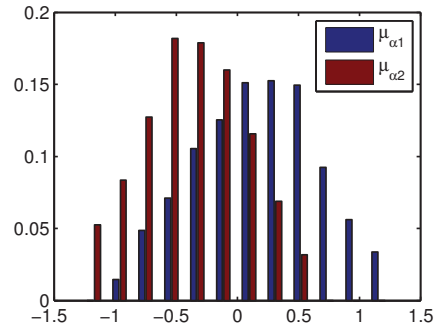
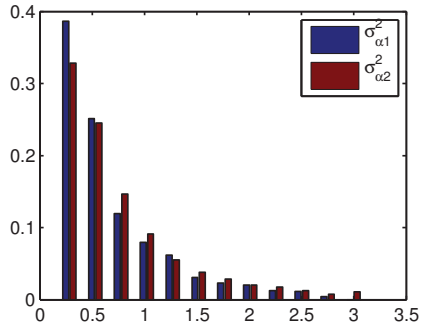
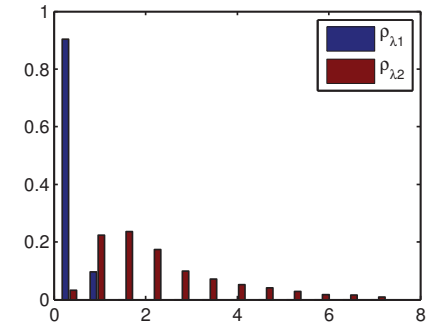
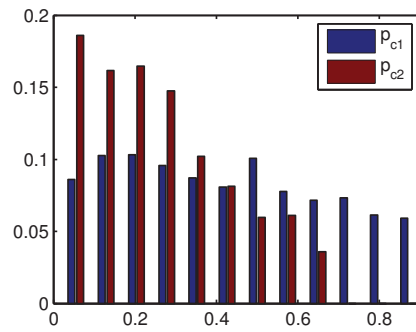
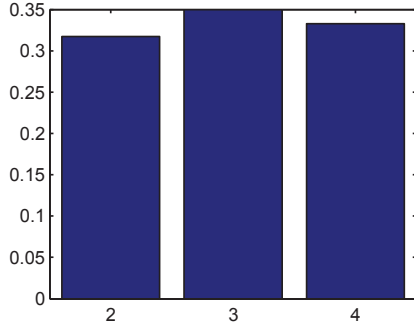
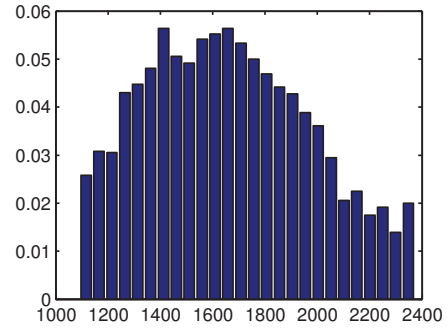
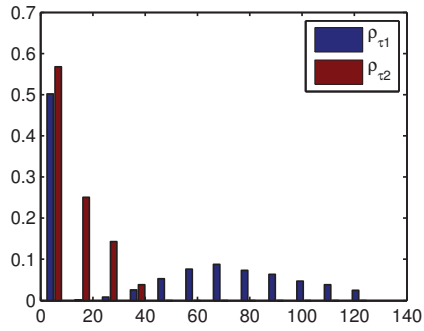
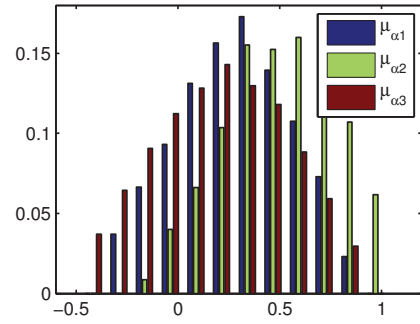
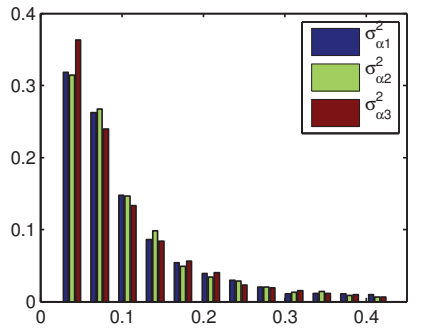
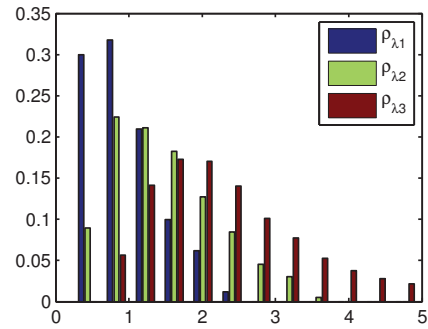
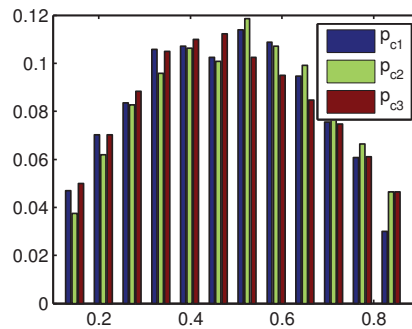
(a) Distribution *a posteriori* de K (b) Distribution *a posteriori* de ρ_σ (c) Distribution *a posteriori* de μ_{α_K} (d) Distribution *a posteriori* de $\sigma_{\alpha_K}^2$ (e) Distribution *a posteriori* de ρ_{λ_K} (f) Distribution *a posteriori* de p_{c1}

FIGURE 5.6: Distribution *a posteriori* des hyperparamètres de la classe Bouilloire conditionnellement à $K = \hat{K}_{\text{map}} = 1$

(a) Distribution *a posteriori* de K (b) Distribution *a posteriori* de ρ_σ (c) Distribution *a posteriori* de $\rho_{\tau_K} | K = 2$ (d) Distribution *a posteriori* de $\mu_{\alpha_K} | K = 2$ (e) Distribution *a posteriori* de $\sigma_{\alpha_K}^2 | K = 2$ (f) Distribution *a posteriori* de $\rho_{\lambda_K} | K = 2$ (g) Distribution *a posteriori* de $p_{c_K} | K = 2$ FIGURE 5.7: Distribution *a posteriori* des hyperparamètres de la classe LampeHalogene conditionnellement à $K = \hat{K}_{\text{map}} = 2$.

(a) Distribution *a posteriori* de K (b) Distribution *a posteriori* de ρ_σ (c) Distribution *a posteriori* de $\rho_{\tau_K} | K = 3$ (d) Distribution *a posteriori* de $\mu_{\alpha_K} | K = 3$ (e) Distribution *a posteriori* de $\sigma_{\alpha_K}^2 | K = 3$ (f) Distribution *a posteriori* de $\rho_{\lambda_K} | K = 3$ (g) Distribution *a posteriori* de $p_{c_K} | K = 3$ FIGURE 5.8: Distribution *a posteriori* des hyperparamètres de la classe MicroOndes conditionnellement à $K = \hat{K}_{\text{map}} = 3$.

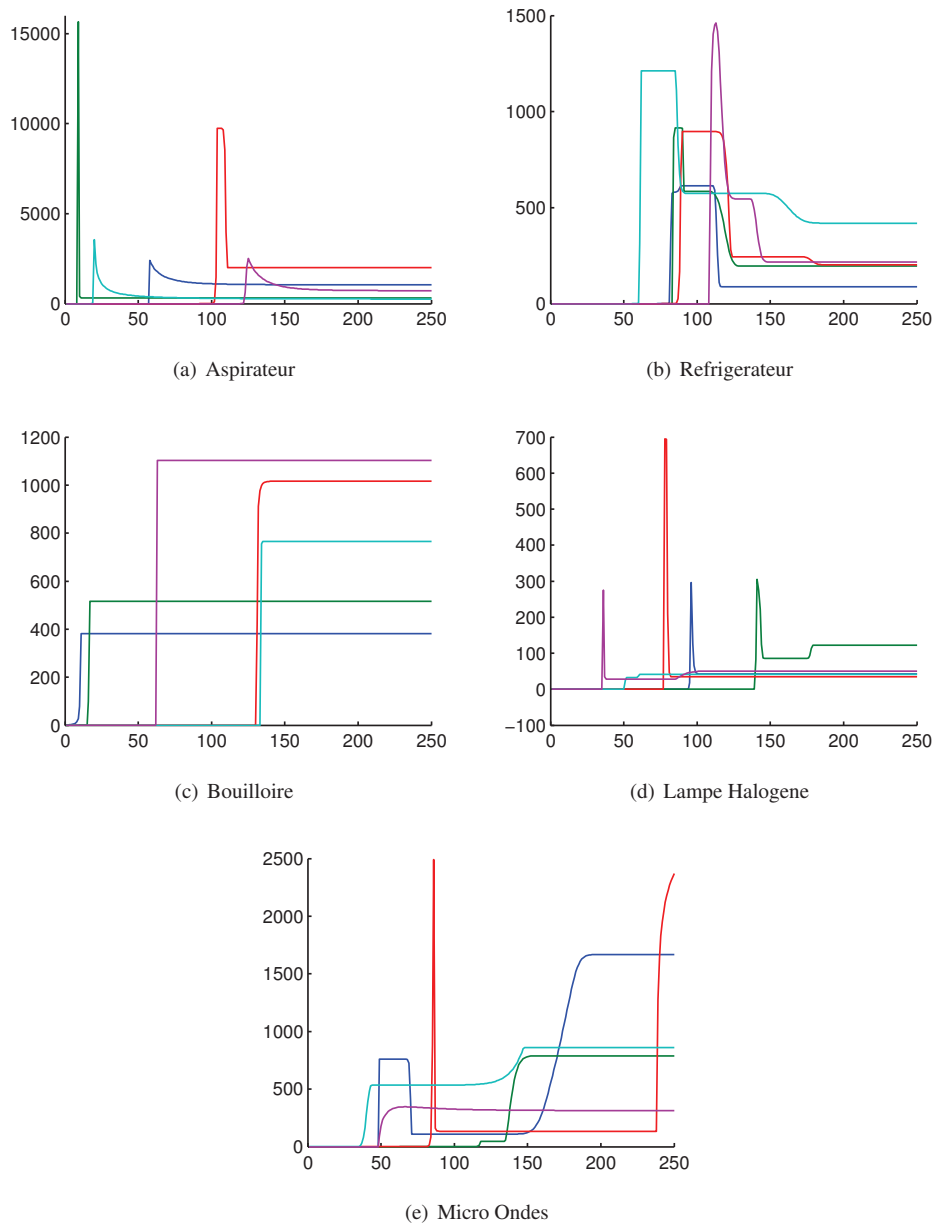


FIGURE 5.9: Exemples de signaux synthétiques tirés selon la distribution apprise de chaque classe.

Une fois la convergence de l'algorithme atteinte, la proportion de fois où l'échantillonneur visite une classe donnée c constitue un estimateur de la probabilité marginale $p(c|\mathbf{x})$ que le signal \mathbf{x} appartienne à la classe c . D'une part, cette approche résout naturellement le problème de la classification multiple en évitant les inconvénients des approches de type test d'hypothèses multiples. D'autre part, elle permet d'obtenir une information plus riche que la seule classification en fournissant toute la loi $p(c|\mathbf{x})$ qui modélise les probabilités des différentes classes pour un signal donné. Cette information peut, par exemple, être utilisée afin de calculer des statistiques de comparaison de modèles telles que les facteurs de Bayes [Han 2001] et interpréter les résultats obtenus.

Loi *a priori* sur les usages c

Dans le cadre de cette thèse, on supposera simplement que cet *a priori* est uniforme : $p(c) = \frac{1}{N_{\text{class}}} \mathbb{I}_C(c)$. Cependant, il serait aisé de prendre en compte des informations *a priori* sur la fréquence d'apparition des transitoires de chaque classe d'usage, pour fixer la densité $p(c)$. On pourrait en particulier se fonder sur les propriétés macroscopiques des usages étudiés pour fixer cette densité *a priori* : par exemple, les transitoires de réfrigérateurs, correspondant à l'enclenchement régulier du compresseur, apparaissent plus fréquemment dans la courbe de charge que les transitoires d'aspirateur, correspondant uniquement au démarrage du moteur de l'appareil.

Loi cible de l'algorithme MCMC pour la sélection de modèle

Pour tout signal \mathbf{x} dont on souhaite estimer la classe c , la loi cible de l'algorithme MCMC est donc la loi *a posteriori*

$$\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, c | \mathbf{x}, \mathcal{X}_{\text{BDD}}) \propto p(\mathbf{x} | \boldsymbol{\theta}) p_{\text{app}}(\boldsymbol{\theta}, \boldsymbol{\phi} | c) p(c). \quad (5.49)$$

où $p(\mathbf{x} | \boldsymbol{\theta})$ est la vraisemblance de \mathbf{x} pour le modèle de régression par transitions régulières paramétré par $\boldsymbol{\theta}$ (3.14), $p_{\text{app}}(\boldsymbol{\theta}, \boldsymbol{\phi} | c)$ est la loi des paramètres et hyperparamètres issue de l'apprentissage sur la base de donnée (5.47), et $p(c)$ désigne la loi *a priori* sur les classes d'usages $c \in \mathcal{C}$.

Échantillonnage de $p_{\text{app}}(\boldsymbol{\phi} | c)$ Les lois des hyperparamètres $p_{\text{app}}(\boldsymbol{\phi} | c)$ ont été échantillonnées pour chaque classe c lors de la phase d'apprentissage. Malheureusement, il n'est pas possible d'obtenir une expression analytique explicite de la loi *a posteriori* jointe de ces hyperparamètres. Cependant, pour simuler cette loi, nous pouvons sélectionner uniformément l'un des échantillons générés lors de la phase d'échantillonnage. Ceci revient à remplacer la loi $p_{\text{app}}(\boldsymbol{\phi} | c)$ par sa loi empirique $p_{\text{emp}}(\boldsymbol{\phi} | \boldsymbol{\phi}^{(N_{\text{chauffe}}+1)}, \dots, \boldsymbol{\phi}^{(N_{\text{iter}})}, c)$ déduite des échantillons générés par l'algorithme MCMC. Cette approche permet d'éviter d'avoir à ajuster une loi paramétrique sur les échantillons générés. Elle se justifie par les propriétés suivantes :

- la loi empirique converge vers la loi théorique $p_{\text{app}}(\boldsymbol{\phi} | c)$ lorsque le nombre d'itérations N_{iter} augmente. C'est une conséquence de la convergence en variation totale de l'algorithme MCMC vers la loi cible,
- la loi du nombre de ruptures K apprise sur la base de données est à support borné supérieurement. Conditionnellement aux valeurs possibles, notées $I_{K,c}$, du nombre de ruptures K , les vecteurs d'hyperparamètres $(\boldsymbol{\phi}_k)_{k \in I_{K,c}}$ sont mutuellement indépendants et de dimensions raisonnables. Il est donc possible d'obtenir une approximation empirique admissible de cette loi à partir d'un nombre raisonnable d'échantillons,
- la loi sur les hyperparamètres $p_{\text{app}}(\boldsymbol{\phi} | c)$ se situe à un niveau d'abstraction élevé de notre modèle hiérarchique. Par conséquent, l'approximation ainsi commise ne doit impacter que très faiblement la loi cible.

Par ailleurs, puisque cette loi empirique est une approximation de la loi cible, cette loi instrumentale peut être identifiée à la loi cible $p_{\text{app}}(\phi|c)$ ce qui conduit à un mouvement de type Gibbs. toujours accepté.

L'algorithme MCMC utilisé pour l'échantillonnage reprend les mouvements utilisés pour l'estimation des paramètres du modèle. Cependant, il convient de définir des mouvements supplémentaires permettant de modifier le paramètre c définissant la classe du signal. L'algorithme est présenté ci-dessous sous forme de pseudo-code (Algorithme 3). Nous ne détaillons pas le mouvement d'échantillonnage des paramètres du modèle qui correspond au même noyau de proposition que celui défini pour l'estimation du modèle (naissance/mort d'une transition, fusion/division de deux transitions et mise à jour simple). Nous nous intéresserons plutôt aux mouvements permettant de changer la classe du signal c . La probabilité p_{chgClass} de tenter un changement de classe c du signal est égale à la probabilité $1 - p_{\text{chgClass}}$ de simplement mettre à jour les paramètres du modèle.

```

Initialisation  $c \sim \mathcal{U}_c, \phi \sim p(\phi|c), \theta \sim p(\theta|\phi)$ 
for  $n = 1 : N_{\text{iter}}$  do
   $u \sim \mathcal{U}_{[0,1]}$ 
  if  $u < p_{\text{chgClass}}$  then
     $w \sim \mathcal{U}_{[0,1]}$ 
    if  $w < p_{\text{chgPar}}$  then
      Exécuter un mouvement de mise à jour de la classe  $c^{(n)}$ 
      Échantillonner les hyperparamètres  $\phi^{(n)} \sim p(\phi|K^{(n)}, c^{(n)})$ 
      Échantillonner les paramètres du signal  $\theta^{(n)} \sim p(\theta|\phi^{(n)})$ 
    else
      Exécuter un mouvement de mise à jour de la classe  $c^{(n)}$ 
      Échantillonner les hyperparamètres  $\phi^{(n)} \sim p(\phi|K^{(n)}, c^{(n)})$ 
    end
  else
    Exécuter un mouvement d'échantillonnage des paramètres du modèle  $\theta^{(n)}$ 
  end
end

```

Algorithme 3: Algorithme RJMCMC pour l'estimation de la classe d'un signal

Mouvement de mise à jour de la classe du modèle c

Le mouvement de mise à jour de la classe du modèle se décompose en deux parties. La première possibilité consiste à échantillonner intégralement un transitoire de classe quelconque :

$$\tilde{c} \sim \mathcal{U}_c, \quad (5.50)$$

$$\tilde{\phi} \sim p(\phi|\tilde{c}), \quad (5.51)$$

$$\tilde{\theta} \sim p(\theta|\tilde{\phi}), \quad (5.52)$$

puis à tenter de remplacer les valeurs courantes des grandeurs échantillonnées $c^{(n)}, \phi^{(n)}, \theta^{(n)}$ par les valeurs échantillonnées $\tilde{c}, \tilde{\phi}, \tilde{\theta}$.

Ce mouvement correspond en fait à la combinaison d'un mouvement de mort de chacune des ruptures de $\theta^{(n)}$ et d'un mouvement de naissance de chacune des ruptures de $\tilde{\theta}$. Le ratio de Green

d'un tel mouvement s'écrit donc :

$$r_1 = \frac{\pi(\tilde{\theta}, \tilde{\phi}, \tilde{c})q(\tilde{u})}{\pi(\theta^{(n)}, \phi^{(n)}, c^{(n)})q(u^{(n)})}, \quad (5.53)$$

où les variables auxiliaires sont respectivement $\tilde{u} = (\tilde{\theta}^{(n)}, \tilde{\phi}^{(n)}, \tilde{c}^{(n)})$ et $u^{(n)} = (\theta^{(n)}, \phi^{(n)}, c^{(n)})$. La loi de proposition q se décline en :

$$q(\theta, \phi, c) = p(\theta|\phi)p(\phi|c)p(c). \quad (5.54)$$

Compte-tenu de l'expression de la loi cible (5.49), ce ratio se simplifie :

$$r_1 = \frac{p(\mathbf{x}|\tilde{\theta})}{p(\mathbf{x}|\theta^{(i)})}. \quad (5.55)$$

La seconde possibilité consiste à tenter de modifier la classe c et les hyperparamètres ϕ les paramètres du signal restant inchangés $\theta = \tilde{\theta} = \theta^{(n)}$:

$$\tilde{c} \sim \mathcal{U}_C, \quad (5.56)$$

$$\tilde{\phi} \sim p(\phi|\tilde{c}). \quad (5.57)$$

Les variables auxiliaires sont alors respectivement $\tilde{u} = (\phi^{(n)}, c^{(n)})$ et $u^{(n)} = (\tilde{\phi}, \tilde{c})$. Le ratio de Green r_2 de ce mouvement se simplifie alors en :

$$r_2 = \frac{p(\mathbf{x}|\tilde{\theta})p(\tilde{\theta}|\tilde{\phi})}{p(\mathbf{x}|\theta^{(n)})p(\theta^{(n)}|\phi^{(n)})} = \frac{p(\theta|\tilde{\phi})}{p(\theta|\phi^{(n)})}. \quad (5.58)$$

L'intérêt de cette seconde possibilité apparaît dans le cas où la valeur des paramètres du signal est proche d'un optimum (en général grâce aux mouvements d'échantillonnage des paramètres du modèle θ) mais que la classe du signal n'a pas été modifiée en conséquence. Cela permet d'échantillonner efficacement la densité de probabilité conditionnelle $p(c, \phi|\mathbf{x}, \theta)$.

5.2.9 Résultats : estimation de la classe des signaux tests

Dans cette partie nous présentons les résultats obtenus en utilisant l'algorithme 3 pour estimer la classe des signaux tests laissé de côté durant la phase d'apprentissage. Pour chaque signal, l'algorithme 3 a été utilisé pour échantillonner $N_{\text{iter}} = 1e4$ valeurs des paramètres, des hyperparamètres et de la classe du signal. Les $N_{\text{chauffe}} = 1e3$ premiers échantillons correspondant à la période de chauffe sont mis de côté. Nous calculons alors la distribution *a posteriori* du paramètre c pour chaque signal. Pour chaque classe, nous présentons ces résultats sous forme de tableau, chaque colonne correspondant à un signal test, chaque ligne correspondant à une classe. Pour chaque signal (colonne), la valeur du paramètre c ayant la plus forte probabilité *a posteriori* est écrite en gras. La ligne correspondant à cette valeur correspond à l'estimateur (5.1) de la classe du signal.

De plus, pour chaque classe nous présentons respectivement le signal test \mathbf{x} dont la probabilité *a posteriori* $p(\mathbf{x}|c)$ associée à sa propre classe c est la plus importante (le signal le mieux classé) et le signal test \mathbf{x} dont la probabilité *a posteriori* $p(\mathbf{x}|c)$ associée à sa propre classe c est la plus faible (le signal le moins bien classé). Notons que dans certains cas le signal « le moins bien classé » peut être correctement classé si tous les signaux tests de sa classe sont également correctement classés. Il correspond alors au signal pour lequel la classification est la plus incertaine. Avec chacun de ces signaux, nous présentons le modèle à transitions régulières tracé pour la valeur des paramètres θ

	1	2	3	4	5	6	7	8
Aspirateur	49	42	50	33	51	28	39	46
Refrigerateur	26	23	18	32	30	3	6	23
Bouilloire	0	0	0	0	0	8	1	0
LampeHalogene	24	32	30	34	18	24	18	31
MicroOndes	0	3	2	1	2	37	35	0

TABLE 5.1: Résultats de la classification des signaux tests de la classe Aspirateur. Les lignes correspondent aux classes, les colonnes aux signaux tests. L'intersection ligne/colonne donne la probabilité estimée que le signal test appartienne à la classe correspondante.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Aspirateur	13	14	13	19	17	14	19	13	10	10	17	13	10	11	11	11	14	13
Refrigerateur	67	72	73	66	71	70	65	69	71	71	61	67	68	74	43	40	29	66
Bouilloire	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LampeHalogene	16	13	12	14	11	12	13	17	16	14	20	19	19	13	45	46	54	14
MicroOndes	4	1	1	2	1	4	3	2	3	5	2	2	3	3	1	2	3	8

TABLE 5.2: Résultats de la classification des signaux tests de la classe Réfrigérateur. Les lignes correspondent aux classes, les colonnes aux signaux tests. L'intersection ligne/colonne donne la probabilité estimée que le signal test appartienne à la classe correspondante.

estimée au sens du maximum *a posteriori*, conditionnellement à sa propre classe. Ces résultats sont présentés sur les figures 5.10, 5.11, 5.12, 5.13 et 5.14.

Les résultats obtenus sont corrects sur l'ensemble des classes. Pour la classe Aspirateur, seuls deux signaux tests sont mal classés par l'algorithme (cf tableau 5.1). Cependant, les signaux de cette classe présentent une probabilité importante d'appartenir à la classe Réfrigérateur et Lampe Halogène. De la même manière, les signaux de la classe Réfrigérateur sont bien classés à l'exception de trois signaux tests (cf tableau 5.2). Là encore, la probabilité pour ces signaux d'appartenir à la classe Aspirateur ou Lampe Halogène est non négligeable, bien que cette tendance soit moins forte que pour les signaux de la classe Aspirateur. Les signaux de la classe Bouilloire ne présentent pas de difficulté à être correctement reconnus (cf tableau 5.3). Les signaux tests de la classe Lampe halogène étant très peu nombreux, il n'est pas évident de tirer des conclusions quant à la viabilité de la méthode pour classifier ces signaux (cf tableau 5.4) Enfin les signaux de la classe Micro-Ondes sont tous correctement classifiés avec des probabilités supérieures à 65% (cf tableau tab :classifMicroOndes). Ce résultat implique qu'au regard du modèle proposé pour extraire les caractéristiques des signaux, la classe Micro-Ondes est bien distincte des autres classes de signaux.

5.2.10 Validation croisée

Afin d'estimer la fiabilité de la méthode d'apprentissage proposée, nous mettons en place une procédure de validation croisée, de type *k-fold*, des modèles appris. Pour chaque classe, nous constituons une partition de $k = 5$ sous-ensembles de taille égale (à une unité près) des signaux de la base de données SISED. Pour chaque sous-ensemble, l'apprentissage est réalisé en utilisant les 4 sous-ensembles restants. Puis la classe des signaux du sous-ensemble considéré est estimée. Nous ne retenons pour chaque signal que l'estimé au sens du maximum *a posteriori* de sa classe. Les résultats combinés pour chacun des sous-ensembles de test sont reportés table 5.6 sous forme de

	1	2	3	4	5	6
Aspirateur	11	14	14	14	13	11
Refrigerateur	2	2	3	4	3	2
Bouilloire	73	69	68	67	69	72
LampeHalogene	13	13	14	14	15	14
MicroOndes	1	1	1	1	1	1

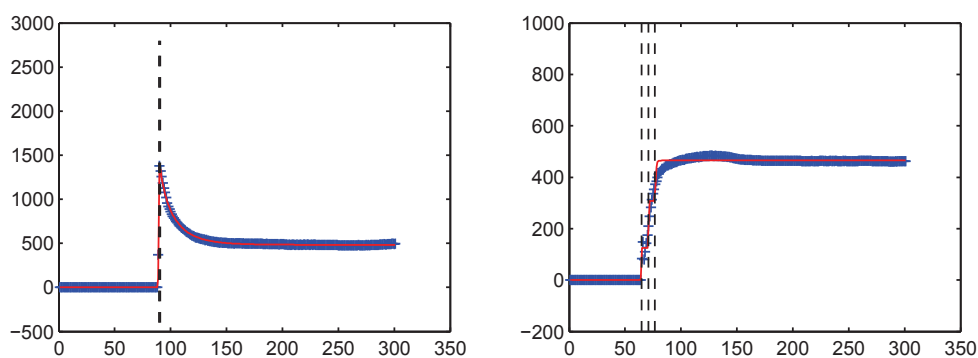
TABLE 5.3: Résultats de la classification des signaux tests de la classe Bouilloire. Les lignes correspondent aux classes, les colonnes aux signaux tests. L'intersection ligne/colonne donne la probabilité estimée que le signal test appartienne à la classe correspondante.

	1	2
Aspirateur	15	13
Refrigerateur	6	16
Bouilloire	7	1
LampeHalogene	39	24
MicroOndes	32	47

TABLE 5.4: Résultats de la classification des signaux tests de la classe Lampe halogène. Les lignes correspondent aux classes, les colonnes aux signaux tests. L'intersection ligne/colonne donne la probabilité estimée que le signal test appartienne à la classe correspondante.

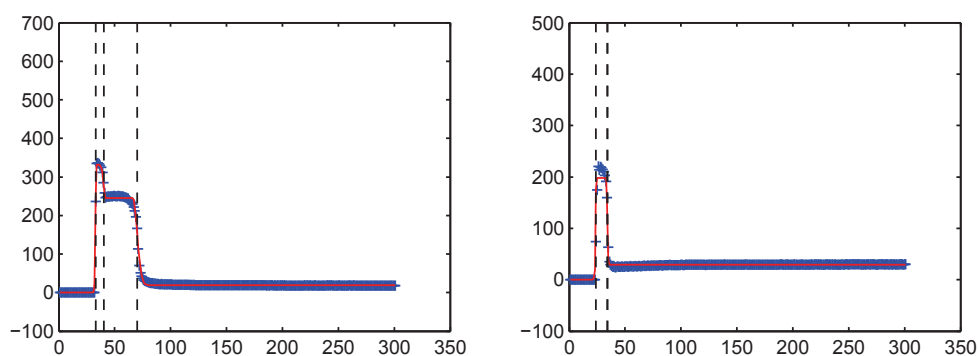
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Aspirateur	15	11	11	7	7	7	6	7	7	7	8	7	9	7	12	13	9
Refrigerateur	2	2	3	0	0	0	1	0	0	2	1	2	0	1	0	0	1
Bouilloire	7	1	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0
LampeHalogene	10	7	7	4	1	2	4	3	3	4	2	3	4	5	4	4	4
MicroOndes	65	80	78	90	92	91	89	91	90	88	88	88	87	87	84	82	87

TABLE 5.5: Résultats de la classification des signaux tests de la classe Four micro-ondes. Les lignes correspondent aux classes, les colonnes aux signaux tests. L'intersection ligne/colonne donne la probabilité estimée que le signal test appartienne à la classe correspondante.



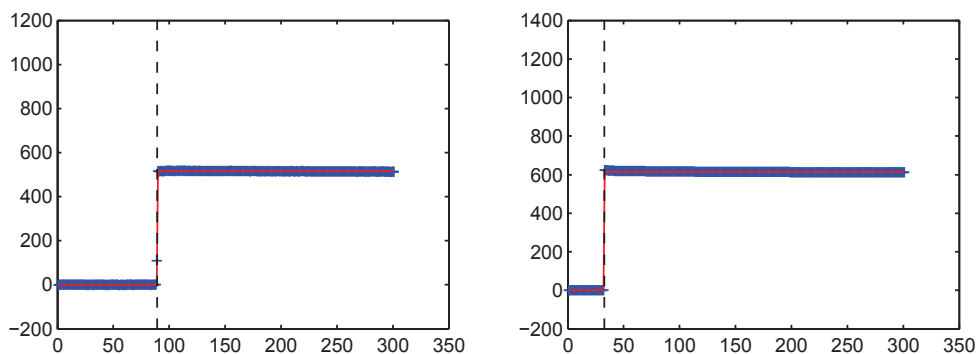
(a) Signal test n°5 de la classe Aspirateur. $P_{\text{Aspirateur}} = 51\%$
 (b) Signal test n°6 de la classe Aspirateur. $P_{\text{Aspirateur}} = 28\%$

FIGURE 5.10: Signaux tests le mieux classé (a), et le moins bien classé (b), de la classe Aspirateur. Signal original (bleu) Modèle estimé conditionnement à la classe Aspirateur (rouge) Position des transitions (pointillés verticaux).



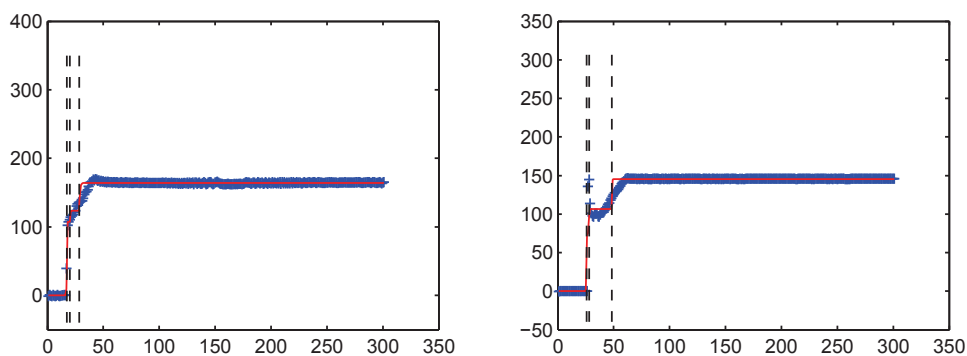
(a) Signal test n°14 de la classe Refrigerateur. $P_{\text{Refrigerateur}} = 74\%$
 (b) Signal test n°17 de la classe Refrigerateur. $P_{\text{Refrigerateur}} = 29\%$

FIGURE 5.11: Signaux tests le mieux classé (a), et le moins bien classé (b), de la classe Refrigerateur. Signal original (bleu) Modèle estimé conditionnement à la classe Refrigerateur (rouge) Position des transitions (pointillés verticaux).



(a) Signal test n° 1 de la classe Bouilloire. $P^{\text{Bouilloire}} = 73\%$ (b) Signal test n° 4 de la classe Bouilloire. $P^{\text{Bouilloire}} = 67\%$

FIGURE 5.12: Signaux tests le mieux classé (a), et le moins bien classé (b), de la classe Bouilloire. Signal original (bleu) Modèle estimé conditionnement à la classe Bouilloire (rouge) Position des transitions (pointillés verticaux).



(a) Signal test n° 1 de la classe LampeHalogene. $P^{\text{LampeHalogene}} = 39\%$ (b) Signal test n° 2 de la classe LampeHalogene. $P^{\text{LampeHalogene}} = 24\%$

FIGURE 5.13: Signaux tests le mieux classé (a), et le moins bien classé (b), de la classe LampeHalogene. Signal original (bleu) Modèle estimé conditionnement à la classe LampeHalogene (rouge) Position des transitions (pointillés verticaux).

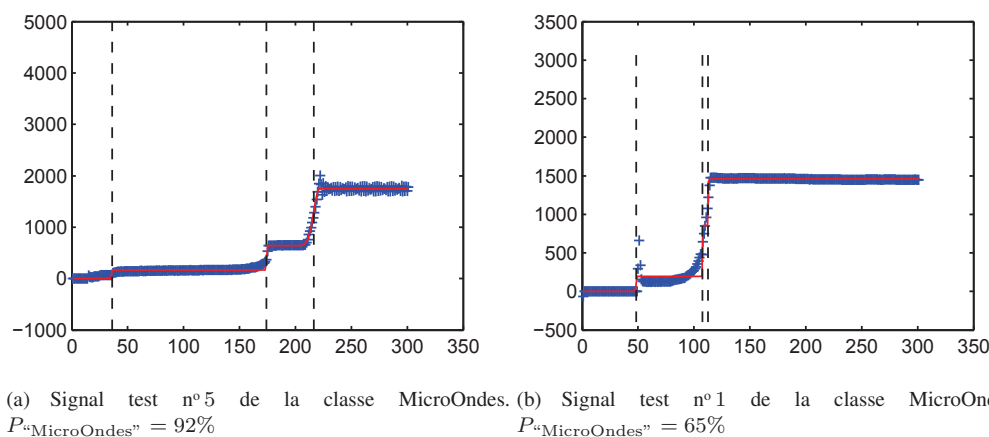


FIGURE 5.14: Signaux tests le mieux classé (a), et le moins bien classé (b), de la classe MicroOndes. Signal original (bleu) Modèle estimé conditionnement à la classe MicroOndes (rouge) Position des transitions (pointillés verticaux).

		Classe estimée				
		Aspirateur	Réfrigérateur	Bouilloire	Lampe Halogène	Micro-ondes
Classe Réelle	Aspirateur	10	0	0	6	0
	Réfrigérateur	0	31	0	5	0
	Bouilloire	0	0	11	1	0
	Lampe Halogène	0	1	0	4	0
	Micro-ondes	0	0	0	0	35

TABLE 5.6: Matrice de confusion de la classification des signaux transitoires obtenue lors de la validation croisée

matrice de confusion. Chaque colonne de la matrice correspond à la classe estimée des signaux, chaque ligne à la classe réelle.

On constate qu'une vaste majorité des signaux sont alors correctement classifiés. Seuls 13 signaux sont mal classifiés, la base SISED comprenant 104 signaux. Cela représente donc un taux de classification de 87,5%. Ce taux s'avère très satisfaisant pour une classification réalisée dans un espace de 5 classes. De plus les résultats se révèlent d'autant meilleurs si l'on remarque que les erreurs de classification résultent toutes d'une confusion avec la classe « Lampe halogène ». Or cette classe est à la fois celle pour laquelle le moins de signaux tests sont disponibles, et la classe pour laquelle les caractéristiques de ces signaux sont les moins homogènes. La classification est pénalisée par cette classe pour laquelle on ne dispose pas de suffisamment d'information dans la base de données.

5.3 Conclusions

Dans ce chapitre, nous avons proposé une méthode d'apprentissage basée sur un modèle bayésien hiérarchique. Cette méthode nous a permis de réaliser conjointement l'extraction des caractéristiques (c-à-d l'estimation des paramètres du modèle à transitions régulières) et l'apprentissage

de la distribution de ces caractéristiques au sein de chaque classe. A partir de cette modélisation, nous avons développé un algorithme RJMCMC pour échantillonner, pour chaque classe, les hyperparamètres qui gouvernent la distribution des caractéristiques. Enfin, nous avons mis en place un algorithme RJMCMC permettant d'estimer la classe d'un signal observé.

Par la suite, nous avons mené des simulations du fonctionnement de la méthode proposée. Dans un premier temps, nous avons découpé en deux la base de données de signaux pour former un ensemble d'apprentissage et un ensemble de test. Dans un second temps, nous avons découpé la base de données en $k = 5$ sous-ensembles et réalisé une validation croisée de la méthode sur cette partition. Dans les deux cas, la classe des signaux tests est correctement estimée pour une majorité de signaux. Le taux de bonne classification total pour la validation croisée est de 87.5%.

Ces bons résultats sont à nuancer au vu de la taille réduite de la base de données et des disparités entre le nombre d'échantillons de chaque classe. Par conséquent, cette base peut s'avérer biaisée et n'est peut être pas suffisamment représentative. Les résultats présentés suggèrent que la méthode proposée permet de classifier efficacement les signaux transitoires électro-domestiques de la base SISED. Mais le travail présenté dans cette thèse doit être vu plus comme une étude de faisabilité que comme une étude de la performance réelle de cette méthode.

Dans le cadre de cette thèse, on suppose que le signal observé est généré par un et un seul signal transitoire. Cela constitue une limite de notre approche. Il serait utile de généraliser la méthode proposée pour réaliser l'identification d'un mélange de signaux transitoires. On supposerait alors que le signal observé peut-être généré par un nombre quelconque de transitoires (éventuellement aucun). Une approche possible serait de porter les algorithmes précédents à un degré d'abstraction supplémentaire : au lieu de manipuler des transitions, l'objet de l'algorithme serait d'échantillonner une séquence de transitoires, eux-mêmes constitués d'une ou plusieurs transitions.

Conclusions et perspectives

L'objectif de cette thèse était de fournir des outils pour la détection et la classification de signaux transitoires électro-domestiques. Ces signaux sont spécifiques aux usages électriques individuels qui composent la courbe de charge globale d'un réseau électrique domestique. Ainsi, l'analyse de ce type de signaux permet d'identifier l'usage qui en est la cause. L'intérêt d'obtenir ce type d'informations est de favoriser une gestion "intelligente" de l'énergie électrique de la part des différents acteurs du marché de l'énergie. En particulier, cette problématique s'inscrit dans le développement de la nouvelle génération de réseaux électriques "intelligents" (« *smart-grid* »).

Dans le chapitre 1 de cette thèse, nous avons développé les tenants et aboutissants de cette problématique, en la replaçant dans son contexte scientifique et industriel. Le chapitre fournit une analyse détaillée des signaux transitoires électro-domestiques de la base de données SISED, qui a servi de fondement aux travaux de cette thèse. Nos premières analyses de ces signaux nous ont conduit à définir un modèle de régression à transitions régulières, détaillé dans le chapitre 3. Ce modèle présente la particularité de découper le signal observé en plusieurs segments. Aussi, nous avons employé une démarche d'inférence bayésienne combiné à un algorithme MCMC pour estimer conjointement le nombre de segments et les paramètres du modèle. Considérant que ce modèle n'est pas restreint à l'application pour laquelle il a été défini, nous l'avons inscrit dans un cadre plus général de régression et lissage de signaux. Le chapitre 3 se conclut donc sur une étude comparative du modèle développé avec d'autres méthodes pour le lissage de signaux standards. Cette étude montre que, pour certains signaux, l'approche proposée se compare favorablement aux autres méthodes. Le modèle de régression à transitions régulières a ensuite été employé pour la modélisation des signaux de la base de données SISED. Les résultats obtenus, rapportés dans le chapitre 4, montrent que le modèle proposé représente efficacement et avec un nombre réduit de paramètres, les signaux de la base SISED. Enfin, dans le chapitre 5, nous nous sommes consacrés à utiliser les paramètres du modèle à transitions régulières pour l'apprentissage et la classification des signaux étudiés. Dans la continuité méthodologique des chapitres précédents, nous nous sommes intéressés aux méthodes d'apprentissage bayésiennes. Considérant qu'il était alors nécessaire de définir un modèle bayésien pour la distribution des paramètres au sein de chaque classe, et que nous avons déjà défini un modèle bayésien pour l'estimation des paramètres à partir des signaux observés, nous avons proposé l'utilisation d'un modèle bayésien hiérarchique pour résoudre conjointement ces deux problèmes. Enfin, la méthode développée est évaluée par validation croisée sur la base de données SISED. Les résultats obtenus, qui closent le chapitre 5, atteignent un taux de bonne classification de 87.5%, les erreurs de classification étant intégralement dues à la classe la plus disparate et la moins représentée dans la base SISED.

Contributions

La première contribution de cette thèse est le développement d'un modèle original de régression à transitions régulières. Bien que ce modèle a été développé initialement pour la représentation de la puissance active des transitoires électro-domestiques, il est adaptable à d'autres applications, la nature des fonctions de transition et des sous-modèles pouvant être modifiée en conséquence. Ce modèle est inspiré d'approches utilisées en économétrie [van Dijk 2002] et de travaux précédents dans le même domaine applicatif que cette thèse [Shaw 2008]. Cependant la méthode proposée étend ces méthodes en permettant l'introduction de n'importe quel type de sous-modèles linéaires

et de fonctions de transitions variées. De plus, l'approche bayésienne utilisée dans cette thèse permet l'estimation jointe du nombre de sous-modèles et des paramètres du modèle.

La seconde contribution est l'utilisation d'un modèle bayésien hiérarchique afin d'estimer conjointement les paramètres du modèle à transitions régulières et réaliser l'apprentissage statistique supervisé de différentes classes de signaux. Ce type d'approche est similaire à celle proposée dans [Davy 2002], cependant la méthodologie présentée dans cette thèse permet de généraliser cette approche au cas où l'espace des caractéristiques (c'est-à-dire l'espace des paramètres du modèle de régression) est à dimension variable.

Perspectives

Sur le plan méthodologique, cette thèse a abouti à une méthode d'extraction de caractéristiques et d'apprentissage conjoint dans un espace de caractéristiques à dimension variable. L'approche proposée possède deux limitations principales. La première limitation est que nous avons choisi de représenter la distribution de chaque caractéristique au sein d'une classe par des mélanges de lois paramétriques de dimension finie. Ce choix présente l'avantage de limiter le phénomène de sur-apprentissage car la distribution des caractéristiques au sein de chaque classe est alors peu flexible et ne risque pas de sur-ajuster les exemples appris. Cette stratégie est pertinente lorsque la base de données est de taille réduite, comme dans le cadre de cette thèse. Cependant, pour des ensembles d'apprentissage plus grand, il sera sans doute nécessaire d'augmenter la flexibilité de ces distributions. Une piste intéressante serait alors l'emploi de méthodes bayésiennes non-paramétriques, afin que le nombre de paramètres définissant la distribution des caractéristique augmente avec le nombre d'exemples appris. Cette piste permettrait en outre de mieux exploiter le modèle de régression à transitions régulières utilisé pour représenter chaque signal, ce modèle s'inscrivant déjà dans un cadre bayésien non-paramétrique. En effet, les paramètres des transitions, qui sont modélisés comme les réalisations d'un processus ponctuel marqué, appartiennent à l'espace des configurations finies de points, espace de dimension infinie.

La seconde limitation de notre approche est que, pour gérer l'apprentissage des caractéristiques dans un espace à dimension variable, nous avons choisi de découpler les distributions conditionnelles des caractéristiques sachant leur nombre. Ce choix peut s'avérer limitant car, en pratique, il peut exister des dépendances entre chaque sous-espace au sein des exemples appris. L'exploitation de ces dépendances est un problème ardu, lié aux problèmes d'interprétation des chaînes de Markov à dimension variables. Notons que des premières approches permettant de s'affranchir de ce problème ont été récemment proposées. Par exemple dans [Roodaki 2012], une approche a été proposée afin d'interpréter *a posteriori* une chaîne de Markov déjà générée dans un espace de dimension variable. De telles approches pourraient maintenant être utilisées lors de l'échantillonnage de la chaîne de Markov, conjointement à la méthodologie d'apprentissage bayésien hiérarchique.

Sur le plan industriel, cette thèse a abouti à une méthode d'apprentissage et de classification des classes de signaux transitoires électro-domestiques présentes dans la base de données SISED. Les résultats obtenus constituent plus une étude de faisabilité qu'une étude de performance de la méthode proposée. D'une part, nous nous sommes limités à 5 classes de signaux, d'autre part le nombre de signaux étudiés était trop faible pour garantir la fiabilité des résultats obtenus dans un contexte plus large. Ainsi, il serait pertinent de tester la méthode proposée sur une base de donnée plus grande et pour de nouvelles classes d'usages. Malheureusement, les bases de données publiques disponibles à l'heure actuelle sont plutôt orientées vers la mise en oeuvre de méthodes macroscopiques alors que la méthode développée dans cette thèse est résolument microscopique.

Il est également important de noter qu'une problématique laissée ouverte par cette thèse est

l'identification d'un mélange de signaux transitoires. Nous nous sommes en effet limités à l'analyse d'un et un seul transitoire à la fois. Tant sur le plan méthodologique qu'industriel, il est intéressant de pouvoir identifier un mélange de signaux transitoires ainsi que d'identifier les cas où aucun transitoire n'est présent. Les outils méthodologiques développés dans cette thèse nous semblent pertinents pour répondre à cette problématique. En effet, les méthodes développées dans cette thèse permettent l'identification et la classification d'une séquence de fonctions de transitions. Il serait donc possible de porter ces méthodes à un degré d'abstraction supplémentaire pour identifier une séquence de transitoires, eux-même constitués de séquence de fonctions de transitions.

Notons finalement que la démarche proposée s'intéresse uniquement à l'identification des usages et non à la problématique de leur consommation individuelle. Or cette question est primordiale pour le développement des compteurs et plus largement des réseaux électriques intelligents. Cette question ne peut être résolue qu'en passant à l'analyse macroscopique des courbes de charges. De plus, il est certain que certaines caractéristiques du fonctionnement des usages, utiles à l'identification de ces appareils, ne peuvent être perçues que par une analyse macroscopique. Par exemple, les cycles de fonctionnement d'un réfrigérateur ou d'un micro-ondes ne peuvent être analysés qu'en observant les signaux à des échelles plus grandes que celles utilisées dans cette thèse. En conséquence la méthode proposée devrait s'inscrire en complément de méthodes macroscopiques afin de fournir une identification complète des usages et une évaluation de la consommation individuelle de chacun.

Publications

Actes de conférences

- M. Sanquer, F. Chatelain, M. El Guedri et N. Martin, *Bayesian curve fitting for transient signals by using smooth transition regression models*, Proceedings of the Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM and MFPT 2010), Stratford-upon-Avon, Angleterre, juin 2010
- M. Sanquer, F. Chatelain, M. El Guedri et N. Martin, *A reversible jump MCMC algorithm for Bayesian curve fitting by using smooth transition regression models*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, République Tchèque, mai 2011
- M. Sanquer, F. Chatelain, M. El Guedri et N. Martin, *Modèle polynomial par morceaux muni de transitions régulières - Application à la modélisation de signaux transitoires électriques*, Actes du colloque GRETSI 2011, Bordeaux, France, septembre 2011
- M. Sanquer, F. Chatelain, M. El Guedri et N. Martin, *Hierarchical Bayesian learning for electrical transient classification*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (À paraître)

Article de revue avec comité de lecture

- M. Sanquer, F. Chatelain, M. El Guedri et N. Martin, *A smooth transition model for multiple-regime time series*, IEEE Transactions on Signal Processing (À paraître)

A.1 Définition et paramétrisation des fonctions spéciales et des lois de probabilité standards

A.1.1 Fonctions spéciales

Fonction Gamma

La fonction Gamma, notée $\Gamma(\cdot)$ dans ce manuscrit, est la fonction définie de $(0, +\infty) \rightarrow \mathbb{R}$, telle que :

$$\Gamma : x \mapsto \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

Fonction Bêta

La fonction Bêta, notée $B(\cdot, \cdot)$ dans ce manuscrit, est la fonction définie de $(0, +\infty)^2 \rightarrow \mathbb{R}$, telle que :

$$B : (x, y) \mapsto \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

A.1.2 Lois de probabilité standards

Loi normale

La loi normale de moyenne $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+$, notée $\mathcal{N}(\mu, \sigma^2)$, est la loi de probabilité de densité :

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \mathbb{I}_{\mathbb{R}}(x).$$

Loi normale multivariée

La loi normale multivariée de moyenne $\boldsymbol{\mu} \in \mathbb{R}^N$ et de matrice de covariance Σ , notée $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, est la loi de probabilité de densité :

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \mathbb{I}_{\mathbb{R}^N}(\boldsymbol{x}).$$

Par abus de notation, les lois normales univariée et multivariée sont notées de la même manière.

Loi gamma

La loi Gamma de paramètre de forme $\nu > 0$ et de paramètre d'échelle $\theta > 0$, notée $\mathcal{G}(\nu, \theta)$, est la loi de probabilité de densité :

$$p(x) = \frac{x^{\nu-1} e^{-\frac{x}{\theta}}}{\Gamma(\nu)\theta^\nu} \mathbb{I}_{\mathbb{R}_+}(x).$$

Loi inverse-gamma

La loi inverse-gamma de paramètre de forme $\nu > 0$ et de paramètre d'échelle $\theta > 0$, notée $\mathcal{IG}(\nu, \theta)$, est la loi de probabilité de densité :

$$p(x) = \frac{\theta^\nu}{\Gamma(\nu)} \frac{1}{x^{\nu+1}} e^{-\frac{\theta}{x}} \mathbb{I}_{\mathbb{R}_+}(x).$$

Loi bêta

La loi bêta de paramètres de forme $\alpha > 0$ et $\beta > 0$, notée $\mathcal{B}(\alpha, \beta)$, est la loi de probabilité de densité :

$$p(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{I}_{[0,1]}(x).$$

A.2 Absence d'harmoniques pairs dans la forme d'onde du courant

On peut constater expérimentalement que, pour la plupart usages électriques, la décomposition en série de Fourier de la forme d'onde du courant ne comporte que des harmoniques impairs. Cette propriété remarquable découle directement d'une propriété de symétrie commune à la réponse électrique de ces usages.

Plus précisément, supposons que lorsqu'on applique une tension $v(t)$ à un usage il est parcouru par un courant $i(t)$. Alors si la tension $-v(t)$ est appliquée l'usage sera parcouru par un courant $-i(t)$.

Dans le cas de signaux périodiques, cette propriété se traduit par :

$$i\left(t + \frac{T}{2}\right) = -i(t),$$

où $T = \frac{1}{F_f}$.

La décomposition en série de Fourier du courant $i(t)$ est

$$i(t) = \sum_{n=1}^{\infty} c_n e^{j2\pi n \frac{t}{T}},$$

où c_n sont les coefficients complexes de la série de Fourier de $i(t)$:

$$c_n = \frac{1}{T} \int_0^T i(t) e^{-j2\pi n \frac{t}{T}} dt.$$

En intégrant la propriété A.2 dans le calcul des coefficients de Fourier, on obtient :

$$\begin{aligned}
 c_n &= \frac{1}{T} \int_0^{\frac{T}{2}} i(t) e^{-j2\pi n \frac{t}{T}} dt + \frac{1}{T} \int_{\frac{T}{2}}^T i(t) e^{-j2\pi n \frac{t}{T}} dt \\
 \Leftrightarrow c_n &= \frac{1}{T} \int_0^{\frac{T}{2}} i(t) e^{-j2\pi n \frac{t}{T}} dt + \frac{1}{T} \int_0^{\frac{T}{2}} i(t + \frac{T}{2}) e^{-j2\pi \frac{n}{T} (t + \frac{T}{2})} dt \\
 \Leftrightarrow c_n &= \frac{1}{T} \int_0^{\frac{T}{2}} i(t) e^{-j2\pi n \frac{t}{T}} (1 - e^{-j\pi n}) dt
 \end{aligned}$$

Et donc :

$$\forall n \in \mathbb{N} \quad c_{2n} = 0$$

A.3 Preuve de la proposition 4.1.1

Proposition A.3.1. *L'argument du maximum du gradient de la fonction de transition $T_{\pi^{(0)}}(\alpha, \lambda)$ est donné par*

$$T_{\pi^{(0)}}(\alpha, \lambda) = \underset{t}{\operatorname{argmax}} \frac{d\pi_{\lambda, \alpha, c}^{(0)}(t)}{dt} = \begin{cases} 0 & \text{si } \alpha < 1 \\ \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 0 \\ -\lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} & \text{si } \alpha > 1, c = 1 \end{cases}$$

Dans le cas où $\alpha > 1$, ce maximum correspond à un point d'inflexion de la fonction.

Démonstration. Calculons les dérivés de la fonction de transition en dérivant l'expression de la fonction de transition 4.6 que l'on rappelle ici :

$$\pi_{\lambda, \alpha, c}^{(0)}(t) = \begin{cases} 1 - \exp\left(-\left(\frac{t}{\lambda}\right)^\alpha\right) & \text{si } t > 0, c = 0 \\ \exp\left(-\left(\frac{-t}{\lambda}\right)^\alpha\right) & \text{si } t < 0, c = 1, \\ 0 & \text{sinon.} \end{cases}$$

On obtient Pour $t > 0$

$$\frac{d\pi_{\lambda, \alpha, c}^{(0)}(t)}{dt} = \begin{cases} \frac{\alpha}{\lambda} \left(\frac{t}{\lambda}\right)^{\alpha-1} e^{-\left(\frac{t}{\lambda}\right)^\alpha} & \text{si } t > 0, c = 0 \\ \frac{\alpha}{\lambda} \left(\frac{-t}{\lambda}\right)^{\alpha-1} e^{-\left(\frac{-t}{\lambda}\right)^\alpha} & \text{si } t < 0, c = 1 \\ 0 & \text{sinon} \end{cases}$$

$$\frac{d^2\pi_{\lambda, \alpha, c}^{(0)}(t)}{dt^2} = \begin{cases} \frac{\alpha}{\lambda} \left(\frac{t}{\lambda}\right)^{\alpha-2} \left[\frac{\alpha-1}{\lambda} - \frac{\alpha}{\lambda} \left(\frac{t}{\lambda}\right)^\alpha\right] e^{-\left(\frac{t}{\lambda}\right)^\alpha} & \text{si } t > 0, c = 0 \\ \frac{\alpha}{\lambda} \left(\frac{-t}{\lambda}\right)^{\alpha-2} \left[-\frac{\alpha-1}{\lambda} + \frac{\alpha}{\lambda} \left(\frac{-t}{\lambda}\right)^\alpha\right] e^{-\left(\frac{-t}{\lambda}\right)^\alpha} & \text{si } t < 0, c = 1 \\ 0 & \text{sinon} \end{cases}$$

Dans le premier cas $t > 0, c = 0$, la dérivée seconde s'annule si $\alpha > 1$ au point

$$\frac{\alpha-1}{\lambda} - \frac{\alpha}{\lambda} \left(\frac{t}{\lambda}\right)^\alpha = 0 \Rightarrow t = \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}$$

Dans le deuxième cas $t < 0$, $c = 1$, la dérivée seconde s'annule si $\alpha > 1$ au point

$$-\frac{\alpha-1}{\lambda} + \frac{\alpha}{\lambda} \left(\frac{-t}{\lambda}\right)^\alpha = 0 \Rightarrow t = -\lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}$$

Lorsque $\alpha < 1$ quelque soit c , la dérivée seconde ne s'annule jamais. Il n'y a donc pas de point d'inflexion. Par contre le gradient présente une discontinuité $t = 0$ et son maximum est atteint en approchant cette discontinuité par la droite. ■

A.4 Preuve de la proposition 4.1.2

Proposition A.4.1. Soit $A_{\pi^{(0)}}(\lambda, \alpha)$, la somme de l'aire sous la courbe avant l'instant $t = T_{\pi^{(0)}}(\alpha, \lambda)$ et l'aire au dessus de la courbe après l'instant $t = T_{\pi^{(0)}}(\alpha, \lambda)$, définie par :

$$A_{\pi^{(0)}}(\lambda, \alpha) = \int_{-\infty}^{T_{\pi^{(0)}}(\alpha, \lambda)} \left(\pi_{\lambda, \alpha, c}^{(0)}(t)\right) dt + \int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \left(1 - \pi_{\lambda, \alpha, c}^{(0)}(t)\right) dt,$$

L'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ est proportionnelle à λ et vaut

$$A_{\pi^{(0)}}(\lambda, \alpha) = \begin{cases} \lambda \left[\frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)\right] & \text{si } \alpha < 0 \\ \frac{\lambda}{\alpha} \left[\frac{1}{\alpha} \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} + \Gamma\left(\frac{1}{\alpha}\right) - 2\gamma\left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha}\right)\right] & \text{si } \alpha > 0, \end{cases}$$

où Γ est la fonction gamma telle que $\Gamma : \begin{cases} \mathbb{R}_+ & \longrightarrow \mathbb{R}_+ \\ (z) & \longmapsto \int_0^{+\infty} t^{z-1} e^{-t} dt \end{cases}$ et γ est la fonction gamma

incomplète telle que $\gamma : \begin{cases} \mathbb{R}_+ \times \mathbb{R}_+ & \longrightarrow \mathbb{R}_+ \\ (z, x) & \longmapsto \int_0^x t^{z-1} e^{-t} dt \end{cases}$

On note $L_{\pi^{(0)}}(\alpha) = \frac{1}{\lambda} A_{\pi^{(0)}}(\lambda, \alpha)$ le facteur de proportionnalité entre $A_{\pi^{(0)}}(\lambda, \alpha)$ et λ , qui ne dépend donc que de α .

Démonstration. On sépare le calcul en plusieurs cas :

1. Premier cas : $\alpha < 1$, $u = 0$ Dans ce cas $T_{\pi^{(0)}}(\alpha, \lambda)$ vaut 0 et la fonction $\pi_{\lambda, \alpha, u}^{(0)}(t)$ vaut 0 également pour $t < 0$, donc :

$$\int_{-\infty}^{T_{\pi^{(0)}}(\alpha, \lambda)} \left(\pi_{\lambda, \alpha, u}^{(0)}(t)\right) dt = 0$$

Il reste à calculer le second terme de $A_{\pi^{(0)}}(\lambda, \alpha)$:

$$\begin{aligned} A_{\pi^{(0)}}(\lambda, \alpha) &= \int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \left(1 - \pi_{\lambda, \alpha, u}^{(0)}(t)\right) dt \\ &= \int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \exp\left(-\left(\frac{t}{\lambda}\right)^\alpha\right) dt \end{aligned}$$

On réalise le changement de variable $t \rightarrow t' = \left(\frac{t}{\lambda}\right)^\alpha$ ce qui donne

$$\begin{aligned} A_{\pi^{(0)}}(\lambda, \alpha) &= \int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \frac{\lambda}{\alpha} (t')^{\frac{1}{\alpha}-1} \exp(-t') dt \\ &= \frac{\lambda}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) \end{aligned}$$

Où Γ est la fonction Gamma définie par

$$\Gamma : z \rightarrow \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (\text{A.1})$$

2. Deuxième cas : $\alpha < 1, u = 1$ Par symétrie, l'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ dans ce cas est identique au cas précédent.
3. Troisième cas : $\alpha > 1, u = 0$ Dans ce cas $T_{\pi^{(0)}}(\alpha, \lambda) = \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}$. On sépare le calcul des deux termes de $A_{\pi^{(0)}}(\lambda, \alpha)$:
 - Premier terme :

$$\int_{-\infty}^{T_{\pi^{(0)}}(\alpha, \lambda)} \pi_{\lambda, \alpha, u}^{(0)}(t) dt = \int_0^{\lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}} 1 - \exp\left(-\left(\frac{t}{\lambda}\right)^\alpha\right) dt$$

On effectue le changement de variable $t \rightarrow t' = \left(\frac{t}{\lambda}\right)^\alpha$ ce qui donne

$$\begin{aligned} \int_{-\infty}^{T_{\pi^{(0)}}(\alpha, \lambda)} \pi_{\lambda, \alpha, u}^{(0)}(t) dt &= \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} - \int_0^{\frac{\alpha-1}{\alpha}} \exp(-t') \frac{\lambda}{\alpha} t'^{\frac{1}{\alpha}-1} dt' \\ &= \lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} - \frac{\lambda}{\alpha} \gamma\left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha}\right) \end{aligned}$$

où γ est la fonction gamma incomplète définie par

$$\gamma : (z, x) \rightarrow \int_0^x t^{z-1} e^{-t} dt$$

– Second terme :

$$\int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \left(1 - \pi_{\lambda, \alpha, u}^{(0)}(t)\right) dt = \int_{\lambda \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}}^{+\infty} \exp\left(-\left(\frac{t}{\lambda}\right)^\alpha\right) dt$$

On effectue encore une fois le changement de variable $t \rightarrow t' = \left(\frac{t}{\lambda}\right)^\alpha$ ce qui donne

$$\begin{aligned} \int_{T_{\pi^{(0)}}(\alpha, \lambda)}^{+\infty} \left(1 - \pi_{\lambda, \alpha, u}^{(0)}(t)\right) dt &= \int_{\frac{\alpha-1}{\alpha}}^{+\infty} \exp(-t') \frac{\lambda}{\alpha} t'^{\frac{1}{\alpha}-1} dt' \\ &= \frac{\lambda}{\alpha} \left(\Gamma\left(\frac{1}{\alpha}\right) - \gamma\left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha}\right)\right) \end{aligned}$$

En combinant les deux termes calculés, on obtient finalement :

$$A_{\pi^{(0)}}(\lambda, \alpha) = \frac{\lambda}{\alpha} \left[\frac{1}{\alpha} \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}} + \Gamma\left(\frac{1}{\alpha}\right) - 2\gamma\left(\frac{1}{\alpha}, \frac{\alpha-1}{\alpha}\right) \right]$$

4. Quatrième cas : $\alpha > 1, u = 1$ Par symétrie, l'aire $A_{\pi^{(0)}}(\lambda, \alpha)$ dans ce cas est identique au cas précédent. ■

A.5 Marginalisation des hyperparamètres de l'apprentissage bayésien hiérarchique

Marginalisation de l'hyperparamètre μ_K

La loi marginale des nombres de ruptures K_1, \dots, K_{N_c} de chaque signal de la base d'apprentissage se déduit simplement des lois *a priori* de ces paramètres (5.12) et de la loi *a priori* (5.22) de l'hyperparamètre μ_K . On rappelle que :

$$p(K_i|\mu_K) = e^{-\mu_K} \frac{\mu_K^{K_i}}{K_i!} \quad \forall i = 1, \dots, N_c,$$

et que

$$p(\mu_K) \propto \frac{1}{\sqrt{\mu_K}} \mathbb{I}_{\mathbb{R}_+}(\mu_K).$$

On en déduit :

$$\begin{aligned} p(K_1, \dots, K_{N_c}) &= \int p(\mu_K) \prod_{i=1}^{N_c} p(K_i|\mu_K) d\mu_K \\ &\propto \int_{\mathbb{R}_+} \frac{1}{\sqrt{\mu_K}} e^{-N_c \mu_K} \mu_K^{\sum_{i=1}^{N_c} K_i} \prod_{i=1}^{N_c} \frac{1}{K_i!} d\mu_K \\ &\propto \Gamma\left(\frac{1}{2} + \sum_{i=1}^{N_c} K_i\right) N_c^{-\left(\frac{1}{2} + \sum_{i=1}^{N_c} K_i\right)} \prod_{i=1}^{N_c} \frac{1}{K_i!} \end{aligned}$$

Marginalisation des paramètres σ_i^2 et β_i

Les paramètres β_i et σ_i^2 n'influent sur la loi *a posteriori* du modèle que via la vraisemblance du modèle (3.14) et leur loi *a priori* respectives (5.14), (5.13), que l'on rappelle ici :

$$\begin{aligned} \mathcal{L}(\mathbf{x}_i|\boldsymbol{\theta}_i) &= \frac{1}{(2\pi\sigma_i^2)^{\frac{n_i}{2}}} e^{-\frac{1}{2\sigma_i^2} \|\mathbf{x}_i - Z_i \boldsymbol{\beta}_i\|^2}, \\ \boldsymbol{\beta}_i|\sigma_i^2, Z_i^* &\sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \delta^2 [(Z_i^*)^T Z_i^*]^{-1}), \quad \sigma_i^2|\rho_\sigma \sim \mathcal{IG}(1, \rho_\sigma). \end{aligned}$$

Pour chaque signal \mathbf{x}_i de la base d'apprentissage, on a donc :

$$\begin{aligned} &\int \int \mathcal{L}(\mathbf{x}_i|\boldsymbol{\theta}_i) p(\boldsymbol{\beta}_i|\sigma_i^2, \boldsymbol{\tau}_i, \boldsymbol{\eta}_i) p(\sigma_i^2|\rho_\sigma) d\boldsymbol{\beta}_i d\sigma_i^2 \\ &\propto \int_{\mathbb{R}_+} \int_{\mathbb{R}^{K_i}} \frac{|Z_i^T Z_i|^{\frac{1}{2}} e^{-\frac{1}{2\sigma_i^2} \|\mathbf{x}_i - Z_i \boldsymbol{\beta}_i\|^2} e^{-\frac{1}{2\sigma_i^2 \delta^2} (\boldsymbol{\beta}_i - \mu_{\boldsymbol{\beta}_K})^T Z_i^T Z_i (\boldsymbol{\beta}_i - \mu_{\boldsymbol{\beta}_K})}}{(\sigma_i^2)^{\frac{n_i}{2}+2} (\sigma_i^2 \delta^2)^{\frac{K_i}{2}}} e^{-\frac{\rho_\sigma}{\sigma_i^2}} d\boldsymbol{\beta}_i d\sigma_i^2 \\ &\propto \int_{\mathbb{R}_+} \frac{e^{-\frac{1}{2\sigma_i^2} (\mathbf{x}_i^T \mathbf{x}_i + \frac{1}{\delta^2} \mu_{\boldsymbol{\beta}_K}^T Z_i^T Z_i \mu_{\boldsymbol{\beta}_K} - \mu_{\boldsymbol{\beta}_i}^T \Sigma_{\boldsymbol{\beta}_i}^{-1} \mu_{\boldsymbol{\beta}_i})}}{(\sigma_i^2)^{\frac{n_i}{2}+2} (1 + \delta^2)^{\frac{K_i}{2}}} e^{-\frac{\rho_\sigma}{\sigma_i^2}} d\sigma_i^2 \\ &\propto \frac{(\delta^2 + 1)^{-\frac{K_i}{2}}}{\left(\rho_\sigma + \frac{1}{2} \left(\mathbf{x}_i^T \mathbf{x}_i + \frac{1}{\delta^2} \mu_{\boldsymbol{\beta}_K}^T Z_i^T Z_i \mu_{\boldsymbol{\beta}_K} - \mu_{\boldsymbol{\beta}_i}^T \Sigma_{\boldsymbol{\beta}_i}^{-1} \mu_{\boldsymbol{\beta}_i}\right)\right)^{\frac{n_i}{2}+1}}, \\ &\propto \frac{(\delta^2 + 1)^{-\frac{K_i}{2}}}{\left(\rho_\sigma + \frac{1}{2} (\mathbf{x}_i^T \mathbf{x}_i - \mu_{\boldsymbol{\beta}_i}^T \Sigma_{\boldsymbol{\beta}_i}^{-1} \mu_{\boldsymbol{\beta}_i})\right)^{\frac{n_i}{2}+1}}, \end{aligned}$$

Marginalisation des hyperparamètres ρ_{τ_K}

Ces hyperparamètres influent sur la loi *a posteriori* du modèle bayésien hiérarchique via leur loi *a priori* :

$$p(\rho_{\tau_K}) = \prod_{l=1}^{K-1} \Delta \frac{e^{-\frac{\Delta}{\rho_{\tau_l, K}}}}{\rho_{\tau_l, K}^2} \mathbb{I}_{\mathbb{R}_+}(\rho_{\tau_l, K}), \quad \forall K > 1,$$

et via la loi *a priori* des paramètres de position des ruptures τ_i :

$$p(\tau_i | \rho_{\tau_K}) = \prod_{k=1}^{K_i-1} \frac{1}{\rho_{\tau_k, K_i}} \exp\left(-\frac{\tau_{k+1, i} - \tau_{k, i}}{\rho_{\tau_k, K_i}}\right) \mathbb{I}_{\mathbb{R}_+}(\tau_{k+1, i} - \tau_{k, i}).$$

On en déduit la loi marginale des paramètres τ_i pour chaque sous-groupe de signal possédant le même nombre de ruptures. On note $I_c^K = \{i = 1, \dots, N_c \setminus K_i = K\}$ les indices des signaux de ce sous-groupe et on note $N_c^K = \text{card}(I_c^K)$. On en déduit que :

$$\begin{aligned} p((\tau_i)_{i \in I_c^K}) &= \int p(\rho_{\tau_K}) \prod_{i \in I_c^K} p(\tau_i | \rho_{\tau_K}) d\rho_{\tau_K}, \\ &= \prod_{k=1}^{K-1} \int_{\mathbb{R}_+} \Delta \frac{e^{-\frac{\Delta}{\rho_{\tau_k, K}}}}{\rho_{\tau_k, K}^2} \prod_{i \in I_c^K} \frac{1}{\rho_{\tau_k, K}} e^{-\frac{\tau_{k+1, i} - \tau_{k, i}}{\rho_{\tau_k, K}}} d\rho_{\tau_k, K}, \\ &= \prod_{k=1}^{K-1} \int_{\mathbb{R}_+} \frac{\Delta}{(\rho_{\tau_k, K})^{2+N_c^K}} e^{-\frac{1}{\rho_{\tau_k, K}} (\Delta + \sum_{i \in I_c^K} (\tau_{k+1, i} - \tau_{k, i}))} d\rho_{\tau_k, K}, \\ p((\tau_i)_{i \in I_c^K}) &= \prod_{k=1}^{K-1} \frac{\Delta \Gamma(1 + N_c^K)}{(\Delta + \sum_{i \in I_c^K} (\tau_{k+1, i} - \tau_{k, i}))^{1+N_c^K}}, \quad \forall K > 1. \end{aligned}$$

Marginalisation des hyperparamètres ρ_{λ_K}

Les hyperparamètres ρ_{λ_K} n'interviennent dans la loi *a posteriori* du modèle bayésien hiérarchique que via leur loi *a priori* :

$$p(\rho_{\lambda_K}) = \prod_{l=1}^K \Delta \frac{e^{-\frac{\Delta}{\rho_{\lambda_l, K}}}}{\rho_{\lambda_l, K}^2} \mathbb{I}_{\mathbb{R}_+}(\rho_{\lambda_l, K}), \quad \forall K > 0,$$

et via la loi *a priori* des paramètres d'échelle de chaque rupture :

$$\lambda_{k, i} | \rho_{\lambda_{k, K_i}} \sim \mathcal{G}\left(\nu_{\lambda}, \rho_{\lambda_{k, K_i}}\right) \quad \forall k = 1, \dots, K_i,$$

On en déduit que :

$$\begin{aligned}
p((\lambda_i)_{i \in \mathbb{I}_c^K}) &= \int p(\rho_{\lambda_K}) \prod_{i \in \mathbb{I}_c^K} p(\lambda_i | \rho_{\lambda_K}) d\rho_{\lambda_K}, \\
&= \prod_{k=1}^K \int_{\mathbb{R}_+} \Delta \frac{e^{-\frac{\Delta}{\rho_{\lambda_{l,K}}}}}{\rho_{\lambda_{l,K}}^2} \prod_{i \in \mathbb{I}_c^K} \frac{1}{\rho_{\lambda_{k,K}}} e^{-\frac{\lambda_{k,i}}{\rho_{\lambda_{k,K}}}} d\rho_{\lambda_{k,K}}, \\
&= \prod_{k=1}^K \int_{\mathbb{R}_+} \frac{\Delta}{(\rho_{\lambda_{k,K}})^{2+N_c^K}} e^{-\frac{1}{\rho_{\lambda_{k,K}}} (\Delta + \sum_{i \in \mathbb{I}_c^K} \lambda_{k,i})} d\rho_{\lambda_{k,K}}, \\
p((\lambda_i)_{i \in \mathbb{I}_c^K}) &= \prod_{k=1}^K \frac{\Delta \Gamma(1 + N_c^K)}{(\Delta + \sum_{i \in \mathbb{I}_c^K} \lambda_{k,i})^{1+N_c^K}}, \quad \forall K > 0.
\end{aligned}$$

Marginalisation des hyperparamètres μ_{α_K} et $\sigma_{\alpha_K}^2$

Les hyperparamètres μ_{α_K} et $\sigma_{\alpha_K}^2$ qui définissent la loi des paramètres de forme des ruptures α_i au sein d'une classe, n'interviennent dans la loi *a posteriori* du modèle bayésien hiérarchique que via leur loi *a priori* :

$$p(\mu_{\alpha_K}, \sigma_{\alpha_K}^2) = \prod_{l=1}^K \Delta \frac{e^{-\frac{\Delta}{\sigma_{\alpha_{l,K}}^2}}}{(\sigma_{\alpha_{l,K}}^2)^2} \frac{e^{-\frac{\mu_{\alpha_{l,K}}^2}{2\Psi\sigma_{\alpha_{l,K}}^2}}}{(2\pi\Psi\sigma_{\alpha_{l,K}}^2)^{\frac{1}{2}}} \mathbb{I}_{\mathbb{R} \times \mathbb{R}_+}(\mu_{\alpha_{l,K}}, \sigma_{\alpha_{l,K}}^2), \quad \forall K > 0,$$

et via la loi *a priori* des paramètres α_i :

$$\alpha_k | \mu_{\alpha_{k,K_i}}, \sigma_{\alpha_{k,K_i}}^2 \sim \mathcal{N}(\mu_{\alpha_{k,K_i}}, \sigma_{\alpha_{k,K_i}}^2) \quad \forall k = 1, \dots, K_i. \quad (\text{A.2})$$

On en déduit que :

$$\begin{aligned}
p((\alpha_i)_{i \in \mathbb{I}_c^K}) &= \int \int p(\mu_{\alpha_K}, \sigma_{\alpha_K}^2) \prod_{i \in \mathbb{I}_c^K} p(\alpha_i | \mu_{\alpha_K}, \sigma_{\alpha_K}^2) d\mu_{\alpha_K} d\sigma_{\alpha_K}^2, \\
&= \prod_{k=1}^K \int_{\mathbb{R}_+} \int_{\mathbb{R}} \Delta \frac{e^{-\frac{\Delta}{\sigma_{\alpha_{l,K}}^2}}}{(\sigma_{\alpha_{l,K}}^2)^2} \frac{e^{-\frac{\mu_{\alpha_{l,K}}^2}{2\Psi\sigma_{\alpha_{l,K}}^2}}}{(2\pi\Psi\sigma_{\alpha_{l,K}}^2)^{\frac{1}{2}}} \prod_{i \in \mathbb{I}_c^K} \frac{1}{(2\pi\sigma_{\alpha_{k,K}}^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_{\alpha_{k,K}}^2} (\alpha_{k,i} - \mu_{\alpha_{k,K}})^2} d\mu_{\alpha_{k,K}} d\sigma_{\alpha_{k,K}}^2, \\
&= \prod_{k=1}^K \int_{\mathbb{R}_+} \frac{\Delta (2\pi)^{-\frac{N_c^K}{2}}}{(N_c^K + \frac{1}{\Psi})^{\frac{1}{2}} \Psi^{\frac{1}{2}}} \frac{1}{(\sigma_{\alpha_{k,K}}^2)^{2+\frac{N_c^K}{2}}} e^{-\frac{1}{\sigma_{\alpha_{k,K}}^2} \left(\Delta + \frac{1}{2} \left(\sum_{i \in \mathbb{I}_c^K} \alpha_{k,i}^2 - \left(\frac{1}{N_c^K + \frac{1}{\Psi}} \sum_{i \in \mathbb{I}_c^K} \alpha_{k,i} \right)^2 \right) \right)} d\sigma_{\alpha_{k,K}}^2, \\
p((\alpha_i)_{i \in \mathbb{I}_c^K}) &= \prod_{k=1}^K \frac{\Delta \Gamma(\frac{N_c^K}{2} + 1)}{\Psi^{\frac{1}{2}} (N_c^K + \frac{1}{\Psi})^{\frac{1}{2}} (2\pi)^{\frac{N_c^K}{2}}} \left(\Delta + \frac{1}{2} \left(\sum_{i \in \mathbb{I}_c^K} \alpha_{k,i}^2 - \left(\frac{1}{N_c^K + \frac{1}{\Psi}} \sum_{i \in \mathbb{I}_c^K} \alpha_{k,i} \right)^2 \right) \right)^{-\frac{N_c^K}{2} - 1}, \quad \forall K > 0.
\end{aligned}$$

Marginalisation des hyperparamètres $p_{\mathbf{c}_K}$

Les hyperparamètres $p_{\mathbf{c}_K}$ n'interviennent dans la loi *a posteriori* du modèle bayésien hiérarchique que via leur loi *a priori* :

$$p(p_{\mathbf{c}_K}) = \prod_{l=1}^K \mathbb{I}_{[0,1]}(c_{l,K}) \quad \forall K > 0.$$

et via la loi *a priori* des paramètres c_k :

$$\begin{cases} p(c_k = 0 | p_{k,K}, K) &= p_{k,K}, \\ p(c_k = 1 | p_{k,K}, K) &= 1 - p_{k,K}. \end{cases}$$

On en déduit la loi marginale des paramètres \mathbf{c}_i :

$$\begin{aligned} p((\mathbf{c}_i)_{i \in \mathbb{I}_c^K}) &= \int p(p_{\mathbf{c}_K}) \prod_{i \in \mathbb{I}_c^K} p(\mathbf{c}_i | p_{\mathbf{c}_K}) dp_{\mathbf{c}_K}, \\ &= \prod_{k=1}^K \int p_{c_{k,K}}^{\sum_{i \in \mathbb{I}_c^K} c_{k,i}} (1 - p_{c_{k,K}})^{N_c^K - \sum_{i \in \mathbb{I}_c^K} c_{k,i}} dp_{c_{k,K}}, \\ p((\mathbf{c}_i)_{i \in \mathbb{I}_c^K}) &= \prod_{k=1}^K B \left(1 + \sum_{i \in \mathbb{I}_c^K} c_{k,i}, 1 + N_c^K - \sum_{i \in \mathbb{I}_c^K} c_{k,i} \right), \quad \forall K > 0. \end{aligned}$$

Bibliographie

- [Akaike 1974] Hirotugu Akaike. *A New Look at the Statistical Model Identification*. IEEE Transactions on Automatic Control, vol. 19, no. 6, pages 716–723, 1974. (Cité en page 50.)
- [Anderson 2012] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe et M. Berges. *BLUED : A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research*. In Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD), Beijing, China, 2012. (Cité en page 29.)
- [Anderssen 2004] R.S. Anderssen, S.A. Husain et R.J. Loy. *The Kohlrausch function : properties and applications*. ANZIAM Journal, vol. 45, no. 0, 2004. (Cité en page 83.)
- [Andrieu 1999] C. Andrieu et A. Doucet. *Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC*. Signal Processing, IEEE Transactions on, vol. 47, no. 10, pages 2667–2676, 1999. (Cité en pages 50 et 57.)
- [Baladan. 2005] V. Baladan., B.K. Mallick et R.J. Carrol. *Spatially Adaptive Bayesian Penalized Regression Splines (P-splines)*. Journal of Computational and Graphical Statistics, vol. 14, no. 2, pages 378–394, 2005. le nom complet du premier auteur est Baladandayuthapani. (Cité en pages 77 et 80.)
- [Baranski 2003] M. Baranski et Voss J. *Nonintrusive appliance load monitoring based on an optical sensor*. In Power Tech Conference Proceedings, 2003 IEEE Bologna, volume 4, page 8 pp. Vol.4, june 2003. (Cité en page 23.)
- [Baranski 2004a] M. Baranski et J. Voss. *Detecting patterns of appliances from total load data using a dynamic programming approach*. In Data Mining, 2004. ICDM '04. Fourth IEEE International Conference on, pages 327 – 330, nov. 2004. (Cité en page 23.)
- [Baranski 2004b] M. Baranski et J. Voss. *Genetic algorithm for pattern detection in NIALM systems*. In Systems, Man and Cybernetics, 2004 IEEE International Conference on, volume 4, pages 3462 – 3468 vol.4, oct. 2004. (Cité en page 23.)
- [Basseville 1993] Michle Basseville et Igor V. Nikiforov. *Detection of abrupt changes : Theory and application*. Prentice-Hall, 1993. (Cité en page 49.)
- [Bishop 1999] C.M. Bishop. *Latent variable models*, pages 371–403. Learning in graphical models. MIT press, 1999. (Cité en page 103.)
- [Bishop 2006] C.M. Bishop. *Pattern recognition and machine learning*. Information Science and Statistics. Springer-Verlag, 2nd édition, 2006. (Cité en page 98.)
- [Bouloutas 1991] A. Bouloutas, G.W. Hart et M. Schwartz. *Two extensions of the Viterbi algorithm*. Information Theory, IEEE Transactions on, vol. 37, no. 2, pages 430 –436, mar 1991. (Cité en page 23.)
- [Brand 1997] M. Brand, N. Oliver et A. Pentland. *Coupled hidden Markov models for complex action recognition*. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, pages 994 –999, jun 1997. (Cité en page 25.)
- [Carlin 1997] B.P. Carlin et T.A. Louis. *Bayes and emprirical bayes methods for data analysis*. Statistics and Computing, vol. 7, pages 153–154, 1997. (Cité en page 61.)
- [Carluccio 2011] D. Carluccio et S. Brinkhaus. *Smart hacking for privacy*. In 28th Chaos Communication Congress, 2011. (Cité en page 18.)

- [Chang 2008] H-H Chang, C-L Lin et H-T Yang. *Load recognition for different loads with the same real power and reactive power in a non-intrusive load-monitoring system*. In Computer Supported Cooperative Work in Design, 2008. CSCWD 2008. 12th International Conference on, pages 1122–1127, 2008. (Cit  en page 27.)
- [Chang 2010] H-H Chang, C-L Lin et J-K Lee. *Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms*. In Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on, pages 27–32, 2010. (Cit  en page 27.)
- [Cole 1998] A.I. Cole et A. Albicki. *Data extraction for effective non-intrusive identification of residential power loads*. In Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings. IEEE, volume 2, pages 812–815 vol.2, 1998. (Cit  en page 27.)
- [Cox 2006] R. Cox, S.B. Leeb, S.R. Shaw et L.K. Norford. *Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion*. In Applied Power Electronics Conference and Exposition, 2006. APEC '06. Twenty-First Annual IEEE, page 7 pp., march 2006. (Cit  en pages 19 et 28.)
- [Daley 2002] D.J. Daley et D. Vere-Jones. *An Introduction to the Theory of Point Processes. Probability and its Applications*. Springer-Verlag, 2002. (Cit  en page 57.)
- [Darby 2006] S. Darby. *The Effectiveness Of Feedback On Energy Consumption*. Change, vol. 22nd, no. April, pages 33–35, 2006. (Cit  en page 17.)
- [Davy 2002] M. Davy, C. Doncarli et J.-Y. Tournet. *Classification of chirp signals using hierarchical Bayesian learning and MCMC methods*. Signal Processing, IEEE Transactions on, vol. 50, no. 2, pages 377–388, 2002. (Cit  en pages 101 et 134.)
- [Denison 1998] D. G. T. Denison, B. K. Mallick et A. F. M. Smith. *Automatic Bayesian curve fitting*. Journal of the Royal Statistical Society : Series B (Statistical Methodology), vol. 60, no. 2, pages 333–350, 1998. (Cit  en pages 73, 74, 77 et 80.)
- [Dobigeon 2007a] N. Dobigeon, J.-Y. Tournet et M. Davy. *Joint Segmentation of Piecewise Constant Autoregressive Processes by Using a Hierarchical Model and a Bayesian Sampling Approach*. Signal Processing, IEEE Transactions on, vol. 55, no. 4, pages 1251–1263, 2007. (Cit  en pages 58, 61 et 65.)
- [Dobigeon 2007b] N. Dobigeon, J.-Y. Tournet et J.D. Scargle. *Joint Segmentation of Multivariate Astronomical Time Series : Bayesian Sampling With a Hierarchical Model*. Signal Processing, IEEE Transactions on, vol. 55, no. 2, pages 414–423, 2007. (Cit  en page 65.)
- [Domingos 1997] Pedro Domingos et Michael Pazzani. *On the optimality of the simple Bayesian classifier under zero-one loss*. Machine Learning, vol. 29, page 103–137, 1997. (Cit  en page 107.)
- [Donoho 1994] David L. Donoho et Iain M. Johnstone. *Ideal Spatial Adaptation by Wavelet Shrinkage*. Biometrika, vol. 81, no. 3, pages 425–455, 1994. (Cit  en pages 74 et 77.)
- [Du 2010] Y. Du, L. Du, B. Lu, R. Harley et T. Habetler. *A review of identification and monitoring methods for electric loads in commercial and residential buildings*. In Energy Conversion Congress and Exposition (ECCE), 2010 IEEE, pages 4527–4533, sept. 2010. (Cit  en page 20.)
- [Durand 2003] J. B. Durand, L. Bozzi, G. Celeux et C Derquenne. *Analyse de courbes de consommation  lectrique par cha nes de Markov cach es*. Rapport technique 4858, Institut National de Recherche en Informatique et en Automatique (INRIA), 2003. (Cit  en page 24.)

- [DYK] (Cité en page 71.)
- [El Guedri 2009] M. El Guedri, J. Bect, C. Lajaunie, G. Fleury, R. Seraoui, T. Garcia et A. Girard. *RJMCMC point process sampler for single sensor source separation : an application to electric load monitoring*. In Proceedings of the 17th European Signal Processing Conference, Glasgow, Royaume-Uni, 2009. (Cité en pages 24 et 30.)
- [Farinaccio 1999] L. Farinaccio et R. Zmeureanu. *Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses*. Energy and Buildings, vol. 30, no. 3, pages 245 – 259, 1999. (Cité en page 23.)
- [Fearnhead 2005] P. Fearnhead. *Exact Bayesian curve fitting and signal segmentation*. Signal Processing, IEEE Transactions on, vol. 53, no. 6, pages 2160–2166, 2005. (Cité en page 65.)
- [Gelfand 1990] A.E. Gelfand et A.F.M. Smith. *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, vol. 85, no. 410, pages 398–409, 1990. (Cité en page 65.)
- [Geyer 1994] C.J. Geyer et J. Møller. *Simulation Procedures and Likelihood Inference for Spatial Point Processes*. Scandinavian Journal of Statistics, vol. 21, no. 4, pages 359–373, 1994. (Cité en page 67.)
- [Ghahramani 1997] Z. Ghahramani et M.I. Jordan. *Factorial Hidden Markov Models*. Machine Learning, vol. 29, pages 245–273, 1997. 10.1023/A :1007425814087. (Cité en pages 24 et 25.)
- [Green 1995] P.J. Green. *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination*. Biometrika, vol. 82, no. 4, pages 711–732, 1995. (Cité en pages 65 et 67.)
- [Gupta 2010] S. Gupta, M.S. Reynolds et S.N. Patel. *ElectriSense : single-point sensing using EMI for electrical event detection and classification in the home*. In Proceedings of the 12th ACM international conference on Ubiquitous computing, Ubicomp '10, pages 139–148, New York, NY, USA, 2010. ACM. (Cité en page 29.)
- [Gustafsson 2000] Fredrik Gustafsson. Adaptive filtering and change detection. Wiley, 2000. (Cité en page 49.)
- [Han 2001] Cong Han et Bradley P Carlin. *Markov Chain Monte Carlo Methods for Computing Bayes Factors*. Journal of the American Statistical Association, vol. 96, no. 455, pages 1122–1132, 2001. (Cité en page 124.)
- [Hart 1992] G.W. Hart. *Nonintrusive appliance load monitoring*. Proceedings of the IEEE, vol. 80, no. 12, pages 1870 –1891, dec 1992. (Cité en pages 16, 20, 22, 23 et 30.)
- [Hart 1993] G.W. Hart et A.T. Bouloutas. *Correcting dependent errors in sequences generated by finite-state processes*. Information Theory, IEEE Transactions on, vol. 39, no. 4, pages 1249 –1260, jul 1993. (Cité en page 23.)
- [Hastie 2009] T. Hastie, R. Tibshirani et J. Friedman. The elements of statistical learning : Data mining, inference, and prediction. Springer-Verlag, 2nd édition, 2009. (Cité en page 98.)
- [Jeffreys 1946] Harold Jeffreys. *An invariant form for the prior probability in estimation problems*. Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences, no. 1007, pages 453–461, 1946. (Cité en page 60.)
- [Jiang 2011] L. Jiang, J. Li, S. Luo, J. Jin et S. West. *Literature review of power disaggregation*. In Modelling, Identification and Control (ICMIC), Proceedings of 2011 International Conference on, pages 38 –42, june 2011. (Cité en page 20.)

- [Kail 2012] G. Kail, J.-Y. Tournéret, F. Hlawatsch et N. Dobigeon. *Blind Deconvolution of Sparse Pulse Sequences Under a Minimum Distance Constraint : A Partially Collapsed Gibbs Sampler Method*. Signal Processing, IEEE Transactions on, vol. 60, no. 6, pages 2727–2743, 2012. (Cité en page 66.)
- [Kemp 2007] Charles Kemp, Amy Pefors et Joshua B. Tenenbaum. *Learning Overhypotheses with Hierarchical Bayesian models*. Developmental Science, vol. 10, no. 3, pages 307–321, 2007. (Cité en page 102.)
- [Kim 2011] H. Kim, M. Marwah, M.F. Arlitt, G. Lyon et J. Han. *Unsupervised Disaggregation of Low Frequency Power Measurements*. In SDM, pages 747–758. SIAM / Omnipress, 2011. (Cité en pages 24 et 25.)
- [Kirkpatrick 1983] S. Kirkpatrick, C. Gelatt et M. Vecchi. *Optimization by Simulated Annealing*. Science, no. 220, pages 671–680, 1983. (Cité en page 51.)
- [Kohlrausch 1854] R. Kohlrausch. *Theorie des Elektrischen Rückstandes in der Leidener Flasche*. Pogg. Ann. Phys. Chem, vol. 91, pages 179–214, 1854. (Cité en page 83.)
- [Kolter 2012] Z. Kolter et T. Jaakkola. *Approximate inference in additive factorial hmms with application to energy disaggregation*. Journal of Machine Learning Research, no. 22, pages 1472–1482, 2012. (Cité en pages 22, 24, 26 et 29.)
- [Laughman 2003] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford et P. Armstrong. *Power signature analysis*. Power and Energy Magazine, IEEE, vol. 1, no. 2, pages 56 – 63, mar-apr 2003. (Cité en pages 18 et 26.)
- [Lavielle 1998] M. Lavielle. *Optimal segmentation of random processes*. Signal Processing, IEEE Transactions on, vol. 46, no. 5, pages 1365–1373, 1998. (Cité en page 65.)
- [Lavielle 2001] M. Lavielle et E. Lebarbier. *An application of MCMC methods for the multiple change-points problem*. Signal Process., vol. 81, no. 1, pages 39–53, 2001. (Cité en page 50.)
- [Leeb 1995] S.B. Leeb, S.R. Shaw et J.L. Kirtley Jr. *Transient event detection in spectral envelope estimates for nonintrusive load monitoring*. Power Delivery, IEEE Transactions on, vol. 10, no. 3, pages 1200 –1210, 1995. (Cité en pages 27 et 28.)
- [Liang 2010] J. Liang, S. Ng, G. Kendall et J. Cheng. *Load Signature Study - Part I : Basic Concept, Structure, and Methodology*. Power Delivery, IEEE Transactions on, vol. 25, no. 2, pages 551 –560, avril 2010. (Cité en page 20.)
- [Lubrano 2000] Michel Lubrano. *Bayesian Analysis of Nonlinear Time Series Models with a Threshold*, pages 79–118. Cambridge University Press, 2000. (Cité en page 50.)
- [Lundbergh 2003] Stefan Lundbergh, Timo Teräsvirta et Dick van Dijk. *Time-Varying Smooth Transition Autoregressive Models*. Journal of Business & Economic Statistics, vol. 21, no. 1, pages 104–121, 2003. (Cité en page 51.)
- [Marceau 2000] M.L. Marceau et R. Zmeureanu. *Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings*. Energy Conversion and Management, vol. 41, no. 13, pages 1389 – 1403, 2000. (Cité en page 23.)
- [Maringer 2008] D.G. Maringer et M. Meyer. *Smooth Transition Autoregressive Models – New Approaches to the Model Selection Problem*. Studies in Nonlinear Dynamics & Econometrics, vol. 12, no. 1, page 5, 2008. (Cité en page 51.)
- [McLaughlin 2011] S. McLaughlin, P. McDaniel et W. Aiello. *Protecting consumer privacy from electric load monitoring*. In Proceedings of the 18th ACM conference on Computer and

- communications security, CCS '11, pages 87–98, New York, NY, USA, 2011. ACM. (Cité en page 23.)
- [Medeiros 2005] M.C. Medeiros et A. Veiga. *A flexible coefficient smooth transition time series model*. Neural Networks, IEEE Transactions on, vol. 16, no. 1, pages 97–113, 2005. (Cité en page 51.)
- [Murata 2001] H. Murata et T. Onoda. *Applying Kernel Based Subspace Classification to a Non-intrusive Monitoring for Household Electric Appliances*. In Georg Dorffner, Horst Bischof et Kurt Hornik, éditeurs, Artificial Neural Networks — ICANN 2001, volume 2130 of *Lecture Notes in Computer Science*, pages 692–698. Springer-Verlag, 2001. (Cité en page 24.)
- [Najmeddine 2008] H. Najmeddine, K. El Khamlichi Drissi, C. Pasquier, C. Faure, K. Kerroum, A. Diop, T. Jouannet et M. Michou. *State of art on load monitoring methods*. In Power and Energy Conference, 2008. PCon 2008. IEEE 2nd International, pages 1256 –1258, dec. 2008. (Cité en page 20.)
- [Neenan 2009] B. Neenan et T. Robinson. *Residential electricity use feedback : A research synthesis and economic framework*. Rapport technique, Electricity Power Institute, 2009. (Cité en page 17.)
- [Norford 1996] L.K. Norford et S.B. Leeb. *Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms*. Energy and Buildings, vol. 24, no. 1, pages 51 – 64, 1996. (Cité en pages 27 et 28.)
- [Onoda 2000] T. Onoda, G. Rätsch et K.R. Müller. *Applying Support Vector Machines and Boosting to a Non-Intrusive Monitoring System for Household Electric Appliances with Inverters*, 2000. (Cité en page 24.)
- [Park 2009] Taeyoung Park et David A. van Dyk. *Partially Collapsed Gibbs Samplers : Illustrations and Applications*. Journal of Computational and Graphical Statistics, vol. 18, no. 2, pages 283–305, 2009. (Cité en page 71.)
- [Parson 2011] O. Parson, S. Ghosh, M. Weal et A. Rogers. *Using Hidden Markov Models for Iterative Non-intrusive Appliance Monitoring*. In Neural Information Processing Systems workshop on Machine Learning for Sustainability, December 2011. Event Dates : 17 December 2011. (Cité en pages 24 et 25.)
- [Parson 2012] O. Parson, S. Ghosh, M. Weal et A. Rogers. *Non-intrusive load monitoring using prior models of general appliance types*. In Twenty-Sixth Conference on Artificial Intelligence (AAAI-12), 2012. (Cité en pages 24 et 25.)
- [Patel 2007] S.N. Patel, T. Robertson, J.A. Kientz, M.S. Reynolds et G.D. Abowd. *At the flick of a switch : detecting and classifying unique electrical events on the residential power line*. In Proceedings of the 9th international conference on Ubiquitous computing, UbiComp '07, pages 271–288, Berlin, Heidelberg, 2007. Springer-Verlag. (Cité en pages 19 et 28.)
- [Phillips 1996] J.C. Phillips. *Stretched exponential relaxation in molecular and electronic glasses*. Reports on Progress in Physics, vol. 59, no. 9, page 1133, 1996. (Cité en page 83.)
- [Pollard 1946] H. Pollard. *The representation of e^{-x^λ} as a Laplace integral*. Bulletin of the American Mathematical Society, vol. 52, no. 10, pages 908–910, 1946. (Cité en page 83.)
- [Powers 1991] J.T. Powers, B. Margossian et B.A. Smith. *Using a rule-based algorithm to disaggregate end-use load profiles from premise-level data*. Computer Applications in Power, IEEE, vol. 4, no. 2, pages 42 –47, april 1991. (Cité en page 23.)

- [Punskaya 2002] Elena Punskaya, Christophe Andrieu, Arnaud Doucet et Fitzgerald William. *Bayesian Curve Fitting Using MCMC With Application to Signal Segmentation*. IEEE Transactions On Signal Processing, vol. 50, no. 3, pages 747–758, 2002. (Cit  en pages 50, 57, 58, 61, 77 et 80.)
- [Rabiner 1989] L.R. Rabiner. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE, vol. 77, no. 2, pages 257–286, 1989. (Cit  en page 24.)
- [Rasmussen 2006] C.E. Rasmussen et C.K.I. Williams. Gaussian processes for machine learning. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, 2006. (Cit  en page 89.)
- [Robert 2004] C. Robert et G Casella. Monte Carlo statistical methods. Springer-Verlag, 2nd  dition, 2004. (Cit  en page 65.)
- [Robert 2006] C. Robert. Le choix bay sien : principes et pratique. Springer-Verlag, 2006. (Cit  en page 50.)
- [Roodaki 2012] A. Roodaki, J. Bect et G. Fleury. *Summarizing posterior distributions in signal decomposition problems when the number of components is unknown*. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, pages 3873 – 3876, march 2012. (Cit  en pages 104 et 134.)
- [Russell 1985] M. Russell et R. Moore. *Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition*. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85., volume 10, pages 5 –8, april 1985. (Cit  en page 25.)
- [Schwarz 1978] Gideon Schwarz. *Estimating the Dimension of a Model*. The Annals of Statistics, vol. 2, no. 6, pages 461–464, 1978. (Cit  en page 50.)
- [Shaw 2008] S.R. Shaw, S.B. Leeb, L.K. Norford et R.W. Cox. *Nonintrusive Load Monitoring and Diagnostics in Power Systems*. Instrumentation and Measurement, IEEE Transactions on, vol. 57, no. 7, pages 1445–1454, 2008. (Cit  en pages 27, 28 et 133.)
- [Srinivasan 2006] D. Srinivasan, W.S. Ng et A.C. Liew. *Neural-network-based signature recognition for harmonic source identification*. Power Delivery, IEEE Transactions on, vol. 21, no. 1, pages 398 – 405, jan. 2006. (Cit  en page 24.)
- [Steinwart 2008] I. Steinwart et A. Christmann. Support vector machines. Information science and statistics. Springer-Verlag, 2008. (Cit  en page 98.)
- [Storn 1995] Rainer Storn et Kenneth Price. *Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces*. Rapport technique, International Computer Science Institute, Berkeley, 1995. (Cit  en page 51.)
- [Sultanem 1991] F. Sultanem. *Using appliance signatures for monitoring residential loads at meter panel level*. Power Delivery, IEEE Transactions on, vol. 6, no. 4, pages 1380 –1385, oct 1991. (Cit  en page 27.)
- [Tikhonov 1977] A. N. Tikhonov et V. Y. Arsenin. Solution of ill-posed problems. Winston & Sons, 1977. (Cit  en page 59.)
- [Unser 1999] Michael Unser. *Splines : A Perfect Fit for Signal & Image Processing*. IEEE Signal Processing Magazine, vol. 16, pages 22–38, 1999. (Cit  en page 49.)
- [van Dijk 2002] Dick van Dijk, Timo Ter svirta et Philip Hans Franses. *Smooth transition autoregressive models : a survey of recent developments*. Econometric Reviews, 2002. (Cit  en pages 50 et 133.)

- [Vapnik 2000] V.N. Vapnik. The nature of statistical learning theory. Statistics for Engineering and Information Science. Springer-Verlag, 2nd édition, 2000. (Cité en page 98.)
- [Varodayan 2011] D. Varodayan et A. Khisti. *Smart meter privacy using a rechargeable battery : Minimizing the rate of information leakage*. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pages 1932–1935, may 2011. (Cité en page 18.)
- [Viterbi 1967] A. Viterbi. *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm*. Information Theory, IEEE Transactions on, vol. 13, no. 2, pages 260–269, 1967. (Cité en pages 23 et 24.)
- [Wasserman 2000] Larry Wasserman. *Bayesian Model Selection and Model Averaging*. Journal of Mathematical Psychology, vol. 44, no. 1, pages 92–107, 2000. (Cité en page 50.)
- [Yi-xin 2008] Y. Yi-xin, L. Peng et Z. Chun-liu. *Non-intrusive method for on-line power load decomposition*. In Electricity Distribution, 2008. CIGRE 2008. China International Conference on, 2008. (Cité en pages 24 et 30.)
- [Zeifman 2011] M. Zeifman et K. Roth. *Nonintrusive appliance load monitoring : Review and outlook*. Consumer Electronics, IEEE Transactions on, vol. 57, no. 1, pages 76–84, february 2011. (Cité en page 20.)
- [Zellner 1986] Arnold Zellner. Bayesian inference and decision techniques : Essays in honor of bruno de finetti, chapitre On assessing prior distributions and Bayesian regression analysis with g-prior distributions, pages 233–243. North-Holland/Elsevier, 1986. (Cité en page 59.)