

REVUE DE STATISTIQUE APPLIQUÉE

M. MISITI

Y. MISITI

G. OPPENHEIM

J.-M. POGGI

Décomposition par ondelettes et méthodes comparatives : étude d'une courbe de charge électrique

Revue de statistique appliquée, tome 42, n° 2 (1994), p. 57-77

http://www.numdam.org/item?id=RSA_1994__42_2_57_0

© Société française de statistique, 1994, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

DÉCOMPOSITION PAR ONDELETTES ET MÉTHODES COMPARATIVES : ÉTUDE D'UNE COURBE DE CHARGE ÉLECTRIQUE*

M. Misiti (2), Y. Misiti (1), G. Oppenheim (1, 3), J.-M. Poggi (1, 4)

(1) Université Paris XI – Orsay

Laboratoire de Modélisation Stochastique et Statistique URA 743

Bât. 425, Mathématiques, 91405 Orsay cedex

(2) Ecole Centrale de Lyon

(3) Université de Marne la Vallée

(4) Université Paris X – Nanterre

RÉSUMÉ

Ce texte présente la description statistique d'une courbe de charge électrique et montre l'apport de la transformée en ondelettes pour décrire l'évolution temporelle d'un signal à différentes échelles de temps. La méthodologie statistique utilisée est décrite et des conclusions quant à l'extraction du signal et la détermination d'un pas d'échantillonnage adéquat pour la description et la modélisation, sont proposées. La stratégie de description procède essentiellement par l'utilisation de méthodes comparatives appliquées aux signaux obtenus par décomposition multiéchelle.

Mots-clés : *Courbe de charge électrique, statistique appliquée, méthodes comparatives, ondelettes, analyse multiéchelle.*

SUMMARY

This paper deals with the problem of statistical description of an electrical load consumption and shows an useful application of wavelets decomposition as multiscale analysis. The statistical methodology used is described and conclusions of two problems are given : the signal and noise analysis, and the choice of the sampling period with respect to description and modelization objectives. The description strategy proceeds essentially by successive uses of various comparative methods applied to signals obtained by multiscale decomposition.

Key-words : *Electrical load consumption, applied statistics, comparative methods, wavelets, multiscale analysis.*

* Ce travail a été réalisé dans le cadre d'un contrat de recherche avec Electricité de France, Dept. Consommations et Clientèle, Clamart.

1. Introduction

Ce texte porte sur la description statistique d'une courbe de charge électrique et vise à montrer dans ce cadre, l'apport de la transformée en ondelettes conjuguée à la mise en œuvre de méthodes comparatives variées appliquées aux signaux issus de décompositions multiéchelles de la courbe de charge. En particulier on présente la méthodologie utilisée pour mettre en évidence les informations contenues dans la courbe de charge et propose des conclusions quant à l'extraction du signal et la détermination d'un pas d'échantillonnage adéquat pour la description et la modélisation.

On dispose de mesures faites toutes les minutes d'un phénomène complexe et agrégé : la consommation d'électricité sur l'ensemble du territoire français métropolitain. On désigne cette suite de nombres par enregistrement. La partie utile, signifiante de cet enregistrement est appelée signal tandis que la partie inutile, non signifiante est appelée bruit. D'un point de vue électrique, la principale originalité de ces données est le pas de temps inhabituel.

Le contexte du travail est celui de la prévision et de modélisation de la consommation électrique. En effet, ce travail est extrait d'une étude visant l'amélioration des méthodes mises en œuvre par EDF, pour la prévision à court terme de la consommation d'électricité. Plus précisément, la prévision est actuellement opérée sur des données échantillonnées à 30 minutes. Ces données, après suppression de diverses composantes (liées à la température, aux effets calendriers, aux points aberrants et à des actions externes connues comme les modifications de tarifs), sont modélisées par un processus SARIMA. Ce modèle permet de prédire à des horizons multiples de 30 minutes jusqu'à 48 heures. L'objectif d'EDF est de fournir des prévisions aux horizons multiples de 5 minutes. La démarche SARIMA précédente n'a pu, pour le moment, être reconduite pour ce pas de temps.

Les données disponibles au pas de 1 minute font donc l'objet de cette étude. Ces données, pour lesquelles aucun modèle n'est construit, permettent d'envisager d'améliorer la prévision actuelle par insertion des informations extraites de la description et de la modélisation aux pas d'échantillonnage inférieurs à 30 minutes. L'objectif est de déterminer ces informations. L'utilisation de la décomposition en ondelettes est alors, dans ce contexte, très naturelle. En effet elle fournit une décomposition multiéchelle du signal de consommation électrique en une somme de signaux orthogonaux correspondant à des échelles de temps différentes. Ceci permet d'apprécier l'apport d'information de chacune d'elles.

La difficulté d'analyse de la courbe de charge échantillonnée à la minute provient de deux sources. D'une part, les informations externes, hormis celles constituées par la courbe elle-même, sont peu nombreuses. D'autre part, les informations disponibles ne sont pas homogènes aux questions posées. La stratégie de description procède alors essentiellement par l'utilisation de méthodes comparatives appliquées aux signaux obtenus par décomposition multiéchelle par ondelettes de la courbe de charge. La comparaison s'appuie sur un témoin issu d'informations externes sur le phénomène électrique sous-jacent.

Le plan proposé est le suivant : après cette introduction fixant le contexte du travail, les données ainsi que les informations électriques disponibles font l'objet du

paragraphe 2 qui précise en outre les questions posées et les objectifs de la description. Les outils et les méthodes, qui sont caractéristiques de ce type de travail de statistique appliquée, sont définis dans le paragraphe 3. Les principaux éléments de conclusions en insistant surtout sur la démarche mise en œuvre et les aspects méthodologiques, font l'objet du dernier paragraphe.

2. Le problème, les données

2.1. Le problème de description

Deux questions sont abordées, la première concerne l'extraction du signal de l'enregistrement et la seconde vise à déterminer un pas d'échantillonnage adéquat pour la description et la modélisation.

Problème 1 : Signal et bruits

- Quelles sont les informations contenues dans l'enregistrement? Lesquelles sont utiles?
- Y a-t-il plusieurs types de bruits et peut-on les distinguer?
- Peut-on séparer le bruit du signal?

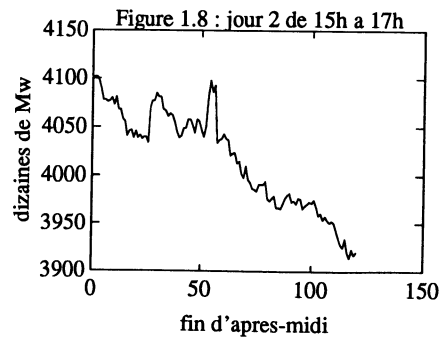
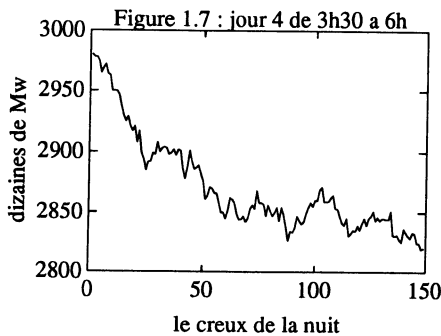
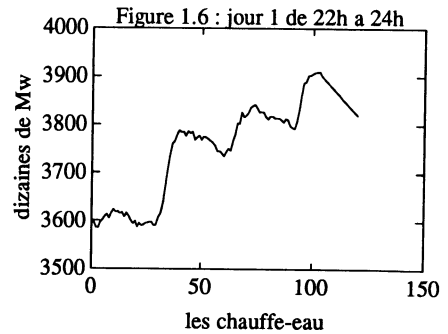
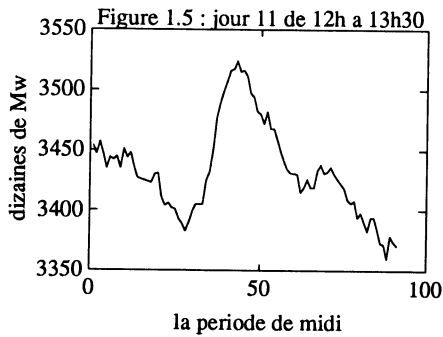
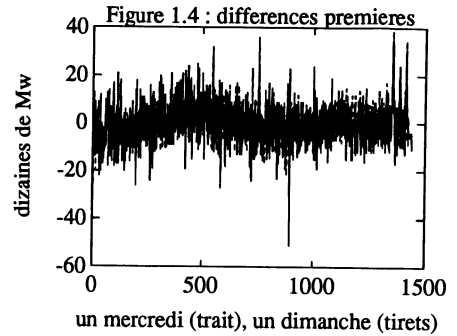
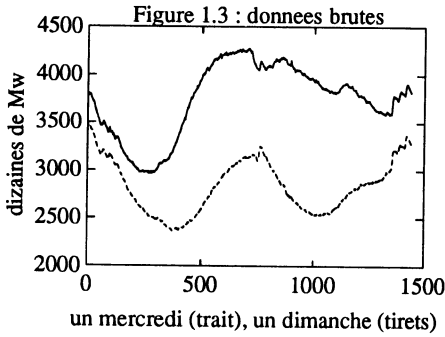
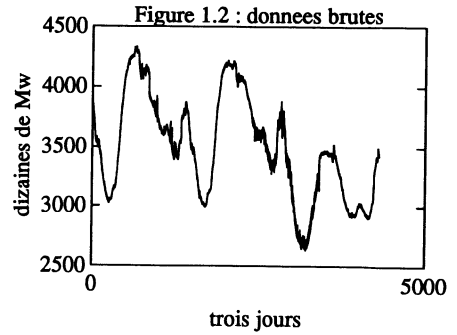
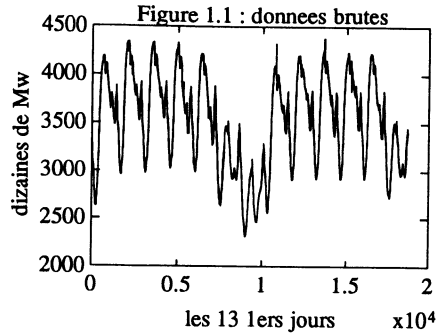
Problème 2 : Le pas d'échantillonnage

- L'utilisation d'un pas de 4, 5 ou 10 minutes ne serait-il pas suffisant pour la description et la modélisation, en vue de la prévision par pas de 5 minutes?

2.2. Les données

Les données traitées sont l'enregistrement de la consommation électrique sur l'ensemble du territoire français métropolitain, échantillonné à la minute durant cinq semaines des mois de juin et juillet 1990, soit une chronique de 50400 points qui est présentée dans la figure 2.2 (les figures sont regroupées en plusieurs planches et sont indexées par le n° de planche suivie du n° du graphique). La figure 1.1 présente un extrait de 13 jours des données illustrant la périodicité hebdomadaire et la pseudo-périodicité journalière. La figure 1.3 est la superposition d'un mercredi et d'un dimanche, sur laquelle on peut noter aussi la régularité globale de la série. Elle est accompagnée de la figure 1.4 constituée de leurs différences premières. On remarque ici la forte irrégularité locale de la série. Les figures 1.5 à 1.8 présentent des formes caractéristiques du signal. Les deux premières : la période de midi et celle des chauffe-eau, seront très utiles dans la suite puisqu'elles correspondent à des périodes dans lesquelles le signal présentent des fluctuations à la fois fortes et rapides.

Des informations externes fournies par les ingénieurs électriciens (des compléments peuvent être trouvés dans Karanta, Ruusunen 91), sont disponibles et utilisées dans la suite :



FIGURES 1.1 À 1.8

1) Une variation instantanée de l'ordre de 0,3% (soit environ 100 Mw) est considérée comme significative.

2) L'enregistrement est l'agrégation des mesures de plusieurs centaines de capteurs, engendrant des erreurs de mesure.

3) La consommation est due environ pour moitié à l'industrie et pour moitié aux particuliers.

4) Il y a environ 25 millions de particuliers. L'enregistrement des particuliers est l'agrégation de ces enregistrements individuels.

5) L'enregistrement des activités industrielles est constante par morceaux et ses variations sont à basse fréquence. Les variations imputables aux particuliers peuvent être à haute fréquence.

6) Les périodicités fondamentales, quotidienne et hebdomadaire, liées au rythme de l'activité économique sont connues.

7) Les instants journaliers, de changements de la consommation liés à des décisions tarifaires sont connus, par exemple les démarrages de chauffe-eau en fin de soirée.

8) Certaines données (des jours 17 et 20 surtout) sont manquantes. Elles ont été remplacées par estimation des données demi-horaires et lissage par des fonctions splines pour les données intermédiaires. En outre, les données aberrantes n'ont pas été corrigées.

9) L'erreur de mesure est, à cause d'un dysfonctionnement massif des capteurs, anormalement grande du jour 10 à 16h30 au lendemain à 9h.

2.3. Essais de définitions

On précise ici ce que nous entendons par signal, bruit de mesure et bruit de dynamique.

Sans pouvoir définir le signal, on peut néanmoins dire que toute partie répétitive est affectée au signal, plus précisément : les aspects périodiques sont du signal; les répétitions, à des dates périodiques, d'enregistrements de caractéristique constante sont du signal.

D'autre part, il y a deux types de bruits affectant le signal électrique : un bruit d'état, appelé aussi bruit de dynamique, et un bruit de mesure. Le bruit de mesure est lié aux capteurs. Le bruit de dynamique ou bruit d'état, est constitué des variations rapides et considérées par les ingénieurs électriciens comme non significatives de la consommation.

La séparation entre ces deux bruits est souhaitable mais vraisemblablement impossible à effectuer dans la plupart des cas. Cependant on peut penser que les bruits attribuables aux capteurs comprennent plus de hautes fréquences que les bruits de dynamique. La suppression conjointe des deux bruits dans les hautes fréquences n'est pas gênante et est même considérée comme souhaitable.

3. Outils et aspects méthodologiques

On détaille dans ce paragraphe les aspects méthodologiques en commençant par une présentation succincte de l'outil essentiel de l'étude : la décomposition en ondelettes. Des détails et références pourront être trouvés dans deux articles des auteurs publiés dans cette revue (*cf.* Misiti *et al.* 93, 2 et 3). Ensuite on présente le squelette de la méthode articulée autour d'une stratégie de comparaison des analyses, puis trois outils de description accompagnés des méthodes de comparaison associées.

3.1. Un outil d'analyse multiéchelle : la décomposition en ondelettes

Les séries temporelles présentent en général, de nombreuses caractéristiques non-stationnaires. Ces aspects constituent une part importante de l'information contenue dans la série : tendance, composante saisonnière, ruptures, début et fin d'événements, phénomènes transitoires. La plupart des approches mathématiques utilisées sont adaptées aux processus stationnaires. De nouvelles méthodes élaborées et mises au point ces dernières années, unifient, généralisent des idées et des pratiques empiriques plus anciennes, et permettent d'analyser des signaux non-stationnaires. La transformée en ondelettes fait partie de ces dernières, et son principe est de décrire, en fournissant des informations sur la régularité locale, l'évolution temporelle d'un signal à différentes échelles de temps.

De nombreux textes de présentation de la théorie des ondelettes sont disponibles, on peut citer entre autres les trois premiers exposés dans Lemarié 90, Mallat 89, Daubechies 88, Strang 89, Gasquet-Witowski 90 et bien sûr le livre d'Y. Meyer (*cf.* Meyer 90). Pour l'analyse par ondelettes de signaux aléatoires, on pourra consulter Cohen 90, Cohen *et al.* 91, Froment 90 et Istat 92.

Au moyen de la décomposition en ondelettes et selon la démarche usuelle de la statistique descriptive des suites d'observations dans le temps, on recherche une description globale de l'évolution (tendances, phénomènes de grande période) complétée par des bruits et des effets locaux : ruptures, discontinuités ainsi que les phénomènes de courte période, répétitifs ou transitoires, c'est-à-dire localisés dans le temps et s'estompant plus ou moins rapidement. Ces évolutions seront classées en détails par opposition aux premières rangées dans la tendance.

Les liens entre les analyses multirésolution dyadiques et les bases orthonormées d'ondelettes permettent de décomposer un enregistrement S échantillonné au pas p , en une somme de signaux orthogonaux associés à des échelles de temps différentes, c'est-à-dire :

$$S = T(-j) + D(-1) + D(-2) + \dots + D(-j) \quad (1)$$

où j est un entier positif, $T(-j)$ est la tendance de niveau $-j$, contenant les composantes de S de «période» supérieure à $p \cdot 2^j$, et où $D(-k)$ est le détail de niveau $-k$, contenant les composantes du signal de «période» comprise entre $p \cdot 2^{k-1}$ et $p \cdot 2^k$.

La formule (1) est obtenue en appliquant itérativement la décomposition suivante :

$$T(-k) = T(-k - 1) + D(-k - 1),$$

au signal S assimilé à $T(0)$.

Les ondelettes utilisées dans ce travail sont les ondelettes à support compact dues à I. Daubechies (cf. Daubechies 88). Les algorithmes pratiques de décomposition et de reconstitution, qui ont été programmés en Matlab (cf. Matlab), sont ceux proposés par S. Mallat (cf. Mallat 89).

Pour mener à bien cette étude, nous avons été amené à construire un logiciel de décomposition en ondelettes incluant des fonctionnalités graphiques de haut niveau (cf. Misiti *et al.* 93, 1).

3.2. Méthode générale de description

On présente le squelette de la méthode, puis trois outils de description accompagnés des méthodes de comparaison associées.

La démarche de travail consiste en des applications répétées des trois étapes suivantes :

1) Préciser les informations utilisées et calculer les données à analyser, typiquement par décomposition en ondelettes, par sous-échantillonnage ou par extraction locale de signaux de détails et de tendances.

2) Une stratégie comparative des analyses est mise en œuvre. Elle revêt deux aspects différents :

– comparaison des analyses à l'analyse d'un témoin déduit des informations externes disponibles, par exemple : périodicités, démarrages de chauffe-eau ou des événements particuliers comme le dysfonctionnement des capteurs;

– lorsqu'il n'existe pas de témoin, une comparaison entre elles de plusieurs analyses est pratiquée, en choisissant *a posteriori* l'une d'elles pour témoin si cela est pertinent.

3) Une méthode de modélisation vient éventuellement compléter la description ou aider à la comparaison.

3.3. Outils de description et méthodes comparatives associées

3.3.1. Modèles ARMA

Principe : les signaux de détails des décompositions en ondelettes étant raisonnablement simples et stationnaires, au moins pour les hauts niveaux de résolution, on peut parfois recourir à des modèles ARMA pour les décrire localement, à l'aide du logiciel Mandrake (cf. Azencott *et al.* 88). Les résultats sont alors utilisables à des fins de comparaison et d'étude de stabilité.

Comparaison de modèles ARMA : la situation typique est la suivante : on désire comparer les modèles ARMA adéquats pour deux intervalles de temps distincts, disons les périodes 1 et 2, par exemple la période de midi pour deux jours différents.

Pour chacune des périodes, le logiciel Mandrake permet de déterminer une liste ordonnée par un critère de type Akaike, de modèles adéquats.

La structure de modèle M1, classée en premier pour la période 1, est proposée à Mandrake comme modèle à interclasser pour la période 2. On désigne par P1/2 son classement. On procède de même avec le modèle M2 et on obtient P2/1.

Les modèles M1 et M2 sont dits compatibles lorsque l'une des positions P1/2 et P2/1, est voisine de 1.

Une description spectrale des signaux analysés peut aussi être utile.

3.3.2. Spectres

Principe : le spectre d'un signal est calculé en deux étapes : une transformée de Fourier fournit un périodogramme qui est ensuite lissé par une fenêtre triangulaire, pour obtenir l'estimateur de la densité spectrale.

Comparaison de spectres : la situation typique est la suivante : on désire comparer les estimateurs des spectres pour deux intervalles de temps distincts. La comparaison des estimateurs se fait d'abord visuellement puis est quantifiée par l'évaluation d'une distance basée sur la distance de Hellinger entre deux densités (cf. Dacunha-Castelle, Duflo 82 p. 160).

Néanmoins le contenu spectral des signaux de détails étant largement déterminé par le filtre associé à l'ondelette analysante et le niveau de résolution, on a recours à des repérages de formes.

3.3.3. Analyse de formes

Principe : l'analyse de formes recherche une forme typique. Par exemple, le démarrage des chauffe-eau à 22h30 est la réponse d'un système passif à des décisions tarifaires. Le signal ressemble à la réponse à une impulsion d'un modèle d'équation linéaire du second ordre amorti. On recherche cette forme dans l'enregistrement ou dans un signal du même type.

Comparaison de formes : la ressemblance est quantifiée par un coefficient de similarité entre signaux construit à l'aide d'un noyau gaussien (cf. Duflo 90 p. 188). Plus précisément, soit X et Y deux blocs de r observations consécutives, issus d'un même signal S de longueur $N \gg r$. On pose alors :

$$\text{sim}(X, Y) = c \cdot \exp(-\|X - Y\|^2 / 2 \cdot h^2) \quad (2)$$

où c est une constante ne dépendant que de r ,

$h = \sigma(S) / N^{1/(r+4)}$, où $\sigma(S)$ est l'écart-type de S et

$\|\cdot\|$ est la norme euclidienne sur \mathbb{R}^r .

Ce choix de h est issu d'un réglage effectué pour une étude préliminaire de prévision non paramétrique par des techniques de noyau, de la courbe de charge. Sa forme est néanmoins classique (cf. Bosq, Lecoutre 92).

4. Démarche de solution et conclusions proposées

Ce paragraphe se compose de trois parties. La première renvoie à deux exemples d'analyses parmi celles qui ont servi à calibrer les outils utilisés. Les deux dernières parties de ce paragraphe fournissent des éléments de réponse aux deux problèmes de description posés en mettant l'accent sur la démarche mise en œuvre.

Dans ce qui suit, on note t le temps, $-k$ le niveau de décomposition (parfois noté k sur les graphiques), $D(-k)$ et $T(-k)$ le détail et la tendance de niveau $-k$ et $D(-k, t)$ le détail de niveau $-k$ à l'instant t , d le pas d'échantillonnage. Tous les signaux, détails, tendances, sont représentés dans le temps initial de l'enregistrement ou de l'enregistrement échantillonné.

4.1. Ondelettes et description locale

Pour les analyses locales, on choisit $j = 5$ dans (1) (car $2^5 = 32$) de façon à étudier les composantes du signal de période inférieure à la demi-heure. Dans le § 3.2 de Misiti *et al.* 93, 3, deux exemples d'analyses portant sur les périodes présentées dans les figures 1.5 et 1.7, sont détaillés et commentés, il ne seront pas repris ici. Des études locales de ce type ont permis en particulier de choisir l'ondelette analysante utilisée ici, à savoir l'ondelette de I. Daubechies d'ordre 3 notée $daub3$.

4.2. Signal et bruits

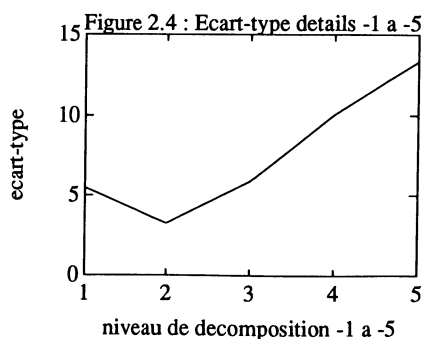
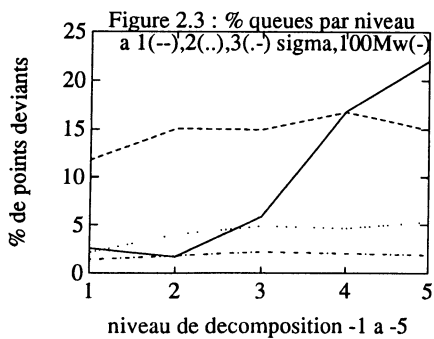
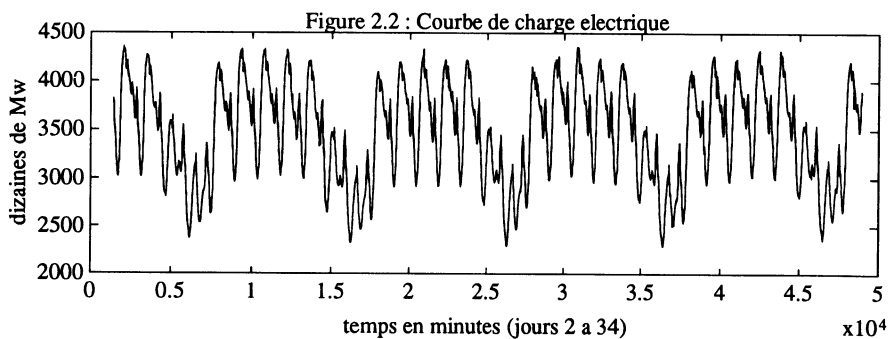
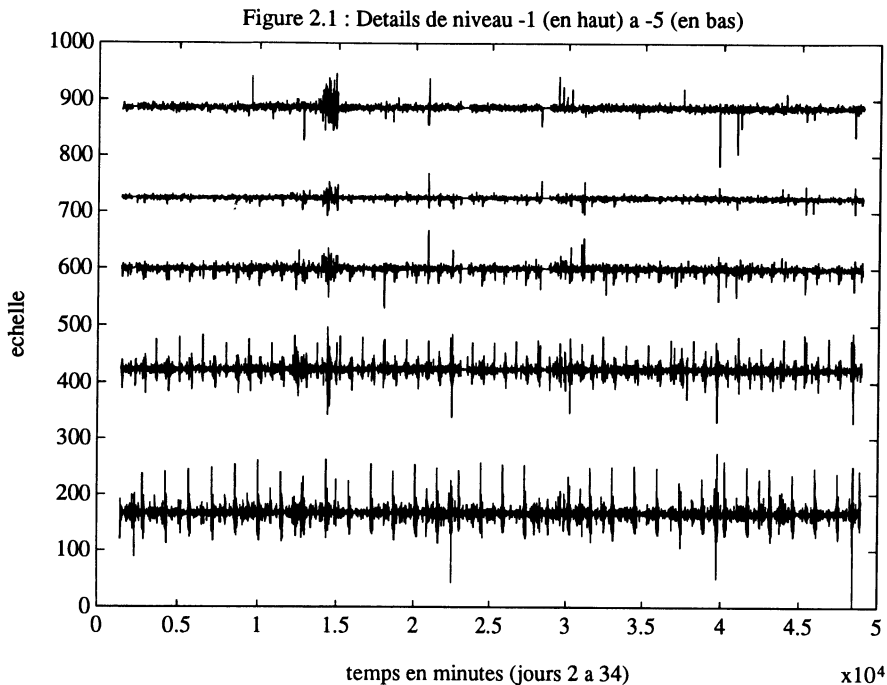
On aborde maintenant le premier problème posé qui porte sur la séparation dans l'enregistrement, du signal et des bruits. Trois types d'approches sont considérées : une approche globale, une approche locale à l'aide des Arma et des spectres et enfin une approche basée sur la reconnaissance des formes qui fournit les indications les plus intéressantes.

On ne considèrera que les cinq premiers niveaux des décomposition, c'est-à-dire que l'on se restreindra à l'étude des phénomènes de période inférieure à 32 minutes pour l'étude du bruit.

4.2.1. Analyses préliminaires

On trouve dans les figures 2.1 et 2.2 la décomposition par ondelettes de l'enregistrement tout entier, ainsi que cet enregistrement. L'enregistrement décomposé et les signaux de détails sont dans la même base de temps, conformément à (1). Ceci permet une lecture synchrone de ces signaux. Les tendances correspondantes sont omises, en effet l'ordre de grandeur des détails relativement à celui de la chronique, fait que les graphiques associés sont peu porteurs d'informations.

On peut formuler quelques commentaires :



FIGURES 2.1 À 2.4

1) On retrouve sur la série initiale (figure 2.2) les périodicités quotidienne et hebdomadaire;

2) On retrouve le fait que la série n'est pas corrigée des points aberrants qui se lisent grâce aux très grandes valeurs de $D(-1)$, lequel se trouve en haut de la figure 2.1;

3) On retrouve nettement les deux longues séquences de données reconstituées pour les jours 17 et 20 qui se lisent grâce aux très faibles valeurs de $D(-1)$, $D(-2)$ et $D(-3)$ pour les périodes correspondantes; on distingue un phénomène semblable dans la nuit du jour 1 au jour 2.

4) Enfin on distingue nettement la période de dysfonctionnement des capteurs correspondant à une longue séquence de valeurs anormalement grandes de $D(-1)$, $D(-2)$ et plus faiblement, $D(-3)$.

5) Pour le niveau de détail -1 , il semble que, sauf valeurs exceptionnelles (cf. remarques 2 et 4 ci-dessus), toutes les valeurs sont dans une «bande»;

6) Il semble que des points ou des séquences de points sortent nettement d'une bande, de façon périodique pour $D(-3)$, $D(-4)$ et $D(-5)$. Visuellement ce n'est pas le cas pour $D(-1)$ et $D(-2)$.

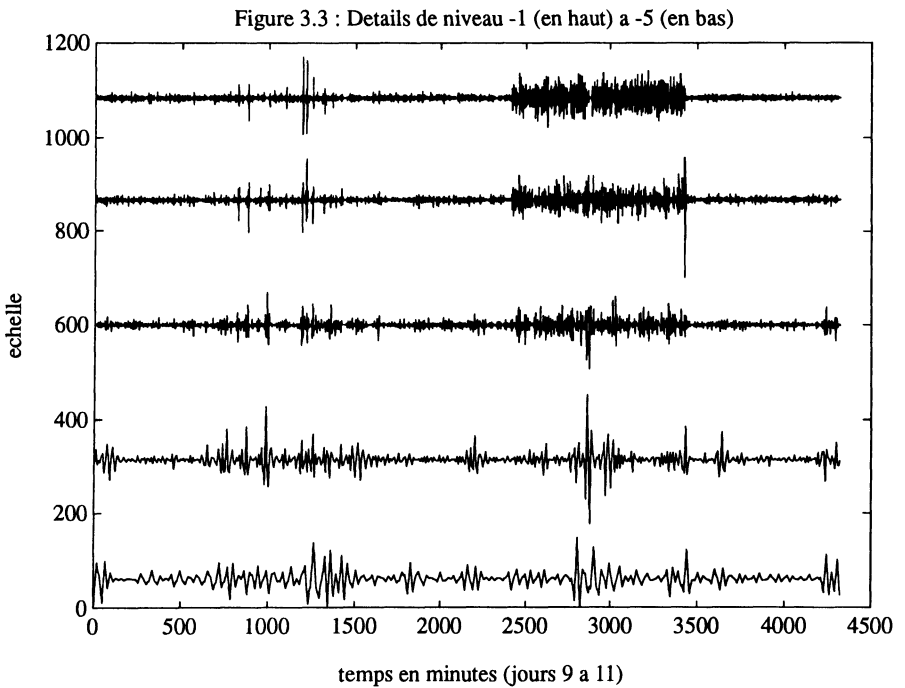
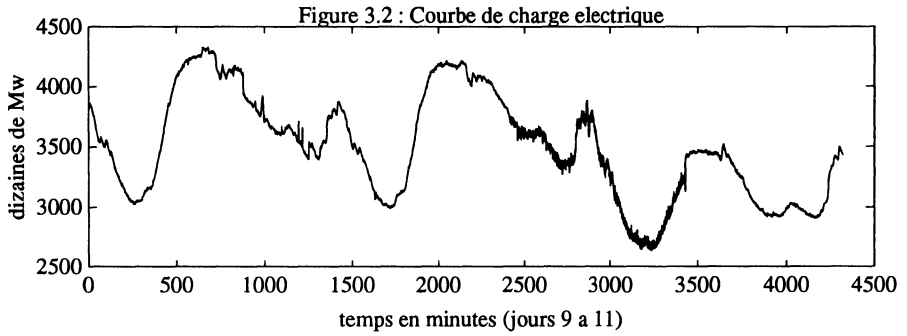
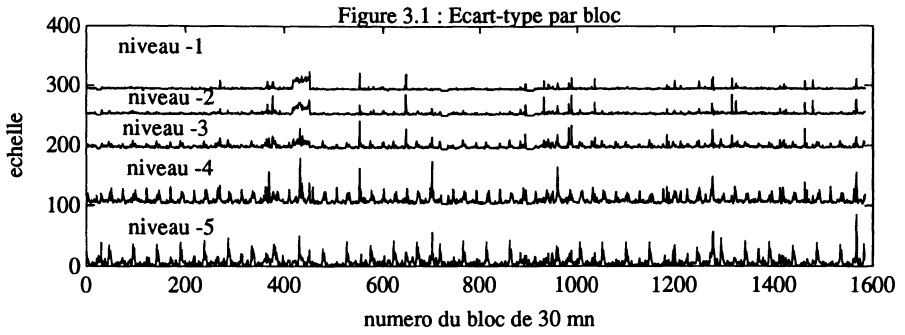
Les deux derniers points sont étayés par les figures 2.3 et 2.4. Sur la figure 2.3, on lit en abscisse le niveau et en ordonnée quatre courbes contenant : le pourcentage de points hors de l'intervalle $\pm a\sigma(D(-i))$ (où $\sigma(D(-i))$ est l'écart-type empirique de $D(-i)$), pour $a = 1$ (courbe en tirets), $a = 2$ (courbe en pointillés), $a = 3$ (courbe en $-$) ainsi que le pourcentage de points déviants de ± 100 Mw (courbe en trait plein) correspondant à 0,3% du niveau moyen : le seuil en deça duquel le signal est considéré comme non significatif. Sur cette dernière courbe on constate une rupture pour le niveau -3 . Des conclusions semblables sont tirées de l'examen de l'évolution en fonction du niveau de l'écart-type des détails (cf. figure 2.4) ainsi que de l'écart-type local sur 30 points (cf. figure 3.1).

4.2.2. Analyses locales

On extrait de la décomposition précédente la décomposition de l'enregistrement des trois jours 9, 10 et 11, au pas $d = 1$ mn (cf. figures 3.2 et 3.3). On retrouve bien sûr, les points 2 et 4 du paragraphe précédent que l'on peut maintenant essayer de préciser. L'effet du dysfonctionnement des capteurs se lit sans peine sur les détails de niveau -1 à -3 . En revanche, il n'est plus visible pour les niveaux -4 et -5 .

Ceci suggère de faire une étude locale du bruit en utilisant l'information *a priori* sur le dysfonctionnement des capteurs. L'hypothèse sous-jacente est que les deux bruits (d'état et de mesure) sont simultanément présents. Le bruit d'état est le même pour une période donnée de la journée, pour deux jours distincts alors que le bruit de mesure de la nuit des jours 10 à 11 est différent de celui des autres jours. Deux voies sont explorées :

- pour les niveaux -1 à -3 , analyses spectrales locales et comparaison;
- pour les niveaux -1 à -4 , modélisations ARMA locales et compatibilité, sur la période [21h-03h], ainsi que deux sous-périodes : [22h-00h], période contenant un



FIGURES 3.1 À 3.3

signal fort (les chauffe-eau) et [01h-03h], période à signal faible (descente vers le «creux de la nuit»).

Comparaison des spectres

Pour ôter l'effet de variance, on détermine d'abord l'énergie par évaluation de l'écart-type de la suite considérée et on norme le signal. Le spectre est ensuite calculé comme indiqué en 3.3.2. On étudie pour chacune des périodes considérées, les 34 ou 35 spectres (un par jour) puis on compare la nuit 10 à toutes les autres. Chaque niveau de -1 à -3 est étudié séparément.

La nuit 10, après suppression de l'effet de variance massif, ressort essentiellement au niveau -3 du point de vue du spectre. Cette constatation plaide en faveur de l'idée qu'il y a peu de signal dans les détails de niveaux -1 et -2 .

Comparaison des modèles ARMA

On compare une nuit «standard» : la nuit 4, à la nuit 10. Les résultats sont les suivants :

- période d'ensemble [21h-03h] : la structure du modèle est presque toujours un AR(10). Quel que soit le niveau, les modèles sont compatibles. Au niveau -3 , le même modèle est spontanément classé en tête pour les 2 jours. Au niveau -4 , il en est de même pour une liste ordonnée de cinq modèles. Remarquons que ces modèles sont non seulement de même structure mais de coefficients proches. D'autre part les polynômes AR ont des zéros proches du cercle unité, ce qui explique les très forts taux de variance expliquée (de 90 à 99% suivant les niveaux).

- période [22h-00h] (signal fort) : les modèles obtenus sont peu ou pas compatibles.

- période [01h-03h] (signal faible) : au niveau -1 , les modèles sont faiblement compatibles, tandis que pour les autres niveaux les modèles sont compatibles.

Les modèles ARMA sont compatibles, sauf lorsque le signal est fort. Ceci peut s'expliquer par l'importance du bruit de mesure dans la nuit 10 aux hautes fréquences relativement au signal, ce dernier est «masqué».

Malheureusement, les tentatives d'extension des modèles obtenus pour les hauts niveaux de résolution, à des périodes plus longues ou adjacentes, révèlent une instabilité. Ce manque de stabilité provient des points aberrants et des données reconstituées. Il est malaisé de tirer des conclusions globales de ces deux études, c'est pourquoi on se concentre ensuite sur l'analyse et la comparaison des formes.

4.2.3. Analyses de formes

Dans cette partie on cherche à mettre en évidence trois aspects : $D(-1)$ ne contient pas de signal; $D(-2)$ contient du signal mais seulement localement et faiblement; $D(-3)$, $D(-4)$ et $D(-5)$ contiennent du signal. Ces éléments lisibles sur les décompositions en ondelettes ne peuvent pas être étayés par des approches globales comme l'analyse spectrale. En effet le signal se manifeste, dans les hauts

niveaux de résolution, non pas par des composantes périodiques mais plus faiblement par des motifs qui se répètent à des dates périodiques (*i.e.* tous les jours à la même heure) qui n'engendrent pas en général de composantes périodiques. On adopte donc une optique de type reconnaissance des formes : pour une séquence de la série, choisie comme témoin à partir d'informations *a priori*, on recherche parmi tous les blocs de même longueur ceux qui lui sont les plus semblables. S'ils apparaissent à des dates périodiques, il y a du signal. Plus précisément on procède comme suit :

- on choisit une période de temps de longueur r pas de base, notée I , dans le jour 1;
- pour chaque niveau $-k$:
 - on choisit pour bloc témoin $BT = (D(-k, t), t \in I)$;
 - on calcule pour tous les blocs de longueur r issus de $D(-k)$ restreint aux jours 2 à 34, la similarité (au sens de (2)) avec BT , on obtient ainsi la suite des $s(j) = \text{sim}(BT, (D(-k, t), t = j, j + 1, \dots, j + r - 1))$
 - on calcule ensuite pour chaque minute m de la journée, la moyenne des similarités sur les 33 jours, notée $\text{sbar}(m)$ pour $m = 1, 2, \dots, 1440$
 - les grandes valeurs de sbar indiquent les dates de début de blocs les plus semblables à BT , qu'il reste à comparer à la date de début de BT : si celle-ci ressort nettement alors la forme se répète tous les jours et il y a du signal, sinon il n'y en a pas.

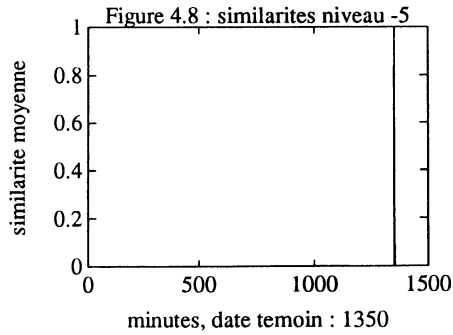
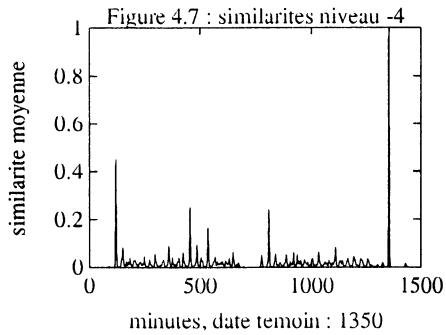
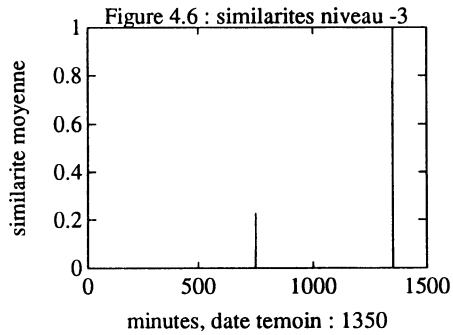
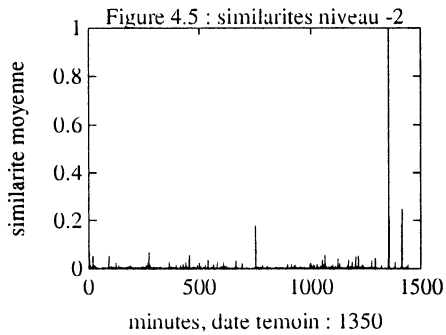
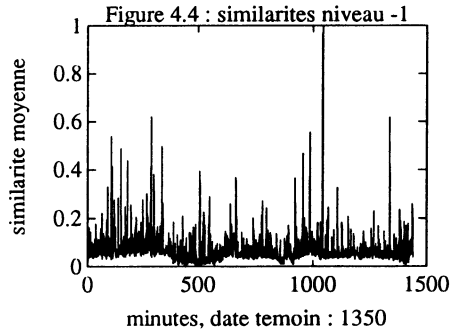
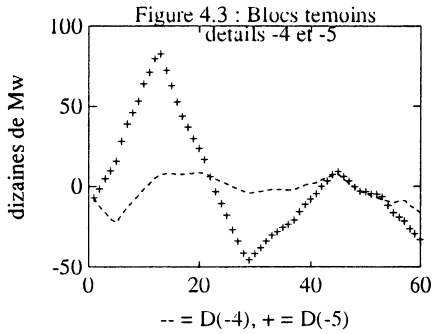
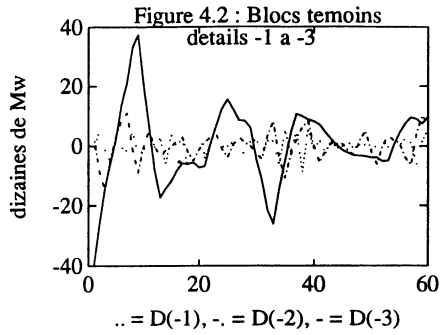
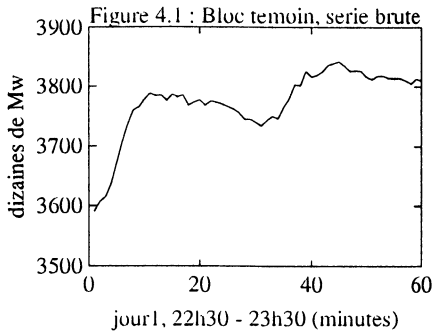
Pour chacune des analyses, on trace les blocs témoins de la série brute et des détails -1 à -5 ainsi que, pour chaque niveau, les $(\text{sbar}(j), j)$ correctement normalisés.

L'étude comprend trois phases. Les deux premières consistent à calibrer les outils décrits ci-dessus sur les deux périodes de la journée contenant des composantes de courte période clairement attribuables au signal. D'après les analyses locales, ces périodes sont celle des démarrages des chauffe-eau en fin de soirée et celle de midi. Ceci permettra de conclure quant au signal. La dernière phase étudie la sensibilité des conclusions relatives aux bruits.

L'étude concernant la période des chauffe-eau est regroupée dans les figures 4.1 à 4.8. On trouve dans la figure 4.1 le bloc témoin de la série brute le jour 1 pour la période témoin 22h30-23h30 ($r = 60$) qui débute à la date témoin de l'analyse 1350. Dans les deux figures suivantes (4.2 et 4.3) sont tracés les blocs correspondant à la même période témoin, pour les détails de niveau -1 à -3 (fig. 4.2) puis -4 et -5 (fig. 4.3). Enfin dans les figures 4.4 à 4.8 on trouve pour chaque niveau de résolution un graphique comportant en abscisse les dates de 1 à 1440 et en ordonnée la moyenne des similarités des blocs débutant au même instant de la journée.

Commentons ces cinq dernières figures :

- 1) Au niveau -1 (fig. 4.4), la date 1350 ressort moins que beaucoup d'autres;
- 2) Au niveau -2 (fig. 4.5), la date 1350 ressort. Ceci est lié aux démarrages très rapides des chauffe-eau, il y a donc du signal dans $D(-2)$ mais faible comme on peut le voir sur le bloc témoin de $D(-2)$ (dessiné dans la figure 4.2). En outre d'autres similarités ne sont pas négligeables;



FIGURES 4.1 À 4.8

3) Au niveau -3 (fig. 4.6), deux dates ressortent : 1350 (le plus fortement) et 750 (12h30), ce qui est conforme aux études locales : les creux ressortent au niveau -3 ;

4) Au niveau -4 (fig. 4.7), la date 1350 ressort. De nombreuses autres ne sont pas négligeables car la morphologie des bosses ressort au niveau -4 , mais pas leurs amplitudes (cf. figure 4.3);

5) Au niveau -5 (fig. 4.8), seule ressort la date 1350 car les bosses ressortent au niveau -5 (cf. figure 4.3).

L'étude relative à la période de midi est quant à elle, présentée dans les figures 5.1 à 5.8 selon un schéma identique à la précédente. La période témoin est 12h-13h ($r = 60$) et la date témoin est 720 (12h). Les conclusions sont sensiblement les mêmes que pour la période précédente, sauf l'inversion pour les niveaux -4 et -5 .

On déduit de ces deux études que $D(-1)$ semble ne pas contenir de signal; $D(-2)$ contient du signal mais seulement localement et faiblement; $D(-3)$, $D(-4)$ et $D(-5)$ contiennent du signal.

Des études complémentaires relatives au détail de niveau -1 , pour diverses périodes de la journée et pour des blocs de longueurs variées : 10, 20, 30 et 60 minutes, ont été effectuées. Elles montrent que la conclusion précédente concernant $D(-1)$, à savoir que la date témoin ne ressort jamais nettement, ne dépend ni de la date de début du bloc témoin, ni de sa longueur. Ceci confirme que le détail de niveau -1 ne contient pas de signal.

4.3. Le pas d'échantillonnage

On cherche à fixer une borne plausible notée d_{\min} , valeur du pas d en dessous de laquelle il paraît inutile d'échantillonner pour la description et la modélisation. La détermination d'une borne inférieure pourrait se faire grâce à la recherche de ruptures en fonction du niveau d'une ou plusieurs caractéristiques des détails, en étudiant de plus près les éléments fournis par les analyses préliminaires en 4.2.

On préfère ici utiliser les conclusions tirées du thème «signal et bruits» pour construire une situation témoin. Notons que l'étude précédente suggère que d_{\min} se situe entre 4 et 8 minutes, car le niveau -3 est le premier pour lequel le détail correspondant contient des informations significatives clairement attribuables au signal.

On se sert pour construire le témoin, des conclusions suivantes :

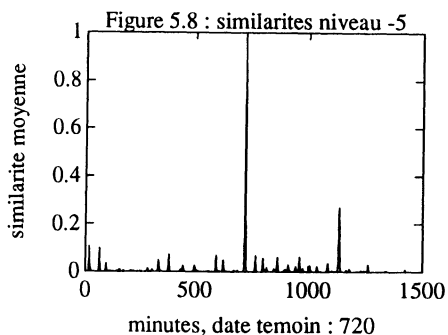
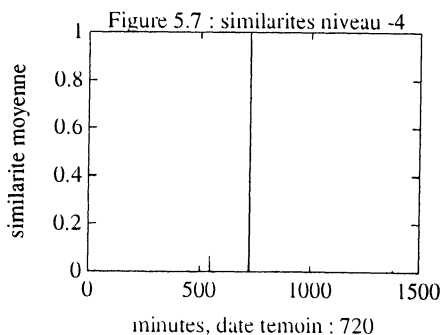
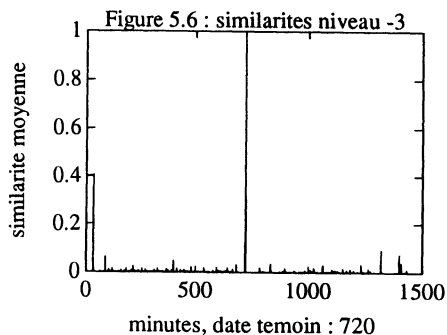
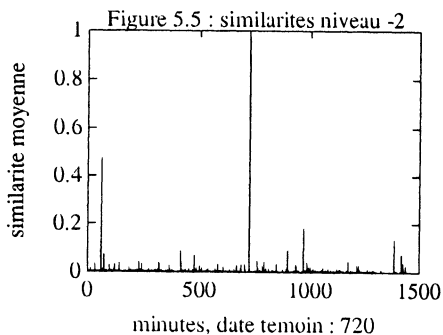
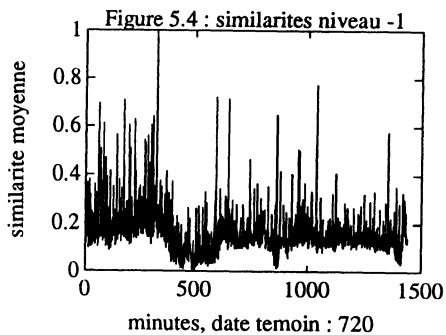
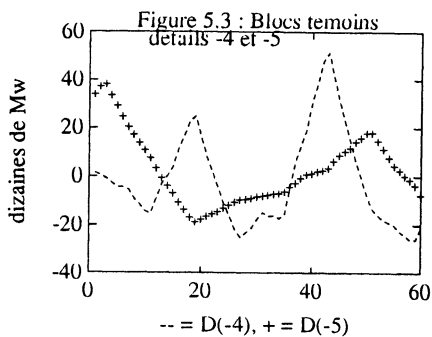
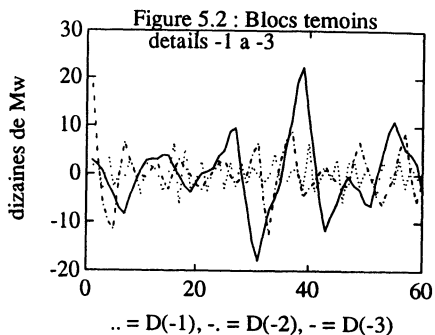
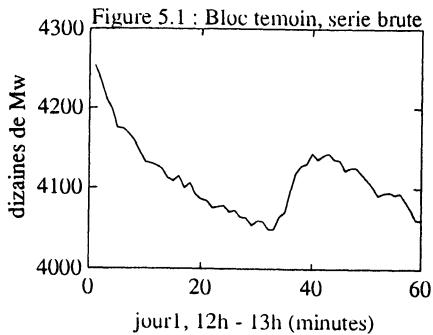
- au pas $d = 1$ mn, on met en évidence que $D(-1)$ ne contient pas de signal, accompagné d'un effet de «bande»;

- les composantes décelables du signal, qui sont à plus haute fréquence sont situées lors de la période des chauffe-eau et celle de midi.

Une procédure de détermination du pas minimum s'en déduit naturellement. Elle consiste à augmenter le pas jusqu'à ce que le détail de niveau -1 de l'enregistrement échantillonné contienne du signal. Plus précisément on applique le schéma d'algorithme suivant :

$$d = 0;$$

faire



FIGURES 5.1 À 5.8

$$d = d + 1;$$

calculer le détail -1 de l'enregistrement échantillonné à d minutes, noté $D_d(-1)$;

étudier la forme des chauffe-eau par la procédure précédente;

jusqu'à ce que $D_d(-1)$ contienne du signal significatif;

$$d_{\min} = d;$$

Lorsque l'on examine $D_d(-1)$ pour $d = 2, 3, 4$, et 5 minutes, il semble visuellement y avoir une rupture pour $d = 4$ ou 5 minutes. Pour étayer cet indice on étudie alors une période particulière dans laquelle le signal à haute fréquence est fort : la période des chauffe-eau.

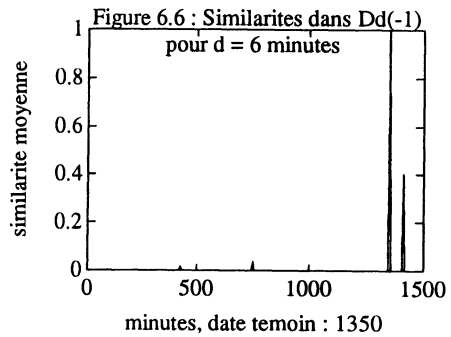
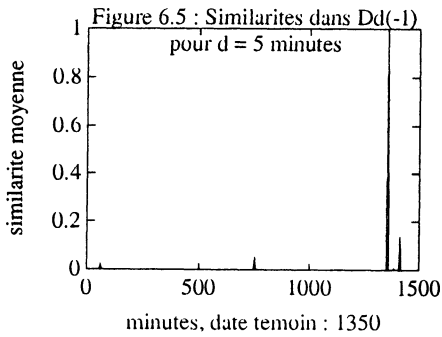
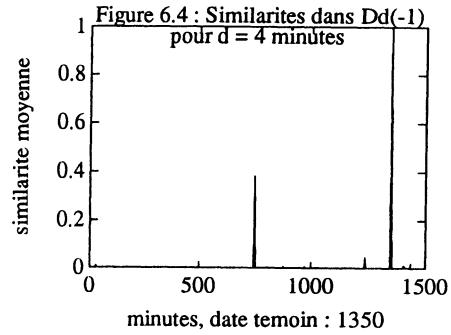
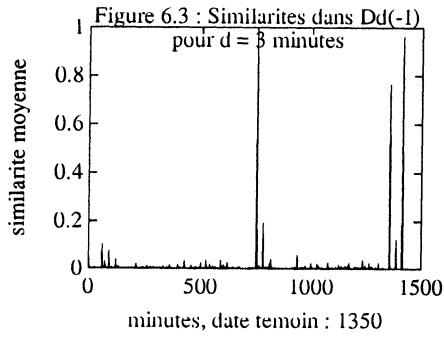
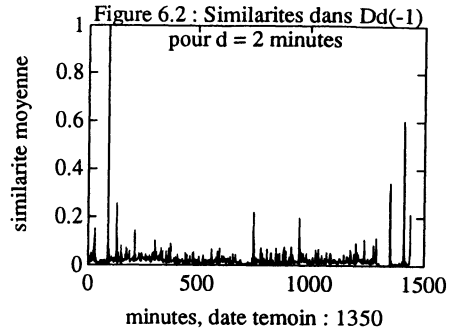
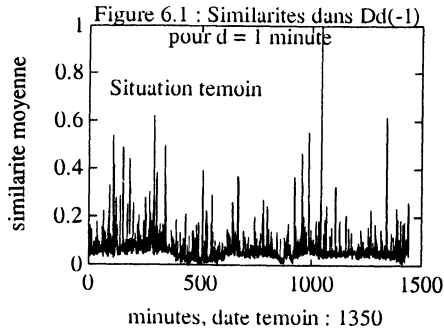
La reconnaissance de la forme des chauffe-eau pour ces signaux (avec donc un bloc témoin de longueur $r_d = 60/d$) avec $d = 1, \dots, 6$ minutes est illustrée par les figures 6.1 à 6.6. L'analyse témoin pour $d = 1$ est rappelée dans la figure 6.1. On constate une rupture par rapport à ce témoin pour $d = 4$ (figure 6.4) avec néanmoins, une persistance d'une similarité moyenne notable autour de midi. La rupture est en revanche très nette pour $d = 5$ (figure 6.5). L'application de la procédure fournit donc d_{\min} de l'ordre de 4 à 5 minutes.

Conclusion

La description statistique d'une courbe de charge électrique montre l'apport de la transformée en ondelettes. Elle fournit en effet, une décomposition multiéchelle du signal de consommation électrique en une somme de signaux orthogonaux correspondant à des échelles de temps différentes. Ceci permet d'apprécier l'apport d'information de chacune d'elles et de les rattacher à la dynamique propre du système de consommation électrique ou aux bruits affectant le signal.

La méthode utilisée pour mettre en évidence les informations contenues dans une courbe de charge électrique procède par l'utilisation de méthodes comparatives variées appliquées aux signaux obtenus par décomposition multiéchelle par ondelettes de la courbe de charge. La comparaison s'appuie systématiquement sur une analyse témoin issue d'informations externes sur le phénomène électrique sous-jacent. Les techniques usuelles pertinentes pour l'analyse des signaux stationnaires sont ici largement inopérantes. La courbe de charge échantillonnée à la minute présentant de nombreuses instationnarités très difficiles à supprimer, la méthode comparative la plus efficace parmi celles examinées, s'avère être la comparaison de formes des signaux issus de l'analyse multiéchelle.

Les principales conclusions électriques sont les suivantes. Pour l'analyse de la courbe, $D(-1)$ ne contient pas de signal, $D(-2)$ contient du signal mais seulement localement et faiblement, $D(-3)$, $D(-4)$ et $D(-5)$ contiennent du signal. En d'autres termes le signal est essentiellement contenu dans les composantes de l'enregistrement de période plus grandes que 4 minutes. L'utilisation d'une méthode semblable appliquée à la décomposition de l'enregistrement échantillonné à divers pas nous fait préconiser un pas d'échantillonnage de l'ordre de 4 à 5 minutes.



FIGURES 6.1 À 6.6

Bibliographie

- AZENCOTT R., DACUNHA-CASTELLE D. (1984), *Séries d'observations irrégulières*, Masson, 1984.
- AZENCOTT R., GIRARD Y., GIRARD B., ASTIER R., JAKUBOWICZ P., BAUDIN M., MARTIN M.M. (1988), *Mandrake : Logiciel expert d'analyse de séries chronologiques*, Int. Symposium on Forecasting, Amsterdam, june 1988.
- BOSQ D., LECOUTRE J.P. (1992), *Analyse et prévision des séries chronologiques. Méthodes paramétriques et non paramétriques*, Masson, 1992.
- BOX G., JENKINS G. (1992), *Time series analysis : forecasting and control*, Holden Day, 1976.
- COHEN A. (1990), *Ondelettes, analyses multirésolutions et traitement numérique du signal*, Thèse de doctorat, Université Paris IX Dauphine, 1990.
- COHEN A., FROMENT J., ISTAS J. (1991), *Analyse multi-résolution de signaux aléatoires*, C.R.A.S., 1991.
- DACUNHA-CASTELLE D., DUFLO M. (1982), *Probabilités et statistiques*, tome I, Masson, 1982.
- DAUBECHIES I. (1988), *Orthonormal basis of compactly supported wavelets*, Comm. Pure Appl. Math., vol. XLI, 909-996, 1988.
- DUFLO M. (1990), *Méthodes récursives aléatoires*, Masson, 1990.
- FROMENT J. (1990), *Traitement d'images et applications de la transformée en ondelettes*, Thèse de doctorat, Université Paris IX Dauphine, 1990.
- GASQUET C., WITOMSKI P. (1990), *Analyse de Fourier et Applications*, Masson, 1990.
- ISTAS J. (1992), *Statistique des processus gaussiens stationnaires continus par méthodes d'ondelettes*, Thèse de doctorat, Université Paris VII, fev. 92.
- KARANTA I., RUUSUNEN J. (1991), *Short term load forecasting in communal electric utilities*, Research report, A40, Helsinki University of Technology, may 1991.
- LEMARIÉ P.G. (1990), (ed.), *Les ondelettes en 1989*, Lect. Notes in Math., 1438, Springer Verlag, 1990.
- LEMARIÉ-RIEUSSET P.G. (1992), *Existence de «fonction-père» pour les ondelettes à support compact*, C.R.A.S. Paris, t. 314, série I, p. 17-19, 1992.
- MALLAT S. (1989), *A theory for multiresolution signal decomposition : the wavelet representation*, IEEE Trans. on PAMI, vol. 2, n°7, 674-693, 1989.
- MATLAB, Mathworks Inc., *Matlab reference manual*, 1990.
- MEYER Y. (1990), *Ondelettes et opérateurs*, Tome 1, Actualités mathématiques, Hermann, 1990.
- MISITI M., MISITI Y., OPPENHEIM G., POGGI J.M. (1993,1), *Micronde : un logiciel d'analyse et de synthèse par ondelettes, de signaux discrets*, Actes des XXV^e Journées de Statistique, Vannes, 106, mai 93.

- MISITI M., MISITI Y., OPPENHEIM G., POGGI J.M. (1993,2), *Analyse de signaux classiques par décomposition en ondelettes*, Revue de Statistique Appliquée, Vol. XLI, n°4, p. 5-32, 1993.
- MISITI M., MISITI Y., OPPENHEIM G., POGGI J.M. (1993,3), *Ondelettes en statistique et traitement du signal*, Revue de Statistique Appliquée, Vol. XLI, n°4, p. 33-43, 1993.
- STRANG G. (1989), *Wavelets and dilations equations : A brief introduction*, SIAM Review, vol. 31, 614-627, 1989.