HAL open science

Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité.

Anestis Antoniadis, Xavier Brosat, Jairo Cugliari, Jean-Michel Poggi

To cite this version:

Anestis Antoniadis, Xavier Brosat, Jairo Cugliari, Jean-Michel Poggi. Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité.. [Rapport de recherche] RR-7982, INRIA. 2012. hal-00703570

HAL Id: hal-00703570

https://inria.hal.science/hal-00703570

Submitted on 4 Jun 2012

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.

Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité.

Anestis Antoniadis, Xavier Brossat, Jairo Cugliari, Jean-Michel Poggi

RESEARCH REPORT
N° 7982
Juin 2012
Project-Team Select

ISSN 0249-6399 ISRN INRIA/RR--7982--FR+ENG

Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. Application à la consommation d'électricité.

Anestis Antoniadis*, Xavier Brossat†, Jairo Cugliari‡, Jean-Michel Poggi§

Équipe-Projet Select

Rapport de recherche n° 7982 — Juin 2012 — 28 pages

Résumé : Nous traitons le problème de la prévision d'un processus stochastique à valeurs fonctionnelles. Nous commençons par étudier le modèle proposé par Antoniadis et al. (2006) dans le cadre d'une application pratique -la demande d'énergie électrique en France- où l'hypothèse de stationnarité semble ne pas se vérifier. Le caractère non stationnaire est double: d'une part, le niveau moyen de la série change dans le temps, d'autre part il existe des groupes dans les données qui peuvent être vus comme des classes de stationnarité.
Nous explorons diverses variantes et corrections qui améliorent la performance de prédiction. Les corrections visent à prendre en compte la présence de ces caractéristiques non stationnaires. En particulier, pour prendre en compte l'existence de groupes, nous avons contraint le modèle de prévision à n'utiliser que les données qui appartiennent au même groupe que celui de la dernière observation disponible. Si le groupe est connu, un simple post-traitement suffit pour obtenir des meilleures performances de prédiction.

Mots-clés : Prévision non paramétrique, Données fonctionnelles, Non stationnaire, Consommation d'électricité

* Université Joseph Fourier, Laboratoire LJK, Tour IRMA, BP53, 38041 Grenoble Cedex 9, France, anestis.antoniadis@imag.fr

† EDF R&D, 1 avenue du Général de Gaulle, 92141 Clamart Cedex, France, xavier.brossat@edf.fr

‡ Inria Select, Univérsité Paris Sud, Bât. 425, 91405 Orsay Cedex, France Jairo.Cugliari@inria.fr

§ Université Paris Descartes, Université Paris Sud, Bât. 425, 91405 Orsay Cedex, France Jean-Michel.Poggi@paris-descartes.fr

RESEARCH CENTRE SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Forecasting non stationary function-valued process. Application to the electricity load demand.

Abstract: We study here the problem of predicting a functional valued stochastic process. We first explore the model proposed by Antoniadis et al. (2006) in the context of a practical application -the french electrical power demand- where the hypothesis of stationarity may fail. The departure from stationarity is twofold: an evolving mean level and the existence of groups that may be seen as classes of stationarity.

We explore some corrections that enhance the prediction performance. The corrections aim to take into account the presence of these nonstationary features. In particular, to handle the existence of groups, we constraint the model to use only the data that belongs to the same group of the last available data. If one knows the grouping, a simple post-treatment suffices to obtain better prediction performances.

Key-words: Nonparametric forecasting, Functional data, Non stationarity, Electricity load curve

1 Introduction.

Notre point de départ est une application industrielle en vraie grandeur. Le but est de prévoir la demande d'électricité de l'opérateur EDF (*Électricité de France*) d'un jour pour tout moment de la journée du lendemain. Sur le plan économique, des bonnes prévisions sont nécessaires pour réduire les coûts de production et profiter des opportunités du marché.

S'il est évident qu'à chaque instant de la journée, il existe un niveau de consommation électrique, les mesures dont on dispose ne sont prises qu'au pas demi-horaire et pas en temps continu. En général, seul cet échantillonnage est pris en compte dans les modèles de prévision. Néanmoins, nous pouvons nous poser la question de la pertinence de considérer la structure continue sous-jacente aux données. Nous nous plaçons dans le contexte de la prévision d'un processus stochastique à temps continu, $X = (X(t), t \in \mathbb{R})$. Étant donnée une trajectoire de X observée sur l'intervalle $[0, T]$ à une fréquence relativement haute, nous sommes intéressés par la prévision du comportement de X sur tout l'intervalle $(T, T + \delta]$, avec $\delta > 0$, plutôt qu'en certains instants. Pour ce faire, Bosq (1991) propose de diviser l'intervalle $[0, T]$ en sous-intervalles de la forme $[(l-1)\delta, l\delta]$, $l = 1, \ldots, n$, où $\delta = T/n$, et de considérer le processus stochastique $Z = (Z_i\ ; i \in \mathbb{N})$ où $\mathbb{N} = \{1, 2, \ldots, n\}$, défini de la façon suivante

$$Z_i(t) = X\{t + (i-1)\delta\} \qquad i \in \mathbb{N}, \qquad t \in [0, \delta).$$

Ce découpage est particulièrement utile quand le processus X présente un cycle saisonnier de période δ. Dans ce cas, cette procédure peut rendre raisonnable l'hypothèse de stationnarité du processus fonctionnel Z puisque nous ne faisons pas d'hypothèse de stationnarité sur le comportement "à l'intérieur" de chaque fonction aléatoire $Z_i, i \in \mathbb{N}$. Alors, le problème de prévision posé équivaut à prévoir Z_{n+1} étant données les observations fonctionnelles $Z_1, \ldots, Z_n$.

Les objets mathématiques à traiter imposent l'adaptation de méthodes de prévision classiques à des variables aléatoires de dimension infinie. Pendant les deux dernières décennies, l'utilisation des variables aléatoires à valeurs fonctionnelles a reçu un intérêt croissant dans la littérature. Ramsey et Dalzell (1991) ont forgé le terme *Analyse des Données Fonctionnelles* pour englober l'ensemble des méthodes et techniques traitant des variables aléatoires fonctionnelles. Citons comme référence les ouvrages de Ramsay et Silverman (1997, 2002) qui donnent une introduction détaillée des aspects tant théoriques que pratiques pour les cas des données indépendantes et identiquement distribuées. Le cas des données dépendantes est abordé dans Bosq (2000) par l'étude des processus linéaires et en particulier le processus Autoregressif Hilbertien (ARH).

Si Z est un processus centré ARH(1) d'ordre 1, un prédicteur de Z_{n+1} étant donné l'historique $Z_n, Z_{n-1}, \ldots, Z_1$ est

$$\widetilde{Z}_{n+1} = \mathbb{E}(Z_{n+1} | Z_n, Z_{n-1}, \ldots, Z_1) = \rho(Z_n) \tag{1}$$

où ρ est un opérateur linéaire borné associé au processus ARH(1). De cette manière, on peut prévoir Z_{n+1} si on peut avoir une manière efficace d'estimer

l'opérateur ρ. Une stratégie alternative consiste en estimer directement la fonction $\rho(Z_n)$ à partir de l'information contenue dans l'historique $Z_n, Z_{n-1}, \ldots, Z_1$.

Or, dans certaines applications de la vie réelle cette modélisation peut s'avérer trop restrictive. C'est le cas de la consommation journalière d'électricité où les transitions entre les successifs jours sont plus riches que celles décrites par le modèle ARH.

Antoniadis *et al.* (2006) proposent un modèle plus général pour le processus fonctionnel Z, dans lequel chaque élément Z_k dépend des valeurs retardées de la variable endogène au travers d'un opérateur non paramétrique m. Si l'on dispose d'une estimation de m obtenue à partir des observations $Z_1, \ldots, Z_n$, alors la prévision de Z_{n+1} est donné par l'évaluation de l'estimation de m dans la dernière valeur observée Z_n. L'estimation de m peut s'obtenir en utilisant des méthodes non paramétriques plus ou moins classiques, comme par example l'estimateur à noyau de Nadaraya-Watson. Si c'est le cas, le prédicteur a une interprétation similaire à celle du modèle utilisé dans Poggi (1994). Pour éviter le problème de la dimension infinie de chaque segment, ils utilisent la *transformée discrète en ondelettes* (DWT pour Discrete Wavelet Transform) pour représenter les trajectoires sur les segments temporels . Sous des hypothèses assez générales, les auteurs prouvent le bien fondé théoriques de leur méthode de prévision avec quelques expériences numériques préliminaires.

Dans ce papier nous montrons que, même en adaptant soigneusement la méthode aux particularités les plus importantes des données, la méthode ne peut pas être utilisée telle quelle sur l'application industrielle que nous intéresse. Nous trouvons l'origine du problème dans l'hypothèse de stationnarité du processus fonctionnel qui n'est pas vérifiée sur les données de consommation d'électricité française. D'une part, le niveau moyen de la série change dans le temps présentant en hiver sa plus grande variabilité, ce qui est d'autant plus grave que c'est sur cette période que les erreurs coûtent le plus cher car le prix de production est plus élevé. D'autre part ils existent des groupes des jours qui ont une structure de consommation très différente d'autres groupes de jours. Nous explorons dans l'article diverses variantes et corrections à la méthode d'origine qui permettent de fournir de prévisions fonctionnelles de grande qualité même en présence de ce type de non stationnarités.

La Section 2 de cet article lie le modèle de prévision basé dans la régression à noyau d'un processus unidimensionnel avec celle d'un processus fonctionnel au travers de la transformée en ondelettes. De plus, nous mettons en évidence le rôle de chacun des paramètres du prédicteur. La Section 3 est dédié à des aspects industriels. Nous y présentons la problématique des données de la consommation d'électricité au travers d'une description des principales caractéristiques des données et d'une courte bibliographie des méthodes de prévision alternatives. Ensuite, dans la Section 4 on applique, après les premières adaptations nécessaires, la méthode d'origine pour prévoir la consommation d'électricité à tout moment de la journée du jour pour le lendemain. Elle nous servira de référence à améliorer et l'analyse de ses performances fondera le choix des modifications à apporter aux modèles. Ainsi, dans la Section 5, des extensions et des corrections sont proposées de sorte que les données d'électricités soient raisonnablement

compatibles avec les hypothèses du modèle prédictif. Une analyse détaillée et motivée des gains apportés par chacune des corrections est menée.

2 Prévision par régression à noyau: du cas univarié au cas fonctionnel.

Dans cette section nous rappelons la construction du prédicteur proposé dans Antoniadis *et al.* (2006). Tout d'abord, nous rappelons le cas d'un processus univarié. Ensuite, nous rappelons les éléments de la transformée discrète en ondelettes, nécessaires pour la représentation des trajectoires d'un processus fonctionnel. Puis, le prédicteur pour un processus fonctionnel est introduit. Enfin, nous discutons du réglage des paramètres du prédicteur (pour une discussion plus détaillée, voir Cugliari (2011)).

2.1 Le cas des processus univariés.

Considérons un processus stochastique stationnaire à valeurs réelles à temps discret $Y = (Y_i, i \in \mathbb{N})$. Si l'on suppose que le processus vérifie la propriété de Markov, on peut écrire pour un entier d

$$\mathbb{E}[Y_n|Y_{n-1}, \ldots, Y_0] = \mathbb{E}[Y_n|Y_{n-1}, \ldots, Y_{n-d}].$$

Nous allons noter $\mathcal{Y}_{n,d} = (Y_n, Y_{n-1}, \ldots, Y_{n-d+1})$. Nous supposons le modèle suivant pour le processus

$$Y_{n+1} = m(\mathcal{Y}_{n,d}) + \epsilon_n,$$

où $m(y) = \mathbb{E}(Y_{n+1}|\mathcal{Y}_{n,d} = y)$ avec $y \in \mathbb{R}^d$ et $(\epsilon_n, n \in \mathbb{N})$ est un processus d-dimensionnel de bruit blanc (i.e. une suite de variables aléatoires réelles avec $\mathbb{E}\epsilon_n = 0$ pour tout $n \in \mathbb{N}$ et $\mathbb{E}(\epsilon_i\epsilon_j) = \sigma^2$ si $i = j$ et $\mathbb{E}(\epsilon_i\epsilon_j) = 0$ sinon) indépendant de $\mathcal{Y}_{n,d}$.

Nous sommes intéressés par le comportement futur de Y à l'horizon s, $Y_{n+s}, s > 0$, étant donné les observations $Y_1, \ldots, Y_n$, pour $n > d$. Nous appelons $\widetilde{Y}_{n+s}$ le prédicteur de Y_{n+s} construit à partir de ces observations. Un critère pour l'obtenir est de minimiser l'erreur quadratique moyenne, c'est-à-dire prendre $\widetilde{Y}_{n+s} = y_0$, où y_0 est l'argument qui minimise $w \mapsto \mathbb{E}[(w - Y_{n+s})^2|\mathcal{Y}_{n,d}]$. La solution s'avère être $y_0 = m(\mathcal{Y}_{n,d})$. Bien sûr ceci n'est pas un prédicteur statistique car il dépend de la loi inconnue du processus sous-jacent.

Comme m est une espérance conditionnelle (au moins pour tout y dans le support de la densité f du vecteur $\mathcal{Y}_{n,d}$), il est habituel de l'estimer par une technique non paramétrique car la régression entre Y_s et $\mathcal{Y}_{0,d}$ peut être non linéaire et assez complexe. Nous allons utiliser un estimateur de Nadaraya-Watson appelé estimateur à noyau (voir par example Härdle (1990)).

L'estimateur à noyau $\hat{m}_n$ de m peut s'écrire

$$\hat{m}_n(y) = \sum_{i=d}^{n-s} w_{n,i}(y) Y_{t+s},$$

où $w_{n,i}(y) = \frac{K_{h_n}(\mathcal{Y}_{i,d}-y)}{\sum_{t=d}^{n-s} K_{h_n}(\mathcal{Y}_{i,d}-y)}$ avec $K_{h_n}(.) = K(./h_n)$, $h_n > 0$ est le noyau K : $\mathbb{R}^d \mapsto \mathbb{R}$ est, pour simplifier, une fonction de densité de probabilité multivariée symétrique autour de l'origine. Le prédicteur à noyau s'en déduit

$$\widehat{Y_{n+s}} = \hat{m}_n(\mathcal{Y}_{n,d}).$$

Les $\{w_{n,i}, i = 1, \ldots, n-s\}$ sont des quantités positives dont la somme fait 1. Ainsi, la prévision à l'horizon s du processus Y étant donné l'historique s'écrit

$$\widehat{Y_{n+s}} = \sum_{t=d}^{n-s} w_{n,i}(\mathcal{Y}_{n,d}) Y_{i+s},$$

et peut être vue comme un barycentre des futurs d'horizon s des blocs du passé pondérés par la similarité entre le bloc présent et ses homologues dans le passé. Ainsi, le poids de chaque bloc $\mathcal{Y}_{i,d}$ est croissant avec sa similarité par rapport au bloc actuel $\mathcal{Y}_{n,d}$, l'idée étant que des passés semblables ont des futurs similaires.

Pour un échantillon de taille n donnée, la quantité h_n appelée *largeur de fenêtre* est cruciale pour la performance de la méthode (Härdle (1990)). Elle est un compromis entre le biais et la variance de l'estimateur. Une valeur trop faible conduira à un estimateur très local et en conséquence avec une variabilité importante. Si l'on considère des valeurs de plus en plus grandes (on dit alors qu'on élargit la fenêtre), on permet à de plus en plus d'observations de contribuer à l'estimation ce qui diminue la variance mais peut conduire à un estimateur plus biaisé. Le réglage de ce paramètre se fait habituellement par validation croisée.

2.2 Transformée en ondelettes.

La *Transformée en Ondelettes* est une technique de décomposition hiérarchique des signaux d'énergie finie qui permet de représenter un signal dans le domaine temps-échelle, où l'échelle joue un rôle analogue à celui de la fréquence dans l'analyse de Fourier (Mallat (1999)). Elle permet de décrire une fonction à valeurs réelles au travers de deux objets: une approximation de cette fonction et un ensemble de détails. La partie approximation résume la tendance globale de la fonction, alors que les changements localisés (en temps et fréquence) sont capturés dans les composantes de détails à différentes résolutions.

L'analyse des signaux est réalisée par des fonctions analysantes appelées *ondelettes* obtenues à partir de transformations simples d'une *ondelette mère*. Une ondelette est une fonction assez régulière, oscillante dans le domaine temporel avec une rapide décroissance vers zero et elle est localisée dans le domaine fréquentiel et le domaine temporel.

Dans la suite, notre attention sera portée par des fonctions définies sur un intervalle compact qui, sans perte de généralité, sera $[0, 1]$.

Nous définissons des versions périodiques de l'ondelette mère ψ^P et de la fonction d'échelle ϕ^P (associé à une analyse multirésolution de $L_2([0, 1])$),

$$\phi(t) = \sum_{l \in \mathbb{Z}} \phi^P(t-l) \qquad \text{et} \qquad \psi(t) = \sum_{l \in \mathbb{Z}} \psi^P(t-l), \qquad t \in [0, 1],$$

et des versions dilatées et translatées de ϕ et ψ définies par

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \qquad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k).$$

Enfin, pour tout $j_0 \in \mathbb{Z}$, la collection suivante fournit une base orthonormale de l'espace $L_2([0,1])$,

$$\{\phi_{j_0,k}, k = 0, 1, \ldots, 2^{j_0} - 1; \psi_{j,k}, j \geq j_0, k = 0, 1, \ldots, 2^j - 1\}.$$

En conséquence, tout $z \in L_2([0,1])$ admet un développement en ondelettes de la forme

$$z = \sum_{k\in\mathbb{Z}} c_{j_0,k}\phi_{j_0,k} + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}, \tag{2}$$

où les coefficients sont définis par

$$c_{j,k} =< z, \phi_{j,k} > \text{ et } d_{j,k} =< z, \varphi_{j,k} >,$$

et on les appelle respectivement *coefficients d'approximation (ou d'échelle)* et *coefficients d'ondelettes (ou de détails)* associés à l'échelle j et la position k.

Dans l'expression (2) le premier et le second termes à droite de l'égalité sont, respectivement, l'approximation au niveau de résolution j_0 et l'erreur d'approximation composée de l'agrégation des détails des niveaux d'échelles $j \geq j_0$. Ces deux composantes, l'approximation et les détails, peuvent être vues comme une partie d'approximation lisse non stationnaire qui contient les basses fréquences, et une composante qui garde l'information de détails localisés dans le temps pour les petites échelles. Enfin, le paramètre j_0 détermine la séparation entre ces composantes.

En outre, on a la conservation de l'énergie des signaux de carré intégrable pour la transformée en ondelettes orthogonale. Ainsi, l'équation (2) permet d'obtenir la fonction originale à l'aide des coefficients d'ondelettes et d'échelle (par la transformée inverse) sans perte d'information.

2.3 Le cas des processus fonctionnels.

Nous allons procéder par analogie au cas univarié pour présenter le modèle de prévision proposé par Antoniadis *et al.* (2006) que nous appelons KWF (pour *Kernel Wavelet Functional*). Maintenant nous considérons un processus stochastique supposé dans ce cadre, stationnaire $Z = (Z_i)_{i\in\mathbb{Z}}$ à valeurs dans un espace fonctionnel H (par exemple $H = L_2([0,1])$). Nous disposons d'un échantillon de n courbes $Z_1, \ldots, Z_n$ et l'objectif est de prévoir Z_{n+1}. La méthode de prévision présentée peut se décomposer en deux phases. D'abord, trouver parmi les blocs du passé ceux qui sont le plus semblables au dernier bloc observé. Ensuite construire un vecteur de poids $w_{n,i}, i = 1, \ldots, n-1$ pour obtenir la prévision souhaitée en moyennisant les futurs des blocs correspondant aux indices $2, \ldots, n$ respectivement.

Première phase. La méthode de prévision nécessite une distance appropriée entre les objets observés qui sont des courbes plus ou moins régulières. Pour prendre en compte la dimension infinie de ces objets, le modèle KWF représente chaque segment $Z_i, i = 1, \ldots, n$, par son développement sur une base d'ondelettes tronqué à une échelle $J > j_0$. Ainsi, chaque observation Z_i est décrite par sa version tronquée

$$Z_{i,J}(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}^{(i)} \phi_{j_0,k}(t) + \sum_{j=j_0+1}^{J} \sum_{k=0}^{2^j-1} d_{j,k}^{(i)} \psi_{j,k}(t), \qquad t \in [0,1].$$

Le premier terme de l'équation est une approximation lisse à la résolution j_0 du comportement global de la trajectoire. Elle contient les composantes non stationnaires associées à des basses fréquences ou encore à une tendance. Le deuxième terme conserve l'information de la structure locale de la fonction. Cette information, décrite en fonction de l'échelle j et la position temporelle k dans l'échelle, sera exploitée pour construire une distance.

Pour deux segments observés $Z_i(t)$ et $Z_{i'}(t)$, nous utilisons la distance euclidienne entre les vecteurs des coefficients d'ondelettes à chaque échelle j avec $j_0 < j \leq J$,

$$\mathrm{dist}_j(Z_i, Z_{i'}) = \left(\sum_{k=0}^{2^j-1} (d_{j,k}^{(i)} - d_{j,k}^{(i')})^2 \right)^{1/2},$$

puis on agrège les distance des échelles en prenant en compte le nombre de coefficients par échelle

$$D(Z_i, Z_{i'}) = \sum_{j=j_0+1}^{J} 2^{-j/2} \mathrm{dist}_j(Z_i, Z_{i'}).$$

Les coefficients d'approximation ne contiennent pas d'information utile pour la prévision (ils fournissent des moyennes locales), car le processus Z est supposé ici stationnaire. De ce fait, ils ne sont pas pris en compte dans la distance proposée. En d'autres termes, la distance D permet de trouver de motifs similaires entre courbes même si elles ont des approximations très différentes. De plus, la capacité de la DWT de détecter des caractéristiques locales d'une fonction motive l'utilisation de cette distance, car les structures locales ne s'expriment qu'au travers des détails.

Deuxième phase. Notons $\Xi_i = \{c_{J,k}^{(i)} : k = 0, 1, \ldots, 2^J - 1\}$ l'ensemble des coefficients d'échelle du i-ème segment Z_i à la résolution J, la plus fine. La prévision des coefficients d'échelle (à l'échelle J) $\widehat{\Xi_{n+1}}$ de Z_{n+1} est donnée par

$$\widehat{\Xi_{n+1}} = \frac{\sum_{m=1}^{n-1} K_{h_n}(D(Z_{n,J}, Z_{m,J}))\Xi_{m+1}}{1/n + \sum_{m=1}^{n-1} K_{h_n}(D(Z_{n,J}, Z_{m,J}))}.$$

Notons que la distance D est calculée sur les trajectoires approchées. De ce fait, nous pouvons coder dans un premier temps les courbes Z_i par Ξ à l'aide

de la DWT. Puis, en utilisant l'algorithme pyramidal sur Ξ, nous obtenons l'ensemble des coefficients d'ondelettes à utiliser dans la distance. La fonction noyau utilisée ici a les mêmes propriétés que celle du cas univarié, à la différence que maintenant K est définie sur $\mathbb{R}$.

Finalement, nous appliquons la transformée inverse de la DWT sur $\widehat{\Xi_{n+1}}$ pour obtenir la prévision de la courbe Z_{n+1} dans le domaine temporel

$$\widehat{Z_{n+1}(t)} = \sum_{k=0}^{2^J-1} \widehat{c_{J,k}^{(n+1)}} \phi_{J,k}(t). \tag{3}$$

Nous pouvons réécrire le prédicteur (3) comme un barycentre des futurs des segments du passé

$$Z_{n+1}(t) = \sum_{m=1}^{n-1} w_{n,m} Z_{m+1}, \tag{4}$$

avec les poids

$$w_{n,m} = \frac{K_{h_n}(D(Z_{n,J}, Z_{m,J}))}{\sum_{m=1}^{n-1} K_{h_n}(D(Z_{n,J}, Z_{m,J}))}. \tag{5}$$

2.4 Les paramètres de réglage du prédicteur.

Afin d'utiliser la méthode de prévision KWF, des paramètres doivent être choisis, ils concernant,

- le découpage en blocs pour définir le processus Z,
- la transformation dans le domaine des ondelettes,
- l'estimateur à noyau.

2.4.1 Paramètres liés au découpage en blocs.

Taille des blocs: paramètre δ. Pour obtenir le processus à temps discret $Z = (Z_i(t), t \in [0, \delta])_{i \in \mathbb{N}}$, nous avons procédé au découpage d'une trajectoire d'un processus stochastique continu à valeurs réelles X, en blocs de taille δ. Même s'il n'y a pas d'éléments théoriques qui puissent guider le choix de δ, nous pouvons néanmoins le choisir de manière à modéliser une saisonnalité. Ainsi, par exemple pour les données de la consommation d'électricité qui présentent une périodicité journalière, le fait de découper en courbes journalières rend plus raisonnable l'hypothèse de stationnarité du processus fonctionnel Z.

2.4.2 Paramètres liés à la DWT.

L'ondelette. La transformée en ondelettes joue plusieurs rôles dans la méthode de prévision KWF. Primo, elle est utilisée pour représenter les données fonctionnelles (de dimension infinie) par des vecteurs (de dimension finie) des coefficients d'échelle à une résolution J. Secondo, comme il est habituel dans

l'analyse des données fonctionnelles, nous ne disposons que d'un échantillonnage des fonctions. Dans le cas de la consommation d'électricité les observations sont faites au pas demi-horaire. La DWT est utilisée pour approcher les trajectoires plus ou moins régulières de la consommation journalière d'électricité à partir de 48 mesures. Tertio, la prévision est obtenue dans le domaine des ondelettes, et c'est grâce à la transformée inverse (et aux bonnes propriétés d'approximation) que l'on obtient la prévision dans le domaine temporel. Nous allons voir par la suite que la transformée en ondelettes aura encore un quatrième rôle quand on voudra adapter la technique à un contexte non stationnaire.

Dans les expériences numériques nous utilisons l'ondelette appelée *Symmlet 6*, avec un filtre de taille 6, qui parait un bon compromis compte tenu de la taille de nos segments. Nous avons observé d'autre part que la sensibilité du prédicteur au choix de l'ondelette est faible (voir Cugliari (2011, p. 55)).

Interpolation à 2^J points. Afin d'utiliser l'algorithme pyramidal de Mallat pour le calcul de la transformée en ondelettes (voir Mallat (1999)), nous devons disposer de vecteurs de 2^J points avec J un entier non nul. Si ce n'est pas le cas, nous utilisons une interpolation par splines naturelles pour le plus proche entier J qui vérifie $2^{J-1} < N < 2^J$.

Résolution de l'approximation. Pour une ondelette et une taille de filtre données, nous approchons les courbes à la résolution j_0 qui doit être choisie. Toutes les résolutions $j < j_0$ ne feront pas partie de la prévision. C'est justement la capacité des ondelettes de faire une analyse à plusieurs résolutions qui nous intéresse. L'échelle j_0 fait la séparation entre les échelles de l'analyse multirésolution qui sont liées aux basses fréquences, associées à des composantes non stationnaires comme la tendance, et celles liées aux hautes fréquences, associées à la partie stationnaire du processus. Il y a un compromis à faire: d'un côté une fréquence trop basse pourrait inclure dans la comparaison des phénomènes non stationnaires; d'une autre côté des fréquences trop hautes pourraient contenir trop de bruit et assez peu de signal utile. Pour commencer nous allons utiliser $j_0 = 0$, c'est à dire l'approximation au niveau de résolution le plus grossier.

2.4.3 Paramètres liés à l'estimateur à noyau.

Distances entre des courbes échantillonnées. Du fait que les courbes ne sont observées qu'au travers d'un échantillonnage, la distance D dépend de l'échantillonnage. Néanmoins, son utilisation en remplaçant la vraie distance théorique peut être justifié (voir Cugliari (2011, p. 35)).

Le noyau et la largeur de fenêtre. Comme en général les méthodes à noyau sont assez robustes par rapport au choix du noyau, le choix est encore une fois guidé par des aspects pratiques (support et régularité du noyau). En revanche, le paramètre h_n joue un rôle crucial pour calibrer la méthode à noyau car son réglage contrôle le compromis entre le biais et la variance du prédicteur.

Dans Antoniadis *et al.* (2006), les auteurs proposent de calculer h_n par *validation croisée* mais elle ne repose que sur des bases empiriques. Dans un second article, les auteurs proposent un calcul fondé sur des éléments théoriques (Antoniadis *et al.* (2009)) qui se fait par minimisation d'une fonction de *risque empirique* sur un échantillon d'apprentissage sur une grille de valeurs possibles. Si la grille contient la vraie valeur du paramètre, alors cette stratégie ne peut pas conduire à un choix très éloigné de l'optimal.

Quand le modèle de prévision est utilisé tout au long d'une période d'étude (par exemple si l'on fait de prévisions journalières pendant toute une année), on peut alors définir deux stratégies apparaissant naturellement pour calculer h_n. La première option est de calculer et fixer la valeur h_n avant la période de prévision (FIX). L'option alternative consiste en mettre à jour cette valeur avant la prévision à chaque date (DYN).

3 Présentation de la problématique de la consommation d'électricité.

Cette section est dédiée à la présentation du problème industriel. Nous commençons par décrire les données au travers de leurs caractéristiques les plus remarquables. Ensuite, nous menons une courte revue des méthodes de prévision utilisées pour des données de la consommation d'électricité française métropolitaine. Enfin, nous discutons sur les performances de prévision attendues.

3.1 Les données.

Nous commençons par énumérer les caractéristiques de la consommation d'électricité à l'aide des données françaises. Dans le Figure 1 nous pouvons observer l'évolution sur le long terme de la demande nationale d'électricité. Malgré la crise économique des dernières années, nous remarquons une tendance ascendante. Le cycle annuel est aussi clairement marqué, présentant les plus importants niveaux de consommation d'électricité pendant l'hiver. Deux faits l'expliquent, d'une part la forte dépendance de la consommation d'électricité aux conditions météorologiques; d'autre part l'activité industrielle qui a un caractère saisonnier.

Lorsque nous faisons un zoom, nous pouvons distinguer une périodicité hebdomadaire (Figure 2). Le profil économique des jours ouvrés et des week-ends est reproduit par la demande avec une forte hausse durant les jours ouvrés. Il existe d'autres artefacts de l'activité socio-économique. À titre d'exemple, durant la période estivale, nous observons deux semaines durant lesquelles la demande en électricité est extrêmement basse, correspondant aux vacances d'été. Il est à noter aussi que le profil de la demande d'électricité en hiver est plus complexe à cause d'une grande variation de la demande. L'impact sur la prévision est une nette dégradation de la performance de prévision en hiver. Malheureusement, cela arrive durant la période où les erreurs de prédiction ont un coût plus élevé pour les fournisseurs d'électricité.

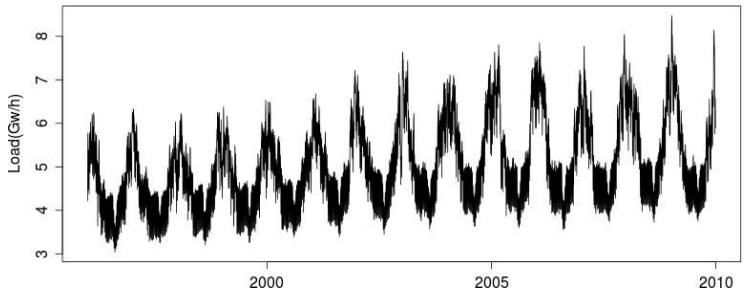

Figure 1: Évolution de la consommation d'électricité moyenne journalière (en Gwh) de 1996 à 2010.

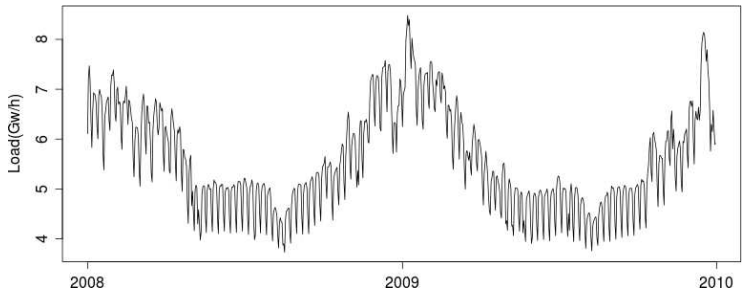

Figure 2: Évolution de la consommation d'électricité moyenne journalière (en Gwh) de 2008 à 2010.

La Figure 3 représente une courbe journalière. Notons que même à cette résolution temporelle, nous pouvons identifier des motifs: la consommation d'électricité est plus faible la nuit, elle augmente entre 5 heures et 9 heures du matin, et présente un pic en fin d'après-midi, etc. Ces caractéristiques sont identifiables sur chaque courbe journalière. En effet, les jours pour lesquels il est difficile d'avoir une bonne prédiction de la consommation/demande d'électricité sont ceux avec des caractéristiques atypiques ou assez rares. De manière générale, ces jours sont aussi parmi ceux qui coûtent le plus en termes d'erreurs de prédiction.

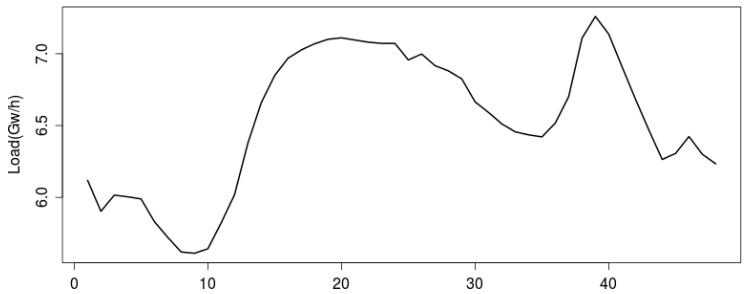

Figure 3: Une courbe de charge journalière.

Nous décrivons à présent la partie de la consommation électrique qui dépend des conditions climatiques. Même si, nous nous centrerons uniquement sur la dépendance vis-à-vis de la température, cependant, d'autres phénomènes météorologiques -comme la couverture nuageuse- ont aussi une incidence sur la demande d'électricité. Dans le cas français, cette demande est connue pour être hautement thermosensible (Figure 4). Deux faits sont à remarquer. D'une part, la température utilisée dans le graphique est construite en faisant une moyenne des températures records de quelques stations météorologiques françaises convenablement choisies. Rappelons que la sensibilité de la demande d'électricité à la température s'exprime notamment dans le chauffage des bâtiments. Ainsi, un obstacle additionnel s'y rajoute, la température à utiliser doit incorporer l'inertie thermique des bâtiments. D'autre part, la dépendance est très complexe et a certainement un comportement non linéaire et asymétrique par rapport aux hautes et basses températures respectivement.

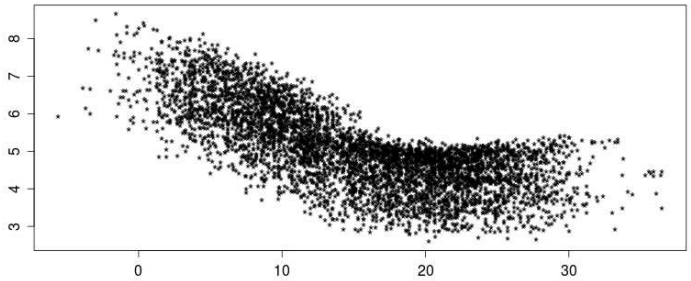

Figure 4: Relation entre demande d'électricité (en Gwh) et température (en ° Celsius).

Une quantification très précise de la part des experts d'EDF montre qu'entre deux seuils de température (situés à 14 et de 23°Celsius), la demande d'électricité est insensible aux changements de température. En ce qui concerne les températures inférieures à 14°C ou supérieures à 23°C, les systèmes de chauffage et climatisation respectivement se mettent en marche.

3.2 Une courte revue de méthodes alternatives.

Dans la littérature sur la prévision de la consommation d'électricité, les méthodes peuvent être groupés par type d'approche: d'analyse de séries temporelles, d'apprentissage automatique, de régression et de recherche de similarité. Pour fixer les idées, nous nous bornons à analyser les références décrivant des travaux qui portent sur des données de consommation d'électricité française sur l'aire métropolitaine.

Dans le groupe d'*analyse de séries temporelles*, Taylor (2010) utilise un lissage exponentiel en prenant en compte la structure de saisonnalités. Dordonnat

et al. (2008, 2011) propose un modèle à espace d'états qui permet suivre les changements de la relation entre les facteurs exogènes (principalement la température) et la demande d'électricité.

Les modèles basés dans l'*analyse de la régression* permettent d'utiliser la connaissance à priori pour modéliser la structure interne de la consommation et la dépendance aux facteurs exogènes. Dans Bruhns *et al.* (2005) cette dépendance est modélisée par une régression non linaire sur la température qui dépend du mois, du jour de la semaine et du moment de la journée. Une version non paramétrique de cette stratégie a été proposée récemment dans Pierrot et Goude (2011).

Parmi les méthodes d'*apprentissage automatique*, Devaine *et al.* (2011) utilisent un mélange de prédicteurs en ligne pour obtenir des prévisions s'adaptant aux non stationnarités.

Le dernier groupe de modèles, basées sur la *recherche de similarités*, est une alternative à la modélisation de la structure de dépendance des cycles saisonnières. Le postulat de base est de dire que de causes similaires dans le passé ont des conséquences futures semblables. Par exemple dans Poggi (1994) la trajectoire de la consommation d'électricité est divisée en blocs de taille journalière. Puis, à l'aide de mesures de dissimilarité introduites par l'auteur, des blocs similaires au dernier observé sont cherchés dans le passé et un vecteur de poids est construit. Enfin, la prévision des deux prochains jours est obtenue par une moyenne pondérée des futurs des jours le plus similaires où les pondérations sont données par le vecteur de poids. Du point de vue statistique, le modèle correspond à une estimation de la régression par la méthode du noyau, du dernier bloc contre tous les blocs du passé. Antoniadis *et al.* (2006) étendent ce modèle au cas de variables aléatoires fonctionnelles.

3.3 Quelle performance de prévision est attendue?

Ces caractéristiques expliquent la part plus importante de la variabilité de la consommation d'un jour typique. Tout modèle de prévision qui prend en compte ces éléments aura une performance moyenne acceptable. La difficulté réelle se trouve dans la prévision de jours de consommation atypique comme les jours fériés, les jours de pont associés à un long week-end, les jours de tarification spéciale, etc. D'autres phénomènes externes peuvent modifier la distribution ou la demande d'électricité, e.g. une crise économique ou des dommages produits par des orages. Les modèles opérationnels à EDF sont capables de gérer la prévision de jours typiques et la plupart de jours atypiques. Parmi l'essentiel, il s'agit de modèles paramétriques incorporant une régression non linéaire pour quantifier la sensibilité de la consommation à la température ressentie.

Nous sommes intéressés par des prévisions journalières du jour pour le lendemain. La performance de prévision est calculée sur les 48 mesures journalières par la moyenne des erreurs relatives absolues (MAPE) qui est définie comme suit:

$$\text{MAPE} = \frac{1}{48}\sum_{k=1}^{48}\left|\frac{Z(t_k) - \widehat{Z(t_k)}}{Z(t_k)}\right|.$$

Les modèles opérationnels d'EDF donnent un niveau de MAPE moyen sur une année à horizon journalier d'environ 1.5%.

4 Prévision de la consommation d'électricité par la méthode KWF.

Les expériences ont été réalisées à l'aide du logiciel `R`. Nous avons développé le package `kerwavfun` qui permet d'obtenir la prévision par la méthode de base, ainsi que pour toutes les variantes présentées par la suite. La DWT est calculée en utilisant le package `wavethresh` (Nason (2010)), qui fournit une implémentation de l'algorithme pyramidal de Mallat.

Pour les expériences numériques, nous avons utilisé un signal contenant des mesures en temps réel (postérieurement consolidées) de la consommation électrique. Les données sont échantillonnées toutes les 30 minutes depuis le 1 septembre 1996 jusqu'au 31 août 2006. Nous mesurons la performance en prévision au pas journalier du 1 septembre 2005 jusqu'au 31 août 2006.

Avec ces données, nous obtenons des blocs de taille $\delta = 48$ points représentant les courbes de charge journalière entre minuit et 23h30. Chaque vecteur de longueur 48 est interpolé à 2^6 points et transformé à l'aide de la DWT pour un niveau de résolution de l'approximation de $j_0 = 0$ en utilisant l'ondelette *Symmlet 6*. Quant au calcul de la largeur de fenêtre h_n, nous avons testé deux options introduites précédemment: calculer un seul h_n une seul fois (FIX) ou la mettre à jour avant chaque prévision (DYN). Le noyau utilisé est le noyau gaussien.

Les résultats des performances de prévision sont présentés dans le Tableau 1 en termes de MAPE (en %) pour les variantes FIX et DYN concernant le calcul de la fenêtre. Nous sommes intéressés par la qualité de la prévision globale mais aussi par la qualité par type de jour et la distribution des erreurs de prévision tout au long de l'année. Ce tableau sera le point de départ pour la comparaison de la méthode KWF de base original (que nous appellerons BASE) et les variantes que nous utiliserons.

De façon générale, les niveaux d'erreur global de la technique sont plus importants que les niveaux observés pendant la période estivale (environ quatre fois plus grands). Avant d'essayer d'expliquer pourquoi, nous nous centrons sur les différences entre les variantes pour le calcul de h_n. Nous pouvons noter que la méthode FIX a une performance supérieure pour chaque type de jour (sauf les samedi) que la méthode DYN pour la moyenne annuelle. L'évolution de la fenêtre mobile calculée par la méthode DYN est affichée dans la Figure 5. Nous pouvons noter une forte et rapide augmentation du paramètre à partir du mois de novembre, après un saut à fin octobre dû au passage à l'heure d'hiver. La fenêtre s'ouvre pour calibrer l'effet des jours fériés et les premiers jours froids qui ont tous les deux une dynamique de consommation très différente des autres types de jours. Cette introduction provoque des fortes perturbations qui altèrent de manière importante la performance de la méthode. Nous continuerons à ex-

Bloc	MAPE	MAPE
lundi	7.1	7.48
mardi	8.11	8.36
mercredi	7.63	7.94
jeudi	8.03	8.28
vendredi	8.72	8.87
samedi	7.62	7.58
dimanche	8.51	8.66
férié	12.9	12.85
Global	8.11	8.31

Table 1: MAPE par type de jour et global pour la période 1/9/05 - 31/08/06. Le paramètre h_n à été réglé selon la méthode FIX (à gauche) avec $h_n = 2155$, et la méthode DYN où h_n est calculée jour après jour (à droite).

aminer cette dichotomie (fenêtre fixe - fenêtre mobile) dans la suite.

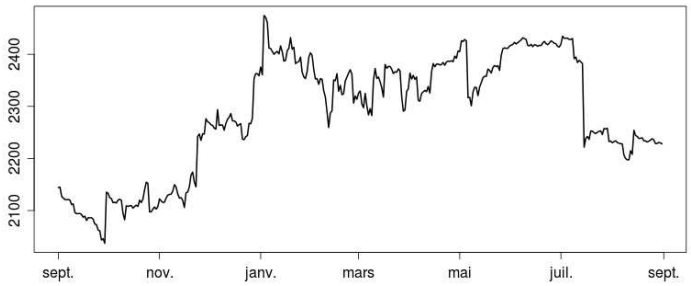

Figure 5: Évolution du paramètre h_n.

En ce qui concerne les hauts niveaux d'erreur de prévision, rappelons que nous n'avons introduit aucune information exogène: la prévision est faite par régression sur les blocs observés du passé. En particulier, nous n'avons pas d'informations sur l'effet sur la consommation d'électricité provoquées par la structure calendaire ou l'effet des variables météorologiques (e.g. la température ou la nébulosité). Toutefois il est souhaitable d'avoir des niveaux d'erreurs globalement faibles et d'analyser sa distribution dans l'année. Pour approfondir, nous analysons l'évolution du MAPE. Deux grands problèmes qui concernent la stationnarité du processus Z se lisent sur la Figure 6:

- la méthode conduit à la pire des performances quand le niveau est plus difficile à prévoir (pendant l'hiver), et
- quand le niveau moyen est assez stable (voir entre juin et octobre), les

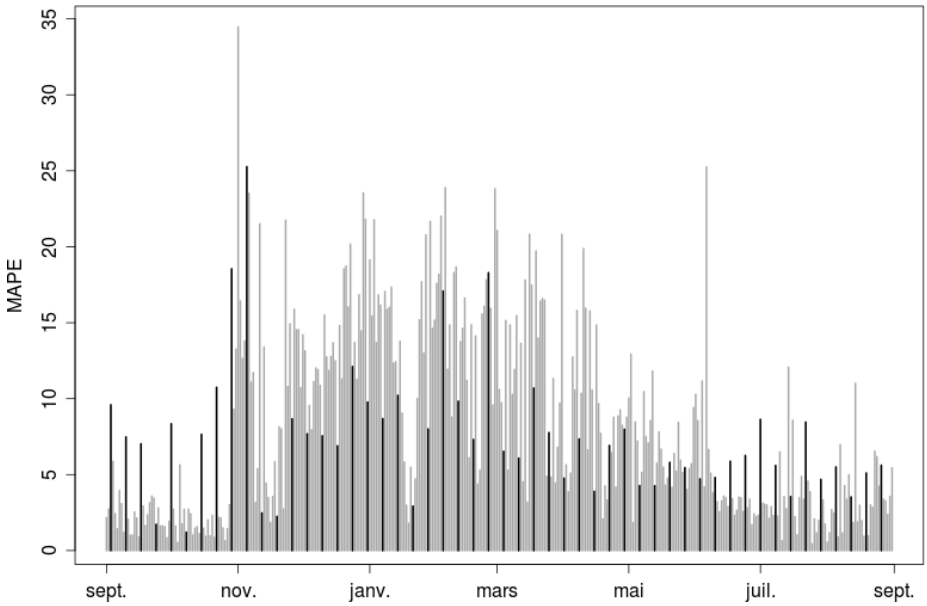

Figure 6: Évolution du MAPE journalier pour la méthode de BASE. Les barres en gras correspondent aux samedis.

barres en gras qui représentent les samedis sont sensiblement plus hautes.

Analysons ces deux points. Pour le premier, il est assez clair que l'hypothèse de stationnarité du processus n'est pas valide pour les données de consommation d'électricité alors que la méthode de base le suppose. C'est justement lorsque le niveau moyen varie de manière plus importante que nous observons les erreurs les plus grandes. Au sujet du deuxième problème, la méthode ne peut pas faire la distinction à partir des données réelles de consommation entre les différents types de jours déterminés par la structure du calendrier. Par exemple, si on doit prévoir un samedi à partir d'un vendredi, la ressemblance des vendredis aux lundis, mardis, mercredis et aux autres vendredis fait que le samedi est prévu par un mélange de jours de la semaine et de jours du week-end. Il serait souhaitable de ne prévoir un samedi qu'à partir de samedis.

Dans le prochaine paragraphe nous essayons de mettre en oeuvre des corrections au modèle de base KWF tenant compte des points ci-dessus: un niveau moyen qui évolue tout au long de l'année et la présence de groupes de jours.

5 Corrections pour prévoir en présence de non stationnarités.

Nous allons explorer des pistes pour corriger les problèmes rencontrés lors de l'application de la méthode de prévision fonctionnelle KWF que nous venons

d'explorer. D'abord, nous traitons le problème d'un niveau moyen non constant par le centrage des segments du passé et la remise à niveau du segment prévu. Ensuite, nous travaillerons sur l'existence de groupes de segments.

Pour faciliter la discussion nous allons noter les parties approximation et détails de chaque fonction Z_i comme $\mathcal{S}_i(t)$ et $\mathcal{D}_i(t)$ respectivement, donc

$$\begin{aligned} Z_i(t) &= \sum_k c^{(i)}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j\geq j_0}\sum_k d^{(i)}_{j,k}\psi_{j,k}(t) \\ &= \mathcal{S}_i(t) + \mathcal{D}_i(t). \end{aligned}$$

5.1 Centrage et remise à niveau des courbes.

5.1.1 Le principe.

Le problème apparaît dès lors qu'interviennent dans le prédicteur $\widehat{Z_{n+1}}(t) = \sum_{m=1}^{n-1} w_{m,n} Z_{m+1}$ des courbes Z_{m+1} pressentant des niveaux moyens très différents. Dans ce cas, il est utile de centrer les courbes avant de calculer la prévision (centrée), puis remettre à niveau la prévision.

Les deux phases de la méthode ont un rôle différent face au centrage de courbes. Dans la première phase, recherche des blocs du passé les plus semblables au bloc présent Z_n, la distance proposée consiste à centrer les données de façon implicite. En effet, les coefficients d'approximation ne sont pas pris en compte ce qui conduit à un centrage implicite.

Cependant, quand l'on combine les futurs $Z_{m+1}, m = 1, \ldots, n-1$ dans la deuxième phase, le centrage doit être fait explicitement. Nous allons utiliser la méthode de base pour prévoir les coefficients d'ondelettes associés à la partie détail $\mathcal{D}_{n+1}$ de la courbe de charge; puis nous allons prévoir les coefficients d'échelle, liés à la partie d'approximation $\mathcal{S}_{n+1}$. Une fois les deux ensembles de coefficients déterminés, la transformée inverse en ondelettes permet d'obtenir la prévision du bloc futur. En définitive, la prévision pour le segment $n+1$ s'écrit $\widehat{Z_{n+1}(t)} = \widehat{\mathcal{S}_{n+1}(t)} + \widehat{\mathcal{D}_{n+1}(t)}$. Le processus fonctionnel $\mathcal{D}_{n+1}(t)$ étant centré, nous pouvons utiliser la méthode de base pour obtenir sa prévision

$$\widehat{\mathcal{D}_{n+1}(t)} = \sum_{m=1}^{n-1} w_{m,n}\mathcal{D}_{n+1}(t),$$

où les poids $w_{m,n}$ sont donnés par l'équation (5). Ensuite, pour la prévision de $\mathcal{S}_{n+1}(t)$ nous allons explorer différentes variantes:

BASE Lorsque nous avons utilisé le modèle KWF dans la section précédente, nous avons prévu la partie approximation par une moyenne pondérée des parties approximations du passé. Les poids étant les mêmes que ceux calculés à partir de la ressemblance entre les détails des courbes. La prévision de $\mathcal{S}_{n+1}$ est obtenue par $\widehat{\mathcal{S}_{n+1}(t)} = \sum_{m=1}^{n-1} w_{m,n}\mathcal{S}_{m+1}(t)$. Si bien que cette variante a l'avantage d'être simple et d'être une extension naturelle de la méthode originale, cependant elle peut s'avérer insuffisante car elle revient

à ne pas centrer les courbes lors de la deuxième phase de la méthode. Les autres variantes testées devront donc être meilleures que celles-ci. Pour cette variante, la prévision globale est

$$\widehat{Z_{n+1}(t)} = \sum_{m=1}^{n-1} w_{m,n}\mathcal{S}_{m+1}(t) + \sum_{m=1}^{n-1} w_{m,n}\mathcal{D}_{m+1}(t).$$

PRST Une alternative simple est de considérer la persistance: nous supposerons alors que le niveau moyen du lendemain est le même que celui d'aujourd'hui, soit $\widehat{\mathcal{S}_{n+1}(t)} = \mathcal{S}_n(t)$. La prévision pour le segment Z est alors:

$$\widehat{Z_{n+1}(t)} = \mathcal{S}_n(t) + \sum_{m=1}^{n-1} w_{m,n}\mathcal{D}_{m+1}(t).$$

Après utilisation de cette voie nous avons constaté que le niveau de la veille n'est pas forcément la meilleure référence pour corriger la moyenne. Bien que ce soit vrai pour les jours de la semaine, pour les jours de week-end la meilleure référence est celle de la semaine antérieure. Nous utilisons donc, la stratégie hybride de choisir pour les jours de la semaine le niveau de la veille et pour les jours week-end le niveau de la dernière semaine.

DIFF Dans l'article Poggi (1994) sur la prévision non paramétrique de la consommation électrique en tant que processus multivarié, l'auteur propose de corriger le niveau moyen de la prévision en considérant la transition des niveaux moyens entre un jour et son lendemain. Dans notre notation, cela revient à prévoir le niveau moyen à partir du niveau du jour témoin plus une correction par différences premières des niveaux du passé: $\widehat{\mathcal{S}_{n+1}(t)} = \mathcal{S}_n(t) + \sum_{m=1}^{n-1} w_{m,n}\Delta(\mathcal{S}_n)(t)$. Nous pouvons alors écrire la prévision de Z comme

$$\widehat{Z_{n+1}(t)} = \underbrace{\mathcal{S}_n(t) + \sum_{m=2}^{n-1} w_{m,n}\Delta(\mathcal{S}_n)(t)}_{\widehat{\mathcal{S}_{n+1}(t)}} + \sum_{m=1}^{n-1} w_{m,n}\mathcal{D}_m(t).$$

SAR Dans la dernière variante, nous exploitons le fait d'avoir choisi $j_0 = 0$ (e.g. l'approximation à la résolution la plus grossière.). Dans ce cas, les fonctions d'approximation sont constantes et proportionnelles au niveau moyen de chaque fonction. De plus, la suite $\{c_{0,0}^{(i)}\}, i = 1, \ldots, n$ est simplement une série temporelle unidimensionnelle. Grâce aux ondelettes cette série est plus simple que sur la série originale. Nous allons utiliser une modélisation classique des séries temporelles du type Box-Jenkins par exemple. Cette voie apporte une alternative paramétrique mais rend plus complexe la prévision car il faut régler deux modèles qui ne "communiquent pas entre eux". Nous avons testé plusieurs variantes. Le modèle retenu pour les comparaisons est assez simple. Il consiste en une modélisation SAR avec 2 coefficients saisonniers (hebdomadaires) et quatre journaliers.

5.1.2 Les résultats.

Nous reprenons les données de la consommation d'électricité pour tester les corrections que nous venons de décrire. Dans un premier temps nous nous concentrons sur la comparaison des différentes corrections par centrage en ne regardant que les résultats à fenêtre fixe.

Bloc	BASE	PRST	DIFF	SAR
lundi	7.1	4.75	2.07	3.89
mardi	8.11	4.49	2.11	3.63
mercredi	7.63	2.56	2.1	2.51
jeudi	8.03	2.06	1.97	2.43
vendredi	8.72	2.59	2.06	2.84
samedi	7.62	5.88	6.94	4.18
dimanche	8.51	5.14	1.51	2.26
férié	12.9	14.74	9.85	13.54
Global	8.11	4.24	2.91	3.42

Table 2: MAPE par type de jour et global pour la méthode de base BASE, et les variantes de centrage: PRST le niveau moyen est mis au niveau du jour témoin, DIFF le niveau moyen est une moyenne pondérée des différences premières des niveaux) et SAR modèle autorégressif saisonnier. La valeur optimale de h_n pour l'estimation FIX de la fenêtre est 2173.

Les résultats (cf. Tableau 2) montrent une amélioration dès lors que l'on corrige le niveau moyen, vis-à-vis de l'option de ne pas corriger (méthode de BASE). Parmi les options présentées, la correction par différences premières DIFF s'avère la plus performante pour tous les types de jour sauf les samedis. Comme auparavant, nous sommes intéressés par la distribution des erreurs dans l'année. L'évolution du MAPE par jour selon type de jour est présentée dans la Figure 7. Par rapport à la Figure 6 nous nous apercevons que globalement les niveaux d'erreur sont plus bas (l'étendue des ordonnées est presque divisé par deux). Deux remarques: d'une part les plus hauts niveaux d'erreurs pendant l'hiver sont plus faibles, d'autre part, le problème signalé précédemment concernant les grandes erreurs de prévision des jours samedi (en gras dans le graphique) est encore plus évident. Avec les jours fériés, les samedis présentent les erreurs les plus importantes (voir Tableau 2).

Afin de comparer les performances entre l'utilisation de la fenêtre FIX (calculée une seule fois avant la période de prévision) et la fenêtre DYN (recalculée pour chaque jour), le Tableau 3 présente le MAPE moyen par type de jour pour ces variantes. Nous y voyons que l'utilisation d'une fenêtre FIX entraîne une très légère dégradation globale de la méthode, sans vraie portée. Cependant, la dégradation ne se produit qu'à partir du moment où des jours avec une dynamique de consommation très différente (comme les jours fériés et les premiers jours de froid) entrent dans l'historique récent du prédicteur.

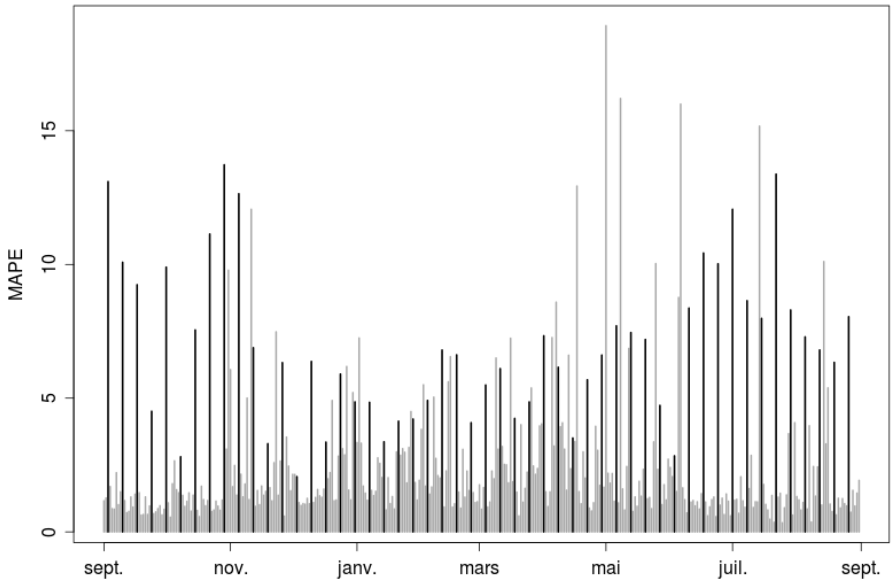

Figure 7: Evolution du MAPE journalier pour le modèle avec centrage DIFF. Les barres en gras correspondent aux samedis.

.

Bloc	FIX	DYN
lundi	2.07	2.09
mardi	2.11	2.12
mercredi	2.1	2.13
jeudi	1.97	1.94
vendredi	2.06	2.09
samedi	6.94	7.05
dimanche	1.51	1.52
férié	9.85	10
Global	2.91	2.94

Table 3: MAPE par type de jour et global pour la méthode DIFF. La valeur de h_n est calculée par la variante FIX à gauche et par DYN à droite.

Nous avons donc réussi à contrôler le problème du niveau variable. Le centrage de chaque courbe dans la construction du prédicteur plus la prévision du niveau par différences premières s'avère une solution prometteuse. Nous avons ainsi gagné en performance globale et nous avons déjà atteint des niveaux

d'erreurs de prévision qui ne sont pas très éloignés de ceux des modèles opérationnels (particulièrement pour les dimanches).

Passons au problème des samedis. Comme nous avons déjà dit, la transition des jours de semaine à un jour du week-end est une information que nous n'introduisons pas dans le modèle. Néanmoins, c'est une information que nous avons au moment de faire la prévision et que nous pouvons incorporer à notre prédicteur. Nous allons aborder ce problème (désormais nous présenterons les prévisions uniquement avec la correction par niveau DIFF).

5.2 Correction par groupes.

5.2.1 Le principe.

L'éventuelle existence de groupes de jours est le deuxième point à introduire pour rendre plus raisonnable l'hypothèse de stationnarité. Nous avons abordé le problème des niveaux différents dans la suite d'approximations $\{\mathcal{S}_n\}_n$. Nous nous concentrons maintenant sur la série de détails $\{\mathcal{D}_n\}_n$. L'ensemble des détails contribuent à l'allure de la courbe. Il est bien connu que la forme de la courbe journalière dépend du calendrier: les jours de week-end présentent des allures très différentes de celles des jours de semaine car l'activité des consommateurs change. La forme de la courbe dépend aussi de variables climatiques comme la température ou la nébulosité entre autres.

Jusqu'ici, la méthode utilise tous les segments du passé pour produire une prévision. Si des groupes des jours existent, nous voudrions utiliser dans la prévision seulement les jours qui appartiennent au même groupe que le dernier jour observé. Dès lors que nous utiliserons moins d'observations, deux effets contraires se produiront. D'un côté, la variance du prédicteur sera plus grande. De l'autre côté, si les observations appartiennent à une même classe, nous nous attendons une diminution du biais du prédicteur. Nous espérons que l'erreur quadratique moyenne du prédicteur aura une diminution nette grâce à une plus forte diminution du biais. L'intuition de ce point est de considérer des classes de stationnarité. Alors, l'hypothèse de stationnarité du processus fonctionnel nécessaire pour appliquer la méthode originale doit être valide mais à l'intérieur de chacun des groupes de jours.

Du point de vue pratique, l'incorporation de l'information des groupes dans l'étape de la prédiction se fait avec un traitement dans le calcul de la ressemblance. Les poids $w_{m,n}$ sont redéfinis en fonction de l'appartenance du jours m et n au même groupe:

$$\widetilde{w}_{m,n} = \frac{w_{m,n}\,\mathbb{1}_{[gr(m)=gr(n)]}}{\sum_{m=1}^{n} w_{m,n}\,\mathbb{1}_{[gr(m)=gr(n)]}},$$

où $\mathbb{1}_{[gr(m)=gr(n)]}$ vaut 1 si les groupe $gr(n)$ du n-ème jour est égal au groupe du $m-$ème jour et zéro sinon. Ceci équivaut à mettre à zéro la valeur de l'indice pour tous les jours m qui n'appartiennent pas au même groupe que le jour n, puis à renormaliser les poids pour avoir une somme égal à 1.

Bien entendu, nous avons supposé que l'information des groupes de jours est disponible au moment de faire la prévision. C'est le cas si nous basons les groupes sur la structure du calendrier. Sinon, les groupes peuvent être déterminés à partir d'une méthode de classification non supervisée (voir Cugliari (2011), partie II).

5.2.2 Groupes de jours.

Nous explorons ici des groupes de jours basés sur l'information du calendrier. L'option la plus simple est d'utiliser les jours de la semaine comme groupes. La connaissance acquise par les experts d'EDF sur la courbe de charge nous amène aux groupes suivants: {lundi}, {mardi, mercredi, jeudi}, {vendredi}, {samedi}, {dimanche} et {fériés}. En effet, les jours du milieu de la semaine ont une forme très similaire. Ceux-ci se différencient de la forme des lundis notamment dans la montée du matin, et de la forme des vendredis dans la forme de l'après midi.

Bien que ce groupement soit simple, il ne prend en compte que la classification du jour témoin. Or, les jours fériés ne seront jamais bien prévus car rien dans la veille n'informe la méthode que le jour à prévoir est particulier. D'ailleurs, nous avons remarqué l'importance d'inclure l'information des transitions entre jours quand nous avons développé la variante de centrage par différences premières. En conséquence, nous allons utiliser comme groupes les transitions entre les types de jours du calendrier. La classification de jours selon les transitions possibles, que nous appellerons CALEN-TR, est l'objet de la Figure 8.

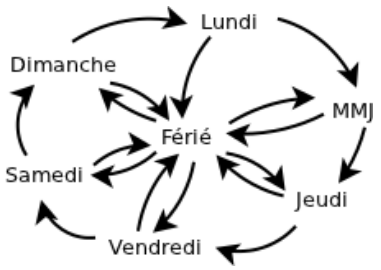

Figure 8: Les transitions possibles entre types de jours. Le sigle MMJ est utilisé pour indiquer le groupe formé par les mardis, mercredis et jeudis.

5.2.3 Les résultats.

Les résultats des performances de prévision pour les données réelles de la consommation d'électricité sont dans le Tableau 4. La prévision est faite par correction de niveau avec le centrage DIFF et avec correction de groupes par le traitement exposé ci-dessus. Pour le calcul de la largeur de fenêtre, nous nous centrons dans un premier temps sur les résultats de performance de prévision pour la fenêtre estimée par la méthode FIX. Remarquons que l'introduction de groupes

de jours a rendu plus complexe le modèle et maintenant nous avons une fenêtre par groupe de jour. Cependant, une conséquence latérale de l'introduction de groupes a été la réduction du temps de calcul. La méthode de base KWF et les variantes issues de la correction par niveau ont pris environ 10 fois plus de temps pour obtenir les prévision de toute l'année que lorsque des groupes sont introduits.

Bloc	DIFF	CALEN-TR (FIX)	CALEN-TR (DYN)
lundi	2.07	2.11	1.95
mardi	2.11	1.66	1.67
mercredi	2.1	1.60	1.54
jeudi	1.97	1.24	1.23
vendredi	2.06	1.66	1.66
samedi	6.94	1.57	1.57
dimanche	1.51	1.60	1.55
férié	9.85	2.59	3.33
Global	2.91	1.66	1.64

Table 4: MAPE par type de jour et global pour la corrigée par centrage DIFF (sans correction par groupes), la méthode avec groupe calendaire CALEN-tr et fenêtre fixe (FIX) et enfin la méthode avec groupe calendaire CALEN-tr et fenêtre dynamique (DYN).

A partir du Tableau 4, nous observons une amélioration globale pour la méthode de prévision avec correction par groupes. Il faut remarquer que nous avons réussi à corriger les hauts niveaux d'erreur pour les samedis. La solution proposée est efficace car on ne prévoit les samedis qu'avec des samedis. De plus, la prévision des jours fériés se voit améliorée. La méthode de prévision semble être appropriée même pour ce type de jours qui sont souvent très difficiles à prévoir.

Passons maintenant à la comparaison avec la version de calcul de fenêtre DYN (colonne à droite du Tableau 4). Nous observons ici une performance supérieure de la version à fenêtre mobile par rapport à l'alternative FIX. Dans la Figure 9 nous montrons l'évolution de la fenêtre selon les groupes de jours. Les lignes correspondent à des trajectoires lissées de la fenêtre en fonction du temps par jour de la semaine. La fenêtre calibrée par la méthode dynamique paraît suivre une évolution stable tout au long de l'année pour la plupart des types de jours. Cependant, les lundis affichent une augmentation importante autour l'hiver, souligné par une bosse dans le graphique, qui est due à des lendemains des jours fériés d'hiver. En effet, les jours fériés présentent un comportement très différent selon que le jour férié est en saison froide ou chaude.

L'évolution du MAPE par type de jour pour la prévision avec correction de centrage DIFF et groupes CALEN-TR est présenté dans la Figure 10. Cette évolution est plus homogène dans l'année que celle de la méthode de base (Figure

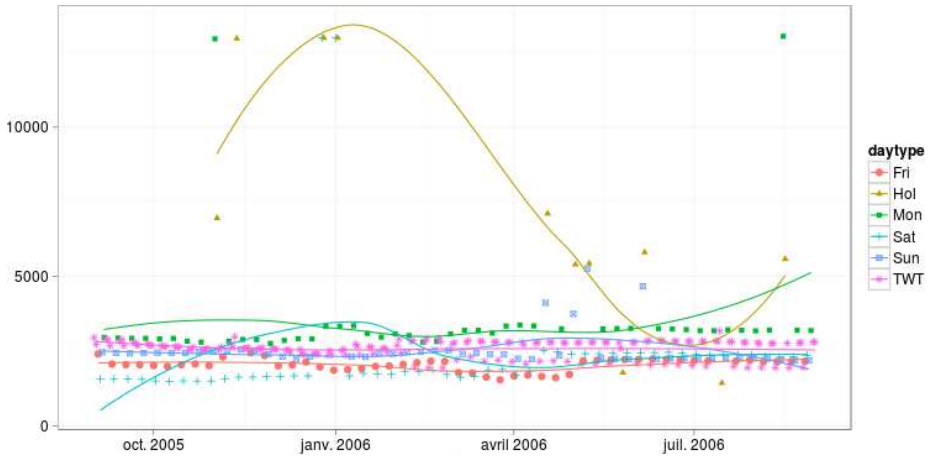

Figure 9: Valeur de fenêtre calibrée par la variante DYN dans la prévision avec centrage DIFF et groupes CALEN-TR.

6). Les problèmes signalés: niveau moyen non constant et présence de groupes de jours, semblent être résolus. La plupart des jours qui présentent encore de hauts niveaux d'erreur sont des jours fériés.

5.3 Analyse de quelques éléments du prédicteur.

Après avoir retenu un modèle qui fournit des prévisions globalement satisfaisantes, nous examinons le rôle de du vecteur de poids et le paramètre j_0 qui détermine les échelles informatives.

Répartition du vecteur de poids. Nous pouvons nous interroger sur le nombre et le type de blocs que la méthode sélectionne pour construire la prévision. Analysons la répartition des poids dans le temps. Considérons l'exemple de la Figure 11 où nous cherchons les jours qui ressemblent au dimanche 2 juillet 2006. La forme de la répartition montre que les poids des jours croissent à proximité du jour en question. En plus, les poids non nuls sont attribués seulement aux jours qui correspondent à la même position du calendrier que le jour de référence.

Le nombre de jours voisins, c'est-à-dire ceux qui participent avec un poids non nul dans la prévision, dépend du type de jour et de la position dans l'année. Pour chaque type de jour, le nombre de jours voisins décroît à fur et à mesure que nous nous approchons de l'hiver. Puis, il y a une rapide augmentation de ce nombre vers les jours d'été. Ceci reflète la plus importante hétérogénéité des jours d'hiver où l'effet de la température sur la consommation produit des trajectoires plus particulières.

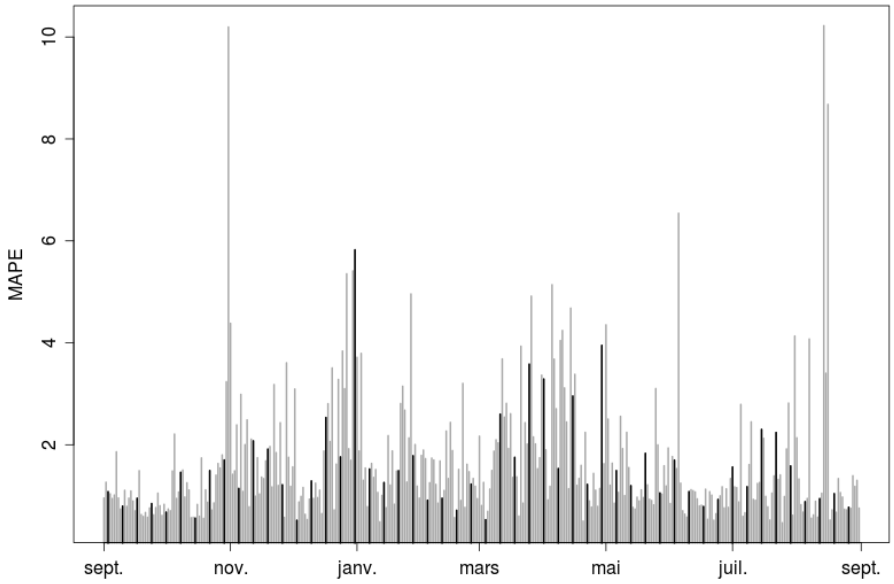

Figure 10: Évolution du MAPE journalier pour la méthode avec correction par centrage DIFF et correction par groupes CALEN-TR et la fenêtre par DYN. Les barres en gras correspondent aux samedis.

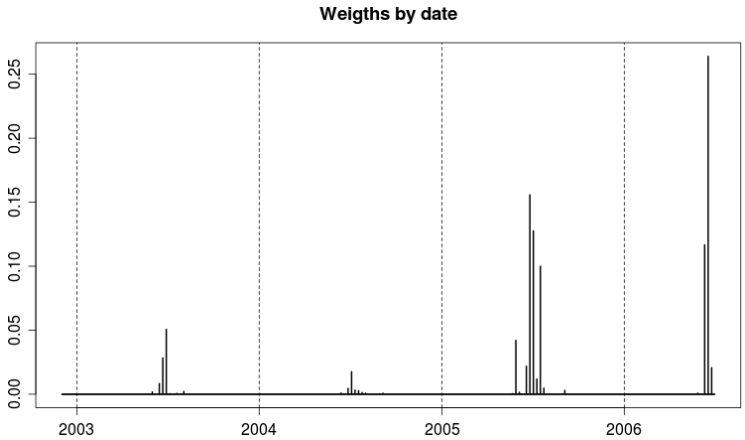

Figure 11: Vecteur de poids pour le dimanche 2 juillet 2006.

Échelles informatives. L'indice de similarité (Équation (5)) calculé dans la première phase de la méthode est basé sur les coefficients d'ondelettes. La

distance proposée (basée sur la DWT) entre deux blocs $Z_m(t)$ et $Z_{m'}(t)$ est

$$D(m,m') = \sum_{j=j_0+1}^{J} \left\{ 2^{-j} \sum_{k=0}^{2^j-1} (d_{j,k}^m - d_{j,k}^{m'})^2 \right\}^{1/2} .$$

Le niveau j_0 est l'échelle à partir de laquelle nous séparons approximation et détails de la trajectoire temporelle en concentrant dans l'approximation l'essentiel des comportements non stationnaires. Pour mieux comprendre cette distance, nous affichons dans la Figure 12 les contributions de chaque échelle de la DWT à l'énergie globale pour quatre semaines d'historique de la consommation d'électricité.

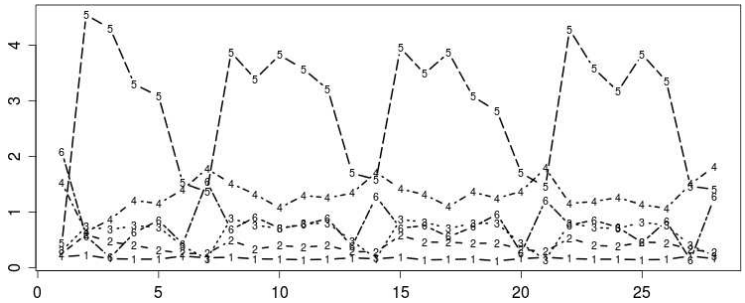

Figure 12: Contributions à l'énergie globale de la courbe de charge journalière par échelle sur quatre semaines. Chaque échelle est représente par une trajectoire.

Il est facile de voir que seules les échelles associés aux résolutions les plus grandes sont informatives. Les échelles plus fines n'apportent pas d'information significative, leurs contributions à l'énergie globale sont faibles et constantes. Donc, elles ne semblent pas informatives dans le calcul de la distance. Pour le confirmer, nous allons faire la prévision avec la méthode retenue (centrage DIFF et groupes CALEN-TR). Dans l'expérience nous calculons les distances seulement à partir de l'échelle j^* avec $j^* = 1, 2 \ldots$, et nous faisons la comparaison avec le calcul par défaut ($j^* = 0$). Les résultats (Table 5) montrent que les deux échelles les plus fines n'apportent effectivement pas d'information significative.

6 Conclusion et perspectives

Dans l'article nous proposons une manière efficace de prévoir une série temporelle continue mais échantillonnée, en tenant compte de la structure continue sous-jacente du processus. L'utilisation d'une suite de fonctions permet d'obtenir une modélisation souple car aucune hypothèse n'est faite sur la structure de dépendance à l'intérieur de chaque fonction, sur chacun des segments temporels associés. En autre, des corrections et des extensions du modèle sont nécessaires pour rendre l'hypothèse de stationnarité du processus fonctionnel

Bloc	CALEN-TR	j^*
lundi	1.95	2.06
mardi	1.67	1.68
mercredi	1.54	1.56
jeudi	1.23	1.34
vendredi	1.66	1.7
samedi	1.57	1.48
dimanche	1.55	1.6
férié	3.33	2.7
Global	1.64	1.66

Table 5: MAPE par type de jour et global pour le modèle retenu (centrage DIFF et groupes calendaire CALEN-TR et) la version avec $j^* = 2$. La fenêtre h_n est réglé par DYN.

plus raisonnable. Cette stratégie s'avère utile pour le problème de prévision de la consommation d'électricité, mais aussi pour une tâche de classification non supervisée associée (voir Antoniadis *et al.* (2010)).

En particulier, il faut noter que la démarche suivie n'utilise que l'historique de la consommation électrique. Ceci est d'autant plus remarquable que le modèle qu'il est bien connu que la consommation d'électricité française est très sensible à des conditions météorologiques. Malgré le fait que dans certains situations il peut être souhaitable d'utiliser un modèle de ces caractéristiques, il peut être utile d'incorporer de l'information exogène dans le modèle de prévision fonctionnel. Une première piste dans cette direction est le travaux mené dans Cugliari (2011) pour incorporer dans le modèle autoregressif hilbertien une dépendance de l'opérateur linéaire associé au processus à une donnée exogène.

Une deuxième perspective de travail consiste à fournir une région de confiance pour la prévision. Dans la proposition originale du modèle KWF pour des processus fonctionnels stationnaires, les auteurs proposent de construire cette région par une technique de rééchantillonnage. Bien entendu, des corrections similaires à celles proposées dans la Section 5 devront être mises au point pour s'adapter au cas considéré ici.

References

[1] A. Antoniadis, B. Brossat, J. Cugliari, and J.-M. Poggi. Clustering functional data with wavelets. In *Proceedings of COMPSTAT 2010*, pages 697–704, 2010.

[2] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *Journal-Royal Statistical Society Series B Statistical Methodology*, 68(5):837, 2006.

[3] A. Antoniadis, E. Paparoditis, and T. Sapatinas. Bandwidth selection for functional time series prediction. *Statistics & Probability Letters*, 79(6):733 – 740, 2009.

[4] D. Bosq. Modelization, nonparametric estimation and prediction for continuous time processes. In George Roussas, editor, *Nonparametric functional estimation and related topics*, pages 509–529. NATO ASI Series, 1991.

[5] D. Bosq. *Linear processes in function spaces: Theory and applications.* Springer-Verlag, New York, 2000.

[6] A. Bruhns, G. Deurveilher, and J.S. Roy. A non linear regression model for mid-term load forecasting and improvements in seasonality. In *Proceedings of the 15th Power Systems Computation Conference*, pages 22–26, 2005.

[7] J. Cugliari. *Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité.* PhD thesis, Université Paris Sud, 2011.

[8] M. Devaine, Y. Goude, and G. Stoltz. Forecasting the electricity consumption by aggregating specialized experts; a review of the sequential aggregation of specialized experts, with application to slovakian and french country-wide one-day-ahead (half-)hourly predictions. preprint: hal-00484940, 2011.

[9] V. Dordonnat, S. J. Koopman, M. Ooms, A. Dessertaine, , and J. Collet. Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics and Data Analysis (In press).*, 52, 2011.

[10] V. Dordonnat, S. J. Koopman, M. Ooms, A. Dessertaine, and J. Collet. An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24:566–587, 2008.

[11] W. Härdle. *Applied nonparametric regression.* Cambridge University Press, 1990.

[12] S.G. Mallat. *A wavelet tour of signal processing.* Academic Press, 1999.

[13] G. Nason. *wavethresh: Wavelets statistics and transforms.*, 2010. R package version 4.5.

[14] A. Pierrot and Y. Goude. Short-term electricity load forecasting with generalized additive models. In *Proceedings of ISAP power 2011*, 2011.

[15] J.-M. Poggi. Prévision non paramétrique de la consommation électrique. *Rev. Statistique Appliquée*, XLII(4):93–98, 1994.

[16] J.O. Ramsay and C.J. Dalzell. Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society. Series B*, 53(3):539–572, 1991.

[17] J.O. Ramsay and B.W. Silverman. *Functional data analysis.* Springer-Verlag, New Yok, 1997.

[18] J.O. Ramsay and B.W. Silverman. *Applied functional data analysis: methods and case studies.* Springer Verlag, 2002.

[19] J.W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204:139–152, 2010.

RESEARCH CENTRE
SACLAY – ÎLE-DE-FRANCE

Parc Orsay Université
4 rue Jacques Monod
91893 Orsay Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399