HAL
open science

Une approche fonctionnelle pour la prévision non-paramétrique de 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. Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité. Journal de la Societe Française de Statistique, 2014, 155 (2), pp.202- 219. hal-00814530v2

HAL Id: hal-00814530

https://hal.science/hal-00814530v2

Submitted on 17 Apr 2013

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.

Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité

Anestis Antoniadis
anestis.antoniadis@imag.fr

Xavier Brossat
xavier.brossat@edf.fr

Jairo Cugliari
Jairo.Cugliari@inria.fr

Jean-Michel Poggi
Jean-Michel.Poggi@paris-descartes.fr

17 avril 2013

Résumé

Dans un travail récent (Antoniadis *et al.* (2012)), les auteurs ont proposé un modèle de prévision pour des séries chronologiques fonctionnelles en présence de non stationnarités. Ce modèle a été appliqué à la demande d'électricité d'EDF. Le principe général du modèle de prévision consiste à trouver dans le passé des contextes similaires à la situation présente et prévoir le futur par une combinaison linéaire des futurs des passés les plus semblables au présent. La notion de similarité est basé sur les ondelettes et plusieurs stratégies sont mises en oeuvre pour prendre en compte les diverses sources d'instationnarités.

Nous explorons dans ce second article, trois aspects de ce modèle qui complètent la méthodologie originale tout en mettant en évidence l'utilité industrielle de la méthode. D'abord nous abordons la construction d'un intervalle de confiance pour le prédicteur fonctionnel. Ensuite, nous examinons la flexibilité et la simplicité du modèle pour fournir, sans effort supplémentaire, des prévisions à des horizons de plus en plus éloignés. Enfin, dans le contexte applicatif spécifique, nous examinons la capacité de la méthode à fournir de prévisions de bonne qualité en présence d'instationnarités subtiles du signal, conséquences de pertes de clients selon divers scénarios.

1 Introduction

Les développements théoriques et les applications pratiques associées à l'analyse de données fonctionnelles ont été majoritairement guidées par le cas des observations

indépendantes. Or, il existe un vaste champ d'applications dans lequel cette hypothèse n'est pas raisonnable. Notamment, dès lors que l'on considère des enregistrements sur une grille de temps de plus en plus fine et que l'on suppose qu'ils sont un échantillonnage d'un signal à temps continu inconnu.

Formellement, le problème peut s'écrire en considérant un processus stochastique continu $X = (X(t), t \in \mathbb{R})$. Alors, l'information contenue dans une trajectoire de X observée sur l'intervalle $[0, T], T > 0$ est aussi représentée par un processus discret $Z = (Z_k(t), k = 0, \ldots, n; t \in [0, \delta])$ où $Z_k(t) = X((\delta - 1)k + t)$ est l'objet de la segmentation de la trajectoire de X en n blocs de taille $\delta = T/n$ (Bosq (1991)). Ainsi, le processus Z est une série temporelle de fonctions. Nous pouvons par exemple prévoir $Z_{n+1}(t)$ à partir des données $Z_1, \ldots, Z_n$. Ce qui revient à prévoir le comportement futur du processus X sur tout l'intervalle $[T, T + \delta]$ en ayant observé X sur $[0, T]$. Notons que par construction, les $Z_1, \ldots, Z_n$ sont en général des variables aléatoires fonctionnelles dépendantes.

Bien que toutes les données fonctionnelles dépendantes ne s'obtiennent pas par segmentation d'un processus à temps continu (voir par exemple Kokoszka (2012)), ce cas est d'un intérêt particulier dans l'étude de la consommation d'électricité. En effet, les ingénieurs de ce domaine considèrent naturellement les enregistrements discrets de consommation comme un échantillonnage de la courbe de charge d'un système électrique. La taille usuelle des segments, 1 jour, prend en compte le cycle journalier de la consommation.

Dans un travail récent (Antoniadis *et al.* (2012)), nous avons utilisé l'approche des séries chronologiques de fonctions pour mettre à point un modèle de prévision en présence de non stationnarités et nous l'avons appliqué à la demande d'électricité d'Électricité de France (EDF). Son principe consiste à trouver dans le passé des contextes similaires à la situation présente. Ainsi, on construit un vecteur de poids reflétant la similarité entre passés et présent. Puis, les trajectoires futures de ces situations passées sont moyennées en les pondérant par les similarités pour construire la prévision. Poggi (1994) propose pour le même problème d'utiliser un prédicteur de nature similaire mais appliqué à un processus multivarié. Ensuite, Antoniadis et al. (2006) proposent un cadre approprié pour des processus fonctionnels stationnaires à l'aide de la transformée en ondelettes. Ce dernière modèle est adapté et étendu au cas des processus fonctionnels non stationnaires (Cugliari (2011)).

Ainsi, on obtient une qualité de prévision qui est du même niveau que d'autres modèles utilisés par EDF alors même que notre modèle permet de représenter la série de manière simple et parcimonieuse. Nous évitons ainsi de modéliser de manière explicite le lien entre la consommation et les covariables météorologiques, connues pour être importantes dans la modélisation et souvent considérées comme indispensables à prendre en compte. Un autre avantage du modèle fonctionnel est sa capacité à fournir des prévisions à plusieurs horizons de manière simultanée en s'appuyant sur toute une portion de trajectoire du passé récent, plutôt que sur certains points comme le font les modèles univariés.

Nous prolongeons ici notre premier article en nous centrant sur l'application in-

dustrielle en explorant trois aspects qui complètent la méthodologie originale tout en mettant en évidence l'utilité industrielle de la méthode. Plus précisément, on aborde dans cette situation non stationnaire la construction d'un intervalle de confiance pour la prévision, puis on examine le lien entre la taille du bloc de référence et l'horizon de prévision, et enfin on examine la capacité de la méthode à fournir de prévisions de bonne qualité en présence d'instationnarités subtiles du signal conséquences de pertes de clients selon divers scénarios. Détaillons un peu plus chacun de ces points.

Tout d'abord, nous construisons un intervalle de confiance pour la prévision. Du point de vue industriel, l'intervalle permet de rendre compte de l'incertitude autour d'une prévision ponctuelle. L'analyse dans notre premier article, des erreurs de prévision de la méthode conçue pour le cas stationnaire nous a permis de mettre en évidence l'écart que les données de consommation d'électricité présentent vis-à-vis de l'hypothèse de stationnarité : l'évolution du niveau moyen et la présence de groupes de jours expliqués par la structure calendaire. À l'image du chemin parcouru dans la construction du prédicteur fonctionnel, nous sommes amenés à introduire le même type de correctifs afin de tenir compte de la présence de tels éléments instationnaires.

Ensuite, le lien entre la taille du bloc de référence (ou témoin) et l'horizon de prévision est étudié. Pour ce faire, nous soumettons la prévision de base, c'est-à-dire celle obtenue à partir d'un bloc de référence et horizon de prévision égal à un jour, à des modifications progressives des longueurs δ et h. En effet, la segmentation du processus X en blocs de taille journalière nous amène à une prévision à horizon d'un jour, ce qui revient à obtenir des prévisions pour les heures $T + h$ avec $0 \leq h \leq 24$. Bien qu'utile et fondé dans la prévision d'électricité, ce choix demeure arbitraire. Nous étudions donc d'abord la capacité de la méthode à continuer à fournir de bonnes prévisions pour diverses modifications de la taille du bloc témoin δ. Puis, nous analysons le lien entre ces deux éléments en les faisant varier en même temps. Ainsi, nous étudions de manière empirique la capacité du modèle à fournir de prévisions de qualité jusqu'à l'horizon hebdomadaire ou de prévoir en considérant des blocs témoins d'un durée infrajournalière.

Enfin, nous examinons la capacité de la méthode de prévision à s'adapter à de nouveaux contextes. Pour ce faire, l'application pratique nous guidera. La récente ouverture du marché électrique peut produire une variation du nombre et du type de clients de chacun des producteurs, dont EDF, du marché français de l'électricité. Ainsi, des clients partent ou arrivent, modifiant ainsi la structure de consommation, le degré de dépendance de la consommation à des variables météorologiques ainsi que la forme de la courbe de charge. Identifier l'information dans le passé utile pour prévoir dans le nouveau contexte devient un point central. À l'aide de trajectoires artificielles créées pour refléter les changements dans la série temporelle de la demande d'électricité, nous mesurons la capacité de la méthode à s'adapter à divers scénarios de pertes de clients.

Nous rappelons dans la section suivante les éléments nécessaires pour la compréhension de ce travail, mais si le lecteur a lu le premier article, sa lecture n'est pas indispensable. Nous y décrivons les données de consommation d'électricité et rappelons le prédicteur appelé KWF ainsi que les correctifs introduits pour prévoir une série fonctionnelle non stationnaire.

2 Les données et le modèle fonctionnel

2.1 Les données électriques

Les données sont les enregistrements de la consommation d'électricité française depuis 1996 au pas demi-horaire. Les caractéristiques principales – une tendance de long terme, plusieurs cycles saisonniers, une forte dépendance aux variables exogènes– peuvent être saisies grâce aux quatre graphiques de la Figure 1.

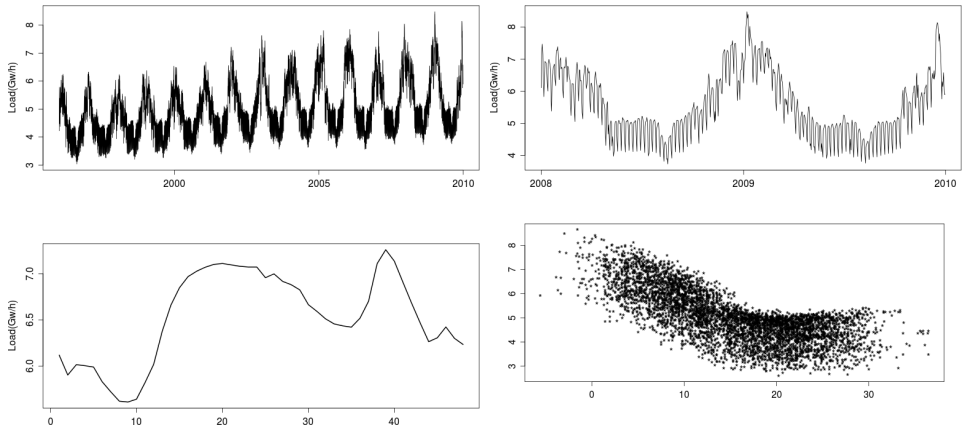

FIGURE 1 – Caractéristiques principales de la consommation d'électricité française.

Pour modéliser ces données, EDF utilise de modèles de régression où à chaque moment de l'année la consommation d'électricité est décrite par deux composantes. La première regroupe les facteurs indépendants des conditions météorologiques (tendance, saisonnalité annuelle, jour de la semaine, moment dans la journée), tandis que le deuxième capture les effets climatiques et dépend entre autres d'une température ressentie à l'intérieur des bâtiments. Le modèle de régression associé comporte un nombre élevé de paramètres (un millier environ) et nécessite comme historique, une longue période d'estimation pour garantir la robustesse de son estimation.

Ces modèles très performants en pratique permettent de satisfaire, aujourd'hui encore, les exigences et les besoins d'EDF. Cependant, les experts les considèrent comme trop lourds et insuffisants pour faire face aux nouveaux enjeux du domaine. En effet, la libéralisation récente du marché électrique et l'avènement de sources renouvelables dans les mélanges de production peuvent induire des non stationnarités dans le signal rendant difficile l'adaptation de tels modèles trop complexes.

2.2 Le modèle fonctionnel KWF

Nous considérons un processus stochastique supposé pour le moment, stationnaire $Z = (Z_k)_{k\in\mathbb{Z}}$ 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 se décompose 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 $w_{n,m}, m = 1, \ldots, n-1$ et obtenir la prévision souhaitée en moyennant les futurs des blocs correspondant aux indices $2, \ldots, n$ respectivement.

Première phase. Pour prendre en compte dans la dissimilarité la dimension infinie des objets à comparer, 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 une version tronquée de son développement obtenue par la transformée en ondelettes discrète (DWT)

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

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 contient l'information de la structure locale de la fonction. Pour deux segments observés $Z_i(t)$ et $Z_{i'}(t)$, nous utilisons la dissimilarité D définie comme

$$D(Z_i, Z_{i'}) = \sum_{j=j_0+1}^{J} 2^{-j} \sum_{k=0}^{2^j-1} (d_{j,k}^{(i)} - d_{j,k}^{(i')})^2. \tag{1}$$

Le processus Z étant supposé ici stationnaire, les coefficients d'approximation ne contiennent pas d'information utile pour la prévision puisqu'ils fournissent des moyennes locales. De ce fait, ils ne sont pas pris en compte dans la distance proposée. En d'autres termes, la dissimilarité D permet de trouver de motifs similaires entre courbes même si elles ont des approximations très différentes.

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

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

où K est un noyau de probabilité.

Finalement, nous pouvons appliquer 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. Si l'on note

$$w_{n,m} = \frac{K_{h_n}(D(Z_{n,J}, Z_{m,J}))}{\sum_{m=1}^{n-1} K_{h_n}(D(Z_{n,J}, Z_{m,J}))}, \tag{2}$$

ces poids permettent de réécrire le prédicteur comme un barycentre des futurs des segments du passé

$$\widehat{Z_{n+1}}(t) = \sum_{m=1}^{n-1} w_{n,m} Z_{m+1}(t). \tag{3}$$

Dans le cas où Z n'est pas un processus fonctionnel stationnaire, des correctifs dans le prédicteur (3) doivent être apportés pour prendre en compte la non stationnarité. Dans Antoniadis *et al.* (2012) des corrections sont proposées et leur efficacité est étudiée pour deux types de non-stationnarités : la présence d'une évolution du niveau moyen des approximations de la série et l'existence de classes de segments.

Il est pratique d'exprimer chaque courbe Z_i en fonction de deux termes $\mathcal{S}_i(t)$ et $\mathcal{D}_i(t)$ décrivant respectivement l'approximation et la somme des détails,

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

Quand les courbes Z_{m+1} présentent des niveaux moyens très différents apparaît le premier problème. Dans ce cas, il est utile de centrer les courbes avant de calculer la prévision (centrée), puis remettre à niveau la prévision lors de la seconde phase. Alors, 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

$$\widehat{\mathcal{D}_{n+1}(t)} = \sum_{m=1}^{n-1} w_{n,m}\mathcal{D}_{n+1}(t), \tag{4}$$

où les poids $w_{n,m}$ sont donnés par l'équation (2). Ensuite, pour la prévision de $\mathcal{S}_{n+1}(t)$ nous utilisons

$$\widehat{\mathcal{S}_{n+1}(t)} = \mathcal{S}_n(t) + \sum_{m=1}^{n-1} w_{n,m}\Delta(\mathcal{S}_{m+1})(t), \tag{5}$$

où $\Delta(\mathcal{S}_{m+1})(t) = \mathcal{S}_{m+1}(t) - \mathcal{S}_m(t)$. Pour résoudre le deuxième problème, nous incorporons l'information des groupes dans l'étape de la prédiction par la redéfinition des poids $w_{n,m}$ en fonction de l'appartenance des fonctions m et n au même groupe :

$$\widetilde{w}_{n,m} = \frac{w_{n,m}\mathbf{1}_{\{gr(m)=gr(n)\}}}{\sum_{m=1}^{n} w_{n,m}\mathbf{1}_{\{gr(m)=gr(n)\}}}, \tag{6}$$

où $\mathbf{1}_{\{gr(m)=gr(n)\}}$ vaut 1 si les groupes $gr(n)$ du n-ème segment est égal au groupe du $m-$ème segment et zéro sinon. Si les groupes ne sont pas connus, ils peuvent être déterminés à partir d'une méthode de classification non supervisée (voir Antoniadis *et al.* (2013)).

3 Intervalle de confiance pour la prévision

Nous abordons ici le problème de la construction d'un intervalle de confiance pour la prévision dans le modèle KWF. Plus précisément, il s'agit d'intervalles ponctuels et non d'une bande de confiance. Nous procéderons comme dans le cas de la mise au point du prédicteur, en supposant dans un premier temps que la suite fonctionnelle Z est un

processus stationnaire puis en apportant des correctifs à la première proposition. Ainsi, nous rappelons le cadre proposé dans Antoniadis *et al.* (2006) pour cette construction initiale. Ensuite nous traitons le cas plus général d'une suite Z instationnaire en proposant des correctifs à l'image de ceux proposés dans le premier article en les adaptant à la construction d'un intervalle de confiance pour la prévision.

3.1 Le cas stationnaire

Nous cherchons les quantités $L_{n+1,\alpha}(t_i)$ et $U_{n+1,\alpha}(t_i)$ telles que, pour chaque instant d'observation $t_i, i = 1, \ldots, N$, nous avons

$$P(L_{n+1,\alpha}(t_i) \leq Z_{n+1}(t_i) \leq U_{n+1,\alpha}(t_i)) \geq 1 - 2\alpha.$$

Les quantités doivent être estimées à partir de l'historique $Z_1, \ldots, Z_n$. Elles sont alors identifiées comme les quantiles α et $1 - \alpha$ de la distribution conditionnelle de Z_{n+1} étant données $Z_1, \ldots, Z_n$. Cette distribution est à la base même du modèle KWF car le prédicteur (3) est un estimateur de l'espérance conditionnelle $\mathbb{E}[Z_{n+1}|Z_1, \ldots, Z_n]$. Ainsi, il semble pertinent d'estimer les quantiles qui nous intéressent en utilisant le vecteur des poids qui interviennent dans le prédicteur ponctuel.

Les poids $w_{n,m}, m = 1, \ldots, n-1$ induisent une loi de probabilité discrète sur les observations $Z_2, \ldots Z_n$ puisqu'ils sont non négatifs et de somme 1. Dans un cadre similaire, Poggi (1994) propose un estimateur de la variance de la loi conditionnelle de $Z_{n+1|n}(t_i)$ de la forme $\sum_{m=1}^{n-1} w_{n,m}(Z_{m+1}(t_i) - \widehat{Z}_{n+1|n}(t_i))^2$. Puis, l'intervalle est obtenu comme le prédicteur ponctuel en rajoutant et retranchant 2 ou 3 fois la racine carrée de l'estimateur de la variance.

De manière alternative, les poids peuvent être utilisés pour faire un échantillonnage bootstrap des résidus de prévision pour ensuite estimer des quantiles sur les distributions boostrap à la place de la distribution empirique des données.

La procédure peut se résumer de la façon suivante :

1. Obtention de la prévision ponctuelle $\widehat{Z}_{n+1}$ par (3).
2. Génération de B pseudo réalisations bootstrap $Z_{n+1}^{(b)}$ telles que $P(Z_{n+1}^{(b)} = Z_{m+1}|Z_n) = w_{n,m}$.
3. Obtention des pseudo résidus $\widehat{R}_{n+1}^{(b)}(t_i) = Z_{n+1}^{(b)}(t_i) - \widehat{Z}_{n+1}(t_i), b = 1, \ldots, B$.
4. Pour chaque t_i, définir $\widehat{R}_{n+1,\alpha}(t_i)$ et $\widehat{R}_{n+1,1-\alpha}(t_i)$ les quantiles d'ordre α et $1 - \alpha$ de $\{\widehat{R}_{n+1}^{(b)}(t_i); b = 1, \ldots, B\}$.

Enfin l'intervalle à $(1 - \alpha)x100\%$ est donné par $[\widehat{L}_{n+1,\alpha}(t_i), \widehat{U}_{n+1,\alpha}(t_i)]$ où

$$\widehat{L}_{n+1,\alpha}(t_i) = \widehat{R}_{n+1,\alpha}(t_i) + \widehat{Z}_{n+1}(t_i) \qquad \text{et} \qquad \widehat{U}_{n+1,\alpha}(t_i) = \widehat{R}_{n+1,1-\alpha}(t_i) + \widehat{Z}_{n+1}(t_i).$$

3.2 Le cas instationnaire

Afin de tenir compte de la nature non stationnaire d'un tel processus fonctionnel, nous adaptons le schéma précédent pour la construction d'un intervalle de prévision. Il est bien sûr souhaitable que les correctifs apportés au prédicteur ponctuel soient de nature similaire à ceux utilisés pour l'intervalle de confiance.

Rappelons que nous traitons deux sources de non stationnarité : la présence de groupes de fonctions et l'évolution du niveau moyen. Pour la première, nous utilisons les poids définis dans (6) à la place de (2) pour générer les pseudo réalisations $Z^*_{n+1,b}, b = 1, \ldots, B$. Il ne reste qu'à établir les groupes, qui dans le cas que nous étudions sont déduites directement du calendrier.

À ce stade, si l'on calcule un intervalle de confiance dans notre application pratique, on s'aperçoit que son amplitude est vraisemblablement trop grande, vis-à-vis de ce qui est attendu. Il en est ainsi car les segments du passé qui interviennent dans le prédicteur (3) (avec les poids redéfinis) peuvent avoir des composantes d'approximation $\mathcal{S}_{m+1}$, des niveaux moyens très différents. Il semble approprié de quantifier seulement la variabilité due aux termes des incréments $\Delta(\mathcal{S}_{m+1})$. Pour ce faire, nous exploitons le fait que dans la pratique nous choisissons $j_0 = 0$ dans l'équation (1) et en conséquence la partie approximation de Z_i est codée par un seul coefficient d'échelle, $c^i_{0,0} = c^i_0$. Nous quantifions la variabilité associée à $\Delta(\mathcal{S}_{m+1})$ par l'estimation des quantiles empiriques de $\{\Delta(c_0^{m+1}), m = 2, \ldots, n\}$. Concrètement, la construction de l'intervalle de confiance se fait comme suit.

1. Obtention de la prévision ponctuelle $\widehat{Z}_{n+1} = \widehat{\mathcal{S}}_{n+1} + \widehat{\mathcal{D}}_{n+1}$ en utilisant les equations (5) et (4) respectivement avec les poids (6).
2. Génération de B pseudo réalisations bootstrap $Z^{(b)}_{n+1}$ telles que $P(Z^{(b)}_{n+1} = Z_{m+1}|Z_n) = \widetilde{w}_{n,m}$ (voir equation (6)). Nous notons $\mathcal{S}^{(b)}_{m+1}$ et $\mathcal{D}^{(b)}_{m+1}$ les parties approximation et détails respectivement, associées à $Z^{(b)}_{n+1}$.
3. Pour $b \in \{1, \ldots, B\}$, obtenir les pseudo résidus liés aux détails $\widehat{R}^{(b)}_{n+1}(t) = Z^{(b)}_{n+1}(t) - \widehat{Z}_{n+1}(t)$; et ceux liés aux approximations $\widehat{q}^{(b)}_{n+1} = \Delta c^{(b)}_0 - \Delta(\widehat{c_0^{n+1}})$ où $\widehat{c_0^{n+1}} = c^n_0 + \sum_{m=1}^{n-1} \widetilde{w}_{m,n} \Delta(c_0^{m+1})$.
4. Définir $\widehat{q}_{n+1,\alpha}$ et $\widehat{q}_{n+1,1-\alpha}$ comme les quantiles empiriques associées à la partie approximation de niveau α et $1-\alpha$ de $\{\widehat{q}^{(b)}_{n+1}, b = 1, \ldots, B\}$. Ensuite, nous appelons $\widehat{Q}_{n+1,\alpha}(t) = (c^n_0 + \widehat{q}_{n+1,\alpha})\phi_{0,0}(t)$ et $\widehat{Q}_{n+1,1-\alpha}(t) = (c^n_0 + \widehat{q}_{n+1,1-\alpha})\phi_{0,0}(t)$ le résultat d'appliquer l'inverse de la transformée en ondelettes sur les coefficients d'échelles estimés. De manière similaire, nous définissons sur la partie détails pour chaque t_i, $\widehat{R}_{n+1,\alpha}(t_i)$ et $\widehat{R}_{n+1,1-\alpha}(t_i)$ comme les quantiles empiriques au niveau α et $1-\alpha$ de $\{\widehat{R}^{(b)}_{n+1}(t_i); b = 1, \ldots, B\}$.

L'intervalle de confiance est donné par $[\widehat{L}_{n+1,\alpha}(t_i), \widehat{U}_{n+1,\alpha}(t_i)]$, où

$$\begin{aligned}\widehat{L}_{n+1,\alpha}(t_i) =& \widehat{Q}_{n+1,\alpha}(t_i) + \widehat{R}_{n+1,\alpha}(t_i) + \widehat{Z}_{n+1}(t_i) \qquad \text{et}\\ \widehat{U}_{n+1,\alpha}(t_i) =& \widehat{Q}_{n+1,1-\alpha}(t_i) + \widehat{R}_{n+1,1-\alpha}(t_i) + \widehat{Z}_{n+1}(t_i).\end{aligned}$$

3.3 Résultats numériques

La procédure décrite pour le cas instationnaire est utilisée pour construire des intervalles de confiance pour la prévision de la consommation d'électricité. En utilisant les données demi-horaires du jour J nous prévoyons les 48 demi-heures de jour $J+1$. L'exercice est répété tout au long d'une année. Nous considérons trois valeurs pour le seuil de confiance $1-\alpha = 0.95, 0.9$ et 0.8. Tout d'abord nous étudions le taux de couverture des intervalles obtenus. Il est défini par la proportion des points de la courbe de charge journalière qui sont inclus dans l'intervalle $I_{1-\alpha}(t_i)$ donné par $\{\widehat{L}_{n+1,\alpha}(t_i), \widehat{U}_{n+1,\alpha}(t_i)\}$,

$$\frac{1}{48}\sum_{i=1}^{48} \mathbf{1}_{\{Z(t_i)\in I_{1-\alpha}(t_i)\}}.$$

Le taux de couverture moyen est de 0.89, 0.85 et 0.8 respectivement pour les niveaux de 95%, 90% et 80%. Ce qui est satisfaisant.

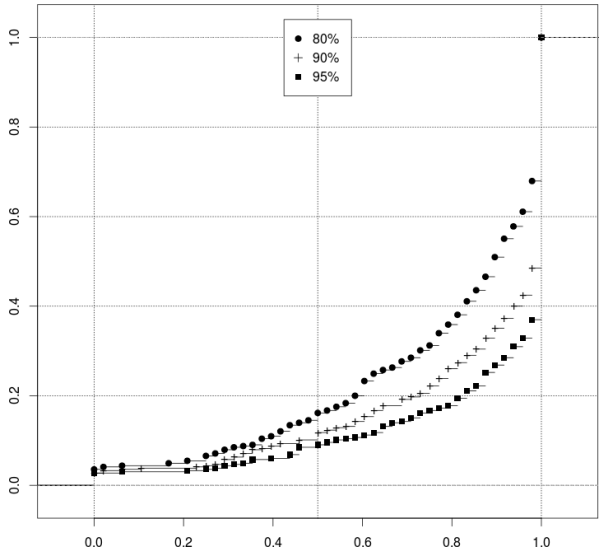

FIGURE 2 – Distribution cumulée du taux de couverture pour $1-\alpha = 0.8, 0.9$ et 0.95.

La Figure 2 présente les distributions cumulées des taux de couverture journalière. Nous pouvons constater que la proportion de jours de mauvaise couverture est relativement faible. A titre d'exemple, ces proportions pour une couverture de moins de 50%, représentent seulement 10%, 12% et 17% des jours respectivement pour les niveaux de 95%, 90% et 80%.

Le taux de couverture est cependant une mesure parfois trop synthétique de la qualité des intervalles. En effet, une très bonne couverture au prix des intervalles trop larges sont d'une utilité faible dans la pratique. Une façon alternative de procéder consiste à utiliser les pseudo réalisations bootstrap construites auparavant comme un outil de diagnostic visuel de variabilité. Nous pouvons ainsi superposer l'ensemble des réalisations bootstrap au prédicteur et son intervalle de confiance. La Figure 3 présente quatre cas concrets de prévisions avec leur tube de confiance et les pseudo réalisations associées.

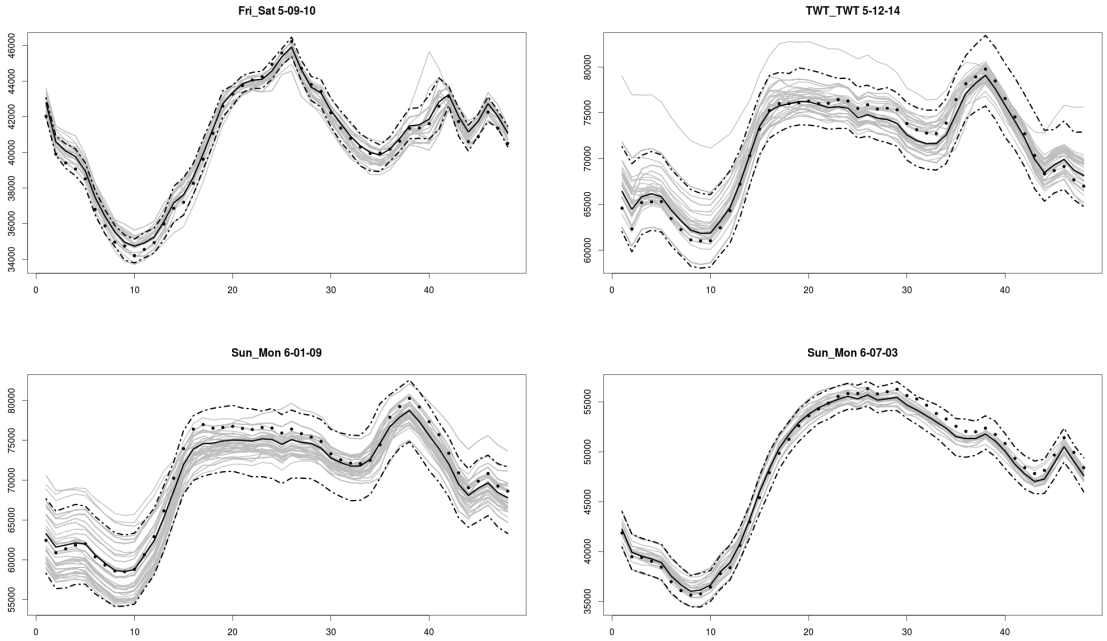

FIGURE 3 – Intervalles de confiance pour quatre jours de l'année.

Les deux cas de la diagonale principale présentent des propriétés de couverture souhaitables avec une amplitude qui semble raisonnable pour une prévision faite la veille. En revanche, les deux autres cas montrent des situations où les intervalles sont plus grands. Ceci arrive dès lors que la fonction de référence (le bloc témoin) a peu de similarité avec les courbes de l'historique, comme c'est le cas par exemple des jours fériés. Un deuxième facteur est la variabilité naturelle du phénomène. En effet, dans les périodes d'automne et de printemps la variabilité des conditions climatiques est plus importante que dans d'autres parties de l'année.

4 Lien entre taille du bloc témoin et horizon de prévision

Concrètement, après la segmentation de la chronique d'origine conduisant à une séquence de courbes de charge journalières, la méthode de prévision se résume en deux phases. D'abord la construction d'un indice de similarité $w_{n,m}$ entre la courbe de charge journalière actuelle $Z_n(t)$ et toutes les courbes de charge journalières du passé $Z_m(t), m = 1, \ldots, n-1$. Puis, les courbes futures $Z_{m+1}(t), m = 2, \ldots, n$ sont moyennées en utilisant les poids $w_{n,m}$ pour obtenir une prévision $\widehat{Z}_{n+1}$ de la courbe journalière du lendemain.

Deux tailles de segments interviennent en fait lors de ces deux phases : celle des segments des courbes des passés et celles des courbes du futur. En effet, lors de la

première phase la taille de la segmentation sert à fixer la largeurs des segments du passé à comparer au bloc témoin. Tandis que dans la deuxième phase δ détermine le plus grand horizon de prévision à travers la taille des segments futurs. Jusqu'à présent, ces deux tailles étaient égale à la taille de la segmentation qui est déterminée par le paramètre δ. Nous allons découpler ces deux quantités que nous appellerons désormais δ_P pour la taille des segments passés et δ_F pour la taille des segments futurs.

Notre objectif de prévision, si l'on revient au processus sous-jacent $X(t)$ s'exprime ainsi. A partir de la trajectoire observée sur $[0, n\delta]$ nous cherchons à prévoir $X(n\delta + t)$ avec $t \in [0, \delta_F]$. Notons les segments futur et passé du jour m respectivement par

$$\begin{aligned} Z_m^F(t) &= X(m\delta + t) && t \in [0, \delta_F], \\ Z_m^P(s) &= X(m\delta - \delta_P + s) && s \in [0, \delta_P]. \end{aligned}$$

Alors le jour n, nous prévoyons $Z_n^F(t)$ avec

$$\widehat{Z}_n^F(t) = \sum_{m=1}^{n-1} w_{n,m} Z_m^F(t) \qquad , t \in [0, \delta_F],$$

où à l'image de l'équation (2), $w_{n,m} \propto K_{h_n}(D(Z_m^P, Z_n^P))$.

Dans la suite, nous faisons varier les deux quantités δ_P et δ_F et mesurons leur impact sur la qualité de prévision de la consommation d'électricité.

4.1 Sensibilité à l'horizon de prévision

Commençons par fixer la taille des segments du passé $\delta_P = 1$ jour, pour faire varier le paramètre δ_F, la taille du segment à prévoir. Dans un contexte de prévision classique, cela revient à prévoir de manière simultanée à plusieurs horizons de prévision. Dans ce cas, δ_F détermine l'horizon maximal de la prévision simultanée fournie par KWF.

Pour l'expérience numérique, δ_F varie de 3 heures à une semaine. La qualité de prévision est mesurée par la moyenne des erreurs relatives (MAPE) sur le segment à prévoir, qui sont reportées dans la Figure 4 à droite. Chaque point représente la moyenne des MAPE sur l'année utilisée en fonction de δ_P. Sans surprise, l'erreur de prévision tend à être plus grande au fur et mesure que l'on produit des prévisions à des horizons de plus en plus lointains. Toutefois, la qualité à horizon hebdomadaire demeure tout à fait acceptable, alors même qu'aucune adaptation n'a été opérée sur le prédicteur. En effet, le MAPE moyen sur l'année pour la prévisions des prochaines 336 points demi horaires (une semaine) est de 3,35%, qui peut être favorablement comparé à bien des alternatives paramétriques nécessitant une adaptation bien plus lourde.

4.2 Sensibilité à la taille du bloc témoin

Fixons maintenant la taille du segment à prévoir à $\delta_F = 1$ jour et faisons varier la taille du segment de référence, encore appelé bloc témoin, soit la taille des segments du passé δ_P.

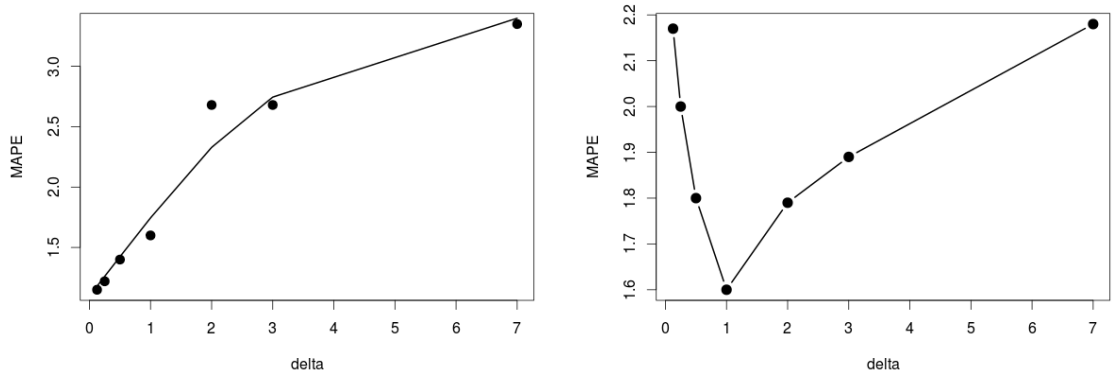

FIGURE 4 – MAPE en fonction des segments futurs δ_F (à droite) et en fonction de la taille des segments passés δ_P (à gauche) .

La qualité de prévision sur une année à l'horizon journalier avec des valeurs de δ_P allant de 3 heures jusqu'à une semaine est présentée dans la Figure 4 à gauche. Le choix de δ_P égal à un jour est a posteriori "justifié" par ce graphique, car cette valeur est associée à la meilleure performance en prévision. Dans ce type de données, très fortement reliée à l'activité socio-économique globale, la structure journalière est si importante que ne pas l'incorporer au modèle conduit nécessairement une dégradation notable de la performance globale mais néanmoins cela peut être utile pour la prévision des jours rares, c'est-à-dire aux voisins similaires peu nombreux dans l'historique.

4.3 Lien entre taille du bloc de référence et l'horizon de prévision.

Faisons varier δ_P et δ_F en même temps. Pour chaque combinaison de ces deux paramètres, nous obtenons des prévisions tout au long d'une année et mesurons leur qualité à l'aide du MAPE. Les résultats sont reportés dans le Tableau 1. La ligne et la colonne centrales contiennent les résultats commentés dans les deux paragraphes précédents.

Les colonnes présentent toutes le même comportement : dès que l'on augmente la taille de δ_F pour obtenir des prévisions à des horizons de plus en plus lointains, la qualité se dégrade. Le comportement des lignes diffère si l'on est en dessous de la taille journalière pour le segment de référence ou au dessus. Pour les prévisions à horizon infra journalier ($\delta_F < 1$ jour), le choix à privilégier pour le segment de référence est $\delta_P = 1$. En revanche, si les prévisions sont à horizon de plusieurs jours, le choix de δ_P doit être adaptée à la taille du segment à prévoir δ_F.

δ_F δ_P	1/8	1/4	1/2	1	2	3	...	7
1/8	1,88	1,92	1,68	1,15	1,18	1,31		1,69
1/4	1,92	1,97	1,72	1,22	1,33	1,43		1,77
1/2	2,00	1,96	1,74	1,40	1,55	1,67		2,04
1	2,17	2,00	1,80	1,60	1,79	1,89	...	2,18
2	2,88	2,79	2,81	2,68	2,46	2,37		2,54
3	2,84	2,72	2,76	2,68	2,76	2,54	...	2,67
...				...				...
7	3,46	3,39	3,41	3,35	3,35	3,21	...	3,15

Tableau 1 – MAPE par taille du segment témoin (δ_P) et horizon de prévision (δ_F) exprimées en jours.

5 Prévision de la consommation d'électricité en univers concurrentiel

Nous nous livrons dans cette section à la prévision d'une courbe de charge en présence de pertes de clients. Dans un premier temps, nous utilisons l'analyse multirésolution des ondelettes afin de décrire les données que nous utilisons. Ensuite, nous appliquons directement le modèle KWF sur ces courbes et vérifions qu'il conserve tout son intérêt.

5.1 Description des données à l'aide des ondelettes

Les données de consommation en présence de pertes de clients ont été fournies par EDF et elles sont différentes des données de demande nationale considérées jusqu'ici par deux aspects : elles ne couvrent plus la demande nationale mais seulement une partie de celle-ci et, d'autre part, on a retiré la consommation de quelques clients pour construire des trajectoires de la demande nationale en présence de pertes de clients.

Sur une période de deux ans et demi, nous disposons de quatre trajectoires de consommation avec pertes, que nous appelons `perteA`, `perteB`, `perteC` et `perteD` en ordre croissant d'intensité, associées à des scénarios détaillés plus bas, et la consommation sans pertes sur la même période (`Effect`). Le pas d'échantillonnage des données est toujours la demi-heure.

Les caractéristiques de ces nouvelles données sont similaires à celles des données nationales décrites dans la section 2.1. On y retrouve une tendance de long terme, une évidente saisonnalité annuelle en plus d'un motif hebdomadaire et une forte dépendance aux conditions climatiques. Cependant quelques différences existent, comme nous pouvons l'apercevoir dans Figure 5 qui représente une semaine de la courbe `Effect` qui a été normalisée pour des raisons de confidentialité. En effet, les nouvelles courbes de charge semblent être plus irrégulières et elles présentent une allure presque constante et de niveau très bas pendant les jours du week-end. On peut aussi remarquer un saut significatif de la consommation tous les jours à partir de 22 heures induit par le contrat

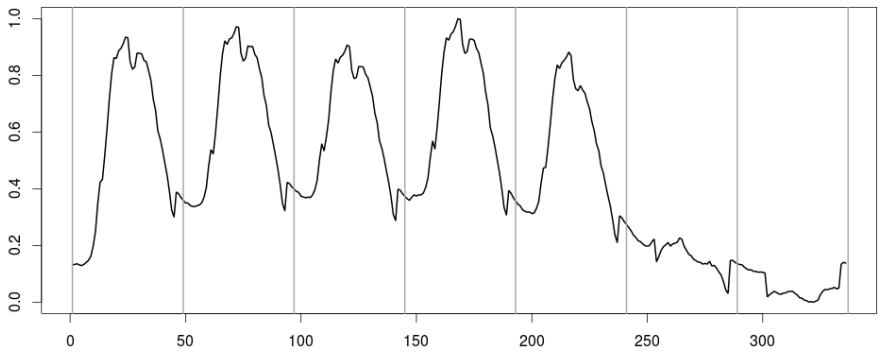

FIGURE 5 – Une semaine de la courbe `Effect`, les barres verticales séparent les jours.

tarifaire de ces clients.

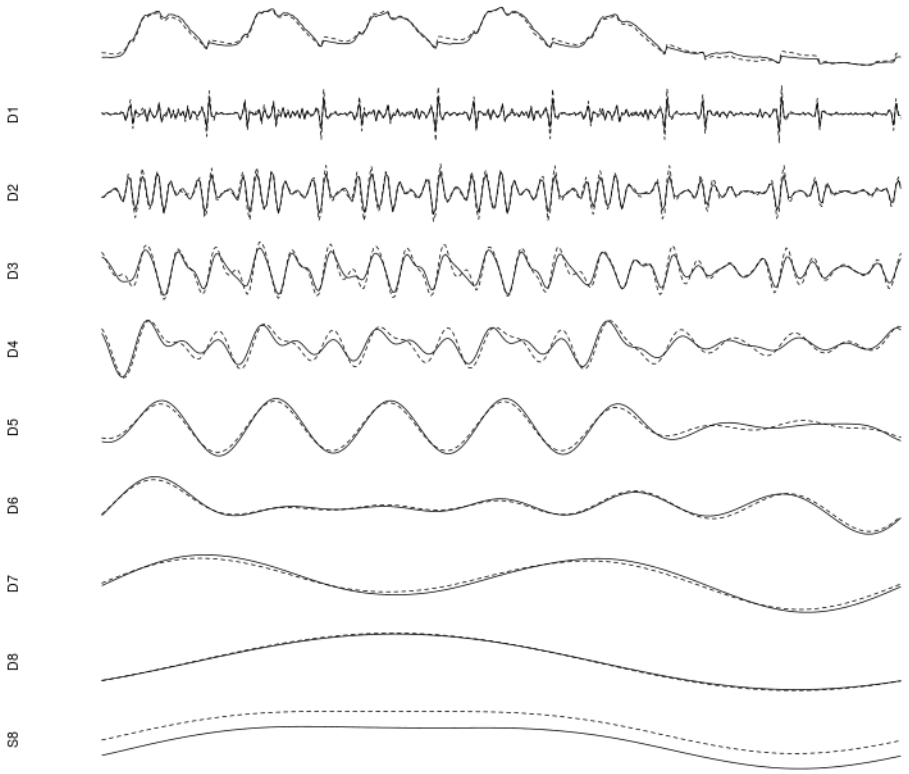

FIGURE 6 – Analyse multirésolution en ondelettes de la courbe `Effect` (trait continu) et de la courbe `perteD` (trait pointillé) sur une semaine.

Étant donnée que notre méthode de prévision utilise une distance basée sur les ondelettes, il est pertinent de comparer les nouvelles données dans le domaines des ondelettes. La Figure 6 contient l'analyse multirésolution en ondelettes de la courbe associée aux pertes les plus intenses `perteD`, et la courbe sans pertes `Effect`. Les deux

analyses sont très similaires pour les différents niveaux de détails mais aussi pour la forme de l'approximation (dernière ligne). Le fait que les deux signaux présentent autant de ressemblances dans le domaine en ondelettes est remarquable. Si les distributions des coefficients des détails des courbes de demande avec et sans pertes sont similaires, la méthode de prévision pourra utiliser les données historiques, i.e. observées sans pertes, pour fournir de prévisions dans le régime actuel, i.e. avec pertes. Il n'y aura donc aucun réglage nouveau à faire et aucune rupture dans le nombre de voisins utiles dans la passé pour construire les prévisions.

5.2 Construction des scénarios de pertes

Nous utilisons les courbes de pertes simulées pour construire deux scénarios pour les pertes de clients. Puisque nous ne disposons que d'un historique de deux années et demi, nous allons prévoir la dernière année en utilisant des configurations de pertes différentes sur la première année et demie. Deux configurations sont retenues pour faire l'analyse :

Perte brusque Nous créons un départ de clients soudain et brusque : tous les clients qui partent le font au même moment et la perte est d'intensité constante. Nous prenons alors la courbe `Effect` pendant les trois premiers semestres. Puis, au moment même que nous commençons à évaluer la performance de prévision, nous passons à la courbe `perteD`.

Perte graduelle Nous utilisons toutes les courbes pour créer un scénario des pertes graduelles et d'intensité croissante de clients. Les départs arrivent à chaque début de semestre, comme l'indique la dernière colonne de la table suivante, autant sur les semestres de calibration de la méthode ainsi que les semestres de validation.

Semestre	Perte brusque	Perte graduelle
1	`Effect`	`Effect`
2	`Effect`	`perteA`
3	`Effect`	`perteB`
4	`perteD`	`perteC`
5	`perteD`	`perteD`

5.3 Expériences numériques

Pour étudier l'impact des scénarios de perte de clients sur la méthode de prévision KWF, il faut une référence pour la performance prédictive de chaque scénario. Pour la construire, évaluons la performance prédictive de chacune des cinq courbes considérées (`Effect` et les 4 courbes de perte) pour la méthode KWF. La mesure de qualité de prévision que nous utilisons est la moyenne des MAPE à horizon de prévision journalier. Les MAPE pour chaque courbe sont disponibles dans la Table 2. La qualité de prévision pour chaque courbe est similaire sur la moyenne de l'année quand la même courbe est utilisée dans les périodes de calibration et de validation. Il faut noter que par rapport aux niveaux d'erreur observés sur les courbes de consommation nationales (MAPE

d'environ 1.5%), les MAPE de ces courbes sont sensiblement supérieurs, le problème de prévision est donc plus difficile. La moindre régularité de ces courbes et la structure très particulière des jours de week-end expliquent cette plus grande difficulté.

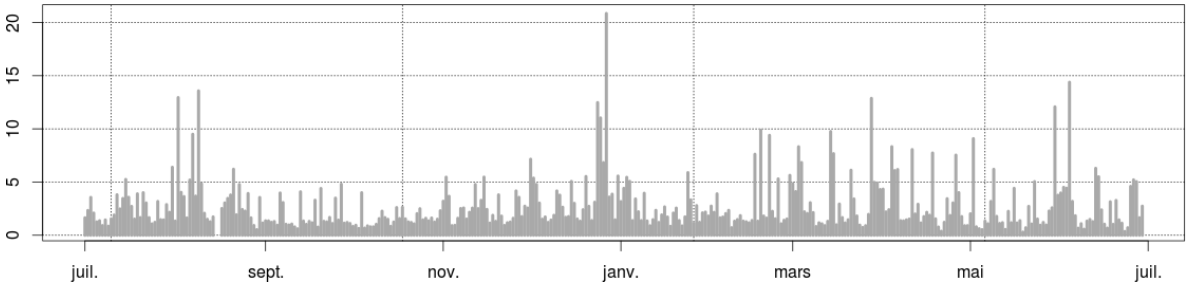

FIGURE 7 – MAPE journalière pour la courbe sans pertes (`Effect`).

Courbe	MAPE	Courbe	MAPE
`Effect`	2.80 %	.	.
`perteA`	2.81 %	`perteC`	2.79 %
`perteB`	2.79 %	`perteD`	2.81 %

TABLEAU 2 – MAPE moyenne sur la dernière année pour chaque courbe de consommation d'électricité avec et sans pertes de clients.

Scénario	MAPE
Sans pertes	2.80 %
Perte brusque	3.06 %
Pertes graduelles	2.89 %

TABLEAU 3 – MAPE moyenne sur la dernière année pour les scénarios de pertes de clients.

Ensuite, nous considérons les prévisions journalières pour les scénarios correspondant aux deux configurations de perte de clients, brusque ou graduelle. Les résultats de performance globale, mesurés par l'erreur moyenne sur toute la dernière année, sont présentés dans la Table 3. Les deux configurations de pertes engendrent des dégradations sur l'erreur moyenne annuelle, avec un plus large impact pour le scénario de perte brusque. Afin d'étudier la manière dont ces dégradations se produisent, nous présentons dans la Figure 8 pour chaque configuration et chaque jour de l'année la différence entre son MAPE et celui du même jour sur la courbe `Effect`. Nous avons rajouté à chaque graphique de la figure l'erreur moyenne sur les jours du passé. Remarquons que les valeurs négatives montrent une dégradation de la qualité de prévision de la situation avec pertes par rapport à la situation sans pertes. L'impact de la perte brusque de clients (graphique de gauche) au début de la période d'analyse est mise en évidence

par une grande dégradation qui représente 6 fois le MAPE de ce jour là pour la courbe `Effect`. Néanmoins, cette dégradation diminue de manière significative au fur et mesure que nous avançons dans l'année test. En revanche, dans le cas d'une perte graduelle de clients (graphique à droite) il n'y a pas d'impact ponctuel significatif.

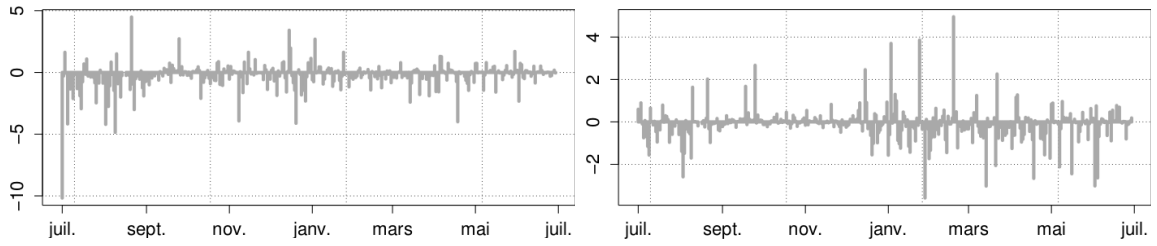

FIGURE 8 – Différence journalière entre MAPE de la courbe sans pertes (`Effect`) et pour les deux configurations de perte de clients : perte brusque (à gauche) et perte graduelle (à droite).

La Figure 9 rassemble les deux courbes de dégradation moyenne des graphiques précédents mais elles sont exprimées comme une proportion du MAPE de la courbe `Effect`. Les deux évolutions sont convergentes vers un niveau de perte moyen de 15% pour la configuration de perte brusque de clients et de 7% pour la perte graduelle.

Ainsi donc la méthode réussit à s'adapter au nouveau contexte dans les deux configurations de perte de clients. L'historique s'enrichit rapidement de plus en plus de courbes qui appartiennent au nouveau contexte et la méthode peut alors trouver relativement vite un plus grand nombre de courbes similaires ce qui permet d'améliorer progressivement la qualité des prévisions. Une partie de cette réussite peut être expliquée par le fait que les représentations en ondelettes des courbes sans et avec pertes (voir Figure 6) sont semblables.

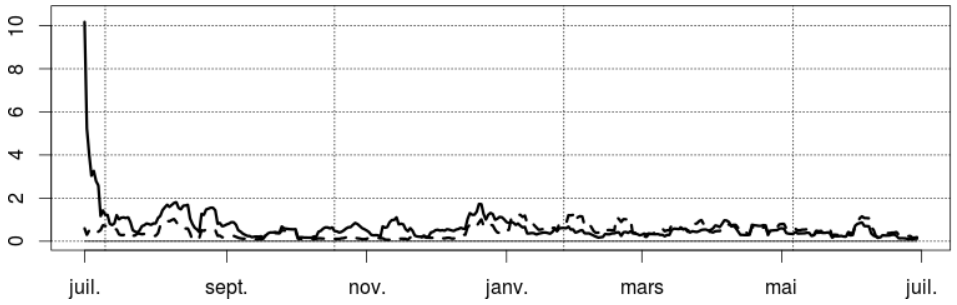

FIGURE 9 – Dégradation moyenne du MAPE pour les deux configurations de perte de clients retenues, exprimée comme pourcentage du MAPE journalier sur la courbe sans pertes `Effect`.

L'évolution du paramètre de largeur de fenêtre apporte une lecture complémentaire. Ce paramètre est adapté au fur et à mesure du temps et contrôle le nombre de voisins considérés, c'est-à-dire le nombre de jours du passé entrant dans la prévision avec un

poids notable. La Figure 10 présente l'évolution de ce paramètre pour la courbe sans perte de clients, qui est notre référence, et celles correspondants aux deux configurations de perte de clients. Les largeurs de fenêtre ont été lissées en utilisant une moyenne pondérée glissante de largeur 7 afin d'éliminer la variabilité due au calendrier. Nous pouvons constater que dans les premières semaines de juillet les évolutions sont très similaires. Il n'y a pas eu encore assez de données dans l'historique pour induire un changement dans le comportement de la fenêtre. Ceci change à partir du mois d'août ou la largeur de fenêtre pour la configuration de perte brusque (en trait noir continu) augmente considérablement par rapport aux deux autres courbes. Ce comportement, qui se constate aussi à d'autres moments de l'année, peut s'expliquer par la difficulté de la méthode à trouver des segments similaires. Afin de réduire la variance du prédicteur, la fenêtre est ouverte en détriment d'un biais plus important tout en conduisant à réduire l'erreur quadratique. Des comportements inattendus apparaissent aussi, comme celui observé pendant le mois d'octobre : l'effacement de certains clients facilite la tâche du prédicteur car plus de segments sont similaires, ce qui conduit à une réduction de la largeur de fenêtre.

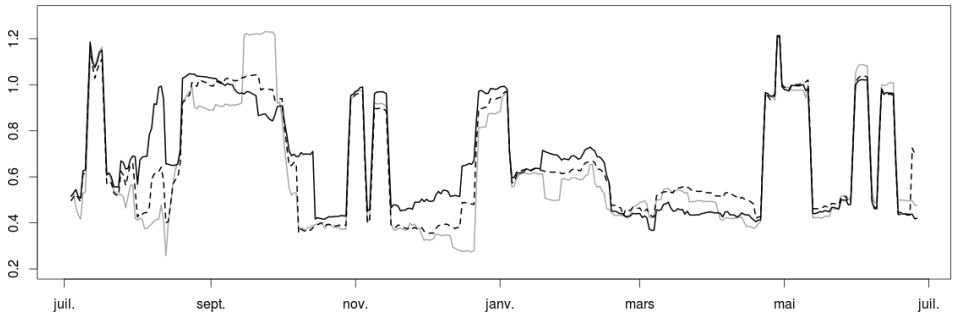

FIGURE 10 – Évolution (lissée) de la largeur de fenêtre pour la courbe `Effect` (en gris), la courbe de perte brusque (en trait noir continu) et la courbe de pertes graduelles (en pointillés noir).

6 Conclusion

Dans cet article, qui fait suite à (Antoniadis *et al.* (2012)), nous avons exploré, trois aspects du modèle KWF, qui est une approche pour la prévision de séries chronologiques de fonctions en présence de non stationnarités. Après avoir complété la prévision ponctuelle par la construction d'un intervalle de confiance pour le prédicteur fonctionnel dans la cas non stationnaire nous examinons la flexibilité et la simplicité du modèle pour fournir, sans effort supplémentaire, des prévisions de bonne qualité en faisant varier trois aspects. Le premier consiste à examiner la prévision à des horizons de plus en plus éloignés, sans rien modifier au prédicteur, la qualité se dégrade comme attendu mais reste très concurrentielle et serait en outre assez facilement améliorable. Un autre point est l'étude conjointe de l'horizon et de la taille du bloc témoin, résumant le contexte

présent. Enfin, le dernier point est très étroitement lié au contexte applicatif spécifique, puisque nous avons constaté la capacité de la méthode à fournir de prévisions de bonne qualité en présence d'instationnarités subtiles du signal conséquences de pertes de clients selon divers scenarios.

Références

[1] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. application á la consommation d'électricité. *Journal de la Société Française de Statistique*, 153(2) :52 – 78, 2012.

[2] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(5) :837–857, 2006.

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

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

[5] P. Kokoszka. Dependent functional data. *ISRN Probability and Statistics*, 2012(Article ID 958254) :30 pages, 2012.

[6] J.-M. Poggi. Prévision non-paramétrique de la consommation électrique. *Revue de Statistique Appliquée*, xlii :93 – 98, 1994.

[7] H.-L. Shang. Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics*, 40(1) :152–168, 2013.