

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

Title: A functional approach to nonparametric forecasting of electricity consumption

Anestis Antoniadis¹, Xavier Brossat², Jairo Cugliari³ et Jean-Michel Poggi⁴

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 basée 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.

Abstract: In a recent work (Antoniadis *et al.* (2012)), the authors proposed a prediction model for functional time series in the presence of nonstationarities. This model has been applied to the electricity demand of Electricité de France (EDF). The general principle of the forecasting model is to find in the past similar situations to the present and linearly combine their futures to build the forecast. The concept of similarity is based on wavelets and several strategies are implemented to take into account the various sources of non stationarity.

We explore in this second article, three aspects of this model that complement the original methodology while highlighting the industrial usefulness of the method. First we discuss the construction of a confidence interval for the predictor function. Next, we examine the flexibility and simplicity of the model to provide, without extra effort, forecasts horizons further and further away. Finally, in the specific application context, we examine the ability of the method to provide good predictions in the presence of subtle signal nonstationarities induced by loss of customers coming from various scenarios.

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

Keywords: Nonparametric forecasting, Functional data, Nonstationarity, Electricity load curve

Classification AMS 2000 : 62G08, 62M20, 62P30

¹ Université Joseph Fourier, Laboratoire LJK, Tour IRMA, BP53, 38041 Grenoble Cedex 9, France

E-mail : anestis.antoniadis@imag.fr

² EDF R&D, 1 avenue du Général de Gaulle, 92141 Clamart Cedex, France

E-mail : xavier.brossat@edf.fr

³ Inria Select, Université Paris Sud, Bât. 425, 91405 Orsay Cedex, France

E-mail : Jairo.Cugliari@math.u-psud.fr

⁴ Université Paris Descartes, Université Paris Sud, Bât. 425, 91405 Orsay Cedex, France

E-mail : Jean-Michel.Poggi@paris-descartes.fr

1. Introduction

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

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 à temps discret $Z = (Z_k(t), k = 0, \dots, n; t \in [0, \delta])$ 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} à partir des données Z_1, \dots, 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, \dots, 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 au 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 dernier modèle est adapté et étendu au cas des processus fonctionnels non stationnaires (Cugliari (2011)).

Ainsi, on obtient une qualité de prévision qui est similaire à celles obtenues par 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 industrielle 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 station-

naire 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 (Antoniadis *et al.* (2012)), 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 δ , la taille du bloc de référence (ou témoin), et l'horizon de prévision h 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'une durée inférieure à un jour.

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. Nous y décrivons les données de consommation d'électricité et rappelons le prédicteur appelé KWF (*Kernel Wavelet Functional*) ainsi que les correctifs introduits pour prévoir une série fonctionnelle non stationnaire.

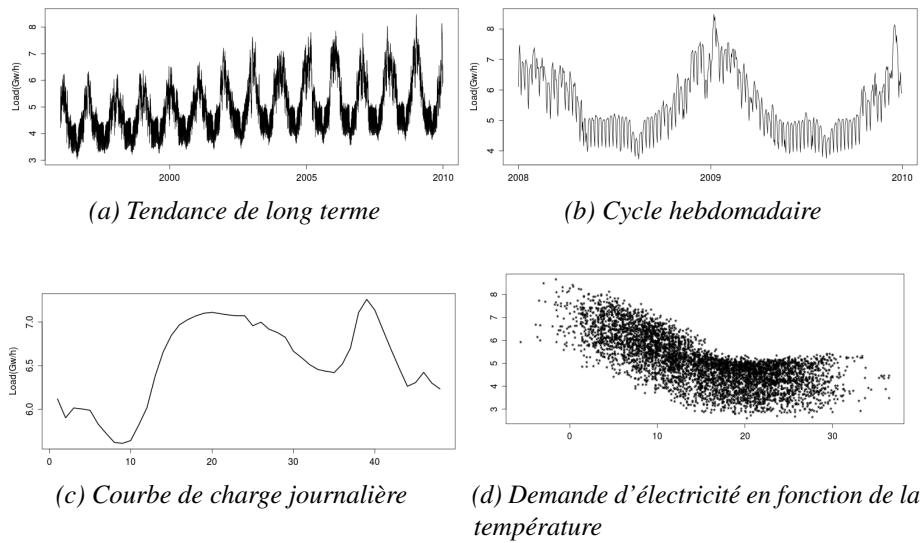


FIGURE 1: Caractéristiques principales de la consommation d'électricité française.

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 chaque 30 minutes. 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.

Pour modéliser ces données, EDF utilise des 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 $Z = (Z_k)_{k \in \mathbb{Z}}$ à valeurs dans l'espace fonctionnel $H = L_2([0, 1])$. Nous disposons d'un échantillon de n courbes Z_1, \dots, Z_n et l'objectif est de prévoir

Z_{n+1} . Pour prendre en compte la dimension infinie des objets à traiter, le modèle KWF représente chaque segment $Z_i, i = 1, \dots, n$, par son développement sur une base d'ondelettes,

$$Z_i(t) = \sum_k c_{j_0,k}^{(i)} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k d_{j,k}^{(i)} \psi_{j,k}(t) \quad (1)$$

$$= \mathcal{S}_i(t) + \mathcal{D}_i(t). \quad (2)$$

Le premier terme de chaque courbe $Z_i(t)$ 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 $\mathcal{D}_i(t)$ contient l'information de la structure locale de la fonction.

En pratique, nous n'observons les courbes $Z_i(t)$ qu'à travers une grille finie $\{t_1, \dots, t_N\}$ de l'intervalle $[0, 1]$. Alors, pour prendre en compte la dimension infinie des segments Z_i , nous tronquons le développement en ondelettes à une échelle $J > j_0$. De plus, nous travaillons avec le choix $j_0 = 0$ ce qui donne l'approximation la plus grossière possible et garde toute l'information de l'allure de la courbe dans $\mathcal{D}_i(t)$.

Avec ce découpage, la prévision pour le segment $n+1$ s'écrit $\widehat{Z}_{n+1}(t) = \widehat{\mathcal{S}}_{n+1}(t) + \widehat{\mathcal{D}}_{n+1}(t)$. La méthode de prévision se décompose alors en deux phases. D'abord, trouver parmi les blocs centrés du passé ceux qui sont le plus semblables au dernier bloc centré observé. Ensuite construire un vecteur de poids $w_{n,m}, m = 1, \dots, n-1$ et obtenir la prévision souhaitée en moyennant les futurs des blocs correspondant aux indices $2, \dots, n$ respectivement. Si les courbes Z_i présentent des niveaux moyens très différents il est utile de centrer les courbes avant de calculer la prévision (centrée), puis remettre à niveau la prévision lors de la seconde phase.

Le processus fonctionnel $\mathcal{D}_{n+1}(t)$ étant supposé centré, nous le prévoyons par

$$\widehat{\mathcal{D}}_{n+1}(t) = \sum_{m=1}^{n-1} w_{n,m} \mathcal{D}_{m+1}(t), \quad (3)$$

où les poids $w_{n,m}$ sont donnés par la suite dans l'équation (7). Ensuite, pour la prévision de $\mathcal{S}_{n+1}(t)$ nous utilisons

$$\widehat{\mathcal{S}}_{n+1}(t) = \mathcal{S}_n(t) + \sum_{m=1}^{n-1} w_{n,m} \Delta(\mathcal{S}_{m+1})(t), \quad (4)$$

où $\Delta(\mathcal{S}_{m+1})(t) = \mathcal{S}_{m+1}(t) - \mathcal{S}_m(t)$. Dans le cas où Z est un processus fonctionnel stationnaire, la moyenne pondérée dans (4) est négligeable et alors le prédicteur de Z_{n+1} s'écrit

$$\widehat{Z}_{n+1}(t) = \sum_{m=1}^{n-1} w_{n,m} Z_{m+1}(t). \quad (5)$$

La construction du vecteur des poids $w_{n,p}$ peut se faire en deux phases.

Première phase. Pour deux segments observés $Z_i(t)$ et $Z_{i'}(t)$, nous utilisons la dissimilarité D définie comme

$$D(Z_i, Z_{i'}) = \sum_{j=1}^J 2^{-j} \sum_{k=0}^{2^j-1} (d_{j,k}^{(i)} - d_{j,k}^{(i')})^2. \quad (6)$$

Puisque les coefficients d'approximation ne sont pas pris en compte dans D , cette dissimilarité permet de trouver de motifs similaires entre courbes même si elles ont des approximations très différentes.

Deuxième phase. Notons $\Xi_i = \{c_{J,k}^{(i)} : k = 0, 1, \dots, 2^J - 1\}$ 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 $K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n)$, K est un noyau de probabilité et h_n la largeur de fenêtre associé à la régression à noyau.

Finalement, nous pouvons appliquer la transformée inverse de la DWT sur $\widehat{\Xi}_{n+1}$ pour obtenir la prévision de la courbe \mathcal{D}_{n+1} dans le domaine temporel. Si l'on note

$$w_{n,m} = \frac{K_{h_n}(D(Z_{n,J}, Z_{m,J}))}{\sum_{m=1}^{n-1} K_{h_n}(D(Z_{n,J}, Z_{m,J}))}, \tag{7}$$

ces poids permettent de réécrire le prédicteur comme un barycentre des futurs des segments du passé (voir (3)).

Même si la démarche présentée permet de prévoir le processus fonctionnel Z en présence d'une évolution de sa partie approximation, d'autres types de non stationnarités sont observées en pratique. 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.

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 n et m au même groupe :

$$\widetilde{w}_{n,m} = \frac{w_{n,m} \mathbf{1}_{\{gr(m)=gr(n)\}}}{\sum_{m=1}^n w_{n,m} \mathbf{1}_{\{gr(m)=gr(n)\}}}, \tag{8}$$

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 $t_i, i = 1, \dots, N$, nous avons

$$P(L_{n+1,\alpha}(t_i) \leq Z_{n+1}(t_i) \leq U_{n+1,\alpha}(t_i)) \geq 1 - 2\alpha, \quad 0 < \alpha < 1.$$

Les quantités doivent être estimées à partir de l'historique Z_1, \dots, 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, \dots, Z_n . Cette distribution est à la base même du modèle KWF car le prédicteur (5) est un estimateur de l'espérance conditionnelle $\mathbb{E}[Z_{n+1}|Z_1, \dots, Z_n]$. 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 $w_{n,m}, m = 1, \dots, n - 1$ induisent une loi de probabilité discrète sur les observations Z_2, \dots, 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 $\hat{\sigma}_{n+1|n}^2(t_i) = \sum_{m=1}^{n-1} w_{n,m} (Z_{m+1}(t_i) - \hat{Z}_{n+1|n}(t_i))^2$. Puis, l'intervalle est obtenu comme

$$\hat{Z}_{n+1|n}(t_i) \pm k \hat{\sigma}_{n+1|n}^2(t_i),$$

avec k une constante égale à 2 ou 3 par exemple.

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 bootstrap à la place de la distribution empirique des données.

La procédure peut se résumer de la façon suivante :

1. Obtenir la prévision ponctuelle \hat{Z}_{n+1} par (5).
2. Générer B pseudo réalisations bootstrap $Z_{n+1}^{(1)}, \dots, Z_{n+1}^{(B)}$ par tirage aléatoires des segments du passé selon la loi induite par le vecteur de poids $w_{n,m}, m = 1, \dots, n - 1$. Alors, pour chaque $b \in \{1, \dots, B\}$ nous avons

$$P(Z_{n+1}^{(b)} = Z_{m+1} | Z_n) = w_{n,m}.$$

3. Obtenir les pseudo résidus $\hat{R}_{n+1}^{(b)}(t_i) = Z_{n+1}^{(b)}(t_i) - \hat{Z}_{n+1}(t_i), b = 1, \dots, B$.
4. Pour chaque t_i , définir $\hat{R}_{n+1,\alpha}(t_i)$ et $\hat{R}_{n+1,1-\alpha}(t_i)$ les quantiles d'ordre α et $1 - \alpha$ de $\{\hat{R}_{n+1}^{(b)}(t_i); b = 1, \dots, B\}$.

Enfin l'intervalle à $(1 - \alpha) \times 100\%$ est donné par $[\hat{L}_{n+1,\alpha}(t_i), \hat{U}_{n+1,\alpha}(t_i)]$ où

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

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 (8) à la place de (7) pour générer les pseudo réalisations $Z_{n+1,b}^*$, $b = 1, \dots, B$. Il ne reste qu'à établir les groupes, qui dans le cas que nous étudions sont déduits 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 (5) (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 (6) et en conséquence la partie approximation de Z_i est codée par un seul coefficient d'échelle, $c_{0,0}^i = c_0^i$. Nous quantifions la variabilité associée à $\Delta(\mathcal{S}_{m+1})$ par l'estimation des quantiles empiriques de $\{\Delta(c_0^{m+1}), m = 2, \dots, n\}$. 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 (4) et (3) respectivement avec les poids (8).
2. Générer B pseudo réalisations bootstrap $Z_{n+1}^{(1)}, \dots, Z_{n+1}^{(B)}$ par tirage aléatoires des segments du passé selon la loi induite par le vecteur de poids $\widetilde{w}_{n,m}, m = 1, \dots, n - 1$ (cf. équation (8)). Alors, pour chaque $b \in \{1, \dots, B\}$ nous avons

$$P(Z_{n+1}^{(b)} = Z_{m+1} | Z_n) = \widetilde{w}_{n,m}.$$

A l'image de l'équation (1), chaque $Z_{n+1}^{(b)}$ est décomposé dans une partie approximation et une partie détail que nous notons $\mathcal{S}_{m+1}^{(b)}$ et $\mathcal{D}_{m+1}^{(b)}$ respectivement.

3. Pour chaque $b \in \{1, \dots, B\}$, on calcule le pseudo résidu lié à la partie détails s'obtient

$$\widehat{R}_{n+1}^{(b)}(t) = Z_{n+1}^{(b)}(t) - \widehat{Z}_{n+1}(t),$$

et le pseudo résidu lié à la partie approximation

$$\widehat{q}_{n+1}^{(b)} = \Delta c_0^{(b)} - \Delta(c_0^{n+1}),$$

où la prévision du coefficient d'approximation s'écrit $\widehat{c}_0^{n+1} = c_0^n + \sum_{m=1}^{n-1} \widetilde{w}_{m,n} \Delta(c_0^{m+1})$.

4. Définir $\widehat{q}_{n+1,\alpha}$ et $\widehat{q}_{n+1,1-\alpha}$ comme les quantiles empiriques associés à la partie approximation de niveau α et $1 - \alpha$ de $\{\widehat{q}_{n+1}^{(b)}, b = 1, \dots, B\}$. Ensuite, nous appliquons l'inverse de la transformée en ondelettes sur les coefficients d'échelles estimés et nous appelons le résultat

$$\widehat{Q}_{n+1,\alpha}(t) = (c_0^n + \widehat{q}_{n+1,\alpha})\phi_{0,0}(t) \quad \text{et} \quad \widehat{Q}_{n+1,1-\alpha}(t) = (c_0^n + \widehat{q}_{n+1,1-\alpha})\phi_{0,0}(t).$$

De manière analogue, nous définissons les quantiles au niveau α et $1 - \alpha$ de $\{\widehat{R}_{n+1}^{(b)}(t_i); b = 1, \dots, B\}$ comme étant les quantiles empiriques $\widehat{R}_{n+1,\alpha}(t_i)$ et $\widehat{R}_{n+1,1-\alpha}(t_i)$ de la partie détails pour chaque t_i .

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

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

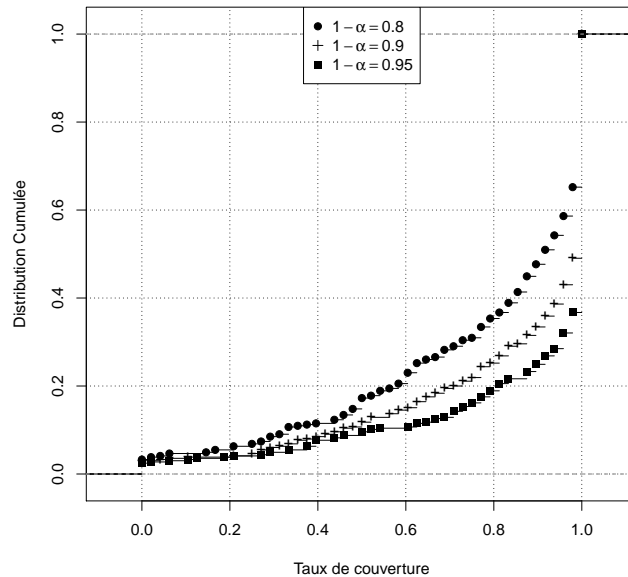


FIGURE 2: Distribution cumulée du taux de couverture pour $1 - \alpha = 0.8, 0.9$ et 0.95 .

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 et nous fixons le nombre de répliques $B = 100$. 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 $\{\hat{L}_{n+1,\alpha}(t_i), \hat{U}_{n+1,\alpha}(t_i)\}$,

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

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

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 un taux de couverture de moins de 50%, représentent seulement 10%, 12% et 17% des jours respectivement pour les niveaux de confiance 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

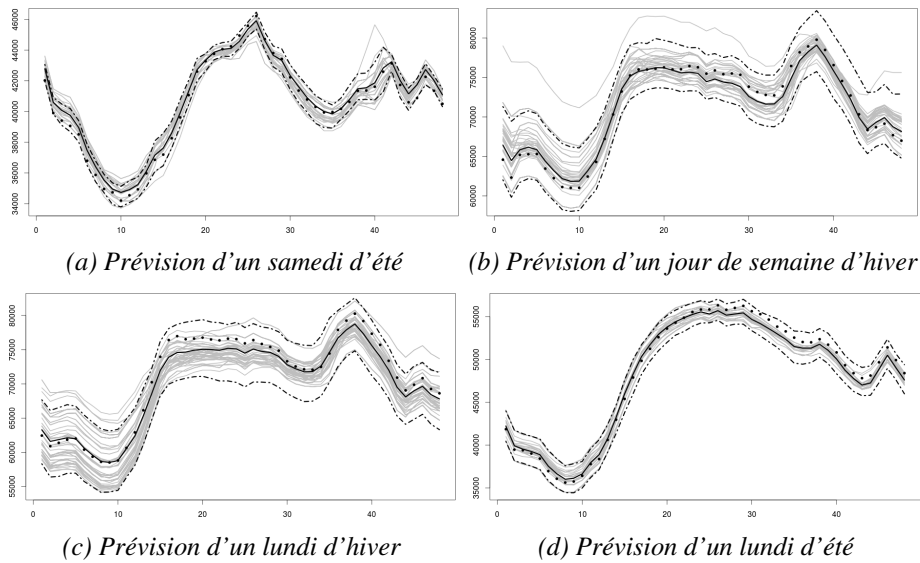


FIGURE 3: *Intervals de confiance pour quatre jours de l'année.*

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. 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 $Z_m(t), m = 1, \dots, n-1$. Puis, les courbes futures $Z_{m+1}(t), m = 2, \dots, n$ sont moyennées en utilisant les poids $w_{n,m}$ pour obtenir une prévision \hat{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 largeur 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 égales 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

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

où à l'image de l'équation (7), $w_{n,m} \propto K_{h_n}(D(Z_m^P, Z_n^P))$.

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)

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\widehat{Z}_{n+1|n}(t_i) - Z_{n+1}(t_i)|}{Z_{n+1}}$$

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 .

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

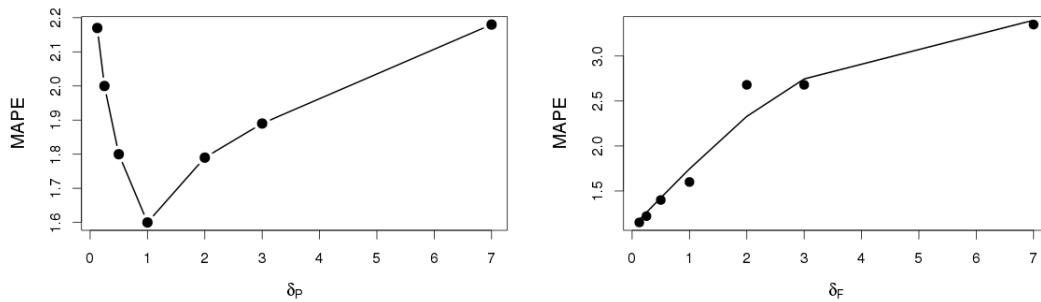


FIGURE 4: MAPE (en pourcentage) en fonction en fonction de la taille des segments passés δ_P (à gauche) et des segments futurs δ_F (à droite).

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 liée 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

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

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

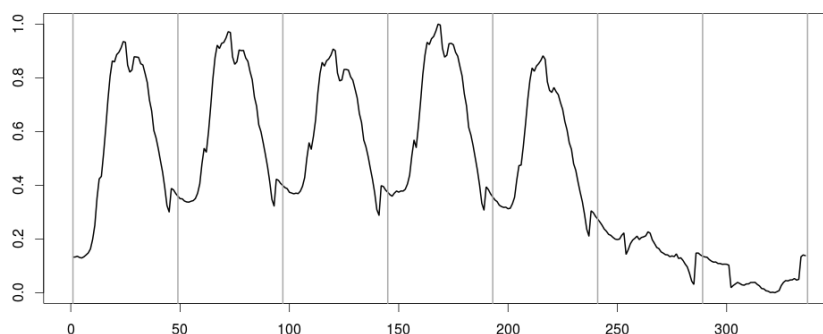


FIGURE 5: Une semaine de la courbe *Effect*, les barres verticales séparent les jours.

de plusieurs jours, le choix de δ_P doit être adaptée à la taille du segment à prévoir δ_F .

5. Prédiction de la consommation d'électricité en univers concurrentiel

Nous nous livrons dans cette section à la prédiction d'une courbe de charge en présence de pertes de clients. Les données que nous utiliserons sont créées à partir d'un signal de consommation différent de celui utilisé auparavant. Dans un premier temps, nous décrivons les données que nous utiliserons ainsi que les scénarios de pertes de clients qui ont été construits pour cette expérience. Ensuite, nous appliquons directement le modèle KWF sur ces données et vérifions qu'il conserve tout son intérêt. Pour finir nous illustrons les résultats grâce à l'analyse multirésolution des ondelettes et de la largeur de fenêtre.

5.1. Description des données

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, les consommations individuelles des clients sont disponibles. Ainsi, EDF peut construire des trajectoires de la demande en présence de pertes de clients si l'on retire la demande d'un sous-ensemble de consommateurs.

Plus précisément nous disposons de quatre trajectoires de consommation avec pertes, que nous appelons *perteA*, *perteB*, *perteC* et *perteD* en ordre croissant d'intensité de pertes et la consommation effective (*Effect*), *i.e.* sans pertes. Les données sont étalées sur une période de deux ans et demi et le pas d'échantillonnage 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* (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 tarifaire de ces clients.

5.2. Construction des scénarios de pertes

Nous combinons 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 obtenir des prévisions journalières sur 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	<i>Effect</i>	<i>Effect</i>
2	<i>Effect</i>	<i>perteA</i>
3	<i>Effect</i>	<i>perteB</i>
4	<i>perteD</i>	<i>perteC</i>
5	<i>perteD</i>	<i>perteD</i>

5.3. Expériences numériques

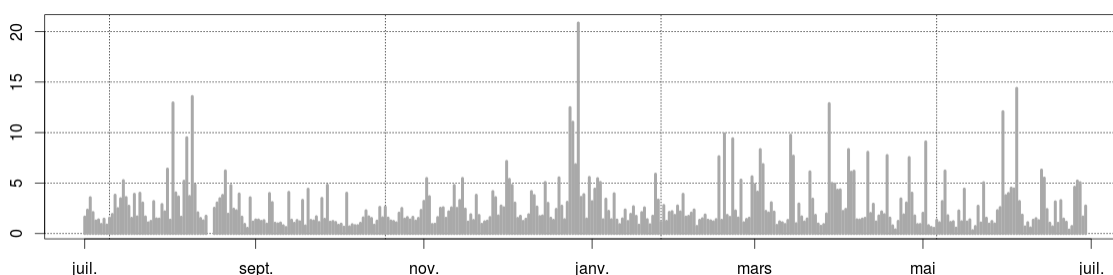
Afin étudier l'impact des scénarios de perte de clients sur la méthode de prévision KWF, il nous faut une référence pour la performance prédictive de chaque scénario. Pour la construire, nous évaluons la performance prédictive de chacune des cinq courbes considérées (*Effect* et les 4 courbes de perte) 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 le Tableau 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

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

Courbe	MAPE	Courbe	MAPE
<i>Effect</i>	2.80 %	.	.
<i>perteA</i>	2.81 %	<i>perteC</i>	2.79 %
<i>perteB</i>	2.79 %	<i>perteD</i>	2.81 %

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

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

FIGURE 6: MAPE journalière pour la courbe sans pertes (*Effect*).

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.

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 le Tableau 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 à la Figure 7 pour chaque configuration et chaque jour de l'année la différence entre son MAPE et celui du même jour sur la courbe *Effect*. 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.

La Figure 8 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.

5.4. Discussion

Nous avons vu que la méthode réussit à s'adapter au nouveau contexte de prévision 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

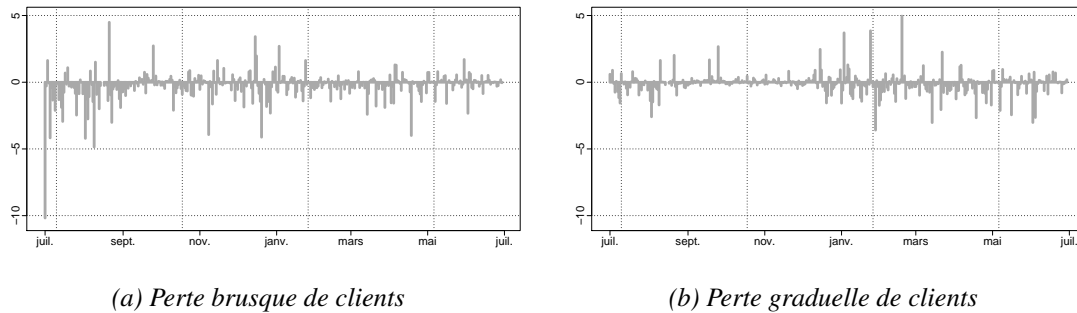


FIGURE 7: Différence journalière entre MAPE de la courbe sans pertes (*Effect*) et pour les deux configurations de perte de clients .

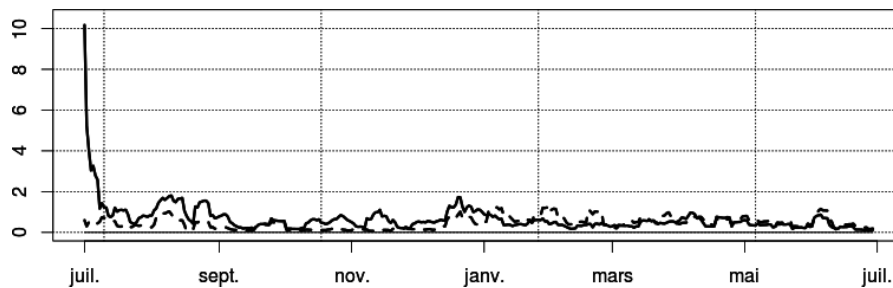


FIGURE 8: Dégradation moyenne du MAPE pour les deux configurations de perte de clients retenues, exprimée comme pourcentage du MAPE journalier sur la courbe sans pertes *Effect*.

grand nombre de courbes similaires ce qui permet d'améliorer progressivement la qualité des prévisions. Ce bon comportement peut être expliqué par la représentation en ondelettes utilisée dans la notion de dissimilarité de la méthode de prévision.

La Figure 9 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.

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

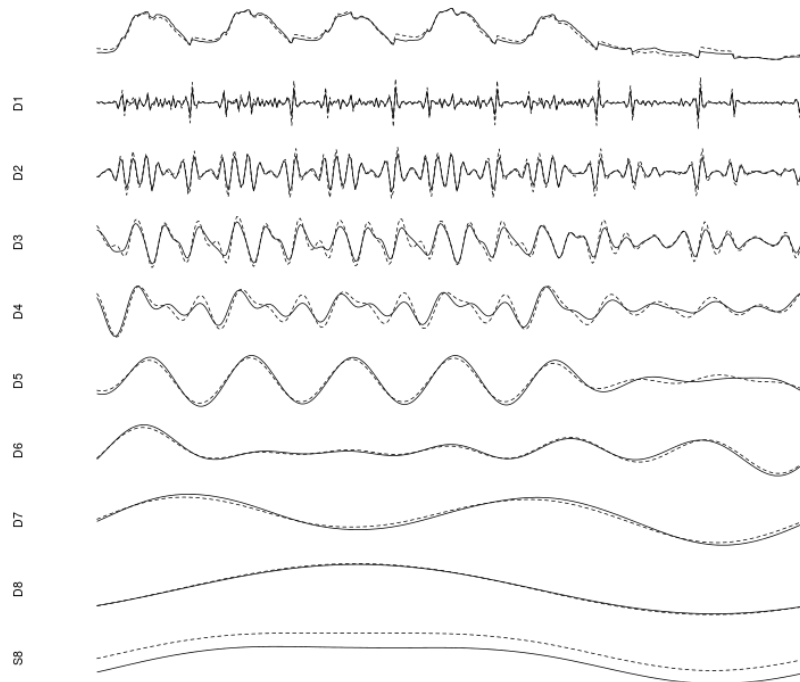


FIGURE 9: Analyse multirésolution en ondelettes de la courbe *Effect* (trait continu) et de la courbe *perteD* (trait pointillé) sur une semaine.

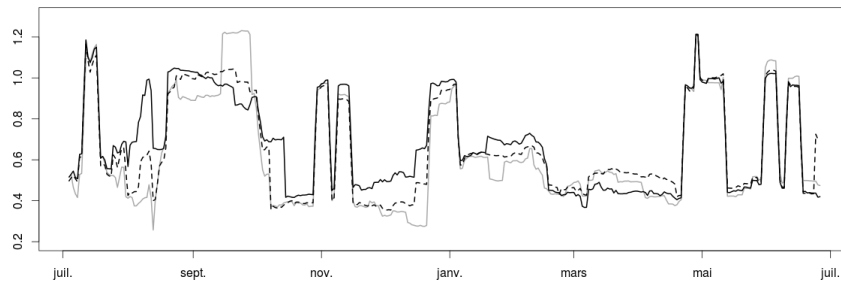


FIGURE 10: Évolution (lissée) de la largeur de fenêtre pour la courbe *Effect* (en gris), la courbe de perte brusque (en trait noir continu) et la courbe de pertes graduelles (en pointillés noirs).

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.

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 des prévisions de bonne qualité en présence d'instationnarités subtiles du signal conséquences de pertes de clients selon divers scénarios.

Remerciements

Les auteurs remercient les deux rapporteurs anonymes de leurs remarques et suggestions qui ont contribué à clarifier et améliorer la première version du manuscrit.

Références

- Antoniadis, A., Brossat, X., Cugliari, J., and Poggi, J.-M. (2012). Prévision d'un processus à valeurs fonctionnelles en présence de non stationnarités. application à la consommation d'électricité. *Journal de la Société Française de Statistique*, 153(2) :52 – 78.
- Antoniadis, A., Brossat, X., Cugliari, J., and Poggi, J. M. (2013). Clustering functional data using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(1).
- Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(5) :837–857.
- Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric functional estimation and related topics*, NATO ASI Series, pages 509–529. Springer.
- Cugliari, J. (2011). *Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité*. PhD thesis, Université Paris Sud.
- Kokoszka, P. (2012). Dependent functional data. *ISRN Probability and Statistics*, 2012(Article ID 958254) :30 pages.
- Poggi, J.-M. (1994). Prévision non-paramétrique de la consommation électrique. *Revue de Statistique Appliquée*, xlii :93 – 98.
- Shang, H.-L. (2013). Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics*, 40(1) :152–168.