

REVUE DE STATISTIQUE APPLIQUÉE

J. M. POGGI

Prévision non paramétrique de la consommation électrique

Revue de statistique appliquée, tome 42, n° 4 (1994), p. 83-98

http://www.numdam.org/item?id=RSA_1994__42_4_83_0

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

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

NUMDAM

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

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

PRÉVISION NON PARAMÉTRIQUE DE LA CONSOMMATION ÉLECTRIQUE¹

J.M. Poggi

Université Paris Sud-Orsay

Laboratoire de Modélisation Stochastique et Statistique URA 743

Bât. 425, Mathématiques, 91 405 Orsay cedex

(et Université Paris X - Nanterre)

RÉSUMÉ

Ce texte porte sur la prévision non paramétrique de la consommation électrique française demi-heure. On présente un prédicteur basé sur l'estimateur à noyau de la fonction d'autorégression et suggéré par un modèle autorégressif non linéaire. On dégage la notion de similarité, qui permet à la fois de mettre au point les paramètres et la forme du prédicteur, et *a posteriori* de valider et d'interpréter en termes électriques la qualité des prévisions.

Mots-clés : *Prévision non paramétrique, estimateur à noyau, statistique appliquée, courbe de charge électrique.*

SUMMARY

This paper deals with the problem of nonparametric prediction of the french electrical load consumption. The predictor based on the kernel estimation of the autoregression function is suggested by a nonlinear autoregressive model. The notion of similarity is used in order to tune predictor's parameters, and *a posteriori* to validate and interpret from an electrical viewpoint the goodness of fit.

Keywords : *Nonparametric prediction, kernel estimates, applied statistics, electrical load consumption.*

Introduction

Un intérêt croissant se manifeste pour l'étude et l'utilisation des modèles non linéaires, paramétriques ou non. On peut citer, entre autres, les modèles bilinéaire, ARCH, GARCH, AR exponentiel, AR à paramètres aléatoires, modèles à seuil, réseaux de neurones, AR fonctionnel. De très nombreux travaux portent sur ces modèles sous des angles variés : étude des propriétés probabilistes et statistiques

¹ Ce travail a été réalisé dans le cadre d'un contrat de recherche avec Electricité de France (Direction Etudes et Recherche, Clamart, France)

(en particulier critères de stabilité, de stationnarité et de mélange), des techniques d'identification, d'estimation et de test, des études (souvent comparatives) des performances des modèles en estimation et plus rarement, en prévision. On en trouvera des éléments ainsi que de nombreuses indications bibliographiques dans les documents de synthèse suivants : Ango Nze, Doukhan 93, Bosq, Lecoutre 92, Dufflo 90, Priestley 88 et Tong 90.

Cet article présente la mise au point d'un prédicteur non paramétrique de la consommation électrique française. Ce travail est issu d'une collaboration avec EDF (*cf.* Misiti *et al.* 93). Le prédicteur utilisé est basé sur l'estimateur à noyau de la fonction d'autorégression et est suggéré par un modèle autorégressif non linéaire. Dans ce contexte, des travaux, parmi lesquels on peut citer Carbon, Delecroix 87, Gannoun 90 ainsi que Yakowitz 85 et 87, mettent en concurrence des modèles paramétriques et non paramétriques. Ils montrent dans de nombreux cas la supériorité des seconds vis à vis des premiers.

Pour la courbe de charge électrique, en dehors de ses très bonnes performances, le prédicteur non paramétrique présente deux avantages décisifs : il admet une interprétation directe en termes électriques par le biais des similarités et il est simple et souple à mettre en oeuvre.

Ce texte est organisé comme suit. Les données et les problèmes font l'objet du premier paragraphe. Les méthodes sont développées dans le paragraphe 2. Le prédicteur à noyau est présenté et la notion de similarité est dégagée. Celle-ci permet à la fois de mettre au point les paramètres et la forme du prédicteur, et a posteriori de valider et d'interpréter en termes électriques la qualité des prévisions. Afin de prendre en compte l'instationnarité de la courbe de charge, deux indices de similarité sont proposés. Enfin le problème de la construction d'un intervalle de prévision conditionnel est abordé. Le paragraphe 3 est consacré à la statistique des jours semblables à l'aide des similarités. Le paragraphe 4 présente la démarche de construction du prédicteur non paramétrique de la courbe de charge électrique corrigée du chauffage, lors des périodes normales. Le paragraphe 5 contient de brefs commentaires sur les résultats des prévisions.

1. Les problèmes, les données

Ce travail s'intègre dans une étude visant l'amélioration des prévisions de la consommation d'électricité. La consommation de l'ensemble du territoire français est prise en compte. Actuellement les méthodes utilisées par EDF fournissent de bonnes prévisions à l'horizon de la demi-heure, lors d'un jour normal. La prévision à un horizon plus grand que la demi-heure (jusqu'à 24 heures) conserve une qualité suffisante. Bien que les taux d'erreurs soient difficiles à quantifier en un seul chiffre, 1.5 à 3 % représente un ordre de grandeur traditionnellement admis de l'erreur relative moyenne (*cf.* Karanta, Ruusunen 91 et Harvey, Koopman 93). La consommation actuelle atteint le niveau moyen d'environ 40000 Mw. La technique actuelle de prévision (*cf.* Ernoult, Mattatia 84) fonctionne de façon satisfaisante depuis plusieurs années et utilise principalement des modèles de la famille SARIMA de Box-Jenkins, modèles paramétriques ayant fait leurs preuves. Une fonction de transfert entre la température et la consommation permet de tenir compte des effets du climat (sur ce

point, on trouvera dans Engle *et al.* 86, une approche par régression semi-paramétrique de ce problème). Cependant la procédure actuelle de prévision présente de sérieuses difficultés pour tenir compte des périodes dites spéciales : les jours fériés, les vacances, les jours de tarification spéciale (EJP) et plus généralement les périodes de rupture vis à vis du régime «normal» de consommation électrique. Ces problèmes affectent environ 25 à 30% des jours et conduisent à des procédures ad hoc pour les périodes spéciales ainsi qu'à des procédures d'invalidation et de reconstitution des périodes spéciales du passé avant la prévision de périodes normales. La prévision non paramétrique permet une approche plus unifiée et plus simple car s'adaptant mieux à la diversité des profils de consommation qu'un modèle SARIMA s'avérant trop rigide. Notons en outre que la solution présentée ici est plus simple que l'approche préconisée dans Harvey, Koopman 93 utilisant des fonctions splines variant dans le temps.

Les données étudiées sont des enregistrements de la courbe de charge. Deux suites sont disponibles : les consommations brutes et les corrections de chauffage issues d'une nouvelle technique de prise en compte des effets de l'aléa climatique (*cf.* Papon 92). Les enregistrements sont donnés 1/2 heure par 1/2 heure, jour par jour. Les données étudiées sont corrigées du chauffage. La période analysée commence le 1er août 1985 et se termine le 4 juillet 1992, soit 2530 jours ou 121440 demi-heures. On trouvera dans les figures 1 et 2 des extraits de cette chronique, montrant une grande régularité globale, une périodicité hebdomadaire massive ainsi qu'une pseudo-périodicité journalière.

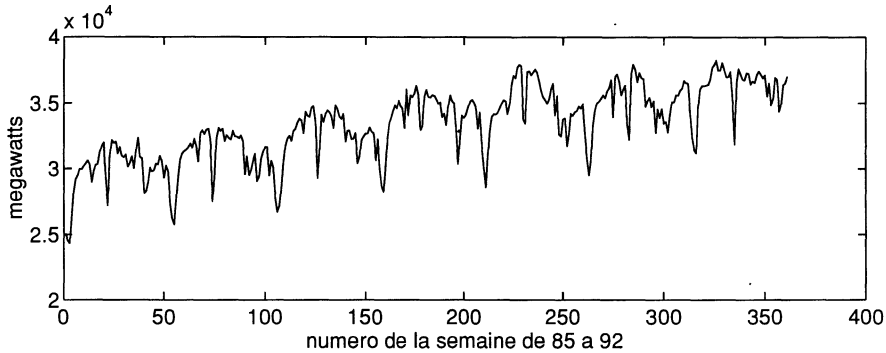


FIGURE 1

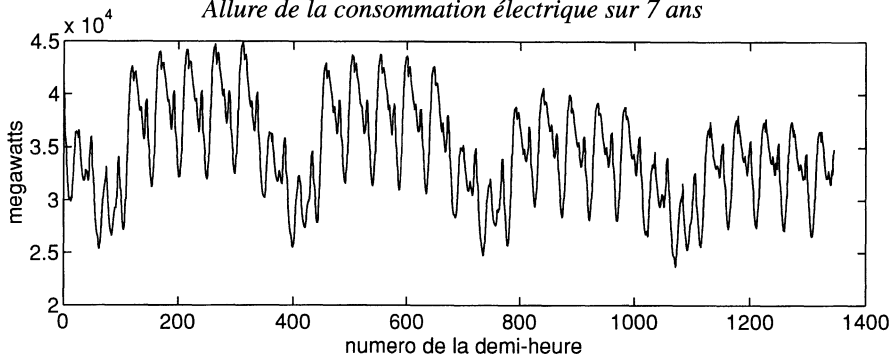
Allure de la consommation électrique sur 7 ans

FIGURE 2

Allure de la consommation électrique sur un mois de 91

2. Les méthodes

2.1. Prédicteur à noyau

Dans la suite nous utiliserons les notations suivantes : $X = (X_n)_{n \in \mathbb{N}}$ désigne le processus de consommation sous-jacent, dont la courbe de charge électrique est une réalisation désignée par $(x_n; n = 1, \dots, T)$. On note $X_{n,(r)}$ le r -uple de variables aléatoires suivant :

$$X_{n,(r)} = (X_n, X_{n-1}, \dots, X_{n-r+1}) \quad (1)$$

Pour un horizon de prévision fixe noté s , la fonction d'autorégression f est définie par :

$$f(x) = E(X_{n+s} | X_{n,(r)} = x) \quad \text{pour } x \in \mathbb{R}^r \quad (2)$$

Remarquons que (X_n) n'étant pas stationnaire on ne peut pas vraiment parler ici d'autorégression. On utilise en fait un estimateur de la «régression moyenne» (pour des résultats théoriques sur le sujet, on peut consulter Bosq 91). L'estimateur à noyau de la fonction f basé sur (X_1, X_2, \dots, X_n) , noté f_n , est défini par :

$$f_n(x) = \frac{\sum_{t=r, \dots, n-s} K((x - X_{t,(r)})/h_n) X_{t+s}}{\sum_{t=r, \dots, n-s} K((x - X_{t,(r)})/h_n)} \quad (3)$$

ou 0 si le dénominateur est nul, où K est un noyau positif, et où $h = (h_n)_n$ est une suite réelle positive décroissante : la fenêtre. Le prédicteur à noyau s'en déduit :

$$X_{n+s|n} = f_n(X_{n,(r)}) \quad (4)$$

et sa valeur est $x_{n+s|n} = f_n(x_{n,(r)})$.

On dispose de nombreux résultats sur le comportement asymptotique de l'estimateur f_n de la fonction f . En particulier, des résultats de convergence presque sûre de f_n vers f et de convergence en loi de $f_n(x) - f(x)$ convenablement normalisée, sont disponibles. Ces résultats ont été montrés essentiellement dans deux cadres : sous des hypothèses de mélange de X (cf. Doukhan 94, Ango Nze, Doukhan 93, Bosq, Lecoutre 87 p.105-135, Robinson 83, Roussas, Tran 92, Truong, Stone 92), et lorsque X est engendré par un AR non linéaire du type $X_{n+1} = f(X_n) + \eta_{n+1}$, sous des hypothèses convenables sur f et la loi de η (cf. Dufflo 90 p. 188-196, Yakowitz 89).

L'estimateur $f_n(x)$ donné par (3) peut aussi s'écrire lorsque n et s sont fixés, sous la forme suivante :

$$f_n(x) = \sum_{t=r, \dots, n-s} w_{n,t} X_{t+s} \quad (5)$$

où $w_{n,t}$ est défini par :

$$w_{n,t} = \frac{K((x - X_{t,(r)})/h_n)}{\sum_{m=r,\dots,n-s} K((x - X_{m,(r)})/h_n)} \quad (6)$$

C'est-à-dire que $X_{n+s|n}$ s'obtient comme le barycentre des $(X_{t+s})_t$ affectés des poids $(w_{n,t})_t$. Si K est un noyau positif choisi de sorte que $K(x)$ décroisse lorsque $\|x\|$ croît, (par exemple le noyau gaussien sur \mathbb{R}^r), alors $w_{n,t}$ admet l'interprétation suivante :

- $X_{n,(r)}$ est le bloc témoin auquel les autres blocs $(X_{t,(r)})_t$ sont comparés;
- $w_{n,t}$ peut alors s'interpréter comme un indice de similarité du bloc $X_{t,(r)}$ au bloc témoin.

La prévision $X_{n+s|n}$ s'obtient donc comme le barycentre des futurs d'horizon s de tous les blocs de longueur r du passé affectés des coefficients de similarité $w_{n,t}$ de ces blocs au bloc témoin.

En outre, on peut remarquer dans (6) que les numérateurs de ces poids $(w_{n,t})_t$ sont indépendants de l'horizon de prévision s (pourvu que la fenêtre h le soit). Les mêmes poids, calculés une fois pour toutes, n étant donné, sont utilisés pour calculer les prévisions à un horizon quelconque, après une troncature dans le temps et une normalisation convenable. A des fins d'interprétation de la prévision, il convient d'examiner les instants t auxquels correspondent les grandes valeurs des poids (cf. figures 6 et 8).

2.2. Similarités

La prédiction se décompose donc, d'après ce qui précède, en deux phases :

- rechercher des périodes semblables aux r dernières observations : on utilise les poids;
- calculer le barycentre des futurs de ces blocs affectés des poids.

La première phase mérite d'être isolée et permet l'étude des périodes semblables, à deux fins : description de la chronique et extraction des connaissances issues de la description (classes de périodes, décroissance des poids,...) pour aider au calibrage des paramètres du prédicteur.

Dans cette phase, afin de prendre en compte l'instationnarité de la chronique, deux indices de similarité sont considérés. Ils sont basés sur l'indice (6) et sont définis comme suit, ayant fixé un instant n .

Le premier (méthode 1 : similarité entre blocs bruts) s'en déduit naturellement :

$$\text{sim}_1(x, X_{t,(r)}) = K((x - X_{t,(r)})/h) \quad (7)$$

Pour ôter l'effet d'une instationnarité affine en moyenne (cf. figure 1), on peut comparer les deux blocs en les ayant préalablement ramené à une moyenne commune

(méthode 2 : similarité entre blocs centrés) :

$$\text{sim}_2(x, X_{t,(r)}) = K((c(x) - c(X_{t,(r)}))/h) \quad (8)$$

où $c(y) = y - m(y)$ avec $m(y)$ la moyenne du bloc y .

Dans ce cadre, pour construire la prévision, on remplacera dans (5), X_{t+s} par Z_{t+s} avec

$$Z_{t+s} = X_{t+s} + (m(X_{n,(r)}) - m(X_{t,(r)})) \quad (9)$$

et dans $w_{n,t}$, sim_1 par sim_2 .

2.3. Intervalle de prévision

Ce qui suit est justifié (sous des hypothèses adéquates) pour la méthode 1 de calcul des poids (cf. (7)); elle sera aussi utilisée dans le cadre de la méthode 2, sans justification pour le moment. L'objectif est de construire un intervalle de confiance conditionnel à l'instant n , pour chaque horizon de prévision s fixé. L'intervalle noté I_s est une fonction de s, n, α et x tel que :

$$P(X_{n+s} \in I_s | X_{n,(r)} = x) \geq 1 - \alpha \quad (10)$$

Ceci peut se faire en construisant au préalable un estimateur de la densité conditionnelle sous-jacente (en supposant qu'elle existe), à savoir $g_s(y|x)$ la densité de la loi L de X_{n+s} conditionnée par $X_{n,(r)} = x$.

Le prédicteur $X_{n+s|n}$ n'est autre que $E(X_{n+s} | X_{n,(r)} = x)$, c'est-à-dire l'espérance de la loi L . Pour mener à bien la construction de I_s , il suffit de disposer d'un estimateur $g_{\text{est},s}(y|x)$ convergeant presque sûrement vers $g_s(y|x)$. On trouvera dans Roussas 91, Youndjé *et al.* 93 (sous des hypothèses de mélange) et Senoussi 91 (sous des hypothèses markoviennes), de tels résultats. Pour le moment, seul est estimé l'écart-type de cette loi conditionnelle :

$$S_{\text{est},s} = \left(\sum_t w_{n,t} e_t^2 \right)^{1/2} \quad (11)$$

où

$$e_t = X_{t+s} - X_{t+s|t} \quad (12)$$

est l'erreur de prévision d'horizon s à l'instant t . Ceci correspond à l'estimateur de la loi conditionnelle constitué par la mesure chargeant les points $(X_{t+s})_t$ avec des poids $(w_{n,t})_t$:

$$g_{\text{est},s} = \sum_t w_{n,t} \delta(X_{t+s}) \quad (13)$$

où $\delta(x)$ est la masse de Dirac en x .

En fait, le coût de calcul de (12) pour tous les instants t disponibles conduit à considérer un deuxième estimateur obtenu en remplaçant dans (12) $X_{t+s|t}$ par $X_{n+s|n}$, pour tout t :

$$\sigma_{\text{est},s} = \left(\sum_t w_{n,t} \varepsilon_t^2 \right)^{1/2} \quad (14)$$

où

$$\varepsilon_t = X_{t+s} - X_{n+s|n} \quad (15)$$

qui est raisonnable puisque $X_{n+s|n}$ est d'autant plus proche de $X_{t+s|t}$ que $w_{n,t}$ est grand. Cet estimateur, moins coûteux à calculer, a en outre l'avantage d'admettre une interprétation intuitive permettant de visualiser la variabilité des futurs des blocs du passé prépondérants dans la prévision. En effet, par définition du prédicteur $X_{n+s|n}$ (cf. (5)) et de ε_t (cf. (15)), il vient que :

$$\sum_{t=r, \dots, n-s} w_{n,t} \varepsilon_t = 0 \quad (16)$$

On peut donc associer à l'instant n , pour un horizon s fixé, l'ensemble des :

$$(X_{t+s} - \varepsilon_t)_t \quad (17)$$

munis des poids $(w_{n,t})_t$ dont le prédicteur $X_{n+s|n}$ est le barycentre.

En procédant ainsi pour chaque horizon s de 1 à S , on obtient un écheveau de portions de trajectoires de longueur S du passé dont chaque filament est muni du poids $w_{n,t}$ correspondant, tel que la trajectoire prévue en est le barycentre.

D'un point de vue de la qualité de la prévision, la dispersion de l'écheveau ou des filaments les plus pesants (mesurée par (14)) est un bon indicateur de la plus ou moins grande homogénéité des futurs des blocs les plus ressemblants au bloc témoin du présent. On se contente ici, pour se donner une idée de celle-ci, de construire autour de la trajectoire prévue l'intervalle $X_{n+s|n} \pm 3\sigma_{\text{est},s}$ pour chacun des horizons de prévision (cf. figures 5 et 7).

3. Statistique des jours semblables

Dans toute la suite, K le noyau introduit en (3) est le noyau gaussien sur \mathbb{R}^r , c'est-à-dire que :

$$K(y) = c \cdot \exp(-\|y\|^2/2), y \in \mathbb{R}^r \quad \text{où } c = (2\pi)^{-r/2}. \quad (18)$$

Il s'agit ici d'utiliser la notion de similarité pour décrire la chronique. On s'intéressera essentiellement à associer à une période témoin, en général un jour, les périodes qui lui ressemblent le plus au sens de l'ordre induit par les similarités par rapport au témoin.

A cet égard, il faut noter que si les valeurs des similarités $(\text{sim}(x, X_{t,(r)}))_t$ dépendent de h_n , en revanche leur ordre n'en dépend pas. En effet par (18) on a que $K(y/h) = c.\exp(-\|y\|^2/2h^2)$, par conséquent l'ordre des $(\text{sim}(x, X_{t,(r)}))_t$ n'est autre que l'ordre inverse des distances entre les blocs $(\|x - X_{t,(r)}\|)_t$, qui est le même pour tout choix de h_n .

L'analyse consiste en l'examen pour un jour donné, des périodes de 48 demi-heures les plus ressemblantes dans l'ensemble de la série notée $(x_i; i = 1, \dots, T)$. Pour cela, on calcule les quantités suivantes (pour les 2 méthodes) :

$$s_{n,t} = \text{sim}(x_{n,(r)}, x_{t,(r)})/c \quad \text{pour } t = r, \dots, T \text{ et } r = 48, \quad (19)$$

de sorte que l'on ait : $\max_t s_{n,t} = s_{n,n} = 1$.

Sur les blocs les plus ressemblants à un jour donné, on s'intéresse aux répartitions par an, par mois, par jour et par statut-fête.

On examine aussi des effets moyens, sur tous les jours d'un certain type. Ceci se fait en calculant (19) pour chacun des jours dans un ensemble précisé *a priori* présentant une caractéristique calendaire ou électrique, en calculant ensuite la répartition des p jours les plus ressemblants vis-à-vis d'une variable intéressante, puis en construisant l'histogramme moyen.

Les principales conclusions sont les suivantes. Les similarités des blocs asynchrones au bloc témoin (*i.e.* lorsque $n - t$ n'est pas multiple de 48) sont négligeables. Les similarités permettent de retrouver des connaissances *a priori*, d'origine électrique, par exemple : la distinction été-hiver est nette, la décroissance des similarités en tendance lorsque $|t - n|$ croît, la classification des jours normaux en 4 ou 5 classes constituées ainsi : {sam}, {dim}, {lun}, {mar, mer, jeu} et {ven} que l'on pourrait agglomérer à la précédente. La méthode 1 de calcul des similarités privilégie trop les périodes les plus récentes.

4. Démarche de construction du prédicteur

Ce paragraphe présente la structure retenue pour le prédicteur, en l'accompagnant de justifications préalables. On ne présente ici que le travail sur les jours normaux (environ 80 à 85% des jours pour notre prédicteur) qui est achevé et qui permet de calibrer les deux paramètres fondamentaux de la méthode : la mémoire r et la largeur de la fenêtre $(h_n)_n$.

Le paramètre r doit être réglé en premier, car il est très lié à la connaissance de la série, qui doit permettre d'en faire un choix raisonnable. Le second paramètre en revanche, est plus lié à la technique d'estimation qu'aux données.

On peut s'interroger sur l'opportunité de travailler sur les données elles-mêmes ou sur les données stationnarisées. Bien que la procédure de stationnarisation soit classique, l'intuition actuelle est que c'est précisément la stabilité globale et la périodicité hebdomadaire qui permettront d'obtenir les meilleurs résultats en prévision non paramétrique.

Nous travaillons donc sur les données corrigées du chauffage sans les transformer.

Les études sur les similarités conduisent à ne pas retenir la méthode 1 pour laquelle les périodes récentes sont trop privilégiées. Seule la méthode 2, travaillant sur les blocs centrés (cf. (8)), sera considérée dans la suite.

Le premier problème clé est le choix de la taille r de la mémoire. Rappelons que si le processus X est r -markovien alors le prédicteur (2) à (4) est le meilleur prédicteur au sens qu'il minimise par rapport à g le critère quadratique suivant :

$$C(g) = E[g(X_n, X_{n-1}, \dots, X_1) - X_{n+s}]^2 \quad (20)$$

La détermination automatique de r pour un modèle non linéaire commence à être examinée (cf. Auestad, Tjostheim 90 et 91), mais ces procédures encore très partielles, n'ont pas été utilisées ici. Deux indications sont plus utiles : la première se réfère au modèle actuellement utilisé faisant ressortir essentiellement la périodicité hebdomadaire et la pseudo-périodicité journalière; la deuxième est liée à l'étude des similarités basée sur les jours. Ceci conduit à considérer qu'il convient de choisir r multiple de 48 et compris entre 48 (un jour) et 336 (une semaine).

Un bloc témoin de longueur une semaine présenterait l'avantage de régler le problème de l'hétérogénéité calendaire des futurs des blocs du passé. Mais cette procédure a l'inconvénient d'être exagérément sélective puisque la présence d'un jour spécial dans une semaine suffit à l'exclure de fait, de la prévision du futur d'une semaine normale. La solution traditionnelle dans les modèles paramétriques, consiste alors à invalider les réalisations anormales du passé et les remplacer par leurs prévisions normales. Ceci est inutile dans le cadre non paramétrique.

On peut alors penser à choisir $r = k.48$ où k est le nombre minimal de jours suffisant pour induire l'homogénéité calendaire évoquée plus haut. D'après les études de similarités, k vaut 3 ou 4 (suivant le statut que l'on réserve au vendredi). La question qui se pose alors est de savoir s'il est vraiment raisonnable de faire «découvrir» au prédicteur la périodicité hebdomadaire. On pense que non. On choisira donc $r = 48$. En effet, si $r = 48$ et que l'on prévoit 96 demi-heures (2 jours), en se basant sur les classes de jours issues du travail sur les similarités : {lu}, {ma, me, je}, {ve}, {sa} et {di}, on constate que les transitions de classes entre les différents jours de la semaine pris comme témoins et leurs lendemain-surlendemain sont alors toutes distinctes. La forme du prédicteur s'en déduit :

$$X_{n+s|n} = \sum_{t \in J(n)} \alpha_{n,t} Z_{t+s} \quad (21)$$

(où $J(n)$ est la sous-grille temporelle ne comprenant que les instants issus de n par pas de 336 (une semaine), où les $\alpha_{n,t}$ sont issus de la méthode 2 (cf. (8)) avec $\sum_{t \in J(n)} \alpha_{n,t} = 1$, et où $Z_{t+s} = X_{t+s} + [m(X_{n,(48)}) - m(X_{t,(48)})]$, i.e. X_{t+s} convenablement corrigé d'un effet de moyenne conformément à la méthode 2.

Ayant dégagé la structure du prédicteur, il reste maintenant à régler le deuxième paramètre : la fenêtre h du prédicteur à noyau. La technique utilisée est inspirée de la validation croisée en régression (cf. Vieu 91 et 92, Härdle, Vieu 92, Härdle 90). Le principe de celle-ci consiste, pour un horizon de prévision s fixé, en la minimisation d'une mesure de déviation quadratique entre f et son estimateur f_n

vu comme une fonction de h , que nous noterons ici simplement f_h . On montre alors, sous des hypothèses peu restrictives sur le noyau K , des hypothèses de régularité sur f et g (la densité marginale de X) et de mélange de X , qu'il est asymptotiquement équivalent de minimiser le critère dit de validation croisée suivant :

$$CV_s(h) = \sum_{i=r, \dots, n-s} \{X_{i+s} - f_{h,-i}(X_{i,(r)})\}^2 / (n-s) \quad (22)$$

où $f_{h,-i}(x)$ est l'estimateur $f_h(x)$ calculé à l'aide de toutes les données sauf la $i^{\text{ème}}$. La procédure utilisée s'inspire de ce qui précède sauf sur un point : on opère une optimisation simultanée sur tous les horizons s de 1 à 96, le critère est $\sum_{s=1, \dots, 96} CV_s(h)$, afin que la fenêtre h soit la même pour tout horizon s , bien que les fonctions de régression soient différentes. On choisit donc de chercher h_n de la forme suivante :

$$h_n = h.(s(x_1, x_2, \dots, x_n)n^{-1/(r+4)}) \quad (23)$$

où $n^{-1/(r+4)}$ résulte du compromis entre le carré du biais et la variance asymptotiques lorsque f et g sont supposées 2 fois différentiables (cf. Bosq, Lecoutre 87 p. 119), où $s(x_1, x_2, \dots, x_n)$ l'écart-type empirique des (x_i) , constitue un choix classique de première normalisation (cf. Bosq, Lecoutre 92 p. 165 et Carbon, Delecroix 87); et où h est la variable sur laquelle porte l'optimisation, à rechercher dans un segment $[a, b]$ convenablement choisi.

Le plan utilisé pour déterminer h est le suivant : on considère la population des N jours normaux de l'année 91, le passé commun est constitué des trois années précédentes : 1988 à 90, et on calcule :

$$cv_i(h) = \sum_{s=1, \dots, 96} (X_{i+s} - X_{i+s|i})^2 / 96 \quad (24)$$

$$\text{puis } cv(h) = \sum_{i=1, \dots, N} cv_i(h) / N \quad (25)$$

$$\text{et } h = \text{argmin } cv(h). \quad (26)$$

Complémentairement on s'intéressera aussi à un critère d'erreur relative moyenne :

$$\text{Crit}_i(h) = \sum_{s=1, \dots, 96} \{|X_{i+s} - X_{i+s|i}| / X_{i+s}\} / 96 \quad (27)$$

$$\text{puis } \text{Crit}(h) = 100 \sum_{i=1, \dots, N} \text{Crit}_i(h) / N \quad (28)$$

Les résultats sont les suivants : $h = 0.36$ est la valeur retenue, les minimums des deux critères (25) et (28) sont atteints pour des valeurs de h très voisines : 0.36 et 0.37 (cf. figures 3 et 4).

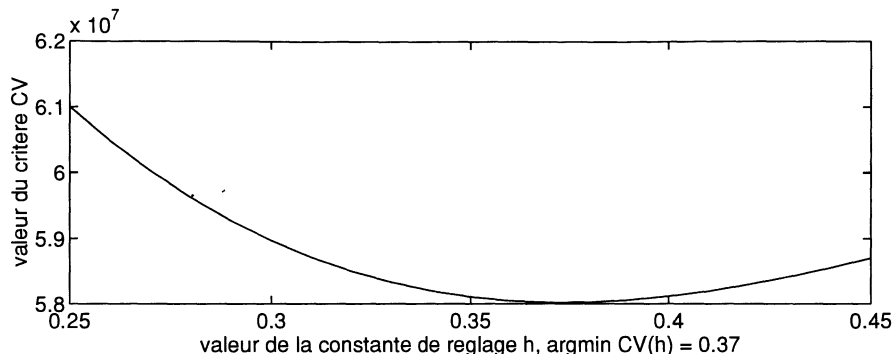


FIGURE 3

Validation croisée sur 91, critère en moyenne quadratique

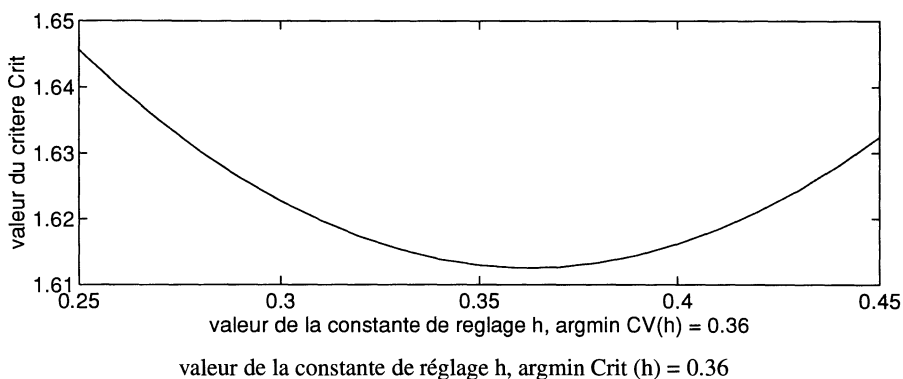


FIGURE 4

Validation croisée sur 91, critère en erreur relative

5. Résultats de prévision

Pour les jours normaux, nous proposons la variante très naturelle suivante du prédicteur (21) : elle consiste à ne pas considérer pour élaborer la prévision, les observations appartenant à des périodes spéciales (fériées, pseudo-fériées ou à tarification spéciale : EJP). Ceci se fait par la simple adjonction d'une information calendaire, sans modifier les données.

Les performances globales (qui sont identiques pour les années 90, 91, données ayant permis d'élaborer le modèle, et 92 qui sont les données tests) peuvent être résumées par les quelques chiffres suivants portant sur les critères d'erreur (27) et (28) : la moyenne est de 1.6 %, la médiane de 1.4 %, les pourcentages des erreurs inférieures à 1, 2, 3 et 4 % sont respectivement d'environ 20, 80, 95 et 97 %. Elles

sont très satisfaisantes et améliorent parfois notablement les résultats obtenus par la méthode actuelle, par exemple en août.

Commentons pour finir, deux exemples de prévision de deux périodes de deux jours, l'une de bonne qualité en février et l'autre de qualité médiocre en août. Deux graphiques présentent, pour chacune des deux périodes, les informations suivantes :

- graphique 1 (figures 5 et 7) : la réalisation (en trait plein), la prévision (en pointillé), une bande (en tirets) qui, pour chaque horizon s , a une largeur de $\pm 3\sigma_{\text{est},s}$ cf. (14) centrée sur la prévision.

- graphique 2 (figures 6 et 8) : en abscisse, le numéro de la semaine depuis début 1988 et en ordonnée la similarité associée, celle-ci est normalisée de sorte que la similarité maximale (obtenue pour le témoin) soit égale à 100.

La première période (figures 5 et 6) concerne la prévision des 4 et 5 février 91 à partir du 3 pris comme témoin. La qualité de la prévision sur la figure 5 est bonne comme en atteste la performance d'environ 1% et la faible amplitude de l'intervalle de prévision. L'examen des similarités sur la figure 6 montre qu'essentiellement la prévision est basée sur les mêmes périodes lors des deux années précédentes.

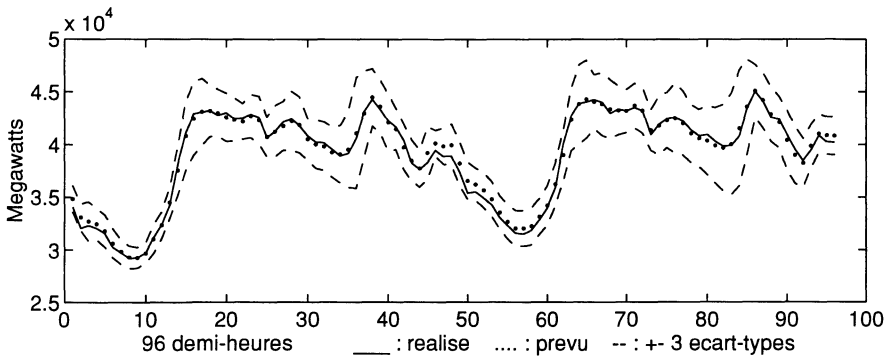


FIGURE 5

Prévision des 4 et 5 février 91. Erreur relative = 0.8986%

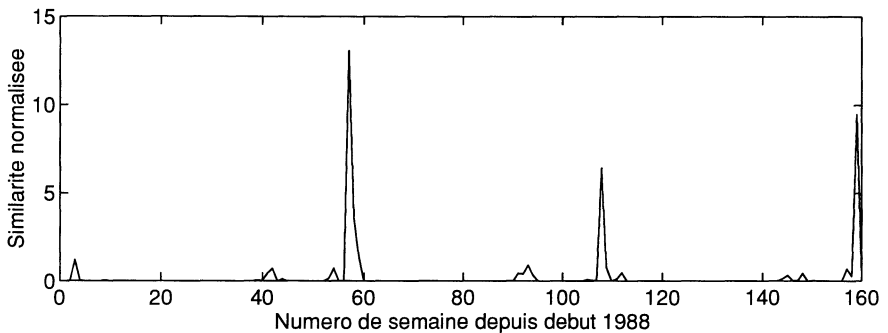


FIGURE 6

Similarités au jour témoin : le 3 février 91

La seconde période (figures 7 et 8) concerne la prévision des 26 et 27 août 91 à partir du 25 pris comme témoin. La qualité de la prévision sur la figure 7 est médiocre comme en atteste la performance d'environ 2.3% et l'assez grande amplitude de l'intervalle de prévision. L'examen des similarités sur la figure 8 montre que la prévision est basée sur beaucoup plus de périodes que dans l'exemple précédent et que des périodes plus anciennes interviennent.

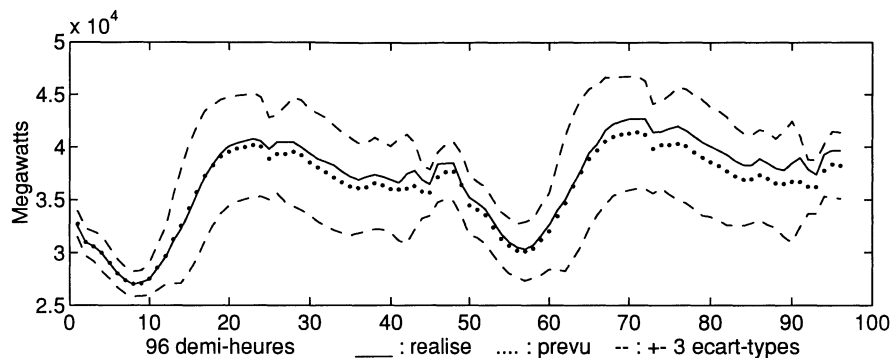


FIGURE 7

Prévision des 26 et 27 août 91. Erreur relative = 2.159%

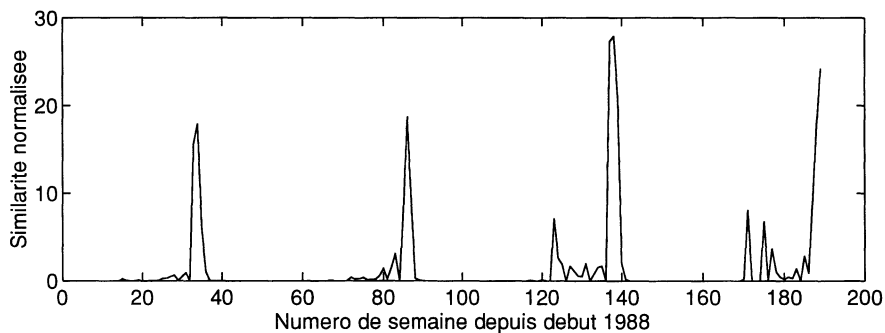


FIGURE 8

Similarités au jour témoin : le 25 août 91

6. Conclusion

Le prédicteur non paramétrique donne des résultats intéressants pour les jours normaux, au moins aussi bons que ceux obtenus par EDF.

En dehors de ses bonnes performances, le prédicteur non paramétrique présente deux avantages décisifs : il admet une interprétation directe en termes électriques par le biais des similarités et il est extrêmement simple à mettre en oeuvre. Le prédicteur travaille en effet directement sur les données (heureusement corrigées du chauffage) sans invalidation ni recalculs des périodes spéciales.

En dehors de l'application à la charge électrique, ce travail illustre une méthode, qui bien que délicate à mettre au point, permet de faire de la statistique descriptive, de la modélisation et de la prévision des séries temporelles à l'aide d'un seul et même outil élémentaire issu de la notion de similarité. Du point de vue de la diffusion des méthodes de prévision, celle-ci semble prometteuse puisqu'elle permet de marier une classe de modèles très riche avec des outils simples permettant la mise au point, la validation et l'explication des prévisions avec des arguments aisément insérables dans le contexte scientifique et technique concerné.

Références Bibliographiques

- ANGO NZE P., DOUKHAN P. (1993), *Functional estimation for time series : a general approach*, (à paraître), Prepublication mathématique d'Orsay, 93-43.
- ANGO NZE P., DOUKHAN P. (1993), *Estimation fonctionnelle de séries temporelles mélangées*, C.R.A.S. Paris, t. 317, Serie I, p. 405-408.
- AUESTAD B., TJOSTHEIM D. (1990), *Identification of non linear time series : First order characterization and order determination*, *Biometrika*, 77, 4, p. 669-687.
- AUESTAD B., TJOSTHEIM D. (1991), *Functional identification in non linear time series*, in G. Roussas (ed.), *Non parametric functional estimation and related topics*, NATO ASI Series, Kluwer Acad. Pub., p. 493-507.
- BOSQ D. (1991), *Non parametric prediction for unbounded almost stationary processes*, in G. Roussas (ed.), *Non parametric functional estimation and related topics*, NATO ASI Series, Kluwer Acad. Pub., p. 389-403.
- BOSQ D., LECOUTRE J.P. (1987), *Théorie de l'estimation fonctionnelle*, *Economica*.
- BOSQ D., LECOUTRE J.P. (1992), *Analyse et prévision des séries chronologiques ; méthodes paramétriques et non paramétriques*, Masson.
- CARBON M., DELECROIX M. (1987), *Comparaison de méthodes paramétrique et non paramétrique de prévision*, Pub. IRMA, Lille, vol. 9, n° 11.
- DOUKHAN P. (1994), *Mixing : properties and examples*, lecture note in *Statistics*, n° 85, Springer Verlag.
- DUFLO M. (1990), *Méthodes récursives aléatoires*, Masson.
- ENGLE R.F., GRANGER C.W.J., RICE J., WEISS A. (1986), *Semiparametric estimates of the relation between weather and electricity sales*, *J.A.S.A.*, vol 81, 394, p. 311-320, june.
- ERNOULT M., MATTATIA R. (1984), *Prévision à court terme de la courbe de charge de la consommation électrique*, EDF Bulletin de la DER, Ser. B, 3, p. 37-44.
- GANNOUN A. (1990), *Estimation non paramétrique de la médiane conditionnelle : médianogramme et méthode du noyau. Application à la prévision des processus*, Pub. de l'I.S.U.P., XXXV, 1, p. 11-22.
- GYÖRFI L., HÄRDLE W., SARDA P., VIEU P. (1989), *Non parametric curve estimation from time series*, *Lecture Notes in Statistics*, 60, Springer Verlag.

- HÄRDLE W. (1990), *Applied nonparametric regression*, Cambridge University Press.
- HÄRDLE W., VIEU P. (1992), *Kernel regression smoothing of time series*, Journal of Time Series Analysis, vol. 13, n 3, p. 209-232.
- HARVEY A., KOOPMAN S. J. (1993), *Forecasting hourly electricity demand using time-varying splines*, J.A.S.A., vol 88, 424, p. 1228-1236, dec.
- KARANTA I., RUUSUNEN J. (1991), *Short term load forecasting in communal electric utilities*, Helsinki Univ. of Technology, Research report A40, may .
- MISITI M., MISITI Y., OPPENHEIM G., POGGI J.M. (1993), *Prévision fonctionnelle de la courbe de charge électrique*, Rapport de contrat de recherche EDF, sept.
- PAPON M. (1992), *Evaluer l'impact de l'aléa climatique sur la consommation d'électricité*, EDF-DER, Epure, 36, oct.
- PRIESTLEY M.B. (1988), *Non linear and non stationary time series analysis*, Academic Press, 1988.
- ROBINSON P.M. (1983), *Nonparametric estimators for time series*, Journal of Time Series Analysis, 4, p. 185-207.
- ROUSSAS G. (1991), *Recursive estimation of the transition distribution function of a Markov process : asymptotic normality*, Statistics and Probability Letters, 11, p. 435-447.
- ROUSSAS G. (1991), *Estimation of transition distribution function and its quantiles in Markov processes : strong consistency and asymptotic normality*, in G. Roussas (ed.), *Non parametric functional estimation and related topics*, NATO ASI Series, Kluwer Acad. Pub., p. 443-462.
- ROUSSAS G., TRAN L.T. (1992), *Asymptotic normality of the recursive kernel regression estimate under dependence conditions*, Annals of Statistics, vol. 20, n 1, p. 98-120.
- SENOUSSI R. (1991), *Loi du log itéré uniforme pour des martingales et estimation récursive de modèles markoviens contrôlés*, Thèse d'état, Univ. Paris-Sud.
- TRUONG Y.K., STONE C.J. (1992), *Nonparametric function estimation involving time series*, Annals of Statistics, 20, 1, p. 77-97.
- TONG H. (1990), *Non linear time series. A dynamical approach*, Oxford Science Publications.
- VIEU P. (1991), *Quadratic errors for nonparametric estimates under dependence*, Journal of Multivariate Analysis, vol. 39, n 2, p. 324-347.
- VIEU P. (1991), *Nonparametric regression : optimal local bandwidth choice*, J.R.S.S., Ser. B, 53, n 2, p. 453-464.
- YAKOWITZ S. (1985), *Markov flow models and the flood warning problem*, Water resources research, vol. 21, n1, p. 81-88.

- YAKOWITZ S. (1987), *Nearest-neighbour methods for time series*, *Journal of Time Series Analysis*, vol. 8, p. 235-247.
- YAKOWITZ S. (1989), *Nonparametric density and regression*, *Journal of Multivariate Analysis*, vol. 30, p. 124-136.
- YOUNDJÉ E., SARDA P., VIEU P., (1993), *Estimateur à noyau d'une densité conditionnelle : choix de la fenêtre pour des observations dépendantes*, CRAS, Paris, Ser. I, t. 316, p. 935-938.