Prédiction de la demande énergétique à court terme par régression non paramétrique

Mathieu Le Cam[1], Ahmed Daoud[2]
[1]École Polytechnique de Montréal, Montréal, Québec, Canada
[2]Laboratoire des Technologies de l'Énergie, institut de recherche d'Hydro-Québec, Shawinigan, Québec, Canada

Abstract

Cet article présente un modèle de prédiction de la demande énergétique sur les vingt-quatre prochaines heures par régression non paramétrique. La demande énergétique est estimée par une moyenne locale, dans un historique de deux années, des appels de puissance enregistrés dans des conditions similaires de climat et d'occupation. Un algorithme génétique est ensuite utilisé afin d'optimiser la taille de la plage de variation de chaque paramètre caractéristique de la similarité des conditions de façon à minimiser l'erreur de prédiction de la demande sur une semaine. La méthode de prédiction est tout d'abord appliquée sur des données synthétiques obtenues par simulation sur EnergyPlus avec un modèle de maison typique du Québec. L'algorithme est ensuite employé sur des données réelles mesurées sur un bâtiment résidentiel québécois.

1 Introduction

La prédiction de la demande énergétique est d'une importance conséquente pour les utilités responsables de la production, du transport et de la distribution de l'énergie. Une prédiction fiable est nécessaire à une bonne gestion de la demande et permet également une planification optimisée de la production.

Une analyse préliminaire de l'évolution de la demande énergétique pour le chauffage d'une maison révèle qu'en l'absence de changement dans les propriétés thermiques et les équipements de chauffage et ventilation de celle-ci, la variation de la demande dépend essentiellement du climat et du comportement des habitants, c'est-à-dire de leurs habitudes de vie. Ainsi, les excitations de type comportemental sont directement liées à l'heure du jour et au jour de la semaine. En effet, on différencie clairement les profils de demande des jours ouvrables et des jours de fin de semaine. Concernant les variables représentatives du climat, la température est le facteur prépondérant, l'humidité est à prendre en compte l'été et la vitesse du vent l'hiver ainsi que l'ensoleillement pour considérer les gains solaires.

Dans la littérature, (Alfares & Nazeeruddin, 2002; Feinberg, Genethliou, Chow, Wu, & Momoh, 2005; Hahn, Meyer-Nieberg, & Pickl, 2009; Hippert, Pedreira, & Souza, 2001; Kyriakides, Polycarpou, Chen, & Wang, 2007) ont décrit et classé les modèles de prédiction à court terme de la demande énergétique dans les catégories suivantes : les méthodes de régression, les séries temporelles, les approches de jour-similaire, les réseaux de neurones et enfin les techniques de logique floue. Une autre méthode de prédiction de la demande énergétique basée sur une régression non-paramétrique a été présentée par (Agarwal et al., 2006; Asber, Lefebvre, Asber, Saad, & Desbiens, 2007; Charytoniuk, Chen, & Van Olinda, 1998; Kramer et al., 2010). Celle-ci peut-être

rapprochée de la méthode du jour-similaire : elle estime la demande énergétique future en calculant une moyenne locale pondérée des consommations passées dont les conditions sont similaires. Cette méthode de modélisation inverse ne requiert pas de connaissances physiques préalables sur le bâtiment résidentiel dont on cherche à estimer la demande. Cette technique permet d'extraire de l'information à partir des données, elle est particulièrement efficace dans les cas de régression multiple pour lesquels le nombre d'observations est important ainsi que le nombre de variables explicatives. Cependant, cette technique d'estimation requiert un grand nombre de données et c'est une modélisation de type boîte noire : il est donc difficile de donner une explication physique à une valeur d'estimation de la demande. Le modèle de prédiction développé dans cette étude s'appuie sur cette dernière méthode, mais il se distingue d'une part par l'utilisation d'un algorithme génétique pour optimiser la largeur des fenêtres de variation des paramètres représentatifs de la similarité de la journée considérée avec certaines de l'historique. D'autre part, cette étude se distingue également par l'application du modèle développé sur un cas concret : un bâtiment résidentiel québécois implanté dans le climat caractéristique d'Amérique du Nord.

Cet article présente les résultats de prédiction de la demande à court terme pour le chauffage d'un bâtiment résidentiel par une méthode de régression non paramétrique. La première section décrit les techniques de régression non paramétrique mono-variable et multi-variable ainsi que l'algorithme génétique employé pour ajuster les paramètres du modèle afin de minimiser l'erreur de prédiction. La deuxième section présente les résultats de l'application de la méthode d'estimation sur des données synthétiques pour deux scénarios et sur des données réelles mesurées.

2 Méthodologie

La prédiction de la demande énergétique pour le chauffage d'un bâtiment par une méthode inverse requiert l'utilisation de données climatiques et d'un historique des appels de puissance issus de mesures. L'utilisation de données mesurées implique l'existence de bruit dans les mesures, d'erreurs et de données manquantes. Or une prédiction précise de la demande requiert une connaissance préalable de la densité de probabilité de ces erreurs. Dans le cas où l'on ne possède pas d'information a priori sur l'allure d'évolution des variables explicatives et des bruits de mesure, l'estimateur à noyau est une méthode de régression non-paramétrique pertinente à utiliser. Ce type d'estimateur est intéressant lorsque l'on dispose d'un grand volume de données et qu'on ne veut pas ou qu'on est incapable d'imposer un modèle paramétrique pour caractériser l'évolution d'une variable aléatoire. Une revue de littérature complète sur les techniques d'estimation non-paramétrique a été réalisée dans (Duda, Hart, & Stork, 2001).

2.1 L'estimateur à noyaux

2.1.1 Estimateur à noyau simple

L'estimateur à noyau, également appelé fenêtre de Parzen-Rosenblatt, a d'abord été introduit par Rosenblatt en 1956. C'est une méthode non-paramétrique pour estimer la densité de probabilité d'une variable aléatoire. L'estimateur à noyau simple et multi-variables est présenté dans (Martinez, 2001), il est défini par la relation suivante :

$$\hat{f}_{Ker}(x) = \frac{1}{n.h}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \quad \textbf{(1)}$$

Dans cette relation, la fonction K(x) est la fonction à noyau, elle doit satisfaire deux conditions : $\int K(x)dx = 1$ et être symétrique. On peut définir la fonction : $K_h(x - X_i) = K\left(\frac{x-X_i}{h}\right)/h$, l'estimateur à noyau peut alors être réécrit de la manière suivante :

$$\hat{f}_{Ker,h}(x)=\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i) \quad \textbf{(2)}$$

Il existe plusieurs types de fonction à noyau pouvant satisfaire les conditions définies précédemment : les fonctions Uniforme, Triangle, Epanechnikov, Quadratique, Cubique, Gaussienne, Circulaire… Selon (Atkeson, Moore, & Schaal, 1997), lorsqu'on dispose d'un grand volume de données, le type de noyau choisi n'est pas primordial. Pour l'étude présente, une fonction à noyau gaussienne sera retenue ; elle est définie par la relation suivante :

$$K(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right) \quad \textbf{(3)}$$

L'estimateur possède donc les propriétés de continuité et de dérivabilité de la fonction à noyau retenue. L'équation (13) montre que la fonction d'estimation de la densité de probabilité est calculée en introduisant la fonction à noyau gaussienne pondérée, centrée en chaque point des données et en prenant ensuite la moyenne locale en chaque point.

Les paramètres importants à déterminer pour une bonne évaluation par une estimation par noyau sont le type de fonction à noyau choisie d'une part et la largeur h de la fenêtre de Parzen d'autre part. Le paramètre h caractérise le nombre d'observations présentes dans la fenêtre sur laquelle la moyenne pondérée est réalisée, il est représentatif du degré de lissage de l'estimation. Ainsi, une faible valeur de h implique un lissage moins tolérant que pour une valeur plus importante de h. La calibration du degré de lissage est donc très importante pour une bonne évaluation avec l'estimateur à noyau.

2.1.2 Estimateur à noyau multi-variable

Le produit des noyaux est le moyen le plus simple d'obtenir une estimation avec l'estimateur à noyau dans le cas où chaque observation est caractérisée par plusieurs variables. Ainsi, pour n observations X_i, i=1...n, de dimension d chacune, le produit des noyaux est défini par la relation suivante :

$$\hat{f}_{Ker}(x)=\frac{1}{n.h_1\ldots h_d}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K\left(\frac{x_j-X_{i,j}}{h_j}\right)\right\} \quad \textbf{(4)}$$

Ou bien encore :

$$\hat{f}_{Ker,h}(x)=\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_h\left(x_j-X_{i,j}\right)\right\} \quad \textbf{(5)}$$

où $X_{i,j}$ est la $j^{ème}$ composante de la $i^{ème}$ observation.

2.2 Régression non-paramétrique

La méthodologie d'estimation par régression a pour but d'établir une relation entre une variable à estimer et une ou plusieurs variables explicatives dont elle dépend. L'estimateur de Nadaraya-Watson a été retenu pour effectuer une régression non-paramétrique permettant de réaliser une prédiction de la demande énergétique. (Nadaraya, 1964) et (Watson, 1969) ont développé un estimateur à noyau équivalent à la même période ; celui-ci permet d'établir une estimation locale

d'une fonction à un point donné x. L'estimateur de Nadaraya-Watson est défini par la relation suivante :

$$\hat{f}_{NW}(x)=\frac{\sum_{i=1}^{n} K_h(X_i-x).Y_i}{\sum_{i=1}^{n} K_h(X_i-x)} \quad \textbf{(6)}$$

2.2.1 Régression non paramétrique simple

Dans le cas où la variable à estimer y_i ne dépend que d'une seule variable explicative x_i, l'équation de régression peut alors être écrite sous la forme :

$$y_i=m(x_i)+\varepsilon_i, \quad \textbf{(7)}$$

où : $E(\varepsilon_i)=0$ et $Cov(m(x_i),\varepsilon_i)=0$, dans le cas considéré la fonction $m(x)$ peut être non-linéaire. Le terme ε_i est un bruit blanc aléatoire, il a une moyenne nulle et sa variance est notée σ^2. Il représente la variation de y_i autour de sa moyenne $m(x_i)$. Cette moyenne est l'espérance conditionnelle de y_i sachant que $x_i = x$, c'est une fonction de la densité f :

$$m(x_i)=E(y_i|x_i=x)=\frac{\int y.f(x,y)dy}{f(x)} \quad \textbf{(8)}$$

L'estimateur de Nadaraya-Watson présenté précédemment permet d'approcher la fonction $m(x)$:

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K_h(X_i-x).Y_i}{\sum_{i=1}^{n} K_h(X_i-x)},$$
$$\text{avec } K_h(X_i - x) = \frac{1}{h.\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{X_i-x}{h}\right)^2\right) \quad \textbf{(9)}$$

La variable y est donc estimée par une moyenne pondérée des observations autour du point x ; le poids de chaque observation i est donné par le terme : $\frac{K_h(X_i-x)}{\sum_{i=1}^{n} K_h(X_i-x)}$. Le dénominateur est introduit pour normaliser les poids. C'est en agissant sur la largeur h de la fenêtre de Parzen que l'on peut donner plus ou moins d'importance à certaines observations. Les poids donnés aux observations sont d'autant plus importants lorsque X_i est proche de x, ceux-ci diminuent progressivement selon le profil de la gaussienne. Ainsi, pour une faible valeur de h, les poids attribués aux observations diminuent rapidement et moins d'observations sont incluses dans la fenêtre.

2.2.2 Régression non paramétrique multi-variable

La demande énergétique dépend de plusieurs variables explicatives représentatives du climat d'une part et du comportement des habitants d'autre part.

Les variables climatiques retenues pour expliquer la demande énergétique pour le chauffage du bâtiment résidentiel sont la température, la vitesse du vent et l'ensoleillement. On considère également la température sol-air moyenne du jour à prédire et des trois jours précédents afin de tenir compte de l'ensoleillement ainsi que de « l'effet tampon » de l'enveloppe thermique du bâtiment. On calcule une température sol-air moyenne sur la zone géographique considérée avec la relation suivante :

$$T_{sol\text{-}air}(t)=T_{ext}(t)+I(t)/h_{out} \quad \textbf{(10)}$$

où T_{ext} est la température extérieure, I correspond à l'ensoleillement global mesuré au niveau de la station météorologique et le coefficient de transfert thermique est supposé constant et évalué

par (American Society of Heating Refrigerating and Air-Conditioning Engineers., 2009), pour un mur de brique et de plâtre rugueux, à : $h_{out}=12.5\ W/m^2K$.

Concernant les variables représentatives de l'effet comportemental, en lien direct avec l'occupation du bâtiment, celles-ci sont directement dépendantes du temps : heure du jour, jour de la semaine. On peut ainsi distinguer les jours ouvrables des jours de fin de semaine, dont les profils d'appel de puissance sont complètement différents. Cela permet aussi de distinguer les lundis et vendredis, juxtaposés à la fin de semaine, des autres jours ouvrables. Les puissances moyennes appelées les trois jours précédant le jour dont la demande est estimée sont également considérées de façon à tenir compte de l'effet tampon de l'enveloppe thermique du bâtiment.

Les variables explicatives décrites précédemment permettent de juger des similarités des conditions afin d'estimer la demande énergétique par une moyenne locale pondérée des appels de puissance passés.

Pour une régression non-paramétrique multi-variable, on utilise donc l'estimateur de Nadaraya-Watson avec l'estimateur à noyau multi-variables pour attribuer une pondération à chaque observation. Avec les couples $(X_i, Y_i),\ i = 1 \dots n$, où chaque observation X_i est un vecteur de dimension d, l'estimation de la demande énergétique en x est donnée par la relation suivante :

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} Y_i * \left(\prod_{j=1}^{d} K_h(X_{i,j} - x_j)\right)}{\sum_{i=1}^{n} \left(\prod_{j=1}^{d} K_h(X_{i,j} - x_j)\right)},$$

$$\text{avec } K_h(x_j - X_{i,j}) = \frac{1}{h_j.\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{X_{i,j} - x_j}{h_j}\right)^2\right) \quad \textbf{(11)}$$

Dans notre modèle, l'estimateur de la demande énergétique est donc défini par la relation suivante :

$$\hat{y}_i = \widehat{m}(x) + \varepsilon \quad \textbf{(12)}$$

Où l'estimateur m(.) dépend de la largeur h de la fenêtre de Parzen.

2.3 Calibration des largeurs des fenêtres de Parzen

Pour une utilisation efficace de l'outil de régression non-paramétrique et donc une estimation plus précise de la demande, les largeurs des fenêtres de Parzen doivent être calibrées. Les fenêtres de Parzen correspondent aux plages de variation des paramètres retenus, ceux-ci caractérisent la similarité des conditions climatiques et comportementales au moment de la prédiction de la demande par rapport à l'historique mémorisé.

Le vecteur des paramètres optimaux $h=[h_1 \dots h_d]$, est obtenu en minimisant l'erreur quadratique moyenne sur la prédiction de la demande pour une semaine ; la fonction à minimiser est définie par la relation suivante :

$$\min_h \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - \hat{y}_i(x_i, h)\right]^2 \quad \textbf{(13)}$$

La minimisation de la fonction précédemment définie est réalisée à l'aide d'un algorithme génétique qui présente l'avantage d'être une méthode stochastique d'ordre 0 : seules les valeurs de la fonction à optimiser aux points d'échantillonnage sont requises (pas de dérivées). Cette technique d'optimisation est ainsi utilisable pour les fonctions très irrégulières ou complexes à calculer. Cependant elle présente un coût calculatoire qui peut devenir important. Les méthodes standard sont

certes plus rapides en temps de calcul, mais elles se retrouvent piégées dans les optima locaux, alors que les AGs sont efficaces pour les vastes espaces de recherche et les fonctions irrégulières.

2.3.1 Algorithme Génétique (AG)

Les AGs ont été construits de façon à imiter et reproduire le comportement d'une population d'organismes vivants qui s'adaptent à leur environnement par l'intermédiaire des mécanismes naturels que sont la sélection et l'héritage génétique. Le principe de fonctionnement global des AGs est illustré sur la **Figure *1***, (Krarti, 2003; Lutton, 2006) l'expliquent de manière plus détaillée. Le processus global de fonctionnement d'un AG repose sur une boucle générationnelle d'individus menant chacun à une solution d'optimisation, les principales étapes sont les suivantes. Tout d'abord, l'AG doit être initialisé : la génération de la première population est habituellement aléatoire, cependant des connaissances préalables sur le problème peuvent être introduites dans cette première étape. Chaque individu de la population représente une solution au problème et est caractérisé par un chromosome codé en langage binaire. Le nouvel ensemble de solutions, la population d'individus « enfants », est généré par les chromosomes « parents » par des processus de croisement et de mutation. Alors que l'évolution par croisement passe par la combinaison des chromosomes de deux parents, la mutation, elle, consiste en de légères modifications d'un unique chromosome. La sélection des individus d'une population aptes à évoluer est réalisée sur la base d'une fonction d'évaluation appelée « fitness ». Le paramètre principal de cette étape de sélection est nommé pression sélective : c'est le quotient de la probabilité de sélection de chaque individu sur la probabilité de sélection de l'individu moyen de la population courante. L'évolution de la population est arrêtée lorsqu'on a atteint le seuil de convergence ou bien le nombre maximum de générations.

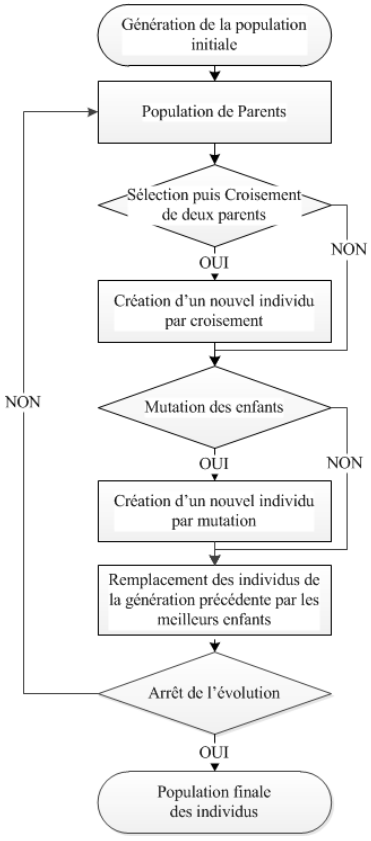

Figure 1 Schéma d'évolution d'un AG

3 Simulation sur données synthétiques

L'algorithme de prédiction de la demande énergétique est testé dans un premier temps sur des données synthétiques d'appel de puissance issues de la simulation d'un modèle de maison typique québécoise développé sur EnergyPlus.

3.1 Description du bâtiment simulé

Le bâtiment modélisé sur EnergyPlus est une maison québécoise typique dont l'architecture est basée sur celle des maisons d'expérimentation construites au LTE. Il s'agit d'un cottage avec trois chambres à coucher, deux salles de bain et un sous-sol habitable.

Le bâtiment a une surface au sol de 53 m^2 et une surface chauffée totale de 160 m^2. La surface totale de fenestration est de 19 m^2 et le bâtiment est orienté avec sa façade plein nord. Le niveau d'isolation de la maison virtuelle correspond à celui d'une habitation construite après l'entrée en vigueur du code de 1983.

Le système de chauffage du bâtiment résidentiel est composé intégralement de plinthes électriques situées dans les pièces principales. Les systèmes de chauffage sont contrôles par des thermostats électroniques qui maintiennent, dans un premier scénario, une température de consigne de 22 °C toute la journée hormis entre 5h et 8h et entre 16h et 24h en semaine où la température est descendue à 18 °C. En fin de semaine, la température de consigne reste de 22 °C toute la journée. Dans un deuxième scénario simulé, la température de consigne est maintenue constante à 22 °C toute la journée, toute la semaine. Les gains internes dus à l'occupation du bâtiment, à l'éclairage, ou aux autres équipements (cuisinière, téléviseur, ordinateur, chauffe-eau…) ne sont pas incluses dans les données synthétiques ; le modèle du bâtiment EnergyPlus est seulement utilisé pour déterminer la demande énergétique en chauffage.

3.2 Prédiction de la demande pour le bâtiment modélisé

La simulation est effectuée sur une période de trois ans : entre 1995 et 1997, la maison virtuelle est supposée être implantée à Montréal. Les données climatiques proviennent d'une station météorologique située à l'aéroport Pierre Elliott Trudeau à Montréal, disponibles sur le site internet : www.simeb.ca. Le modèle développé sur EnergyPlus fournit les données de consommation énergétique pour le chauffage du bâtiment sur les trois années considérées pour les deux types de scénarios présentés précédemment.

Sur les trois années de données climatiques et d'appels de puissance énergétique, deux années sont utilisées comme historique pour le modèle de prédiction de la demande. Les largeurs des fenêtres de Parzen sont ensuite calibrées à l'aide de l'AG de manière à minimiser l'erreur de prédiction sur la semaine d'apprentissage suivant les deux années d'historique. Enfin une prédiction de la demande énergétique pour le chauffage de la maison est réalisée jour par jour sur l'année suivant la semaine d'apprentissage. Les prédictions sont comparées aux demandes énergétiques obtenues par simulation pour validation.

Scénario 1 : abaissement de la température de consigne pendant les périodes d'absence et la nuit

Une optimisation des paramètres est réalisée à l'aide de l'AG sur la première semaine du mois de janvier. Les largeurs des fenêtres de Parzen après calibration sont présentées dans le **Tableau 1**.

Tableau 1 Largeurs optimales des fenêtres de Parzen (Bâtiment modélisé - Scénario 1)

h1 [°C]	3.4	**h9 [°C]**	8.06
h2 [%]	100	**h10 [°C]**	16.02

h3 [km/h]	406	**h11 [°C]**	18.67
h4 [h]	0.01	**h12 [°C]**	1.43
h5 [h]	1.41	**h13 [W]**	666.76
h6 [j]	3.4	**h14 [W]**	10000
h7 [j]	0.1	**h15 [W]**	2666.74
h8 [j]	6.4	**h16 [W]**	4666.72

Les valeurs numériques des largeurs des fenêtres de Parzen attribuées par l'AG suite à l'optimisation peuvent être expliquées « physiquement » : la largeur de la fenêtre peut être vue comme un indice de l'importance donnée au paramètre. En effet, plus la largeur de la fenêtre du paramètre sera élevée et moins le paramètre aura d'importance puisque le nombre de voisins sera plus grand. Par exemple, si l'on considère les paramètres climatiques que sont la température, le taux d'humidité et la vitesse du vent, d'après les valeurs du **Tableau 1** seul l'effet de la température est considéré. Ce résultat est justifié par le fait que l'appel de puissance pour le chauffage est plus dépendant de la température que des autres paramètres climatiques.

De la même manière, l'analyse du **Tableau 1** montre que la température sol-air du jour précédent ($\mathbf{h_{12}}$) est impactante mais que celles des deux jours précédents ($\mathbf{h_{10}}, \mathbf{h_{11}}$) peuvent être négligées. Pour les paramètres représentatifs de l'heure du jour et du jour de la semaine ($\mathbf{h_4} \dots \mathbf{h_8}$), ceux-ci doivent être conservés en intégralité, de même pour les paramètres décrivant les puissances moyennes appelées les trois jours précédents et l'heure précédente ($\mathbf{h_{13}} \dots \mathbf{h_{16}}$).

Les **Figure *2***, **Figure *3*** et **Figure *4*** présentent les résultats de la prédiction jour par jour de la demande énergétique pour le chauffage du bâtiment résidentiel sur une période d'une semaine respectivement au cours des mois de janvier, avril et octobre 1997. L'analyse du profil d'appel de puissance prédit montre que la calibration des paramètres sur la première semaine de janvier permet une bonne estimation de l'appel de puissance sur les semaines suivantes au cours de l'année. D'un point de vue général, le profil prédit possède une allure très similaire au profil réel pour le scénario considéré. On peut noter que les pics d'appel de puissance correspondant au redémarrage du système de chauffage sont estimés avec une erreur relative moyenne de 6.5 % sur la semaine de janvier, 35.6 % pour avril et 72.6 % pour octobre. L'erreur quadratique moyenne sur la semaine complète de janvier est de 31.4 %, 15.7 % pour avril et 13.5 % pour octobre.

Ainsi lorsque l'on s'éloigne de la semaine d'apprentissage, on devient beaucoup moins précis sur la prédiction de la demande, cependant on est quand même capable de garder une bonne estimation de l'allure du profil de la demande sur la semaine complète y compris la fin de semaine.

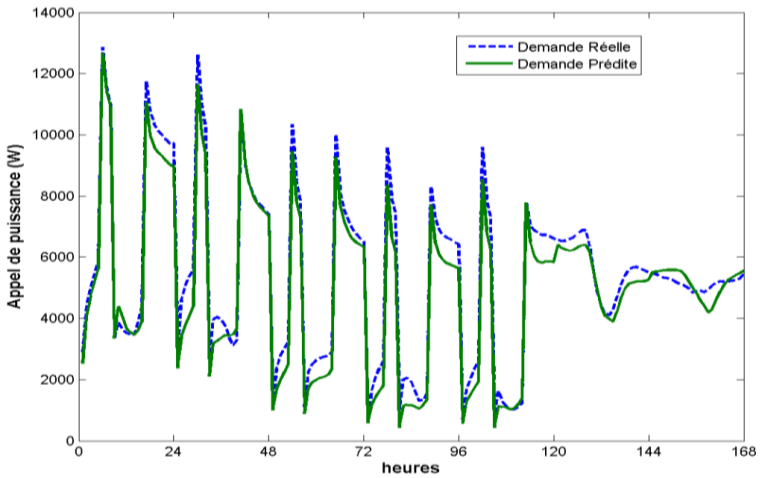

Figure 2 Profil prédit dans le scénario 1 (semaine du mois de janvier 1997)

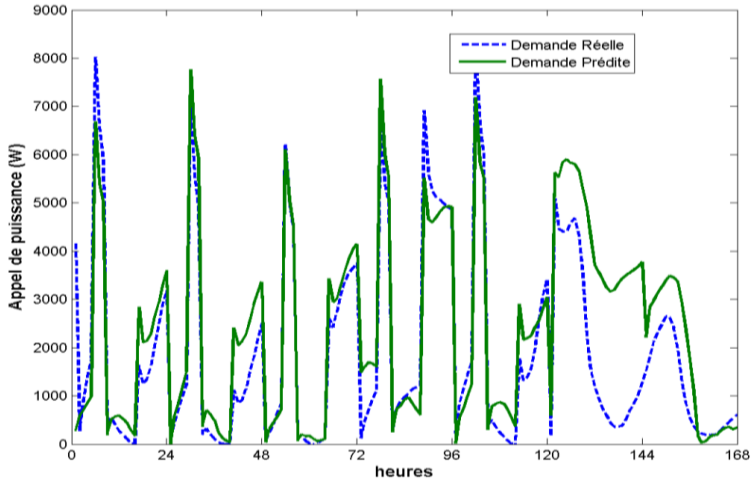

Figure 3 Profil prédit dans le scénario 1 (semaine du mois d'avril 1997)

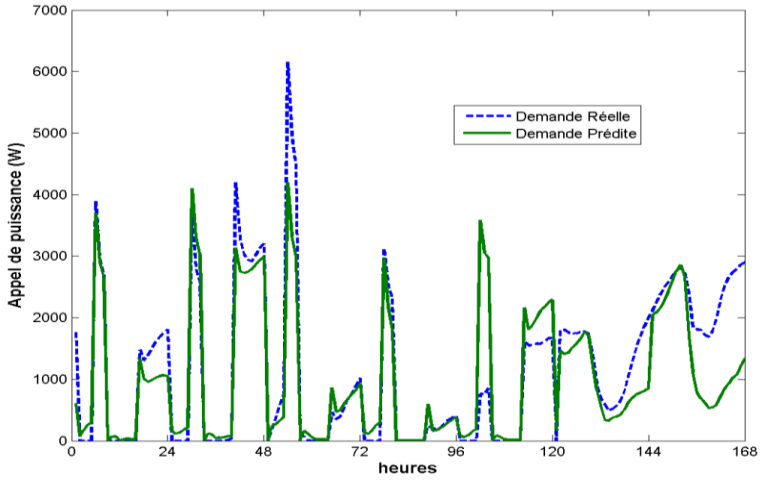

Figure 4 Profil prédit dans le scénario 1 (semaine du mois d'octobre 1997)

Scénario 2 : température de consigne constante

Une méthodologie identique est appliquée dans le cas du scénario 2 : on calibre tout d'abord les paramètres avec l'AG, ceux-ci sont présentés dans le **Tableau 2**. Les résultats de la prédiction de

la demande énergétique sont présentés sur les **Figure *5*, Figure *6*** et **Figure *7***. Ces figures correspondent aux demandes énergétiques d'une semaine au cours du mois de janvier, avril et octobre 1997 pour le chauffage d'un bâtiment résidentiel dans le cas du scénario 2.

Pour ce scénario, l'allure du profil d'appel de puissance prédit est encore très similaire au profil réel. L'erreur quadratique moyenne sur la semaine complète de janvier est de 20.3 %, 19.5 % pour avril et 20.0 % pour octobre.

Tableau 2 Largeurs optimales des fenêtres de Parzen (Bâtiment modélisé - Scénario 2)

h1 [°C]	1.42	**h9 [°C]**	18.67
h2 [%]	47.2	**h10 [°C]**	17.35
h3 [km/h]	340	**h11 [°C]**	18.67
h4 [h]	2.2	**h12 [°C]**	8.06
h5 [h]	1.8	**h13 [W]**	666.76
h6 [j]	6.4	**h14 [W]**	7333.36
h7 [j]	8.02	**h15 [W]**	3333.4
h8 [j]	4.6	**h16 [W]**	666.76

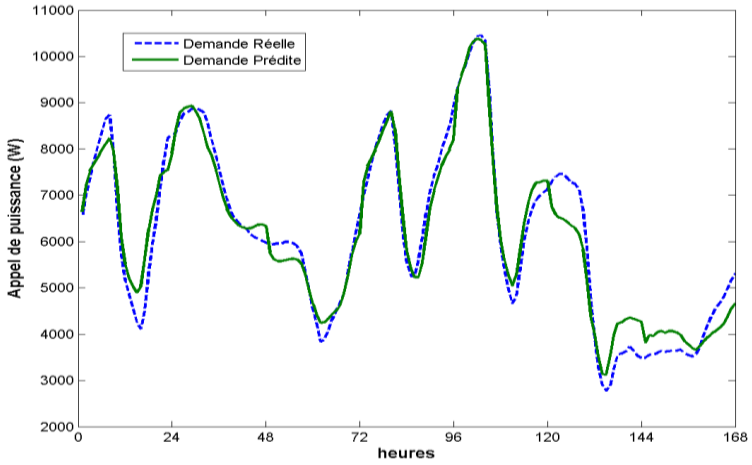

Figure 5 Profil prédit dans le scénario 2 (semaine du mois de janvier 1997)

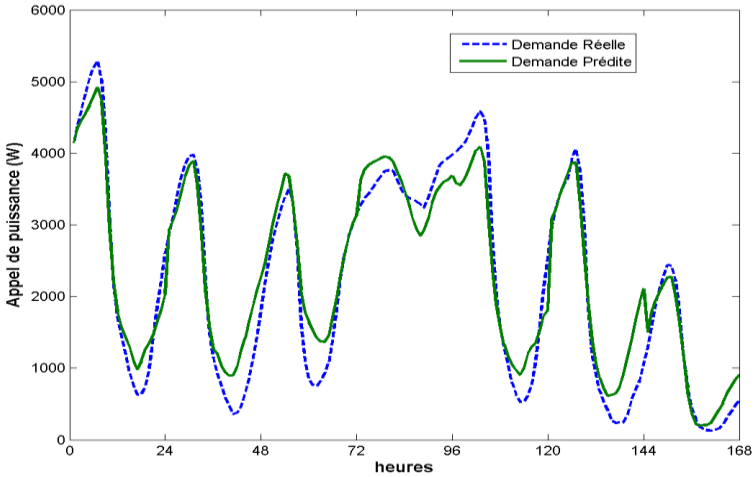

Figure 6 Profil prédit dans le scénario 2 (semaine du mois d'avril 1997)

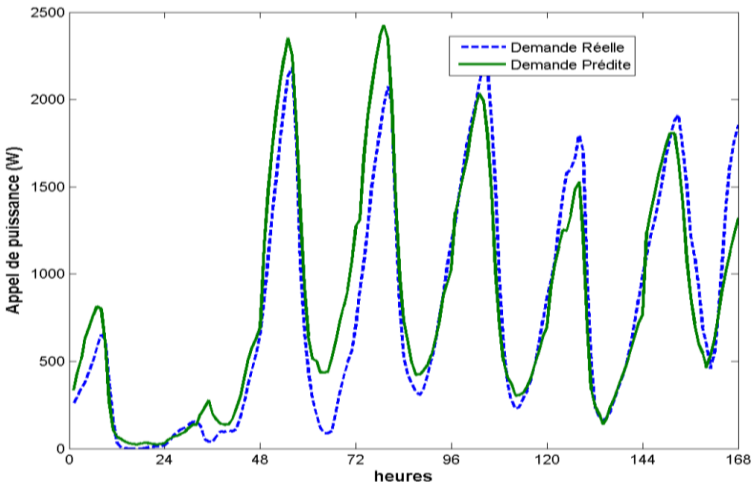

Figure 7 Profil prédit dans le scénario 2 (semaine du mois d'octobre 1997)

4 Simulation sur données réelles

L'algorithme de prédiction de la demande énergétique est utilisé dans un deuxième temps sur des données réelles de consommation énergétique issues de mesures sur des habitations de la région de Montréal.

4.1 Description du bâtiment, objet de la prédiction

Les données de consommation énergétique utilisées pour tester le modèle dans des conditions réelles proviennent du projet d'établissement des profils de consommation (PEPC). Les mesures ont été effectuées aux quinze minutes sur vingt-neuf maisons pendant un an, en 1994, avant qu'elles ne reçoivent des thermostats électroniques et pendant un an et demie à partir de 1995. Des mesures ont également été effectuées sur vingt-huit autres maisons pour lesquelles la période de pré-installation des thermostats a duré un à deux mois. L'équipe de projet d'établissement des profils de consommation avait identifié les mesures potentiellement défectueuses et 0.2 % sur la consommation totale ont été rejetés. Ainsi, les données d'appel de puissance de trente-trois habitations dans la région de Montréal sont disponibles. Les données climatiques employées proviennent de la station météorologique située à l'aéroport Pierre Elliott Trudeau à Montréal, elles sont disponibles sur le site internet : www.simeb.ca à partir de 1995, et les données de l'année 1994 peuvent être extraites mensuellement à partir du site internet : www.climat.meteo.gc.ca.

4.2 Prédiction de la demande énergétique

Dans cette section, les résultats de prédiction de la demande énergétique d'une habitation résidentielle parmi les 29 maisons du projet PEPC seront présentés en détail. La même méthodologie pour l'estimation de la demande que pour les données synthétiques est utilisée : deux années de données sont employées comme historique et les paramètres sont optimisés à l'aide de l'AG pour une erreur de prédiction minimale sur la première semaine du mois de janvier. Les largeurs des fenêtres de Parzen après calibration sont présentées dans le Tableau 3.

Tableau 3 Largeurs optimales des fenêtres de Parzen (Bâtiment réel)

h1 [°C]	3.4	**h9 [°C]**	20
h2 [%]	53.8	**h10 [°C]**	20
h3 [km/h]	538	**h11 [°C]**	9.39

h4 [h]	0.81	**h12 [°C]**	1.43
h5 [h]	2.2	**h13 [kW]**	666.76
h6 [j]	7.6	**h14 [kW]**	6666.7
h7 [j]	3.4	**h15 [kW]**	9333.34
h8 [j]	5.2	**h16 [kW]**	0.1

Les **Figure *8***, **Figure *9*** et **Figure *10*** présentent les résultats de l'estimation jour par jour de la demande énergétique pour le chauffage du bâtiment résidentiel sur une période d'une semaine respectivement au cours des mois de janvier, février et mars 1997. L'analyse des figures présentant les résultats des estimations permet de retrouver des conclusions similaires aux simulations sur données synthétiques. En effet, l'allure générale du profil d'appel de puissance prédit est très similaire au profil réel. On retrouve également la perte de précision dans l'estimation au fur et à mesure que l'on s'éloigne de la semaine d'apprentissage sur laquelle l'optimisation a été réalisée.

Le **Tableau 4** présente l'erreur relative moyenne et l'erreur quadratique moyenne normalisée pour les semaines des mois de janvier, février et mars.

Tableau 4 Erreur de prédiction de la demande énergétique (Bâtiment réel)

Erreur de Prédiction	MAPE	NRMSE
Semaine de janvier	10.1 %	20.5 %
Semaine de février	22.7 %	17.3 %
Semaine de mars	27 %	15.5 %

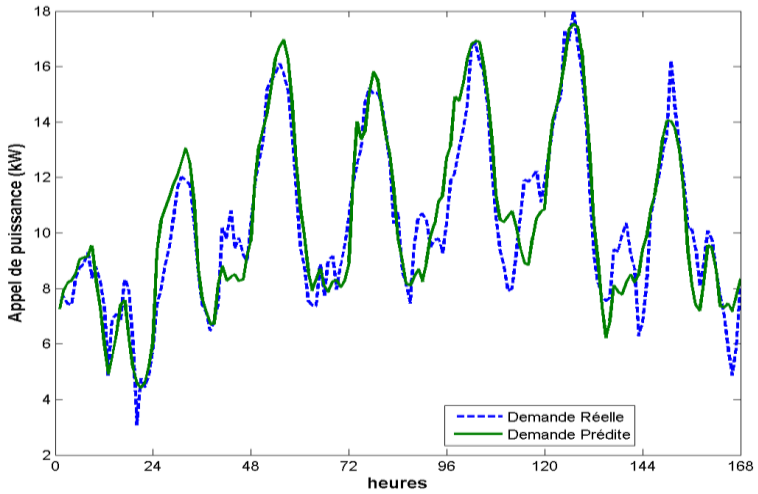

Figure 8 Profil prédit : semaine du mois de janvier 1997

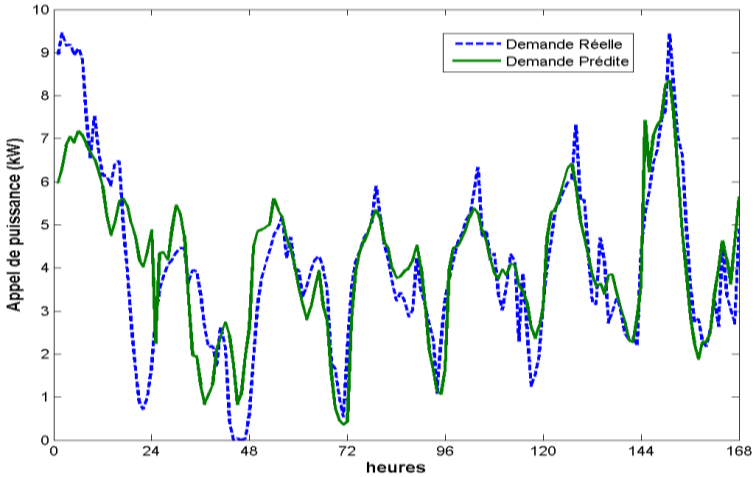

Figure 9 Profil prédit : semaine du mois de février 1997

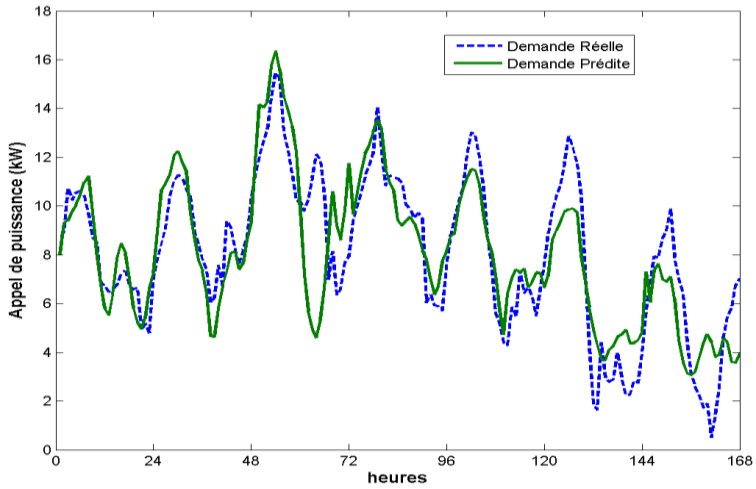

Figure 10 Profil prédit : semaine du mois de mars 1997

Il est à noter qu'un certain pourcentage d'erreur doit être attribué au fait qu'un historique de deux années de mesures a été utilisé ; alors que l'allure de l'appel de puissance a été légèrement modifié suite au changement des thermostats ou bout d'un an et demi. Afin d'éviter cette erreur de prédiction due au changement des thermostats, un test a été réalisé en utilisant une seule année de données d'historique. Cependant, cet échantillon de données s'est révélé de taille trop restreinte et n'est donc pas adapté pour cette technique d'estimation non-paramétrique. Elle présente en effet des limites : l'échantillon est trop pauvre et il n'est pas assez représentatif de l'étendue de variation des paramètres de similitudes, les estimations de la demande se révèlent donc souvent impossibles.

D'autre part, il faut prendre en compte que les données climatiques utilisées pour l'estimation de la demande sont des données réelles mesurées. Or, dans le cas d'un modèle utilisé pour de la prédiction en temps réel, les données météorologiques employées seraient alors des estimations ce qui impliquerait un pourcentage d'erreur supplémentaire sur la prédiction de la demande énergétique.

5 Conclusions

Cet article présente une méthode inverse d'estimation de la demande énergétique à court terme par régression non paramétrique pour le chauffage d'un bâtiment résidentiel. La méthodologie présentée permet d'établir une prédiction sur les vingt-quatre prochaines heures en utilisant les données de consommation horaire et climatiques des deux années précédentes. Un algorithme génétique est également employé afin de calibrer les fenêtres de variation des paramètres caractéristiques de la similitude de manière à minimiser l'erreur de prédiction. Cette méthodologie a été testée sur des données synthétiques et des données réelles mesurées. Elle a montré une capacité à estimer la demande énergétique avec précision dans plusieurs scénarios, autant d'un point de vue similitude de l'allure du profil que d'estimation des pics d'appel de puissance. Cependant elle présente une diminution de la précision de la prédiction au fur et à mesure que l'on s'éloigne de la semaine d'apprentissage. Ainsi, cette technique de modélisation présente plusieurs avantages : elle ne requiert pas de connaissances physiques préalables sur le bâtiment résidentiel dont on cherche à estimer la demande. Elle permet d'extraire de l'information à partir des données et elle est particulièrement efficace dans les cas où le nombre d'observations est important ainsi que le nombre de variables explicatives. Cependant, ce type de modèle requiert la disponibilité d'un grand nombre de données pour de plus solides résultats. De plus, c'est une modélisation de type boîte noire : il est donc difficile de donner une explication physique à une valeur d'estimation de la demande.

Des travaux futurs chercheront à améliorer la prédiction par l'utilisation d'un outil de classification permettant d'obtenir les profils typiques d'appel de puissance pour chaque jour de la semaine. Ainsi, l'allure d'une semaine typique obtenue en termes de profil d'appel de puissance pourra être utilisée pour calibrer les paramètres avec l'algorithme génétique. D'autre part, le modèle développé pourrait devenir adaptatif et gagner en expérience si l'on ajoute progressivement à l'historique des données, les appels de puissance horaire et que l'on recalibre périodiquement les largeurs des fenêtres de Parzen pour tenir compte des nouvelles données.

6 Références

Agarwal, V., Bougaev, A., Tsoukalas, L., Kollias, S., Stafylopatis, A., Duch, W., & Oja, E. (2006). Kernel Regression Based Short-Term Load Forecasting. In A. N. N. I. 2006 (dir.). (Vol. 4132, pp. 701-708): Springer Berlin / Heidelberg.

Alfares, H. K., & Nazeeruddin, M. (2002). Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science, 33*(1), 23-34.

American Society of Heating Refrigerating and Air-Conditioning Engineers. (2009). *2009 ASHRAE handbook - Fundamentals* (Inch-pound^{e} éd.). Atlanta, GA.: ASHRAE.

Asber, D., Lefebvre, S., Asber, J., Saad, M., & Desbiens, C. (2007). Non-parametric short-term load forecasting. *29*(8), 630-635.

Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally Weighted Learning. *Artificial Intelligence Review, 11*(1), 11-73.

Charytoniuk, W., Chen, M. S., & Van Olinda, P. Y. A. (1998). Nonparametric regression based short-term load forecasting. *IEEE Transactions on Power Systems, 13*(3).

Duda, R., Hart, P., & Stork, D. (2001). *Pattern Classification (2nd Edition)*: Wiley-Interscience.

Feinberg, E. A., Genethliou, D., Chow, J. H., Wu, F. F., & Momoh, J. (2005). Load Forecasting. In A. M. f. R. E. P. Systems (dir.). (pp. 269-285): Springer US.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *199*(3), 902-907.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. Y. F. (2001). Neural networks for short-term load forecasting : a review and evaluation. *IEEE Transactions on Power Systems, 16*(1).

Kramer, O., Satzger, B., Lässig, J., Graña Romay, M., Corchado, E., & Garcia Sebastian, M. (2010). Power Prediction in Smart Grids with Evolutionary Local Kernel Regression. In H. A. I. Systems (dir.). (Vol. 6076, pp. 262-269): Springer Berlin / Heidelberg.

Krarti, M. (2003). An Overview of Artificial Intelligence-Based Methods for Building Energy Systems. *Journal of Solar Energy Engineering, 125*(3), 331-342.

Kyriakides, E., Polycarpou, M., Chen, K., & Wang, L. (2007). Short Term Electric Load Forecasting: A Tutorial. In T. i. N. Computation (dir.). (Vol. 35, pp. 391-418): Springer Berlin / Heidelberg.

Lutton, E. (2006). Algorithmes génétiques et algorithmes évolutionnaires. *S1*(S7218).

Martinez, W. (2001). *Computational statistics handbook with MATLAB*: Chapman & Hall/CRC.

Nadaraya, E. (1964). On estimating regression. *Theory of Prob. and Appl., 9*, 141-142.

Watson, G. (1969). Smooth regression analysis. *Sankhya, Series, A*(26), 359-372.