Prévision de la consommation d'électricité par correction itérative du biais

Journées MAS 01/09/2010

V. Lefieux, L. Maillard-Teyssier
RTE / DMA

En collaboration avec
P. A. Cornillon, N. Hengartner, E. Matzner-Løber

Plan

Contexte

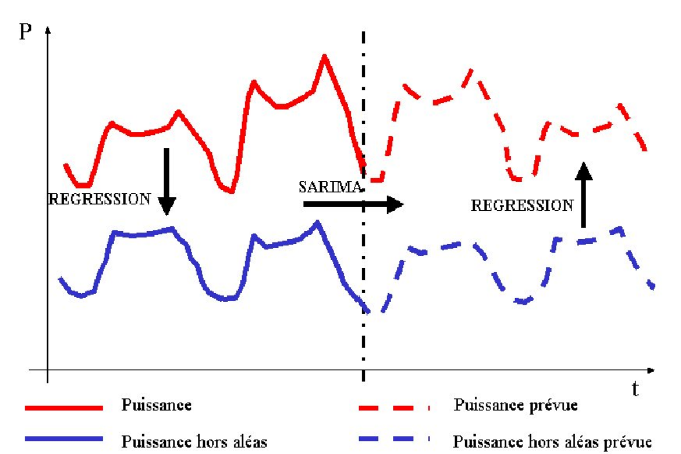
Les enjeux

- L'électricité ne se stocke pas à l'échelle industrielle
- A tout instant, la production d'électricité doit être égale à celle qui est consommée (fréquence de 50 Hz)

La solution adoptée

- Prévision de la "courbe de charge" nationale (ou régionale) : caractère autorégressif et variables exogènes (ex : météo, tarifs et prix)
- Juxtaposition de deux modèles paramétriques (régression et SARIMA)
- Problèmes : spécification du modèle, intervalles de confiance...

Principe de la prévision a court terme



Etat de l'art

- Modélisation paramétrique : rigide et pas adaptée à la structure de certaines séries temporelles
- Modélisation non-paramétrique : fléau de la dimension
- D'autres points de vue : méthode de réduction de la dimension (semi-paramétrique), correction itérative du biais

Lutter contre le "fléau de la dimension"

Approximation fonctionnelle

- Les *modèles additifs généralisés*, étudiés entre autres par Auestad *et al.* (1990), Härdle *et al.* (1996).
  En notant $Y \in \mathbb{R}$ la variable à expliquer et $X = \left(X^1, \ldots, X^p\right)^T \in \mathbb{R}^p$ les p variables explicatives, on suppose que :
  $$\mathbb{E}\left(Y/X\right) = \sum_{j=1}^{p} g_j\left(X^j\right)$$
  où $g_j : \mathbb{R} \to \mathbb{R}$ sont p fonctions mesurables inconnues.
- Les *modèles autorégressifs à coefficients fonctionnels*, étudiés par Chen *et al.* (1993).
  On suppose que :
  $$X_t = f_1\left(X^*_{t-1}\right) X_{t-1} + \ldots + f_r\left(X^*_{t-1}\right) X_{t-r} + \varepsilon_t$$
  où $f_i : \mathbb{R} \to \mathbb{R}$ sont r fonctions mesurables inconnues et $X^*_{t-1} = \left(X_{t-i_1}, \ldots, X_{t-i_k}\right)^T$ (avec $i_j \in \mathbb{N}^*$ pour $j \in \{1, \ldots, k\}$).

Lutter contre le "fléau de la dimension"

Réduction de la dimension 1/2

- Les modèles de *"Projection Pursuit Regression"* (PPR), étudiés entre autres par Huber (1985), Hall (1989), Chen (1991), Xia *et al.* (1997).
  On suppose que :
  $$\mathbb{E}\left(Y/X\right) = \sum_{j=1}^{D} g_j\left(\beta_j^T X\right)$$
  où $g_j : \mathbb{R} \to \mathbb{R}$ sont D fonctions mesurables inconnues et $\beta_j \in \mathbb{R}^p$ sont D paramètres (vectoriels) inconnus.
- Les *modèles à direction révélatrice unique* (SIM : Single Index Model), étudiés entre autres par Härdle *et al.* (1993), Ichimura (1993), Carroll *et al.* (1997), Delecroix *et al.* (1999), Xia *et al.* (1999), Delecroix *et al.* (2006).
  On suppose que :
  $$\mathbb{E}\left(Y/X\right) = g\left(\beta^T X\right)$$
  où $g : \mathbb{R} \to \mathbb{R}$ est une fonction mesurable inconnue et $\beta \in \mathbb{R}^p$ est un paramètre (vectoriel) inconnu.

Lutter contre le ”fléau de la dimension”

Réduction de la dimension 2/2

- Les *modèles à directions révélatrices multiples*, étudiés entre autres par Li *et al.* (2001), Xia *et al.* (2002), Li *et al.* (2004).
  On suppose que :
  $$\mathbb{E}\left(Y/X\right) = g\left(B^T X\right)$$
  où $g : \mathbb{R}^D \to \mathbb{R}$ est une fonction mesurable inconnue et $B = \left(\beta^1, \ldots, \beta^D\right)_{p \times D}$ est un paramètre (matriciel) inconnu.

Modèle

On se place dans un contexte de régression.
Supposons que les données $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ sont liées via le modèle de régression

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

$m(\cdot)$ est la fonction inconnue.
Les erreurs sont indépendantes, de moyenne nulle et de variance σ^2.
Ce modèle peut s'écrire sous forme matricielle

$$Y = m + \varepsilon.$$

Lisseurs

Un lisseur linéaire s'écrit simplement sous la forme

$$\widehat{m} = SY.$$

S est matrice de lissage de taille $n \times n$.
$\widehat{m} = \widehat{Y} = (\widehat{Y}_1, \dots, \widehat{Y}_n)^t$, sont les valeurs ajustées.

Le lisseur est paramétré par un paramètre de lissage λ (la fenêtre pour un lisseur à noyau, la pénalité pour les splines de lissage...).
Nous écrirons le lisseur de façon à qu'une valeur élevée de λ corresponde à un estimateur très lisse.
Il faut sélectionner le paramètre λ.

Lisseurs classiques

- Moyenne mobile : $S_{ij} = 1/\text{nbr de } X \text{ dans le voisinage}$
- Lisseur par cellule : $S_{ij} = 1/\text{nbr de } X \text{ dans la cellule}$
- Lisseur à noyau : $S_{ij} = K_h\left(X_i - X_j\right) / \sum_l K_h\left(X_i - X_l\right)$
- K-pp voisins : $S_{ij} = 1/K \text{ si } X_j \in K - pp(X_i)$
- Spline de régression : $S = B\left(B'B\right)^{-1} B'$
- Spline de lissage : $S = N(N'N + \lambda\Omega_N)^{-1}N'$

Méthode IBR 1/2

Nous partons d'un estimateur lisse (λ grand), et calculons les résidus à cette première étape et nous corrigeons l'estimateur initial ou pilote en enlevant les résidus lissés (par simplicité nous conservons le même lisseur à chaque étape) :

$$\begin{aligned} R_1 &= Y - \widehat{m}_1 = (I - S_1)Y \\ \widehat{m}_2 &= \widehat{m}_1 + S_1 R_1 \\ &= (S_1 + S_1(I - S_1))Y. \end{aligned}$$

Nous itérons cette procédure et obtenons à l'étape k :

$$\begin{aligned} \widehat{m}_k &= S_1 Y + S_1(I - S_1)Y + \cdots + S_1(I - S_1)^{k-1} Y \\ &= (I - (I - S_1)^k)Y \end{aligned}$$

On peut montrer que le comportement de l'estimateur dépend des valeurs singulières de $I - S_1$.

Méthode IBR 2/2

L'estimateur peut également s'écrire sous la forme

$$\widehat{m}_k = S_1[Y + (I - S_1)Y + \cdots + (I - S_1)^{k-1}Y].$$

Le problème maintenant consiste à choisir k.

Critères d'arrêt classiques

$$\hat{k}_{AIC} = \arg\min_{k \in \mathcal{K}} \left\{ \widehat{\sigma}^2 + 2\frac{trace\,(S_k)}{n} \right\},$$

$$\hat{k}_{AIC_C} = \arg\min_{k \in \mathcal{K}} \left\{ \log \widehat{\sigma}^2 + 1 + \frac{2\,(trace\,(S_k) + 1)}{n - trace\,(S_k) - 2} \right\},$$

$$\hat{k}_{GCV} = \arg\min_{k \in \mathcal{K}} \left\{ \log \widehat{\sigma}^2 - 2 \log \left(1 - \frac{trace\,(S_k)}{n} \right) \right\}.$$

Propriétés théoriques

Le lisseur obtenu avec des splines plaques minces et avec GCV admet des propriétés d'adaptativité.
Cornillon P. A., Hengartner N., Matzner-Løber E. (2009) Recursive Bias Estimation for high dimensional regression smoothers, soumis.
Le problème des splines plaques minces est que le nombre minimum de valeurs propres égales à 1 est fonction de la dimension d.

Formalisation du problème

Dans notre problème, nous souhaitons prévoir en premier lieu la série des consommations corrigées des aléas climatiques et tarifaires pour laquelle on dispose de n observations $\{X_t\}_{t=1}^n$:

$$X_{t+l} = f_l(X_t, \ldots, X_{t-47}) + \varepsilon_t$$

où $f_l : \mathbb{R}^{48} \mapsto \mathbb{R}$ est une fonction inconnue dépendant de l'horizon l (72 au maximum).
Pour estimer cette fonction, nous utilisons la méthode des noyaux.

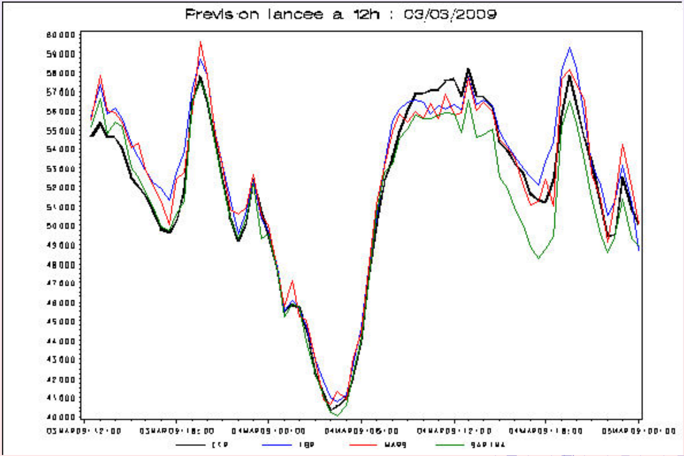
Formalisation du problème

Nous utilisons le noyau produit gaussien :

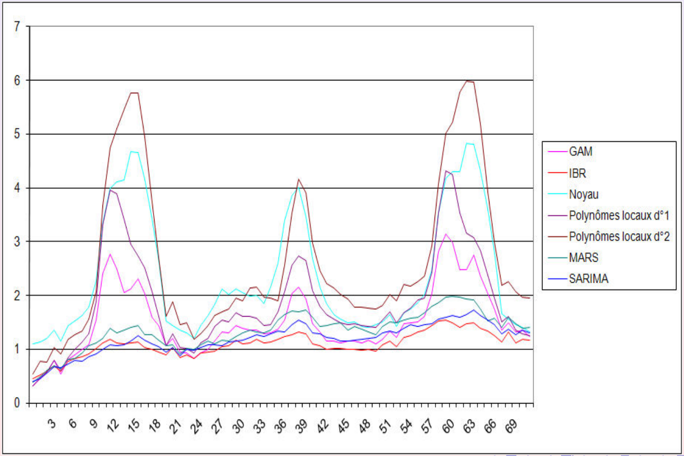
$$K_h(x - X_{i.}) = \prod_{j=1}^{48} \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{(x_j - X_{ij})^2}{h_j} \right\}$$

où $X_{i.}$ représente la i-ème ligne de la matrice X et $h_j \in \mathbb{R}^+$ est la largeur de fenêtre pour la variable X_j.
Pour choisir cette valeur, nous fixons pour chaque lisseur univarié un ddl (le même en général pour chaque variable).

Exemple : prévisions SARIMA, MARS et IBR lancées le 03/03/2009 à 12h



Quelques résultats : MAPE des prévisions lancées du 10 janvier 2009 au 20/12/2009 à 12h



Variables explicatives

- Données calendaires : type de jour, jours fériés et particuliers, vacances
- Données météorologiques : température, nébulosité, vent
- Pas de temps : 10 min, 30 min, 1h
- Maille géographique : stations, points de grille

Sélection des variables explicatives

- Dans IBR : procédure Forward avec un critère de type AIC.
- Autre piste en cours d'exploration : utilisation de dictionnaires (bases de Fourier, de Haar, ondelettes...) et de procédures efficaces en grande dimension (Dantzig, LOL, Lasso...)