Utilisation d'estimateurs en plusieurs étapes appliqués à des modèles additifs modélisant la prévision de consommation électrique

Vincent Thouvenot [1] & Anestis Antoniadis [2] & Xavier Brossat [1,3] & Yannig Goude [1,4] & Jean Michel Poggi [5]

[1] *1 avenue Général de Gaulle, 92140 Clamart, France, e-mail: vincent.thouvenot@edf.fr*
[2] *Université Joseph Fourier Tour IRMA, B.P.53, 38041 Grenoble CEDEX 9, France, e-mail: Anestis.Antoniadis@imag.fr* [3] *e-mail: xavier.brossat@edf.fr*
[4] *e-mail: yannig.goude@edf.fr*
[5] *Université Paris Descartes et Université Paris-Sud, Bâtiment 425 91405 Orsay Cedex, France, e-mail: Jean-Michel.Poggi@math.u-psud.fr*

Résumé. La prévision de consommation d'électricité est une activité majeure pour EDF. Afin de satisfaire la demande à chaque instant et d'ajuster le planning de production, EDF se doit d'estimer avec une grande précision la consommation future de son portefeuille de clients. L'ouverture du marché de l'électricité en France, l'évolution des usages (nouvelles technologies, nouvelles normes de construction, ...) ainsi que le développement des smart grid induisent de nouveaux problèmes de prévisions. Des questions statistiques se posent telles que la calibration d'un grand nombre de modèles (prévision pour le réseau de distribution) ou l'inclusion de nouvelles variables explicatives dans des modèles additifs. L'objectif de ce travail est de présenter des procédures automatiques de sélection et d'estimation de composantes d'un modèle additif. Nous utilisons du Group LASSO, qui est, sous certaines conditions, consistant en sélection, et des P-Splines, qui sont consistantes en estimation. Les procédures sont illustrées sur des applications pratiques.

Mots-clés. Group LASSO, Estimateurs en plusieurs étapes, Modèle Additif, Prévision de charge électrique, P-Splines

Abstract. French electricity load forecasting encounters major changes since the past decade. These changes are, among others things, due to the opening of electricity market (and economical crisis), which asks development of new automatic time adaptive prediction methods. The advent of innovating technologies also needs the development of some automatic methods, because we have to study some thousands or tens of thousands chronological series. In this paper, we adopt for time prediction a semi-parametric approach based on additive models. We present an automatic procedure for covariate selection in a additive model. We combine Group LASSO, which is selection consistent, with P-Splines, which are estimation consistent. Real applications are provided.

Keywords. Additive Model, Group LASSO, Load Forecasting, Multi-step estimator, P-Splines,

1 Modèle additif

Comme l'électricité ne se stocke pas, ou très difficilement, EDF doit toujours équilibrer sa production d'électricité avec la consommation de celle-ci. Ceci amène à un besoin de modélisation et de prévision de consommation d'électricité à différents horizons.

Pour répondre à la question de prévision de consommation à court et moyen terme, nous avons choisi d'utiliser des modèles additifs, introduits par Hastie et Tibshirani [3], parce qu'ils réalisent un bon compromis entre complexité et flexibilité.

$$E\left(Y|(X_1,\ldots,X_d)=(x_1,\ldots,x_d)\right)=\beta_0+\sum_{i=1}^{d} f_i(x_i), \tag{1}$$

où Y est la variable à expliquer (par exemple la charge consommée), $(X_1,\ldots,X_d)$ les variables explicatives (par exemple la température, le moment dans l'année,...), f_i la composante associée à la *ieme* variable explicative, β_0 une constante.

Nous utilisons des bases de B-Splines pour approcher localement les composantes f_i. Cette projection permet de nous ramener au cas du modèle linéaire. En effet, le modèle (1) peut être approché par :

$$E\left(Y|(X_1,\ldots,X_d)=(x_1,\ldots,x_d)\right)=\beta_0+\sum_{j=1}^{d}\sum_{k=1}^{m_j}\beta_{j,k}B_{j,k}^{q_j}(x_j), \tag{2}$$

avec $\mathbf{B_j}(-)=\left(B_{j,k}^{q_j}(-)|k=1,\ldots,K_j+q_j=m_j\right)$ la base de B-Splines de degré q_j et K_j noeuds, où est projetée la variable X_j. Soit $\mathbf{B_j}=(\mathbf{B_j}(x_{1j})^T,\ldots,\mathbf{B_j}(x_{nj})^T)^T\in\mathbb{R}^{n\times m_j}$. Les paramètres à estimer du modèle sont alors β_0 et $\beta=(\beta_{1,1},\ldots,\beta_{d,m_d})^T$.

2 Méthode de régression pénalisée

2.1 Estimateur du MCO

Nous nous plaçons dans le cas de l'estimation du modèle 2. L'estimateur des MCO (moindres carrés ordinaires) minimise la fonction objective suivante:

$$Q^{MCO}(\beta)=\sum_{i=1}^{n}\left(Y_i-\beta_0-\sum_{j=1}^{d}\sum_{k=1}^{m_j}\beta_{j,k}B_{j,k}^{q_j}(X_{i,j})\right)^2,$$

Nous appelons alors estimateurs des moindres carrés ordinaires (EMCO) les estimateurs $\hat{\beta}$ de β réalisant cette minimisation, i.e., $\hat{\beta}=\arg\min_\beta Q^{MCO}(\beta)$. Remarquons toutefois que la méthode par MCO ne permet pas de sélectionner les composantes, et de plus, nécessite que $\sum_{j=1}^{d} m_j\leq n$.

2.2 Group LASSO

Pour sélectionner les composantes additives nous utilisons le Group LASSO, introduit par Yuan et Lin [7], car il permet de sélectionner groupe de variables par groupe de variables. Appliquée au modèle (2), la fonction objective est donnée par l'équation (3).

$$Q^{GroupLASSO}(\beta) = Q^{MCO}(\beta) + \lambda \sum_{j=1}^{d} \sqrt{m_j}||\beta_{\mathbf{j}}||_2, \tag{3}$$

avec $\beta_{\mathbf{j}} = (\beta_{j1}, \ldots, \beta_{jm_j})^T$ et m_j l'ordre de la base de B-Splines où est projetée la $j$ième variable. Il est à noter que la constante n'est pas pénalisée. L'expression (3) ne tient pas compte des échelles et nous utilisons donc une version standardisée du Group LASSO. Cette procédure a cependant les mêmes défauts que le LASSO, en particulier, il y a un biais dans l'estimation, car elle correspond à un seuillage doux des coefficients.

2.3 P-Splines

Pour estimer les approximations des composantes additives, nous utilisons des estimateurs P-Splines, introduits par Eilers et Marx [2]. La fonction objective à optimiser est donnée dans l'équation 4:

$$Q^{pspline}(\beta) = Q^{MCO}(\beta) + \sum_{j=1}^{d} \lambda_j ||D_{2,j}\beta_{\mathbf{j}}||_2^2, \tag{4}$$

où $D_{2,j}$ est la représentation matricielle de la différence d'ordre 2. Si $j > 2$, nous avons $D_{2,j}\beta_{\mathbf{j}} = \beta_j - 2\beta_{j-1} + \beta_{j-2}$.
Lorsque les noeuds sont équi-répartis, cela revient à minimiser l'équation 5:

$$Q^{PSPline}(\beta) = Q^{MCO}(\beta) + \lambda \int_0^1 \left\{ \left(\sum_{j=1}^{d} \sum_{k=1}^{m_j} \beta_k B_k^{q_j}(x_i) \right)^{(2)} \right\}^2 dx, \tag{5}$$

Si l'estimateur P-Spline est efficace en termes d'estimation, il ne permet pas de sélectionner les covariables. La section suivante décrit la manière avec laquelle nous combinons les critères présentés précédemment afin de sélectionner et estimer pertinente les composantes additives du modèle.

3 Estimateurs en trois étapes

3.1 Combinaison des méthodes de régression pénalisée

Soit $S = \{1, \ldots, d\}$ l'ensemble contenant les numéros associés aux variables explicatives présentes dans le dictionnaire de covariables. Pour une grille donnée et choisie adéquatement Λ_{GrpL} de valeurs du paramètre de régularisation λ, nous utilisons la méthode suivante:

1. **Première étape: Construction de sous-plans d'expériences candidats:** Pour chaque $\lambda_i \in \Lambda_{GrpL}$
   - Notons S_{λ_i} le sous-ensemble des numéros des covariables sélectionnées par l'estimateur associé au Group LASSO de paramètre λ_i
   - Résoudre $\tilde{\beta}_{\lambda_\mathbf{i}} = \arg\min \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{d} \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}^{q_j}(X_{i,j})\right)^2 + \lambda_i \sum_{j=1}^{d} \sqrt{m_j} ||\beta_\mathbf{j}||_2$
   - Pour chaque $j \in S$, si $||\tilde{\beta}_{\lambda_\mathbf{i},\mathbf{j}}||_2 \neq 0$, ajouter j dans S_{λ_i}, sinon ne pas considérer j dans S_{λ_i}.

   Pour faciliter les notations, nous supposons que tous les sous-plan d'expériences candidats sont différents.

2. **Seconde étape: Estimation des modèles candidats:** Pour chaque $S_{\lambda_s} \in \{S_{\lambda_{min}}, \ldots, S_{\lambda_{max}}\}$
   - Résoudre un problème du type $\hat{\beta}_{\mathbf{S}_{\lambda_\mathbf{s}}} = \arg\min Q$ où Q est la fonction objective des moindres carrés ordinaires (MCO) ou des P-Splines [2]
   - Calcul du critère de sélection de modèle (CSM), typiquement le BIC , l'AIC ou le GCV de l'estimateur $\hat{\beta}_{\mathbf{S}_{\lambda_\mathbf{s}}}$

3. **Troisième étape: Sélection de l'estimateur final:** Sélectionner $\hat{\beta}_{\mathbf{S}_{\lambda_\mathbf{b}}}$ qui minimise le CSM choisi.

3.2 Résultats théoriques

Dans le rapport technique [1], inspirés du travail de Wang *et al.* [6], nous montrons la consistance en sélection et en estimation lorsque le BIC et les MCO sont utilisés. Une difficulté de la démonstration réside dans le fait que nous travaillons sur une approximation du modèle (1). Parmi nos hypothèses, nous supposons que la norme euclidienne des composantes non nulles est supérieure à une borne qui peut tendre vers 0 avec la taille de l'´échantillon, et ceci est fondamentalement différent de l'hypothèse classique faite dans ce type de modèles, qu'il existe une constante strictement supérieure à 0 qui minore la norme euclidienne des composantes non nulles.

4 Expériences numériques

4.1 Application nationale

Cette application est explicitée dans [5]. Nous avons à disposition les données entre septembre 2007 et juillet 2013 des demandes d'électricité des clients d'EDF. Il s'agit d'un niveau agrégé de la consommation électrique. Ce portefeuille est soumis à différentes variations: arrivées/départs de clients, évolution des usages de consommation etc, pouvant être prises en compte dans un modèle additif. Pour répondre à cette demande, un expert d'EDF R&D a construit un large dictionnaire de covariables, comportant des variables calandaires et des variables météorologiques instantanées, journalières et lissées. Pour sélectionner son modèle additif final, l'expert teste un nombre très important de modèles plausibles. La méthode est longue, fastidieuse et ne peut être facilement généralisée à d'autres jeux de données. Ce besoin d'automatisation motive l'application de nos procédures sur ce jeu de données. Nous travaillons à la fois sur du court terme (jour pour le lendemain) et du moyen terme (prévision de quelques mois à une année).

4.2 Application locale

4.2.1 GEFCom 2012

Cette application est aussi explicitée dans [5]. The Global Energy Forecasting Competition 2012 (GEFCom, voir [4] pour plus de détails) est une compétition mise en ligne par le site Kaggle qui propose de prévoir et reconstruire la charge horaire consommée en kW de 20 zones américaines. Les participants de la compétition ont à disposition les courbes de charges et les températures de 11 stations météorologiques entre 2004 et juin 2008. Ce qui est important de noter est qu'il n'y a aucune information géographique fournie: les participants n'ont ni la localisation des zones de consommation électrique à prévoir, ni sur la localisation des stations météorologiques. Notre objectif est de prévoir les mois de 2008. Nous appliquons nos procédures pour sélectionner les stations.

4.2.2 Poste source

Avec la création d'un marché européen de l'électricité, EDF a été contraint de séparer ses fonctions de production, de transport et de distribution d'électricité. La liaison entre les réseaux de transport et de distribution a lieu au niveau des postes sources (environ 2200 en France). Il y a un besoin de procédures permettant de prévoir la charge consommée à court et moyen terme pour chaque poste source pour gérer la grille de distribution, quantifier les contraintes réseau et optimiser la configuration du réseau. Les prévisions court et moyen terme sont nécessaires pour des raisons financières et de fiabilité de réseau. Il y a plus de 2200 postes sources, et donc plus de 2200 séries chronologiques à étudier. En moyenne, chaque poste source est relié à 40 gros consommateurs et 16000 petits consommateurs.

Cependant, les profils des postes sources peuvent complètement changer selon le type de consommateurs qui lui sont reliés (un poste source relié principalement à des gros clients industriels est plus dépendant du moment dans l'année par exemple) ainsi que de la localisation (dans le Sud, nous pouvons penser qu'il y a plus de climatiseurs installés, et donc le poste source risque d'être plus sensible aux fortes chaleurs par exemple).

References

[1] A. Antoniadis, Y. Goude, J-M. Poggi, and V. Thouvenot. Sélection de variables dans les modèles additifs avec des estimateurs en plusieurs étapes. February 2015.

[2] P.H.C. Eilers and B.D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 05 1996.

[3] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*. London: Chapman & Hall, 1990.

[4] T. Hong, P. Pinson, and S. Fan. Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357 – 363, 2014.

[5] V. Thouvenot, A. Pichavant, Y. Goude, A. Antoniadis, and J-M Poggi. Electricity forecasting using multi-step estimators of nonlinear additive models. *Submitted*, March 2015.

[6] H. Wang, B. Li, and C. Leng. Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):671–683, 2009.

[7] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68:49–67, 2006.