

Prévision avec les modèles GAM : Application aux données de la compétition GEFCOM'14

Amandine Pierrot, EDF R&D

1 La compétition GEFCOM'14

La compétition GEFCOM'14 - Global Energy Forecasting COMpetition 2014 - s'est déroulée de septembre à décembre 2014. Organisée par l'IEEE Power & Energy Society, hébergée par le site CrowdANALYTIX.com, elle offrait 4 sujets aux compétiteurs :

- Prévision de consommation : 362 inscriptions
- Prévision de prix : 287 inscriptions
- Prévision éolienne
- Prévision photovoltaïque

Le thème de cette édition 2014 était la prévision probabiliste : prévoir les 99 quantiles de la distribution pour chacun des 4 sujets.

Les données sont des données américaines.

Sujet 1 : La prévision de consommation électrique La consommation électrique - colonne "LOAD" - est fournie au pas horaire, du 1er janvier 2005 00 : 00 au 31 décembre 2011 23 : 00 - colonnes "heure" et "date". Aucune information n'est disponible sur le consommateur : type de client ? localisation géographique ?

25 séries de température - colonnes " $w1$ " à " $w25$ " - sont également fournies au pas horaire du 1er janvier 2001 00 : 00 au 31 décembre 2011 23 : 00. Chaque série correspond à une station météorologique, la localisation géographique des stations étant inconnue.

L'objectif de ce premier sujet est de prévoir la consommation électrique **au pas horaire sur l'année 2011**. Contrairement au thème de la compétition, on prévoira la consommation **en espérance** et non pas en distribution. Pour ce faire, on pourra utiliser des variables calendaires (type de jour, position du jour dans l'année, heure ...) et les variables *Température* appropriées (choix

des chroniques de température, variables instantanées, lissées, retardées ...). Les prévisions de consommation seront effectuées à température **réalisée** - on ne demande pas de prévoir la température - et à horizon $M+1$:

- Pour prévoir janvier 2011, du 1er janvier 2011 00 : 00 au 31 janvier 2011 23 : 00, on dispose des données jusqu'au 31 décembre 2010 23 : 00
- Pour prévoir février 2011, du 1er février 2011 00 : 00 au 28 février 2011 23 : 00, on dispose des données jusqu'au 31 janvier 2011 23 : 00
- etc

Sujet 2 : La prévision des prix de l'électricité Le prix de l'électricité - colonne "Zonal Price" - est fourni au pas horaire, du 1er janvier 2011 00 : 00 au 17 décembre 2013 23 : 00 - colonne "timestamp". Aucune information n'est disponible sur la zone dont est issu le prix.

En plus de la chronique de prix, deux chroniques de prévision de consommation électrique (Total + Zone) - colonnes "Forecasted Total Load" et "Forecasted Zonal Load" - sont fournies au pas horaire sur la même période.

L'objectif de ce deuxième sujet est de prévoir le prix de l'électricité **au pas horaire sur les journées suivantes** :

- **6 juin 2013**
- **17 juin 2013**
- **24 juin 2013**
- **4 juillet 2013**
- **9 juillet 2013**
- **13 juillet 2013**
- **16 juillet 2013**
- **18 juillet 2013**
- **19 juillet 2013**
- **20 juillet 2013**
- **24 juillet 2013**
- **25 juillet 2013**
- **6 décembre 2013**
- **7 décembre 2013**
- **17 décembre 2013**

Contrairement au thème de la compétition, on prévoira le prix **en espérance** et non pas en distribution. Pour ce faire, on pourra utiliser des variables calendaires (type de jour, position du jour dans l'année, heure ...) et les variables de prévision de consommation à disposition.

Les prévisions de prix seront effectuées à horizon $J+1$:

- Pour prévoir le 6 décembre 2013, de 00 : 00 à 23 : 00, on dispose des données jusqu'au 5 décembre 2013 23 : 00
- Pour prévoir le 7 décembre 2013, de 00 : 00 à 23 : 00, on dispose des données jusqu'au 6 décembre 2013 23 : 00
- etc

2 Objectif

2.1 Objectif général

Par groupe de deux ou trois étudiants, vous étudierez ces données et mettrez en oeuvre les techniques de modélisation vues en cours. L'objectif de votre étude est de fournir un modèle GAM de prévision de consommation électrique ou de prix de l'électricité, en fonction du sujet choisi.

2.2 Étapes à suivre

- 1 Effectuer une recherche bibliographique sur le sujet de la prévision de consommation électrique ou des prix de l'électricité (voir par exemple [1], [2], [3]).
- 2 Analyser les données, comprendre et représenter les interactions entre variables à l'aide de modèles GAM (package R *mgcv* de S.N. Wood, [4]).
- 3 Mettre en place un modèle de prévision : conception du modèle, sélection de variables, validation.
- 4 Rédiger un rapport (15-20 pages) sous forme d'article scientifique (en Français) reprenant les points précédents. Proposer un code R lisible et commenté.

2.3 Critères d'évaluation

L'évaluation sera faite sur la base du rapport, ainsi que sur le code R et les résultats obtenus en prévision. Votre capacité à mettre en oeuvre les méthodes vues en cours sera appréciée, ainsi que votre rédaction et la pertinence de vos analyses.

Références

- [1] Shu Fan and Rob J. Hyndman. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1) :134–141, 2012.
- [2] A. Pierrot and Y. Goude. Short-term electricity load forecasting with generalized additive models. In *Proceedings of ISAP power*.
- [3] F. Serinaldi. Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. *Energy Economics*, 33 :1216–1226, 2011.
- [4] S.N. Wood. *Generalized Additive Models, An Introduction with R*. Chapman and Hall, 2006.