

Apprentissage du signal prix de l'électricité. Arbres de régression, séries temporelles et prédictions à long terme

Louis-Victor PASQUIER* Antoine CORNUEJOLS* Suzanne PINSON**

*Agroparistech, UMR 518 Statistiques et Génomes. F-75231 Paris, France

**Université Paris-Dauphine, LAMSADE, CNRS-UMR 7243, PSL 75775 Paris Cedex 16

1 Données et méthode d'apprentissage

Le signal à prédire est le prix spot de l'électricité en €/MWh. Les données se présentent sous la forme d'une série temporelle échantillonnée au pas de temps horaire mesurée sur une année complète, soit un total d'environ 100 000 points relevés. Afin de normaliser les données, le prix d'une heure de consommation est exprimé par rapport au prix moyen de l'année étudiée.

	date	dayofyear	month	dayofweek	hour	we	hol	target
1	2005-01-01 00:00:00	1	1	5	0	1	1	0.3181746
2	2005-01-01 01:00:00	1	1	5	1	1	1	0.3813810
3	2005-01-01 02:00:00	1	1	5	2	1	1	0.3218170

FIG. 1 – Échantillon de l'ensemble d'apprentissage.

Une moyenne pondérée des prédictions de chacun des arbres de régression appris par une méthode de *Random Forests* est employée pour déterminer la variable d'intérêt y .

2 Résultats expérimentaux

La performance des modèles appris est mesurée grâce à la variance expliquée, similaire à un écart quadratique moyen normalisé : $Var.exp. = 1 - \frac{Variance(y_{observée} - y_{modèle})}{Variance(y_{observée})}$. Une variance expliquée de $\approx 0,6$ indique un bon degré de prédictivité.

Ens. app	Ens. test	Nb arbres	Nb attr.	Min. ex./feuille	Var. exp. app.	Var. exp. test
2005-2012	2014	200	5	50	82%	62%
2010-2012	2014	500	5	50	83%	61%
2005-2012	2014	200	5	1	76%	62%

TAB. 1 – Quelques résultats en fonction de divers paramétrages.

Sur une journée Le modèle appris pour une journée (figure 2) prédit bien les heures les moins chères (5h la nuit et 15h en journée) et les heures les plus chères (8h et 18h).

Apprentissage du signal prix de l'électricité

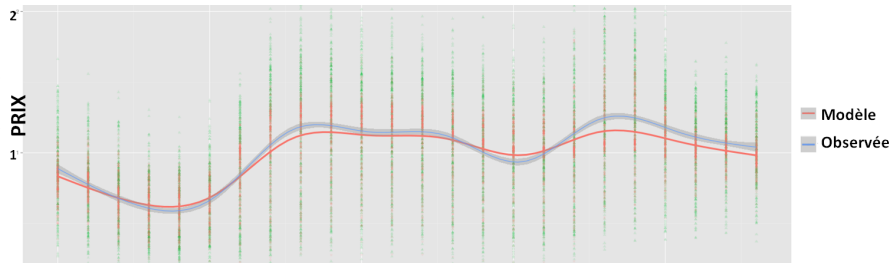


FIG. 2 – Comportement du modèle (courbe en rouge) sur un jour moyen de l'année. L'ensemble de test est représenté par la courbe en bleu (année 2014).

Sur une semaine L'amplitude de variation élevée en semaine et moindre en week-end est bien prédite, tandis qu'est mise en évidence une différence significative entre les samedis (prix intermédiaire) et les dimanches (prix les plus faibles). Le modèle prédit aussi correctement que les pointes (celle de 18h) soient de plus forte amplitude de variation en hiver qu'en été.

Sur une année Les prédictions du modèle s'accordent bien avec la réalité dans les grandes variations annuelles. Ainsi la forte chute du prix de l'électricité fin décembre est bien prédite, correspondant aux départs massifs en congé et la libération de l'essentiel des bâtiments de bureaux à cette période.

3 Conclusion

La méthode présentée s'applique directement à toutes les grandeurs variant pseudo-périodiquement au cours du temps en fonction de nombreuses variables à valeurs numériques. De plus, la méthode des random forests est peu coûteuse à mettre en œuvre en terme de calculs.

Références

- Karakatsani, N. et D. Bunn (2010). Fundamental and behavioural drivers of electricity price volatility. *Studies in Nonlinear Dynamics and Econometrics*.
- Weron, R. (2014). Electricity price forecasting : A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30(4), 1030–1081.

Summary

Predicting the price of the electricity commodity in the long term is a challenge that current techniques do not meet satisfactorily (Karakatsani et Bunn, 2010; Weron, 2014). In this paper, we introduce a new regression tree based model that yields good predictions on a long-term period with low computational resources requirements. Our approach is validated by temporal series collected from an electricity provider.