

# MODÈLES DE PRODUCTION D'ÉNERGIE ÉOLIENNE ET PRÉVISION

Lucie Montuelle <sup>1</sup> & Aurélie Fischer <sup>2</sup> & Mathilde Mougeot <sup>2</sup> & Dominique Picard <sup>2</sup>

<sup>1</sup> *Agrocampus Ouest*, 65 rue de Saint-Brieuc, 35042 Rennes Cedex, France.  
*lucie.montuelle@agrocampus-ouest.fr*

<sup>2</sup> *LPMA, bâtiment Sophie Germain*, 75205 Paris Cedex 13, France.  
*fischer@math.univ-paris-diderot.fr*, *mougeot@math.univ-paris-diderot.fr*,  
*picard@math.univ-paris-diderot.fr*.

**Résumé.** Ce travail étudie et compare différents modèles de prévision temps réel de l'énergie éolienne pour 3 parcs éoliens situés dans le nord et l'est de la France. Nous étudions l'impact de la précision de la mesure de vent sur les performances de prévision. Nous montrons que les modèles à base d'apprentissage statistique ont des performances supérieures aux modèles inspirés d'équations physiques habituellement utilisés. Les arbres de régression combinés par la méthode de bagging fournissent les meilleurs résultats.

**Mots-clés.** production électrique éolienne, modèles de régression, prévision, data mining, forêts aléatoires, bagging.

**Abstract.** We focus on short-term wind power forecast using physical, parametric and machine learning techniques. The impact of the wind speed measure sharpness over the wind power forecast is also considered. We show on real data from three farms located in the North and the East of France that parametric models, even following closely the physical equation relating wind production to wind speed are outperformed by intelligent learning algorithms. In particular, the algorithm combining CART and Bagging gives very stable and promising results.

**Keywords.** wind power forecast, modeling, data mining, random forest, bagging.

## 1 Introduction

Aujourd'hui, la production d'énergie éolienne française représente environ 4% de la production nationale d'électricité. L'objectif annoncé lors de la COP21 fin 2015 et relayé par le syndicat des énergies renouvelables (2015), est d'atteindre une production de 30% à l'horizon 2020.

L'électricité est difficile à stocker. Il faut donc constamment équilibrer l'offre et la demande. Pour assurer cet objectif au quotidien, il est essentiel de pouvoir disposer

d'outils de prévision, à la fois pour la consommation et pour les différentes sources de production d'énergie. En France, la production éolienne provient de 1400 parcs éoliens, répartis sur tout le territoire national. La production de chaque parc est très dépendante des conditions météorologiques locales, et en particulier des conditions de vent. Des régimes de vent très contrastés sont observés d'une région française à l'autre, dépendants fortement à la fois de la topographie et du climat. Pour garantir une prévision robuste au niveau national, il est essentiel de pouvoir disposer de modèles de prévision pour chaque parc éolien.

## 2 Présentation des données

Le travail présenté ici a été effectué en collaboration avec Maïa Eolis, société productrice d'énergie éolienne en France. La base de données disponible comprend 4 années de production pour 3 parcs éoliens, installés dans le Nord et l'Est de la France, regroupant chacun de 4 à 6 éoliennes. Au niveau de chaque éolienne, des capteurs fournissent, avec un pas d'échantillonnage de 10 minutes, des mesures de température, de vitesse et d'orientation du vent pour deux anémomètres installés en haut du mât (dont un anémomètre résistant au gel), de production d'énergie et une variable contextuelle indiquant les périodes de non-fonctionnement de l'éolienne.

## 3 Contexte et problématique

Historiquement, les modèles proposés pour les équipements industriels, comme les éoliennes, sont le plus souvent issus de la physique, comme le montre les articles de revue de Costa et al. (2008) et Lydia et al. (2014). C'est le cas, par exemple, pour la production d'énergie éolienne modélisée à l'aide de *la courbe de puissance* élaborée par le constructeur de turbines, reposant sur la puissance du vent et différents paramètres représentant les conditions météorologiques, souvent difficilement calculables en pratique, comme illustré par Carillo et al. (2013). Des modèles basés sur des méthodes d'apprentissage, faisant intervenir un nombre de paramètres plus important et souvent différents des modèles physiques, peuvent également être proposés, comme chez Giebel et al. (2011) et Kusiak et al. (2009). Dans une optique de modélisation puis de prévision de l'énergie éolienne, une question qui nous a paru intéressante et à ce jour ouverte, est de comparer les apports respectifs de ces deux types de modèles. Pour cela, nous avons quantifié leurs performances respectives sur la base de données de Maïa Eolis.

## 4 Résultats obtenus

Dans ce travail, onze modèles distincts ont été mis en oeuvre et étudiés : un modèle inspiré de la physique basé sur le fonctionnement de la turbine, trois modèles statistiques paramétriques, et six modèles basés sur des méthodes d'apprentissage statistique, ainsi que le modèle naïf persistant. Ce dernier suppose une certaine stabilité temporelle de la production éolienne et consiste à prédire par la dernière valeur de production électrique observée.

Parmi les méthodes statistiques testées, la meilleure est non-paramétrique et combine des arbres de régression (Breiman (1984)) à l'aide du bagging (Breiman (1996)). Elle équivaut à un changement de calibration des paramètres des forêts aléatoires, introduites par Breiman (2001). De façon intéressante, nous avons observé que les performances obtenues par les différents types de modèles varient d'un parc à l'autre. Pour les deux parcs éoliens situés dans le Nord de la France, les arbres de régression combinés à l'aide du bagging fournissent les meilleures performances de prédiction tandis que pour le parc éolien situé à l'Est, le modèle physique apporte le meilleur résultat. En fonction de la position géographique des parcs éoliens, le modèle le mieux adapté utilisant des données locales diffère d'un parc à l'autre.

Dans un objectif de prévision de la production d'énergie éolienne, les prédictions de vent le plus souvent utilisées sont fournies par des modèles météorologiques, uniquement disponibles en des points de grille. La distance entre les points est fonction de la résolution du modèle de simulation utilisé. Le modèle le plus précis fournit des données régulièrement espacées d'une distance de 6 km. Afin d'évaluer l'impact de la perte de précision des données de vent dans un contexte de prédiction, nous avons simulé une dégradation des données de vent en calculant la moyenne de ces données. Nous avons comparé les performances obtenues pour les trois parcs pour l'ensemble des modèles proposés en utilisant les données de vent locales puis des données de vent agrégées.

En nous plaçant dans un contexte de prédiction, il est intéressant de noter que les modèles les plus performants sur l'ensemble des parcs sont les modèles basés sur les arbres de régression avec bagging.

## Bibliographie

- [1] Syndicat des Énergies Renouvelables (2015), Panorama de l'électricité renouvelable en 2015, *Technical Report*, Syndicat des Énergies Renouvelables.
- [2] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E (2008), A review on the young history of the wind power short-term prediction, *Renewable and Sustainable Energy Reviews*, 12(6):1725-1744.
- [3] Lydia M, Kumar SS, Selvakumar AI, Kumar GEP (2014), A comprehensive review on wind turbine power curve modeling techniques, *Renewable and Sustainable Energy*

*Reviews*, 30:452-460.

[4] Carrillo C, Montano AO, Cidras J, Diaz-Dorado E (2013), Review of power curve modelling for wind turbines, *Renewable and Sustainable Energy Reviews*, 21:572–581.

[5] Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C (2011), The state-of-the-art in short-term prediction of wind power: A literature overview, *Technical Report, ANEMOS plus*.

[6] Kusiak A, Zheng H, Song Z (2009), Wind farm power prediction: a data-mining approach, *Wind Energy*, 12(3):275-293.

[7] Breiman L, Friedman JH, Olshen RA, Stone CJ (1984), Classification and regression trees, *Wadsworth Statistics/Probability Series*, Wadsworth Advanced Books and Software, Belmont, CA.

[8] Breiman L (1996), Bagging predictors, *Machine Learning*, 123-140.

[9] Breiman L (2001), Random forests, *Machine Learning*, 45(1):5-32.