

PRÉVISION DE PRODUCTION ÉOLIENNE PAR FORÊTS ALÉATOIRES, AGRÉGATION ET ALERTE DE RAMPES

Mamadou DIONE ¹

¹ *ENGIE Green*

Le Triade II, Parc d'activités Millénaire II, 215 rue Samuel Morse, 34967 Montpellier

² *CREST, ENSAE ParisTech*

5 Avenue Henry Le Chatelier, 91120 Palaiseau

Email : mamadou.dione@engie.com

Résumé. La rupture des contrats d'obligation d'achat avec la loi de transition énergétique définie par l'État français implique la vente d'électricité éolienne sur le marché. Pour cette vente, il est nécessaire d'annoncer la quantité d'électricité à produire. Nous avons donc besoin de prévisions de la production. Nous proposons dans cet article un modèle de forêt aléatoire puis une agrégation des prévisions de plusieurs parcs éoliens pour réduire les incertitudes et des alertes de rampes pour prévenir les périodes de grosses erreurs.

Mots-clés. Apprentissage statistique, Prévision d'énergie éolienne, Alerte de rampes, Agrégation de prévisions d'énergie éolienne.

Abstract. The breaking of purchase obligation contracts, a consequence of the french state's transition law, involves the sale of wind power on the market. For this sale it is necessary to announce the quantity of electricity to produce. So we need forecasts of production. A random forest model is proposed, followed by aggregation of forecasts from several wind farms to reduce uncertainties and ramp alerts to prevent periods of large errors.

Keywords. Machine learning, Wind power forecast, Ramp alert, Wind power forecast aggregation.

1 Introduction

La loi de transition énergétique définie par l'État français a des implications précises sur les énergies renouvelables, en particulier sur son mécanisme de rémunération. Jusque-là, un contrat d'obligation d'achat permettait de vendre l'électricité d'origine éolienne selon un tarif fixe. Demain, il faudra vendre cette électricité sur le marché (selon des tarifs variables) avant d'obtenir un complément de rémunération destiné à diminuer le risque. Cette vente sur le marché requiert d'annoncer à l'avance (environ 24h avant) la production

qui sera livrée sur le réseau et une surestimation ou une sous-estimation sera exposée à des pénalités. Il existe donc un besoin énorme de prévisions précises. A cet effet, nous nous sommes servi des prévisions météorologiques pour les transformer en prévision de production éolienne par des modèles statistiques.

La prévision de la production de parcs éoliens n'est cependant pas nouvelle ([1, 2, 3, 4]) et elle est abordée depuis des décennies à l'aide de différents modèles et techniques. Les forêts aléatoires [5] sont souvent utilisées en apprentissage statistique notamment en prévision d'énergie éolienne [3]. Cependant l'agrégation des prévisions de parcs éoliens est peu étudiée. Nous proposons une prévision (horizon 24h à 48h : days-ahead) de production de parcs éoliens par forêt aléatoire, une agrégation des prévisions et la prévision d'alerte de rampes (ramp forecasting [6]).

2 Les données

Les données proviennent du modèle européen ECMWF et du système de contrôle et d'acquisition de données (scada).

- **Les données ECMWF** : ECMWF (European Center for Medium-Range Weather Forecasts) est une organisation intergouvernementale indépendante financée par 34 États. ECMWF est à la fois un institut de recherche et un service opérationnel, qui produit et diffuse des prévisions météorologiques numériques à ses états membres. Les données du modèle ECMWF utilisées sont les prévisions horaires du run de 00h avec un horizon de prévision de 48h et une résolution de 0,125 deg. A chaque point de grille nous avons les variables suivantes.

Table 1: Données météorologiques

variables	descriptions	unités
UGRD_10m	composante est-ouest du vent à 10m du sol	m/s
VGRD_10m	composante nord-sud du vent à 10m du sol	m/s
UGRD_100m	composante est-ouest du vent à 100m du sol	m/s
VGRD_100m	composante nord-sud du vent à 100m du sol	m/s
TMP_2	température à 2m du sol	deg C
SP	pression à la surface	Pa
TP	précipitation à la surface	m

Nous considérons les 16 points de grille du maillage ECMWF les plus proches de chaque parc éolien. A partir de *UGRD* et *VGRD* nous avons calculé la vitesse W du vent et sa direction à 10m et à 100m. Le nombre de variables d'entrée est donc égale à 176 (11 variables*16 points de grille). Nous avons un historique de 3 ans (2015, 2016 et 2017) de ces données.

- **Les données réelles des parcs éoliens (données SCADA) :** les données historiques des parcs éoliens d’ENGIE Green sont utilisées. Elles couvrent la période de janvier 2015 à décembre 2017. Il s’agit de la production de chaque éolienne, de ses conditions de fonctionnement (indice de fonctionnement) et de la vitesse du vent mesurée par un anémomètre sur sa nacelle à chaque minute. L’indice de fonctionnement (égal à 0 ou 1) dépend de l’état de l’éolienne à tout moment (marche, urgence ou arrêt par exemple), de la capacité de l’éolienne à produire et du fait que la valeur de la production n’est pas fixe (la scada ne répète pas la même mesure par dysfonctionnement). Cet indice est défini de manière spécifique à Engie Green et permet de savoir à tout moment si la valeur de la production fournie par la scada est correcte ou incorrecte. Les moyennes horaires sont ensuite calculées pour

Table 2: Les données du scada

variables	description	unités
P	production de chaque éolienne	Kw
$W_{Nacelle}$	vitesse réelle du vent sur chaque éolienne	m/s
F	indice de fonctionnement de chaque éolienne	

réduire la variabilité et avoir la même fréquence que les données météorologiques. La variable à prédire est la production d’un parc éolien qui est la somme de la production des éoliennes du parc. Cependant nous avons appliqué un modèle indirect qui prévoit le vent sur le parc et transforme la prévision du vent en prévision de production par une courbe de puissance réelle estimée par une spline. Les données sont divisées en deux ans d’apprentissage et un an de test. Nous avons utilisé le $NMAE = (1/PI * N) * \sum_{i=1}^N |\hat{P}_i - P_i|$ comme indicateur de performance, où PI est la puissance installée d’un parc éolien.

3 Méthodologie et résultats

Dans cette partie, nous avons fait la prévision indirecte (décrit ci-dessous) par forêts aléatoires de trois parcs éoliens.

- Apprentissage sur 2 ans :
 - estimation de la courbe de puissance réelle : $P = g(W_{nacelle}) + \eta$ par spline de lissage pour obtenir \hat{g} ,
 - prévision indirecte : $W_{nacelle} = h(X) + \zeta$, apprendre le modèle de prévision de la vitesse du vent sur le parc et obtenir l’estimateur (ou le modèle de prévision) \hat{h} par forêt aléatoire.

- Test sur 1 an : chaque jour, utiliser les données météorologiques (horizon 24h à 48h) pour prévoir le vent sur le site $\hat{W}_{nacelle} = \hat{h}(X)$ et le transformer en prévision de production $\hat{P} = \hat{g}(\hat{W}_{nacelle})$.

3.1 Forêts aléatoires

Depuis son introduction par L. Breiman [5] en 2001, beaucoup de publications ont abordé la théorie des forêts aléatoires avec des applications dans plusieurs domaines. Le principe des forêts aléatoires est tout d’abord de générer plusieurs échantillons bootstrap $\mathcal{B}_n^{\Theta_1}, \dots, \mathcal{B}_n^{\Theta_q}$. Ensuite, sur chaque échantillon $\mathcal{B}_n^{\Theta_l}$, une variante de CART (Classification And Regression Tree [7]) est expliquée. En d’autres termes, un arbre est construit de façon suivante. Le découpage d’un nœud se fait par un tirage aléatoire de m variables et la recherche de la meilleure coupure suivant les m variables sélectionnées. De plus, l’arbre construit est complètement développé (arbre maximal) et non élagué. L’élagage consiste à chercher le meilleur sous-arbre élagué de l’arbre maximal (meilleur au sens de l’erreur de généralisation). La collection d’arbres obtenus est enfin agrégée (moyenne en régression, vote majoritaire en classification) pour donner le prédicteur de forêts aléatoires.

Table 3: NMAE pour chaque parc éolien

parcs éoliens	NMAE
parc éolien 1	7.64
parc éolien 2	8.12
parc éolien 3	8.32

Même si les résultats sont convaincants par rapport à la littérature où le NMAE tourne autour de 8, 9 voire 10% ([9, 3]), il existe des périodes où les prévisions sont très mauvaises (périodes de rampe). Nous avons observé également que la qualité des prévisions dépend du parc éolien (par exemple la production du parc 1 est plus facile à prédire). Après analyse, ces erreurs sont principalement dues à de mauvaises prévisions de la météorologie. C’est pourquoi nous avons abordé la prévision de rampe pour alerter des périodes de grosses erreurs de prévision.

3.2 Alerte de rampes

Il existe plusieurs définitions d’événement de rampe (voir [8]) qui sont basées sur les variations de la production. Nous souhaitons étudier les périodes de grosses erreurs du modèle qui ne sont pas forcément dues à une variation brusque de la production. Par conséquent, nous avons défini les événements de rampes ($Ramp_t$), comme étant les périodes des écarts en valeur absolue ($|\hat{P}_t - P_t|$) supérieurs au troisième quantile des écarts absolus. Ce qui correspond à un $NMAE > 9.92\%$. A chaque instant t , nous avons $Ramp_t = 0$ s’il y a un

événement de rampe et $Ramp_t = 1$ sinon. Pour prévoir une période de rampe nous avons défini :

$$\Delta t = (P_{t-\beta+2} - P_{t-\beta+1}, \dots, \hat{P}_{t+1} - P_t, \dots, \hat{P}_{t+H} - \hat{P}_{t+H-1}),$$

où β définit le nombre de prévisions passées à prendre en compte et H , l'horizon maximum de prévision. Ensuite, Δ_t est utilisé comme entrée pour prédire les rampes ($Ramp_t$) avec une forêt aléatoire. Nous avons utilisé deux métriques d'évaluation des prévisions de rampes basées sur le tableau de contingence : *ratio de succès* = $1 - FP/(FP + VP)$ et *probabilité de détection* = $VP/(VP + FN)$, où FP est le nombre de faux positifs, VP le nombre de vrais positifs et FN le nombre de vrais négatifs (voir [8]).

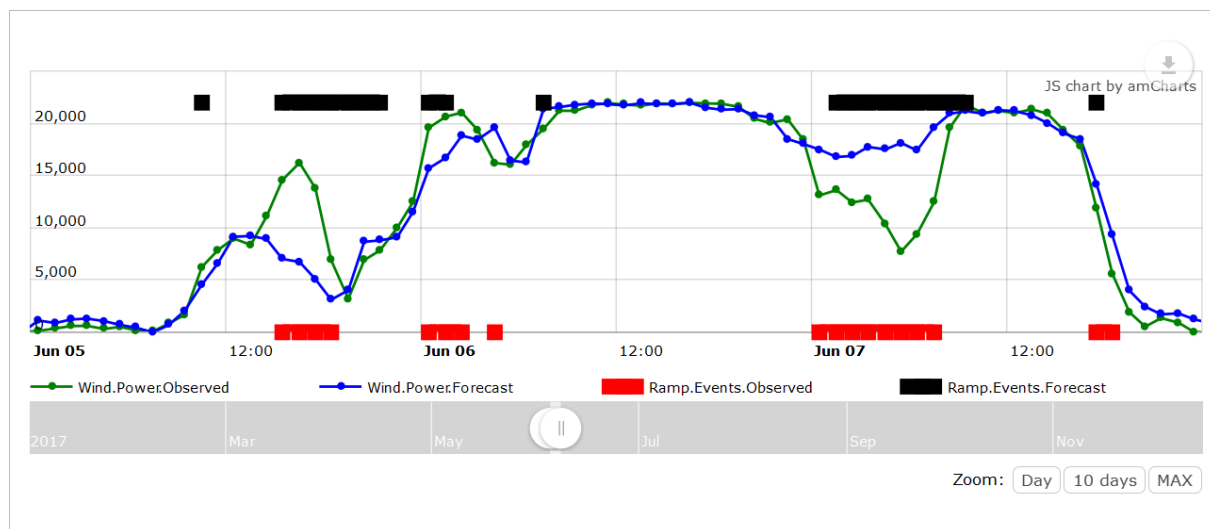


Figure 1: Prédiction de production éolienne du parc 1 et alerte de rampes : zoom sur la période du 5 au 7 juin 2017.

Sur l'échantillon teste nous avons obtenu une probabilité de détection de 86.6% et un ratio de succès de 87.7%.

3.3 Agrégation de prévisions de parcs éoliens

Dans cette partie nous proposons une agrégation des prévisions qui réduit les erreurs. Pour ce faire nous agrégeons par la somme des prévisions des trois parcs et de leurs productions réelles pour évaluer les erreurs de prévision. En agrégeant les prévisions des trois parcs éoliens nous avons obtenu une réduction du MAE qui est passé à 6.29%.

4 Conclusion

Nous avons proposé dans cet article une méthode de prévision de la production éolienne par forêt aléatoire. Comme source d'entrée nous avons utilisé les prévisions météorologiques

sur plusieurs points de grille pour prendre en compte l'aspect spatial du vent. Nous avons ensuite étudié les périodes où le modèle faisait de grosses erreurs (périodes de rampes) et nous avons constaté que ces erreurs sont dues à de mauvaises prévisions météorologiques. Nous avons ensuite proposé un autre modèle d'alerte des périodes de rampes et une agrégation des prévisions de trois parcs qui a réduit considérablement le *MAE*.

Ces résultats sont très concluants par rapport à la littérature aussi bien sur la prévision de la production ([9]) que sur la prévision de périodes de rampes ([8]). Il convient cependant de mieux comprendre les périodes de rampes voire ajuster un modèle de prévision de la production pour ces périodes pour réduire les grosses erreurs et donc de manière générale l'incertitude sur les prévisions de la production éolienne. D'autres méthodes d'agrégation des prévisions sont aussi envisageables (agrégation par poids exponentiels) de même que la construction d'intervalles de confiance (dans le cas d'une distribution non symétrique) en lieu et place des alertes de rampes.

References

- [1] E. K. Kotwa, H. Madsen, J. Tastu, P. Pinson and H. A. Nielsen.: Spatiotemporal analysis and modeling of short term wind power forecast errors. *Wind Energy*, **14**(1), 43–60, (2011)
- [2] H. Zheng, A. Kusiak and Z. Song.: Wind farm power prediction: A data-mining approach. *Wind Energy* **12**(3), 275–293 (2008)
- [3] J. Juban, L. Fugon and G. Kariniotakis.: Data mining for wind power forecasting. *European Wind Energy Conference 2008*, p. 6, EWEC, Brussels, Belgium (2008). <hal-00506101>
- [4] H. Madsen, A. Costa A. Crespo J. Navarron G. Lizcano and E. Feitosa.: A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, **12**(6), 1725–1744, (2008)
- [5] L. Breiman. *Random forest. : Machine Learning*, (2001)
- [6] C. Gallego-Castillo , A. Cuerva-Tejero, O. Lopez-Garcia.: A review on the recent history of wind power ramp forecasting.: *DAVE-ETSIAE*, Universidad Politecnica de Madrid. Pza Cardenal Cisneros, 3, 28040 Madrid, Spain
- [7] J. Friedman, C. J. Stone, L. Breiman and R.A. Olshen.: *Classification and regression trees*. Chapman&Hall/CRC, pp. 5–32, (1984)
- [8] J. Zhang, A. Florita, B. M. Hodge and J.Freedman.: Ramp Forecasting Performance from Improved Short-Term Wind Power Forecasting. *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2014)*, New York, August 17–20, 2014
- [9] A. Gensler, B. Sick and S. Vogt: A Review of Deterministic Error Scores and Normalization Techniques for Power Forecasting Algorithms. In *Computational Intelligence (SSCI)*, 2016 IEEE Symposium Series on, pages 1-9. IEEE, 2016.