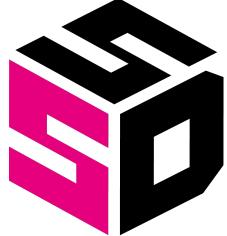




Université de Montpellier



Master 2 - Statistiques et Sciences
des Données

Année Universitaire 2024 - 2025



Développement d'un jumeau numérique pour la modélisation de production des fermes éoliennes

Auteur :

Guillaume Bernard-Reymond

Encadrant :

Paul Mazoyer

Mars 2025 - Août 2025

Remerciements

Reprendre les études à presque 38 ans a été pour moi un véritable défi. Réussir à concilier vie professionnelle, vie étudiante et surtout vie familiale n'a clairement pas été de tout repos. Ce parcours marque cependant un nouveau départ, qui se conclut par ce stage au sein de l'entreprise Engie Green, que je tiens à remercier sincèrement.

Mathias, un ami, m'a permis de concrétiser ma volonté de mettre mes compétences au service de la transition énergétique, en me présentant les bonnes personnes au bon moment : je lui exprime ici toute ma gratitude.

L'équipe *Ressources et Productibles*, que j'ai rejointe durant ce stage, m'a immédiatement considérée comme un membre à part entière. Benoît Buffard, je te suis reconnaissant de m'avoir mis à l'aise dans cet environnement de travail dès les premiers jours.

Dans cette équipe, j'ai eu la chance d'avoir Paul Mazoyer pour maître de stage. À chacun de nos points hebdomadaires, tu as su faire preuve de bienveillance face à mon manque initial de connaissances sur le contexte éolien, tout en étant exigeant, tant sur le fond scientifique que sur la forme et la communication. À chaque discussion, tu as su me relancer, me pousser à explorer de nouvelles pistes pour toujours améliorer ce travail de recherche. Un grand merci à toi.

Je ne peux conclure sans exprimer toute ma gratitude à Camille, mon amie. Ta présence tout au long de ce master a fait toute la différence, et je suis convaincu que mes résultats n'auraient pas été les mêmes sans ton soutien.

Enfin, je tiens à remercier tout particulièrement ma compagne Sandra, qui m'a soutenue, crue en moi, tout en me rappelant qu'il existait autre chose que les études et le travail. Merci infiniment.

Résumé

Résumé : Dans le cadre d'une étude de potentiel éolien, les ingénieurs s'appuient sur divers types de données afin d'estimer la production long-terme. Des mesures sur site, réalisées à l'aide de mâts de développement, sont couplées à des données météorologiques pour évaluer la ressource en vent du site candidat. Dans cette approche, la vitesse du vent constitue la variable principale, car elle conditionne directement la production d'énergie. Toutefois, d'autres variables, bien que souvent négligées dans les études classiques, jouent également un rôle significatif dans la production du parc.

Par ailleurs, les estimations de puissance sont généralement effectuées à partir de la courbe de puissance fournie par les constructeurs. Pourtant, la littérature montre que des modèles de courbes de puissance issus de l'apprentissage statistique (Machine Learning) offrent souvent de bien meilleures performances.

Ce rapport s'articule autour de trois axes de recherche.

Dans un premier temps, nous nous intéressons à la modélisation long terme de la ressource en vent d'un site éolien en mobilisant la méthodologie "Measure-Correlate-Predict". D'autres variables, autres que la vitesse de vent ou sa direction, sont modélisées par ce processus, allant au delà de ce qui est fait lors d'une étude de potentiel éolien. Une attention particulière est portée au choix des types de données disponibles pour ces modélisations, dans le but d'identifier celles permettant d'obtenir les résultats les plus robustes.

Dans un second temps, nous développons et comparons plusieurs modèles de courbe de puissance, incluant notamment des approches fondées sur l'apprentissage statistique. Le choix des données d'entraînement, élément déterminant pour la qualité de la modélisation, fait l'objet d'une analyse approfondie dans ce rapport. Les performances des différents modèles sont ensuite comparées à celles des courbes de référence issues des standards industriels.

Enfin, le troisième axe est consacré à l'étude des transferts de modèles entre éoliennes. L'objectif est de déterminer dans quelle mesure ces transferts peuvent compléter, voire partiellement remplacer, la courbe de puissance constructeur dans le cadre d'un nouveau développement de parc.

Abstract : As part of a wind resource assessment study, engineers rely on various types of data to estimate long-term energy production. On-site measurements, conducted using meteorological masts, are combined with meteorological data to evaluate the wind resource of the candidate site. In this approach, wind speed is the main variable, as it directly determines energy production. However, other variables—often overlooked in standard studies—also play a significant role in the overall production of the wind farm.

Power estimates are generally derived from the power curve provided by the turbine manufacturers. However, the literature shows that power curve models based on statistical learning (Machine Learning) often deliver significantly better performance.

This report is structured around three main research axes.

First, we focus on the long-term modeling of the wind resource at a wind farm site by applying the "Measure-Correlate-Predict" methodology. Variables other than wind speed or direction are also modeled through this process, extending beyond the scope of typical wind potential studies. Particular attention is given to the selection of available data types for these models, with the aim of identifying those that yield the most robust results.

Second, we develop and compare several power curve models, including approaches based on statistical learning. The selection of training data—an essential factor in the quality of the modeling process—is thoroughly analyzed in this report. The performance of the different models is then compared to reference curves derived from industry standards.

Finally, the third axis is dedicated to the study of model transfers between wind turbines. The objective is to determine the extent to which such transfers can complement or even partially replace the manufacturer's power curve in the context of a new wind farm development.

Table des matières

1 Présentation de l'entreprise	13
2 Introduction générale	14
2.1 Contexte éolien	14
2.1.1 Le fonctionnement d'une éolienne	14
2.1.2 Évaluation du productible	15
2.1.3 Principales variable d'influence sur la production électrique d'une éolienne	17
2.1.4 Production annuelle d'énergie - Annual Energy Production (AEP) .	21
2.2 Objectifs du stage	21
2.2.1 Cas d'usage	21
2.2.2 Modélisation longue durée	22
3 Campagnes de mesures et type de données utilisées	24
3.1 Présentation détaillée des parcs étudiés	24
3.1.1 Avesnes	24
3.1.2 Chemin des Haguenets	25
3.1.3 Couturelle	25
3.1.4 Prévoterie	25
3.2 Types de données disponibles	26
3.2.1 Données de Campagne de Mesure de Vent (CMV)	26
3.2.2 Données opérationnelles SCADA	27
3.2.3 Données de réanalyses	28
4 Méthodologie Measure-Correlate-Predict et modèles de courbe de puissance	30
4.1 L'approche Measure-Correlate-Predict pour l'estimation de la ressource . .	30
4.1.1 Principe général et objectifs de prédiction	30
4.1.2 Les différents modèles de prédiction testés	31
4.1.3 Critères de choix du modèle	33
4.1.4 Les cas d'entraînement	34
4.1.5 Cas de test	34
4.1.6 Utilisation des données de mât	35
4.2 Estimation du productible	36
4.2.1 État de l'art	36
4.2.2 Différents modèles de courbes de puissance	36
4.2.3 Cas d'entraînement : développement des modèles	40
4.2.4 Cas de test : prédiction et comparaison aux données opérationnelles	41
5 Résultats et interprétations	43
5.1 Choix des modèles et résultats de la modélisation long-terme	43
5.1.1 Avec utilisation des données LiDAR	43

5.1.2	L'utilisation des données de mâts	46
5.1.3	Conclusion	48
5.2	Évaluation du productible	48
5.2.1	Choix du modèle	48
5.2.2	Écart aux valeurs mesurées des différents modèles	49
5.2.3	Conclusion	52
5.3	Comparaison des productions d'énergie annuelle estimées	53
5.3.1	Production annuelle de référence	53
5.3.2	Comparaison pour les différents modèles	54
5.3.3	Conclusion	56
5.4	Transfert de modèles	57
5.4.1	Combinaisons testées et méthodologie	57
5.4.2	Résultats obtenus	57
5.4.3	Conclusion	57
6	Conclusion et perspectives	59
6.1	Bilan du stage	59
6.2	Perspectives d'amélioration	59
A	Normalisation de l'intensité de turbulence	62
B	Sensibilité au cisaillement comme prédicteur dans la prédiction long-terme	66
B.1	Avec utilisation des données LiDAR	66
B.2	Avec utilisation des données de mât	67

Sigles et Acronymes

- k-NN** *k*-Nearest Neighbors. 33, 48
- ABH** Avesnes et Bosc-Hyons. 24, 35, 46, 50
- AEB** Avesnes et Beauvoir. 9, 24, 30, 35, 46, 50
- AEP** Annual Energy Production. 3, 10, 21, 22, 27, 53–56
- CDH** Chemin des Haguenets. 25, 34, 35, 44, 46, 50, 57
- CDS** Chemin des Haguenets Sud. 10, 25, 34, 35, 44–46, 50
- CMV** Campagne de Mesure de Vent. 3, 12, 21, 26, 27, 29–31, 34
- COF** Couturelle. 10, 25, 35, 44–47, 50, 54, 57
- EDS** Prévôterie 2. 25, 26, 35, 44, 46, 50, 54
- EnR** Energies Renouvelables. 13
- EPE** Etude de Potentiel Eolien. 21, 38
- GBM** Gradient Boosting Model. 33, 48
- IEC** International Electrotechnical Commission. 18, 19, 37
- LiDAR** Light Detection And Ranging. 3, 4, 9, 10, 21, 27–31, 34, 35, 38–40, 43, 45–52, 59, 66
- MCP** Measure-Correlate-Predict. 10–12, 22, 28, 39–41, 44, 47–50, 52, 53, 55–57, 59, 66–68
- NRMSE** Normalized Root Mean Squared Error. 9–11, 33, 41, 43, 44, 47–52, 55–59, 66–68
- PC** Power Curve. 16
- PRP** Prévôterie 1. 25, 26, 35, 46, 50, 55, 57
- RF** Random Forest. 12, 33, 36, 38, 43, 44, 46, 48, 49, 52, 55
- RMSE** Root Mean Square Error. 9, 33, 39
- SCADA** Supervisory Control and Data Acquisition. 3, 9–12, 27, 28, 30, 35, 36, 38, 40–42, 45, 46, 50, 53, 55, 67
- TI** intensité de turbulence (Turbulence Intensity). 18, 31, 34, 35, 38, 44, 47, 48
- WA** direction du vent (Wind Angle). 35, 39, 48
- WS** vitesse de vent (Wind Speed). 17, 34–36, 38, 44, 45, 47, 48

Glossaire

shear cisaillement. 10, 11, 20, 29, 31, 34, 35, 38, 44, 47, 48, 66, 67

veer déviation verticale du vent. 29, 39, 44

Liste des symboles

α_h	Coefficient de cisaillement
Δ_P	Écart de puissance
$\overline{P}_{i,j}$	Puissance moyenne pour l'heure i contenant j mesures
\bar{y}	Moyenne arithmétique de la variable y
ρ	Densité de l'air
ρ_0	Densité de l'air de référence
$\rho_{10\text{min}}$	Densité de l'air moyennée sur 10 minutes
$\sigma_{v_{10\text{min}}}$	Ecart-type temporelle de la vitesse du vent sur 10 minutes
C_P	Coefficient de puissance
D_{rotor}	Diamètre du rotor
D_{KL}	Divergence de Kullback-Leibler
h	Hauteur
h_{hub}	Hauteur de la nacelle
i	Indice temporel
N_h	Nombre d'heures dans une année
P	Puissance produite par l'éolienne
P_i	Puissance produite par l'éolienne à l'heure i
P_{rated}	Puissance nominale de l'éolienne
S	Surface du rotor
$T_{10\text{min}}$	Température de l'air moyennée sur 10 minutes
T_{ref}	Température de l'air de référence
v	vitesse du vent
v_h	Vitesse du vent à la hauteur h
$v_{\text{cut-in}}$	Vitesse de démarrage de l'éolienne
$v_{\text{cut-out}}$	Vitesse d'arrêt de l'éolienne
v_{hub}	Vitesse du vent à hauteur de nacelle

v_{norm} Vitesse normalisée par la densité de l'air

v_{rated} Vitesse nominale de l'éolienne

Liste des figures

1.1	Bureaux et parcs d'Engie-Green.	13
2.1	Fonctionnement et éléments d'une éolienne.	14
2.2	Vue aérienne d'une partie du site de test d'Osterild (Danemark) avec à gauche le mât de mesures et à droite l'éolienne testée. Images <i>Google Earth</i>	16
2.3	Courbe constructeur pour une éolienne de puissance nominale de 2050 kW.	16
2.4	Distribution de la vitesse du vent avec la fonction densité de Weibull associée.	17
2.5	Mise en évidence de l'effet de l'intensité de turbulence sur la production de puissance.	19
2.6	Influence du cisaillement sur la vitesse du vent.	20
2.7	Cas d'usage d'une étude de potentiel éolien.	22
2.8	Cas d'étude du stage.	23
3.1	Position géographique des parcs étudiés.	24
3.2	Schéma d'implantation des éoliennes des parcs d'Avesnes.	24
3.3	Schéma d'implantation des éoliennes des parcs du Chemin des Haguenets. .	25
3.4	Schéma d'implantation des éoliennes du parc Couturelle.	25
3.5	Schéma d'implantation des éoliennes des parcs de Prévoterie.	26
3.6	Lidar à quatre faisceaux lasers sur la nacelle d'une éolienne.	27
3.7	Nettoyage de données Supervisory Control and Data Acquisition (SCADA). .	28
4.1	Écart à la moyenne pluriannuelle des moyennes annuelles de la vitesse de vent issue des données SCADA pour l'éolienne Avesnes et Beauvoir (AEB). .	30
4.2	Principe de prédiction Measure-Correlate-Predict.	31
4.3	Ecart moyen annuel pour l'intensité de turbulence et le cisaillement obtenu à partir des données de réanalyse.	32
4.4	Comparaison entre la puissance observée et la puissance prédictive selon trois techniques de modélisation de la puissance.	36
4.5	Comparaison entre l'approche indirecte à gauche et directe à droite pour la prévision de la puissance P	37
4.6	Courbe de puissance constructeur et son interpolation.	37
4.7	Comparaison des courbes de puissance constructeur, par tranches à partir des données de la campagne Light Detection And Ranging (LiDAR) (CP binning) et celle obtenue par normalisation des turbulences (CP TI=0 théorique).	38
4.8	Diagrammes en boîtes obtenus à partir 10 valeurs de Root Mean Square Error (RMSE) en kW obtenues après la validation croisée effectuée sur chaque modèle représenté en abscisse, ainsi que la valeur du RMSE du modèle standard pour la prédiction de la puissance.	39
4.9	Comparaison des distributions de la puissance pluri-annuelle et celle la plus représentative sur un an des données complètes	41
5.1	Normalized Root Mean Squared Error (NRMSE) par parc et par modèle pour la vitesse de vent normalisée.	43
5.2	NRMSE par parc et par modèle pour l'intensité de turbulence.	44
5.3	NRMSE par parc et par modèle pour le cisaillement.	44

5.4	Ecart moyen normalisé de la vitesse modélisée à l'aide des données LiDAR.	45
5.5	Distributions modélisée et observée de la vitesse normalisée pour l'éolienne Chemin des Haguenets Sud (CDS).	45
5.6	Comparaison des vitesses de vent issues des données SCADA et LiDAR.	46
5.7	Comparaison des écarts moyens de la vitesse issue de la Measure-Correlate-Predict (MCP) sur données LiDAR (ws_SLT) et données de mât (ws_mat).	47
5.8	Comparaison des NRMSE de la vitesse issue de la MCP sur données LiDAR (ws_SLT) et données de mât (ws_mat).	47
5.9	Comparaison de différents modèles statistiques pour la prédiction de P	49
5.10	Comparaison de l'écart moyen normalisé aux mesures opérationnelles SCADA de la puissance P prédicté à partir des données provenant de la MCP avec entraînement sur les données LiDAR.	50
5.11	Comparaison des modèles de courbe de puissance entraînés sur données de la campagne LiDAR (P_{SLT_SLT}) et sur l'année la plus représentative ($P_{SLT_best_year}$).	51
5.12	Comparaison des écarts normalisé aux données opérationnelles pour la meilleure année selon l'utilisation de données LiDAR ($P_{SLT_best_year}$) ou de mât ($P_{SLT_mat_best_year}$) lors de la MCP.	52
5.13	Comparaison du NRMSE pour la "meilleure" année selon l'utilisation de données LiDAR ($P_{SLT_best_year}$) ou de mât ($P_{SLT_mat_best_year}$) lors de la MCP.	52
5.14	Ecart en pourcentage à l'Annual Energy Production (AEP) calculé en gardant les données ayant 6 mesures par heures pour l'éolienne Couturelle (COF)1. var_i désignant l'écart où l'on a gardé les horodatages contenant au moins i mesures par heure.	54
5.15	Comparaison des écarts moyens normalisés des modèles de courbe de puissance constructeur construits à partir de données issues de la MCP sur données de mât.	55
5.16	Comparaison du NRMSE de différents modèles de courbe puissance utilisant des données issues de la MCP sur données de mât.	56
5.17	Comparaison des courbes, en bleu "constructeur", en vert (Binning scada) des données opérationnelles (v_{scada}, P_{scada}) regroupées par tranche et moyennées , en gris des données ($v_{mat_norm}, P_{mat_best_year}$) moyennées par tranche donnant la courbe rouge (Binning SLT).	56
5.18	Ecart moyen normalisé pour le transfert de modèle de courbe de puissance de l'éolienne 1 avec les données de l'éolienne 2.	58
5.19	NRMSE pour le transfert de modèle de courbe de puissance de l'éolienne 1 avec les données de l'éolienne 2.	58
A.1	Processus pour obtenir la courbe de puissance initiale avec zéro-turbulence.	62
A.2	Processus pour obtenir la courbe de puissance théorique avec zéro-turbulence.	63
A.3	Processus pour obtenir la courbe de puissance finale avec zéro-turbulence.	64
A.4	Application de la courbe de puissance à turbulence nulle à un nouveau site.	65
B.1	Ecart moyen normalisé entre données issues de la MCP sans (ws_SLT) ou avec (ws_SLT_shear) prise en compte shear dans les données de réanalyse.	66
B.2	NRMSE entre données issues de la MCP sans (ws_SLT) ou avec (ws_SLT_shear) prise en compte shear dans les données de réanalyse.	67

Liste des tableaux

3.1	Caractéristiques des différentes éoliennes étudiées.	26
3.2	Récapitulatifs des différents dispositifs utilisés lors d'une Campagne de Mesure de Vent (CMV).	29
4.1	Synthèse des données Lidar disponibles pour les éoliennes étudiées.	34
4.2	Synthèse des données Lidar disponibles pour les éoliennes étudiées.	34
4.3	Données de mât disponibles pour les différents parcs étudié.	35
4.4	Statistiques des mesures de la vitesse opérationnelle SCADA.	36
4.5	Année dont la distribution de la puissance est la plus proche de la distribution pluri-annuelle.	40
4.6	Résumé des informations concernant la puissance opérationnelle SCADA. .	41
5.1	Biais pour les différents parcs pour les modèles de Random Forest (RF) utilisés pour la MCP.	44
5.2	Biais des prédictions issues de la MCP sur données de mât.	48
5.3	Nombre de données avant et après filtrage sur les heures ayant 6 mesures et la perte en pourcentage du nombre de données.	54

1

Présentation de l'entreprise

Dans un contexte de réchauffement climatique, et d'indépendance énergétique, le développement des Energies Renouvelables (EnR) est une nécessité. En croissance régulière depuis plusieurs années en France, les EnR représentent 15,4 % de la consommation d'énergie primaire en 2023 avec pour objectif d'atteindre 30 % en 2030. C'est au travers de cet objectif nécessaire et ambitieux que la société *Engie Green* s'emploie à leur développement en France et dans le monde.

Fondée en 2016 et filiale à 100 % de l'entreprise *Engie*, *Engie Green* est donc spécialisée dans la production et l'exploitation d'énergies renouvelables telles que l'éolien, le solaire ou la biomasse. En 2024, comme l'illustre la figure 1.1, la production des différents parcs d'*Engie Green* a représenté l'équivalent de la consommation de 3 millions de foyers français.

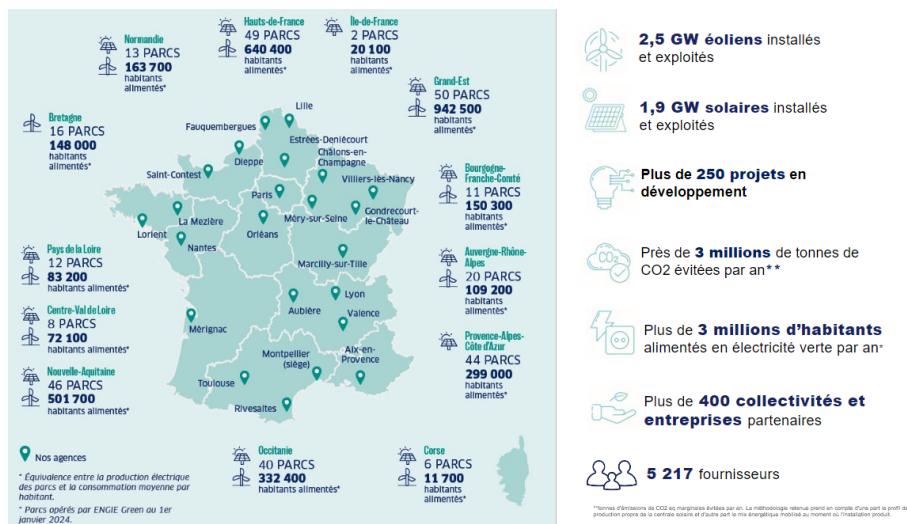


Figure 1.1: Bureaux et parcs d'*Engie-Green*.

Plus de 600 collaborateurs situés dans 23 agences sur l'ensemble du territoire français, permettent le développement et l'exploitation des différents projets indispensables à l'objectif d'une société décarbonée. Dans cette optique, le pôle *Ressources et Productibles* actuellement sous la direction de Benoît Buffard, est en charge d'évaluer les ressources disponibles pour un projet éolien ou solaire dans un objectif d'étudier la faisabilité économique d'un projet, mais aussi d'analyser et modéliser la performance des différents parcs. C'est donc au sein de ce pôle et plus précisément dans l'équipe de *Data Science et Modélisation* que j'ai effectué mon stage sous la direction de Paul Mazoyer, le responsable de cette équipe.

Introduction générale

2.1 Contexte éolien

2.1.1 Le fonctionnement d'une éolienne

Une éolienne produit de l'électricité grâce au vent qui met en mouvement un rotor, permettant sa transformation en énergie mécanique (Fig. 2.1). La vitesse de rotation de l'arbre entraîné par le mouvement des pales est accélérée par un multiplicateur. Cette énergie mécanique est ensuite transmise au générateur.

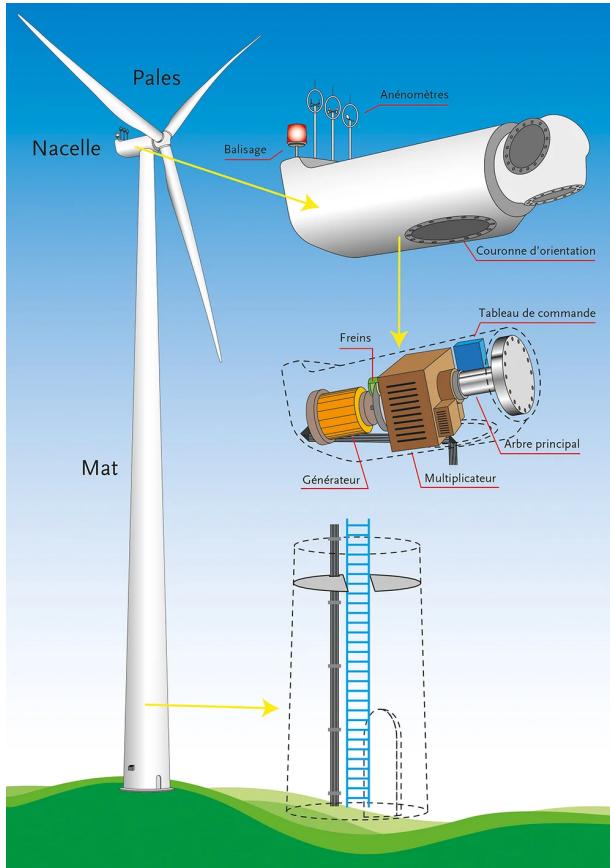


Figure 2.1: Fonctionnement et éléments d'une éolienne.

Un transformateur situé à l'intérieur du mât élève la tension du courant électrique produit par l'alternateur, pour qu'il puisse être plus facilement transporté dans les lignes à moyenne tension du réseau électrique.

Un parc éolien est constitué de plusieurs éoliennes espacées de plusieurs centaines de mètres et connectées entre elles par un réseau interne souterrain et raccordées au réseau public par l'intermédiaire d'un poste de livraison.

Une turbine éolienne moderne a une puissance comprise entre 2 MW et 21.5 MW pour les derniers modèles d'éoliennes en mer. Pour pouvoir démarrer, une éolienne a besoin d'une vitesse de vent minimale d'environ 3 m/s, et pour des questions de sécurité, l'éolienne s'arrête automatiquement de fonctionner lorsque le vent dépasse 25 m/s. La vitesse nominale, à partir de laquelle l'éolienne ne produit pas plus qu'à une vitesse supérieure, est d'environ 13 m/s.

Une éolienne peut être soumise à différents types de bridages, entraînant une réduction, voire un arrêt temporaire de sa production. Ces bridages peuvent répondre à des contraintes environnementales, acoustiques, techniques ou de réseau :

- **Bridage avifaune ou chiroptère** : des capteurs visuels ou sonores détectent la présence d'oiseaux ou de chauves-souris à proximité. L'éolienne est alors arrêtée temporairement afin d'éviter tout risque de collision et de préserver la faune.
- **Bridage acoustique** : afin de limiter les nuisances sonores pendant certaines périodes (notamment la nuit ou à proximité de zones habitées), la vitesse de rotation de l'éolienne est réduite, voire stoppée, pour diminuer le niveau sonore émis.
- **Bridage réseau** : pour maintenir en permanence l'équilibre entre la production et la consommation d'électricité, certaines éoliennes peuvent être ralenties ou arrêtées afin d'éviter une surcharge du réseau électrique.
- **Bridage dû au sillage** : lorsqu'une éolienne se trouve dans le sillage d'une autre (en fonction de la direction du vent), elle peut être arrêtée pour éviter les effets de turbulence, qui peuvent engendrer une usure prématuée ou une baisse de rendement.

2.1.2 Évaluation du productible

Une éolienne sert à récupérer l'énergie cinétique de l'air situé en amont du rotor. Toutefois, elle ne peut pas récupérer l'énergie totale fournie par le vent. Cela reviendrait à arrêter le vent. Lorsque l'énergie cinétique du vent est convertie en énergie mécanique par le rotor, le vent est freiné par celui-ci, la vitesse du vent en amont du rotor est toujours supérieure à celle en aval. Or, la masse d'air qui traverse la surface balayée par le rotor est identique à celle sortante (le débit massique est conservé). Il en résulte un élargissement de la veine d'air à l'arrière du rotor. Ce freinage du vent est progressif, jusqu'à ce que la vitesse de l'air à l'arrière du rotor devienne à peu près constante. L'air conserve donc une énergie cinétique résiduelle pour qu'il subsiste un écoulement.

Albert Betz a démontré que la puissance maximale récupérable est égale aux $\frac{16}{27}$ de la puissance incidente. La puissance théorique P en watts (W) d'une éolienne est calculée par :

$$P = \frac{1}{2} \rho S C_P v^3 \quad (2.1)$$

où :

- ρ : densité de l'air (kg/m^3) ;
- S : surface balayée par les pales (m^2) ;

- C_P : coefficient de puissance dont la valeur est comprise entre 0 et $\frac{16}{27} \approx 0.59$;
- v : vitesse du vent en amont au niveau de la nacelle (m/s)

La courbe de puissance, ou Power Curve (PC), est un graphe représentant la puissance de sortie d'une éolienne à différentes vitesses de vent à hauteur de moyeu, moyennée sur 10 minutes.

Les constructeurs de turbines éoliennes en fournissent une aux développeurs de fermes éoliennes. Celle-ci est élaborée à partir de mesures effectuées sur site en fixant un anémomètre sur un mât situé à proximité de l'éolienne et non directement sur l'éolienne même ou trop près de celle-ci, car le vent est ralenti à l'approche de la turbine mais aussi afin de limiter les turbulences qui nuiraient à la fiabilité des mesures. Les fabricants d'éoliennes utilisent par exemple le site d'Osterild au Danemark pour effectuer les tests et fournir aux exploitants une garantie sur la production d'énergie. (photo 2.2)

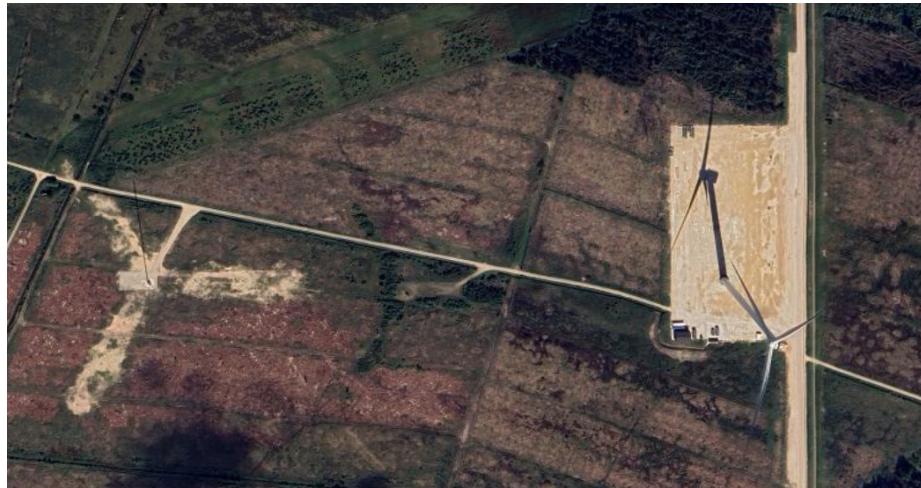


Figure 2.2: Vue aérienne d'une partie du site de test d'Osterild (Danemark) avec à gauche le mât de mesures et à droite l'éolienne testée. Images *Google Earth*.

Suite à ces tests, les fabricants de turbines sont en mesure de fournir une courbe semblable à celle présente sur la figure 2.3.

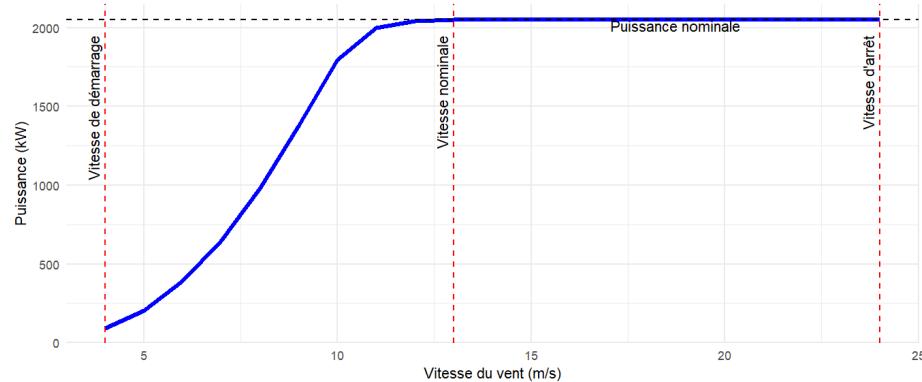


Figure 2.3: Courbe constructeur pour une éolienne de puissance nominale de 2050 kW.

On peut y observer plusieurs phases. La première phase se trouve avant la vitesse de démarrage ($v_{\text{cut-in}}$) où la vitesse du vent est trop faible pour mettre en mouvement l'éolienne. Vient ensuite la phase classique de l'éolienne où le calcul de la puissance en sortie se comporte comme l'équation (2.1). À partir de la vitesse nominale (v_{rated}), la puissance générée par l'éolienne n'augmente plus, car elle a atteint la puissance nominale (P_{rated}). Cette puissance, limitée par le générateur, représente le maximum que l'éolienne peut produire dans des conditions optimales de fonctionnement. Elle est déterminée par le fabricant afin d'assurer un rendement élevé tout en respectant les contraintes techniques et de sécurité de la machine. Enfin, à partir de la vitesse d'arrêt ($v_{\text{cut-out}}$), aux alentours de 24 – 25 m/s, l'éolienne est mise à l'arrêt pour éviter toute détérioration liée aux survitesses de rotation du rotor.

2.1.3 Principales variables d'influence sur la production électrique d'une éolienne

2.1.3.1 Vitesse du vent

La vitesse de vent (Wind Speed) (WS) est évidemment la variable privilégiée dans la production de puissance par une éolienne. Il s'agit plus précisément de la vitesse à hauteur de nacelle moyennée 10 minutes . On observe en Europe de l'Ouest que la distribution de la WS correspond une loi de Weibull dont la fonction de densité peut-être mise sous la forme :

$$f(v, k, \lambda) = \frac{k}{\lambda} \left(\frac{v}{\lambda} \right)^{k-1} e^{-\left(\frac{v}{\lambda}\right)} \mathbb{1}_{\mathbb{R}_+} \quad (2.2)$$

où :

- v désigne la vitesse du vent en m/s,
- k est un paramètre de forme (shape),
- λ est un paramètre d'échelle (scale).

Le graphique 2.4 permet d'observer cette distribution de Weibull de la vitesse du vent :

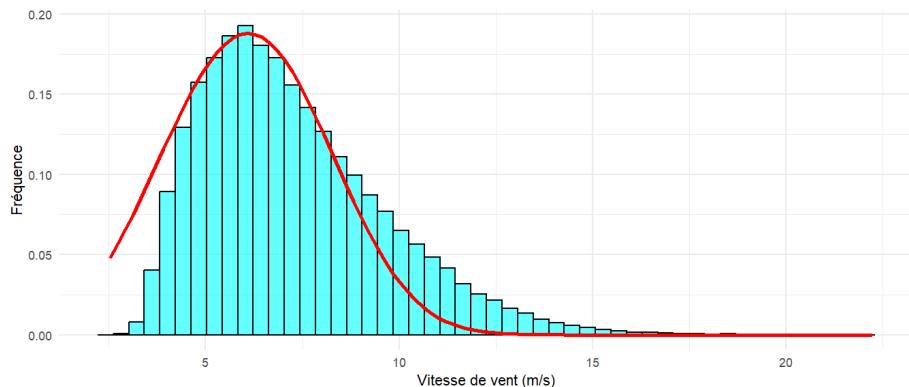


Figure 2.4: Distribution de la vitesse du vent avec la fonction densité de Weibull associée.

2.1.3.2 La densité de l'air

La densité de l'air, exprimée en kg/m^3 , est inversement liée à la température. L'augmentation de la température entraîne une réduction de l'énergie cinétique. La norme définie par l'International Electrotechnical Commission (IEC) [1] recommande de normaliser la vitesse du vent en tenant compte de la densité de l'air, en utilisant une formule qui intègre la température, la pression atmosphérique et l'humidité pour estimer la densité moyenne sur une période de 10 minutes (Voir [1]). Cependant, en l'absence de toutes ces données ou pour simplifier les calculs, une approximation couramment utilisée pour la vitesse normalisée v_{norm} est la suivante :

$$v_{\text{norm}} = v \times \left(\frac{\rho_{10\text{min}}}{\rho_0} \right)^{1/3} = v \times \left(\frac{T_{\text{ref}}}{T_{10\text{min}}} \right)^{1/3} \quad (2.3)$$

où :

- $\rho_{10\text{min}}$ est la densité de l'air moyennée sur 10 minutes,
- T_{ref} est la température de référence en Kelvin ($288^\circ\text{K} = 15^\circ\text{C}$),
- $T_{10\text{min}}$ qui est la température en Kelvin moyennée sur 10 minutes.

La vitesse de vent utilisée dans les différents modèles correspond à une vitesse normalisée par la densité de l'air à hauteur de moyeu.

2.1.3.3 Intensité de turbulence

L'intensité de turbulence intensité de turbulence (Turbulence Intensity) (TI) caractérise les fluctuations du vent au cours du temps. Elle correspond à l'écart-type relatif de la vitesse du vent sur 10 minutes :

$$TI = \frac{\sigma_{v_{10}}}{\bar{v}} \quad (2.4)$$

Pour mieux comprendre l'influence de cette TI, supposons que la fluctuation de la vitesse autour de la moyenne suive une loi Gaussienne centrée en 0 et d'écart-type σ_v . Nous avons vu que, dans la région de la courbe de puissance entre la vitesse de démarrage et le changement de courbure (Fig. 2.3) autour de 10 m/s, P était proportionnelle à v^3 (Eq. (2.1)) et en notant v' la fluctuation de la vitesse autour de la moyenne sur 10 minutes, il vient :

$$\begin{aligned} \bar{P} &\propto \bar{v}^3 = \overline{(\bar{v} + v')^3} \\ &= \overline{(\bar{v})^3} + 3\overline{(\bar{v})^2 v'} + 3\overline{\bar{v} v'^2} + \overline{v'^3} \\ &= \bar{v}^3 + 3\bar{v}^2 \bar{v}' + 3\bar{v} \bar{v}'^2 + \bar{v}'^3 \end{aligned}$$

Or, par hypothèse, $\bar{v}' = 0$ et par symétrie de la distribution $\overline{v'^3} = 0$ d'où :

$$\bar{P} \propto \bar{v}^3 + 3\bar{v}\bar{v}'^2 = \bar{v}^3 (1 + 3TI^2) \quad (2.5)$$

L'équation (2.5) suggère donc un effet positif de l'intensité de turbulence dans la région considérée. On trouvera des explications complémentaires dans [2].

Suivant la résolution temporelle des mesures, ce paramètre sera plus ou moins important. Toutefois, dans les cas rencontrés qui sont ceux définis par les standards internationaux de IEC [1], les intervalles de temps considérés sont de 10 minutes. Il est admis que ce moyennage sur 10 minutes représente 50% de l'effet de turbulence total sur les courbes de puissance [3].

Il y a donc une dépendance importante de la puissance à ce phénomène physique comme le montre la figure 2.5. Sur ce graphique tiré de [4], on peut observer de quelle manière l'intensité de turbulence modifie la production de puissance. Entre la vitesse de départ et environ 9 m/s, on voit un effet légèrement bénéfique de la turbulence : les différentes courbes sont au-dessus de celle de référence pour l'intensité nulle. On retrouve donc ce qui a été mentionné à partir de l'équation (2.5). Au-delà de ce seuil de 9 m/s, plus l'intensité augmente, plus la production est dégradée. Une intensité de turbulence maximale de 30% est un niveau au-delà duquel on pourra considérer que les mesures sont aberrantes.

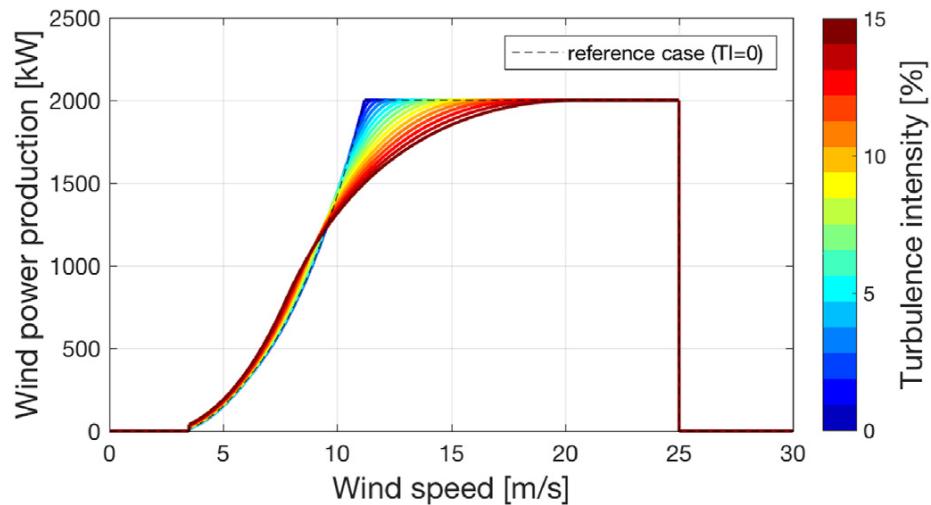


Figure 2.5: Mise en évidence de l'effet de l'intensité de turbulence sur la production de puissance.

2.1.3.4 Cisaillement du vent

Le vent ne s'écoule pas de la même façon au sol ou en altitude. L'influence du type de terrain est caractérisée par un coefficient de rugosité représentant la couverture du terrain. Un terrain recouvert de forêt aura un coefficient de rugosité élevé, à la différence de la mer où ce coefficient est considéré comme nul.

Il en ressort que la vitesse d'écoulement du vent évolue avec l'altitude. La force s'exerçant sur l'éolienne n'est donc pas la même en bas ou en haut des pâles et la puissance produite s'en trouve modifiée. C'est un modèle exponentiel qui est dans la plupart des cas considéré pour décrire le profil de vent sur un terrain simple :

$$v_h = v_{\text{hub}} \times \left(\frac{h}{h_{\text{hub}}} \right)^{\alpha_h} \iff \alpha_h = \frac{\ln \left(\frac{v_h}{v_{\text{hub}}} \right)}{\ln \left(\frac{h}{h_{\text{hub}}} \right)} \quad (2.6)$$

où :

- α_h est le coefficient de cisaillement (ou *shear*) à la hauteur h ;
- h_{hub} est la hauteur de la nacelle ;
- v_{hub} est la vitesse du vent à hauteur de nacelle ;
- v_h désigne la vitesse du vent à la hauteur h .

La figure 2.6 compare deux profils de vent, l'un sans cisaillement ($\alpha = 0$) et l'autre avec ($\alpha > 0$). On note qu'avec la hauteur, la vitesse du vent augmente et ainsi donc l'énergie cinétique. Cela va donc amener une contrainte mécanique sur les pâles soumises à une force plus élevée en haut qu'en bas. Généralement, la valeur du cisaillement est comprise entre 0 et 0.3.

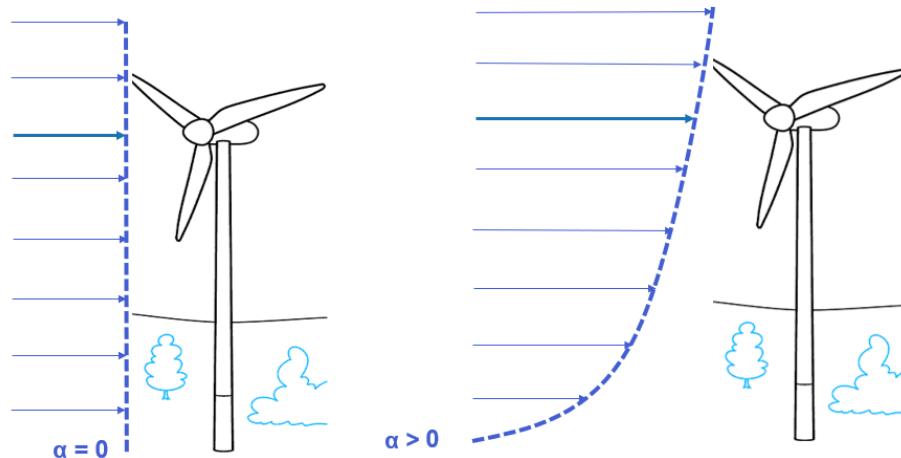


Figure 2.6: Influence du cisaillement sur la vitesse du vent.

2.1.3.5 Variation de direction avec l'altitude

Comme vu précédemment, le vent ne souffle pas à la même vitesse selon l'altitude. Il ne souffle pas non plus dans la même direction horizontale. Ce coefficient traduisant cette variation de direction en fonction de la hauteur est communément appelé *veer* et s'exprime en °/m. Il quantifie la différence de direction entre le vent à hauteur de nacelle et le vent à une autre hauteur.

2.1.3.6 Autres facteurs

D'autres paramètres vont influer sur la production de puissance, parmi lesquels on peut citer :

- l'humidité qui va augmenter la densité de l'air ;
- la composante verticale de la vitesse du vent.

Ne sont pas listées la température et la pression, nous avons vu à la section 2.1.3.2 qu'elles étaient très liées à la densité de l'air. Tous ces facteurs cités ne sont pas considérés dans la suite de ce rapport du fait de leur faible influence lorsque l'on utilise la vitesse normalisée par la densité de l'air.

2.1.4 Production annuelle d'énergie - Annual Energy Production (AEP)

L'AEP est une estimation annuelle de la production d'énergie. Elle correspond à l'espérance de la courbe de puissance pour la loi de la vitesse du vent :

$$AEP = N_h \int_{\mathbb{R}} P(v) f(v; k; \lambda) dv \quad (2.7)$$

où :

- N_h est le nombre d'heures dans une année : 8766 en moyenne ;
- P la fonction correspondant à la courbe de puissance de l'éolienne ;
- $f(v; \lambda; k)$ la distribution de Weibull (Eq. (2.2)).

L'AEP est un indicateur primordial dans le développement d'un projet éolien et sa détermination doit être la plus réaliste possible. Si l'on dispose d'une valeur de puissance par heure sur toute l'année, l'AEP se calculera simplement en faisant la somme de toutes ces puissances :

$$AEP = \sum_i P_i \quad (2.8)$$

où P_i désigne la puissance produite à l'heure i

Le calcul sera donc sujet à fluctuation si on ne dispose justement pas de ces valeurs horaires de puissance.

2.2 Objectifs du stage

2.2.1 Cas d'usage

Dans le cadre d'une Etude de Potentiel Eolien (EPE), une CMV est menée afin de caractériser le régime éolien local à l'aide d'un dispositif de mesure tel qu'un mât instrumenté (doté d'anémomètres et de girouettes à différentes hauteurs) ou un LiDAR déployé au sol. La durée recommandée de cette campagne est d'au moins une année complète, afin de capter l'ensemble des variations inter-saisonnieres des paramètres atmosphériques d'intérêt, notamment la vitesse, la direction du vent, et la turbulence.

Une fois la série de données mesurées consolidée et validée, une corrélation de long terme est réalisée avec des données météorologiques historiques provenant de modèles météorologiques, dits de réanalyse, (type ERA5, MERRA-2) ou de données satellitaires. Cette étape s'appuie généralement sur une méthodologie MCP ayant pour but de "corriger" les données de réanalyse afin de s'ajuster au mieux aux conditions réelles, permettant ainsi de produire une série synthétique de longue durée (typiquement 10 à 20 ans), représentative des conditions de vent sur site.

À partir de cette série reconstruite, on calcule les statistiques clés (moyennes, distributions, percentiles, etc.). Des indicateurs énergétiques sont aussi calculés, tels que le vent moyen à hauteur de nacelle, le facteur de charge, ou l'AEP. L'évaluation de cette dernière grandeur se fait par l'utilisation de la courbe constructeur en convertissant la ressource modélisée en vent en puissance potentiellement délivrée par l'éolienne sur le long terme. Ces indicateurs servent de base à l'évaluation technico-économique du projet et à l'estimation de sa rentabilité sur l'ensemble de sa durée de vie.

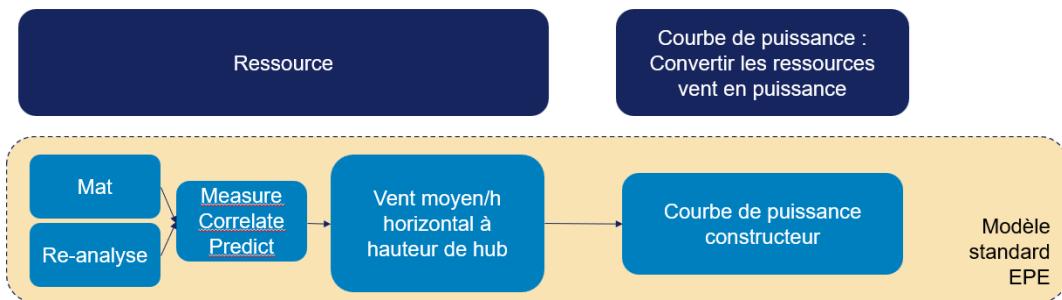


Figure 2.7: Cas d'usage d'une étude de potentiel éolien.

2.2.2 Modélisation longue durée

Toutefois, les modèles de corrélation long terme actuellement utilisés dans le cadre des méthodologies MCP ne tiennent généralement pas compte de l'intensité de turbulence ni du cisaillement, alors même que, comme présenté dans la section 2.1.3, ces deux variables ont un impact significatif sur la courbe de puissance des éoliennes, et donc sur l'estimation de l'AEP.

L'un des objectifs majeurs de ce stage consiste à développer une méthodologie MCP enrichie permettant de générer une série de long terme de la vitesse et de la direction du vent, mais également de l'intensité de turbulence et du cisaillement. À partir de cette série multivariée reconstituée, un modèle de courbe de puissance avancé est calibré de manière à restituer une série de long terme de la puissance attendue, que nous comparons ensuite aux données de production mesurées sur site.

Enfin, nous étudions dans quelle mesure ces modèles de courbe de puissance peuvent être généralisés ou transférés à d'autres éoliennes, qu'elles soient du même modèle ou non, implantées sur le même parc ou sur d'autres sites. L'objectif à long terme étant de renforcer

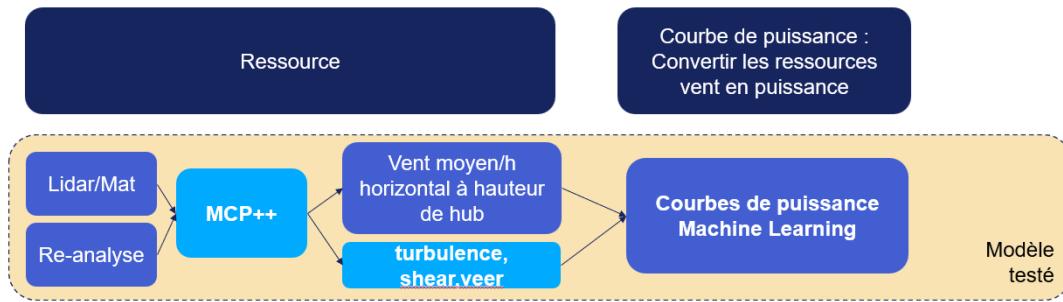


Figure 2.8: Cas d'étude du stage.

la robustesse des évaluations de production afin de s'affranchir de la courbe constructeur pour la remplacer par notre modèle de courbe de puissance.

3

Campagnes de mesures et type de données utilisées

La figure 3.1 présente les positions des différentes fermes éoliennes étudiées. On remarque qu'elles se trouvent toutes dans le Nord de la France où le potentiel en vent est en moyenne plus élevé que sur le reste du territoire.

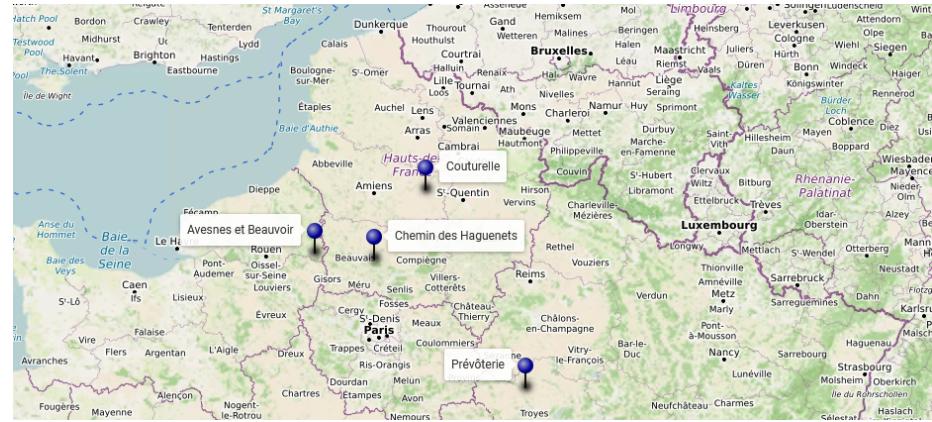


Figure 3.1: Position géographique des parcs étudiés.

Dans chacun de ces parcs, nous regarderons deux éoliennes différentes.

3.1 Présentation détaillée des parcs étudiés

3.1.1 Avesnes

Le site d'Avesnes se situe en Seine-Maritime. Il est composé de deux parcs distincts mais très proches géographiquement : AEB et Avesnes et Bosc-Hyons (ABH). Sur ces parcs, nous nous intéressons, pour AEB, à une éolienne V90 (fabricant Vestas dont le diamètre des pales fait 90 m) de puissance nominale 2 MW, et pour ABH à une éolienne V117 de puissance nominale 3 MW.

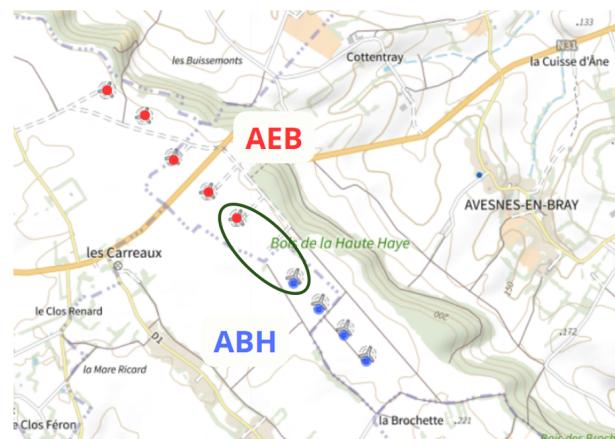


Figure 3.2: Schéma d'implantation des éoliennes des parcs d'Avesnes.

Ces deux éoliennes ont été respectivement installées en 2011 et 2021, et sont distantes d'environ 550 m.

3.1.2 Chemin des Haguenets

Il s'agit ici de nouveau de deux parcs situés en Picardie et très proches géographiquement. Chemin des Haguenets (CDH), sur lequel nous étudions une éolienne MM92 (MM : Senvion) dont la puissance nominale est de 2.05 MW installée en 2009. La seconde éolienne, une V110 mise en service en 2021 et éloignée de 980 m de la première, appartient au parc CDS.

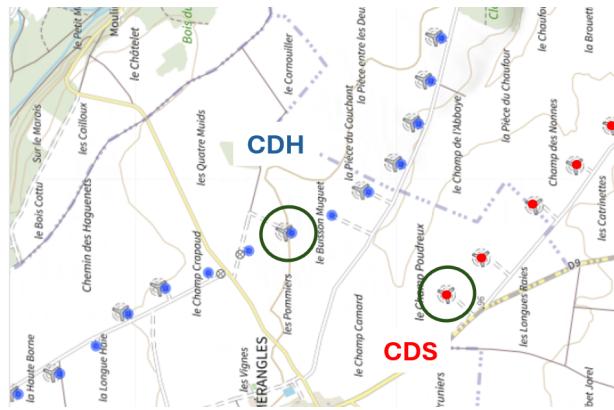


Figure 3.3: Schéma d'implantation des éoliennes des parcs du Chemin des Haguenets.

3.1.3 Couturelle

Sur le parc COF se trouvant dans le département de la Somme, nous nous intéressons aux deux premières éoliennes COF1 et COF2, des MM92 de 2.05 MW écartées de 400 m.

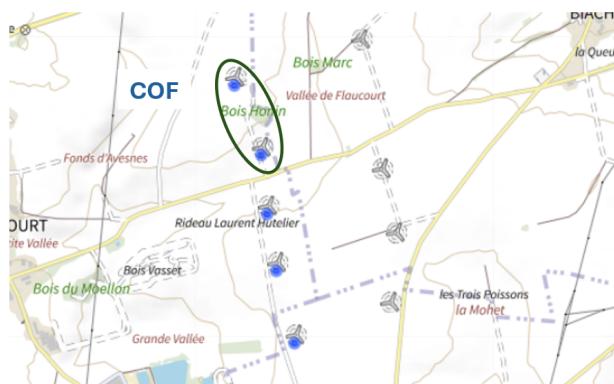


Figure 3.4: Schéma d'implantation des éoliennes du parc Couturelle.

Ces deux éoliennes ont été mises en service en 2017 et 2020.

3.1.4 Prévoterie

Le site Aubois de Prévoterie est composé de deux parcs : Prévôterie 1 (PRP) et de son extension Prévôterie 2 (EDS). Sur PRP, nous aurons à notre disposition les données d'une

MM92 de 2.05 MW de puissance nominale mise en service en 2010. Sur le parc EDS, il s'agira d'une éolienne V100 de 2 MW construite en 2021 qui est écartée de 550 m de l'éolienne de PRP.

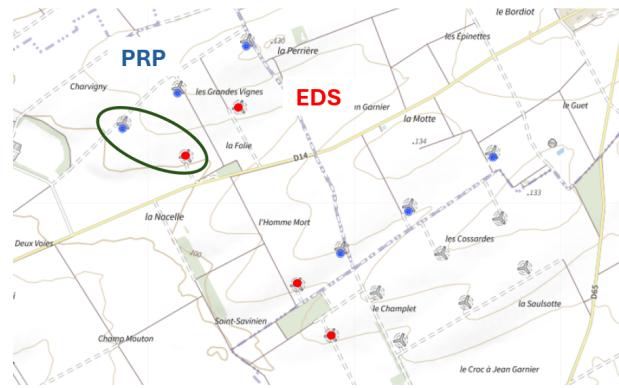


Figure 3.5: Schéma d'implantation des éoliennes des parcs de Prévoterie.

Le tableau 3.1 récapitule les caractéristiques des éoliennes étudiées :

Éolienne	ABH	AEB	CDH	CDS	COF1	COF2	EDS	PRP
Mise en service	2021	2011	2009	2021	2017	2020	2021	2010
Constructeur	Vestas	Vestas	Senvion	Vestas	Senvion	Senvion	Vestas	Senvion
Diamètre du rotor [m]	117	90	92.5	110	92.5	92.5	100	92.5
Puissance nominale [MW]	3	2	2.05	2.2	2.05	2.05	2	2.05
Hauteur de nacelle [m]	80	80	78.5	80	80	80	80	80

Table 3.1: Caractéristiques des différentes éoliennes étudiées.

3.2 Types de données disponibles

3.2.1 Données de Campagne de Mesure de Vent (CMV)

Les CMV remplissent plusieurs rôles et peuvent s'appuyer sur différents dispositifs. Historiquement, l'installation de mâts de mesure a toujours constitué une étape préalable à la construction d'un parc éolien, dans le but d'évaluer le potentiel éolien d'un site. On parle alors de mâts de préconstruction ou de mâts de développement.

Des anémomètres, ainsi que d'autres capteurs (de direction du vent, température, pression atmosphérique, etc.), sont installés à différentes hauteurs sur ces mâts. Ils recueillent des données sur une période idéale d'au moins un an, afin d'avoir une image représentative et fiable du climat venteux local.

En complément, des mâts d'exploitation peuvent également être déployés pendant la phase de fonctionnement du parc. Leur rôle est alors de surveiller en continu les conditions de vent, de vérifier la cohérence entre la production réelle et les prévisions, et de détecter d'éventuelles dérives de performance. Ces mesures permettent ainsi d'assurer une exploitation optimale, d'anticiper la maintenance, et de garantir la rentabilité à long terme du parc.

éolien.

Les exploitants éoliens utilisent aussi des LiDAR qui sont des instruments de mesure pouvant être posés au sol dans une phase de préconstruction ou sur la nacelle (Fig.3.6) en postconstruction. Grâce à plusieurs faisceaux lasers, ils enregistrent de nombreuses données physiques à plusieurs hauteurs et distances. Leur facilité de pose et leur fiabilité en font des outils de plus en plus utilisés par les exploitants.

Le tableau 3.2 récapitule les différents usages de ces dispositifs.



Figure 3.6: Lidar à quatre faisceaux lasers sur la nacelle d'une éolienne.

Dans la section 5.1.1.1, nous utilisons des données de post-construction pour le développement de nos modèles pour finalement adopter (section 5.1.2) des données de pré-construction pour leur amélioration. Le tableau 3.2 présente les différents dispositifs utilisés pour une CMV, leur utilisation dans l'industrie ainsi que leur usage dans ce rapport.

3.2.2 Données opérationnelles SCADA

Les données SCADA (Supervisory Control And Data Acquisition) d'une éolienne sont des données opérationnelles généralement enregistrées à fréquence 1 Hz puis moyennées 10 minutes par le système de contrôle de la machine. Elles comprennent des mesures clés telles que la vitesse du vent au moyeu, la puissance produite, la direction du vent, l'état de fonctionnement de l'éolienne, ainsi que des informations internes comme la température des composants ou la vitesse de rotation du rotor. Ces données sont essentielles pour analyser les performances de l'éolienne, détecter les anomalies, établir une courbe de puissance réelle, et estimer l'AEP en post-construction.

Afin de disposer de données de référence fiables, un nettoyage préalable est effectué à l'aide d'une application interne développée par *Engie Green*. Cette étape vise à ne conserver que les horodatages dits "normaux", en excluant ceux associés à des situations de bridage, de valeurs aberrantes ou d'arrêts d'éoliennes.

La figure 3.7a permet d'observer tous les enregistrements des capteurs de l'éolienne. Un premier nettoyage est effectué à l'aide de formules physiques traduisant un comportement normal de la turbine, (Fig.3.7b) puis les derniers points aberrants sont sélectionnés manuellement avant d'être classés dans une catégorie particulière.

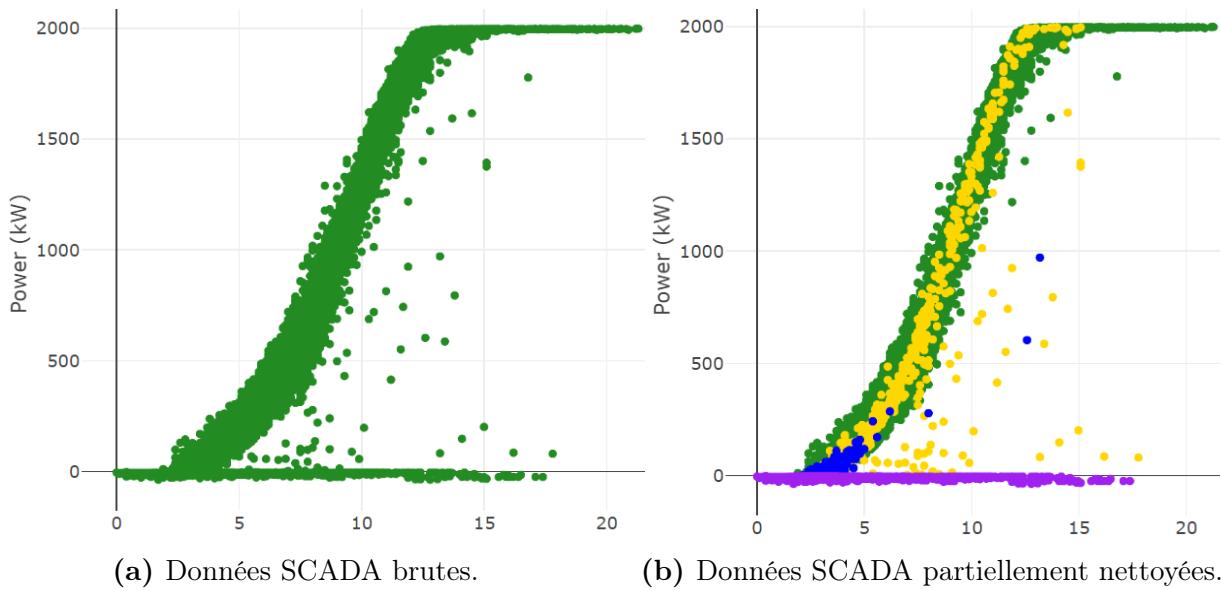


Figure 3.7: Nettoyage de données SCADA.

3.2.3 Données de réanalyses

Les données de réanalyses sont des estimations météorologiques obtenues en combinant des observations historiques (stations météo, satellites, etc.) avec des modèles numériques de prévision du temps. Elles fournissent une description cohérente et continue de l'état de l'atmosphère sur plusieurs décennies, avec une résolution spatiale allant de quelques kilomètres à plusieurs dizaines de kilomètres, et une résolution temporelle horaire à quotidienne. Dans le contexte de l'énergie éolienne, les données de réanalyses — comme celles issues des jeux ERA5 ou MERRA-2 — sont utilisées pour modéliser les conditions de vent à long terme sur un site donné. Elles servent notamment dans les méthodes MCP pour extrapoler les mesures de vent d'un mât ou d'un LiDAR à une série temporelle plus longue. Ces données incluent typiquement la vitesse et la direction du vent à différentes altitudes, la température, la pression ou encore l'humidité, et peuvent aussi intégrer des diagnostics comme l'intensité de turbulence ou le cisaillement vertical du vent selon les modèles.

Phase du projet de fermes éoliennes	Dispositifs de mesure	Variables mesurées	Utilisation dans le processus industriel	Distance par rapport à l'éolienne	Utilisation dans ce rapport
Test	Mât de test	Vent moyen, direction, température, pression, cumul de précipitations, humidité	Utilisé par les constructeurs	$2.5 \times D_{\text{rotor}}$	Pas d'accès
Pré-Construction	LiDAR au sol	Vent moyen, direction, turbulence, <i>veer, shear</i> , température, densité de l'air etc.	Mesurer la ressource en vent sur un site donné	Jusqu'à plusieurs kilomètres	Pas eu d'accès
	Mât de mesure	Vent moyen, direction, température, pression, cumul de précipitations, humidité	Mesurer la ressource en vent sur un site donné	Jusqu'à plusieurs kilomètres	Utilisation pour faire des extrapolations long terme dans un second temps.
Post-construction	Anémomètre	Vent à hauteur de nacelle	Vérifier les conditions de vent sur le site d'installation	0 m	Utilisé comme valeur de référence pour valider les algos.
	LiDAR nacelle	Vent moyen, direction, turbulence, <i>veer, shear</i> , température, densité de l'air etc.	Mesurer la courbe de puissance pour vérifier que la turbine produit au bon niveau	0 m	Utilisation en imitant une phase de pré-construction avec des données de post-construction

Table 3.2: Récapitulatifs des différents dispositifs utilisés lors d'une CMV.

4

Méthodologie Measure-Correlate-Predict et modèles de courbe de puissance

4.1 L'approche Measure-Correlate-Predict pour l'estimation de la ressource

4.1.1 Principe général et objectifs de prédition

4.1.1.1 La méthode "Measure-Correlate-Predict"

La moyenne annuelle de la vitesse est très variable d'une année à l'autre, comme l'illustre la figure 4.1.

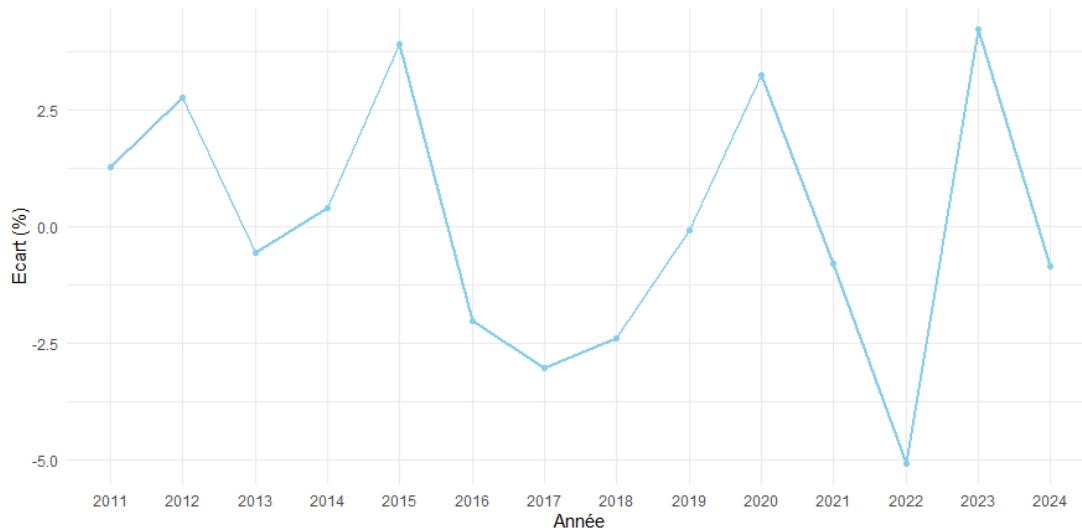


Figure 4.1: Écart à la moyenne pluriannuelle des moyennes annuelles de la vitesse de vent issue des données SCADA pour l'éolienne AEB.

Or il est indispensable pour les exploitants de parcs éoliens d'avoir une bonne estimation de la ressource en vent pour rendre viable le projet aux yeux d'éventuels financeurs. Les données CMV ne suffisent pas car n'intègrent pas la variabilité pluriannuelle des ressources en vent. Ces données vont donc servir à "corriger" celles de réanalyse et ainsi obtenir une série long terme plus réaliste. Le schéma 4.2 illustre cette méthodologie.

Les données issues de la CMV, enregistrées toutes les 10 minutes, sont moyennées pour obtenir des valeurs horaires. Dans le cas d'usage d'un développement d'un parc éolien, toutes les données provenant du mât de pré-construction vont pouvoir être utilisées. Toutefois, dans notre cas de test, les données proviennent de LiDAR nacelle dont les données ont été filtrées pour être exemptes de bridage, de sillage ou d'anomalies. Afin que la moyenne horaire reflète correctement les conditions réelles et dans un souci d'homogénéité, seules les heures comportant six mesures valides sont conservées.

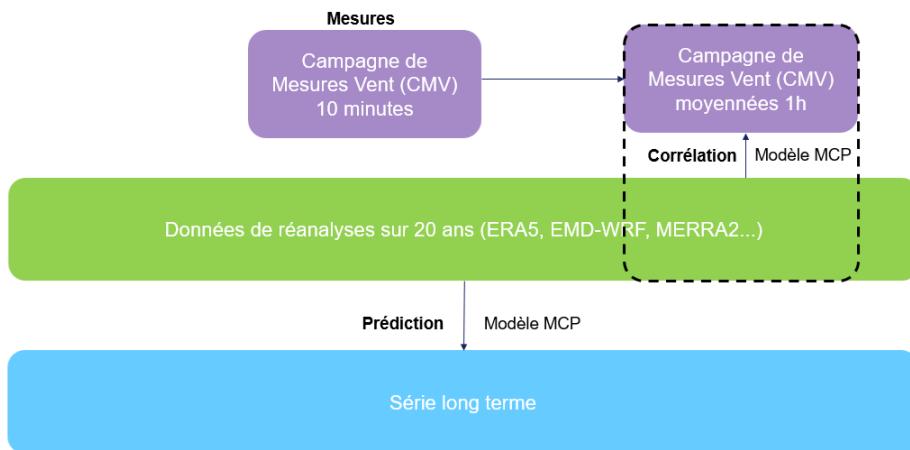


Figure 4.2: Principe de prédiction Measure-Correlate-Predict.

Une fois les données CMV mises en correspondance horaire avec les données de réanalyse, chaque variable issue de la CMV peut être "corrélée" aux variables provenant des données de réanalyse. On utilise pour chaque variable issue de la CMV, l'ensemble des variables choisies de la réanalyse pour effectuer cette "corrélation". Il s'agit ici d'une modélisation univariée : pour chaque variable de la CMV, un modèle individuel est établi, et non de manière conjointe ou multivariée. Ces modèles sont ensuite appliqués à l'ensemble de la série de réanalyse afin de reconstituer une série long terme corrigée des variables d'intérêt.

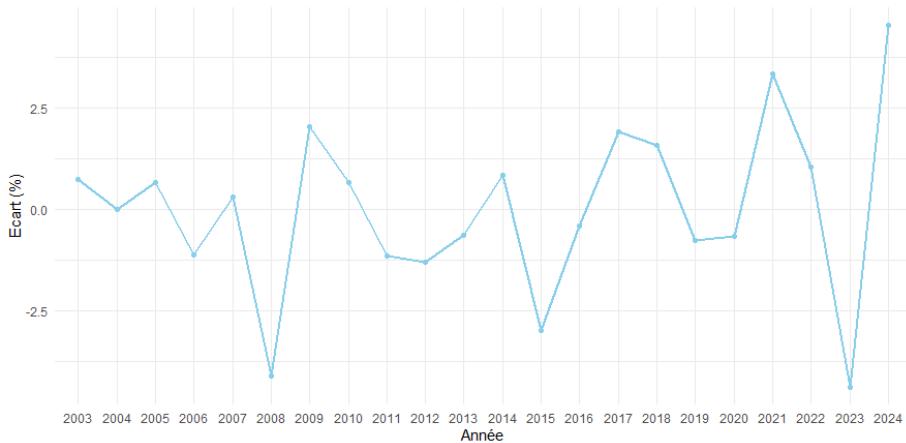
4.1.1.2 Les objectifs de prédictions

La TI, le *shear* et, dans une moindre mesure, la *shear* sont trois variables liées à la vitesse du vent. Il paraît donc opportun d'imaginer qu'elles aient aussi des variations interannuelles non négligeables. Comme on peut le voir sur les graphiques 4.3, l'écart en pourcentage par rapport à la moyenne long terme varie de manière conséquente. Cela nous pousse donc à tenter d'obtenir une série long terme pour les variables d'intensité de turbulence, de cisaillement mais aussi de déviation verticale du vent. Les variations de cette dernière variable n'ont pu être affichées du fait de l'absence de mesures sur le long terme.

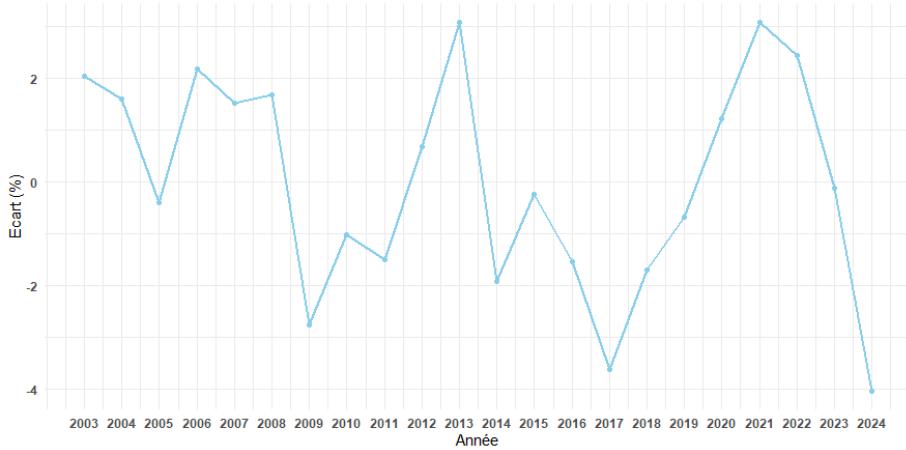
4.1.2 Les différents modèles de prédiction testés

Les variables à prédire sont donc :

- la vitesse du vent normalisée par la densité de l'air à hauteur de moyeu ;
- l'intensité de turbulence ;
- le cisaillement ;
- la déviation verticale de la direction du vent si utilisation de données LiDAR ;



(a) Ecart moyen normalisé à la moyenne pour l'intensité de turbulence.



(b) Ecart moyen normalisé à la moyenne pour le cisaillement.

Figure 4.3: Ecart moyen annuel pour l'intensité de turbulence et le cisaillement obtenu à partir des données de réanalyse.

- la direction du vent si utilisation de données de mât.

Les variables suivantes, choisies à la hauteur la plus proche de celle de la nacelle et issues des données de réanalyse, seront toutes utilisées comme prédicteurs pour chacune des variables à prédire :

- la vitesse du vent ;
- l'intensité de turbulence ;
- la direction du vent ;
- le cisaillement dont on analysera l'apport (Sec. B).

Le cisaillement n'est pas directement présent dans les données de réanalyses, mais peut être estimé du fait des modélisations à plusieurs hauteurs de la vitesse du vent (Sec. 2.1.3.4).

Enfin, différents modèles d'apprentissage statistique sont mis en œuvre à des fins de comparaison. L'analyse reste dans un cadre univarié, c'est-à-dire que les modèles sont développés variable par variable, et non de manière conjointe ou multivariée. Une grille de recherche de paramètres optimaux a été mise en place ainsi qu'une procédure de validation croisée 5-folds répétée 10 fois pour s'assurer de la robustesse des résultats obtenus :

- la régression linéaire classique : un modèle paramétrique simple qui suppose une relation linéaire entre la variable dépendante et les variables explicatives. Il sert de modèle de référence pour évaluer les performances des méthodes plus complexes ;
- les k -Nearest Neighbors (k -NN) : une méthode non paramétrique fondée sur la similarité locale, où la prédiction est basée sur les valeurs des k observations les plus proches dans l'espace des variables explicatives. Le choix a été fait de centrer et réduire les observations et de faire varier le nombre de plus proches voisins k entre 5 et 9 pour en déterminer le nombre optimal ;
- les forêts aléatoires (RF) : une autre méthode d'ensemble reposant sur la construction de multiples arbres de décision construits à partir d'échantillons aléatoires des données, avec une agrégation finale par moyenne. La grille de recherche des paramètres optimaux s'est réduite au nombres d'arbres (500 ou 1000) ainsi au choix du nombre de variables sélectionnées aléatoirement à chaque division lors de la construction d'arbres de décision ;
- le Gradient Boosting Model (GBM) : un algorithme d'ensemble itératif qui combine plusieurs arbres de décision faibles pour construire un modèle robuste et performant, particulièrement adapté aux relations non linéaires et aux interactions complexes. L'ajustement s'est fait sur des données centrées réduites, un nombre d'arbres de 500 ou 1000, et soit 5 ou 10 observations minimum dans les derniers noeuds de l'arbre.

4.1.3 Critères de choix du modèle

Afin d'évaluer la qualité des modèles, c'est-à-dire la fiabilité de leurs prédictions, plusieurs indicateurs peuvent être envisagés. Dans le domaine éolien, l'indicateur le plus couramment utilisé est le RMSE, défini par :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{pred},i} - y_i)^2} \quad (4.1)$$

où y_i désigne la valeur réelle de l'observation i et $y_{\text{pred},i}$ sa valeur prédite.

Toutefois, dans le cadre de la comparaison de différents parcs éoliens exposés à des conditions de vent hétérogènes au cours de l'année, il est plus pertinent d'utiliser une version normalisée du RMSE. Nous utilisons ainsi le NRMSE, obtenu en divisant le RMSE par la moyenne arithmétique des observations \bar{y} , puis exprimé en pourcentage :

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{pred},i} - y_i)^2}}{\bar{y}} \quad (4.2)$$

Enfin, nous évaluons également le biais de la prédition, qui mesure la tendance systématique du modèle à surestimer ou à sous-estimer les valeurs réelles :

$$\text{Biais} = \overline{\hat{y}_{\text{pred}}} - \bar{y} \quad (4.3)$$

où $\overline{\hat{y}_{\text{pred}}}$ désigne la moyenne arithmétique de la variable prédite.

Cet indicateur permet de vérifier que le modèle ne s'écarte pas systématiquement de la réalité dans un sens donné.

4.1.4 Les cas d'entraînement

Les différents modèles testés ont été ajustés à l'aide de données LiDAR issues d'une CMV dont la durée varie de 2 à 9 mois selon les éoliennes considérées (Tab. 4.1).

Afin de réaliser la CMV, ces données doivent être mises en correspondance avec les données de réanalyse, dont les horodatages sont fournis à une fréquence horaire. Il est donc nécessaire d'agrégner les mesures LiDAR à une résolution horaire. Toutefois, seules les heures comportant au moins six mesures distinctes sont retenues dans ce calcul. Nous faisons en effet ce choix dans un souci de cohérence des données, et avoir moins de six mesures par heure pourrait se révéler insuffisant pour refléter la forte variabilité de la vitesse du vent, même à l'échelle de quelques minutes.

Ce filtrage et le moyennage réduisent le volume initial de données d'un facteur d'environ 11. Enfin, parmi les données ainsi conservées, 80 % sont utilisées pour la calibration des modèles, tandis que les 20 % restantes sont réservées à la validation des performances prédictives.

Éolienne	ABH	AEB	CDH	CDS	COF1	COF2	EDS	PRP
Début	04-06-2022	29-03-2022	19-09-2023	18-08-2023	04-03-2022	13-03-2022	09-04-2023	09-04-2023
Fin	24-01-2023	30-12-2023	31-12-2023	27-09-2023	16-05-2022	26-08-2022	24-09-2023	24-09-2023
Nb données	11323	11552	2484	1180	5486	4438	10836	8021
Moyenne WS [m/s]	6.71	6.68	6.97	5.78	6.86	6.16	6.14	6.44
Ecart-type WS [m/s]	2.47	2.53	2.83	1.77	2.27	2.29	2.52	2.28
Moyenne TI [%]	11.04	11.40	11.33	12.45	12.39	11.36	12.83	11.29
Moyenne shear []	0.25	0.25	0.30	0.25	0.24	0.27	0.24	0.25
Nb données d'entraînements	831	1048	191	101	523	379	654	1059

Table 4.1: Synthèse des données Lidar disponibles pour les éoliennes étudiées.

On note dès à présent que pour l'éolienne CDS, la moyenne et l'écart-type de la vitesse sont nettement inférieurs notamment à l'éolienne CDH se trouvant à proximité. Des disparités importantes du nombre de valeurs d'entraînement selon les éoliennes sont également présentes.

4.1.5 Cas de test

Afin de valider notre modèle, nous utilisons des données n'ayant pas servi à son entraînement :

Éolienne	AEB	ABH	CDH	CDS	COF1	COF2	EDS	PRP
Nb données de validation	262	208	48	25	131	95	163	265

Table 4.2: Synthèse des données Lidar disponibles pour les éoliennes étudiées.

Il ressort clairement que le nombre de données de validation pour les éoliennes CDH, CDS, COF1 et COF2 est faible. Des sources de données alternatives, plus nombreuses, sont donc nécessaires pour tenter d'expliquer la variance de ces variables.

4.1.6 Utilisation des données de mât

Nous avons décidé de regarder, en plus des quatre éoliennes mentionnées précédemment (CDH, CDS, COF1 et COF2), les données de mât associées aux quatre autres (ABH, AEB, EDS et PRP). En effet, cela nous permet de nous affranchir des données LiDAR qui ont été considérées au départ car correspondant aux mesures exactes des conditions de vent à hauteur de nacelle mais qui ont le défaut d'avoir été filtrées notamment à basse vitesse. Les données de mât, certes ayant un plus gros volume, proviennent parfois de mâts situés à plusieurs kilomètres de l'emplacement de l'éolienne une fois construite.

Les variables à prédire sont celles présentes dans les données de mât :

- la vitesse de vent, normalisée à posteriori à l'aide des données de réanalyse ;
- l'intensité de turbulence ;
- la direction du vent ;
- le cisaillement, excepté pour COF où une seule hauteur de mesure de vitesse du vent est disponible.

Nous considérons les mêmes prédicteurs issus des données de réanalyse : WS, TI, direction du vent (Wind Angle) (WA) et le shear. Le tableau 4.3 récapitule les données de mât pour les différents parcs. On y note des conditions de vent mesurées très différentes d'un parc à l'autre.

Eoliennes	ABH/AEB	CDH/CDS	COF1/COF2	EDS/PRP
Début	2017-11-17	2016-10-27	2022-02-23	2006-12-19
Fin	2019-06-12	2017-11-07	2025-06-01	2015-05-14
Données disponibles	69280	53074	171609	125570
Moyenne WS [m/s]	6.33	5.02	5.89	5.42
Ecart-type WS [m/s]	2.73	2.58	2.80	3.45
Moyenne TI [%]	11.22	12.82	16.64	15.37
Moyenne shear []	0.26	0.22	NA	0.14

Table 4.3: Données de mât disponibles pour les différents parcs étudié.

Les modèles de prédiction des variables de mât suivent la même procédure que celles issues des données LiDAR, à savoir un 5-folds répété 10 fois. La comparaison se fait avec les vitesses opérationnelles SCADA dont on retrouve un résumé dans le tableau 4.4.

Eolienne	ABH	AEB	CDH	CDS	COF1	COF2	EDS	PRP
Début	2021-12-01	2011-04-01	2017-05-17	2021-03-24	2017-02-21	2022-03-15	2021-12-08	2016-10-06
Fin	2024-12-31	2024-12-31	2024-04-25	2024-04-09	2024-12-31	2024-12-31	2024-01-01	2024-01-01
Nb données	14715	88526	26941	13211	44653	15532	10347	37407
Moyenne WS [m/s]	7.59	7.26	6.97	5.97	6.52	6.31	6.86	6.61
Ecart-type WS [m/s]	2.61	2.38	2.54	2.51	2.22	2.25	2.47	2.18

Table 4.4: Statistiques des mesures de la vitesse opérationnelle SCADA.

4.2 Estimation du productible

4.2.1 État de l'art

Depuis une dizaine d'années, les modèles de machine learning sont utilisés afin d'obtenir des prédictions de la puissance et d'étudier l'influence de certains paramètres. A. Clifton en 2014 dans [3] étudie l'influence de la turbulence sur la production et compare différents modèles de prédictions de puissance sur des données simulées. Il en ressort (Fig. 4.4) une efficacité supérieure des méthodes de prédictions par RF par rapport aux méthodes par tranches ou avec normalisation de la turbulence présente dans l'[1] .

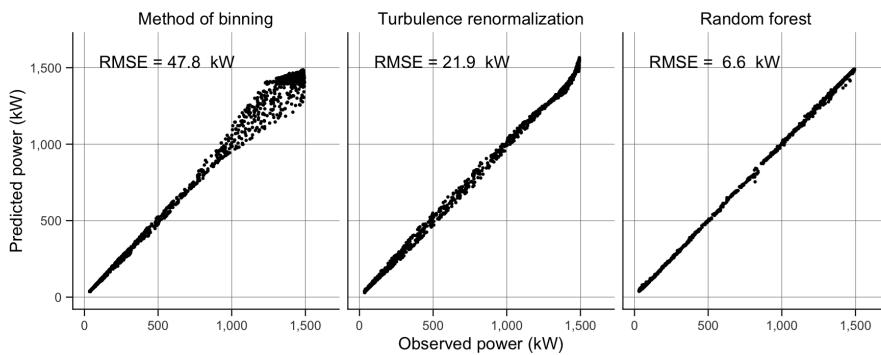


Figure 4.4: Comparaison entre la puissance observée et la puissance prédite selon trois techniques de modélisation de la puissance.

Dans sa thèse [5], Mamadou Dione explore une approche de prévision à court terme basée sur des techniques de Machine Learning. Il y évalue l'efficacité de modèles statistiques classiques en s'appuyant sur différentes données de réanalyse spatialisées autour de la turbine visée par la prévision. Deux stratégies sont comparées : une approche directe, où la vitesse du vent issue des données de réanalyse est utilisée pour prédire directement la puissance, et une approche indirecte, consistant d'abord à estimer la vitesse de vent normalisée à la hauteur de la nacelle, puis à en déduire la puissance. Un des résultats de son travail met en évidence la supériorité de l'approche indirecte (Fig. 4.5), ainsi que la performance notable des forêts aléatoires (RF) pour ce type de tâche prédictive. Nous nous inspirons donc de son approche indirecte pour modéliser la puissance.

4.2.2 Différents modèles de courbes de puissance

Dans cette section, nous présentons les différents modèles de courbes de puissance sur lesquels repose notre analyse. Certains d'entre eux s'appuient uniquement sur des données brutes, tandis que d'autres nécessitent une phase d'apprentissage préalable pour pouvoir estimer la puissance.

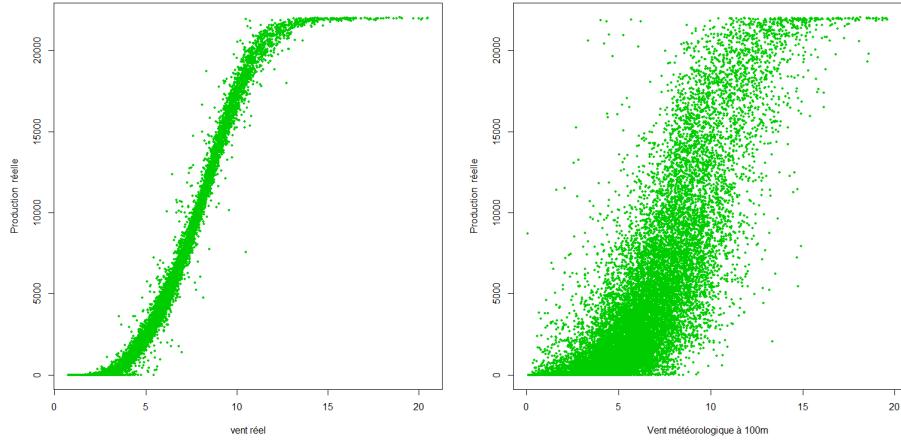


Figure 4.5: Comparaison entre l'approche indirecte à gauche et directe à droite pour la prévision de la puissance P .

4.2.2.1 Courbe de puissance constructeur

Ce modèle, que l'on pourra considérer comme une référence, est un modèle fourni par les constructeurs d'éoliennes. Il s'agit d'un nuage de points $\{(v, P)\}$ que l'on interpole afin qu'à chaque vitesse de vent v corresponde une valeur de puissance P . La figure 4.6 illustre ce propos :

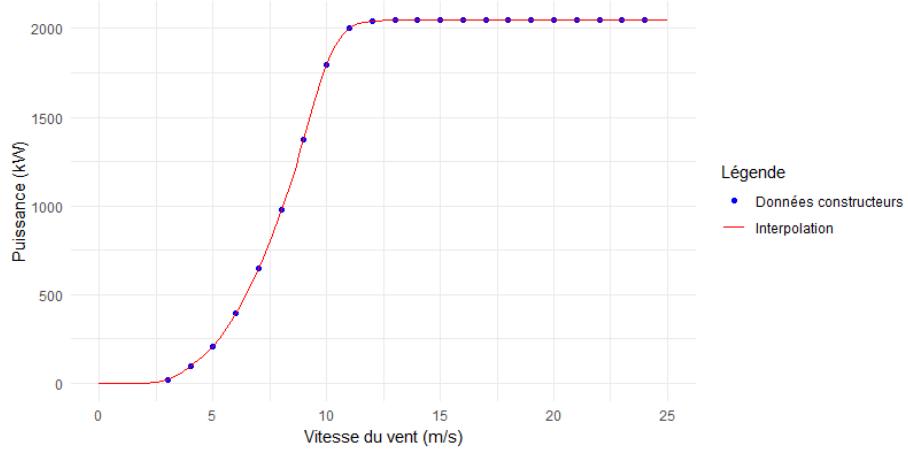


Figure 4.6: Courbe de puissance constructeur et son interpolation.

Dans un souci de simplification, nous ne distinguons plus le nuage de point et son interpolation pour parler de la courbe de puissance constructeur dans toute la suite du rapport.

4.2.2.2 Modèle avec normalisation de l'intensité de turbulence

Le document de référence de l'IEC [1] propose un modèle physique, et donc explicable, de courbe de puissance tenant explicitement compte de l'intensité de turbulence. En pratique, une partie significative de l'effet de la turbulence provient du fait que la puissance

délivrée et la vitesse du vent sont moyennées sur des intervalles de dix minutes. Pour éliminer ce biais, la norme recommande de ramener toutes les mesures à une intensité de turbulence de référence fixée à 10%.

Cette mise à l'échelle s'effectue à l'aide d'un algorithme normatif, schématisé en annexe A, qui transforme les données brutes en une courbe de puissance "normalisée". Sur la figure 4.7, on observe la différence entre la courbe constructeur et la courbe corrigée après normalisation de la turbulence.

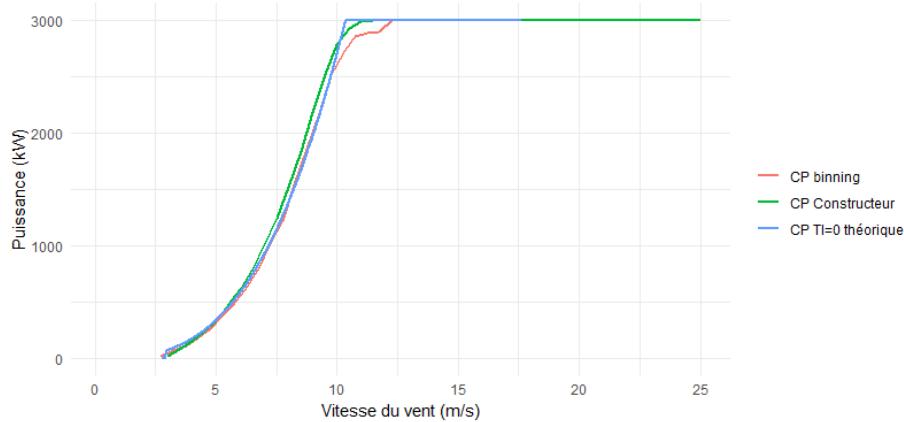


Figure 4.7: Comparaison des courbes de puissance constructeur, par tranches à partir des données de la campagne LiDAR (CP binning) et celle obtenue par normalisation des turbulences (CP TI=0 théorique).

4.2.2.3 Modèles de prédiction basés sur les données de la campagne LiDAR

La précédente stagiaire encadrée par Paul Mazoyer, Léna Lacomme, a développé plusieurs modèles visant à prédire la puissance produite P , à partir de certaines variables mesurées par le LiDAR, en les confrontant aux données issues du système SCADA. Bien que ces modèles soient relativement éloignés du cas d'usage d'une EPE, car construits sur des données post-construction, ils présentent l'intérêt de démontrer l'efficacité de méthodes statistiques fondées exclusivement sur des données météorologiques.

Dans son rapport [6], Léna Lacomme commente la figure 4.8 en indiquant que son choix de modèle aurait pu se porter sur **RF + ATM + OP** (Random Forest avec pour prédicteurs : WS, TI, *shear*, direction du vent, température extérieure, angle de calage des pâles, vitesse de rotation du rotor, et angle d'orientation de la nacelle). Toutefois, la connaissance des trois derniers paramètres opérationnels permet, via les lois de la physique, de déduire directement la puissance. Afin d'éviter d'utiliser des variables correspondant à la réponse dans les prédicteurs, mais aussi parce que les variables opérationnelles sont indisponibles dans les données de pré-construction, l'étude choisit de retenir le modèle **RF + ATM**. Cela met ainsi en évidence à la fois la supériorité des forêts aléatoires sur d'autres approches statistiques et la pertinence d'un modèle reposant uniquement sur des variables atmosphériques.

Nous avons donc décidé de faire quelque chose de semblable en utilisant les RF avec

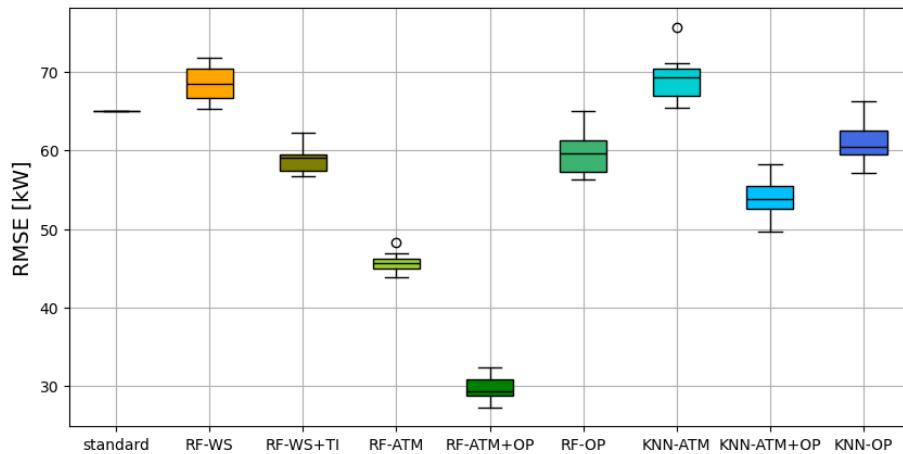


Figure 4.8: Diagrammes en boîtes obtenus à partir 10 valeurs de RMSE en kW obtenues après la validation croisée effectuée sur chaque modèle représenté en abscisse, ainsi que la valeur du RMSE du modèle standard pour la prédiction de la puissance.

une validation croisée répétée 10 fois sur un 5-fold et en prenant comme prédicteurs les variables atmosphériques suivantes :

- la vitesse de vent normalisée à hauteur de nacelle ;
- l'intensité de turbulence ;
- le cisaillement ;
- la déviation verticale ;

4.2.2.4 Modèle de prédiction basé sur les données issues de la MCP

Il s'agit d'un modèle très proche de celui présenté en section 4.2.2.3, dans la mesure où l'objectif reste de prédire la puissance P à partir des données météorologiques. Toutefois ici, il ne s'agira pas de données directement mesurées par un dispositif de mesure du vent, mais de données modélisées par la méthodologie MCP.

Si des données LiDAR sont utilisées pour la phase de "corrélation" (Fig. 4.2) dans la MCP, nous considérerons :

- la vitesse de vent normalisée par la densité de l'air ;
- l'intensité de turbulence ;
- le cisaillement ;
- le *veer*.

Le *veer* n'étant pas présent dans les données de mât, nous le remplacerons par la WA pour la réalisation de la corrélation lors de la MCP.

Que les données proviennent de LiDAR ou de mât, une attention particulière sera portée à ce modèle, dans la mesure où nous en attendons des résultats significatifs. En effet, il pourrait ainsi constituer un point de comparaison pertinent vis-à-vis de la courbe constructeur, voir s'y substituer, tout en étant transposable à d'autres éoliennes.

4.2.3 Cas d'entraînement : développement des modèles

Afin de capturer l'ensemble de la saisonnalité, il est pertinent d'entraîner nos modèles de courbe de puissance basés sur les données issues d'une MCP sur une année complète de données. Ceci est possible grâce à la disponibilité des données SCADA sur plusieurs années. Afin d'avoir un modèle s'ajustant de façon optimale tout en limitant le temps de calcul, nous cherchons donc à identifier l'année dont la distribution de la puissance est la plus représentative de l'ensemble des années disponibles. Pour identifier ce que nous appelons la "meilleure année", nous utilisons la divergence de Kullback-Leibler, définie comme suit :

$$D_{KL}(P|Q_t) = \int_{\mathbb{R}} p(x) \log \left(\frac{p(x)}{q_t(x)} \right) dx \quad (4.4)$$

où p et q_t sont respectivement les fonctions de densité des variables aléatoires continues P de la distribution pluriannuelle et Q_t de la distribution de l'année t . Il s'agit d'obtenir une divergence la plus faible possible en accord avec l'équivalence suivante :

$$P \stackrel{p.s}{=} Q \iff D_{KL}(P|Q) = 0 \quad (4.5)$$

Toutefois, ne disposant pas d'une formule explicite de la distribution, une approche de découpage par tranche s'impose. Nous choisissons de découper une année de puissance par tranche de 50 kW et d'estimer $D_{KL}(P|Q)$ de la façon suivante :

$$\hat{D}_{KL}(P|Q_t) = \Delta_P \sum_k \left((P_k + \varepsilon) \cdot \log \left(\frac{P_k + \varepsilon}{Q_{t,k} + \varepsilon} \right) \right) \quad (4.6)$$

où : $\Delta_P = 50$ kW, $\varepsilon = 10^{-10}$ afin d'éviter certains problèmes d'existence, et P_k , $Q_{t,k}$ sont les fréquences des tranches k pour respectivement l'ensemble des années et pour l'année t . On retrouve dans le tableau 4.5 l'année de la distribution la plus proche de la distribution complète que nous nommons la "meilleure année"

Parc	ABH	AEB	CDH	CDS	COF1	COF2	EDS	PRP
Année	2023	2021	2021	2022	2021	2024	2023	2023

Table 4.5: Année dont la distribution de la puissance est la plus proche de la distribution pluri-annuelle.

La figure 4.9 montre bien une superposition entre la distribution pluriannuelle de la puissance et celle de la "meilleure année" pour deux éoliennes.

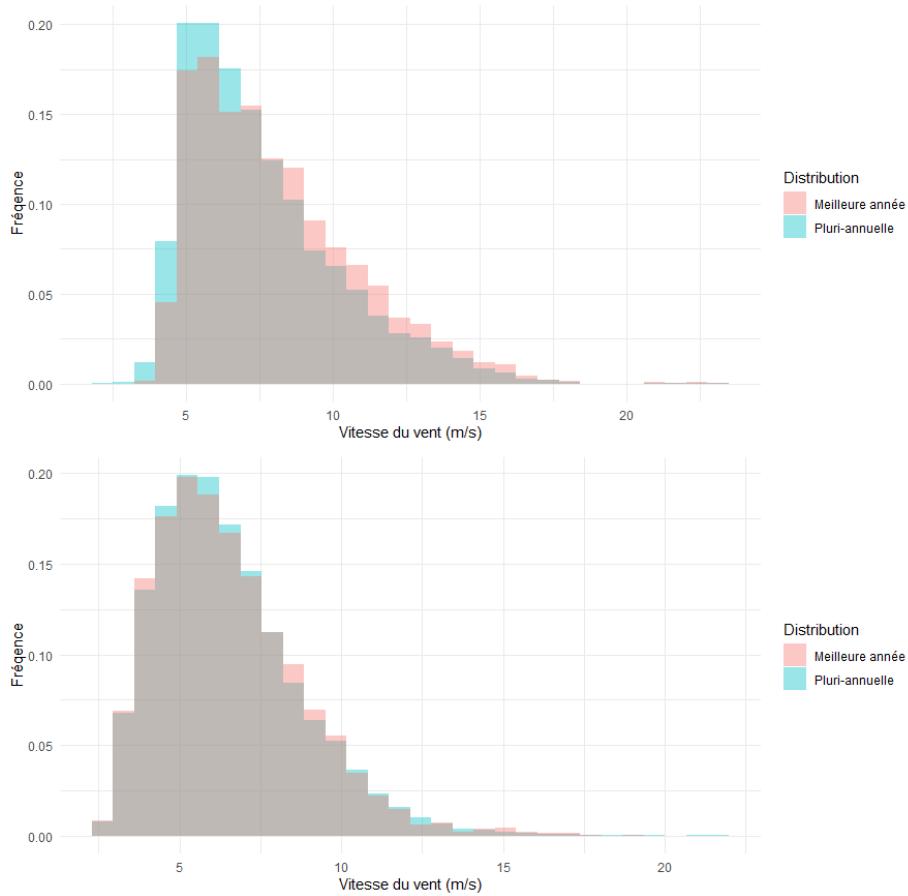


Figure 4.9: Comparaison des distributions de la puissance pluri-annuelle et celle la plus représentative sur un an des données complètes

4.2.4 Cas de test : prédition et comparaison aux données opérationnelles

Disposant de données atmosphériques de long terme par l’intermédiaire de la MCP et de modèles permettant d’obtenir une valeur de la puissance en fonction de ces données, nous pouvons alors obtenir différentes séries de puissance de long terme. L’idée est donc de les comparer aux données SCADA servant de référence et dont on trouve les statistiques dans le tableau 4.6.

Eolienne	ABH	AEB	CDH	CDS	COF1	COF2	EDS	PRP
Début	2021-12-01	2011-04-01	2017-05-17	2021-03-24	2017-02-21	2022-03-15	2021-12-08	2016-10-06
Fin	2024-12-31	2024-12-31	2024-04-25	2024-04-09	2024-12-31	2024-12-31	2024-01-01	2024-01-01
Nb données	14715	88526	26941	13211	44653	15532	10347	37407
Moyenne P [kW]	1273.77	737.94	609.79	698.27	670.77	645.61	752.59	676.12
Ecart-type P [kW]	966.05	579.09	583.83	656.27	580.77	588.93	621.91	593.51

Table 4.6: Résumé des informations concernant la puissance opérationnelle SCADA.

Sur le modèle mentionné en section 4.2.2.4, une fois notre modèle de courbe de puissance entraîné pour une éolienne donnée, nous l’appliquons aux données long-terme de cette même éolienne. Nous évaluons alors la qualité des prédictions horaires à l’aide du NRMSE normalisé (4.2) mais aussi de l’écart moyen normalisé aux données SCADA exprimé en

pourcentage.

Nous appliquons ensuite ce modèle aux données long-terme d'autres éoliennes. Une attention particulière sera portée aux cas où le modèle et les données concernent des éoliennes de même type et/ou situées dans le même parc. Les performances sont systématiquement comparées aux données SCADA de l'éolienne sur laquelle le modèle est initialement construit.

Résultats et interprétations

Dans cette section, nous exposons les résultats des différentes méthodologies et entraînements statistiques évoqués dans la section 4. Au fur et à mesure, nous faisons des choix définitifs en éliminant certains scénarios donnant des résultats médiocres. On trouve donc dans un premier temps les résultats concernant la modélisation long terme des ressources vent avant de regarder leur traduction en puissance par l’intermédiaire des différents modèles présentés en section 4.2.2.

5.1 Choix des modèles et résultats de la modélisation long-terme

5.1.1 Avec utilisation des données LiDAR

5.1.1.1 Performances issues de l’entraînement

Les figures 5.1-5.2-5.3 montrent que, dans chacun des cas : vitesse de vent, intensité de turbulence et cisaillement, le modèle des forêts aléatoires est le plus performant. Toutefois, l’écart est parfois faible avec d’autres modèles statistiques, mais surtout il est à noter que le NRMSE est parfois très élevé, signifiant des écarts horaires très grands entre la valeur réelle et celle prédite.

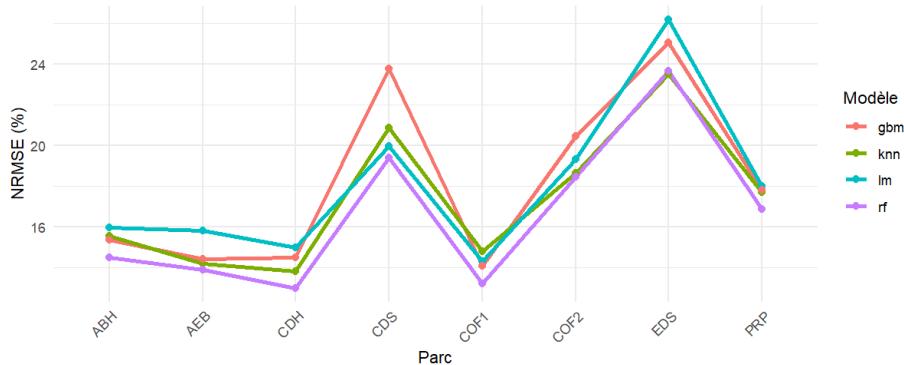


Figure 5.1: NRMSE par parc et par modèle pour la vitesse de vent normalisée.

Il apparaît ainsi que le modèle statistique basé sur les RF offre les meilleures performances en matière de prédiction horaire. Il convient alors de s’interroger sur l’existence d’un biais systématique dans ses estimations. Autrement dit, ce modèle a-t-il tendance à surestimer ou sous-estimer les valeurs prédites de manière récurrente ?

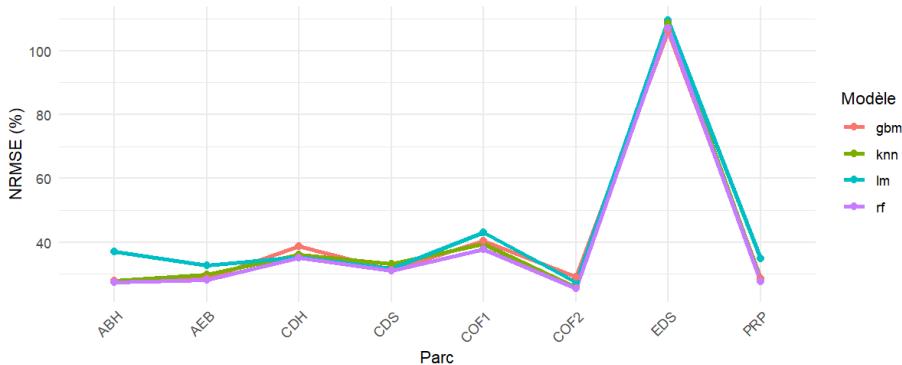


Figure 5.2: NRMSE par parc et par modèle pour l'intensité de turbulence.

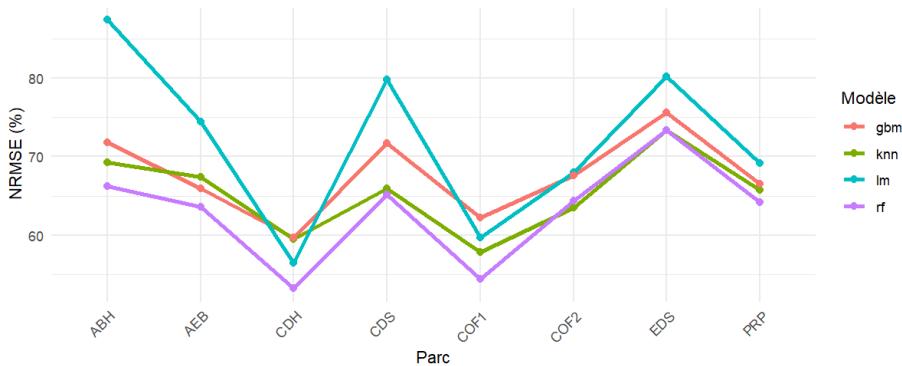


Figure 5.3: NRMSE par parc et par modèle pour le cisaiellement.

Eolienne	WS [m/s]	TI [%]	shear []	veer [$^{\circ}$ /m]
ABH	0.03	-0.26	0.01	0.00
AEB	0.12	-0.06	0.00	-0.00
CDH	0.13	1.03	-0.03	-0.01
CDS	0.33	-0.10	0.01	-0.01
COF1	0.14	-0.20	0.01	-0.00
COF2	-0.08	-0.38	0.01	-0.02
EDS	0.24	-0.79	-0.02	0.01
PRP	-0.03	0.25	-0.01	-0.01

Table 5.1: Biais pour les différents parcs pour les modèles de RF utilisés pour la MCP.

La première chose à observer sur le tableau 5.1, est la quasi-absence de biais pour le shear et *veer*. Pour les éoliennes CDH, CDS, COF1 et COF2, certains biais sont à considérer comme importants, ce qui pourrait être dommageable. Toutefois, ces valeurs sont à mettre en regard avec le faible nombre de données de validation sur lesquelles est calculé ce biais (Tab. 4.2). On note en outre que l'éolienne EDS, avec 26 horodatages servant à la validation, a un biais conséquent pour la vitesse du vent et l'intensité de turbulence.

5.1.1.2 Performances testées sur des données long-terme

Il est aussi nécessaire de comparer les prédictions avec les valeurs de référence que sont les données opérationnelles SCADA. Néanmoins, seule la WS pourra être comparée avec la vitesse opérationnelle normalisée, les autres variables n'étant pas présentes dans les données SCADA. Cette comparaison se fera à l'aide de l'écart moyen normalisé défini par :

$$\text{Ecart} = \frac{\overline{ws_{pred}} - \overline{ws_{scada}}}{\overline{ws_{scada}}} \quad (5.1)$$

et exprimé en pourcentage, l'objectif étant d'obtenir un résultat proche de 0.

La figure 5.4 montre des écarts très importants CDS, COF1 et COF2. La taille réduite de l'ensemble d'apprentissage constitue vraisemblablement le principal facteur explicatif de ce biais. L'écart reste mesuré pour les autres éoliennes.

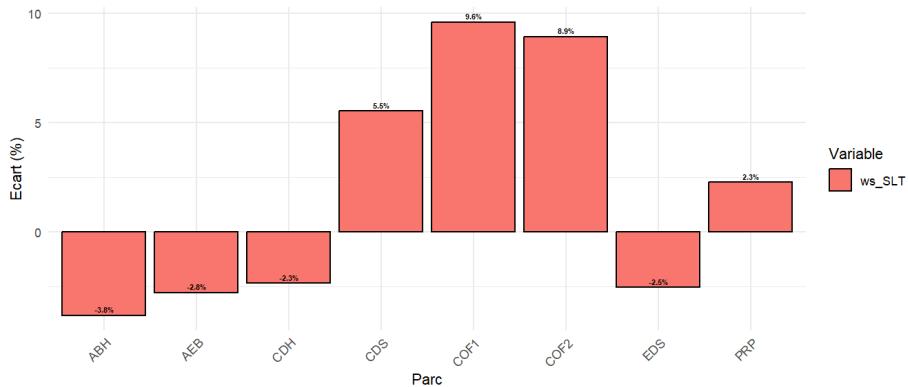
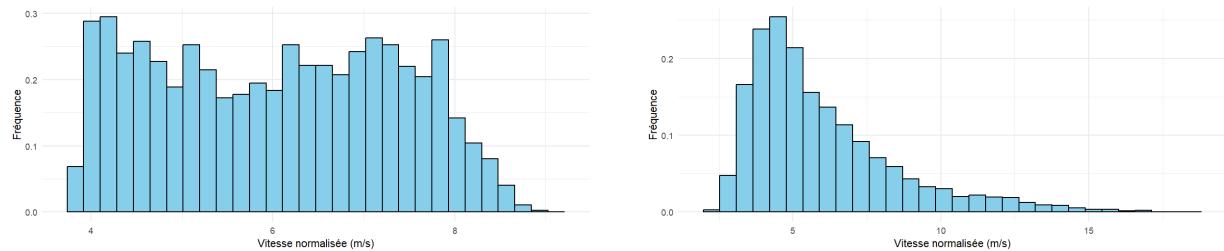


Figure 5.4: Ecart moyen normalisé de la vitesse modélisée à l'aide des données LiDAR.

En observant par exemple pour l'éolienne CDS, la modélisation long-terme de la vitesse (Fig. 5.5), on se rend compte de l'éloignement certain de la distribution observée.



(a) Distribution de la série long-terme de la vitesse normalisée.

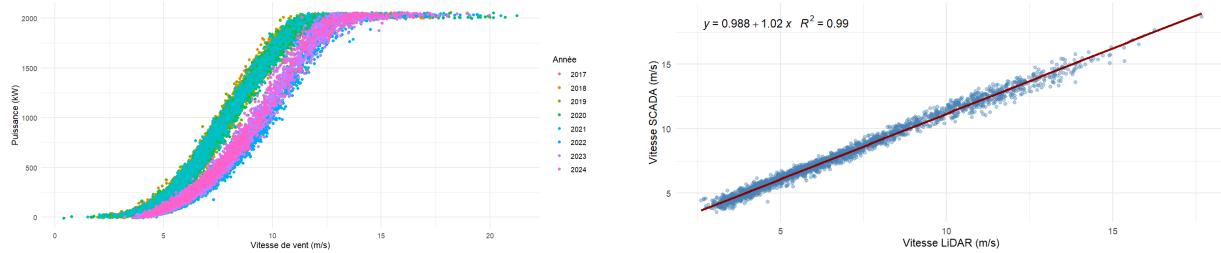
(b) Distribution observée de la vitesse normalisée.

Figure 5.5: Distributions modélisée et observée de la vitesse normalisée pour l'éolienne CDS.

Deux éléments majeurs se dégagent de la figure 5.5a. Premièrement, la distribution observée s'éloigne nettement d'une loi de Weibull, comme le montre la comparaison avec la figure 2.4. Deuxièmement — et c'est sans doute le point le plus préoccupant — les vitesses

de vent élevées ne sont pas représentées : la distribution semble plafonner autour de 9 m/s.

Les résultats pour l'éolienne CDH semblent satisfaisants avec un écart moyen faible mais trompeur. En effet, l'anémomètre de la nacelle semble avoir un problème à partir de l'année 2021 (Fig. 5.6a).



(a) Nuage de points obtenu à partir des données SCADA et coloré par année.

(b) Nuage de points et droite de régression linéaire entre la vitesse mesurée par le LiDAR et par la vitesse SCADA.

Figure 5.6: Comparaison des vitesses de vent issues des données SCADA et LiDAR.

La figure 5.6a montre un décrochage de la courbe de puissance à partir de 2022 et celle 5.6b indique, par l'observation des coefficients de la régression linéaire $y = 0.988 + 1.02x$, un décalage d'environ 1 m/s entre les vitesses mesurées par le LiDAR et celles enregistrées par l'anémomètre en nacelle. En comparant avec la moyenne de la vitesse SCADA avant 2021, on obtient un écart normalisé de 3,1 %, devant être considéré comme un écart important.

Pour les quatre éoliennes présentant un écart trop important avec la vitesse SCADA, l'utilisation de données alternatives paraît être une nécessité. Leur volume plus important et leur fiabilité permettront une correction des données de réanalyse plus représentative de la réalité, contrairement aux données LiDAR, trop peu nombreuses et en outre filtrées aux basses vitesses de vent pour éviter des incohérences. Nous ne l'appliquons pas seulement aux quatre éoliennes dont le nombre de données LiDAR fait défaut mais aussi aux autres.

5.1.2 L'utilisation des données de mâts

Après avoir entraîné nos modèles par RF avec la même procédure 5-folds répétée 10 fois, nous normalisons par la densité de l'air la série de vitesse obtenue. Cette normalisation est nécessaire étant donné que la variable vitesse modélisée n'est pas normalisée par la densité de l'air, au contraire de celle issue des données LiDAR qui l'ont été.

La comparaison avec la vitesse opérationnelle SCADA montre alors des résultats contrastés (Fig. 5.7). En comparaison des résultats obtenus avec l'utilisation des données LiDAR, il y a une véritable amélioration pour CDS et COF1 où, de plus, les écarts sont tout à fait mesurés par rapport à la réalité. Une amélioration est à noter pour COF2, mais l'écart moyen avec la vitesse SCADA reste élevé. Les résultats restent similaires pour ABH et EDS. Enfin, il y a une dégradation forte pour les éoliennes AEB, CDH et PRP.

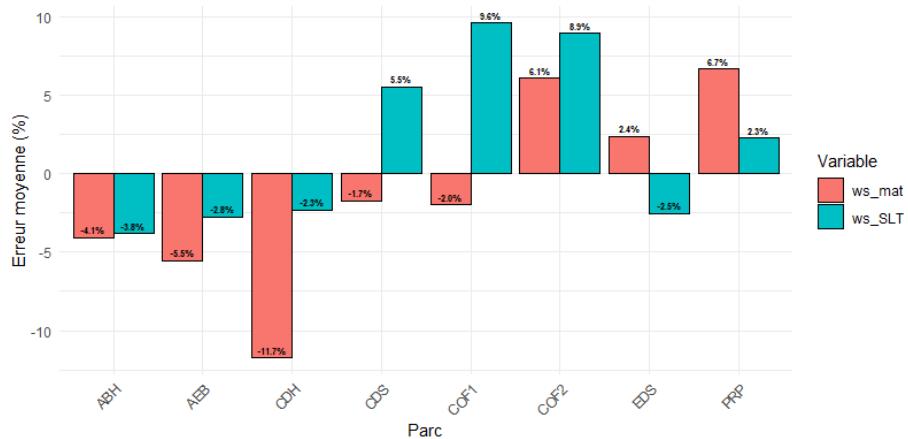


Figure 5.7: Comparaison des écarts moyens de la vitesse issue de la MCP sur données LiDAR (ws_SLT) et données de mât (ws_mat).

En comparant le NRMSE, on aboutit à un constat similaire avec des dégradations pour certaines éoliennes et des améliorations pour d'autres (Fig. 5.8).

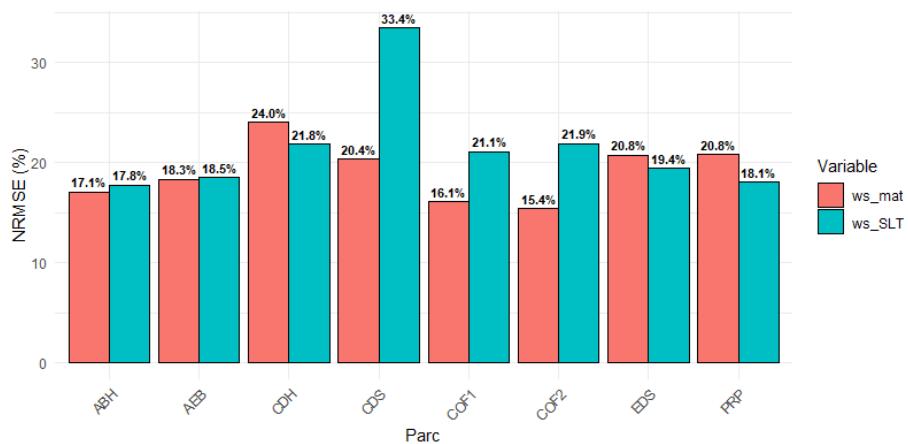


Figure 5.8: Comparaison des NRMSE de la vitesse issue de la MCP sur données LiDAR (ws_SLT) et données de mât (ws_mat).

Comme pour les données LiDAR, nous regardons le biais des prédictions (Tab. 5.2). Il est intéressant de noter que les biais sont considérablement réduits pour la WS, la TI et le shear par rapport à ceux observés lorsque la MCP avait été réalisée à l'aide des données LiDAR (Tab. 5.1). De plus, les biais sont minimaux pour les deux éoliennes ayant le plus grand nombre de données d'entraînement : COF1 et COF2.

Eolienne	WS [m/s]	TI [%]	shear []	WA [°]
ABH	0.07	0.13	0.00	-1.26
AEB	0.07	0.13	0.00	-1.26
CDH	0.11	-0.08	-0.00	0.66
CDS	0.11	-0.08	-0.00	0.66
COF1	-0.00	-0.01	NA	-0.10
COF2	-0.00	-0.01	NA	-0.10
EDS	-0.07	0.20	0.00	0.64
PRP	-0.07	0.20	0.00	0.64

Table 5.2: Biais des prédictions issues de la MCP sur données de mât.

5.1.3 Conclusion

Après avoir identifié les forêts aléatoires comme le modèle statistique offrant les meilleures performances prédictives, l'importance de disposer de données issues de mâts de mesure s'est confirmée. Toutefois, le choix entre données LiDAR et données de mât reste délicat. En effet, les performances de prédiction varient selon les cas : les données LiDAR offrent parfois de meilleurs résultats, et parfois ce sont les données de mât (voir Fig. 5.7 et 5.8).

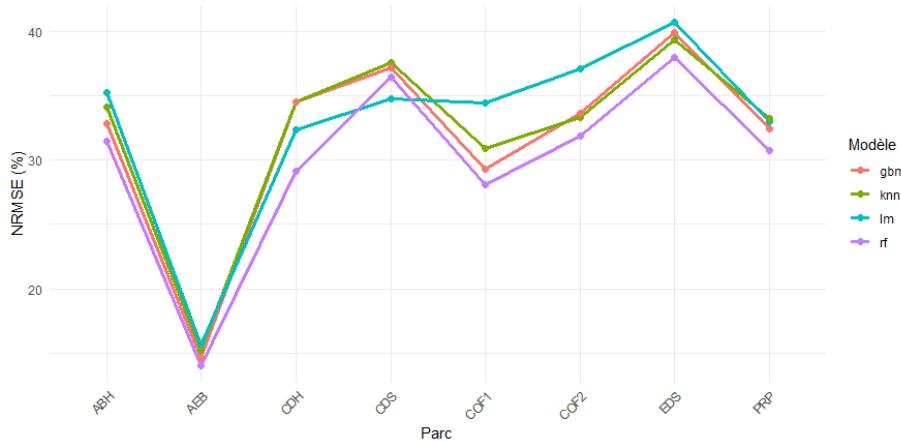
Ces variations peuvent s'expliquer par deux facteurs opposés : le faible volume mais la proximité géographique des données LiDAR d'un côté, et le volume important mais l'éloignement spatial des données de mât de l'autre. Ce compromis entre qualité spatiale et quantité de données demeure donc un point clé de la modélisation.

5.2 Évaluation du productible

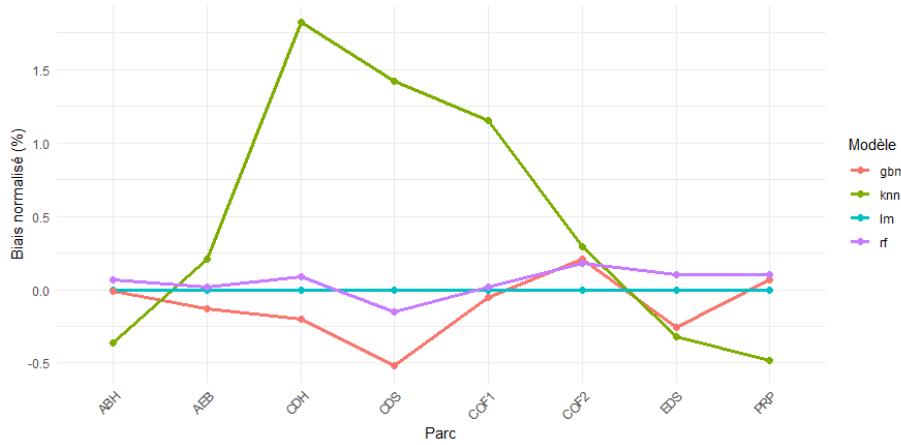
5.2.1 Choix du modèle

De nouveau, un choix s'impose quant au modèle statistique à retenir pour la prédiction de la puissance P . Pour cela, nous mettons en œuvre une procédure de validation croisée de type 5-folds, répétée 10 fois, en comparant plusieurs modèles : les modèles linéaires, la forêt aléatoire (RF), les k -plus proches voisins (k -NN) et le gradient boosting (GBM). Les données d'entrée utilisées sont celles issues de la méthode MCP, entraînée à partir des observations LiDAR.

Les critères de performance retenus pour guider le choix du modèle sont le NRMSE et le biais normalisé.



(a) Comparaison des NRMSE normalisés pour différents modèles statistiques.



(b) Comparaison des biais normalisés pour différents modèles statistiques.

Figure 5.9: Comparaison de différents modèles statistiques pour la prédiction de P .

En accord avec la figure 5.9, notre choix se porte sur les RF, qui ont de nouveau le meilleur pouvoir prédictif dans la grande majorité des cas via l'observation du NRMSE, tout en introduisant peu de biais. On remarque des valeurs sur les écarts horaires déjà importantes.

5.2.2 Écart aux valeurs mesurées des différents modèles

Dans un premier temps, nous comparons les modèles de courbe de puissance, utilisant les données de la MCP entraînée sur les données LiDAR, avec les données opérationnelles (Fig. 5.10). On retrouve donc les modèles évoqués dans la section 4.2.2 et utilisant en entrée les données issues de la MCP sur données LiDAR. On les rappelle succinctement ici :

- P_SLT_CMV : modèle entraîné, à partir des données LiDAR, sur la puissance mesurée lors de la campagne LiDAR ;
- P_SLT_Cons : modèle de référence utilisant la vitesse de vent normalisée pour prédire la puissance à partir de la courbe de puissance constructeur ;

- P_SLT_SLT : modèle entraîné sur les données de puissance de la campagne LiDAR à partir des données issues de la MCP sur données LiDAR ;
- P_SLT_TI : modèle physique de courbe de puissance avec normalisation de la turbulence utilisant la vitesse de vent normalisée et l'intensité de turbulence pour prédire la puissance (Annexe A).

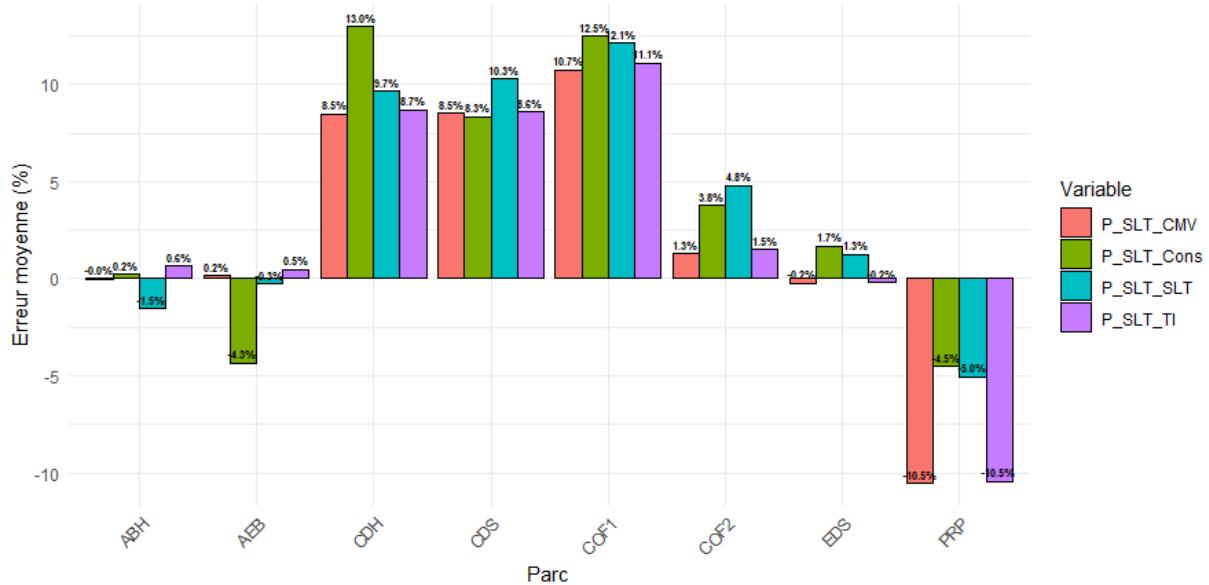


Figure 5.10: Comparaison de l'écart moyen normalisé aux mesures opérationnelles SCADA de la puissance P prédite à partir des données provenant de la MCP avec entraînement sur les données LiDAR.

Premièrement, les résultats sont assez bons pour les éoliennes AEB, ABH et EDS, notamment le modèle de courbe de puissance utilisant les données issues de la MCP et entraîné sur les données LiDAR : P_SLT_SLT. Pour ces trois éoliennes, le nombre de données était relativement important (Tab. 4.1). Toutefois, un nombre important de données ne suffit pas, car pour PRP, les résultats s'éloignent énormément des données opérationnelles. Enfin, pour les quatre éoliennes au nombre de données faible (CDH, CDS, COF1 et COF2), les pourcentages obtenus s'écartent énormément de la réalité. Ceci est en accord avec la figure 5.4, où les vitesses modélisées pour ces mêmes éoliennes étaient loin des vitesses réellement mesurées.

Il convient toutefois de souligner que, dans le cas du modèle de courbe de puissance P_SLT_SLT, les données d'entraînement proviennent exclusivement des mesures de puissance de la campagne LiDAR moyennées sur une heure, dont le volume reste relativement limité (Tab. 4.1).

Une alternative intéressante, présentée dans la section 4.2.3, consiste à entraîner ce modèle sur la "meilleure année". Cette approche s'avère particulièrement prometteuse : les figures 5.11a et 5.11b en témoignent, montrant des améliorations spectaculaires. On observe une diminution significative des écarts normalisés ainsi que du NRMSE sur l'ensemble des parcs étudiés.

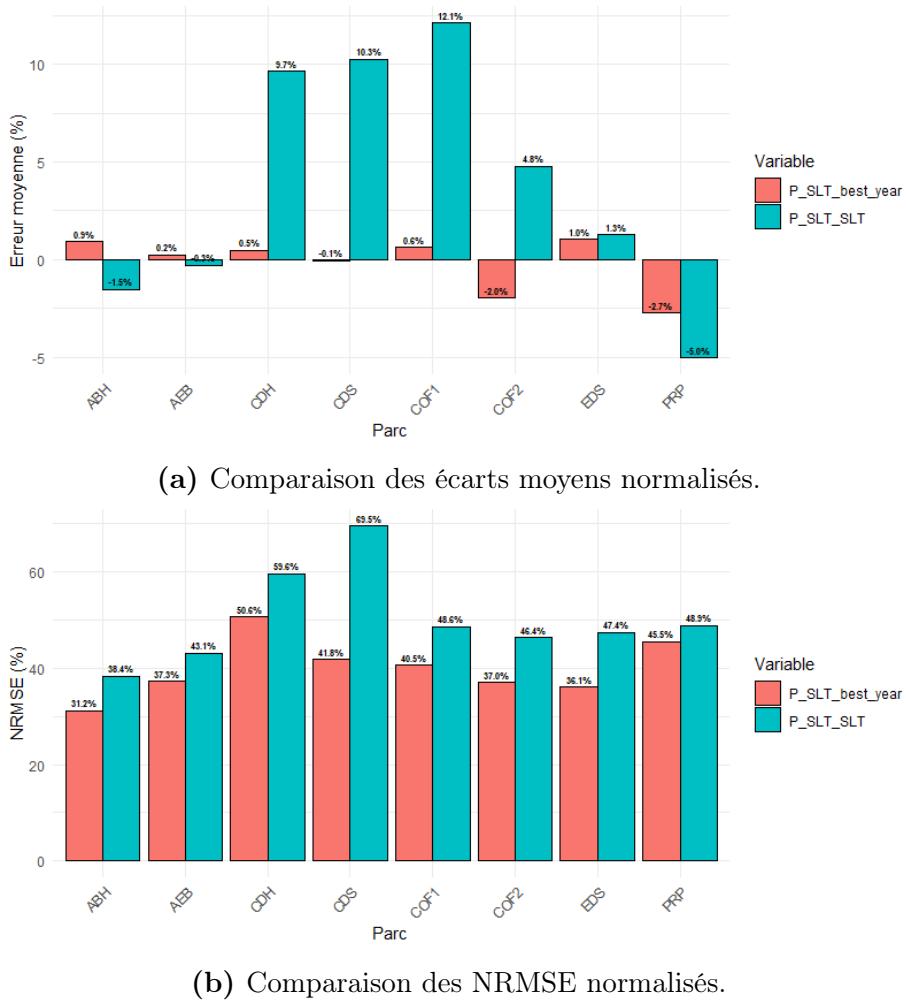


Figure 5.11: Comparaison des modèles de courbe de puissance entraînés sur données de la campagne LiDAR (P_SLT_SLT) et sur l’année la plus représentative (P_SLT_best_year).

En conséquence, nous adoptons cette stratégie basée sur le choix de l’année la plus représentative de la distribution pluriannuelle de la puissance pour la suite du rapport.

La figure 5.12 ne permet toujours pas de trancher clairement entre l’utilisation des données LiDAR ou celles issues des mâts, les résultats étant globalement proches et alternativement en faveur de l’un ou de l’autre. En revanche, l’analyse du NRMSE (Fig. 5.13) semble cette fois-ci indiquer un avantage en faveur des données de mât, en particulier en ce qui concerne la prédition horaire de la puissance. Ce constat est d’autant plus pertinent si l’on garde à l’esprit les performances parfois médiocres observées lors de la prédition de la vitesse du vent avec les données LiDAR, notamment pour certaines éoliennes.

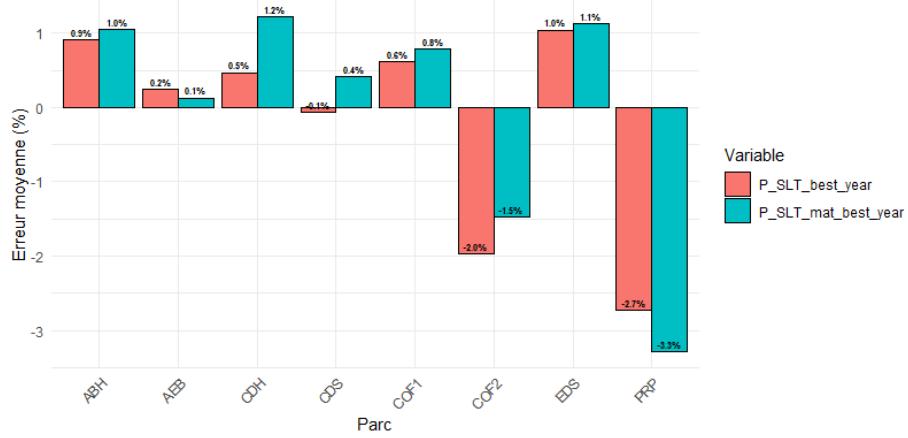


Figure 5.12: Comparaison des écarts normalisé aux données opérationnelles pour la meilleure année selon l'utilisation de données LiDAR (P_SLT_best_year) ou de mât (P_SLT_mat_best_year) lors de la MCP.

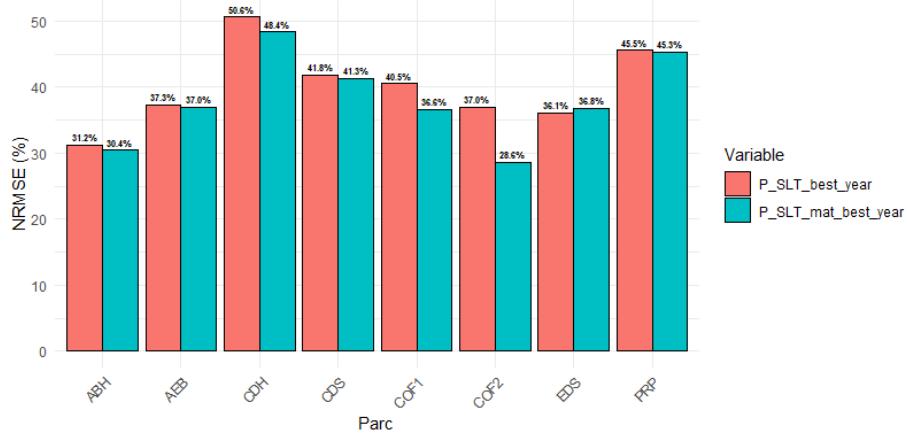


Figure 5.13: Comparaison du NRMSE pour la "meilleure" année selon l'utilisation de données LiDAR (P_SLT_best_year) ou de mât (P_SLT_mat_best_year) lors de la MCP.

5.2.3 Conclusion

Le banc d'essai des différents modèles statistiques utilisés pour la prédiction de la puissance a , une fois de plus, mis en évidence la supériorité des forêts aléatoires (RF) dans cet exercice. Forts de ce constat, nous avons jugé pertinent d'entraîner notre modèle privilégié de courbe de puissance sur l'année la plus représentative de la distribution pluriannuelle de P , conformément à la stratégie évoquée précédemment. Les résultats obtenus témoignent d'une amélioration spectaculaire des performances prédictives.

Enfin, la comparaison entre les données de mât et les données LiDAR a révélé un certain avantage en faveur des premières, notamment en ce qui concerne la précision des prédictions de vitesse, et plus encore, de puissance, comme en atteste le NRMSE.

En conséquence, la suite de notre étude se focalise sur l'utilisation de modèles de courbe de puissance entraînés sur la meilleure année, en s'appuyant sur des données de mât dans le cadre de la méthode MCP.

5.3 Comparaison des productions d'énergie annuelle estimées

Il est dans un premier temps nécessaire de créer une mesure de référence. Ceci se fait de nouveau à l'aide des données SCADA sur le long terme. Une comparaison à cette référence est ensuite faite pour les différents modèles.

5.3.1 Production annuelle de référence

Nous avons déjà vu que la vitesse du vent pouvait varier d'une année à l'autre, et par voie de conséquence, il en est de même pour la puissance. Considérer la puissance produite sur une année n'est donc pas suffisant. C'est pour cela que nous regardons les données SCADA sur toute la période de disponibilité.

Pour obtenir une première mesure de référence, nous considérons les heures, qui après nettoyage (Fig. 3.7), comptent bien six mesures et en faisons la moyenne. Puis nous estimons l'AEP par la formule suivante :

$$\text{AEP} = \frac{\sum_i \overline{P}_{i,6}}{p_{year}} \quad (5.2)$$

où $\overline{P}_{i,6}$ est la puissance moyenne pour l'heure contenant 6 mesures, et p_{year} la proportion d'années des données disponibles.

On pourrait aussi envisager de faire une moyenne sur une heure en pondérant la ou les mesures pour qu'elles comptent comme six. Nous observons sur la figure 5.14 l'influence du nombre de données minimum gardées par heure en le comparant à la valeur obtenue par la formule (5.2). On y voit, pour une même année, que la différence entre le plus petit et le plus grand écart peut dépasser 3%, ce qui est une valeur très importante.

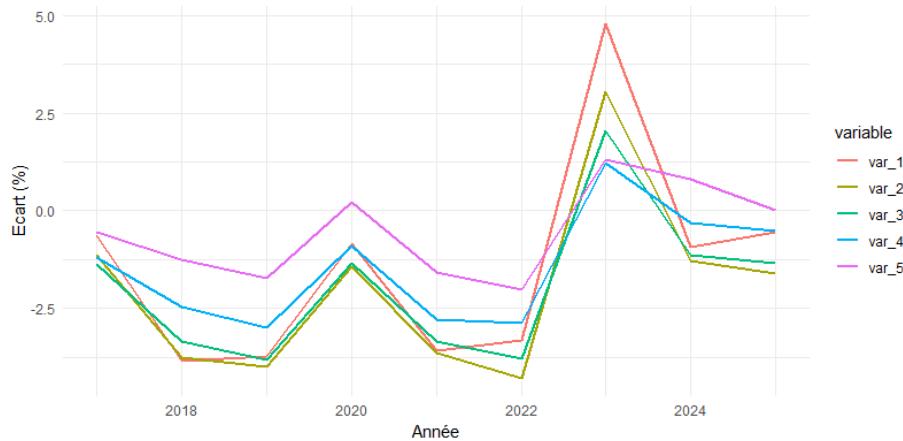


Figure 5.14: Ecart en pourcentage à l'AEP calculé en gardant les données ayant 6 mesures par heures pour l'éolienne COF1. var_{_i} désignant l'écart où l'on a gardé les horodatages contenant au moins *i* mesures par heure.

En raison de cette variabilité, il est nécessaire d'opérer un choix aussi fiable que possible. Les heures comportant moins de six mesures font l'objet d'une pondération qui, bien qu'ajustée, ne permet pas de refléter fidèlement la réalité. Par conséquent, nous choisissons de ne conserver que les heures contenant exactement six données pour l'estimation de l'AEP. Ce choix induit toutefois un biais : une part significative des données est ainsi exclue de l'analyse (Tab. 5.3), et l'estimation de l'AEP repose donc sur un sous-ensemble incomplet. Néanmoins, notre objectif n'étant pas de déterminer une valeur "vraie" de l'AEP, mais plutôt de comparer les valeurs estimées selon les différents modèles, cette simplification reste acceptable pour la suite de l'étude.

Eolienne	Avant filtrage	Après filtrage	Perte (%)
ABH	113981	88290	22.54
AEB	592094	531156	10.29
CDH	201498	161646	19.78
CDS	102304	79266	22.52
COF1	302355	267918	11.39
COF2	114495	93192	18.61
EDS	117354	62082	47.10
PRP	291302	224442	22.95

Table 5.3: Nombre de données avant et après filtrage sur les heures ayant 6 mesures et la perte en pourcentage du nombre de données.

Pour la plupart des parcs, la perte de données n'excède pas 23% hormis pour la turbine EDS qui voit son nombre de données réduit de près de 50%. Il faut garder cette valeur à l'esprit pour cette éolienne en cas d'écart important à la valeur de référence.

5.3.2 Comparaison pour les différents modèles

Du fait des choix faits lors des précédentes sections, choix des données de mâts et entraînements de la prédiction *P* sur une année complète, il ne reste que trois modèles de courbe

de puissance à comparer à la puissance mesurée SCADA :

- P_SLT_mat_cons : modèle de référence utilisant la vitesse prédicté lors de la MCP sur données de mât et la courbe constructeur ;
- P_SLT_mat_TI : modèle prédisant la puissance à partir de la vitesse du vent normalisée et de l'intensité de turbulence au travers de l'algorithme fourni en annexe A ;
- P_SLT_mat_best_year : modèle de machine learning par RF entraîné sur l'année la plus "ressemblante" (Section 4.2.3) aux données opérationnelles pluriannuelles en utilisant les variables de vitesse de vent normalisée, d'intensité de turbulence, de direction du vent et, selon les éoliennes, du cisaillement.

On s'intéresse de nouveau à l'écart moyen normalisé (Fig. 5.15) ainsi qu'au NRMSE (Fig. 5.16). Pour chacun de ces deux critères, le modèle fondé sur la courbe de puissance constructeur génère des prédictions beaucoup plus éloignées des valeurs observées que les deux autres approches. Si l'écart moyen est comparable entre les modèles alternatifs, celui basé sur l'apprentissage par RF se distingue par un NRMSE plus faible, traduisant une meilleure concordance entre les prévisions et les mesures à l'instant horaire. Enfin, on observe que malgré des écarts ponctuels significatifs sur les horodatages, les faibles valeurs globales de l'écart moyen suggèrent une certaine compensation des erreurs sur l'ensemble de la période étudiée.

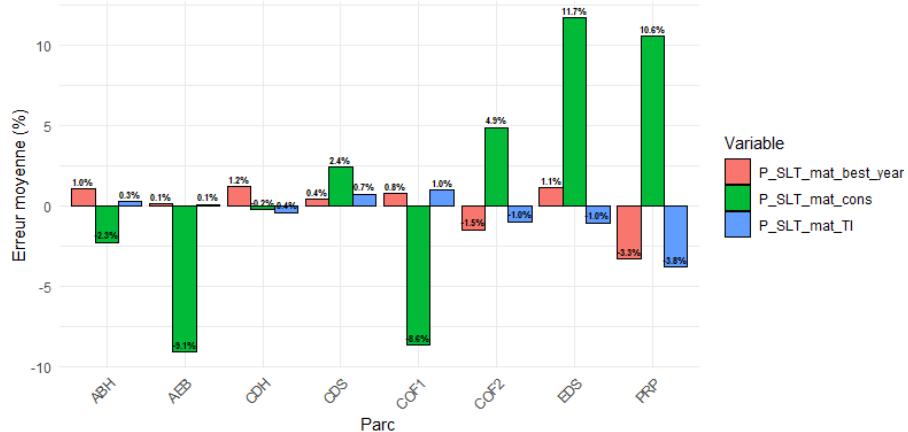


Figure 5.15: Comparaison des écarts moyens normalisés des modèles de courbe de puissance constructeur construits à partir de données issues de la MCP sur données de mât.

La figure 5.15 correspond aussi aux écarts par rapport à l'AEP de référence évoqué dans la section 5.3.1. Presque tous les écarts pour les deux meilleurs modèles restent dans une plage de $\pm 1.5\%$, ce qui reste acceptable pour un éventuel développement d'un projet éolien. Seule l'éolienne PRP, avec un écart au mieux de -3.3% . Notre modèle ici sous-estime par rapport à la puissance opérationnelle. Même si les résultats sont satisfaisants, il faut garder à l'esprit qu'il s'agit là de modèles utilisant les données d'une éolienne pour prédire la puissance de cette même éolienne.

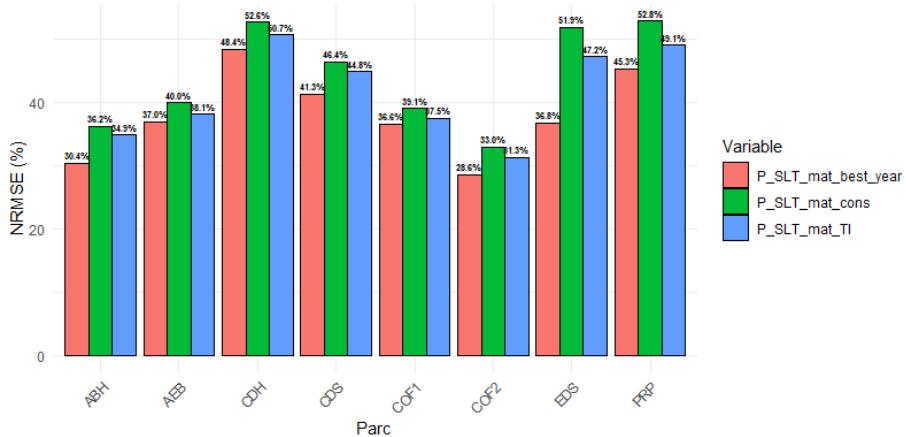


Figure 5.16: Comparaison du NRMSE de différents modèles de courbe puissance utilisant des données issues de la MCP sur données de mât.

5.3.3 Conclusion

Établir une valeur de référence n'est pas chose aisée, notamment quand un moyennage est fait avec plus ou moins de valeurs. L'étude de ce facteur montre que des écarts importants peuvent avoir lieu, et choisir la façon de calculer cet AEP de référence est une nécessité.

En choisissant au mieux notre modèle, nous réussissons à réduire l'écart moyen normalisé et ainsi l'AEP, mais aussi l'écart entre les horodatages. Pour ce dernier point, on reste vigilant car les différences observées sont encore très grandes comme la figure 5.17 l'illustre et où l'on prend garde à l'échelle de l'axe des abscisses qui ne représente pas la même vitesse pour toutes :

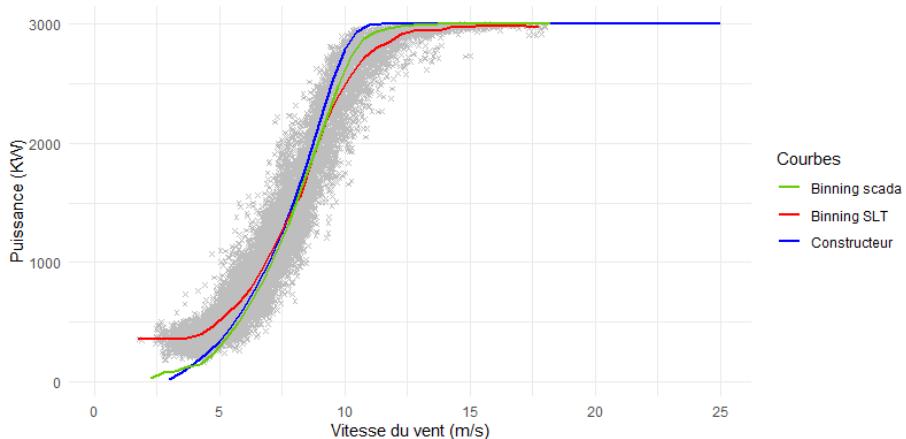


Figure 5.17: Comparaison des courbes, en bleu "constructeur", en vert (Binning scada) des données opérationnelles (v_{scada}, P_{scada}) regroupées par tranche et moyennées , en gris des données ($v_{mat_norm}, P_{mat_best_year}$) moyennées par tranche donnant la courbe rouge (Binning SLT).

5.4 Transfert de modèles

L'objectif est de déterminer si un modèle de courbe de puissance, développé à partir des données long-terme d'une éolienne, peut être généralisé en le testant sur une autre éolienne.

5.4.1 Combinaisons testées et méthodologie

Nous disposons de données provenant de huit éoliennes. Ainsi, nous pouvons tester $2 \times \binom{8}{2} = 56$ transferts mais nous regardons seulement les éoliennes de modèle identique. Il n'y a pas de sens à tenter de faire une prédiction de puissance entre deux éoliennes aux caractéristiques techniques différentes. De ce fait, les transferts de modèle se feront entre éoliennes de type MM92 (Tab 3.1), à savoir CDH, COF1, COF2 et PRP.

Pour chacune des combinaisons possibles, nous choisissons dans un premier temps le modèle de courbe de puissance entraîné à partir des données issues de la MCP sur ressources de mât de l'éolienne 1. Puis, nous considérons les données de l'éolienne 2 provenant de la MCP sur données de mâts. Cela nous donne donc 12 combinaisons possibles pour lesquelles nous regardons l'écart moyen normalisé et le NRMSE par rapport aux données opérationnelles de l'éolienne sur laquelle est fondé le modèle.

5.4.2 Résultats obtenus

On observe tout d'abord, sur les figures 5.18 et 5.19, des valeurs identiques pour certaines combinaisons. Ceci s'explique par le fait que les données long-terme issues de COF1 et COF2 sont strictement les mêmes, entraînant ainsi des résultats équivalents lorsque le même modèle leur est appliqué.

Ensuite, les transferts entre COF1 et COF2 sont, comme attendu, ceux qui donnent les meilleures performances. Les écarts moyens y sont faibles, ce qui traduit une bonne capacité de reproduction des tendances générales. En revanche, les NRMSE restent relativement élevés, ce qui reflète une dispersion notable des valeurs individuelles autour des observations réelles.

Enfin, pour les autres combinaisons de sites, les métriques obtenues sont particulièrement dégradées.

5.4.3 Conclusion

Malgré des résultats encourageants pour la prédiction de leur propre puissance (Fig. 5.15), les modèles développés sur une éolienne semblent difficilement transférables à une éolienne d'un autre parc, malgré un modèle d'éolienne identique. En revanche, cela pourrait être utile sur un même parc, avec des éoliennes identiques, notamment dans le cas d'une extension d'un parc avec des machines identiques. Il serait nécessaire de consolider les résultats en augmentant la taille de l'échantillon d'étude.

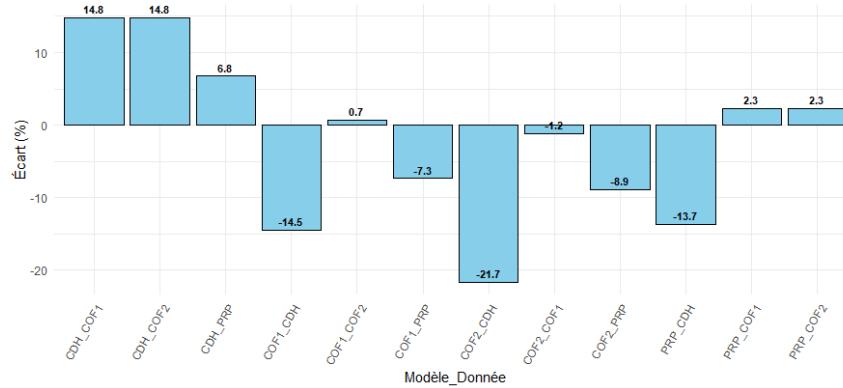


Figure 5.18: Ecart moyen normalisé pour le transfert de modèle de courbe de puissance de l'éolienne 1 avec les données de l'éolienne 2.

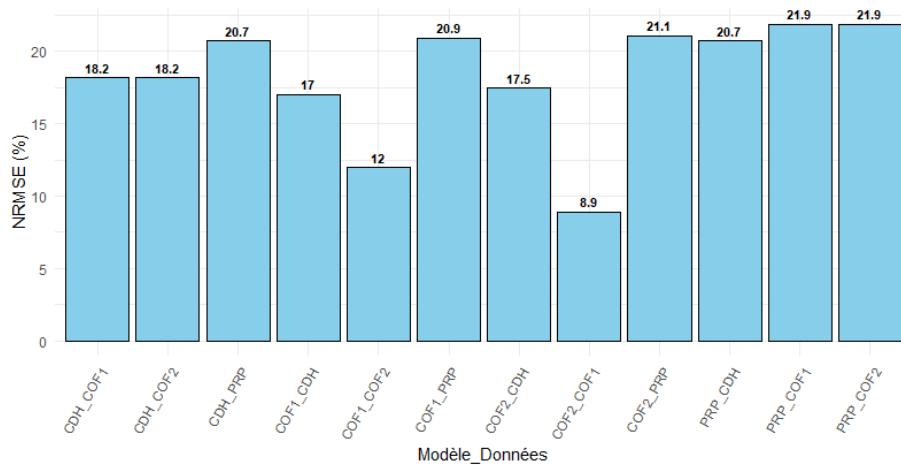


Figure 5.19: NRMSE pour le transfert de modèle de courbe de puissance de l'éolienne 1 avec les données de l'éolienne 2.

Conclusion et perspectives

6.1 Bilan du stage

Afin d'atteindre l'objectif principal de ce stage — la création de jumeaux numériques modélisant la production de puissance d'une éolienne — plusieurs étapes ont été nécessaires, mobilisant des méthodes et approches variées.

Nous avons tout d'abord entrepris la modélisation long terme de variables influant sur la production, en adoptant la méthodologie MCP. Loin de nous limiter au cadre standard (modélisation de la seule vitesse et direction du vent), nous avons souhaité évaluer la pertinence d'approches issues du Machine Learning. Les premiers essais, menés à partir de données LiDAR, ont rapidement montré leurs limites en raison d'un volume insuffisant de données. Cela nous a conduits à nous appuyer sur des mesures issues de mâts de développement, plus riches. Ce changement de source a parfois eu un impact notable sur la modélisation de la vitesse du vent, sans que cette amélioration ne soit systématique à l'ensemble des parcs étudiés.

La modélisation de la courbe de puissance par apprentissage statistique a constitué le second axe majeur du stage. L'utilisation d'une année représentative de la distribution pluriannuelle s'est révélée décisive pour améliorer les performances prédictives. Si l'écart moyen normalisé a pu être réduit sous les 1,5% dans la grande majorité des cas, le NRMSE est resté relativement élevé. Cela traduit des erreurs de prédiction horaires encore importantes, souvent compensées entre elles, mais révélatrices d'une précision limitée à l'échelle fine.

Enfin, le dernier objectif du stage — le transfert de modèle — a mis en lumière les limites actuelles de cette démarche. Ces résultats s'expliquent très probablement par les erreurs accumulées aux étapes précédentes. Néanmoins, certaines perspectives encourageantes se dégagent, notamment dans le cas de transferts entre éoliennes identiques au sein d'un même parc.

Ce stage m'a offert une immersion précieuse dans deux univers : celui de l'entreprise, avec sa dynamique collaborative tournée vers un objectif commun, et celui, passionnant, de l'énergie éolienne, riche de problématiques techniques et scientifiques. La pratique quotidienne du code m'a permis de renforcer mes compétences en programmation, tout en me faisant prendre conscience de l'importance cruciale d'une structuration rigoureuse des données en amont de toute analyse.

6.2 Perspectives d'amélioration

L'un des principaux axes d'amélioration identifiés concerne la prédiction des données long-terme via la méthode MCP. Les écarts moyens normalisés ainsi que les valeurs de NRMSE se sont révélés élevés, traduisant un écart significatif entre les prédictions et les observa-

tions réelles. Ces résultats suggèrent que l'approche actuelle atteint ses limites, et qu'il serait pertinent d'explorer des méthodes plus avancées ou adaptées, afin de mieux capturer la variabilité des phénomènes en jeu et ainsi améliorer la qualité des prédictions. On peut citer par exemple [7] où il est proposé d'introduire une méthode d'interpolation et d'extrapolation temporelle des données de vitesse du vent tout en considérant des incertitudes mixtes de la vitesse du vent.

References

- [1] IEC. International standard iec61400-12-1-2022. Technical report, AFNOR, 2022.
- [2] Branko Kosovic and al. Impact of atmospheric turbulence on performance and loads of wind turbines : Knowledge gaps and research challenges. *EAWE*, 2025.
- [3] Bardal and Saetran. Influence of turbulence intensity on wind turbine power curves. *Energy Procedia*, 137:553–558, 2017.
- [4] Yves Saint-Drenan. A parametric model for wind turbine power curves incorporating environmental conditions. *Renewable Energy*, 157(1):754–768, 2020.
- [5] Mamadou Dione. *Prévision court terme de la production éolienne par Machine learning*. PhD thesis, Institut Polytechnique de Paris, 2020.
- [6] Léna Lacomme. Rapport de stage ingénieur. Master's thesis, Centrale Nantes, 2024.
- [7] Xiao Liu and al. A new mcp method of wind speed temporal interpolation and extrapolation considering wind speed mixed uncertainty. *Energies*, 2017.

ANNEXE A

Normalisation de l'intensité de turbulence

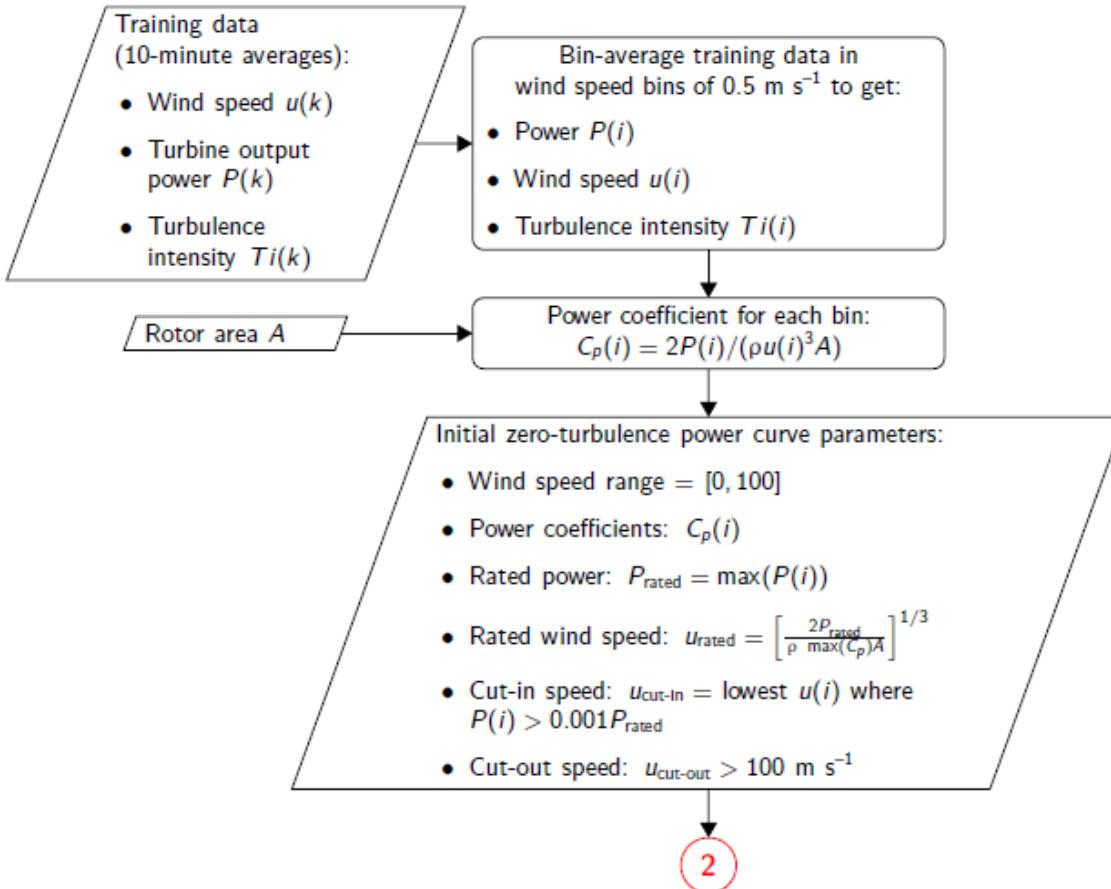


Figure A.1: Processus pour obtenir la courbe de puissance **initiale** avec zéro-turbulence.

Sur le schéma A.1, $u(k)$ désigne la vitesse moyenne sur 10 minutes normalisée par la densité de l'air, $Ti(k)$ est exprimée en pourcentage et une méthode de découpage par tranches de vitesse est appliquée aux données.

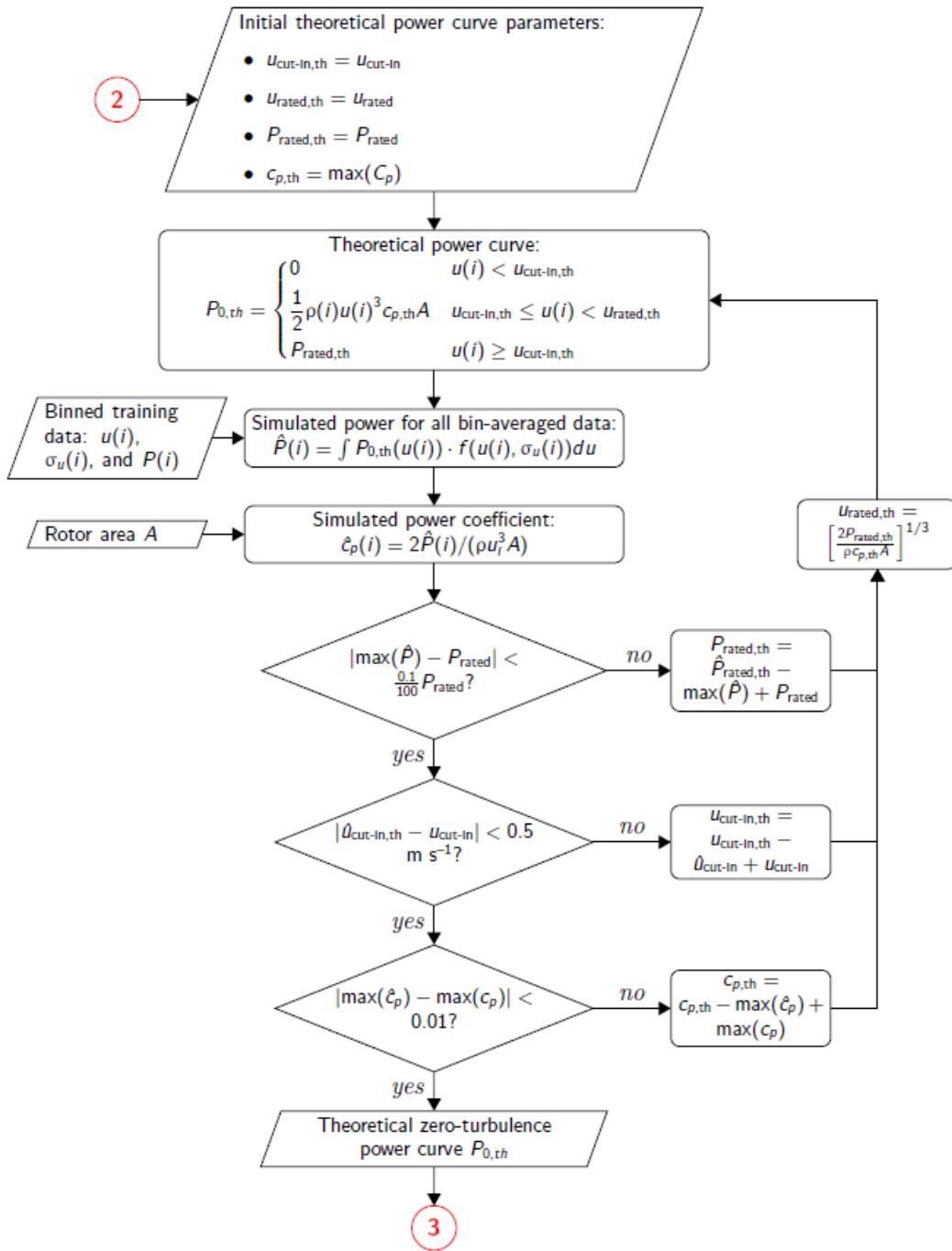


Figure A.2: Processus pour obtenir la courbe de puissance théorique avec zéro-turbulence.

On observe une mise à jour sur la figure A.2 de la vitesse nominale théorique dépendant de l'estimateur de puissance \hat{P} . Il est à noter que la méthode d'intégration choisie pour obtenir \hat{P} peut avoir une influence sur les résultats.

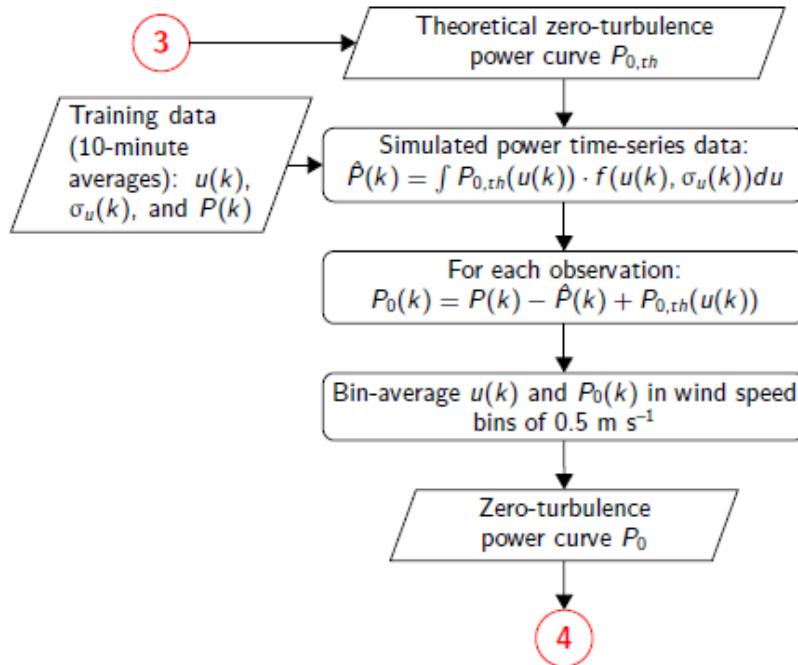


Figure A.3: Processus pour obtenir la courbe de puissance **finale** avec zéro-turbulence.

Ensuite, la courbe de puissance théorique à turbulence nulle est utilisée pour estimer la production d'énergie pour les données d'entraînement (Fig. A.3). Puisqu'il y aura toujours une différence entre la production d'énergie simulée et observée, cette valeur est ajoutée à la courbe de puissance à turbulence nulle.

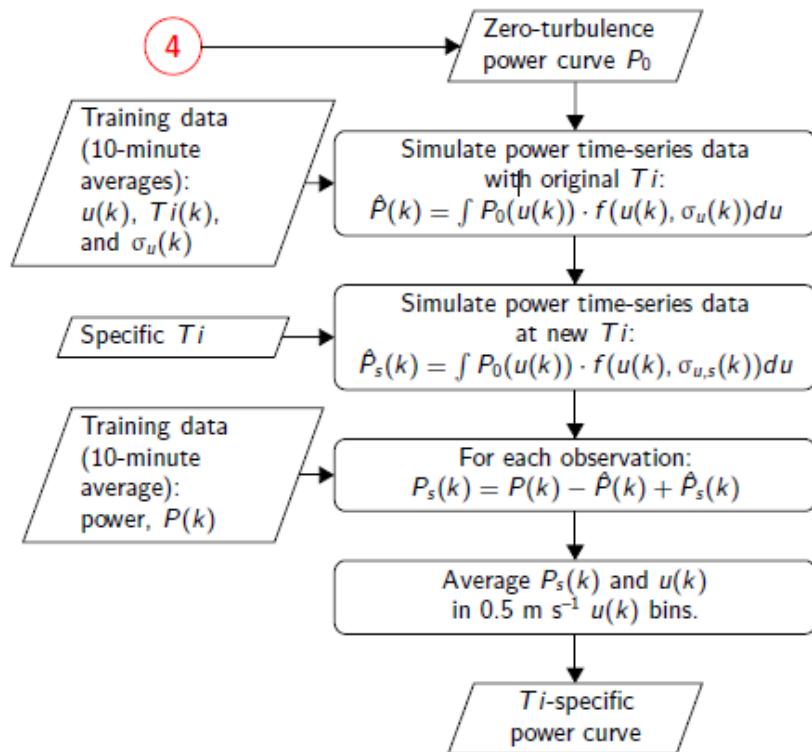


Figure A.4: Application de la courbe de puissance à turbulence nulle à un nouveau site.

Enfin, la puissance produite par l'éolienne pour une combinaison spécifique de vitesse de vent et de turbulence peut être prédite en simulant la courbe de puissance pour chaque nouvelle valeur de turbulence (A.4), puis en interrogeant cette courbe de puissance afin de déterminer la puissance produite à une vitesse de vent donnée.

ANNEXE B

Sensibilité au cisaillement comme prédicteur dans la prédiction long-terme

B.1 Avec utilisation des données LiDAR

Nous avons envisagé l'intégration du shear, initialement absent des données de réanalyse mais estimé à partir de mesures à différentes hauteurs, comme variable explicative dans le cadre de la MCP. La figure B.1 met en évidence des résultats contrastés, sans tendance claire. L'ajout du shear dans l'ensemble des prédicteurs améliore parfois les performances, mais peut également les dégrader, selon les cas. Ainsi, aucune règle générale ne se dégage quant à sa pertinence systématique, et sa contribution à la qualité de la modélisation semble fortement dépendante du site ou du contexte étudié.

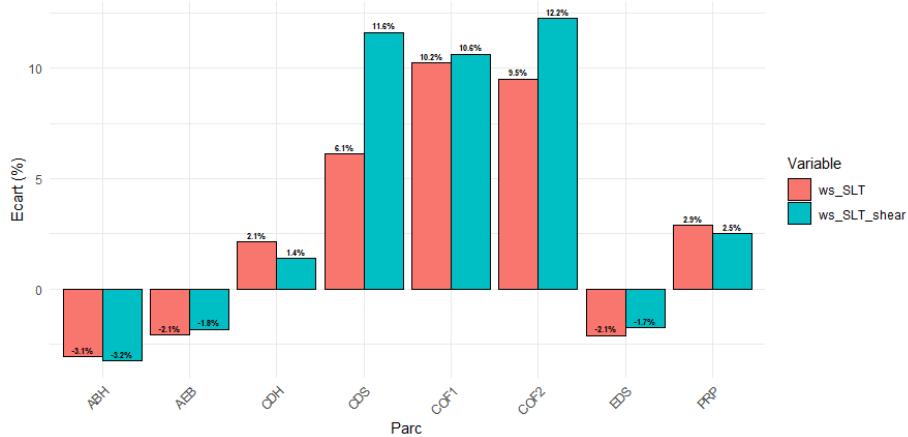


Figure B.1: Ecart moyen normalisé entre données issues de la MCP sans (*ws_SLT*) ou avec (*ws_SLT_shear*) prise en compte shear dans les données de réanalyse.

L'étude du NRMSE (Fig. B.2) n'apporte pas davantage de conclusions définitives quant à sa véritable utilité dans les prédicteurs lors de la MCP.

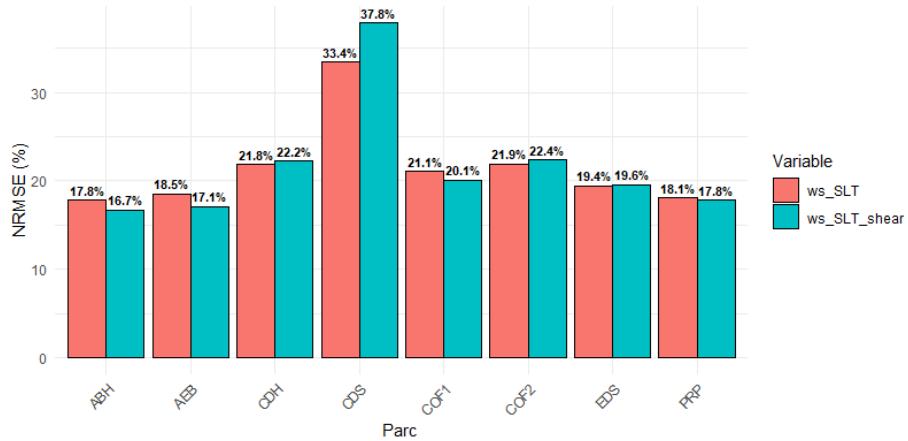


Figure B.2: NRMSE entre données issues de la MCP sans (*ws_SLT*) ou avec (*ws_SLT_shear*) prise en compte shear dans les données de réanalyse.

B.2 Avec utilisation des données de mât

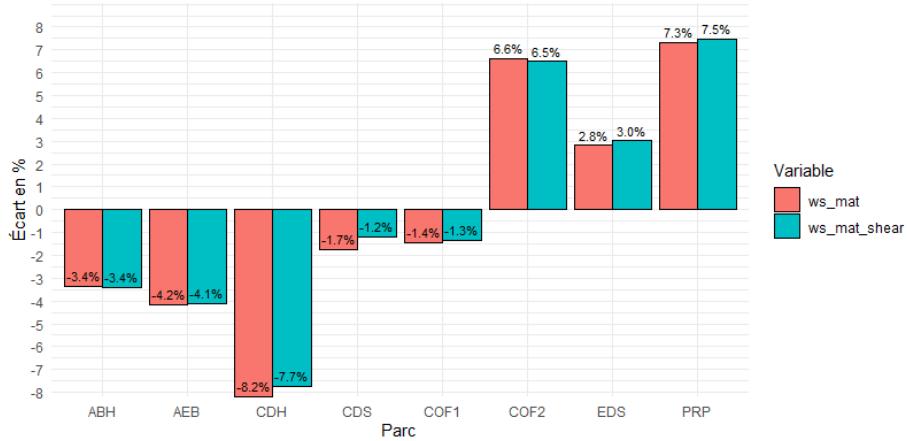


Figure B.3: Ecart moyen normalisé de la vitesse issue de la MCP avec données de mât à la moyenne de la vitesse des données SCADA. Sans prise en compte du shear : (*ws_mat*) ou avec (*ws_mat_shear*).

L'ajout du shear dans les prédicteurs de la MCP avec utilisation des données de mât, n'apporte rien de décisif, malgré le fait d'avoir un NRMSE en baisse pour tous les parcs (Fig. B.4). À des fins de parcimonie, la méthodologie MCP peut donc se faire sans cette variable.

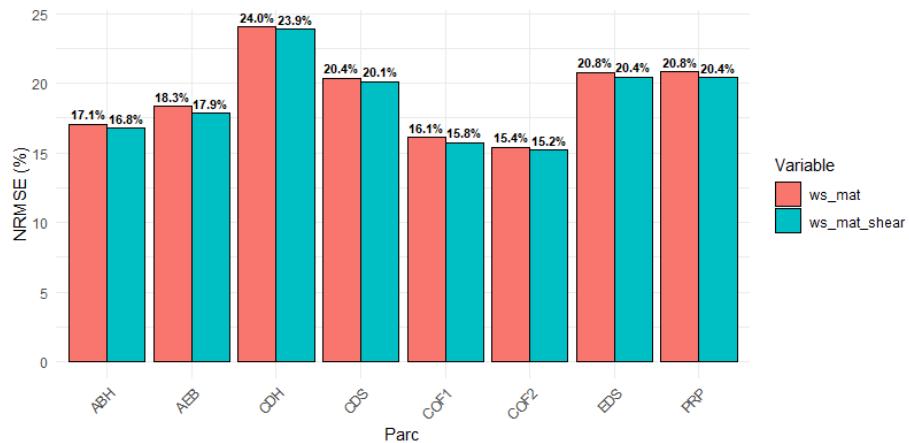


Figure B.4: NRMSE de la vitesse normalisée issue de la MCP sur données de mât.