



HAL
open science

Système de surveillance conditionnelle de turbines éoliennes déployable à l'échelle d'une flotte industrielle

Théodore Raymond

► **To cite this version:**

Théodore Raymond. Système de surveillance conditionnelle de turbines éoliennes déployable à l'échelle d'une flotte industrielle. Automatique / Robotique. Université Grenoble Alpes [2020-..], 2024. Français. NNT : 2024GRALT060 . tel-04845473

HAL Id: tel-04845473

<https://theses.hal.science/tel-04845473v1>

Submitted on 18 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)

Spécialité : Automatique - Productique

Unité de recherche : Grenoble Images Parole Signal Automatique

Système de surveillance conditionnelle de turbines éoliennes déployable à l'échelle d'une flotte industrielle

Wind turbine condition monitoring system deployable on an industrial fleet scale

Présentée par :

Théodore RAYMOND

Direction de thèse :

Sylvie CHARBONNIER
MAITRE DE CONFERENCES, Université Grenoble Alpes
Christophe BERENGUER
PROFESSEUR DES UNIVERSITES, GRENOBLE INP - UGA

Directrice de thèse

Co-directeur de thèse

Rapporteurs :

Pierre BEAUSEROY
PROFESSEUR DES UNIVERSITES, Université de Technologie de Troyes
Antoine PICOT
MAITRE DE CONFERENCES, Toulouse INP

Thèse soutenue publiquement le **18 septembre 2024**, devant le jury composé de :

Sylvie CHARBONNIER, MAITRESSE DE CONFERENCES, Grenoble INP - UGA	Directrice de thèse
Christophe BÉRENGUER, PROFESSEUR DES UNIVERSITES, Grenoble INP - UGA	Co-directeur de thèse
Pierre BEAUSEROY, PROFESSEUR DES UNIVERSITES, Université de Technologie de Troyes	Rapporteur
Antoine PICOT, MAITRE DE CONFERENCES, Toulouse INP	Rapporteur
John-Jairo MARTINEZ-MOLINA, PROFESSEUR DES UNIVERSITES, Grenoble INP - UGA	Examineur
Louise TRAVÉ MASSUYES, DIRECTRICE DE RECHERCHE, CNRS	Présidente

Invités :

Alexis Lebranchu
INGENIEUR DOCTEUR, VALEMO
Frédéric Prevost
INGENIEUR, VALEMO



Sommaire

0.1	Français	1
0.2	Anglais	2
1	Introduction	3
1.1	Crise climatique et transition énergétique	3
1.2	Pressions plurisectorielles	4
1.3	Maintenance conditionnelle	6
1.4	Présentation des travaux de recherche	10
2	Contexte et objectifs	15
2.1	Énergie éolienne	15
2.2	Organisation de la maintenance actuelle	21
2.3	Surveillance conditionnelle	25
2.4	Conclusion	31
3	État de l'art	33
3.1	Type de données utilisées	33
3.2	Génération et analyse de résidus	38
3.3	Verrous à une application industrielle	44
3.4	Conclusion	49
4	Ressources	53
4.1	Description de la flotte d'éoliennes de VALEMO	53
4.2	Échantillon de parcs éoliens	55
4.3	Base de données SCADA	59
4.4	Analyse de la base de données SCADA	63
4.5	Conclusion	69
5	Génération de modèles et détection de défauts	71
5.1	Système de surveillance conditionnelle	71
5.2	Génération de régresseurs	78
5.3	Évaluation des performances	88
5.4	Application sur quatre défauts	98
5.5	Conclusion	110
6	Approche multi-indicateur	113
6.1	Motivations	114
6.2	Génération des résidus et détection de défaut	117
6.3	Application sur quatre défauts	126
6.4	Conclusion	149

7	Généralisabilité	153
7.1	Non-reproductibilité du processus de sélection	154
7.2	Évaluation de la généralisabilité	157
7.3	Analyse des résultats	162
7.4	Conclusion	170
8	Conclusion	173
8.1	Résumé général	173
8.2	Implantation expérimentale	175
8.3	Perspectives	177
A	Description des variables SCADA de la base de données	181

Liste des Figures

1.1	Carte de la répartition des centrales d'énergie renouvelable de l'entreprise VALOREM en France métropolitaine et en France outre-mer en 2022	11
2.1	Composition d'une éolienne on-shore	16
2.2	Courbe de puissance théorique d'une éolienne	17
2.3	Mix de la production électrique française de 1980 à 2018 (DB-nomics et U.S. Energy Information Agency)	18
2.4	Projections du mix de la production électrique française horizon 2060	20
2.5	Organisation de la maintenance industrielle entre VALEMO et le constructeur	27
3.1	Schéma de disposition de capteurs sur une éolienne, et fonction associée [38]	34
3.2	Fréquence et durée d'arrêt par défaillance des composants et systèmes d'une éolienne [62]	36
3.3	Principe de surveillance conditionnelle par génération de résidus issus de modèles construits à partir de données SCADA	38
4.1	Variations de la température du roulement arrière de la génératrice durant le défaut pour les six turbines du parc E_f^{rg2}	57
4.2	Photos d'une intervention de remplacement d'une multiplicatrice d'éolienne	58
4.3	Visualisation du processus de collecte de données SCADA des capteurs installés à la base de données de l'entreprise VALEMO	60
4.4	Localisation des capteurs SCADA dans une nacelle d'éolienne [26]	61
4.5	Illustration de la disponibilité des variables SCADA du parc M_f^{al} durant une année	64
4.6	Matrice de corrélation (R^2) des variables de la base de données SCADA pour des données de bon fonctionnement du parc N_f^{rc}	65
4.7	Relation entre la température de l'huile de la multiplicatrice et la température du roulement avant de la multiplicatrice	66
4.8	Relation entre l'intensité du courant électrique produit par la génératrice et la température du système de refroidissement du convertisseur	67
4.9	Variations de $t_{roulGene2}$, t_{stator} et $p_{apparente}$ durant le défaut pour les six turbines du parc E_f^{rg2}	68

5.1	Schéma de construction de l'indicateur multi-turbine développé par Lebranchu [27]	75
5.2	Schématisation du processus complet de surveillance conditionnelle proposé	78
5.3	Coefficients de régression attribués par l'algorithme LASSO en fonction de la valeur du coefficient de pénalité pour un ensemble de variables	79
5.4	Schématisation du principe de sélection de variable GFS pour deux étapes	81
5.5	Coefficients de régression par variable pour les huit variables de sortie du parc N_f^{rc} par application de la régression pénalisée LASSO	86
5.6	Schématisation simplifiée du processus LASSO-GFS	88
5.7	Densités de probabilité de l'indicateur multi-turbine sur H_0 et H_1 pour la surveillance conditionnelle du roulement arrière de la génératrice de E_f^{rg2}	91
5.8	Courbe de caractéristique de performance, ou courbe ROC TVP(TFP) calculée à partir des données de l'indicateur multi-turbine sur H_0 et H_1 pour la surveillance conditionnelle du roulement arrière de la génératrice du parc E_f^{rg2}	92
5.9	Visualisation du calcul de ADT sur les données binaires d'un détecteur sur la période H_1	95
5.10	Variations temporelles de l'indicateur multi-turbine pour la surveillance du roulement arrière de la génératrice de E_f^{rg2} sur H_0 et H_1	103
5.11	Variations temporelles des indicateurs multi-turbine pour la surveillance du roulement principal de M_f^{al} sur H_0 et H_1	104
5.12	Variations temporelles des indicateurs multi-turbine pour la surveillance thermique de la multiplicatrice de M_f^{rm} sur H_0 et H_1	105
5.13	Variations des indicateurs multi-turbine pour la surveillance du refroidissement du convertisseur de N_f^{rc} sur H_0 et H_1	106
5.14	Courbes ROC des parcs E_f^{rg2} , M_f^{al} , M_f^{rm} , N_f^{rc}	107
6.1	Illustration de la corrélation entre une variable à modéliser ($t_{\text{convertisseur}}$) et les régresseurs générés par le processus de sélection LASSO-GFS ($t_{\text{convertisseur2}}$, t_{transfo})	115
6.2	Processus de génération des jeux de régresseurs X_1 à X_4 par application de l'algorithme GFS	119
6.3	Variations des sept indicateurs définis en section 6.2.1 pour la modélisation du roulement arrière de la génératrice de E_f^{rg2}	128
6.4	Matrices de corrélation des 7 indicateurs définis en section 6.2.1 pour la modélisation thermique du roulement arrière de la génératrice de E_f^{rg2}	129

6.5	Courbe ROC du taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP) pour les sept indicateurs définis en section 6.2.1, appliquée aux données du parc E_f^{rg2}	130
6.6	Répartition des interventions inutiles (NII) et des taux de faux positifs (TFP) par élément modélisé du parc E_f^{rg2} sur H_0 pour les approches de détection mono-indicateur et multi-indicateur	131
6.7	Visualisation des valeurs des détecteurs des trois approches de détection mono-indicateur (en bleu) et des quatre approches de détection multi-indicateur (en rouge) pour la modélisation du roulement arrière de la génératrice de E_f^{rg2} sur H_1	132
6.8	Variations des sept indicateurs définis en section 6.2.1 pour la modélisation thermique du convertisseur de N_f^{rc}	135
6.9	Matrices de corrélation des 7 indicateurs définis en section 6.2.1 pour la modélisation thermique du convertisseur de N_f^{rc}	136
6.10	Courbe ROC du taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP) pour les sept indicateurs proposés, appliquée aux données du parc N_f^{rc}	136
6.11	Répartition des interventions inutiles (NII) et des taux de faux positifs (TFP) par élément modélisé du parc N_f^{rc} sur H_0 pour les approches de détection mono-indicateur et multi-indicateur	138
6.12	Variations des sept indicateurs pour la modélisation thermique du palier de l'arbre lent de N_f^{rc} sur H_L et H_0	139
6.13	Visualisation des détecteurs des approches mono-indicateur (en bleu) et multi-indicateur (en rouge) pour la modélisation thermique du convertisseur de N_f^{rc} sur H_1	140
6.14	Schéma du processus complet de surveillance conditionnelle multi-indicateur proposé	151
7.1	Visualisation de la sélection de variables réalisée par le processus LASSO-GFS pour la modélisation du convertisseur à partir de parcs provenant de la technologie de turbine M	154
7.2	Schéma en coupe d'une éolienne de la technologie de turbine M156	
7.3	Partie hors-ligne de la stratégie d'évaluation de la généralisabilité des modèles générés par LASSO-GFS à l'échelle de la technologie de turbine	158
7.4	Partie en-ligne de la stratégie d'évaluation de la généralisabilité des modèles générés par LASSO-GFS à l'échelle de la technologie de turbine	159
7.5	Courbes ROC de la modélisation du roulement arrière de la génératrice du parc E_f^{rg2} , en utilisant les jeux de régresseurs obtenus par application du processus LASSO-GFS sur les données des parcs de la technologie E	165

7.6	Courbes ROC de la modélisation du roulement principal à partir des données de M_f^{al} , en utilisant les modèles issus des parcs de la technologie M	165
7.7	Courbes ROC de la modélisation de la multiplicatrice à partir des données de M_f^{rm} , en utilisant les modèles issus des parcs de la technologie M	167
7.8	Courbes ROC de la modélisation du convertisseur à partir des données de N_f^{rc} , en utilisant les modèles issus des parcs de la technologie N	168

List des Tableaux

2.1	Définition des besoins utilisateurs	30
2.2	Définition des contraintes des données industrielles	30
4.1	Description de l'échantillon de parcs éoliens utilisé, issu de la flotte VALEMO	55
4.2	Présentation des huit variables de sortie utilisées pour la surveillance de l'état de santé des composants	62
4.3	Disponibilité des variables de sortie associées à chaque composant pour les trois technologies de turbine présentes dans l'échantillon de parcs éoliens utilisé (Tableau 4.1)	63
5.1	Résultat de sélection de la régression LASSO pour les modélisations associées aux 4 défauts présentés dans le chapitre 4	100
5.2	Résultat de sélection de l'algorithme GFS pour les modélisations associées aux 4 défauts présentés dans le chapitre 4	101
5.3	Métriques de performances des indicateurs multi-turbine pour la modélisation des quatre composants en défaut	108
6.1	Caractéristiques des sept indicateurs considérés dans l'approche de surveillance conditionnelle multi-indicateur	122
6.2	Régresseurs générés par le processus de sélection LASSO-GFS pour l'approche multi-indicateur de la modélisation thermique du roulement arrière de la génératrice de E_f^{rg2}	127
6.3	Régresseurs générés par le processus de sélection LASSO-GFS, considérés pour l'approche multi-indicateur de la modélisation thermique du convertisseur de N_f^{rc}	134
6.4	Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation du roulement arrière de la génératrice de E_f^{rg2}	142
6.5	Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation du roulement principal de M_f^{al}	143
6.6	Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation thermique de la multiplicatrice de M_f^{rm}	144
6.7	Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation thermique du convertisseur de N_f^{rc}	144

6.8	Comparaison du nombre d'interventions inutiles à l'échelle du parc NII ^F par approche de surveillance conditionnelle, pour l'ensemble de la sous-flotte	147
7.1	Résultats de sélection du processus LASSO-GFS selon le parc utilisé, par variable de sortie modélisée	163
7.2	Bilan de performances des indicateurs multi-turbine pour la modélisation du roulement arrière de la génératrice de E_f^{rg2} .	169
7.3	Bilan de performances des indicateurs multi-turbine pour la modélisation du roulement principal de M_f^{al}	169
7.4	Bilan de performances des indicateurs multi-turbine pour pour la modélisation du système de refroidissement de la multiplicatrice de M_f^{rm}	169
7.5	Bilan de performances des indicateurs multi-turbine pour la modélisation du système de refroidissement du convertisseur de N_f^{rc}	169
A.1	Description des variables SCADA disponibles	182

0.1 Français

La surveillance conditionnelle permet l'analyse en temps réel de l'état de santé de composants afin de générer des alarmes indiquant la présence de défaut aux utilisateurs experts, permettant la mise en place d'une maintenance préventive.

En vue d'une application à une flotte d'éolienne, une méthode de surveillance conditionnelle doit fournir des informations interprétables, et pouvoir être adaptée à n'importe quel composant présentant un besoin d'optimisation des coûts de maintenance. Dans la littérature, les méthodes de surveillance étudiées en application à l'éolien sont fondées sur une analyse de résidus thermiques générés par des modèles de bon fonctionnement, appris sur des données SCADA acquises pendant des périodes sans défaut. Ces modèles prédisent l'évolution de la température des composants d'une nacelle d'éolienne dans le but de détecter des défauts à l'aide de résidus thermiques.

Alors que les études les plus récentes se focalisent sur l'optimisation de la précision des modèles en utilisant des modèles boîte noire complexes, nous faisons le choix d'utiliser des modèles simples, linéaires et statiques, composés d'un faible nombre d'entrées, afin de respecter les besoins et contraintes propres à un contexte d'application industrielle. Nous nous concentrons de résidus présentant de plus petites valeurs en période de bon fonctionnement que sous l'influence d'un défaut. Toutefois, les modèles étant simples, il est nécessaire de bien choisir les variables d'entrée.

Nous proposons ainsi une méthodologie de génération automatique de modèles linéaires et statiques par application d'un algorithme de sélection des variables d'entrée guidée par les données. Pour pouvoir déployer le système de surveillance à une flotte éolienne, cet algorithme peut être appliqué à tout élément de la flotte d'éolienne considérée, tout en conservant des performances en détection acceptables.

Les modèles utilisés étant simples à générer, nous proposons d'en générer un ensemble et de les combiner pour l'estimation d'une même variable de température, afin d'améliorer les performances en détection du système de surveillance conditionnelle proposé.

La méthode de détection de défaut proposée a été appliquée sur des données industrielles provenant d'une flotte hétérogène composée de 11 parcs éoliens provenant de 4 technologies différentes. Cette mise en oeuvre a permis de valider l'applicabilité du système de surveillance conditionnelle proposé à une échelle industrielle, en vérifiant la satisfaction des contraintes et besoins industriels par l'évaluation de métriques de performances adaptées.

0.2 Anglais

Conditional monitoring enables real-time analysis of the health of components to generate alarms indicating the presence of faults to expert users, enabling the implementation of preventive maintenance.

For application to a fleet of wind turbines, a condition monitoring method must provide interpretable information, and be adaptable to any component requiring maintenance cost optimization. In the literature, monitoring methods studied in application to wind turbines are based on an analysis of thermal residuals generated by normal behavior models, learned from SCADA data acquired during fault-free periods. These models predict the temperature evolution of wind turbine nacelle components, with the aim of detecting faults using thermal residuals.

While the most recent studies focus on optimizing model accuracy by using complex black-box models, we have chosen to use simple, linear and static models, composed of a small number of inputs, in order to meet the specific needs and constraints of an industrial application context. We settle for residuals with smaller values in good operating conditions than under the influence of a fault. However, since the models are simple, the input variables must be carefully selected.

We therefore propose a methodology for automatically generating linear and static models by applying a data-driven input variable selection algorithm. In order to be able to deploy the monitoring system to a fleet of wind turbine, this algorithm can be applied to any element of the fleet under consideration, while maintaining acceptable detection performance.

As the models used are simple to generate, we propose to generate a set of them and combine them for the estimation of the same temperature variable, in order to improve the detection performance of the proposed condition monitoring system. The proposed fault detection method has been applied to industrial data from a heterogeneous fleet of 11 wind farms from 4 different technologies. This application has enabled us to validate the application of the proposed conditional monitoring system on an industrial scale, by verifying that industrial constraints and requirements have been met through the evaluation of appropriate performance metrics.

Introduction

1.1 Crise climatique et transition énergétique

1.1.1 Modèle mondial de l'énergie fossile

Peu après la fin de la révolution industrielle, l'énergie issue de la combustion de décompositions organiques (solides pour le charbon, liquide pour le pétrole, et gazeux pour le gaz naturel), plus connue sous le nom d'énergie "fossile", s'est rapidement imposée comme la principale source d'énergie en France et dans le monde. Ce développement favorable s'explique par une très forte densité énergétique de ce vecteur énergétique (55 mega-joules par kilogramme pour le gaz naturel contre 16 mega-joules par kilogramme pour le bois), couplé à un faible coût d'exploitation et une abondance de gisements.

L'expansion de cette filière de production énergétique est corrélée à une hausse de la concentration en dioxyde de carbone (CO₂) dans l'atmosphère : aujourd'hui à 420 parties par million (ppm), sa concentration était autour de 340 en 1984. Cet accroissement, sans précédent dans l'histoire de notre planète (le dernier cas d'un ampleur comparable s'étant produit il y a 56 millions d'années sur une période "courte" de 5000 ans selon une récente étude d'un équipe de scientifiques de l'université de Genève [66]), perturbe le cycle de son climat naturel : le CO₂ est en effet un gaz à effet de serre, qui contribue au réchauffement global de l'atmosphère. Or, le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) fixe une limite de 450 ppm, au-delà de laquelle il devient difficile et peu probable de conserver un réchauffement planétaire global inférieur à 2°C.

1.1.2 Vers une transition énergétique

Au regard de l'état actuel de la crise climatique, dont les effets sont chaque année plus dévastateurs, fréquents et étendus, et en constatant que la consommation d'énergie est responsable des trois quarts des émissions de CO₂ mondiales, dont plus de 40% pour la production d'énergie électrique selon l'agence internationale de l'énergie (AIE) en 2021, il est ainsi primordial d'accélérer la transition énergétique à l'échelle mondiale.

La transition énergétique peut se définir simplement comme un remplacement des sources d'énergie principales. Pour que cette transition soit cohérente avec la résolution de la crise climatique, et puisse permettre une

adaptation à ses manifestations actuelles et futures, il est nécessaire de réduire au maximum la part de la filière fossile dans la production d'énergie, en faveur des filières dites "renouvelables" (dont les ressources nécessaires à la production d'énergie sont en quantité "infinie") et "bas carbone" (dont l'empreinte carbone¹ est faible en comparaison de la filière fossile²). Concernant la production d'énergie électrique, ces filières sont aujourd'hui principalement représentées par l'énergie solaire photovoltaïque, l'énergie hydraulique, et l'énergie éolienne.

1.1.3 Cas de la filière éolienne

La puissance installée de la filière éolienne augmente de façon constante d'année en année. Cette expansion s'accompagne d'un gain en puissance des nouvelles technologies de turbine (on peut citer l'exemple du constructeur Vestas dont les premières éoliennes terrestres assuraient une puissance nominale de 200 kilowatts en 1991, contre 8 mégawatts pour leur modèle le plus puissant, industrialisé en 2016 : la Vestas V164). Aujourd'hui, pour le cas de la filière éolienne "off-shore" (en mer), les parcs éoliens les plus récents présentent une capacité de production équivalente à la moitié de la puissance installée d'un réacteur nucléaire, avec 500MW pour la centrale de Fécamp inaugurée en Avril 2024, soit l'équivalent de la consommation en électricité de 60% du département de la Seine-Maritime.

Malgré un développement ralenti dans certains pays, à l'échelle mondiale, la filière éolienne a donc le potentiel de se présenter comme un concurrent sérieux à l'énergie électrique d'origine fossile, aux côtés du photovoltaïque et de l'hydraulique, afin d'assurer la transition énergétique. Cependant, une transition énergétique "100% renouvelable" se confronte à des pressions d'origines très diverses.

1.2 Pressions plurisectorielles de la transition énergétique

Il est difficile d'estimer le besoin futur en énergie électrique. Cependant, les trajectoires simulées par différentes expertises convergent vers une très forte augmentation du besoin par rapport à la consommation actuelle. Pour le cas de la France métropolitaine, Réseau de Transport d'Électricité (RTE) estime ainsi que la consommation totale sera de 645 térawatt-heure (TWh) à l'horizon 2050, ce qui représente une croissance de 30% par rapport à la consommation de 2023. En plus d'être ambitieux (atteindre ce chiffre signifie

¹mesure des émissions de gaz à effet de serre qui peuvent être imputées à l'activité humaine étudiée, selon la définition donnée par Wikipédia

²voir le rapport "Carbon Footprint of Electricity Generation" du Parliamentary Office of Science and Technology

une croissance de la production de plus d'un pourcent par an jusqu'en 2050), cet objectif se confronte à plusieurs obstacles.

1.2.1 Taux de retour énergétique

Tout d'abord, le taux de retour énergétique (TRE) de la filière éolienne. Le TRE est défini comme l'énergie disponible fournie par une installation, rapportée à l'énergie dépensée pour l'obtenir. Cette dernière intervient sur l'ensemble des phases du cycle de vie du vecteur énergétique considéré : extraction des matériaux, construction de l'installation énergétique, utilisation et fin de vie. Il existe plusieurs définitions du TRE. Si l'on considère la définition de Hore-Lacy³, le taux de retour énergétique des plus grandes éoliennes est autour de 20 (pour 1 unité d'énergie dépensée, une éolienne en fournit 20 au cours de son utilisation). Cependant, comme tous les moyens de production d'énergie électrique (renouvelables et fossiles), ce chiffre est amené à baisser sur les prochaines décennies, selon une étude des scénarios de transition énergétique "100% renouvelable" réalisée par Capellán en 2019 [7]. Ce constat s'explique notamment par l'empreinte carbone et le besoin en matériaux de toute filière de production d'énergie. Dans le cas de l'énergie éolienne, construire une turbine demande en effet une forte consommation d'énergie (notamment pour la constitution des fondations en béton), et donc des émissions de CO₂ significatives, ainsi que l'extraction de matériaux (acier et béton pour le mât, fibre de carbone ou de verre pour les pales, mais également des terres rares telles que le néodyme, le dysprosium et le terbium) et la fabrication de composants, qui représentent environ 65% du bilan carbone d'une éolienne⁴.

Ainsi, une expansion de la filière éolienne signifie:

- Une **augmentation des émissions de CO₂** de la filière, puisque la production de béton présente une empreinte carbone significative (7% des émissions mondiales de CO₂ selon l'Association mondiale du ciment et du béton), qui s'ajoute à la construction et au transport des pales et gros composants.
- Une **augmentation de l'extraction des matériaux** mentionnés, dont les réserves sont finies et dont l'extraction est amenée à être de moins en moins rentable selon une étude de Stéphant sur le facteur de l'industrie minière dans la transition énergétique [58].
- Une **augmentation supplémentaire du besoin en énergie électrique**, nécessaire à l'extraction et à la transformation des matériaux susmentionnés.

³<https://www.feasta.org/documents/wells/contents.html?one/horelacy.html>

⁴<https://www.hellocarbo.com/blog/calculer/bilan-carbone-eolienne-etude/>

En résumé, l'énergie éolienne sera tendanciellement moins rentable (en termes d'énergie réellement fournie, illustrée par le taux de retour énergétique) durant les prochaines décennies. On peut d'ailleurs noter que son rôle dans l'atténuation du réchauffement climatique dépendra des directions du développement de la transition énergétique : une flotte d'éolienne construite à partir de sources d'énergie principalement fossile n'aura pas la même contribution en termes de bilan d'émission de gaz à effets de serre du cycle de vie total que si elle est pouvant être construite à partir de vecteurs d'énergie 100% renouvelable.

1.2.2 Intermittence de l'énergie éolienne

Un deuxième obstacle à cette expansion est la stabilité relative du réseau électrique dans le monde. En effet, concernant l'énergie éolienne, les installations produisent de l'électricité entre 20 et 30% de l'année. Ce chiffre, qu'on appelle facteur de charge, s'explique par l'intermittence des énergies renouvelables.

Cette intermittence peut être d'origine "naturelle" : une éolienne ne produit de l'électricité que si le vent est suffisamment fort pour faire tourner les pales, mais pas trop turbulent pour ne pas endommager la turbine durant leur rotation. On peut d'ailleurs ajouter qu'un scénario considérant un "status-quo" des efforts de limitation des effets du changement climatique⁵, entraîne une augmentation de l'intermittence naturelle des éoliennes du fait de l'augmentation en fréquence et en intensité des phénomènes météorologiques extrêmes (ici, les tempêtes), d'après l'étude menée par Bonanno en 2022 [5].

Mais une part de cette intermittence peut également provenir de l'exploitation de l'éolienne : les dynamiques de l'environnement dans lequel fonctionne une éolienne (variables météorologiques) en plus de la production électrique très fluctuante car dépendant de la ressource en vent, sont autant de facteurs de dégradation de ses composants, augmentant le risque d'arrêt de la production. Ce constat est accentué dans le cas des turbines proches de leur fin de vie, qui sont plus fragilisées et demandent plus d'entretien.

1.3 Surveillance de composants en vue d'une maintenance préventive conditionnelle

Nous avons vu que l'expansion de la filière éolienne, nécessaire à la mise en place d'une transition énergétique réussie, se confronte à différents challenges. Un moyen d'y répondre est d'optimiser le contrôle de la production, c'est à dire de maximiser l'énergie fournie par une éolienne (et donc son facteur de charge) à moindre coût : de manière évidente, une éolienne qui

⁵scénario élaboré par le GIEC, nommé RCP 8.5, supposant un forçage radiatif supérieur à 8.5 watt par mètre carré en 2100

tombe moins souvent en panne produit plus d'énergie au cours de son cycle de vie. Il en résulte un facteur de charge et un taux de retour de l'énergie éolienne plus élevées.

Nous allons voir dans cette section qu'il existe un potentiel significatif d'optimisation de la production de la filière éolienne, si est mise en place une stratégie de maintenance qui permet la réduction des arrêts fréquents des turbines dus à des défaillances non anticipées et donc des pertes de production importantes.

1.3.1 Limites de la maintenance préventive actuellement en place

Afin de prévenir l'apparition ou le développement d'un défaut futur, des interventions de maintenance préventives sont mises en place. Ces interventions présentent un coût faible, et sont réalisées en périodes de fonctionnement normal de la turbine. Ces interventions sont majoritairement planifiées selon un calendrier défini à l'avance. La fréquence de ces interventions est fixée selon la criticité des systèmes diagnostiqués, et peut aller d'un mois à quelques années.

Or une éolienne est un système complexe et non stationnaire, exposé continuellement à de nombreuses perturbations internes et externes. Ces perturbations entraînent l'apparition de défauts, qui peuvent évoluer vers des défaillances, forçant l'arrêt de la turbine. L'apparition et l'évolution de ces défauts peut se faire en dehors des dates d'interventions de maintenance établies par le calendrier. Par conséquent, les défauts peuvent ne pas être détectés par une approche de maintenance préventive basée sur un calendrier, et peuvent ainsi évoluer vers des défaillances nécessitant une intervention pour remplacer le composant impacté. Ces interventions sont de loin les plus coûteuses à la filière de production d'électricité d'origine éolienne puisqu'elles impliquent : l'arrêt de la production pour la turbine défaillante (pouvant aller d'une semaine à quelques mois selon la disponibilité des pièces et composants de rechange), l'achat et le transport d'un nouveau composant (environ 100.000 euros pour une multiplicatrice), l'utilisation d'appareils de chantier onéreux et la construction d'une infrastructure d'intervention adaptée et temporaire (grues de chantier pour les plus grands composants tels que la multiplicatrice ou la génératrice).

1.3.2 Définition et enjeux de la maintenance préventive conditionnelle

Pour limiter les arrêts pour remplacement de composants, une stratégie de maintenance basée sur une surveillance de l'état de santé (ou de la condition) des composants du système industriel surveillé peut être mise en place.

Afin de comprendre la pertinence d'une telle approche de maintenance

préventive, nous pouvons comparer une éolienne qui produit de l'énergie à un être humain en vie : au cours de notre vie, nous pouvons réaliser de façon régulière des rendez-vous avec notre médecin généraliste pour faire un état des lieux de notre état de santé général afin de détecter et prévenir le développement d'éventuelles maladies (maintenance préventive fondée sur un calendrier). Cependant, nous pouvons recueillir des informations en dehors de ces rendez-vous qui nous permettent de déterminer si nous développons une maladie ou non (notre température, la composition de nos selles, notre état de fatigue...). Plutôt que d'attendre notre prochain rendez-vous chez notre médecin généraliste, nous pouvons alors avancer ce rendez-vous pour traiter au plus tôt cette maladie, sur la base d'une analyse de ces informations. Nous réalisons ainsi une maintenance préventive conditionnelle (basée sur une surveillance de notre condition à partir d'indicateurs de bonne santé) de notre corps humain. Cet effort de surveillance continue de notre condition nous permet ainsi de maximiser notre temps de vie en bonne santé, tout en réduisant notre budget alloué au traitement de nos maladies.

Il en va de même pour la maintenance préventive conditionnelle d'une éolienne. Une éolienne présente l'avantage d'être un système industriel possédant un réseau de capteurs très diversifié. Une exploitation des informations fournies par ce réseau permet de caractériser l'état de bonne santé de ses composants. Ensuite, une analyse en ligne des données issues des capteurs permet de détecter un écart par rapport à cet état de bonne santé (ou bon fonctionnement), symptomatique de potentiels défauts en développement sur ses composants. Il est ainsi possible de générer des alarmes à destination des techniciens exploitant l'éolienne, les informant de la nécessité d'avancer les prochaines interventions de maintenance préventive sur les composants présentant un défaut. Ces nouvelles dates d'interventions permettront ainsi de prendre en charge les défauts en cours afin d'éviter leur développement en défaillance, et donc d'éviter des interventions de remplacement de composant qu'auraient pu engendrer l'application d'une maintenance préventive fondée sur un calendrier.

Dans la suite, nous désignerons par surveillance conditionnelle le processus d'analyse de l'état de santé des composants aboutissant à la génération d'alarmes.

La maintenance préventive conditionnelle présente de nombreux enjeux. L'enjeu le plus évident est économique : pour un constructeur, la bonne application de cette approche de maintenance permet une réduction du nombre d'interventions de remplacement de composants au cours du cycle de vie d'un parc éolien, en adaptant le calendrier de maintenance pour pouvoir les réparer avant la défaillance. Ainsi, le budget de maintenance est réduit, et par conséquent le coût de la production d'énergie électrique d'origine éolienne aussi, ce qui se traduit par une augmentation de la rentabilité du parc éolien. Cette boucle économique vertueuse est par ailleurs favorable à

une accélération du développement de parcs éoliens, qui devient alors moins risqué financièrement pour les différents acteurs de la filière.

De même, la maintenance fondée sur la surveillance conditionnelle des composants permet d'améliorer le taux de retour énergétique. En effet, moins de remplacements de composants signifie moins d'énergie mobilisée (pour la construction de ces nouveaux composants, des infrastructures d'intervention, la logistique de transport), et donc une empreinte carbone et une pression sur l'industrie minière allégées. Ainsi, la généralisation de la maintenance fondée sur l'analyse de l'état de santé des composants est un atout essentiel dans une transition énergétique 100% renouvelable, car elle permet d'alléger les pressions économiques, environnementales et climatiques qui accompagnent la nécessaire accélération de l'expansion de la filière éolienne.

1.3.3 Freins au déploiement de la surveillance conditionnelle en milieu industriel

Au cours des dernières décennies, les recherches portant sur le développement de systèmes de surveillance conditionnelle pour la détection de défaut dans le domaine de l'énergie éolienne ont proposé des solutions enthousiasmantes. Cependant, leur application à une échelle industrielle n'est aujourd'hui pas envisageable pour plusieurs raisons.

Premièrement, leur cadre de développement est souvent réduit à un sous-ensemble de composants et d'éoliennes. La possibilité d'implémenter les méthodes développées pour surveiller d'autres composants, voire pour surveiller les mêmes composants installés sur d'autres types d'éoliennes n'est alors pas assurée. Ceci est un frein à la mise en oeuvre de ces méthodes car les parcs éoliens qui composent une flotte exploitée par un opérateur industriel d'énergie renouvelable sont souvent très différents : les éoliennes qui composent une flotte se différencient structurellement car elles sont la propriété de divers constructeurs qui proposent des "modèles" de turbine aux caractéristiques techniques propres, les parcs n'ont pas la même ancienneté (et donc le même état de dégradation des composants), ils ne sont pas situés dans les mêmes régions et ne sont donc pas soumis aux mêmes conditions climatiques, les capteurs installés ne sont pas les mêmes....

Ensuite, les approches de surveillance conditionnelle proposées dans la littérature présentent des problèmes d'interprétabilité : les indicateurs générés par les méthodes de surveillance sont difficiles à comprendre pour un utilisateur. Or, cet utilisateur sera un technicien d'exploitation, possédant de fortes connaissances sur le fonctionnement physique de l'éolienne. Il aura du mal à faire confiance à des méthodologies complexes, de type "boîte-noire" et présentant peu de cohérences physiques même si ces méthodes obtiennent de bonnes performances de détection pour un cas applicatif précis.

1.4 Présentation des travaux de recherche

La surveillance conditionnelle nécessite de disposer d'informations représentant l'état de santé d'un composant, qui sont ensuite traitées pour générer des alarmes à destination des techniciens exploitant l'éolienne. En pratique, l'état de santé d'un composant peut être analysé à l'aide d'informations issues de capteurs spécifiquement conçus pour la réalisation de cette fonction. L'inconvénient est que ces capteurs ne sont généralement pas installés sur les éoliennes, et demandent donc un surcoût d'installation et d'exploitation.

Ainsi, dans le cadre des travaux de thèse présentés, nous exploiterons les informations provenant de capteurs déjà installés sur les éoliennes. Nous utiliserons ces données pour construire un indicateur de l'état de bon fonctionnement d'un composant prenant la forme d'un résidu. Ce résidu est établi par comparaison de la valeur mesurée par un capteur avec celle prédite par un modèle appris en fonctionnement normal dont les entrées sont des valeurs provenant d'autres capteurs. Il permet de mettre en évidence un écart par rapport à une situation de bon fonctionnement et renseigne donc sur la présence éventuelle d'un défaut, sans fournir d'information sur son évolution. Cet écart est ensuite analysé afin de générer des alarmes, donnant une information binaire à l'opérateur en charge de la surveillance du système sur la présence ou l'absence de défaut.

1.4.1 Objectifs

L'originalité de nos recherches vient du fait que nous nous efforcerons de développer des méthodes permettant de générer des indicateurs de surveillance pouvant être mises en oeuvre par un industriel chargé de l'exploitation de parcs éoliens. Nous aurons à coeur de développer des solutions qui puissent s'appliquer à n'importe quel composant d'une éolienne, quelque soit la marque ou le modèle de l'éolienne, sans nécessité d'adapter la méthode. Cela permettra le déploiement de la méthode à une échelle industrielle, c'est à dire sur n'importe quel parc éolien composant la flotte de l'exploitant industriel. Nous nous attacherons à générer des indicateurs de défaut aisément compréhensibles par un expert, ce qui facilitera l'acceptation de la solution par les opérateurs de maintenance.

Cette thèse a été menée selon le dispositif des Conventions industrielles de formation par la recherche (Cifre), et s'inscrivent dans une collaboration entre l'entreprise VALEMO, réalisant l'exploitation et la maintenance d'énergies renouvelables (éolien, photovoltaïque et hydraulique), et le Gipsalab, un laboratoire spécialisé dans le traitement de signal et l'automatique. Nous allons voir dans la section suivante que ce cadre d'application des travaux de recherche est cohérent avec un objectif d'industrialisation de la méthodologie de surveillance conditionnelle développée, et permet la prise en compte des pressions plurisectorielles à l'expansion de la filière éolienne

mentionnées en section 1.2.

1.4.2 Cadre d'application

VALEMO est une entreprise française O&M, spécialisée en énergies renouvelables (EnR), fondée en 2011. Initialement spécialisée dans la filière éolienne, l'entreprise a progressivement intégré dans ses compétences l'exploitation et la maintenance des principales énergies renouvelables : l'éolien, le solaire, l'hydroélectricité et plus récemment les énergies marines. Elle représente en Mai 2024 plus d'un gigawatt (GW) de puissance installée en exploitation, et 305 mégawatts (MW) en maintenance⁶ (la puissance installée de la filière éolienne terrestre était de 21.39 GW en France en Juin 2023 selon RTE). La Figure 1.1 représente la répartition des centrales d'énergie renouvelable (éolien, solaire, hydroélectrique et stockage par hydrogène) du groupe VALOREM, dont VALEMO est une filiale, en 2022.

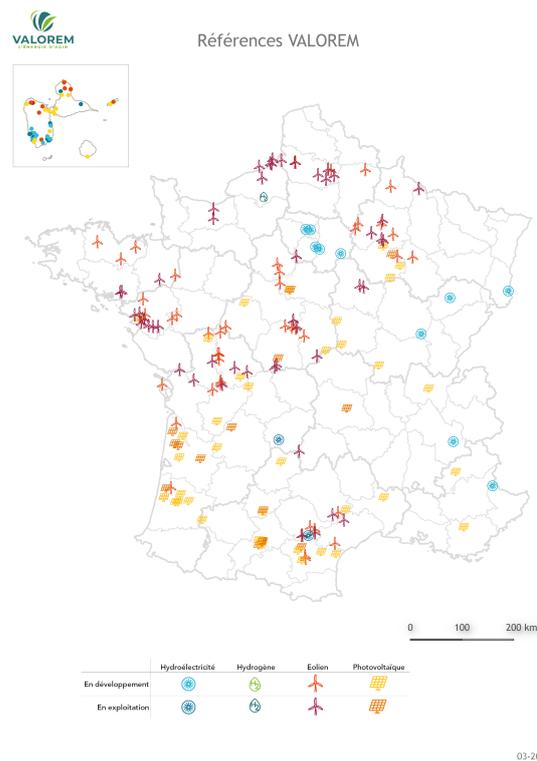


Figure 1.1: Carte de la répartition des centrales d'énergie renouvelable de l'entreprise VALOREM en France métropolitaine et en France outre-mer en 2022

⁶<https://valemo.fr/>

Depuis 2016, des recherches ont été menées chez VALEMO afin de développer une solution de surveillance conditionnelle de l'ensemble des parcs exploités par l'entreprise, dans un objectif d'optimisation de l'efficacité de son service de maintenance conditionnelle. Les travaux de thèse présentés dans ce manuscrit constituent la dernière étape de ce grand projet, avec pour ambition l'implantation de la solution développée au sein des outils d'exploitation de l'entreprise.

Le développement et l'intégration du processus de surveillance conditionnelle proposé présente ainsi des enjeux industriels significatifs, car il permettrait une réduction significative des coûts et efforts de maintenance de l'entreprise VALEMO, par une amélioration considérable de son service de maintenance préventive conditionnelle. Au travers des chiffres sur la capacité de production d'électricité d'origine éolienne de l'entreprise VALEMO sur le territoire français, nous pouvons également affirmer que l'amplitude de réduction des coûts et efforts de maintenance de la méthodologie proposée s'inscrit dans un objectif de réduction du besoin en matériau et de l'empreinte carbone de la filière éolienne en France.

1.4.3 Plan du manuscrit

Ce manuscrit est organisé de la manière suivante. Nous débuterons par le chapitre 2, qui présentera l'état actuel et prospectif de la filière éolienne en France, en prenant le temps d'exposer la composition et le fonctionnement d'une éolienne. Nous ferons un tour d'horizon de la maintenance industrielle de cette filière, et mettrons en évidence les freins au développement et au déploiement d'outils de surveillance conditionnelle. Ceci nous permettra d'identifier les besoins et contraintes de cette application industrielle.

Le chapitre 3 présentera les méthodes de surveillance conditionnelle appliquées à la filière éolienne publiées dans la littérature. Nous identifierons les obstacles à l'utilisation de ces méthodes dans un contexte industriel, en les confrontant aux besoins et contraintes inhérents à ce milieu. Nous pourrions alors exposer les contributions de nos travaux de recherche, ainsi que le cadre d'étude.

Dans le chapitre 4, l'échantillon de parcs éoliens utilisé pour le développement des méthodes proposées dans ce manuscrit sera détaillé. Nous présenterons les quatre cas de défauts qui serviront de support à l'évaluation des performances des méthodes. La base de données utilisée sera également présentée, puis analysée afin de mettre en lumière certaines contraintes inhérentes au contexte industriel.

Dans le chapitre 5, nous présenterons une méthode guidée par les données pour construire de manière automatique un modèle de bon fonctionnement pour tout composant d'une éolienne d'une flotte, dans l'objectif de générer un résidu facilement interprétable pour un opérateur. Nous proposerons aussi un cahier des charges précisant les critères de performances en détection

applicables à un système de surveillance déployable à l'échelle industrielle. Nous analyserons les performances en détection de notre système de surveillance sur la base des quatre défauts présentés au chapitre 4.

L'évaluation des performances en détection ayant mis en évidence certaines limites de la méthode de surveillance conditionnelle fondée sur un seul modèle proposée au chapitre 5, nous présenterons dans le chapitre 6 une approche de surveillance conditionnelle fondée sur la génération automatique de plusieurs modèles de bon fonctionnement pour un même composant. Nous proposerons alors différentes solutions pour fusionner les multiples résidus en découlant.

Un des freins au déploiement industriel des méthodes de surveillance conditionnelle est leur acceptation par les opérateurs de maintenance qui ont besoin de comprendre les décisions prises par le système. Le système ne doit pas alourdir la charge cognitive des opérateurs et donc fournir des informations facilement interprétables. Ainsi, dans le chapitre 7, nous vérifierons que, pour une technologie d'éolienne, et un ensemble de variables d'entrée donnés, les performances de détection du résidu associé ne dépendent pas du parc étudié. Cette propriété de "généralisabilité" du modèle garantit de pouvoir utiliser un même modèle pour tous les parcs de même technologie et donc facilite l'interprétation des alarmes générées par les opérateurs.

Nous finirons par une conclusion générale qui résume le processus complet de surveillance conditionnelle proposé. Nous exposerons alors des recommandations pour les futurs utilisateurs, ainsi que différents champs de recherche prospectifs pour améliorer les performances du système proposé, et pérenniser son intégration au sein d'une entreprise exploitant des parcs éoliens.

Contexte et objectifs

Nous commencerons ce chapitre par une présentation de l'énergie éolienne. Nous expliciterons le principe de fonctionnement d'une éolienne, puis nous reviendrons sur la place de la filière éolienne dans le mix énergétique. Nous définirons notamment les leviers qui ont rendu possible la croissance à l'échelle mondiale de cette filière, avant de mettre en lumière les enjeux de cette croissance dans le cas particulier de la France métropolitaine.

Puis nous présenterons les différentes catégories de maintenance industrielle. Nous expliquerons alors pourquoi la maintenance préventive conditionnelle est l'approche la plus adaptée à un objectif d'optimisation de la production énergétique.

Cependant, nous verrons ensuite que la surveillance conditionnelle, nécessaire à la mise en place d'une maintenance préventive conditionnelle, n'est aujourd'hui que très peu implantée au sein des services de maintenances des entreprises O&M de la filière éolienne. Avec l'exemple de VALEMO, nous verrons que ce constat s'explique par la politique de maintenance en place chez les opérateurs industriels. Enfin, nous identifierons les différents besoins et contraintes inhérents au contexte industriel, à intégrer au développement du système de surveillance conditionnelle proposé dans ces travaux de recherche, afin de rendre cette politique plus favorable au développement de la maintenance préventive conditionnelle.

2.1 Énergie éolienne

2.1.1 Description et fonctionnement d'une éolienne

Composition d'une éolienne

L'aérogénérateur, ou éolienne, a pour fonction la transformation de l'énergie cinétique du vent en mécanique puis en énergie électrique. Dans ce manuscrit, on s'intéressera à l'éolienne à axe horizontal, ou HAWT (horizontal axis wind turbine), à 3 pales, représentant la grande majorité des installations en France. Ce modèle d'éolienne permet d'injecter de l'énergie électrique dans le réseau à l'aide des composants identifiés dans la Figure 2.1, selon le fonctionnement suivant:

1. Les **pales** (A), rattachées au **rotor** (K), entrent en rotation à partir d'une vitesse de vent minimum. Le système **yaw** (G), contrôlant

la rotation du mât de l'éolienne, permet de positionner l'axe du rotor face au vent. La prise au vent est optimisée par l'utilisation du système **pitch**, contrôlant la rotation des pales sur elles-mêmes, en complément.

2. Au sein de la **nacelle** (E), le **rotor** (K), alors en rotation, entraîne l'**arbre principal** (B), également appelé **arbre lent**, tournant à des vitesses faibles de 10 à 20 tours par minute en moyenne.
3. Une boîte à engrenages, ou **multiplicatrice** (C), est rattachée à l'arbre principal pour adapter la vitesse de rotation (multiplication de la vitesse par 100 environ) à la **génératrice** (F) pour la production d'énergie électrique. Pour le cas des éoliennes à entraînement direct, la multiplicatrice n'est pas présente car la génératrice installée est adaptée aux faibles vitesses de rotation.
4. La production d'électricité est ensuite assurée en triphasée par l'utilisation d'un **convertisseur** pour passer d'un courant continu à un courant alternatif, puis d'un **transformateur**, élevant la tension du signal pour faciliter le transport de l'énergie électrique dans les lignes moyenne tension du réseau public. Ces deux composants sont généralement localisés dans une armoire de couplage au réseau électrique (I), positionnée au pied du mât de l'éolienne (H) .

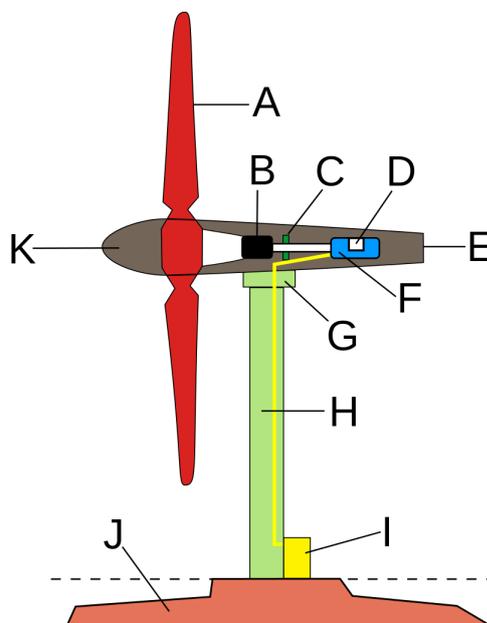


Figure 2.1: Composition d'une éolienne on-shore

Modes de fonctionnement et courbe de puissance

La production d'électricité d'une éolienne se décompose en 4 modes de fonctionnement, représentés par des chiffres romains dans la courbe de puissance en Figure 2.2. Cette courbe représente la relation entre la vitesse du vent et la puissance produite par la turbine. Elle est propre à chaque type d'éolienne, et est utilisée par les opérateurs industriels pour caractériser les performances d'une éolienne.

Pour produire de l'énergie, une éolienne a besoin d'une force de vent minimale, appelée "cut-in". Cette valeur de démarrage varie entre 3 à 4 mètre par seconde (m/s) selon le modèle de turbine considéré. La turbine rentre alors dans un mode de fonctionnement transitoire, où la relation entre la vitesse du vent et la puissance produite est non-linéaire, jusqu'à atteindre une vitesse de vent nominale, à partir de laquelle la production atteint son maximal (puissance nominale de la turbine). Au-delà d'une certaine valeur de vitesse de vent "cut-out", l'éolienne est arrêtée par sûreté car non dimensionnée pour des vents plus forts.

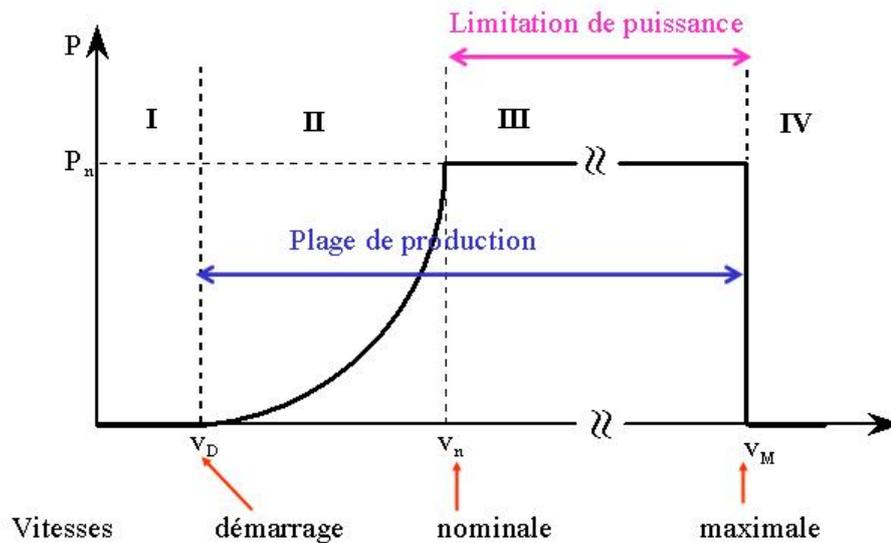


Figure 2.2: Courbe de puissance théorique d'une éolienne

2.1.2 L'énergie éolienne dans la transition énergétique

Présentation du mix énergétique Français

La France métropolitaine présente une diversité de biomes propice à une transition énergétique électrique réussie, réunissant les trois énergies dont les sources sont renouvelables : l'éolien (11% du mix total en France en 2023), le photovoltaïque (7,6% du mix total en France en 2023), et l'hydraulique (18%

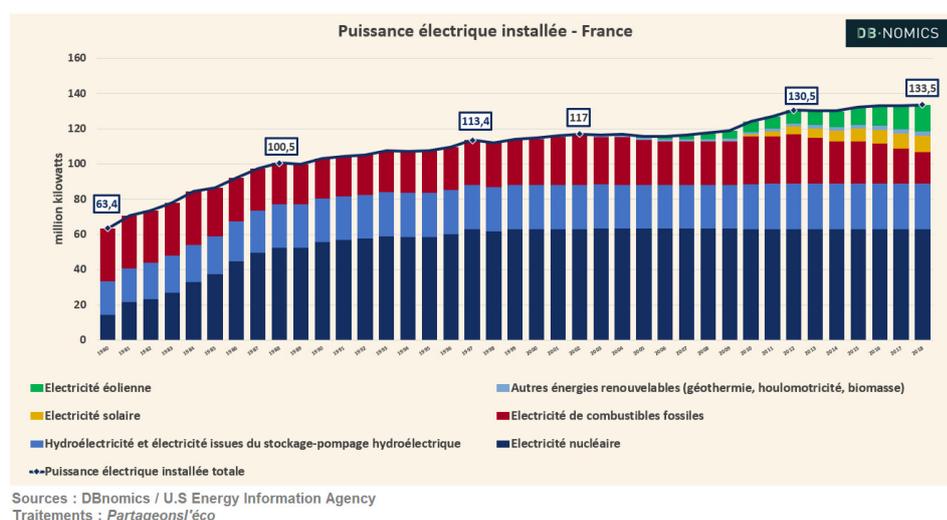


Figure 2.3: Mix de la production électrique française de 1980 à 2018 (DB-nomics et U.S. Energy Information Agency)

du mix total en France en 2022). Cette diversité lui permet de produire aujourd'hui plus d'un quart de son électricité à partir de ces trois filières renouvelables¹, le reste étant majoritairement assuré par l'énergie nucléaire (70%). Du point de vue de la puissance installée totale en énergies renouvelables, la filière éolienne se place ainsi en deuxième position derrière l'énergie hydraulique, encore aujourd'hui principale source d'énergie renouvelable, mais amenée à être détrônée à cause de son faible taux de développement en comparaison des autres énergies renouvelables.

Comme on l'observe sur la Figure 2.3, présentant l'évolution du mix de la production électrique française de 1980 à 2018, établie par l'agence de l'information sur l'énergie à partir de la base de données DBnomics, la filière éolienne terrestre est relativement récente, et sa puissance installée est en continuelle augmentation. Elle a presque triplée durant la décennie 2012-2022, pour atteindre plus de 20 GW en 2022, d'après le panorama produit par le gestionnaire Réseau de transport de l'électricité (RTE) pour l'année 2022.

Leviers de croissance de la filière

À l'échelle mondiale, cette forte croissance de la filière est due à plusieurs facteurs. Le premier facteur est lié au progrès technique et à l'augmentation de la dimension des éoliennes comme la hauteur du mât et la longueur des pâles, proportionnelle à la puissance produite d'après l'équation 2.1, formulant la production de l'énergie éolienne, où ρ est la masse volumique

¹selon le site "eco2mix" de RTE France

de l'air, S la surface d'un cercle de rayon égal à la longueur d'une pôle, et V la vitesse du vent.

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

Mais cette croissance est surtout expliquée par le rôle démontré de l'énergie éolienne dans le besoin de transition énergétique pour l'atténuation et l'adaptation aux effets du changement climatique : en 2021, l'agence internationale de l'énergie (AIE) a établie que 41% des émissions de dioxyde de carbone (CO₂) dues à la combustion d'énergie provenaient du secteur de la production d'électricité². Ce pourcentage est majoritairement expliqué par le secteur de la combustion fossile, notamment le charbon : d'après la sixième édition du rapport du groupe d'experts intergouvernementale sur l'évolution du climat (GIEC)³, la production d'un kilowattheure (kWh) par cette source d'énergie émet plus de 80 fois plus de CO₂ que par l'utilisation d'une éolienne on-shore.

L'éolien est ainsi aujourd'hui publiquement reconnu comme une "énergie d'avenir", et cette reconnaissance se traduit par une amélioration et un développement des formations et compétences dans le domaine de l'énergie renouvelable proposées par les écoles supérieures françaises de renom.⁴ Elle se manifeste également par une croissance en nombre et en chiffre d'affaire du réseau d'entreprises spécialisées dans la gestion de cette énergie (développement, construction, installation, exploitation, fin de vie), et surtout par une forte baisse du coût de production de l'énergie électrique d'origine éolienne, qui est une des alternatives aux filières de production d'énergie fossile les moins chères d'après le rapport de la Cour des comptes de 2021 sur l'analyse des coûts du système de production électrique en France⁵.

Enjeux du développement de la filière éolienne pour la France

Concernant la France, si le secteur de l'énergie éolienne (terrestre et maritime) est considéré comme un enjeu majeur pour une transition énergétique souhaitable, c'est également pour anticiper la fin de vie de la plupart des centrales nucléaires françaises : en faisant l'hypothèse d'un prolongement de cycle de vie à 60 ans, la France perdrait hypothétiquement 56 GW de production nucléaire d'ici 2050 soit environ 40% de sa production d'électricité

²rapport "Key World Energy Statistics 2021" de l'International Energy Agency (IEA)

³<https://www.ipcc.ch/report/sixth-assessment-report-cycle/>

⁴nous pouvons par exemple citer le Mastère spécialisé de l'école nationale supérieure des arts et métiers "expert en projets et production énergies renouvelables", celui de l'école des Mines Paris "énergies renouvelables" ou encore le parcours "énergies renouvelables et efficacité énergétique" de l'université de Lille.

⁵rapport "l'analyse des coûts du système de production électrique en France" de 2021 de la Cour des comptes

totale de 2019, ce qui représente plus de deux fois la puissance installée du parc éolien terrestre à la fin de l'année 2023⁶.

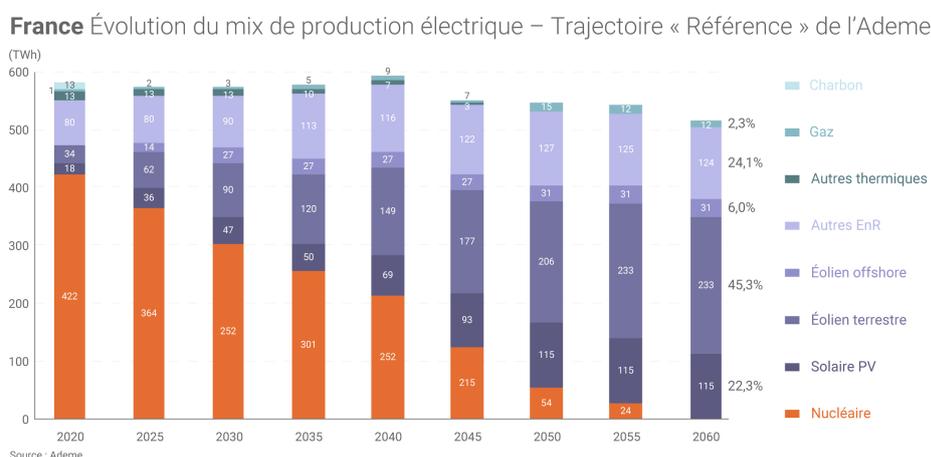


Figure 2.4: Projections du mix de la production électrique française horizon 2060

C'est pour ces raisons que les scénarios à horizon 2050 ou 2060 prévoient une compensation de la production nucléaire par de la production d'électricité d'origine renouvelable, comme on le remarque sur la Figure 2.4, présentant les projections du mix électrique français à horizon 2060 réalisées par l'agence de la transition écologique (ADEME)⁷. C'est également dans ce contexte qu'est né l'objectif de 33% d'énergie renouvelables (biomasse, hydraulique, éolien, solaire) en France en 2030⁸, ce qui représente une croissance d'environ 12% en 7 ans en considérant que la part de production de l'énergie nucléaire se stabilise durant cette période. Sur la base des graphiques présentés dans le rapport RTE de 2022, cette croissance représente une accélération de la production significative par rapport à la dernière décennie, malgré la faiblesse de l'objectif en comparaison d'autres pays Européens, où ce même objectif dépasse 40% en moyenne selon la Commission européenne.

Nous venons de voir que la rapide croissance de la filière éolienne, dans le cas de la France, est nécessaire à la mise en place d'une transition énergétique réussie. Cependant, afin de relever le défi de l'accélération de ce vecteur d'énergie renouvelable, il est nécessaire de développer des stratégies d'exploitation et de maintenance adaptées.

⁶selon le rapport "Futurs énergétiques" 2050 de RTE

⁷voir le rapport de l'ADEME "Trajectoires d'évolution du mix électrique à horizon 2020-2060"

⁸notre-environnement.gouv.fr

2.2 Description et limites de l'organisation de la maintenance actuelle

2.2.1 Objectifs de la maintenance d'éoliennes

Avant de présenter les différentes catégories de maintenance en milieu industriel, nous allons expliciter et contextualiser les missions du service de maintenance en application à l'éolien. Pour cela, nous allons énumérer les causes de défaillance principales, et les différents leviers d'action du service de maintenance pour assurer une minimisation des pertes de production et du coût de maintenance des parcs éoliens. Par la suite, nous parlerons de "pertes de production" pour désigner l'énergie totale non produite durant le cycle de vie d'une défaillance (de la mise à l'arrêt de la turbine à la résolution du problème).

Pour rappel, un **défaut** se définit par un fonctionnement anormal d'un composant ou système, n'empêchant pas la production d'énergie de l'éolienne impactée. Une **défaillance** se traduit par un état de développement d'un défaut suffisamment avancé pour que le composant ou système en question ne puisse plus assurer sa fonction, forçant la mise à l'arrêt de la turbine.

Principales causes de défaillance

Une éolienne peut être mise à l'arrêt par une grande diversité de défaillances. Ces défaillances présentent des causes que Badihi [3] classe en quatre catégories :

- Les **perturbations structurelles** (design défectueux, matériau usé, mauvaise installation...).
- Les **perturbations externes**, principalement causées par une météo défavorable (tempêtes, coup de tonnerre, givre sur les pales...), mais qui peuvent également se traduire par des chocs environnementaux, notamment les rapaces dont la zone de vol se situe dans l'alignement des turbines.
- Les **perturbations électriques** (surcharge électrique, défauts de connexion...).
- Les **défauts d'usure** (détérioration naturelle des composants, corrosion, surchauffe, mauvaise lubrification...).

Ces défaillances peuvent impacter différents composants d'une éolienne, et la revue réalisée par Peng en 2023 [47], détaillant les principales anomalies associées à ces quatre catégories de causes de défaillance, permet de réaliser la grande diversité des modes de défaillance pouvant impacter la productivité d'une telle installation.

D'après la littérature consacrée à la fiabilité du système éolien, le nombre de défaillances augmente de manière globale chaque année [47]. Plus précisément, les études de Carroll [8] et Pfaffel [48], analysant les taux de défaillances des principaux composants de plusieurs parcs éolien, ont montré qu'en moyenne une turbine tombait en panne 10 fois par an.

Rôle de la maintenance sur le contrôle du facteur de charge de l'éolien

Toutes ces défaillances occasionnent des périodes d'arrêts et de ralentissement de la production d'énergie conséquentes, qui s'ajoutent à celles d'origine "naturelle" expliqués par l'intermittence du fonctionnement d'une éolienne. Ces baisses et arrêts de production permettent de définir le **facteur de charge** d'une éolienne : le rapport entre l'énergie électrique produite par la turbine sur une période donnée (généralement un an) et l'énergie qu'elle aurait produit en fonctionnant continuellement à puissance nominale durant cette période. Pour le cas de l'éolien, l'ordre de grandeur de ce paramètre est d'environ 25%. Il est à noter que ce nombre ne signifie pas qu'une éolienne ne produit pas d'énergie pendant 75% de l'année, mais qu'il n'y a pas un vent suffisamment fort sur 3/4 de l'année pour qu'elle produise de l'énergie électrique à "pleine puissance".

Au sein d'une entreprise O&M, le rôle du service maintenance est d'anticiper ces défaillances en intervenant à un état de développement peu avancé des défauts, ou si ce n'est pas possible, d'intervenir au plus tôt sur ces défaillances en vue de valoriser et d'optimiser la durée de vie et de production des parcs. Il joue ainsi un rôle primordial dans l'optimisation de la production de l'énergie éolienne, car il permet de contrôler le facteur de charge de la filière éolienne, et donc de contrôler le coût de production de son énergie (levelized cost of energy, ou LCOE) [43].

En effet, en reprenant les quatre causes de défaillances mentionnées dans la section précédente, nous pouvons déterminer divers leviers d'action du service de maintenance pour la réduction des impacts des perturbations énumérées. Un contrôle de l'installation des composants d'une éolienne, et un suivi continu de leur état permet de limiter les **perturbations structurelles**. Il en va de même pour les **perturbations électriques**, qui peuvent être réduites en contrôlant de façon régulière les signaux électriques induits par le fonctionnement de l'éolienne. Les **défauts d'usure** peuvent être détectés et atténués en assurant une bonne lubrification des composants, et en réalisant un diagnostic régulier de leur état d'usure, ainsi que du fonctionnement des systèmes de refroidissement. Toutefois, il est plus difficile de contrôler les impacts des **perturbations externes** : il n'est pas possible de réduire la force d'une tempête, et les éoliennes ne sont pas équipées de paratonnerre pour les protéger de la foudre. Cependant, des mesures de prévention sont possibles pour réduire les impacts des perturbations ex-

ternes, comme la réduction de la puissance nominale de l'éolienne, ou la mise à l'arrêt durant les périodes de perturbation.

Nous allons maintenant expliciter l'organisation de la maintenance industrielle, en présentant les différentes catégories d'intervention, et les stratégies de planification associées pour mener à bien cet objectif de contrôle du facteur de charge.

2.2.2 Présentation des différentes catégories de maintenance

Dans le cas de la maintenance d'installations éoliennes, on distingue les interventions de **maintenance préventive**, et les interventions de **maintenance corrective**. La norme AFNOR FD X60-000 définit 5 niveaux de maintenance préventive et corrective, selon leur contenu, allant du graissage journalier et remplacement de "petits" composants au démontage complet de la machine.

Maintenance corrective

La maintenance corrective est la plus coûteuse, car elle consiste à remplacer intégralement un composant ou système de l'éolienne défaillant. Une maintenance corrective peut être anticipée, dans le cas où le composant ou système est identifié comme défaillant à un stade de développement où l'éolienne peut encore produire de l'énergie, éventuellement en mode dégradé (réduction de la puissance maximale) jusqu'à la date d'intervention, ou non anticipée si la défaillance est identifiée à un stade de développement trop avancé pour poursuivre la production d'énergie. Cette différence d'anticipation a un fort impact sur le coût total de l'intervention résultante car elle mobilise de nombreuses contraintes industrielles telles que la disponibilité des équipes, des pièces de rechange, la perte de production, ainsi que le potentiel impact de la défaillance sur le fonctionnement des composants environnants [3].

Maintenance préventive planifiée

La maintenance préventive consiste à vérifier l'état d'un composant ou système de l'éolienne, et à éventuellement réaliser des actions mineures pour assurer la bonne reprise de production après intervention (par exemple : graissage des parties tournantes). Cette maintenance est peu coûteuse car elle ne nécessite généralement pas l'arrêt de la production ni le remplacement de composants. Elle est ainsi de courte durée (moins d'une journée), limitant les pertes énergétiques.

La maintenance préventive peut être réalisée à l'aide d'un calendrier, planifiant de manière régulière les interventions. Ce calendrier est généralement établi par le constructeur, propriétaire de l'éolienne, et la fréquence et la durée de l'intervention dépend de la criticité du composant étudié

(fréquence des défaillances et durée moyenne des interventions). Cependant, l'application de cette approche de maintenance n'est pas adaptée à un objectif d'optimisation de la production électrique, car elle ne permet pas de détecter systématiquement l'apparition d'un défaut.

Nous avons vu dans la section 2.2.1 que chaque année, une dizaine de défaillances peuvent impacter une éolienne. Or, la fréquence des interventions de maintenance préventive planifiées est généralement très basse, de l'ordre de quelques mois, voire de l'année. Par conséquent, les défauts peuvent se développer entre deux interventions de maintenance préventive planifiée, jusqu'à un stade où les composants concernés sont considérés comme défaillants, forçant l'arrêt de la turbine. Il devient alors nécessaire de remplacer le composant, ce qui entraîne des pertes énergétiques et financières conséquentes.

Maintenance préventive conditionnelle

L'expansion du parc éolien s'accompagne d'une augmentation du nombre de défaillances, comme l'illustre Olabi [43], ainsi qu'une augmentation du coût de maintenance total, mis en évidence par Peng [47]. Il est ainsi nécessaire de développer au sein des entreprises O&M des stratégies de maintenances permettant la prise en charge au plus tôt des défauts d'une éolienne, afin de prévenir au mieux l'apparition de défaillances.

Une solution triviale serait de revoir le planning de maintenance préventive afin de réaliser une intervention de maintenance préventive quotidienne sur tous les composants d'une flotte d'éoliennes. Cependant, cette solution n'est en pratique pas réaliste vu que le nombre de composants d'une flotte d'éolienne est bien souvent plus important que le nombre de techniciens de maintenance d'un opérateur industriel.

Ainsi, une nouvelle forme de maintenance a été pensée, et est aujourd'hui toujours en plein développement : la maintenance basée sur la surveillance de la condition des composants, ou CBM pour "condition-based maintenance". La norme NF EN 13306 X 60-319 (AFNOR, 2001) définit la maintenance conditionnelle (ou CBM) comme "une maintenance préventive basée sur une surveillance du fonctionnement du bien et/ou des paramètres significatifs de ce fonctionnement, intégrant les actions qui en découlent". La CBM consiste ainsi à intervenir sur un composant sur la base d'informations représentant son état de santé en temps réel : si l'analyse de ces informations indique que le composant présente un défaut, alors une alarme est générée à destination des opérateurs industriels, qui vont alors décider de la mise en place d'une intervention de maintenance préventive. Pour rappel, nous appellerons dans ce manuscrit "surveillance conditionnelle" le processus d'analyse de l'état de santé des composants aboutissant à la génération d'alarmes.

Nous venons de présenter les différentes stratégies de maintenance, qui

différent par le contenu, le coût et la planification des interventions. Parmi elles, la maintenance préventive conditionnelle se présente comme l'approche la plus adaptée à un objectif d'optimisation des coûts de maintenance par la détection précoce des défauts éoliens. Afin de développer cette stratégie de maintenance, il est nécessaire de mettre en place des outils de surveillance conditionnelle adaptés. Nous allons ainsi voir dans les sections suivantes quel est l'état d'implantation de cette approche de surveillance en environnement industriel.

2.3 Surveillance conditionnelle en environnement industriel

2.3.1 Approches traditionnelles

La surveillance conditionnelle de composants d'une éolienne peut être réalisée à partir d'un système de contrôle et d'acquisition de données SCADA (Supervisory Control and Data Acquisition), dont les capteurs sont installés sur ou à proximité des composants de l'éolienne. Ce système génère automatiquement des alarmes lorsqu'un des signaux supervisés dépasse une valeur seuil (par exemple : échauffement anormal d'un roulement de génératrice). Cet outil de surveillance conditionnelle est pratique et peu coûteux, mais il est plus adapté à la détection de défaut de capteur (surchauffe, erreur de mesure...). En effet, les seuils définis dans le système SCADA sont des valeurs physiques représentatives d'un comportement anormal à l'échelle de la dynamique mesurée. Ce système de surveillance conditionnelle en lui-même n'est donc pas adapté à la prévention de défaillances de composant, qui sont le résultat d'un ensemble de relations anormales entre différents systèmes d'une éolienne.

Un autre système de surveillance conditionnelle communément utilisé par les entreprises O&M pour la maintenance préventive conditionnelle est basé sur une analyse fréquentielle des données vibratoires fournies par des accéléromètres. Par abus de langage, la plupart des entreprises utilisent le terme CMS (pour Condition Monitoring System) pour désigner ce processus particulier de surveillance conditionnelle⁹. Cette approche permet la détection de défauts de composants mécaniques à un stade de développement peu avancé, favorisant la mise en place systématique d'interventions de maintenance préventive pour la valorisation de ces composants par l'optimisation de leur durée de vie. Cependant, ce système est plus coûteux à mettre en place, et ne peut s'appliquer qu'aux composants présentant des dynamiques

⁹c'est le cas de l'entreprise VALEMO au sein de laquelle les travaux de thèse sont réalisés. Il sera donc nécessaire de revoir les termes utilisés pour désigner ses outils de surveillance conditionnelle afin d'éviter toute confusion entre le système de surveillance conditionnelle basé sur des données vibratoires et celui proposé

vibratoires mesurables et dont l'analyse permet de révéler des symptômes de défaut en cours, typiquement les roulements de composants tournants.

La maintenance préventive basée sur une surveillance conditionnelle est néanmoins une stratégie très prometteuse pour la stabilisation, voire la réduction du LCOE pour les années à venir car elle permet théoriquement de détecter un défaut de façon précoce, et donc de favoriser et optimiser la planification d'interventions de maintenance préventives à un stade où le composant ou système peut être réparé sans remplacement, permettant l'extension de la durée de vie des composants, et donc la valorisation énergétique et financière de l'éolienne par la réduction des interventions de maintenance corrective. Or, en environnement industriel, la place des systèmes de surveillance conditionnelle parmi les outils d'exploitation reste aujourd'hui marginale.

2.3.2 Contrainte de la politique de maintenance industrielle : exemple de VALEMO

Pour comprendre la difficulté de l'implantation d'un système de surveillance conditionnelle parmi les outils d'exploitation d'une entreprise O&M pour son service de maintenance, nous allons prendre l'exemple de l'entreprise VALEMO au sein de laquelle les travaux de recherche ont été menés.

Répartition des responsabilités de maintenance

Au sein d'une entreprise O&M, les différentes approches de maintenance (corrective, préventive planifiée et conditionnelle) s'articulent autour d'une organisation précise. Dans cette section, nous parlerons de "politique de maintenance industrielle" pour désigner l'organisation des différentes approches de maintenance mises en place par une industrie, c'est à dire les règles qui définissent leurs interactions possible. Nous verrons à travers l'exemple de VALEMO que ces règles peuvent limiter l'efficacité de la maintenance, notamment dans un cadre d'exploitation intégrant plusieurs acteurs, et lorsque la contrainte budgétaire entre en jeu.

L'entreprise VALEMO présente une politique de maintenance industrielle spécifique : bien qu'elle réalise l'exploitation et la maintenance de ses parcs éoliens, sa liberté d'intervention est limitée, car dépendante du constructeur. En effet, contractuellement, VALEMO est chargée de l'ensemble des interventions de maintenance préventive planifiées, dont le calendrier est déterminé par le constructeur. Elle réalise également les interventions de "petit correctif", définies comme les interventions de remplacement de composants dont les pièces de rechange sont incluses dans le stockage interne à l'entreprise (correspondant donc à des bas niveaux, en se référant aux normes AFNOR). Le "gros correctif", représentant les plus lourdes interventions en termes de durée d'arrêt et de pertes, concernant notamment les

composants nécessitant l'utilisation d'un système de levage (grue), est de la responsabilité du constructeur.

La Figure 2.5 résume l'organisation de la maintenance industrielle, dans le contexte de l'entreprise VALEMO.

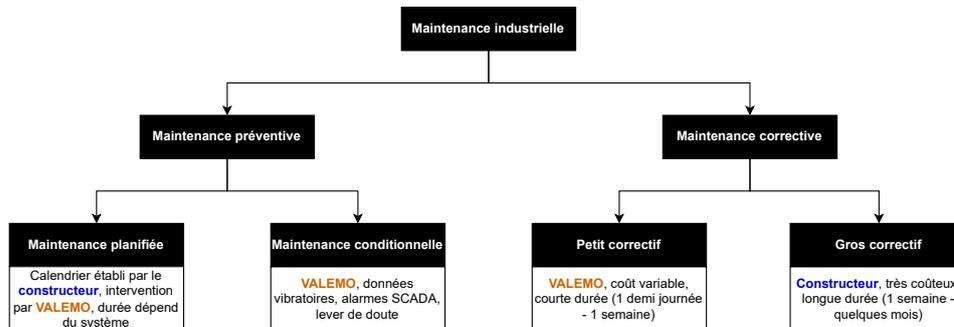


Figure 2.5: Organisation de la maintenance industrielle entre VALEMO et le constructeur

Intégration de la surveillance conditionnelle

Cette répartition des responsabilités de maintenance place la maintenance basée sur des outils de surveillance conditionnelle dans une position défavorable : les interventions associées, déclenchées par des alarmes provenant de divers systèmes (analyse de données vibratoires, supervision SCADA), sont spontanées et initiées par VALEMO, et non le constructeur. Elles sont par conséquent potentiellement définies sur des dates non fixées par le constructeur dans son calendrier de maintenance préventive et VALEMO n'aura donc ainsi pas la possibilité de réaliser son intervention.

Dans le cas où une intervention préventive autorisée par le constructeur est réalisée, il est possible que les techniciens de maintenance concluent à la présence d'un défaut sur le composant surveillé. Si l'état de développement nécessite la mise en place d'une maintenance corrective anticipée à la charge du constructeur, il est possible que ce dernier n'ait pas intérêt à réaliser cette intervention. En effet, un budget fixé de maintenance "gros correctif" est réalisé par le constructeur. Ce budget est établie sur la base d'une estimation du nombre moyen de remplacements de composants prévu par éolienne sur toute la durée de vie du parc. Par conséquent, une maintenance corrective anticipée peut être considérée comme "hors-budget" et non rentable pour un constructeur, notamment durant les dernières années d'activité du parc, où il est financièrement préférable de poursuivre la production en mode dégradé plutôt que de remplacer le composant ou système défectueux.

En conclusion, l'analyse de la politique de maintenance industrielle établie entre un constructeur d'éolienne et VALEMO permet de mettre en lumière

une position défavorable de la maintenance préventive conditionnelle. Il est ainsi nécessaire de proposer aux industriels des systèmes de surveillance conditionnelle suffisamment performants pour qu'ils puissent être valorisés dans les politiques de maintenance industrielle. Dans la section suivante, nous allons définir les besoins et contraintes auxquelles doivent se confronter un système de surveillance conditionnelle pour répondre à cet objectif.

2.3.3 Besoins et contraintes industriels

Grâce au contexte de recherche en convention Cifre, il a été possible de réaliser les travaux de recherche à partir de données industrielles au sein de l'équipe "ingénierie data" de l'entreprise, en plus d'avoir un lien direct avec l'ensemble des experts de la filière éolienne (techniciens d'exploitation et de maintenance, ingénieurs...). Cet environnement de travail nous a permis d'identifier d'une part les besoins des experts de la filière (que nous nommerons par la suite "besoins utilisateurs"), d'autre part les contraintes industrielles liées notamment à l'utilisation de données industrielles pour la mise en place d'une solution de surveillance conditionnelle adaptée à une application industrielle.

Au cours de ce manuscrit, on utilisera le terme "application industrielle" pour faire référence à une application sur des données réelles représentant un échantillon (d'éoliennes ici) intégrant un ensemble varié de parcs éoliens, provenant de différents constructeurs.

Besoins utilisateurs

Pour rappel, la surveillance conditionnelle consiste à analyser un indicateur représentant l'état de santé d'un système ou sous-système de l'éolienne, afin de mettre à jour le planning de maintenance en fonction de cette analyse. En application industrielle, il est important qu'un utilisateur expert du domaine de l'énergie éolienne puisse comprendre les informations fournies par le système de surveillance conditionnelle pour pouvoir prendre une décision sur la nécessité de planification d'une intervention de maintenance préventive. Cette compréhension passe notamment par une transparence et une cohérence physique des modèles utilisés. Ensuite, pour favoriser l'acceptabilité de la solution de surveillance dans un contexte industriel de maintenance conditionnelle, il est important que le processus et les performances en détection puissent s'appliquer à tout élément de la flotte considérée, pour maximiser son potentiel de réduction du LCOE. Enfin, le traitement de l'indicateur doit prendre en compte la ressource de techniciens de maintenance, finie, et dont le déploiement est coûteux pour l'entreprise : les alarmes générées par la méthode de détection de défaut doivent se limiter aux défauts nécessitant une intervention de maintenance préventive, afin d'éviter les déplacements inutiles de techniciens en période de bon fonction-

nement.

Ces réflexions sur les étapes du processus de surveillance conditionnelle permettent de mettre en lumière un ensemble de besoins utilisateurs, qui seront considérés pour le développement de l'approche industrielle proposée dans ce manuscrit:

- Tout d'abord, pour répondre à un besoin de **performance en détection**, le système de surveillance conditionnelle doit engendrer un minimum de fausses alarmes tout en permettant la meilleure distinction possible entre un fonctionnement normal et anormal.
- Ensuite, le système de surveillance conditionnelle doit fournir aux opérateurs une information compréhensible et analysable d'après des connaissances expertes. On parlera alors de besoin d'**interprétabilité**.
- Son **industrialisation** doit être possible, dans le sens où il doit pouvoir être appliqué à tout élément de la flotte d'éolienne considérée sans nécessité d'adaptation de la méthode, et tout en conservant des performances en détection acceptables
- L'**ergonomie** doit également être prise en compte : l'interface de présentation de l'information fournie par le système de surveillance conditionnelle doit être pensée pour limiter la charge mentale additionnelle des utilisateurs. En ce sens, le système de surveillance conditionnelle doit être parcimonieux dans la quantité d'informations à fournir aux opérateurs pour l'aide à la décision.
- Enfin, le système de surveillance conditionnelle devra prendre en compte la limitation et le **coût de la main d'oeuvre** disponible en application réelle, ainsi que le coût de déploiement de cette dernière pour l'entreprise. Cela signifie que le système de surveillance conditionnelle devra limiter au maximum le déploiement "inutile" de techniciens de maintenance.

On peut remarquer que ces besoins semblent difficilement quantifiables pour l'établissement d'un cahier des charges. Ce sont des indicateurs qualitatifs dont l'appréciation dépend de l'expérience de l'opérateur : ils lui permettent de juger de la pertinence de l'outil de surveillance conditionnelle pour une application industrielle.

Contraintes de la donnée industrielle

En travaillant sur un ensemble de base de données industrielles à l'échelle d'une flotte d'éoliennes réparties sur toute la France, nous avons pu identifier des contraintes propres à cette ressource et ce contexte, à considérer pour assurer l'intégration et l'applicabilité d'une solution de surveillance conditionnelle :

- Pour commencer, la contrainte de qualité des données industrielles, qui concerne la **disponibilité**, et la **perturbation** de la donnée.

Contrairement à l'utilisation de données simulées, le processus d'acquisition de données réelles n'assure pas une **disponibilité** des mesures maximale : les signaux obtenus sont couramment exposés à des périodes de pertes de données, de durée pouvant atteindre plusieurs mois.

Ensuite, le système de surveillance conditionnelle doit être robuste aux **perturbations** diverses pouvant impacter les mesures des capteurs. Ces perturbations peuvent provenir d'éléments exogènes (choc sur le capteur), ou endogènes au système d'acquisition des données (mauvaise fixation ou paramétrage du capteur, défaut de communication)

- Enfin, le système de surveillance conditionnelle doit être adapté à la **non-homogénéité** des bases de données au sein d'une flotte d'éoliennes. En effet, chaque constructeur fixe le nombre, les technologies, et la localisation des capteurs à installer sur son modèle d'éolienne. Ainsi, les informations recueillies dans les bases de données de parcs dont les éoliennes proviennent d'un même constructeur sont uniques car de taille propre, et intégrant des signaux aux propriétés statistiques uniques.

Performances en détection	Minimisation du nombre de fausses détection, optimisation de la distinction fonctionnement normal et anormal
Interprétabilité	Information compréhensible et analysable pour un opérateur, modèles simples et physiquement cohérents
Industrialisation	Méthode applicable à tout composant de la flotte d'éoliennes en conservant les mêmes performances en détection
Ergonomie	Minimisation de la charge mentale additionnelle, parcimonie dans la quantité d'informations fournies aux opérateurs
Coût en main d'oeuvre	Prise en compte du caractère limité de la main d'oeuvre (techniciens de maintenance) disponible et déployable

Tableau 2.1: Définition des besoins utilisateurs

Disponibilité	Robustesse aux périodes d'indisponibilité des données
Perturbations	Robustesse aux perturbations diverses impactant les mesures des capteurs
Non-homogénéité	Prise en compte de l'unicité des bases de données des constructeurs

Tableau 2.2: Définition des contraintes des données industrielles

Les tableaux 2.1 et 2.2 résument l'ensemble de ces besoins utilisateurs et contraintes liées aux données industrielles, ainsi qu'une définition concise associée.

2.4 Conclusion

Nous avons débuté ce chapitre en décrivant la filière éolienne et son rôle dans la réalisation d'une transition de l'énergie électrique 100% renouvelable. Nous avons ainsi rappelé la nécessité d'accélérer la croissance de cette filière, afin d'anticiper l'effondrement futur de la production d'énergie électrique d'origine nucléaire dans le cas de la France.

Cet accroissement des efforts de développement de la filière éolienne doit s'accompagner d'une optimisation de la stratégie de maintenance des éoliennes. Nous avons en effet expliqué qu'une stratégie de maintenance efficace permettait une maximisation de la productivité des éoliennes, par une atténuation de l'occurrence des défaillances durant leurs cycles de vie. À travers une analyse des principales approches de maintenance industrielle, nous avons conclu que la maintenance préventive conditionnelle, planifiant des interventions de maintenance préventive sur la base d'une représentation de la condition des composants, était l'approche la plus adaptée à cet objectif de maximisation de la productivité des éoliennes.

Afin de mettre en place une approche de maintenance préventive conditionnelle, il est nécessaire de développer un système de surveillance conditionnelle, réalisant la génération et l'analyse d'indicateurs de défauts. Or dans la dernière partie de ce chapitre, nous avons vu que l'implantation de tels systèmes était marginale au sein des entreprises, notamment à cause d'une politique de maintenance industrielle en défaveur de ces nouvelles méthodes de détection de défaut. Nous avons ainsi conclu ce chapitre en définissant les besoins et contraintes que doit intégrer un système de surveillance conditionnelle afin d'être reconnu comme suffisamment performant par un industriel pour modifier sa politique de maintenance industrielle de manière à donner plus de valeur aux alarmes générées par un tel système.

Dans le chapitre suivant, au travers d'une revue de l'état de l'art sur les systèmes de surveillance conditionnelle en application à l'énergie éolienne, nous allons évaluer la prise en compte de ces besoins et contraintes parmi les solutions proposées dans la littérature. À l'issue de cette analyse, nous pourrons alors positionner les travaux de recherche présentés, et définir les contributions et objectifs proposés.

État de l'art de la surveillance conditionnelle dans l'éolien

Dans ce chapitre, nous allons réaliser un état de l'art des différentes méthodologies développées pour la surveillance conditionnelle en application à l'éolien. Dans un premier temps, nous rappellerons le principe de la surveillance conditionnelle par analyse de résidus, en distinguant les différents courants en application à la filière éolienne. Nous verrons notamment qu'à chaque type de donnée (et donc de capteurs) considéré est associé une approche de surveillance conditionnelle, et que la considération de données SCADA est la plus adaptée à une application industrielle. Nous présenterons alors le principe général de la détection de défaut par utilisation de ces données, en s'appuyant sur la littérature pour développer chaque étape de ce processus. Cette analyse de l'état de l'art nous permettra de réaliser que les récentes avancées en termes d'optimisation de la précision de la prédiction de variables de sortie ont entraîné une complexification des modèles pour la surveillance conditionnelle en application à l'éolien. Or cette complexification engendre une opacité des modèles utilisés, qui sont difficiles à interpréter car non représentatifs de dynamiques physiques connues des experts, Les processus ne sont ainsi pas applicables industriellement, car ne permettent pas le respect des besoins et contraintes définis à l'issue du chapitre 2. Nous exposerons alors en conclusion de ce chapitre le positionnement des travaux de thèse présentés ainsi que les différentes contributions proposées dans le but de lever les verrous de la littérature scientifique identifiés.

3.1 Type de données utilisées pour générer des indicateurs de défaut

3.1.1 Présentation du réseau de capteurs d'une éolienne

La surveillance conditionnelle, permettant la mise en place d'une maintenance préventive conditionnelle, est fondée sur le suivi d'informations représentatives de l'état de santé de composants. Nous nommerons par la suite ces informations "indicateurs de défaut". Les indicateurs de défaut

sont construits à partir de l'exploitation de mesures provenant de capteurs installés sur le composant étudié. Dans le cadre de la surveillance de l'état des différents composants d'une éolienne, ces mesures sont très nombreuses et variées. En effet, une éolienne est un système industriel fortement instrumenté, comme on peut le voir sur la Figure 3.1, illustrant cette diversité en énumérant quelques mesures de dynamiques communément présentes au sein d'une nacelle d'éolienne.

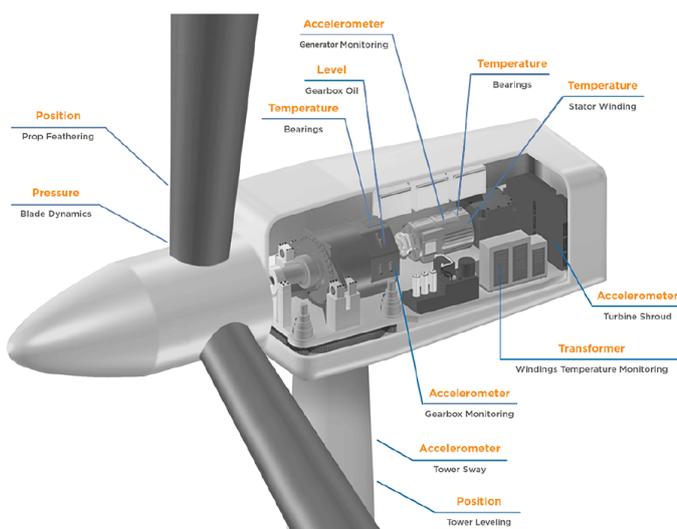


Figure 3.1: Schéma de disposition de capteurs sur une éolienne, et fonction associée [38]

Sur cette Figure, on remarque notamment qu'à chaque type de capteur est associé un objectif de surveillance spécifique : les mesures de températures permettent de surveiller les roulements et enroulements alors que les mesures des accéléromètres sont adaptées à la surveillance des gros composants (multiplicatrice et génératrice) ainsi que du carter de l'éolienne.

L'atlas de l'instrumentation des éoliennes réalisée par Lundquist [34] montre cependant que cette illustration ne représente qu'une partie des capacités d'instrumentation d'une éolienne, et que beaucoup d'autres systèmes de mesure existent. Par conséquent, de nombreuses stratégies de surveillance conditionnelle ont été développées en application à la filière éolienne, et la revue de Tchakoua [62] montre que les méthodologies peuvent être classées selon le type de données utilisées (températures, données vibratoires, vitesses de rotation, vitesse du vent, puissance électrique). Nous allons ainsi présenter les principales mesures considérées dans la littérature, ainsi que leur cadre d'application et de surveillance associé. Cette analyse nous permettra d'identifier le type de mesure le plus adapté au développement d'un outil de surveillance conditionnelle de défauts d'éolienne applicable à un

contexte industriel.

3.1.2 Capteurs de mesures vibratoires

Comme nous l'avons expliqué dans le chapitre 2, les systèmes actuels de surveillance conditionnelle appliqués par la plupart des industriels sont fondés sur une analyse de données vibratoires issues d'accéléromètres. Ce choix de mesures est particulièrement adapté à la surveillance de défauts mécaniques. En effet, pour ce cadre d'application, les premiers symptômes de défaillance sont détectables au niveau des analyses de vibration, comme le rappelle Tchakoua [62].

Nous pouvons ainsi citer Azevedo [2] dont la méthode de surveillance conditionnelle par analyse de données vibratoires a permis la détection d'un défaut au niveau d'un des roulements de la multiplicatrice d'une éolienne quelques mois avant la défaillance du composant.

Cependant, les accéléromètres ne mesurent que des vibrations, et ne peuvent donc pas fournir d'informations sur d'autres variations physiques (par exemple des températures ou des signaux électriques) sensibles au développement d'autres types de défaut. Ainsi, comme l'illustre la revue des analyses vibratoires réalisées dans l'éolien de Teng [63], cette approche de surveillance conditionnelle se limite aux défauts mécaniques impactant les roulements des composants de la chaîne de transmission d'une éolienne (multiplicatrice et génératrice).

3.1.3 Autres capteurs mesurant des variations physiques spécifiques

Les approches de surveillance conditionnelle ne se limitent cependant pas au domaine de la détection de défauts mécaniques par analyse de données vibratoires. Comme l'illustre la Figure 3.1, d'autres types de capteurs peuvent être utilisés pour des objectifs de surveillance différents.

Nous pouvons ainsi mentionner les méthodes de surveillance conditionnelle utilisant des acquisitions de données électriques pour détecter des anomalies liées à la production d'énergie électrique. Ces méthodes visent à contrôler la stabilité du réseau, mais elles permettent aussi de détecter des défauts impactant les composants producteurs d'électricité. Djurovic [11] réalise par exemple la surveillance d'une génératrice d'éolienne par analyse du courant statorique.

D'autres exemples plus marginaux d'analyse de mesures de variations physiques spécifiques pour la surveillance conditionnelle en application à l'éolien peuvent également être mentionnées, comme celles utilisant des jauges de contrainte pour l'identification de défauts de structure [1], ou encore l'analyse acoustique pour la détection de bruits de mesure [60].

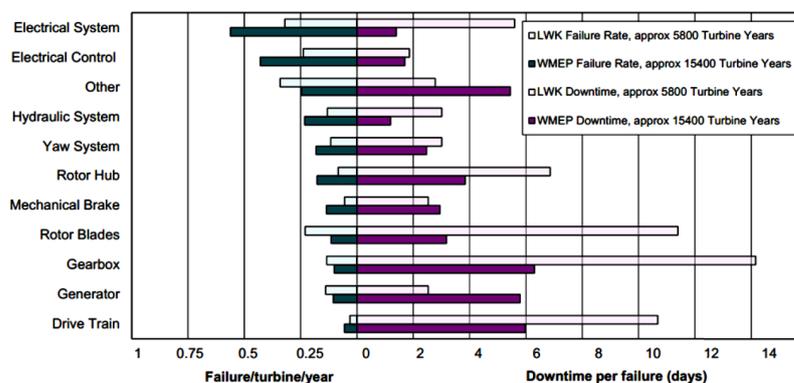


Figure 3.2: Fréquence et durée d'arrêt par défaillance des composants et systèmes d'une éolienne [62]

Comme on le voit dans la Figure 3.2 issue de la revue de Tchakoua [62], illustrant les fréquences annuelles et durées moyennes d'arrêts des défaillances des différents composants et systèmes d'une éolienne, tous les éléments d'une nacelle d'éolienne présentent un coût de maintenance élevé. Par conséquent, une méthodologie de surveillance conditionnelle doit s'appliquer à tous ces éléments pour répondre pleinement à un objectif de réduction significative du coût de maintenance de la filière éolienne.

Ce constat est confirmé par la revue des méthodes de surveillance conditionnelle d'éoliennes de Badihi [3] et l'analyse de défaillance d'un parc éolien réalisée par Lin [32], qui montrent que les défauts les plus critiques (en termes de fréquence d'apparition et de durée de mise à l'arrêt de la turbine occasionnée) ne sont pas limités à la génératrice et à la multiplicatrice d'une éolienne, qui constituent le cadre d'application des études susmentionnées.

Toutefois, il semble difficile d'imaginer un système de surveillance conditionnelle intégrant et fusionnant une grande diversité de capteurs relativement coûteux pour assurer la détection de défaut de la totalité des composants critiques présents sur la Figure 3.2. Nous allons ainsi étudier une autre technologie d'acquisition de données, qui peut être utilisée pour augmenter le potentiel de détection des outils de surveillance conditionnelle.

3.1.4 Données SCADA

Le système de contrôle et d'acquisition de données SCADA (Supervisory Control and Data Acquisition), est une technologie de mesure de variations physiques, représenté par un réseau de capteurs installés sur ou à proximité des composants de l'éolienne. Les données SCADA sont typiquement utilisés dans les systèmes industriels pour fournir aux exploitants des données essentielles à l'analyse de l'état des machines, notamment par l'intégration d'un système de levée d'alarmes par capteur. On peut distinguer deux approches

de traitement des données SCADA pour la surveillance conditionnelle en application à l'énergie éolienne. Premièrement, citons la modélisation de la courbe de puissance d'une éolienne pour la détection des anomalies de production [35]. Comme le montre l'étude de Gonzalez [17], cet objectif est principalement accompli à l'aide de données SCADA acquises à une haute fréquence, de l'ordre de la seconde. Cela permet en effet une plus grande précision dans les résultats de modélisation. L'autre fonction principale de l'analyse des données SCADA est la détection de défauts dont le développement se traduit par des symptômes thermiques, par exemple une usure de roulement qui peut entraîner son échauffement, ou un défaut de système de refroidissement entraînant une mauvaise régulation thermique du composant concerné. Ces défauts font partie des plus fréquents et dont la maintenance est la plus coûteuse d'après la revue de Marugán [39]. Pour ces raisons, nous nous intéressons dans ces travaux de recherche à cette deuxième catégorie d'utilisation, réalisée à l'aide de données SCADA échantillonnées à basse-fréquence (de l'ordre de la minute).

L'utilisation de données SCADA basse-fréquence présente différents avantages qui motivent le développement d'un processus de surveillance conditionnelle à l'échelle d'une éolienne sur la base de ce type de données. Premièrement, les capteurs associés aux données sont peu onéreux, et généralement installés durant la construction et la mise en production d'une éolienne. Le processus d'acquisition et de surveillance des données mesurées par la technologie SCADA est de surcroît simple à appréhender pour un industriel, contrairement aux données vibratoires par exemple, dont l'exploitation nécessite pour une entreprise l'emploi d'un ou plusieurs experts, donc un coût de main d'œuvre supplémentaire. Ensuite, bien que la revue des méthodes de surveillance conditionnelle utilisant des données SCADA réalisée par Maldonado [36] montre qu'il est d'usage dans la littérature de surveiller principalement la génératrice et la multiplicatrice d'une éolienne, il est possible de surveiller d'autres composants critiques identifiés dans la Figure 3.2 en utilisant des données SCADA basse-fréquence, contrairement à l'utilisation de données vibratoires ou autres mesures physiques spécifiques mentionnées dans la section précédente. En effet, le réseau de capteurs associé à cette technologie d'acquisition recouvre l'intégralité de ces composants, comme le montre cette même revue de Maldonado.

Ces constats permettent de positionner la surveillance conditionnelle par analyse de données SCADA comme le courant de recherche le plus pertinent dans un objectif de réduction des coûts de maintenance de la filière éolienne. Nous allons par la suite présenter le principe de surveillance conditionnelle associé à l'utilisation de ce type de données. Nous identifierons et analyserons pour chacune des étapes de ce processus les méthodologies de la littérature en application à la filière éolienne.

3.2 Génération et analyse de résidus à l'aide de modèles de bon fonctionnement

3.2.1 Principe général

Le schéma de la Figure 3.3 représente le principe général de surveillance conditionnelle réalisé par l'analyse de résidus issus de modèles construits à partir de données SCADA. L'idée générale est de prédire la valeur d'une variable y à l'aide d'un modèle décrivant les liens existants entre cette variable et des variables d'entrée quand le système est en bon fonctionnement, et de comparer la valeur prédite avec la valeur mesurée en ligne afin de générer un résidu qui servira d'indicateur de défaut.

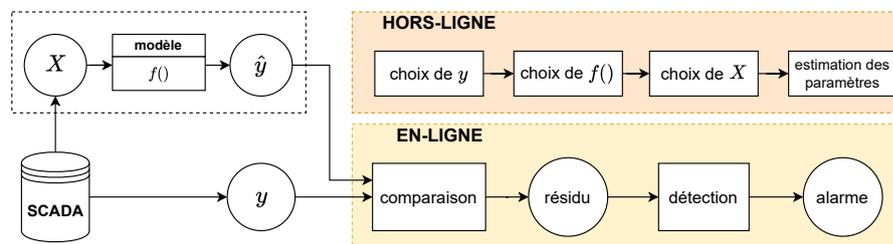


Figure 3.3: Principe de surveillance conditionnelle par génération de résidus issus de modèles construits à partir de données SCADA

Hors-ligne

La première partie de ce processus est réalisée "hors-ligne", c'est-à-dire à partir d'un échantillon de données SCADA issu d'une période temporelle passée, dont on connaît le contenu.

Dans le cadre des travaux de recherche présentés, on considère un problème de régression, consistant à prédire les valeurs d'une variable SCADA y à partir de données d'apprentissage. Ainsi, pour un composant donné à surveiller, la première étape de cette partie hors-ligne consiste à déterminer la variable de sortie y du modèle, c'est-à-dire la variable à estimer. Comme nous l'avons vu dans la présentation des différents courants de surveillance conditionnelle selon le type de données utilisé, l'utilisation de données SCADA est adaptée à la détection de défauts thermiques. Ainsi, les variables de sortie sont généralement des températures, choisies pour leur sensibilité aux défauts thermiques analysés.

Ensuite, une hypothèse sur la complexité du modèle (linéaire, quadratique...) est réalisée, et les variables X de la fonction f associée (les variables d'entrée du modèle), utilisées pour estimer la variable de sortie y , sont déterminées. Cette sélection est réalisée par une analyse des relations entre les variables de la base de données SCADA avec la variable de sortie. Nous

verrons par la suite qu'il existe plusieurs méthodes permettant de réaliser cette sélection.

Finalement, après avoir choisi le type de modèle puis déterminé les variables d'entrée et de sortie du modèle, on procède à l'apprentissage des paramètres du modèle en construction, par analyse des relations liant la variable de sortie aux variables d'entrée du modèle sur des données ne présentant pas de défaut, selon le principe de modélisation en bon fonctionnement, qui sera présenté dans la section suivante.

En-ligne

Une fois les modèles construits et paramétrés à l'aide de données SCADA durant une phase hors-ligne, de nouvelles données SCADA en temps réel sont utilisées pour l'application de la seconde partie du processus de surveillance conditionnelle. Un indicateur de défaut, ou résidu, représentant l'état de santé du composant, est tout d'abord construit. Ce résidu est établi par comparaison entre la valeur de la variable de sortie mesurée par le système SCADA, et celle modélisée à partir des variables d'entrées. Les valeurs de ce résidu sont ensuite analysées afin de détecter l'éventuelle présence de défaut, pour la génération d'alarmes à destination d'un opérateur industriel. Il est à noter que dans le cadre de la détection de défaut, l'analyse de ce résidu est "binaire", et ne permet pas d'obtenir d'informations sur l'état de développement du défaut : son objectif est simplement d'indiquer la présence ou non d'un défaut au niveau du composant surveillé.

Dans la suite de cette section, nous allons développer les différentes phases du processus de surveillance conditionnelle illustré dans la Figure 3.3, en réalisant un état de l'art critique des méthodologies dominantes de la littérature.

3.2.2 Revue des méthodologies par étape

Apprentissage des modèles de bon fonctionnement

Dans un contexte industriel d'utilisation de données SCADA dans le domaine de l'éolien, les données de défaillances sont très limitées. Par conséquent, dans le cadre du développement de méthodes de surveillance conditionnelle en application industrielle, il n'est pas toujours possible d'entraîner et de valider des modèles en utilisant des données présentant des défauts, puisque ces défauts peuvent ne pas exister.

Pour répondre à cette problématique, une approche de modélisation pertinente consiste à construire et entraîner des modèles de régression en utilisant des données sur des périodes où les composants étudiés ne présentent pas de défaut. Cette approche, appelée "normal behaviour modeling" (ou NBM) dans la littérature, permet d'estimer les valeurs d'une variable de sortie à partir des données en condition de bon fonctionnement de variables

d'entrée. Bien que ne considérant pas de données sur des périodes de défaut pour la construction des modèles, l'approche NBM permet d'obtenir des résidus permettant une bonne sensibilité aux défauts, comme le montre le système SIMAP proposé par Garcia en 2006 [16] appliqué à la détection de défaut de roulement d'une multiplicatrice d'éolienne. On peut également mentionner le système proposé par Schlechtingen en 2013 [56], évalué sur cinq types de défaut dans un autre article [55]. Ce système intègre 45 modèles de bon fonctionnement générés par application d'un réseau de neurone artificiel à partir de données SCADA basse-fréquence provenant de 33 signaux.

Il est à noter que dans un contexte industriel, l'hypothèse que le système est en bon fonctionnement pendant l'acquisition des données d'apprentissage est complexe à vérifier. En effet, les données SCADA présentent fréquemment des perturbations mineures ainsi que des valeurs aberrantes, même durant des périodes où aucun défaut n'est relevé par la maintenance. Plusieurs études portent ainsi sur la prise en compte de ces perturbations pour améliorer la qualité des modèles. Nous pouvons par exemple mentionner l'étude de surveillance des performances d'une éolienne de Gonzalez [17], qui utilise le journal des alarmes réalisé par le service de maintenance pour sélectionner les données d'apprentissage des modèles, ou encore Letzguis [29] qui propose une méthode de détection de points de rupture au niveau des signaux SCADA pour améliorer l'annotation des bases de données considérées dans les approches de surveillance conditionnelle.

Types de modèles

Les modèles utilisés dans les systèmes de surveillance conditionnelle de la littérature en application industrielle sont des modèles de bon fonctionnement. Ils décrivent l'évolution attendue d'une variable de sortie en fonction de variables d'entrée quand le système est en bon fonctionnement. Dans la littérature de génération de modèles de bon fonctionnement en application à la filière éolienne, la fonction $f()$ représentant la forme mathématique du modèle peut être déterminée selon différentes stratégies.

Il est tout d'abord possible de l'établir à partir de connaissances expertes issues d'une revue des lois physiques régissant les dynamiques des composants modélisés. C'est le cas de la méthode de détection de défaut de roulement d'arbre lent réalisée par Cambron [6], qui se base sur le premier principe de conservation de l'énergie de la thermodynamique pour modéliser la température du roulement à partir d'autres sources de chaleur selon l'équation 3.1, où C_p est la capacité thermique de la nacelle de l'éolienne. Le premier terme de la partie droite de l'équation représente la conductivité thermique du système étudié, avec $t_{\text{roulement}}$ la température du roulement, t_{interne} la température ambiante dans la nacelle, et R la résistance thermique,

le second la chaleur produite par la rotation de l'arbre lent (ω), avec μ le coefficient de friction, et le dernier l'effet Joule produit par la génératrice, avec α la quantité d'énergie transformée en chaleur par unité d'électricité produite et P la puissance produite par l'éolienne.

$$C_p * \frac{dt_{\text{roulement}}}{dt} \approx R^{-1} * (t_{\text{roulement}} - t_{\text{interne}}) + \mu * \omega^2 + \alpha * P \quad (3.1)$$

Dans le cas de la modélisation thermique du roulement principal d'une éolienne, une fonction $f()$ peut ainsi être trouvée par exploitation de lois physiques, et le modèle obtenu présente l'avantage d'être compréhensible par un expert en plus d'être très précis. Toutefois, une telle fonction ne peut pas être considérée pour tous les composants d'une éolienne, à cause d'un manque de connaissance sur l'existence et la formulation de ces lois pour tous les composants et sous-composants d'une éolienne.

Pour cette raison, il est plus courant de trouver dans la littérature des modèles dont la forme est déterminée par une analyse des données SCADA. Ces modèles peuvent être linéaires et estimer des liens de similarités entre y et les variables d'entrée X . C'est le cas de la méthode de détection de défauts de roulement de génératrice de Lebranchu [27] qui estime une variable de sortie à partir de relations linéaires avec des températures de composant proches. Les modèles linéaires présentent l'avantage d'être interprétables par un utilisateur expert, mais ce ne sont pas les plus précis.

Pour obtenir des modèles plus précis, une possibilité est de considérer des fonctions non-linéaires, obtenues à l'aide de méthodes de modélisation issues du Machine Learning. Comme le montre Stetco [59], les algorithmes de réseaux de neurones (NN), support vector machine (SVM) et de decision tree (DT) sont les approches de modélisation guidées par la donnée les plus utilisées en surveillance conditionnelle d'éolienne. De tels algorithmes peuvent être utilisés pour estimer des relations entre un ensemble de variables d'entrée avec y . Par exemple, Zhang [75] considère un réseau de neurone artificiel pour modéliser la température de roulement arrière de la génératrice d'une éolienne à partir de la vitesse de rotation de l'arbre lent et de la puissance active.

Le réseau utilisé dans cet article est relativement simple, et permet d'obtenir un modèle précis et interprétable lorsqu'aucun modèle mathématique n'existe. Cependant, d'après les revues de 2018 des systèmes de réseaux de neurones appliqués aux systèmes d'énergie éolienne de Marùgan [39] et Helbing [20], la majorité des études de surveillance conditionnelle appliquées à l'éolien utilisent des réseaux de neurone artificiels plus complexes issus de l'apprentissage profond (ou deep learning) tels que le perceptron multicouche [69], l'auto-encodeur [9], les réseaux de neurone convolutifs [73], ou une combinaison de ces approches [64], permettant d'obtenir une meilleure précision de prédiction par la modélisation de relations non-linéaires complexes.

Choix des variables d'entrée des modèles

Les variables d'entrée d'un modèle peuvent être sélectionnées manuellement en considérant les relations de similarité et de causalité avec la variable de sortie du modèle. L'exploitation de lois physiques faisant intervenir y est également un moyen d'identifier les variables permettant sa modélisation.

Comme le montre la revue de Khan de 2024 [23] sur les méthodes d'apprentissage fondées sur la physique pour les systèmes de gestion de santé (SHM), il est possible d'améliorer ces stratégies de sélection de variables d'entrée en considérant en complément de connaissances physiques une exploitation des données utilisées. Nous pouvons ainsi mentionner la méthode de détection d'anomalie proposée par Mian Du [13], dont les variables d'entrée des modèles sont identifiées à l'aide de connaissances physiques et d'une analyse de coefficients de corrélation, ou encore l'analyse de défauts de roulement de Kusiak [25], qui mobilise des connaissances expertes pour réduire le nombre de variables à analyser par l'algorithme de sélection de variables par réseau de neurones à celles ayant un impact sur les températures de roulement.

Cependant, en application industrielle, ces relations et lois physiques ne sont pas toujours connues. Ainsi, une approche plus adaptée à une application industrielle est de sélectionner automatiquement les variables d'entrée en utilisant une démarche entièrement guidée par les données. Cette stratégie de choix de variables d'entrée fait l'objet d'un champ de recherche foisonnant. Ainsi, beaucoup d'algorithmes existent, comme le montre la revue thématique de Kumar [24], et il n'y a aujourd'hui pas d'algorithme de sélection automatique de variables d'entrée consensuel, comme le montre la revue de Nunes [42]. Certaines applications proposent des méthodes très simples à implanter et exécuter pour sélectionner les variables d'entrées. C'est le cas de Lima [31] et Tautz [61], qui réalisent leur choix de variables d'entrée à l'aide d'une analyse de coefficients de corrélation des variables considérées avec la variable de sortie. D'autres approches présentent un temps de calcul bien plus long. C'est le cas des algorithmes génétiques, comme l'analyse Venkatesh [65]. Ces algorithmes sont utilisés dans les méthodologies de détection de défauts d'éolienne de Schlechtingen [56, 55] et Khan [22], et sont particulièrement adaptés aux bases de données contenant un grand nombre de variables représentant des dynamiques physiques très hétérogènes. Enfin, d'autres études proposent des approches de sélection de variables fondées sur des algorithmes de régression pénalisée, permettant de répondre au double objectif de régression et de sélection de variables d'entrée. Par exemple, Reder [53] confronte deux algorithmes de régression pénalisée au même problème de modélisation, et retient l'algorithme Elastic-Net, fournissant un ensemble de variables d'entrée plus réduit (et donc un modèle moins complexe) pour une précision de modélisation équivalente.

Construction du résidu

Une fois les variables d'entrée sélectionnées, et les modèles entraînés à l'aide de données SCADA représentatives de conditions de bon fonctionnement selon l'approche NBM, l'état de santé du ou des composants surveillés est représenté par un résidu. Ce résidu est construit par comparaison entre la valeur de température mesurée par le système SCADA (variable de sortie du modèle) et la valeur de température estimée par le modèle de bon fonctionnement. On trouve dans la littérature de surveillance conditionnelle appliquée à l'éolien plusieurs méthodologies réalisant cette comparaison.

Liu [33] et Yuan [72] utilisent par exemple la distance de Mahalanobis. Cette approche intègre dans son calcul la dispersion des composantes considérées. Elle est ainsi plus adaptée à une application à des données industrielles que la distance Euclidienne car elle est plus robuste aux bruits de mesure des données industrielles. La méthode de détection de défaut de capteurs appliquée à l'éolien d'Harrou [19] propose une autre approche de comparaison en considérant la distance de Kantorovich, fondée sur les distributions statistiques de deux variables aléatoires.

Nous pouvons également citer Lebranchu [27] qui propose un résidu défini à partir des données de toutes les turbines d'un même parc éolien, afin de contrer les sensibilités aux saisonnalités et aux modes de fonctionnement que peuvent engendrer l'utilisation de modèles de régression linéaire. Dans un contexte de génération de résidus pour la détection de défaut en application industrielle à une flotte d'éolienne, cette méthode de calcul de résidu "multiturbine" semble être la plus adaptée, car la plus robuste aux contraintes inhérentes à ce contexte d'application.

Génération des alarmes pour la détection de défaut

La dernière étape du processus de surveillance conditionnelle consiste à analyser les valeurs du résidu construit afin de détecter la présence de défauts pour la génération d'alarmes à destination de l'opérateur industriel. Pour se faire, on distingue dans la littérature deux approches principales : la détection de rupture et le couplage du résidu à un seuil de détection.

L'approche de génération d'alarmes par détection de rupture consiste à détecter les instants où la distribution statistique d'un signal donné présente des changements significatifs. Dans le cas de la surveillance conditionnelle de composants, on considère ainsi que le résidu est une variable aléatoire dont on connaît la distribution statistique quand le système est en bon fonctionnement. On recherche alors un changement dans les paramètres de cette distribution que l'on associe à une alarme, indiquant la présence d'un défaut. On peut citer Yuan [72] qui a appliqué cette approche pour la prédiction de défauts de multiplicatrice. Toutefois, les résultats de cette étude montrent que cette approche est trop sensible aux perturbations, et

peut ainsi engendrer un bilan de fausses alarmes significatif.

Une autre approche d'analyse du résidu pour la détection de défaut est de comparer ses valeurs à un seuil de détection. Ce seuil est communément calibré à partir de la distribution statistique du résidu pendant une période d'apprentissage, nécessairement en bon fonctionnement. Le processus de prédiction de défaut développé par Lima [31] considère ainsi un seuil de détection égal à trois fois l'écart-type des résidus pendant la période d'apprentissage. Bien que présentant des performances de détection peu précises dans le temps, comme rappelé par Yuan [72], cette approche est la plus largement considérée dans la littérature de détection de défaut par analyse de résidus. Ce constat s'explique par le fait que cette approche est significativement moins coûteuse que l'approche de détection de rupture. Elle est également plus simple à interpréter pour un opérateur industriel.

Nous venons de présenter dans cette section le principe général de génération et d'analyse de résidus construits à partir de modèles de bon fonctionnement sur la base de données SCADA. Nous avons par la suite énuméré les méthodologies identifiées dans la littérature scientifique, pour les différentes phases hors-ligne et en-ligne du processus de surveillance conditionnelle en application à l'éolien. Cette analyse nous a permis de mettre en lumière des courants majoritaires sur différentes phases du processus de surveillance conditionnelle. Nous pouvons ainsi conclure qu'il est commun de réaliser l'apprentissage des modèles à partir de données ne présentant pas de défaut, de sélectionner les variables d'entrée de ces modèles à partir d'une analyse des données SCADA, et de comparer les valeurs du résidu à un seuil pour procéder à la détection de défauts.

3.3 Verrous à une application industrielle

Nous avons pu constater dans la section précédente que la littérature scientifique des processus de surveillance conditionnelle appliquées à l'éolien présente une tendance à la complexification des méthodes de modélisation par l'utilisation de réseaux de neurones artificiels complexes, pouvant être catégorisés dans le domaine de l'apprentissage profond (Deep Learning). Nous allons voir dans cette section que ce choix méthodologique n'est pas adapté à certains besoins et contraintes définies dans les Tableaux 2.1 et 2.2 du chapitre 2, contraignant l'application industrielle des processus de surveillance conditionnelle associés.

3.3.1 Optimisation de la précision des modèles

Dans leurs revues de processus de surveillance conditionnelle appliqués à l'éolien, Pandit [44] et Khan [23] concluent que les modèles de bon fonctionnement des méthodes de détection de défauts atteignent aujourd'hui

des niveaux de précision très hauts. Ce constat s'illustre dans la littérature par la prise en compte de la précision des modèles (généralement évaluée à l'aide de mesures d'erreurs de prédiction) comme principale mesure de performance des méthodes de détection de défaut [67, 25, 15]. Ainsi, dans la littérature, l'amélioration de la précision des modèles utilisés pour la surveillance conditionnelle dans l'éolien semble toujours être perçue comme un besoin, sans réflexion sur les implications de ce gain : la considération de la précision des modèles comme unique indicateur de performance occasionne en effet différents freins à une application industrielle.

Sensibilité aux données d'apprentissage

En comparaison d'approches de modélisation plus simples telles que la régression linéaire, les modèles générés à partir de réseaux de neurones exigent une grande quantité de données d'apprentissage "propres", c'est-à-dire ne présentant pas de perturbations. En effet, la précision du modèle généré par réseau de neurone artificiel par Zhang [75] dépend fortement de la qualité et de la quantité des données d'entraînement. Pour cette raison, et comme le fait remarquer la revue de Stetco [59], la majorité des méthodes de la littérature associées sont appliquées à des données simulées, présentant une qualité de mesure largement supérieure aux données réelles, pouvant être substantiellement bruitées voire erronées.

Modèles "boîte noire"

Une conséquence de la concentration d'efforts sur l'optimisation de la précision est le développement de modèles de type "boîte noire" (dont on ne connaît pas la fonction $f()$ reliant les variables d'entrée à y) qui ne sont ni physiquement cohérents ni interprétables par un utilisateur expert du domaine d'application.

Bien que certaines études, analysées par Drobnič [12], tentent d'intégrer le facteur "interprétabilité" dans la construction des modèles, ceux-ci restent trop peu compréhensibles, et le besoin d'**interprétabilité** reste rarement considéré comme un critère de validation, justement parce qu'il ne s'accompagne pas toujours d'un gain de précision. Or, comme nous l'avons expliqué à l'issue du chapitre 2, en application industrielle, un processus de surveillance conditionnelle doit être interprétable, car un opérateur industriel peut manifester une certaine réticence à l'utilisation de méthodes de surveillance conditionnelles fondées sur des modèles qu'il ne comprend pas, et considérer les informations fournies comme non fiables. De plus, comme Khan [23] l'analyse, augmenter l'interprétabilité des résultats de méthodes de modélisation permet de réduire le risque de surapprentissage, qui fait parti des inconvénients des algorithmes de Deep Learning pour la modélisation, comme le rappelle la revue de méthodes d'apprentissage profond pour l'éolien de Helbing [20].

3.3.2 Absence de contrôle sur la sélection de variables d'entrée

Une conséquence de la complexification des modèles pour la surveillance conditionnelle appliquée à l'éolien est l'absence de contrôle sur la sélection de variables d'entrée : les algorithmes et méthodes de sélection considérées ne pénalisent pas le nombre de variables retenues, qui peut être important. En effet, comme le conclut la revue des systèmes de réseaux de neurones appliqués aux systèmes d'énergie éolienne de Marùgan [39], les algorithmes majoritairement utilisés étant rapides à exécuter, les efforts d'optimisation sont concentrés sur la précision des modèles obtenus. Le problème de sélection de variables n'est ainsi pas considéré comme prioritaire dans la majorité des études. Par exemple, la méthode d'analyse de défauts de roulement de Kusiak [25] permet de "réduire" le nombre de variables d'entrées à 18. On peut également mentionner l'approche de détection de défaut de multiplicatrice de Yuan [72] qui considère des modèles construits à partir de 3 ensembles distincts de 10 variables d'entrée. Nous allons voir dans cette section que cette absence de régulation du nombre de variables d'entrée des modèles se confronte à plusieurs besoins et contraintes définis dans les Tableaux 2.1 et 2.2, qui rendent les processus de surveillance conditionnelle associés non adaptés à une application industrielle.

Impact sur la compréhension et l'analyse des modèles par un utilisateur expert

Tout d'abord, peu importe le type de modèle considéré, un modèle contenant 18 variables d'entrée tel que celui proposé par Kusiak [25] peut difficilement être interprété par un utilisateur expert. En effet, lorsqu'une alarme est générée par un système de surveillance conditionnelle, cet utilisateur va procéder à une analyse complémentaire de l'état du composant à l'aide de connaissances expert. Cette analyse peut notamment être réalisée par une analyse rudimentaire des signaux utilisés pour construire le modèle. Si ce modèle contient plus d'une dizaine de variables, cette analyse devient complexe, voire impossible, et le besoin d'**interprétabilité** du processus de surveillance conditionnelle n'est ainsi pas respecté.

Faible reproductibilité des résultats de sélection

En dehors de la difficulté d'analyse et d'interprétation de modèles contenant un grand nombre de variables d'entrée, le résultat de sélection de variables d'entrée doit être reproductible à toutes les turbines d'une flotte d'éolienne pour répondre à un besoin d'**ergonomie** d'un processus de surveillance conditionnelle en application industrielle. Cela permet en effet de limiter la charge mentale des opérateurs industriels en limitant le nombre de variables à analyser pour chaque alarme identifiée par les indicateurs de défauts. Or,

au sein d'une flotte d'éolienne, les instrumentations (nombre et type de capteurs installés) des turbines dépendent du fabricant de la turbine (contrainte de **non-uniformité** des bases de données définie à l'issue du chapitre 2). Les bases de données sont donc distinctes d'un fabricant de turbine à un autre, et, pour reprendre l'exemple de Kusiak [25], il n'est pas certain que les 18 variables qui composent le modèle soient disponibles pour tous les types d'éolienne qui composent une flotte industrielle.

Augmentation du risque d'indisponibilité des résidus

Quand bien même l'ensemble de variables d'entrée d'un modèle de la taille de celui de Kusiak [25] ou Yuan [72] serait disponible dans les bases de données de la totalité des éoliennes d'une flotte, en application industrielle, l'augmentation du nombre de variables d'entrée d'un modèle s'accompagne d'une augmentation des risques d'indisponibilité de la solution de surveillance conditionnelle. En effet, comme nous l'avons expliqué dans le chapitre 2, en application industrielle, la **qualité des données** est bien plus faible que celle des données simulées, à cause de la forte présence de perturbations. Ces perturbations peuvent se traduire par des bruits de mesure (choc ou mauvaise fixation des capteurs) ou des pertes de données. Ainsi, plus un modèle contient de variables d'entrée, plus il est exposé à ces perturbations et plus il risque d'être indisponible sur une période donnée.

3.3.3 Restriction de l'échelle d'application

Les revues d'utilisation de réseaux de neurone artificiels en application à l'éolien de Marugán [39] et Peng [46] ont conclut que de telles méthodes de modélisation s'adaptent difficilement à d'autres problèmes de détection que ceux sur lesquels elles ont été façonnées. En d'autres termes, il est difficile d'adapter une méthode de détection de défaut utilisant un réseau de neurone complexe, développée et validée pour la surveillance du roulement arrière d'une génératrice d'une turbine appartenant à un constructeur précis, à la surveillance d'un autre composant, provenant d'un constructeur différent ou non. Cette limite est définie par les revues de Pandit [44] et Khan [23] parlent ainsi d'un mauvais équilibre entre complexité et échelle d'applicabilité des résultats. On peut notamment remarquer que dans la plupart des études susmentionnées, l'application des méthodes de détection à d'autres composants et d'autres éoliennes que celles utilisées pour l'étude fait l'objet de perspectives de recherche, ce qui implique une perspective d'**industrialisation** très lointaine.

Comme l'analyse Pandit [44], la majorité des études de surveillance conditionnelle appliquées à l'éolien concernent les roulements des composants de la chaîne de transmission d'une nacelle d'éolienne, ainsi qu'à la multiplicative d'après Marugán [39]. Les autres composants dont le coût de

maintenance est conséquent selon la Figure 3.2 ne sont ainsi pas intégrés à cette littérature. Un exemple concret est celui de la génératrice, dont la surveillance conditionnelle est aujourd'hui très peu documentée en comparaison de la multiplicatrice, alors qu'elle fait partie des composants les plus critiques en termes de fréquence et de durée de pannes, selon les revues de Tchakoua [62] et Badihi [3].

Cette limitation du champ d'application des plus récentes méthodes de modélisation concerne également le type d'éolienne étudié : dans tous les articles et revues susmentionnés, les études sont menées sur une ou un ensemble d'éoliennes provenant d'un même constructeur, sans vérification de leur applicabilité sur des turbines provenant d'autres constructeurs. Or, selon la contrainte de **non-homogénéité** des bases de données industrielles (Tableau 2.2), la généralisation de la modélisation d'un composant d'une turbine d'un constructeur d'éolienne à un composant d'une turbine provenant d'un autre constructeur n'est pas assurée, du fait que chaque constructeur d'éolienne conçoit des turbines ayant leurs propres caractéristiques.

En conclusion, le besoin d'**industrialisation** défini à l'issue du chapitre 2, selon lequel un processus de surveillance conditionnelle doit pouvoir être appliqué à tout élément d'une flotte d'éolienne sans adaptation et ajustement de la méthodologie, n'est ainsi pas respecté.

L'analyse de l'état de l'art de la surveillance conditionnelle appliquée à l'éolien réalisée a permis de mettre en évidence un principe général adopté par la majorité des études : utilisation de données SCADA pour la génération de résidus construits à partir de modèles de bon fonctionnement pour la détection de défauts thermiques. Cette fonction est réalisée à partir de méthodes de modélisations complexes, conçues selon un objectif mono-critère d'optimisation de la précision des modèles.

Nous avons montré que ce choix s'accompagnait d'une dégradation de la compréhension des modèles par un utilisateur expert, ainsi qu'à des risques de surapprentissage, et qu'il impliquait la considération de méthodologies dont les performances sont très sensibles à la qualité de la donnée. Cette complexification des modèles s'accompagne de surcroît d'un manque de régulation du nombre de variables d'entrée, diminuant la reproductibilité des modèles, et augmentant le risque d'indisponibilité des indicateurs de défaut associés. Enfin, nous avons montré que les méthodes de modélisation proposées par les plus récentes études de surveillance conditionnelle présentaient des limites d'application, et qu'aucun processus ne présentait l'ambition d'une application à l'ensemble des composants dont le coût de maintenance est le plus critique.

Par conséquent, les solutions de surveillance conditionnelle appliquées à l'éolien issues de la littérature ne sont pas adaptées à une utilisation à l'échelle d'une entreprise réalisant l'exploitation et la maintenance d'une flotte d'éoliennes, car elles ne répondent pas à l'ensemble des besoins et

contraintes définis dans les Tableaux 2.1 et 2.2.

3.4 Conclusion

Bilan de l'état de l'art

Une éolienne est une machine industrielle fortement instrumentée, dont les capteurs mesurent divers comportements internes (températures, vitesses de rotation, vitesse de vent...). Cette hétérogénéité de mesure s'accompagne d'une hétérogénéité de méthodologies de surveillance conditionnelle.

Néanmoins, il ressort de l'analyse de l'état de l'art de la surveillance conditionnelle en application à l'éolien réalisée durant ce chapitre que ces dernières années ont vu émerger un processus dont les choix méthodologiques convergent sur certaines étapes : utilisation de données SCADA pour la construction de modèles de bon fonctionnement pour l'estimation de variables de température, génération de résidus par comparaison entre la valeur prédite par ces modèles et la valeur mesurée, et traitement de ces résidus par application d'un seuil de détection.

Concernant le choix des modèles, quelques auteurs se tournent vers des modèles simples, construits à partir de la connaissance qu'ont les experts du fonctionnement d'un composant, ou vers des modèles de régression linéaire. Ces modèles présentent l'avantage d'être compréhensibles par un utilisateur expert, mais sont souvent conçus pour un composant spécifique et peuvent ne pas exister pour la surveillance d'autres composants où ces connaissances ne sont pas connues. Par conséquent, un grand nombre d'auteurs choisit d'utiliser des méthodes de modélisation plus complexes fondées sur des réseaux de neurone artificiels pour générer à partir de l'exploitation de données SCADA des modèles très précis dans leur prédiction.

Pour obtenir cette qualité de prédiction, ces modèles utilisent souvent un nombre important de variables d'entrée. Ces modèles sont généralement très complexes et se comportent comme des boîtes noires : les relations entre la variable à modéliser et les variables d'entrée ne sont pas connues. Il est donc difficile d'interpréter l'évolution de la variable qu'ils prédisent, notamment pour un opérateur industriel cherchant à développer et optimiser son service de maintenance par l'utilisation d'un processus de surveillance conditionnelle. L'utilisation de ces modèles dans un contexte industriel se heurte ainsi à plusieurs problèmes :

- Le manque de disponibilité des variables d'entrée, due à des problèmes d'acquisition via les capteurs, qui ne permet pas de calculer le résidu pendant de longues périodes.
- La difficulté d'acceptation de la solution par les opérateurs de maintenance qui ne comprennent pas l'origine des alarmes générées.

- Le fait que les modèles sont spécifiques à un composant donné pour une turbine d'une technologie donnée, et ne peuvent donc pas être utilisés tels quels sur d'autres composants ou sur d'autres turbines de la flotte d'éolienne, équipées avec des réseaux de capteur différents. De plus, les modèles proposés par la littérature de surveillance conditionnelle en application à l'éolien se concentrent sur un nombre très réduit de composants, et n'intègrent pas l'ensemble des composants présentant le plus de risque d'être en défauts, occasionnant ainsi les coûts de maintenance les plus élevés.

Nous pouvons donc conclure qu'aucune des solutions proposées dans la littérature ne peut être déployée par une entreprise O&M pour réaliser la surveillance conditionnelle de sa flotte dans l'objectif d'optimiser ses coûts de maintenance.

3.4.1 Positionnement et contributions des travaux de recherche

Dans le cadre des travaux de thèse présentés dans ce manuscrit, notre objectif est ainsi de proposer un système de surveillance conditionnelles pouvant être appliqué à l'ensemble d'une flotte de parcs éoliens en contexte industriel. Conformément au consensus qui se dégage de la littérature, nous adopterons une approche à base de résidus générés par des modèles de bon fonctionnement, appris sur des données SCADA pendant des périodes sans défaut. Nous utiliserons des modèles prédisant l'évolution de la température des composants afin de détecter des défauts à l'aide de résidus thermiques.

Comme nous l'avons démontré par l'analyse de l'état de l'art réalisée, les freins à l'application des processus de surveillance conditionnelle proposés par la littérature réside dans le développement de modèles trop complexes. Ainsi, notre contribution principale est de proposer des modèles simples, linéaires et statiques. Ce choix permet de générer des résidus dont l'évolution est facilement interprétable par les opérateurs : la construction de modèles par régression linéaire permet en effet de réduire la précision de la prédiction en faveur de la compréhension des modèles, comme le met en avant l'analyse comparative de systèmes de surveillance conditionnelle de Schlechtingen [54]. Nous considérons un nombre faible de variables d'entrées afin d'assurer la **disponibilité** et l'**interprétabilité** des modèles. Notre hypothèse est que des modèles simples à peu d'entrées peuvent suffire à prédire l'évolution d'une température si les variables d'entrée sont bien choisies. Par conséquent, nous ne chercherons pas ici à optimiser la précision de la prédiction de la variable de sortie par nos modèles : nous nous contenterons d'une erreur de prédiction plus faible quand le composant surveillé est en bon fonctionnement que lorsqu'il est en défaut.

Pour pouvoir déployer le système de surveillance à une flotte éolienne,

celui-ci doit être applicable sans adaptation de la méthodologie ni modification des performances de détection à tout élément d'une flotte d'éolienne, pour répondre au besoin d'**industrialisation**. En considérant le manque de connaissances sur les dynamiques physiques régissant les comportements des composants d'une éolienne, ainsi que la diversité de modèles à intégrer à ce processus pour répondre à ce critère d'industrialisation, la première partie de notre travail de recherche consistera ainsi à proposer une méthodologie de génération automatique de modèles linéaires et statiques pour tous les composants présentant des coûts de maintenance importants, pour n'importe quelle turbine de la flotte d'éolienne. Cet objectif passe par le développement d'une méthode de sélection automatique de variables d'entrée permettant d'explicitier une variable de sortie. Pour répondre au besoin d'**interprétabilité** de la méthode de détection de défaut, les liens entre ces variables d'entrée des modèle et la température à prédire doivent être compréhensibles par un utilisateur expert. La méthode de sélection doit être guidée par les données et applicable sans conditions sur toute turbine d'une flotte d'éolienne afin de répondre au besoin d'**industrialisation** et à la contrainte de **non homogénéité** des données industrielles.

Les modèles étant simples à générer et très peu complexes, nous améliorerons la méthodologie proposée dans le but de générer un ensemble de modèles linéaires statiques pour réaliser la surveillance d'un même composant. Nous vérifierons que ces modèles permettent de générer des résidus adaptés à la détection de défaut en application industrielle. Pour cela, nous proposerons un cahier des charges des métriques à utiliser pour évaluer les performances d'un système de détection de défaut, et dont le respect assure son acceptabilité par les opérateurs de maintenance. Nous proposerons alors plusieurs stratégies de détection de défaut dont nous comparerons les performances à l'aide de ce cahier des charges : l'une n'utilisant qu'un seul modèle linéaire et statique et l'autre combinant l'information issue de l'ensemble de modèles obtenu.

Nous vérifierons enfin que les modèles appris à partir de données enregistrées sur un parc éolien peuvent être appliqués en conservant les mêmes performances de détection sur des parcs présentant les mêmes caractéristiques technologiques. Cela permettra de répondre au critère d'**ergonomie** défini au chapitre 2 : les mêmes modèles pouvant s'appliquer sur plusieurs parcs, l'analyse des résidus est rendue plus simple pour les opérateurs et la charge mentale n'est pas alourdie.

Ressources : parcs éoliens et bases de données SCADA

Dans ce chapitre, nous allons présenter l'ensemble des ressources qui ont été utilisées pour mener à bien les travaux de recherche. Dans un premier temps, la flotte de parcs éoliens exploitée par VALEMO, au sein de laquelle les études de recherche ont été menées, sera présentée. Nous mettrons alors en lumière la forte hétérogénéité de la flotte de parcs éoliens exploitée par VALEMO, en exposant leurs différentes caractéristiques. Puis nous définirons l'échantillon de parcs utilisé, issu de cette flotte, ainsi que les défauts étudiés. Nous présenterons alors la base de données SCADA utilisée, en détaillant son contenu. Nous identifierons notamment les variables utilisées pour la surveillance de l'état de santé de composants. Ce chapitre sera terminé par une analyse de cette base de données. Nous illustrerons notamment le problème de la qualité de la donnée, ainsi que quelques exemples de relations entre variables, représentant des dynamiques physiques pertinentes pour le cadre de surveillance conditionnelle considéré.

4.1 Description de la flotte d'éoliennes de VALEMO

VALEMO réalise l'exploitation et la maintenance d'une flotte constituée de plus de 70 parcs éoliens, principalement localisés sur la moitié Ouest de la France métropolitaine, mais également sur la France d'outre-mer, avec huit éoliennes installées autour de Sainte-Rose en Guadeloupe.

4.1.1 Caractéristiques des parcs éoliens

L'entreprise propose des contrats d'exploitation et de maintenance à des entreprises qui fabriquent des éoliennes. En conséquence, la flotte exploitée par VALEMO est composée d'un ensemble de turbines très hétérogène, appartenant à huit constructeurs différents. De même que pour un constructeur automobile, chaque constructeur d'éolienne propose plusieurs "gammes" de machine, qui se différencient principalement par leurs performances. Si on considère ce niveau de distinction, alors on dénombre 24 types de machine exploitées et maintenues par VALEMO.

En termes de nombre de turbines et de puissance totale installée, ces parcs sont très différents. Par conséquent, on peut également distinguer les

parcs éoliens composant la flotte de VALEMO selon les enjeux d'optimisation de production : ils ne sont en effet pas les mêmes pour un parc composé de trois éoliennes d'une puissance nominale de 1.5 mégawatts que pour un parc composé de dix éoliennes d'une puissance nominale de 3.6 mégawatts.

Cette hétérogénéité se manifeste par d'autres facteurs :

- **L'âge des parcs** : alors que certains parcs ne produisent de l'énergie que depuis quelques mois, d'autres sont très proches de leur fin de vie (15 ans). Ce facteur a un impact sur l'état de dégradation des composants, et comme nous l'avons vu dans le chapitre 2, sur la politique de maintenance appliquée.
- **La quantité de variables SCADA accessibles** : VALEMO n'est pas propriétaire des données SCADA : elles sont la propriété du fabricant de l'éolienne. Cette spécificité contractuelle explique la contrainte de non-homogénéité définie en section 2.3.3 : pour pouvoir réaliser ses activités de conduite et d'exploitation, l'entreprise VALEMO réalise pour chaque nouvelle "gamme" de turbine exploitée une négociation avec le constructeur sur la quantité de variables qu'elle sera autorisée à utiliser à des fins commerciales.
- **Le choix de composants et capteurs installés**, qui dépend du fournisseur et du stock de l'entreprise réalisant la maintenance.

Cette entreprise n'est pas toujours VALEMO selon le type de maintenance, (comme expliqué en section 2.3.2). Ainsi, selon l'historique d'interventions de maintenance d'un parc, les technologies de composants et de capteurs peuvent changer d'un parc à un autre.

4.1.2 Technologie de turbine

Dans ce manuscrit, on utilisera le terme "technologie de turbine" pour identifier un ensemble d'éoliennes présentant des caractéristiques communes :

- Concrètement, une technologie de turbine est une "gamme" d'éolienne conçue par un fabricant, présentant des caractéristiques de performances propres (par exemple les éoliennes V100 et V126 de Vestas sont deux technologies de turbines différentes, dimensionnées pour une puissance nominale de respectivement 2 MW et 3.45 MW).
- Du point de vue des bases de données associées, une technologie de turbine est un ensemble d'éoliennes présentant un réseau de capteurs SCADA aux caractéristiques (nombre de capteurs, localisation et technologies utilisées) communes. Les bases de données de plusieurs parcs appartenant à la même technologie de turbine possèdent ainsi les mêmes variables.

4.2 Échantillon de parcs éoliens considéré

Les travaux de recherche menés sont appliqués sur une sélection de parcs éoliens de la flotte VALEMO, que l'on nommera "flotte expérimentale" par la suite. L'ensemble de turbines résultant contient quatre cas de défauts sur lesquels les méthodologies de surveillance conditionnelle proposées seront évaluées. Il est à noter que de nombreux défauts ont impacté les différents parcs éoliens retenus sur la période temporelle d'étude, mais seuls ceux intégrant des symptômes thermiques mesurables par des capteurs SCADA installés sur le composant concerné sont retenus.

4.2.1 Flotte expérimentale

La flotte expérimentale utilisée pour l'application et l'évaluation des méthodes de surveillance conditionnelle proposées est composée de 11 parcs éoliens, listés dans le Tableau 4.1.

Nom	N_T	Année	Localisation	Puissance installée
E_f^{rg2}	6	2010	Sud	12MW
E_1	5	2010	Sud	10MW
M_f^{al}	6	2013	Nord-Ouest	12.3MW
M_f^{rm}	5	2015	Nord-Ouest	10.25MW
M_1	5	2010	Nord-Ouest	10MW
M_2	3	2010	Nord-Ouest	6MW
M_3	5	2010	Nord-Ouest	10MW
M_4	6	2016	Nord-Ouest	12MW
M_5	3	2017	Nord-Ouest	6MW
N_f^{rc}	10	2019	Nord	36MW
N_1	4	2021	Nord	9.6MW

Tableau 4.1: Description de l'échantillon de parcs éoliens utilisé, issu de la flotte VALEMO

Cet ensemble de turbine représente une puissance installée totale de plus de 134 MW, soit un peu plus de 10% de la puissance installée de la flotte complète de parcs en exploitation de VALEMO. Chaque parc est désigné par sa technologie, représentée par une lettre majuscule (E,M et N) et par un identifiant, reporté en indice. Les parcs associés à l'identifiant "f" présentent un défaut dont le composant impacté est représenté en exposant : "rc" pour le système de refroidissement du convertisseur, "al" pour le palier de l'arbre lent, "rg2" pour le roulement arrière de la génératrice et "rm" pour le système de refroidissement de la multiplicatrice. Ainsi, M_f^{al} représente un

parc de la technologie de turbine M présentant un défaut sur le palier de l'arbre lent d'une de ses turbines, alors que M_1 est un autre parc de cette même technologie de turbine ne présentant pas de défaut.

Cette flotte expérimentale est composée de 4 parcs présentant un défaut : E_f^{rg2} , M_f^{al} , M_f^{rm} et N_f^{rc} . Tous les autres parcs dont les éoliennes proviennent de la même technologie de turbines (E,M ou N) sont intégrés à cet échantillon. Dans la suite du manuscrit, on désignera par N_T le nombre de turbine d'un parc éolien.

On peut observer que cet échantillon de parcs éoliens est très hétérogène. Les parcs se distinguent selon plusieurs caractéristiques :

- **L'année de début d'exploitation**, qui va de 2010 pour les parcs éoliens en fin de vie à 2021 pour les plus récents.
- La **taille des parcs**, de $N_T = 3$ à $N_T = 10$ turbines.
- La **localisation** : les parcs intégrés à la flotte couvrent la quasi totalité du territoire métropolitain.
- Enfin, la **puissance nominale** des turbines (égale à la puissance installée divisée par N_T), qui va de 2 MW à 3.6 MW pour le parc N_f^{rc} , un des parcs les plus rentables de l'entreprise VALEMO.

Ainsi, cet échantillon est représentatif de l'hétérogénéité de la flotte industrielle de parcs éoliens exploitée par VALEMO. Cette hétérogénéité se traduit dans ce contexte par une diversité de conditions météorologiques, de l'état d'usure naturelle des composants lié à l'âge des parcs ou encore de procédures d'exploitation qui peuvent différer selon le constructeur. Par conséquent, l'analyse des performances de l'outil de surveillance proposé sur cette flotte expérimentale permettra d'évaluer sa capacité d'**industrialisation** (Tableau 2.1).

4.2.2 Défauts étudiés

Parmi l'échantillon de parcs éoliens considéré dans cette étude, quatre parcs présentent une défaillance : E_f^{rg2} , M_f^{rm} , M_f^{al} et N_f^{rc} .

E_f^{rg2} : roulement arrière de la génératrice

Le premier défaut étudié concerne la surchauffe du roulement arrière de la génératrice d'une turbine du parc éolien E_f^{rg2} . Ce défaut s'est développé quelques mois seulement après le début de la mise en production du parc éolien, et a entraîné l'arrêt de l'éolienne pendant près d'un mois pour le remplacement complet de la génératrice.

Ce défaut, également étudié par Lebranchu [27] pour l'application de sa méthodologie de surveillance conditionnelle par analyse de résidus thermiques, a causé les pertes énergétiques et financières les plus lourdes pour l'entreprise parmi les quatre défauts étudiés. Il est apparu durant le mois de Janvier de l'année 2010, et s'est développé jusqu'au 28 Février de la même année, date à laquelle la génératrice est devenue défaillante, et où l'éolienne a été arrêtée pour environ un mois.

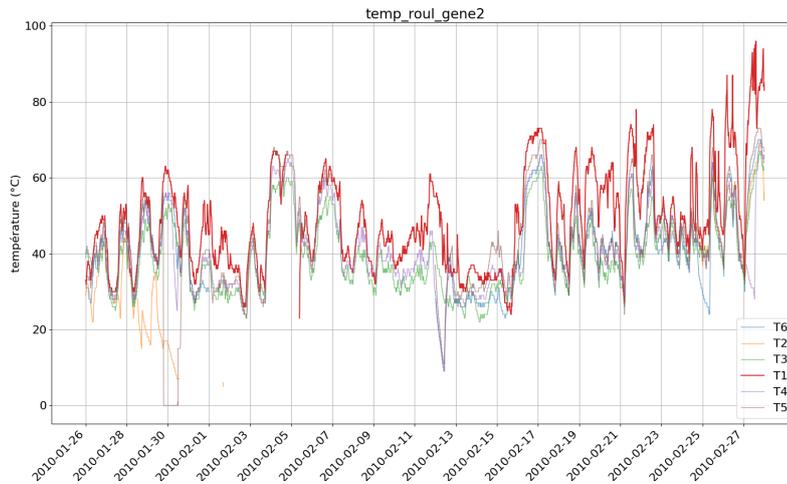


Figure 4.1: Variations de la température du roulement arrière de la génératrice durant le défaut pour les six turbines du parc E_f^{rg2}

La Figure 4.1 illustre les variations de la température du roulement arrière de la génératrice durant la surchauffe du composant pour les six turbines du parc E_f^{rg2} . La courbe épaisse rouge représente les données de la turbine T_1 , impactée par le défaut. On remarque qu'un écart de température est visible avec les autres turbines du parc, mais que cet écart n'est pas constant : d'autres turbines du parc présentent à certaines dates des températures de roulement de génératrice proches, voire supérieures à celles du composant de T_1 . Ce n'est que durant les dernières semaines de la période de défaut que la température du roulement arrière de la génératrice de T_1 atteint des valeurs significativement plus grandes que les autres turbines, alors que le composant est proche de la défaillance.

M_f^{al} : palier de l'arbre lent

Une usure du palier de l'arbre lent (ou roulement principal) a été identifiée sur une turbine du parc éolien M_f^{al} . Cette usure a entraîné la pollution de la graisse du roulement (présence de particules dans le fluide). La progression de cette pollution a entraîné l'échauffement du composant, jusqu'à un stade

de développement où la turbine a dû être arrêtée pendant 10 jours pour remplacement complet de l'arbre lent.

Ce défaut s'est développé sur une longue période de plus d'un an (du début de l'année 2019 au 31 Août 2020, date à laquelle la turbine a été arrêtée), à cause de forts délais d'approvisionnement de composants du côté du fabricant. Ce retard s'est traduit par des surcoûts d'exploitation : pour assurer la production d'électricité dans l'attente des composants, le constructeur a procédé à l'installation et l'utilisation continue d'une pompe de rinçage pour nettoyer la graisse du roulement et ralentir l'usure. Dans le cadre de cette étude, on s'intéressera aux données antérieures à l'installation de cette pompe (jusqu'à Juillet 2019). Ces données correspondent à une période de développement du défaut, non détecté par le fabricant.

M_f^{rm} : refroidissement de la multiplicatrice

Le troisième défaut concerne un échauffement anormal de l'huile et des roulements de la multiplicatrice d'une turbine du parc éolien M_f^{rm} . Ce défaut est apparu à partir de la fin de l'année 2021, jusqu'au 1er Juillet 2022, date à laquelle le liquide de refroidissement de la multiplicatrice a été changé. Il est à noter que contrairement aux deux défauts précédents, ce défaut n'a pas entraîné le remplacement des composants. Il est cependant pertinent de l'inclure dans les travaux de recherche menés car un des objectifs de l'outil de surveillance développé est d'être applicable à tous les composants principaux d'une éolienne.



Figure 4.2: Photos d'une intervention de remplacement d'une multiplicatrice d'éolienne

De plus, comme nous l'avons expliqué dans le chapitre 3, le remplacement d'une multiplicatrice d'éolienne est très coûteux pour une entreprise. En effet, en plus du coût d'achat et du transport du nouveau composant, cette intervention nécessite l'utilisation d'une grue de chantier comme on peut le voir sur la Figure 4.2, affichant des photos prises lors d'une intervention de remplacement d'une multiplicatrice. Il est donc important de vérifier que le système de surveillance conditionnelle proposé permette de détecter de façon précoce les défauts d'une multiplicatrice.

N_f^{rc} : refroidissement du convertisseur

Enfin, une anomalie au niveau du système de refroidissement des convertisseurs du parc N_f^{rc} est étudiée. Cette anomalie est expliquée par une obstruction des grilles des ventilateurs du système de refroidissement du convertisseur, causée par une période de moisson dans le champ agricole sur lequel sont installées les éoliennes du parc. Cette obstruction a été observée sur toutes les éoliennes du parc éolien N_f^{rc} . Dans le cadre des travaux de thèse, seule l'éolienne la plus impactée sera étudiée. Ce défaut est apparu durant l'été 2020, à partir de la fin du mois de Juin, jusqu'à mi-Août, où des opérations de nettoyage des grilles de ventilation et de remplacement des fluides de refroidissement ont été menées.

De nombreux défauts ont impacté les parcs éoliens de l'échantillon présenté dans le Tableau 4.1. Ces quatre cas d'étude ont été retenus car ils présentent des symptômes thermiques identifiés à un stade de développement du défaut non critique (la production de l'éolienne est toujours possible). Ils rentrent ainsi dans le cadre de détection de l'outil de surveillance conditionnelle par analyse de résidus thermiques considéré pour les travaux de recherche présentés. De plus, leur non prise en charge par une stratégie de maintenance préventive conditionnelle adaptée a eu pour conséquences des pertes énergétiques et économiques relativement conséquentes. Étudier les performances de l'application de l'outil de surveillance conditionnelle sur ces défauts est donc pertinent pour un industriel, qui pourrait y voir un gain financier et énergétique significatif.

4.3 Base de données SCADA

4.3.1 Processus d'acquisition des données

Les travaux de recherche présentés sont basés sur des données acquises par le système de contrôle et d'acquisition de données SCADA, implanté sur un automate installé en nacelle. Par simplification, on parlera par la suite de données SCADA. L'utilisation de ce type de données est économique, car les capteurs les générant sont pour la plupart déjà installés sur les éoliennes. Les

données SCADA présentent également l'avantage de mesurer les dynamiques de l'ensemble des composants d'une nacelle d'éolienne dont la maintenance est la plus coûteuse, contrairement à d'autres types de données dont les capteurs ne peuvent être installés que sur un nombre très réduit de composant comme expliqué dans le chapitre 3.

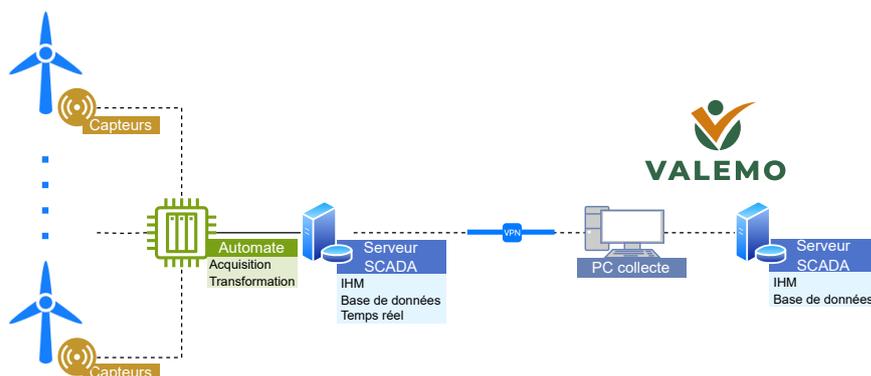


Figure 4.3: Visualisation du processus de collecte de données SCADA des capteurs installés à la base de données de l'entreprise VALEMO

La Figure 4.3 illustre le processus de collecte des données SCADA pour un parc éolien dans le cas de l'entreprise VALEMO. Les mesures issues des capteurs installés sur les éoliennes sont tout d'abord récupérées sur un automate toutes les 40 milli-secondes, puis elles sont transformées : l'automate calcule toutes les 10 minutes la moyenne des données SCADA générées durant les 10 dernières minutes (ce qui représente 15000 échantillons). Ces nouvelles données sont ensuite transférées au serveur SCADA, qui les stocke en temps réel dans une base de données, manipulable par un utilisateur expert via une interface homme-machine (ou IHM). Enfin, les données de ce serveur sont transférées par un VPN au PC de collecte de données localisé dans l'entreprise VALEMO, avant d'être stockées dans le serveur SCADA de l'entreprise. Il est à noter que le processus SCADA intègre également une supervision rudimentaire des composants instrumentés, prenant la forme d'un système de levée d'alarmes automatisée, fournissant des informations aux opérateurs sur l'état des composants en temps réel.

4.3.2 Contenu de la base de données

VALEMO n'est pas propriétaire de l'intégralité des données SCADA des parcs éoliens exploités : la quantité de variables à disposition dépend de la technologie de turbine considérée. Les négociations réalisées entre VALEMO et le constructeur déterminent cette quantité, qui peut aller d'une dizaine à plus de quarante variables. Il est commun d'avoir à disposition pour tous

les parcs de la flotte VALEMO les catégories de variables SCADA suivantes pour un suivi de production efficace [70, 3] :

- **Électriques** : la puissance active et réactive ainsi que les courants et tensions de la génératrice
- **Environnementales** : les différentes mesures de la vitesse du vent par anémomètre, la température ambiante
- **Thermiques** : les températures des roulements et systèmes de refroidissement des composants tournants principaux (multiplicatrice, génératrice, rotor), ainsi que d'autres composants de l'éolienne, comme le convertisseur et le transformateur
- **Contrôle** : les vitesses de l'arbre lent et de l'arbre rapide de la multiplicatrice, les angles d'orientation des pales (pitch) et du mât de la nacelle (yaw)

Parmi toutes ces catégories de variables, les variables de température sont les plus présentes au sein des bases de données SCADA. Cela s'explique par le fait que les capteurs associés (sondes PT-100) sont simples à installer et parmi les moins chers du marché. De plus, l'échantillonnage à 10 minutes est plus adapté aux variables de température, dont les variations s'opèrent sur un horizon temporel bien plus long que les variables électriques et mécaniques.

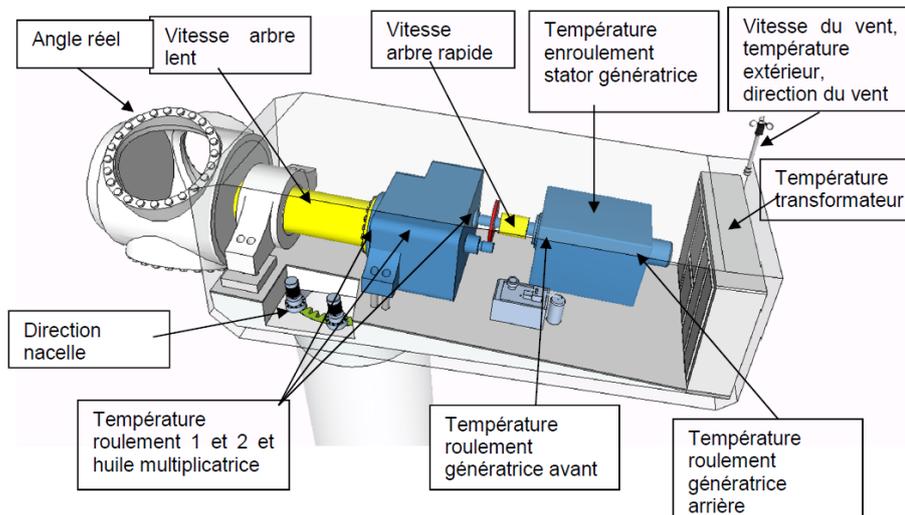


Figure 4.3 Position des capteurs utilisées

Figure 4.4: Localisation des capteurs SCADA dans une nacelle d'éolienne [26]

La Figure 4.4 illustre la localisation des capteurs SCADA dans une nacelle d'éolienne, et la liste complète des variables SCADA utilisées dans cette étude est détaillée en annexe dans le Tableau A.1.

4.3.3 Variables utilisées comme indicateur de défaut des composants

Le processus de surveillance conditionnelle proposé est fondé sur des résidus thermiques, construits à partir de la modélisation de variables de températures remplissant une fonction d'indicateur de défaut. Lors de l'utilisation de données SCADA, la température du fluide de refroidissement est une variable à privilégier pour modéliser l'état d'un composant, car c'est celle dont le comportement est le plus sensible à un défaut aux symptômes thermiques. Par exemple, pour la surveillance de la multiplicatrice, la température de l'huile de ce composant est utilisée dans la littérature, comme l'illustre la revue de Bie [4]. Cependant, le contenu de la base de données SCADA exploitable par VALEMO étant déterminé par un commun accord entre le fabricant d'éolienne et l'entreprise O&M, ces variables ne sont pas toujours disponibles. Le cas échéant, les variables de température dont les capteurs sont localisés au plus proche de la source d'effort (et donc de chaleur) du composant sont alors choisies comme indicateurs de défaut.

Dans le cadre des travaux de thèses présents, nous utilisons ainsi huit variables de température qui permettront de surveiller la présence de défaillance sur les composants associés. Ces variables, que l'on nommera variables de sortie, ainsi que leur composant ou système associé, sont définies dans le Tableau 4.2.

Composant	Variable de sortie	Description
Moyeu rotor	t_{hub}	température du moyeu du rotor, ou "hub"
Roulement principal	$t_{paliArbrLent}$	température du roulement de l'arbre lent
Refroidissement de la multiplicatrice	$t_{huileMulti}$	température de l'huile de la multiplicatrice
Roulement avant de la génératrice	$t_{roulGene1}$	température du roulement arrière de la génératrice
Roulement arrière de la génératrice	$t_{roulGene2}$	température du roulement avant de la génératrice
Refroidissement de la génératrice	$t_{generatrice}$	température du liquide de refroidissement de la génératrice
Refroidissement du convertisseur	$t_{convertisseur}$	température du liquide de refroidissement du convertisseur
Bobine transformateur	$t_{transformateur}$	température de la bobine du transformateur

Tableau 4.2: Présentation des huit variables de sortie utilisées pour la surveillance de l'état de santé des composants

Les défauts étudiés et présentés en section 4.2.2 sont ainsi surveillés à partir de la température du roulement arrière de la génératrice ($t_{roulGene2}$) pour le cas du parc E_f^{rg2} , de la température du palier de l'arbre lent ($t_{paliArbrLent}$) pour le cas du parc M_f^{al} , de la température de l'huile de la multiplicatrice ($t_{huileMulti}$) pour le cas du parc M_f^{rm} et de la température du fluide de refroidissement du convertisseur ($t_{convertisseur}$) pour le cas du parc N_f^{rc} .

Le Tableau 4.3 expose la présence de ces variables de sortie parmi les bases de données associées aux technologies de turbine intégrées dans l'échanti-

llon de parcs éolien considéré (Tableau 4.1).

Composant	Variable de sortie	E	M	N
Moyeu rotor	t_{hub}	Non	Oui	Oui
Roulement principal	$t_{\text{paliArbrLent}}$	Non	Oui	Oui
Refroidissement de la multiplicatrice	$t_{\text{huilMulti}}$	Oui	Oui	Oui
Refroidissement de la génératrice	$t_{\text{generatrice}}$	Non	Non	Oui
Roulement avant de la génératrice	$t_{\text{roulGene1}}$	Oui	Oui	Oui
Roulement arrière de la génératrice	$t_{\text{roulGene2}}$	Oui	Oui	Oui
Refroidissement du convertisseur	$t_{\text{converter}}$	Non	Oui	Oui
Bobine transformateur	t_{transfo}	Oui	Non	Non

Tableau 4.3: Disponibilité des variables de sortie associées à chaque composant pour les trois technologies de turbine présentes dans l'échantillon de parcs éoliens utilisé (Tableau 4.1)

On remarque que les indicateurs de défaut ne sont pas tous disponibles pour les trois technologies de turbine étudiées. On peut cependant observer que la technologie de turbine N présente le plus fort potentiel de surveillance conditionnelle, avec sept composants pouvant être surveillés à l'aide d'une variable de sortie choisie comme indicateur de défaut. Cette constatation est intéressante car le parc N_f^c , appartenant à cette technologie de turbine, fait partie des parcs les plus performants et rentables exploités par VALEMO, avec une puissance installée totalisant 36MW d'après le Tableau 4.1.

4.4 Analyse de la base de données SCADA

Nous venons de présenter la base de données SCADA utilisée pour développer le processus de surveillance conditionnelle. Par une analyse de cette base de données, nous allons définir dans cette section les contraintes auxquelles la méthode de détection développée devra être robuste afin d'assurer son déploiement industriel, et présenter quelques exemples de relation entre variables illustrant les possibilités de modélisation en bon fonctionnement par l'utilisation de cette base de données.

4.4.1 Qualité de la donnée

Assez fréquemment lors de l'utilisation de données industrielles, les données ne sont pas enregistrées par le système d'acquisition du fait de la perte de communication entre le serveur et les postes de supervision de l'entreprise. Ces perturbations peuvent être présentes sur l'ensemble du réseau de capteurs d'une turbine, par exemple dans le cas où une maintenance informatique est en cours sur une ou plusieurs éoliennes d'un parc.

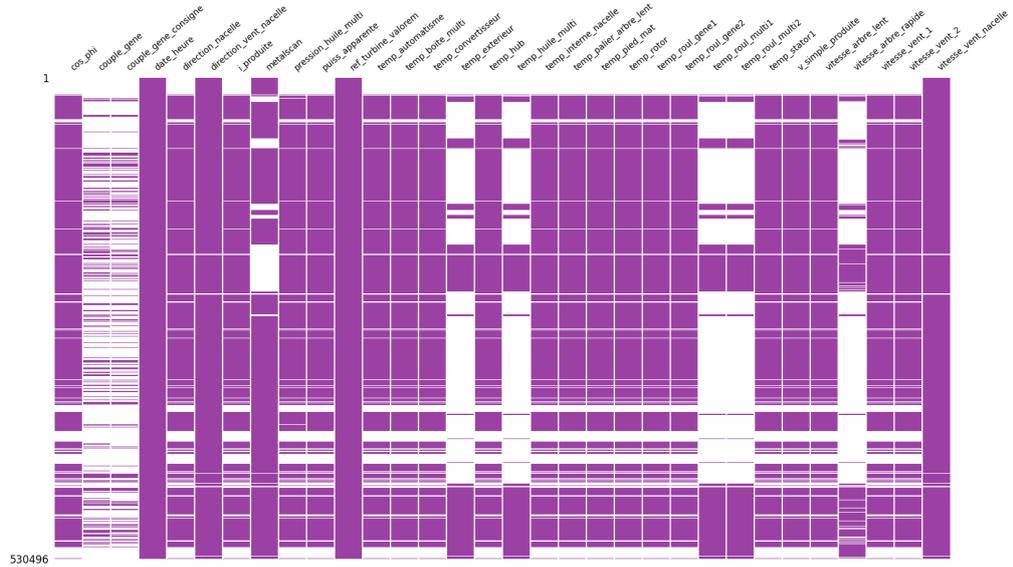


Figure 4.5: Illustration de la disponibilité des variables SCADA du parc M_f^{al} durant une année

Pour illustration, la Figure 4.5 retrace la disponibilité au cours d'une année des variables SCADA du parc M_f^{al} . Chaque colonne représente une variable de la base de données du parc, et l'axe vertical représente toutes les dates de l'année, à un pas de temps de 10 minutes. Une barre violette est dessinée pour chaque date où la variable est disponible. On remarque ainsi que toutes les variables présentent au moins une date d'indisponibilité. Particulièrement, les variables "*temp_exterieur*", "*temp_huile_multi*", "*temp_roul_multi1*", "*temp_roul_multi2*", "*vitesse_arbre_rapide*", "*couple_gene*" et "*couple_gene_consigne*" présentent des pertes de données conséquentes, et sont indisponibles la majeure partie de l'année étudiée. Par conséquent, la surveillance du système de refroidissement de la multiplicatrice ($y = t_{\text{huileMulti}}$) ne sera possible que sur une courte portion de l'année étudiée.

En outre, les capteurs peuvent fournir des valeurs erronées ou fortement bruitées à cause d'une mauvaise fixation ou calibration des capteurs. Par conséquent, la donnée mesurée n'est pas toujours fiable.

Ces perturbations, ainsi que la relative disponibilité de la donnée SCADA, représentent la contrainte de **qualité de la donnée** (Tableau 2.2).

4.4.2 Relation entre variables

La base de donnée SCADA utilisée dans ces travaux de recherche contient un ensemble de signaux très variés, représentant les dynamiques physiques

(thermique, électrique, mécanique...) des composants intégrés dans une nacelle d'éolienne.

Les composants d'une éolienne forment une chaîne cinématique, présentée dans le chapitre 2, et présentent donc une forte interactivité. Cette interactivité se traduit par de nombreuses relations de nature différente entre les différentes variables de la base de données. Dans cette section, nous allons illustrer ces relations, et en analyser quelques-unes. Cette analyse permettra notamment de comprendre comment des variables SCADA peuvent être utilisées pour expliquer les comportements thermiques des variables de sortie (Tableau 4.3).

Matrice de corrélation

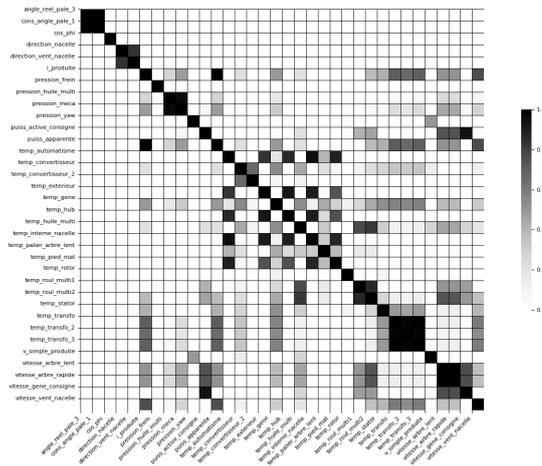


Figure 4.6: Matrice de corrélation (R^2) des variables de la base de données SCADA pour des données de bon fonctionnement du parc N_f^{rc}

Parmi les données SCADA, on trouve des relations linéaires. Ces relations peuvent illustrer des liens de similarité entre variables de même nature. La base de données SCADA peut également contenir des liens de causalité (le changement de comportement d'une variable implique un changement de comportement sur une autre variable), qui s'illustrent par des relations plus complexes, pouvant être non-linéaires.

La Figure 4.6 est une matrice de corrélation qui illustre l'ensemble des relations présentes au sein de la base de données du parc N_f^{rc} , calculées à partir de données enregistrées pendant une période ne présentant pas de défaut. La métrique utilisée est le coefficient de corrélation de Pearson (R^2) entre deux variables de la base de données SCADA. Son expression est donnée par l'équation 4.1, où x_i et y_i représentent des valeurs des variables

X et Y , et \bar{x} et \bar{y} les valeurs moyennes de ces variables. Il est égal à 1 (carré noir sur la Figure 4.6) pour deux variables exactement proportionnelles.

$$R^2 = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}} \quad (4.1)$$

On peut observer qu'il existe des corrélations entre variables de même nature, illustrant des liens de similarité. C'est notamment le cas des variables de température. Toutefois, on observe également des corrélations entre variables de natures différentes : si on se base sur la ligne associée à la puissance apparente, on identifie en effet une forte corrélation avec l'intensité produite par la génératrice et la vitesse du vent par exemple. Nous allons analyser quelques-unes de ces relations, intégrant des variables de sortie du Tableau 4.3.

Relations de similarité

Parmi toutes les relations entre variables observées sur la Figure 4.6, prenons l'exemple de la relation entre la température de l'huile de la multiplicatrice et la température du roulement 1 de la multiplicatrice, illustrée dans la Figure 4.7.

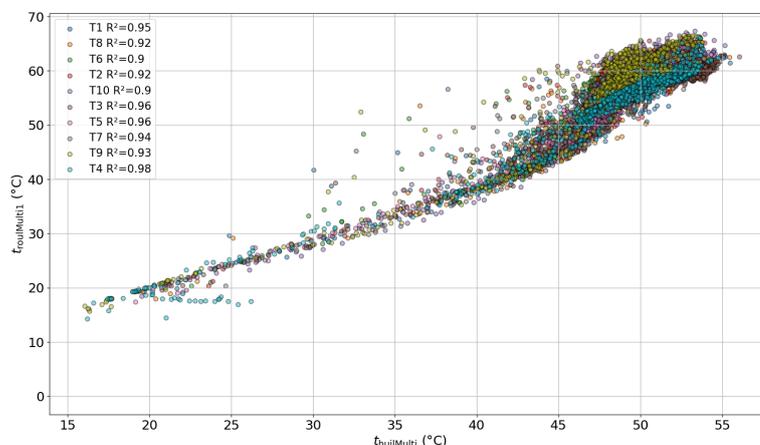


Figure 4.7: Relation entre la température de l'huile de la multiplicatrice et la température du roulement avant de la multiplicatrice

On observe une relation linéaire entre ces deux variables, qui s'explique simplement : les roulements de la multiplicatrice sont partiellement immergés dans l'huile. Ainsi, lorsqu'ils s'échauffent, l'huile de la multiplicatrice s'échauffe par conduction thermique. On peut ainsi expliquer le comportement de l'huile de la multiplicatrice (dont la température fait partie des

variables de sortie listées dans le Tableau 4.2) à partir du roulement avant de la multiplicatrice.

Ce cas est pris pour exemple car il est relativement simple à comprendre, et concerne une des variables de sortie considérée. Toutefois, comme nous pouvons le voir dans la matrice de corrélation de la Figure 4.6, il existe d'autres liens de similarité illustrés par des relations linéaires. La plupart d'entre elles s'expliquent par une proximité des composants au sein de la nacelle : l'échauffement d'un composant implique ainsi l'échauffement des composants voisins par transfert de chaleur.

Relations de causalité

Une relation de causalité se manifeste lorsque les variations d'une variable impliquent un changement dans les variations d'une autre variable. Un exemple classique de relation de causalité est celle liant la vitesse du vent avec la puissance produite par l'éolienne : comme nous l'avons vu dans le chapitre 2 (Figure 2.2), selon la plage de vitesse de vent mesurée par un anémomètre d'une éolienne, l'évolution de la production de puissance ne sera pas la même.

Parmi les liens mis en évidence par la Figure 4.6, nous pouvons observer des liens entre des variables qui présentent des relations de causalité. Pour exemple, la Figure 4.8 illustre la relation entre la température du convertisseur (choisie dans cette étude comme indicateur de défaut de ce composant) et l'intensité du courant électrique produit par la génératrice.

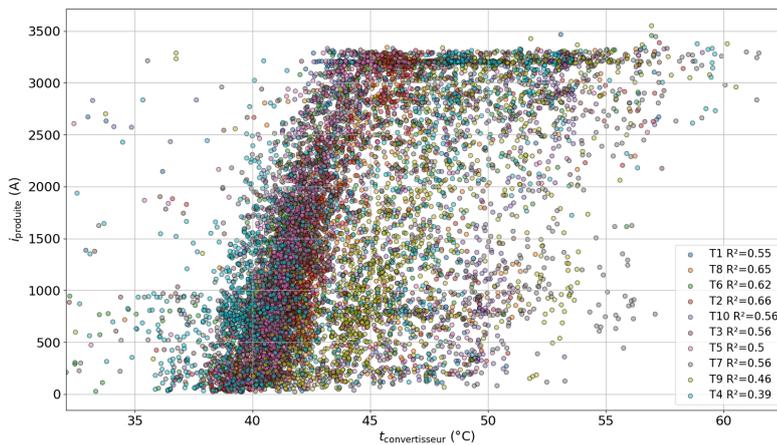


Figure 4.8: Relation entre l'intensité du courant électrique produit par la génératrice et la température du système de refroidissement du convertisseur

L'intensité produite par la génératrice est proportionnelle à la produc-

tion de puissance, et cette intensité est une entrée directe du convertisseur. Ainsi, plus la génératrice est sollicitée, plus l'intensité du courant produit est grande, et plus le convertisseur, qui reçoit ce courant, s'échauffe par effet Joule (augmentation de la température d'un conducteur causé par l'augmentation de son énergie). Ce phénomène physique explique la manifestation d'une relation entre ces deux variables.

La Figure 4.9 illustre les variations de l'indicateur de défaut $t_{\text{roulGene2}}$, ainsi que de la température du stator de la génératrice t_{stator} et de la puissance apparente $p_{\text{apparente}}$ durant le défaut présenté en section 4.2.2 pour les données des six turbines du parc E_f^{rg2} .

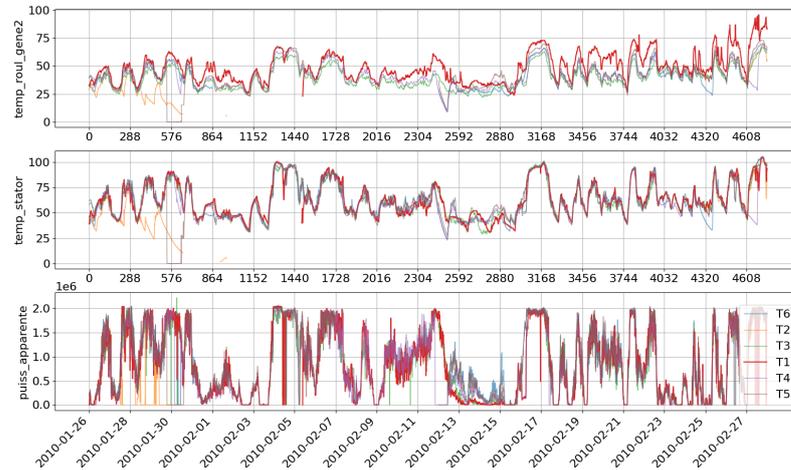


Figure 4.9: Variations de $t_{\text{roulGene2}}$, t_{stator} et $p_{\text{apparente}}$ durant le défaut pour les six turbines du parc E_f^{rg2}

On remarque que, contrairement à $t_{\text{roulGene2}}$, les variables t_{stator} (présentant un lien de similarité avec y) et $p_{\text{apparente}}$ (présentant un lien de causalité avec y), ne sont pas impactées par le défaut, et conservent des variations proches des autres turbines du parc, représentatives d'un fonctionnement normal. Ainsi, en modélisant y à l'aide de ces deux variables, on obtient une bonne précision de prédiction en période de fonctionnement normal car t_{stator} et $p_{\text{apparente}}$ possèdent un lien avec $t_{\text{roulGene2}}$, mais cette précision se dégrade en période de défaut. Par conséquent, l'erreur de prédiction augmente lorsque le défaut apparaît, mais reste faible en période de bon fonctionnement. Ce cas d'étude permet d'illustrer le principe de modèles de bon fonctionnement présenté durant le chapitre 3. Pour conclure, rappelons qu'il est important d'intégrer ces deux types de relations (similarité et causalité) dans les modélisations thermiques, en vue de proposer une solu-

tion reposant sur des modèles physiquement cohérents et donc interprétables par un opérateur expert.

4.5 Conclusion

Nous avons débuté ce chapitre en présentant la flotte d'éoliennes exploitée par l'entreprise VALEMO, en soulignant sa forte hétérogénéité (technologies de turbine considérées, âge des parcs, contenu des bases de données, localisation...).

Nous avons par la suite présenté l'échantillon de parcs éoliens considéré dans le cadre des travaux de recherche menés. Cet échantillon est composé de 11 parcs répartis sur trois technologies de turbine différentes. Quatre de ces parcs présentent un défaut sur des composants distincts. Ces défauts présentent des symptômes thermiques, et leur prise en charge par une maintenance préventive aurait permis une réduction des coûts énergétiques et financiers significative pour l'entreprise.

Les travaux de thèse se basent sur des données basse-fréquence (10 minutes) SCADA exploitées par VALEMO. Ces données sont issues de capteurs localisés sur différents composants d'une éolienne (rotor, arbre lent et rapide, multiplicatrice, génératrice, convertisseur et transformateur). Les bases de données résultantes, établies par négociation avec les constructeurs des turbines, mesurent des dynamiques thermiques, mécaniques, électriques ou encore environnementales. Nous avons ainsi pu définir les variables de température utilisées pour la surveillance de l'état de santé de composants d'une éolienne.

Dans une dernière partie, nous avons mis en évidence la contrainte de la qualité dégradée de la donnée industrielle sur la base d'un exemple concret. Puis nous nous sommes intéressés aux relations linéaires et non-linéaires qui composent la base de données utilisée, et qui permettent de comprendre comment une variable de température peut être expliquée à l'aide de données d'autres variables SCADA en période de bon fonctionnement.

Génération automatique de modèles pour la détection de défauts par analyse de résidus

Dans un premier temps, nous présenterons l'architecture de surveillance conditionnelle utilisée, en détaillant les différentes phases hors-ligne et en-ligne du processus de l'extraction des données à la génération d'alarmes par analyse de résidus. Nous présenterons ensuite en détail le processus de sélection automatique de variables d'entrée permettant la génération automatique de modèles linéaires statiques. Nous proposerons alors un cahier des charges des métriques à utiliser pour évaluer les performances en détection d'un système de détection de défaut acceptable par les opérateur de maintenance et nous définirons ces métriques. Nous évaluerons finalement notre système de surveillance en évaluant le respect du cahier des charges défini, par une analyse de ses métriques sur les quatre défauts présentés au chapitre précédent.

5.1 Élaboration du système de surveillance conditionnelle

Le système de surveillance proposé s'appuie sur l'analyse de résidus thermiques générés à partir de modèles bon fonctionnement. Nous utiliserons la formulation de résidus "multi-turbine" proposée par Lebranchu et al. [27], car elle permet de limiter l'effet des facteurs environnementaux sur les résidus. Pour répondre aux besoins et contraintes de déploiement industriel identifiés à l'issue du chapitre 2, des choix restrictifs sur les modèles de bon fonctionnement utilisés sont réalisés : afin de prédire les variables de sortie, notre contribution principale est de considérer des modèles linéaires statiques à trois variables. Nous considérons un nombre restreint de variables d'entrée pour prendre en compte la **relative qualité** des données SCADA et limiter les risques d'indisponibilité des modèles, et pour conserver des modèles simples, **interprétables** par des utilisateurs experts. Pour des questions

d'**interprétabilité** et d'**ergonomie** du système de surveillance conditionnelle proposé, nous chercherons des ensembles de variables d'entrée pouvant s'appliquer à toutes les turbines d'un parc éolien. Notre hypothèse est que des modèles simples à peu d'entrées peuvent suffire à prédire l'évolution d'une température si les variables d'entrée sont bien choisies. Nous ne cherchons pas ici à obtenir des modèles qui prédisent avec précision l'évolution de la variable de sortie. Il suffit que l'erreur de prédiction de ces modèles quand le système est en bon fonctionnement soit plus faible que lorsque le système est en défaut.

5.1.1 Extraction et pré-traitement des données

Pour un parc éolien donné, les données du processus sont l'ensemble des signaux, ou variables, acquis par le système de supervision SCADA. La base de données associée, présentée dans le chapitre 4, contient un ensemble de mesures issues de capteurs illustrant différentes variations (thermiques, mécaniques, électriques...) des composants d'une éolienne de ce parc.

Dans cette étude, pour des données provenant de la base de données SCADA d'un parc, le pré-traitement est réalisé en plusieurs étapes:

1. Premièrement, les variables dont les valeurs varient très peu sont supprimées de l'analyse (par exemple la fréquence du réseau, autour de 50 Hertz), car ils ne contiennent pas d'informations utiles.
2. Puis, les échelles de valeurs sont uniformisées. Cela concerne les variables angulaires principalement, qui, selon la configuration des capteurs, peuvent mesurer des angles de 0 à 360°, ou de -180 à 180°. On choisit dans cette étude de considérer des valeurs angulaires allant de -180 à 180°.
3. Ensuite, des bornes de valeur basse et haute sont appliquées sur l'ensemble des variables restantes. Ces bornes sont déterminées pour chaque technologie de turbine à partir de connaissances expertes sur les amplitudes de valeurs physique de chaque dynamique de composant. Par exemple, les valeurs de températures "normales" du roulement principal sont comprises entre 0°C et 100°C. Comme l'a mis en avant Marti-Puig [37], choisir ces valeurs de filtrage manuellement permet une bonne identification et suppression des valeurs aberrantes sans compromettre significativement la quantité d'information utile disponible.

Cette méthode de pré-traitement des données SCADA permet d'identifier et d'écartier la plupart des données nuisibles à l'application du processus de surveillance conditionnelle, notamment durant la phase d'apprentissage des modèles. En moyenne, sur une période de données SCADA de un an, moins de 5% des valeurs sont supprimées de l'analyse par application de cette méthode.

5.1.2 Génération du modèle

Nous considérons dans cette étude la modélisation de variables de température, à partir de données provenant de variables présentant des liens avec cette variable de sortie. Le contexte applicatif industriel implique une faible disponibilité de données historiques liées à des défauts. Le système de surveillance conditionnelle inclut ainsi un processus hors-ligne utilisant une période de données de durée 1 an, où le parc éolien surveillé est en fonctionnement normal, selon le principe Normal Behavior Modeling (NBM) [30]. Nous désignerons par H_L cette période de donnée, utilisée pour réaliser le processus de surveillance hors-ligne, comprenant l'apprentissage des modèles et le réglage des seuils de détection.

Sélection automatique des régresseurs

Afin de prédire les variables de sortie, des modèles linéaires statiques à trois variables sont utilisés. Les variables X composant ces modèles, que nous appellerons par la suite régresseurs, sont automatiquement sélectionnées selon un processus qui combine une régression pénalisée avec un algorithme de sélection de variables itératif qui utilise un critère de sélection prenant en compte les données de toutes les turbines du parc éolien, garantissant ainsi la généralisation des régresseurs X à toutes les turbines du parc éolien considéré.

Apprentissage des modèles

Une fois les régresseurs déterminés par le processus de sélection automatique développé, les paramètres (coefficients de régression) du modèle sont appris pour chaque éolienne du parc par application de la méthode des moindres carrés sur les données d'apprentissage.

Il est à noter que les variables SCADA sont standardisées à l'aide de leur moyenne et de leur écart-type calculés sur les données de la période H_L , afin d'assurer la convergence de l'algorithme de régression utilisé.

On considère des modèles linéaires pour limiter le risque de surapprentissage, et assure l'interprétabilité des modèles par un utilisateur expert.

L'équation 5.1 décrit le type de modèle considéré, pour un instant t , où \hat{y}^{T_i} est la variable de température modélisée pour les données de la turbine T_i , (a, b, c) les coefficients de régressions, et $x_1(t)$, $x_2(t)$, $x_3(t)$ les valeurs à l'instant t des régresseurs pour le composant X .

$$\hat{y}^{T_i}(t) = a * x_1^{T_i}(t) + b * x_2^{T_i}(t) + c * x_3^{T_i}(t) \quad a, b, c \in \mathbf{R} \quad (5.1)$$

Période d'apprentissage

On considère qu'un parc est en fonctionnement normal sur une période s'il ne présente pas de maintenance corrective (remplacement de composant) sur cette période. Précisons que, dans un cadre applicatif industriel, il peut y avoir des anomalies sur divers composants et capteurs d'une éolienne pendant une période de bon fonctionnement. Ces perturbations ne sont pas toujours détectées par les systèmes de levée d'alarmes en place au sein des entreprises O&M, et impactent les variations des données SCADA (décalage de la moyenne pour une mauvaise fixation ou calibration de capteur, saut de valeurs pour un choc externe...).

On suppose donc que les anomalies survenant pendant la période d'apprentissage H_L sont suffisamment mineures pour n'avoir aucun impact significatif sur la qualité de la modélisation.

5.1.3 Construction et traitement de l'indicateur multi-turbine

Une fois le modèle de bon fonctionnement généré et ses coefficients paramétrés pour toutes les turbines du parc éolien considéré, les résidus peuvent être construits. Ces résidus sont basés sur la formule résiduelle "multi-turbine" développée par Lebranchu en 2019 [27] pour la détection précoce d'une surchauffe de génératrice d'une éolienne.

La valeur résiduelle à un instant t pour une turbine T_i , $err_y^{T_i}(t)$, est définie selon l'équation 5.2 comme la différence entre la valeur modélisée $\hat{y}^{T_i}(t)$ (équation 5.1) et la valeur mesurée par le SCADA $y^{T_i}(t)$. Pour une turbine donnée T_i , le résidu multi-turbine $\varepsilon_y^{T_i}(t)$ mesure alors l'écart entre $err_y^{T_i}(t)$ et la valeur résiduelle médiane du parc éolien à ce même instant. L'équation 5.3 décrit le calcul du résidu résultant $\varepsilon_y^{T_i}(t)$, que l'on nommera par la suite "indicateur multi-turbine".

$$err_y^{T_i}(t) = \hat{y}^{T_i}(t) - y^{T_i}(t) \quad (5.2)$$

$$\varepsilon_y^{T_i}(t) = err_y^{T_i}(t) - \text{médiane}_{j \in [1; N_T]}(err_y^{T_j}(t)) \quad (5.3)$$

$\varepsilon_y^{T_i}(t)$ est alors l'indicateur multi-turbine calculé à l'instant t , associé à la variable de sortie y pour la turbine T_i , avec N_T le nombre de turbines du parc éolien considéré. La Figure 5.1 illustre le schéma de construction de l'indicateur multi-turbine.

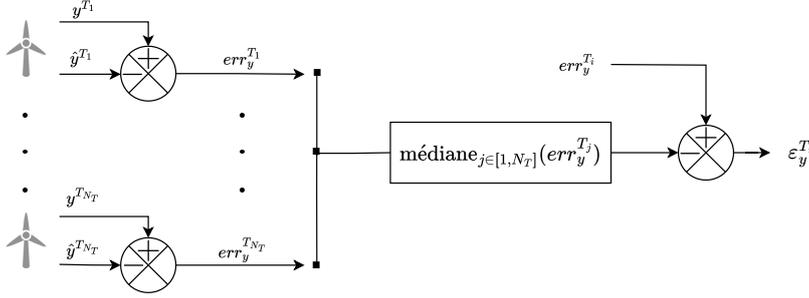


Figure 5.1: Schéma de construction de l'indicateur multi-turbine développé par Lebranchu [27]

Schlechtingen a mis en évidence en 2011 [54] la sensibilité aux saisonnalités des résidus dont les modèles sont construits par régression linéaire, pouvant rendre la détection de défaut difficile. La correction à la valeur résiduelle médiane du parc permet une robustesse à ces variations environnementales ainsi qu'aux modes de production, qui sont également un facteur de sensibilité selon Lebranchu [26], sans augmenter la complexité des modèles.

Dans le cadre d'une application à l'échelle d'une flotte d'éoliennes, beaucoup de modélisations sont nécessaires. Par conséquent, beaucoup de résidus doivent être calculés. Ainsi, l'indicateur multi-turbine est standardisé à l'aide de la valeur moyenne et de son écart-type calculés sur la période d'apprentissage H_L pour faciliter l'analyse comparative des résultats. Enfin, pour traiter les éventuelles valeurs aberrantes restantes, pouvant être causées par les perturbations des données SCADA non traitées par l'application du pré-traitement, les valeurs de l'indicateur multi-turbine sont lissées à l'aide d'une moyenne glissante.

$$\overline{\varepsilon_y^{T_i}(t)} = \frac{1}{f} * \sum_{l=t-f}^t \frac{\varepsilon_y^{T_i}(l) - \mu_{H_L}}{\sigma_{H_L}} \quad (5.4)$$

Le traitement (standardisation, puis filtrage) des données résiduelles de l'indicateur multi-turbine est résumé par l'équation 5.4, où μ_{H_L} et σ_{H_L} sont respectivement la moyenne et l'écart-type de l'indicateur $\varepsilon_y^{T_i}$ calculés sur les données hors-ligne de la période d'apprentissage H_L , et où f est la fenêtre de calcul de la moyenne glissante, égale à 144. D'après l'échantillonnage des données SCADA (10 minutes), la moyenne glissante est ainsi calculée sur un horizon d'une journée de données ($144 * 10 = 60 * 24$ minutes). Par simplification, on utilisera par la suite la notation $\varepsilon_y^{T_i}$ pour désigner l'indicateur multi-turbine standardisé et filtré.

5.1.4 Réglage du seuil de détection

Une fois les résidus générés sur la période d'apprentissage, il reste à fixer le seuil de détection dont le franchissement signale la présence d'un défaut. Rappelons que dans le cas de la maintenance d'une flotte éolienne, le coût de la **main d'oeuvre** est très élevé et qu'il est important de ne pas envoyer sur site des techniciens de maintenance alors que la turbine fonctionne normalement. Ainsi, il est important de fixer des valeurs de seuil de détection permettant de limiter les alertes fournies par le système de surveillance conditionnelle aux défauts avérés.

Définition d'une intervention inutile

On considère qu'un dépassement du seuil de détection par le résidu de manière continue pendant au moins une journée traduit la présence d'un défaut, et qu'il est nécessaire de mettre en place une intervention de maintenance préventive pour réaliser un diagnostic sur le composant ou système surveillé. Nous nommons donc "intervention" une période de temps pendant laquelle les valeurs de l'indicateur multi-turbine $\varepsilon_y^{T_i}(t)$ sont supérieures au seuil de détection pendant au moins une journée (sans compter les dates d'indisponibilité de l'indicateur), ce qui déclenchera une intervention de l'équipe de maintenance. Il est à noter que nous faisons ainsi l'hypothèse que les utilisateurs expert font confiance à l'outil de surveillance conditionnelle développé, et ne réalisent pas d'analyses complémentaires de l'état de santé des composants étudiés pour déterminer la nécessité d'intervenir sur le terrain.

Quand une intervention apparaît pendant une période où il n'y a pas de défaut sur le composant surveillé, elle aura comme conséquence un déplacement d'une équipe de techniciens. On nommera donc une intervention qui apparaît pendant une période sans défaut "intervention inutile" car l'équipe de technicien se déplace inutilement. Ainsi, toute intervention détectée pendant la période d'apprentissage H_L peut être considérée comme intervention inutile, car il s'agit par hypothèse d'une période sans défaut. Il est à noter qu'une intervention inutile est différente d'une fausse alarme, laquelle correspond au franchissement d'un seuil de détection à un instant donné t , alors que l'intervention inutile prend en compte la persistance dans le temps de ce franchissement.

Valeur du seuil de détection

Traditionnellement, le choix de la valeur du seuil de détection est réalisé sur un critère statistique, comme par exemple le quantile [56, 74] : la valeur du seuil de détection est déterminée de manière à ce qu'un pourcentage des valeurs du résidu soit supérieur au seuil pendant une période de bon fonctionnement. Cette méthode revient ainsi à fixer le taux de fausses alarmes

attendu. Dans cette étude, un critère représentatif des attentes industrielles en termes de performances en détection est utilisé : le nombre d'interventions inutiles, noté NII par la suite. Ce critère a été utilisé par Lebranchu [26] pour analyser les performances de détection des indicateurs multi-turbine construits.

Pour un composant d'une éolienne donné, on fixe le seuil de détection comme la valeur à partir de laquelle on obtient n_i interventions inutiles pendant la période d'apprentissage H_L , n_i étant un paramètre de réglage. En pratique, le réglage du seuil de détection se fait de la manière suivante : pour optimiser le temps d'exécution du processus, on utilise les 5% plus grandes valeurs de l'indicateur multi-turbine sur la période d'apprentissage, puis on applique différentes valeurs de seuil de détection de la valeur maximale à la valeur minimale de cette sélection de valeurs. On retient alors la première valeur de seuil fournissant un nombre d'interventions inutiles $n_i \leq 1$.

5.1.5 Fonctionnement en ligne : détection des défauts

Une fois les entrées du modèle sélectionnées, les paramètres appris et le seuil de détection réglé sur la période d'apprentissage, le système de surveillance peut être utilisé en ligne pour détecter la présence de défaut. Pour un indicateur multi-turbine $\varepsilon_y^{T_i}$ donné, on appelle détecteur une fonction binaire basée sur la comparaison des valeurs de $\varepsilon_y^{T_i}$ avec le seuil de détection associé $\tau_{\varepsilon_y}^{T_i}$.

La binarisation réalisée par un détecteur appliqué aux données d'un indicateur multi-turbine $\varepsilon_y^{T_i}$ est définie par l'équation 5.5 : à un instant t , si le résidu est strictement supérieur à $\tau_{\varepsilon_y}^{T_i}$, le composant est défini comme possiblement en défaut à cet instant (la valeur du détecteur est alors de 1), sinon il est défini comme sain (la valeur du détecteur est alors de 0).

$$D_{\varepsilon_y}^{T_i}(t) = \begin{cases} 1 & \text{if } \varepsilon_y^{T_i}(t) > \tau_{\varepsilon_y}^{T_i} \\ 0 & \text{sinon.} \end{cases} \quad (5.5)$$

Si la valeur du détecteur reste à 1 pendant au moins une journée (correspondant à une intervention), alors le système de surveillance génère une alerte à destination des opérateurs industriels.

Nous venons dans cette section de présenter le processus complet de surveillance conditionnelle proposé, repris dans la figure 5.2 de l'extraction des données au traitement des indicateurs multi-turbine. Dans la section suivante, nous allons présenter en détail le processus de sélection automatique de régresseurs développé pour assurer la génération de modèles interprétables pour tous les éléments d'une flotte d'éolienne.

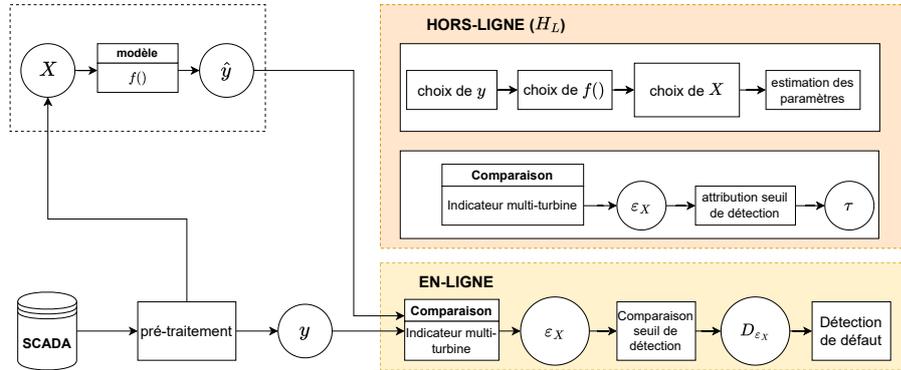


Figure 5.2: Schématisation du processus complet de surveillance conditionnelle proposé

5.2 Sélection automatique des régresseurs

Nous allons dans cette section présenter le processus de sélection automatique de variables d'entrée, ou régresseurs, permettant la modélisation des variables de température. Ce processus a fait l'objet d'une publication lors de la conférence ESREL (European Safety and Reliability Conference) de 2022 [51]. Pour rappel, nous choisissons de travailler avec des modèles linéaires et statiques dont le nombre de variables d'entrée est limité à 3 pour respecter les besoins et contraintes de mise en oeuvre industrielle (Tableau 2.1 et 2.2). Pour que le modèle ait une capacité de prédiction acceptable, il est indispensable de choisir 3 variables pertinentes pour prédire la température d'un composant. Ce choix doit se faire parmi une trentaine de variables disponibles. Pour le réaliser, nous allons nous appuyer sur des algorithmes de sélection de variables existant que nous présenterons dans une première partie. Nous expliciterons ensuite le processus proposé, qui combine deux algorithmes de sélection. Pour des questions d'**interprétabilité** et d'**ergonomie**, nous cherchons une unique sélection de variables par composant, qui s'applique sur toutes les turbines du parc.

5.2.1 Algorithmes de sélection de variables utilisés

Pour une variable de température à modéliser y , on applique sur la période d'apprentissage H_L un processus de sélection automatique de régresseurs, combinant deux algorithmes de sélection de variables : la régression pénalisée de type LASSO (Least Absolute Shrinkage and Selection Operator) [50], et la méthode de sélection par étapes (ou stepwise) GFS (pour Greedy Forward Selection) [71].

Régression pénalisée LASSO

L'approche de régression LASSO fait partie des méthodes dites pénalisées. Ces algorithmes présentent la spécificité de réaliser à la fois la régression et la sélection de variables. La sélection de variables est rendue possible par l'application d'une "pénalité" sur les coefficients de régression, qui permet d'écarter les variables les moins pertinentes à la résolution du problème de régression. La comparaison de ces approches de régression pénalisées réalisée par Greenwood [18] a démontré que LASSO est celle permettant une discrimination (ou sélectivité) la plus forte, en écartant complètement (c'est-à-dire en attribuant un coefficient de régression nul) les variables jugées non pertinentes pour l'estimation de y .

L'algorithme de régression LASSO répond à un problème d'optimisation défini par l'équation 5.6, où y est la variable de sortie, x_i une variable dont on cherche à évaluer le lien avec y , a_i son coefficient de régression et λ le coefficient de pénalité.

$$\min_{a_1, \dots, a_M} \sum_{t=1}^N (y(t) - \sum_{i=1}^M a_i * x_i(t))^2 + \lambda * \sum_{i=1}^M |a_i| \quad (5.6)$$

On remarque que le premier terme représente le problème d'optimisation de la régression des moindres carrés (minimisation des erreurs de prédiction au carré). Le deuxième terme illustre une pénalité sur la pente (en valeur absolue) de la droite de régression, et c'est cette pénalité qui permet de réduire les valeurs des coefficients de régression (jusqu'à zéro pour les variables n'apportant pas d'informations utiles pour l'explication de y).

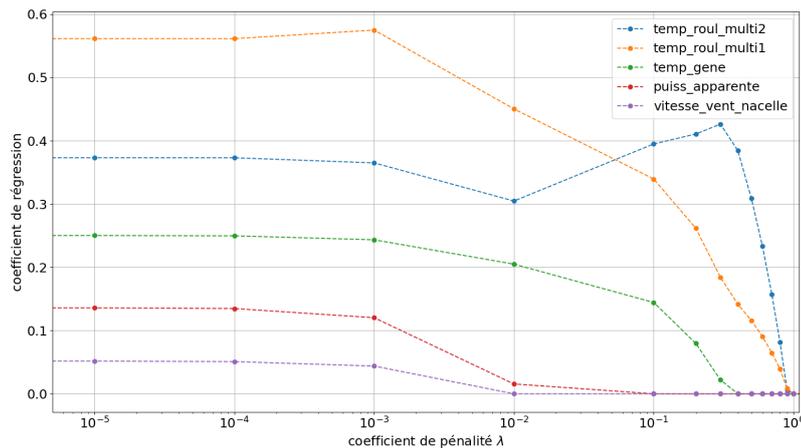


Figure 5.3: Coefficients de régression attribués par l'algorithme LASSO en fonction de la valeur du coefficient de pénalité pour un ensemble de variables

La Figure 5.3 présente l'évolution de valeurs de coefficients de régression en fonction des valeurs du coefficient de pénalité λ , pour la modélisation de la température de l'huile de la multiplicatrice à partir d'un ensemble de cinq variables sur les données H_L du parc N_f^{rc} . Parmi ces cinq variables, trois présentent un lien de similarité avec la température de l'huile (*temp_gene*, *temp_roul_multi1* et *temp_roul_multi2*), et les deux autres un lien de causalité (*puiss_apparente* et *vitesse_vent_nacelle*).

Ce graphique permet d'illustrer la sensibilité de la capacité de réduction des coefficients de régression de la régression LASSO à la valeur de λ : pour des coefficients de pénalité proches de 0 (correspondant au problème de régression des moindres carrés, sans sélection), chaque variable se voit attribuer un coefficient de régression non nul. Pour des λ supérieurs à 10^{-3} , la pénalité permet aux coefficients de décroître à des vitesses variables, et on peut observer la sélectivité de la régression LASSO : les variables *puiss_apparente* et *vitesse_vent_nacelle* sont écartées de la régression pour $\lambda \geq 10^{-1}$, et plus tard *temp_gene* pour $\lambda \geq 4 * 10^{-1}$. Cependant, lorsque le coefficient de pénalité devient trop grand (ici $\lambda = 1$), la sélectivité de l'algorithme est trop forte et toutes les variables sont écartées de la régression, alors que les températures de roulement de multiplicatrice sont des variables pertinentes pour la modélisation de la température de l'huile de la multiplicatrice par leur lien de similarité.

Il est donc nécessaire de paramétrer le coefficient de pénalité λ afin de considérer une régression LASSO plus sélective que la régression des moindres carrés, mais qui permet de conserver les variables les plus pertinentes à la modélisation des variables de sortie.

Greedy Forward Selection (GFS)

La méthode Greedy Forward Selection, ou GFS, est un algorithme de sélection de variable par étape (ou "stepwise") qui évalue à chaque itération plusieurs modèles à l'aide d'un critère défini. Son principe est le suivant : en partant d'un modèle vide pour l'estimation d'une variable de sortie y , un ensemble des variables constituant des régresseurs potentiels est classé à chaque étape selon un critère de sélection. La meilleure variable selon ce critère est alors ajoutée à l'ensemble de régresseurs recherché. Le même processus est repris ensuite de manière itérative en partant d'un modèle constitué des variables sélectionnées à l'étape précédente jusqu'à ce que le critère d'arrêt du processus soit atteint. Le principe de ce processus est schématisé sur deux étapes par la Figure 5.4.

L'approche de sélection de variables GFS présente l'avantage d'être peu coûteuse, ce qui est important lorsqu'on considère une application industrielle sur un grand nombre de données, nécessitant la génération de beaucoup de modèles. De plus, comme le montre les résultats de l'analyse comparative de processus de sélection de variables d'entrée réalisée par

Pohjankukka [49], cette méthode présente des performances similaires à d'autres algorithmes de sélection de caractéristiques plus complexes, tels que l'approche par algorithme génétique.

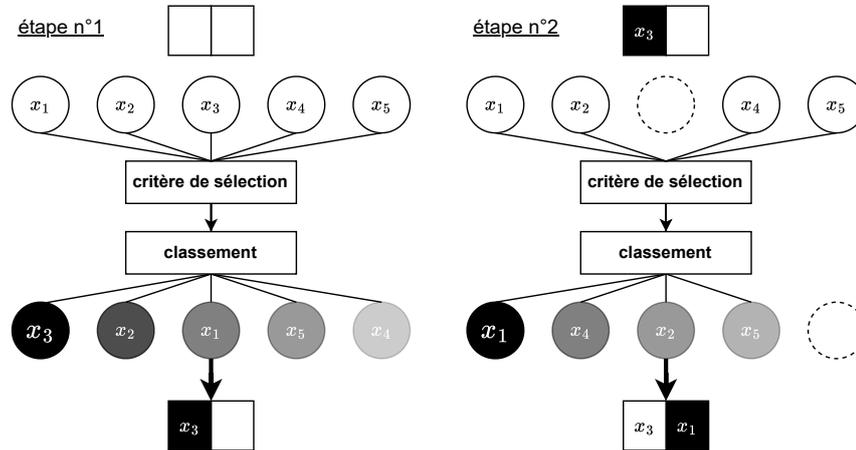


Figure 5.4: Schématisation du principe de sélection de variable GFS pour deux étapes

5.2.2 Présentation du processus LASSO-GFS

Pour un parc éolien F composé de N_T turbines, nous allons détailler le processus de sélection automatique de variables d'entrée proposé dans ces travaux de recherche. Ce processus combine l'utilisation des algorithmes LASSO et GFS que nous venons de présenter. Nous nommerons par la suite ce processus LASSO-GFS.

LASSO : réduction du nombre de régresseurs potentiels

Dans le cadre du processus de sélection automatique de variables d'entrée, la régression pénalisée LASSO est utilisée dans un but de réduction de la dimension du jeu de potentiel régresseurs. En effet, si on considère la totalité de la base de données SCADA, le nombre de variables d'entrées potentielles est très large et l'utilisation de l'algorithme GFS sans sélection préalable peut ainsi être long.

Le protocole suivant est appliqué pour cette première étape du processus de sélection de variables LASSO-GFS :

1. Premièrement, les variables de la base de données SCADA trop peu disponibles (moins de 50% des données disponibles sur H_L) ne sont pas intégrées au processus de génération de régresseurs. On considère en effet que de telles variables ont un contenu informatif trop faible pour être intégrée à la régression réalisée.

2. Pour une variable de température à modéliser y , on applique une régression LASSO avec un coefficient de pénalité de $4 * 10^{-3}$ sur l'ensemble des variables SCADA du parc considéré, pour chaque turbine du parc. La valeur du coefficient de pénalité λ est déterminée expérimentalement : pour les bases de données SCADA considérées elle permet d'exclure les variables jugées non pertinentes tout en conservant un nombre conséquent de potentiels régresseurs pour limiter le risque de sous-optimalité. Pour chaque variable intégrée à la régression LASSO, on obtient alors un ensemble de N_T coefficients de régression, c'est-à-dire un pour chacune des N_T turbines du parc F.
3. Pour chaque variable de régression considérée, on cherche à déterminer une valeur unique (et commune à toutes les turbines) du coefficient de régression correspondant. Pour ce faire, on calcule pour chaque variable la valeur médiane des N_T coefficients de régression calculés. En effet, même si les données considérées pour cette phase de régression sont par hypothèse représentative d'un fonctionnement normal du parc éolien considéré selon le principe NBM, il peut persister des perturbations qui affectent les données, et donc les coefficients de régression associés. On fait l'hypothèse que de telles perturbations impacteront une minorité de turbines. Ainsi, pour avoir une valeur la plus représentative possible d'un fonctionnement normal du parc, on choisit pour chaque variable la valeur médiane des coefficients, moins sensible à ces perturbations que la valeur moyenne.
4. Finalement, on retient les $N = 10$ variables associées aux 10 plus grands coefficients de régression médians. Ce sous-ensemble de variables sera nommé par la suite V_{Lasso} . La taille de V_{Lasso} est fixée car malgré la calibration du coefficient de pénalité λ réalisée pour contrôler la force de la sélectivité, l'algorithme de régression présente une sélectivité variable, qui dépend de la variable de sortie y à modéliser, comme nous le verrons dans la section suivante.

On choisit un nombre fixe de variables selon le résultat de la régression LASSO car malgré la calibration de λ , l'algorithme de régression présente une sélectivité relative : il est parfois trop sélectif et sensible à des variables aux variations très proches de la variable à modéliser, et à contrario peut sélectionner des régresseurs peu pertinents, et donc fournir un résultat peu interprétable. Considérer un sous-ensemble de taille fixe (ici, 10, déterminée de façon expérimentale) des "meilleurs" régresseurs retenus par LASSO permet ainsi de prendre en compte les limites de sélectivité de l'algorithme LASSO en application industrielle afin d'assurer une bonne réduction de dimension du nombre de régresseurs à considérer. La relative sélectivité de la régression LASSO sera analysée à partir d'un exemple concret dans la section suivante.

GFS : sélection automatique des régresseurs

L'application de la régression LASSO sur l'ensemble de la base de données du parc F permet d'obtenir V_{Lasso} , un ensemble de 10 variables potentiellement pertinentes pour la modélisation d'une variable de sortie y . V_{Lasso} contient les variables qui ont obtenues les coefficients de régression les plus grands pour la majorité des turbines du parc. Cet ensemble de variable est alors utilisé en entrée de l'algorithme GFS aux côtés de la variable de sortie y , sur les données du parc F pour la période d'apprentissage H_L .

L'algorithme de sélection de variables GFS permet de générer X , l'ensemble des trois meilleurs régresseurs pour la modélisation de y , selon le processus suivant :

1. À l'étape $s = 1$, pour une variable v de V_{Lasso} , le modèle $\hat{y} = f(v)$ est entraîné sur la période H_L par régression des moindres carrés pour toutes les turbines du parc F.
2. L'erreur moyenne absolue (ou MAE) entre les données SCADA de y et les valeurs estimées par le modèle \hat{y} , est ensuite calculée selon l'équation 5.7, où N est la taille de H_L , et X l'ensemble de régresseurs utilisé pour modéliser y . Lors de cette première étape du processus GFS, X ne contient que v .

$$\text{MAE}_y^{T_i}(X) = \sum_{t=1}^N \frac{|y^{T_i}(t) - \hat{y}^{T_i}(t)|}{N} = \sum_{t=1}^N \frac{|y^{T_i}(t) - f(X^{T_i}(t))|}{N} \quad (5.7)$$

Le MAE est considéré car c'est un critère couramment utilisé pour évaluer la précision de l'estimation des modèles, simple à mettre en œuvre, et surtout plus robuste aux valeurs aberrantes que d'autres mesures équivalentes telles que l'erreur moyenne quadratique (MSE).

L'erreur moyenne absolue est calculée pour toutes les éoliennes du parc F, et la valeur médiane des MAE obtenus est considérée comme critère de sélection.

$$\text{MAE}_y^{\text{F}}(X) = \text{médiane}_{j \in \llbracket 1; N_T \rrbracket}(\text{MAE}_y^{T_j}(X)) \quad (5.8)$$

Ce critère, défini par l'équation 5.8 pour un parc F composé de N_T turbines, sera nommé par la suite MAE^{F}

3. Le critère de sélection MAE^{F} est alors calculé pour chacun des dix modèles $\hat{y} = f(v)$, $v \in V_{\text{Lasso}}$.
4. La variable v^* fournissant le plus petit MAE^{F} est retenue comme première variable de l'ensemble de régresseurs recherché, et est retirée de l'ensemble V_{Lasso}

5. Le processus est ensuite répété ($s = 2$) pour des modèles à 2 variables (ou bivariés), constitués de v^* et des variables restantes de V_{Lasso} :
 $\hat{y} = f(v^*, v), v \in V_{\text{Lasso}} \setminus v^*$
 Le deuxième régresseur recherché est sélectionné en appliquant les quatre points précédents sur ces modèles à deux variables.
6. Le processus est répété une dernière fois à partir du modèle bivarié optimal obtenu à l'issue de l'étape précédente pour trouver le troisième régresseur de X .

L'algorithme 1 résume la routine GFS présentée.

Algorithm 1 Processus de génération automatique de régresseurs LASSO-GFS

```

1:  $X \leftarrow \emptyset$ 
2: for  $s \in \llbracket 1; 3 \rrbracket$  do
3:   for  $v \in V_{\text{Lasso}} \setminus X$  do
4:      $X_v \leftarrow [X; v]$ 
5:     for  $j \in \llbracket 1; N_T \rrbracket$  do
6:        $\text{MAE}_y^{T_j}(X_v) \leftarrow (5.7)$ 
7:     end for
8:      $\text{MAE}_y^{\text{F}}(X_v) \leftarrow \text{médiane}_{j \in \llbracket 1; N_T \rrbracket}(\text{MAE}_y^{T_j}(X_v))$ 
9:   end for
10:   $v^* \leftarrow \min_{v \in V_{\text{Lasso}} \setminus X}(\text{MAE}_y^{\text{F}}(X_v))$ 
11:   $X \leftarrow \llbracket X; v^* \rrbracket$ 
12: end for

```

5.2.3 Discussion sur le processus LASSO-GFS

Combinaison des deux algorithmes de sélection de variable

L'algorithme GFS ne peut être utilisé seul pour réaliser la sélection de régresseurs car il serait trop coûteux de l'exécuter sur l'ensemble des variables SCADA de la base de données du parc F (pour illustration, le Tableau A.1 en annexe liste l'ensemble des variables de la base de données du parc N_f^{rc}). De plus, la base de données d'un parc éolien peut intégrer des variables très corrélées entre elles, pouvant induire un problème de multicollinéarité lors de l'application de la régression des moindres carrés dans l'algorithme GFS.

La réduction de variables réalisée par la régression LASSO permet alors de réduire fortement le temps d'exécution de l'algorithme GFS en limitant le nombre de modèles à entraîner par application du processus de sélection de variables. Elle permet en outre de régler le risque de multicollinéarité des variables en ne sélectionnant qu'une variable d'un groupe de variables intercorrélées.

De même, il n'est pas envisageable d'utiliser la régression LASSO seule pour réaliser cette tâche car sa sélectivité est difficile à contrôler. En effet, comme on le voit sur l'exemple de la Figure 5.5, illustrant le résultat de l'application de cette régression sur toutes les variables SCADA disponibles et pour les 8 cas de modélisations du parc N_f^{rc} (Tableau 4.2), pour une paramétrisation de sélectivité fixe ($\lambda = 4 * 10^{-3}$), on peut distinguer deux cas non exploitables.

Premièrement, l'algorithme LASSO ne retient qu'une variable, proposant un modèle associé potentiellement peu sensible aux défauts. On peut voir sur la Figure 5.5 que c'est le cas pour la modélisation du transformateur uniquement à partir de la température d'un des enroulements de la bobine de ce composant : si un défaut thermique impacte la bobine du transformateur, tous les enroulements de la bobine seront impactés. Ainsi, la valeur de y modélisée à partir de la variable d'entrée sélectionnée par l'algorithme LASSO sera proche de y , et le modèle associé ne sera pas sensible aux défauts impactant le transformateur.

À l'inverse, pour le cas des roulements de la génératrice, beaucoup de variables sont retenues : les modèles associés présentent ainsi un risque d'indisponibilité car pour qu'un modèle soit disponible, il faut que toutes ses variables d'entrée le soient. Or, les données industrielles présentent fréquemment des périodes de perturbations de la mesure et d'erreurs de communication. De plus, parmi les variables retenues par la régression LASSO, certaines sont non pertinentes (par exemple l'angle d'une pale de l'éolienne).

Ainsi, le couplage LASSO-GFS permet de palier les inconvénients d'utilisation de ces deux algorithmes dans un contexte industriel, et de réaliser une sélection automatique de trois régresseurs pertinents pour la modélisation de toute variable de sortie disponible parmi les données d'un parc éolien, tout en optimisant le temps d'exécution.

Critère de sélection

Le critère de sélection MAE^F utilisé pour évaluer la précision des modèles construits durant l'algorithme GFS est un critère "multi-turbine" : il est calculé à l'échelle d'un parc éolien, ce qui permet à l'algorithme GFS de fournir un ensemble de régresseurs applicable à l'ensemble des turbines du parc éolien considéré.

Une telle hypothèse est possible car les éoliennes d'un parc donné sont suffisamment proches géographiquement pour présenter les mêmes propriétés environnementales (ressource en vent, température ambiante, humidité...), et on peut considérer que les composants surveillés par le processus ont des dynamiques (thermiques, mécaniques, électriques) similaires d'une machine à l'autre, car la technologie de turbine est la même (section 4.1.2).

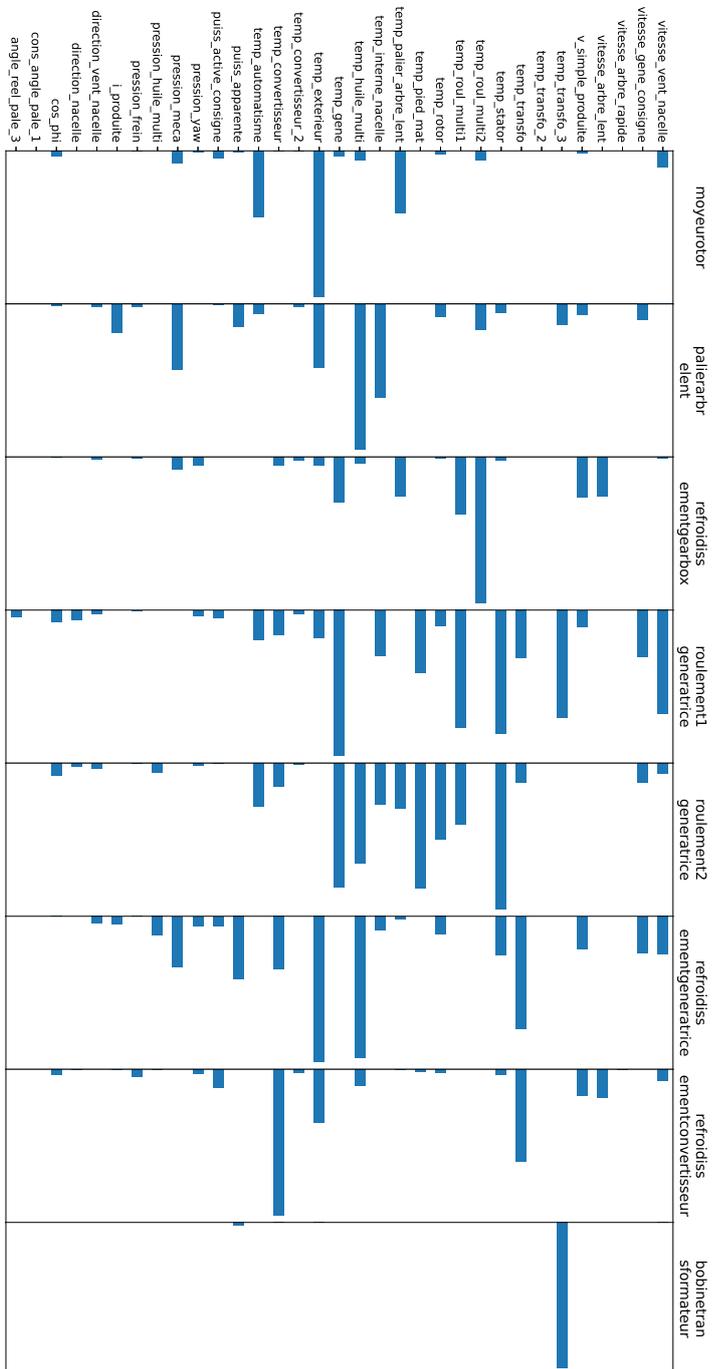


Figure 5.5: Coefficients de régression par variable pour les huit variables de sortie du parc N_{fc} par application de la régression pénalisée LASSO

Malgré l'hypothèse de bon fonctionnement des données de la période d'apprentissage H_L , des perturbations peuvent être observées. Ces perturbations se traduisent par des valeurs de MAE bien plus grandes pour les turbines impactées que pour des turbines ne présentant pas d'anomalie durant la période d'apprentissage. Dans la mesure où l'on fait l'hypothèse que ces perturbations impactent une minorité des éoliennes du parc, la valeur médiane des MAE sera proche de celle des MAE des éoliennes non impactées. Ainsi, la valeur médiane des MAE calculés sur les données de chaque turbine est choisie plutôt que la valeur moyenne par souci de robustesse à la relative qualité des données industrielles.

Industrialisation du processus

Enfin, le processus de génération automatique de régresseur LASSO-GFS proposé permet de répondre au besoin d'**industrialisation** du système de surveillance conditionnelle développé. En effet, il est applicable à la modélisation de n'importe quel élément d'une éolienne, pour toutes les technologies de turbine existantes, sans adaptation de la méthodologie. Sa seule condition d'application est d'avoir accès à une des variables de sortie définies dans le Tableau 4.2.

Ainsi, ce processus peut être appliqué à la surveillance conditionnelle de toute la flotte d'éoliennes exploitée par VALEMO, ainsi qu'à tout nouveau parc intégré à cette flotte dans les années suivant son implantation parmi les outils d'exploitation de l'entreprise.

La Figure 5.6 schématise le processus LASSO-GFS complet, permettant de générer un ensemble de trois régresseurs, pour la modélisation d'une variable y . Ce processus peut être appliqué à toute variable de sortie disponible de la base de données SCADA d'un parc éolien F.

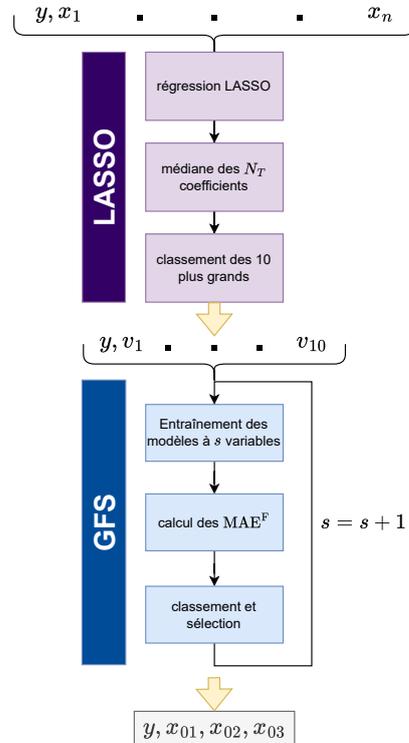


Figure 5.6: Schématisation simplifiée du processus LASSO-GFS

5.3 Méthodes d'évaluation des performances en détection

Nous allons présenter dans cette section la méthode d'évaluation des performances du système de surveillance conditionnelle proposé. Dans un premier temps, nous définirons les outils conventionnels utilisés pour l'évaluation des performances d'un système de surveillance conditionnelle. Puis nous présenterons deux métriques originales d'évaluation des performances en détection. Nous définirons alors le cahier des charges à respecter, fondé sur ces métriques. Nous chercherons à évaluer la capacité de l'outil de surveillance conditionnelle développé à distinguer un défaut en cours d'un comportement normal, ne nécessitant pas d'intervention de maintenance préventive. Ce défaut devra être détecté suffisamment tôt (du point de vue de son développement) afin de permettre la mise en place d'une mainte-

nance préventive, et les fausses alarmes devront être limitée en comportement normal pour éviter les déplacements de techniciens inutiles. Les outils et métriques utilisées pour l'évaluation des performances du système de surveillance conditionnelle seront illustrés et analysés à partir des données du parc E_f^{rg2} .

5.3.1 Définition des périodes de bon fonctionnement et de défaut utilisées pour calculer les performances

Les outils et métriques utilisées pour évaluer les performances du système de surveillance conditionnelle seront calculés sur deux périodes de données : une période où la turbine est en bon fonctionnement H_0 et une période où la turbine est en défaut H_1 .

H_0

La période H_0 est une période de durée un an au cours de laquelle les composants de toutes les éoliennes du parc considéré sont supposées être en état de bon fonctionnement. Elle est différente de la période d'apprentissage H_L , qui est pour rappel une période de durée un an, utilisée pour paramétrer les modèles de bon fonctionnement ainsi que les seuils de détection associés aux résidus construits.

H_1

La période H_1 est une période de durée variable au cours de laquelle un composant de l'une des éoliennes d'un parc éolien présente un défaut. La fin de H_1 est définie comme la date à partir de laquelle le défaut est pris en compte par l'opérateur, soit par une maintenance préventive, soit par une maintenance corrective. Le début de H_1 est plus difficile à définir, car les systèmes de surveillance conditionnelle utilisés par l'entreprise VALEMO ne permettent pas d'identifier le début des défauts étudiés (c'est d'ailleurs la raison pour laquelle on cherche à faire de la surveillance conditionnelle). Nous avons choisi de définir le début des périodes de défaut par une analyse rudimentaire des variations des indicateurs multi-turbine associés, en considérant le commencement théorique d'un défaut (première date de la période H_1) comme la date à partir de laquelle une croissance (non quantifiée) des valeurs résiduelles est observable, et persistante jusqu'à la fin de la période H_1 .

Choix des périodes de données utilisées

Il est à noter que les périodes H_L (période d'apprentissage) H_0 et H_1 ne sont pas nécessairement consécutives. Idéalement, H_0 est choisie juste après la période d'apprentissage H_L , et juste avant la période de défaut H_1 . En

effet, une éolienne étant un système non stationnaire, les modèles appris durant H_L présentent de meilleures performances sur des données proches de l'apprentissage que sur des données localisées quelques années plus tard, où l'état de fonctionnement du parc éolien étudié est potentiellement éloigné de son état juste après l'apprentissage des modèles. Cependant, cet ordonnancement des périodes de données H_L , H_0 et H_1 n'est pas toujours possible. En effet, la période de donnée succédant à H_L peut présenter une maintenance corrective, et donc ne pas être considérée comme une période de "bon fonctionnement". Ce constat est une manifestation de la contrainte de **faible qualité de la donnée** industrielle. Il peut être illustré par le cas du parc E_f^{rg2} , où le remplacement de composant étudié est apparu moins de quatre mois après le début d'exploitation du parc. Par conséquent, il a été nécessaire de choisir les deux années de données succédant à H_1 pour les périodes de bon fonctionnement respectives H_L et H_0 , même s'il faut reconnaître que cette situation n'est sans doute pas optimale.

5.3.2 Outils conventionnels d'évaluation des performances de détection

Disponibilité des indicateurs multi-turbine

L'application de la méthode de surveillance sur une période H_L , H_0 ou H_1 fournit un ensemble de N_d valeurs de l'indicateur multi-turbine. N_d est le nombre de dates où cet indicateur est disponible sur la période, dépendant de la disponibilité des variables d'entrée et de la variable de sortie du modèle considéré.

Ainsi, N_d est inférieur ou égal (pour le cas où les variables d'entrée et la variable de sortie du modèle considéré sont disponibles sur toute la période) à N_t , le nombre total d'échantillons de cette période. On définit alors la disponibilité d'un indicateur multi-turbine sur une période H_L , H_0 ou H_1 comme le nombre d'échantillons où cet indicateur est disponible sur le nombre total d'échantillons de cette période : $\frac{N_d}{N_t}$.

Taux de positifs

La densité de probabilité de l'indicateur multi-turbine sous H_0 ou sous H_1 est estimée à l'aide de la méthode des noyaux [57] sur les N_d valeurs disponibles sur ces périodes.

À titre illustratif, la Figure 5.7 représente la densité de probabilité de l'indicateur multi-turbine issu de la surveillance du roulement arrière de la génératrice de E_f^{rg2} sur les périodes H_0 (courbe verte) et H_1 (courbe rouge).

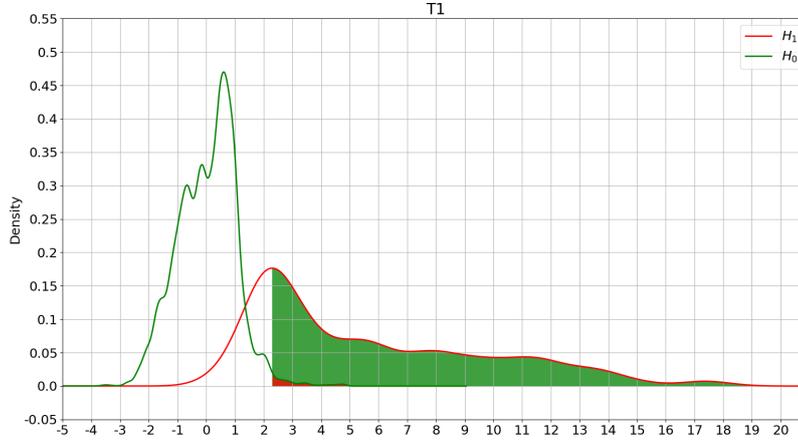


Figure 5.7: Densités de probabilité de l'indicateur multi-turbine sur H_0 et H_1 pour la surveillance conditionnelle du roulement arrière de la génératrice de E_f^{rg2}

L'estimation des deux densités de probabilités d'un indicateur permet de calculer deux métriques conventionnelles de mesure de performances de détection en fonction d'une valeur de seuil de détection τ : le taux de faux positifs (ou TFP) et le taux de vrais positifs (ou TVP). Ceux-ci sont définis par les équations 5.9 et 5.10, où x est une valeur de l'indicateur multi-turbine de la période considérée, ddp sa densité de probabilité et τ un seuil de détection. Le taux de faux positifs est la probabilité que le résidu soit supérieur à τ quand la turbine est en bon fonctionnement (H_0) et le taux de vrais positifs est la probabilité que le résidu soit supérieur à τ quand la turbine est en défaut (H_1).

$$TFP = P(x > \tau | H_0) = \int_{\tau}^{+\infty} ddp_{H_0}(x) dx \quad (5.9)$$

$$TVP = P(x > \tau | H_1) = \int_{\tau}^{+\infty} ddp_{H_1}(x) dx \quad (5.10)$$

Ces métriques sont représentées par les zones colorées de la Figure 5.7 : ce sont les aires sous la courbe des densités de probabilités au-delà du seuil de détection sur les périodes H_0 (pour le taux de faux positifs, représenté par la zone rouge) et H_1 (pour le taux de vrais positifs, représenté par la zone verte).

Courbe ROC

Afin d'évaluer les performances en détection d'un détecteur, nous construisons une courbe ROC ("receiver operating characteristic" [45]).

La courbe ROC considère en ordonnée le taux de vrais positifs (TVP) calculé sur H_1 selon l'équation 5.10, et en abscisse le taux de faux positifs (TFP) calculé sur H_0 selon l'équation 5.9. Ces taux sont calculés pour des valeurs de seuil τ variant de la valeur minimale à la valeur maximale des indicateurs multi-turbine sur les périodes H_0 et H_1 . Pour analyser la courbe ROC, l'aire sous la courbe (AUC) est conventionnellement utilisée.

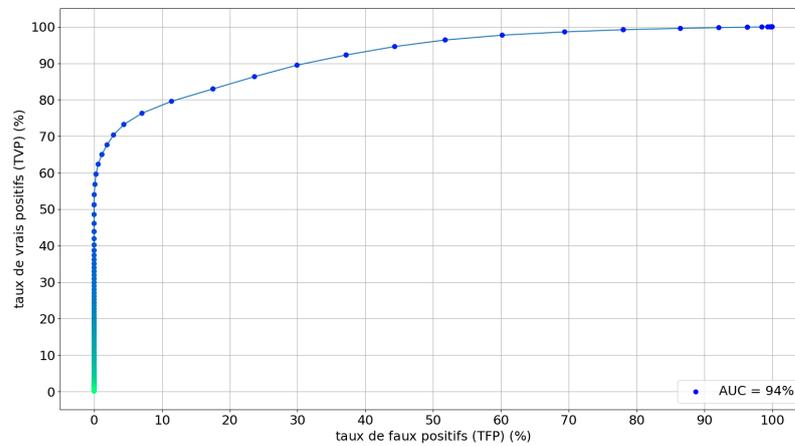


Figure 5.8: Courbe de caractéristique de performance, ou courbe ROC TVP(TFP) calculée à partir des données de l'indicateur multi-turbine sur H_0 et H_1 pour la surveillance conditionnelle du roulement arrière de la génératrice du parc E_f^{rg2}

La Figure 5.8 illustre la courbe ROC obtenue pour la modélisation du roulement arrière de la génératrice du parc E_f^{rg2} . La courbe se lit de droite (bleu foncé pour les plus petites valeurs de seuil de détection) à gauche (bleu cyan pour les plus grandes valeurs de seuil de détection) : plus le seuil de détection est grand, plus les taux de positifs diminuent, le couple optimal étant (100%, 0%). Pour le cas illustré dans la Figure 5.8, l'AUC est de 94%, ce qui signifie que la binarisation des valeurs de l'indicateur multi-turbine par un seuil de détection permet d'obtenir d'excellentes performances en détection selon les métriques considérées.

Cependant, ce résultat n'est pas exploitable pour un industriel. En vue d'une industrialisation, il est nécessaire de prendre en compte la **main d'oeuvre** : ce qui intéresse un industriel, c'est de savoir combien de déplacements de techniciens occasionnera l'application de l'outil de surveil-

lance conditionnelle, et quelle part de ces déplacements correspondra à la prise en charge d'un réel défaut. Or, il n'est pas envisageable d'associer une alarme (au sens d'un franchissement ponctuel du seuil de détection) à un déplacement de technicien de maintenance. En effet, le cas échéant, un taux de faux positifs de 1% sur H_0 correspond à environ 525 alarmes (pour rappel, H_0 dure un an, et est donc composé de 52560 dates), soit 525 déplacements de techniciens inutiles, donc un sur-coût de maintenance important pour un industriel, sur un seul exemple d'indicateur parmi le grand nombre qu'implique une implantation à l'échelle d'une flotte d'éoliennes. Il est donc nécessaire de fonder l'évaluation des performances en détection du système de surveillance conditionnelle sur d'autres métriques plus réalistes pour un déploiement industriel.

5.3.3 Métriques adaptées à un contexte industriel

Prise en compte de l'intervention pour l'évaluation des performances en détection

Durant la période H_0 , aucune maintenance corrective n'est présente de par la méthode de sélection de données de bon fonctionnement proposée. Toutefois, des perturbations peuvent persister, causées par des erreurs de prédiction dûs à des comportements non-linéaires mal modélisés par les modèles linéaires (c'est par exemple le cas des modèles utilisant une variable de puissance pour estimer une variable de température), des chocs sur les capteurs, ou encore des échauffements temporaires représentant des contextes de fonctionnement marginaux, non appris durant H_L (par exemple une période de quelques heures où la turbine fonctionne à puissance nominale en plein été pendant un épisode météorologique d'effet "sèche cheveux"¹). Ces perturbations sont de courte durée (moins d'une journée), et on suppose qu'elles ne nécessitent pas d'actions de maintenance préventive, et ne doivent donc pas être détectées par le processus de surveillance conditionnelle proposé.

Il est donc pertinent de considérer l'objet "intervention", défini sur la base d'un franchissement de seuil de détection d'au moins une journée, pour le processus de détection de défaut, car il permet ainsi de limiter les alertes fournies par le système de surveillance conditionnelle aux utilisateurs finaux aux perturbations assimilables à des défauts. On respecte de fait la variable de **main d'oeuvre**, en permettant une réduction du nombre d'alertes générées par le processus de surveillance conditionnelle, et donc une réduction du nombre de déplacements des techniciens de maintenance.

¹forte chaleur apparaissant en fin de rafale de vent, le dernier épisode ayant eu lieu en France en Juin 2022

Sur H_0 : nombre d'interventions inutiles à l'échelle du parc (NII^F)

On calcule sur les données de H_0 un nombre d'interventions inutiles représentatif du parc éolien. Cette métrique, qu'on désignera par NII^F par la suite, peut être interprétée comme le nombre moyen d'interventions inutiles générées par la surveillance d'un composant d'une turbine du parc F au cours d'une année.

Le NII^F est égal au nombre total d'interventions inutiles générées par la surveillance conditionnelle d'un parc éolien F (somme des NII pour tous les éléments surveillés du parc F) divisé par le nombre d'éléments surveillés du parc F. Pour un parc éolien F, le nombre d'éléments surveillés est égal au produit entre N_T , le nombre d'éoliennes, et N_c , le nombre de composants modélisés. L'arrondi à l'unité du résultat obtenu est conservée pour faciliter l'interprétation.

$$NII^F = \text{arrondi}\left(\frac{\sum_{i=1}^{N_T} \sum_{j=1}^{N_c} NII_{y_j}^{T_i}}{N_T * N_c}\right) \quad (5.11)$$

Le calcul de NII^F est illustré par l'équation 5.11, où $NII_{y_j}^{T_i}$ est le nombre d'interventions inutiles calculé pour la surveillance du composant j , représenté par la variable de sortie y_j , de la turbine T_i .

Sur H_1 : avance à la détection (ADT)

À partir des données de la période H_1 , on évalue l'avance à la détection, ou ADT. ADT est le nombre de jours entre la fin de H_1 , notée t_{H_1} dans l'équation 5.12, et t_d , la date de début de la première intervention, identifiée par le détecteur au cours de cette période.

$$ADT = t_{H_1} - t_d \quad (5.12)$$

Alors que le NII^F , est calculé à partir des données de tous les indicateurs multi-turbine calculés pour la surveillance conditionnelle d'un parc éolien F, l'ADT n'est calculée qu'à partir des données de l'indicateur multi-turbine du composant de l'éolienne impacté par le défaut.

La Figure 5.9 permet de visualiser l'ADT sur les données de la période H_1 d'un détecteur. Les zones blanches représentent des périodes de franchissement de seuil de détection (détecteur égal à 1) inférieures à une journée, alors que les zones colorées en rouge représentent des séquences de plus de 144 valeurs (équivalent à une journée) de résidus supérieures au seuil de détection, donnant lieu au déclenchement d'une intervention.

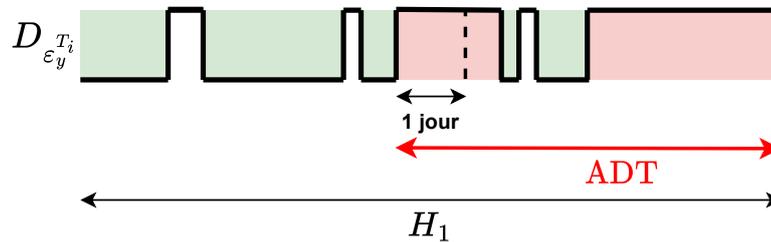


Figure 5.9: Visualisation du calcul de ADT sur les données binaires d'un détecteur sur la période H_1

Dans la littérature, la plupart des études se basent sur la première date de franchissement de seuil pour évaluer les performances de détection de la solution sans analyser l'évolution de l'indicateur au-delà. Or, si ce franchissement de seuil est isolé, et n'est pas suivi par une détection de défaut persistante par le détecteur, alors un utilisateur expert doutera de la nécessité de mobiliser une équipe de techniciens de maintenance. L'utilisation de l'objet "intervention" dans le calcul de l'avance à la détection permet ainsi de prendre en compte la persistance d'une détection en période de défaut.

5.3.4 Cahier des charges des performances en détection d'un système déployable à l'échelle industrielle

Afin de vérifier le respect des besoins et contraintes industrielles (Tableaux 2.1, 2.2), un cahier des charges est défini à partir des métriques suivantes : taux de positifs pour les seuils de détection calibrés à partir de la méthode définie en section 5.1.4, disponibilité des indicateurs, avance à la détection ADT et nombre d'interventions inutiles à l'échelle du parc NII^F.

Avance à la détection (ADT)

Lebranchu [26] a considéré dans son étude que l'application de son processus de surveillance conditionnelle devait permettre la détection du défaut étudié au moins 7 jours avant la défaillance du composant. Cette valeur a été définie d'après une concertation avec différents experts de l'exploitation et la maintenance de la filière éolienne, la considérant comme "la limite minimum pour avoir le temps d'organiser une inspection et mettre la turbine en sécurité". Nous considérons ainsi que le processus de surveillance conditionnelle proposé doit permettre une **avance à la détection d'au moins 7 jours**.

Tant que l'ADT est supérieure ou égale à sept jours, on suppose que le coût de la maintenance préventive est constant (le contenu de l'intervention est inchangé). En revanche, si l'ADT est inférieur à sept jours, la planification d'une maintenance préventive n'est pas possible : la turbine doit être

arrêtée pour éviter le développement du défaut vers une défaillance, et une intervention de maintenance corrective doit être planifiée pour remplacer le composant.

En contexte "normal" de fonctionnement du service de maintenance de VALEMO (c'est-à-dire pas d'intervention de grande envergure en cours), un technicien de maintenance peut être déployé sur le terrain en moins de 2 jours pour une intervention de maintenance préventive (lever de doutes). Cependant, une entreprise O&M peut être confrontée à des périodes de fortes sollicitations de son service de maintenance, notamment à cause de la forte dépendance climatique et météorologique du fonctionnement des installations renouvelables : si un ouragan touche toute la côte Ouest de la France, et a pour conséquence la dégradation d'un tiers de la flotte VALEMO, localisée sur la zone d'activité du phénomène météorologique, alors il sera probablement nécessaire d'attendre la fin des interventions de maintenance occasionnées pour mettre en place l'intervention de maintenance préventive. Dans ce contexte particulier, une avance à la détection de 7 jours n'est pas suffisante pour permettre la prise en charge du défaut identifié avant son développement en défaillance. Cette valeur limite est ainsi à considérer comme une "moyenne" du temps de préparation nécessaire à la mise en place d'une intervention de maintenance préventive. La considération du contexte industriel permettra ainsi de nuancer et approfondir les analyses de performances réalisées dans cette étude.

Taux de vrais positifs (TVP)

Si en application en temps réel sur une période de défaut durant quelques mois le système de surveillance conditionnelle détecte une intervention la première semaine de la période, mais que le taux de positifs (TVP) sur cette même durée est de 10%, alors il est difficile de conclure à la détection d'un défaut avéré. Ainsi, l'analyse de l'avance à la détection (ADT) seule pour l'évaluation des performances en détection sur H_1 peut être incomplète lorsqu'on considère une période de défaut de longue durée.

Par conséquent, il est nécessaire de coupler l'analyse de l'ADT à une analyse du taux de vrais positifs. Ainsi, le taux de positifs doit être significativement plus important en période de défaut H_1 qu'en période de bon fonctionnement H_0 . Toutefois, les indicateurs linéaires étant imparfaits, nous considérons qu'un objectif de plus de 90% de positifs sur H_1 n'est pas réaliste. De plus, comme nous l'avons expliqué lors de la présentation et l'analyse des courbes ROC pour les données du parc E_f^{rg2} (section 5.3.2), cette métrique n'est pas la plus importante pour un industriel, car elle ne dit rien sur les impacts de la surveillance conditionnelle sur les coûts et l'organisation de la maintenance. Nous préférons donc revoir cet objectif à la baisse : nous faisons l'hypothèse qu'un **taux de vrais positifs (TVP) supérieur ou égal à 50%**, est suffisant pour confirmer la sensibilité du

processus de surveillance conditionnelle à un défaut.

Nombre d'interventions inutiles à l'échelle du parc (NII^F)

Les interventions inutiles présentent un surcoût de maintenance significatif pour une entreprise O&M, puisqu'elles impliquent la mobilisation d'un ou plusieurs techniciens de maintenance et l'arrêt temporaire de la production des éoliennes. Il est donc essentiel d'avoir un NII^F le plus proche possible de zéro. Cependant, on sait que des perturbations persisteront sur H_0 . En effet, les périodes de bon fonctionnement (H_0 et H_L) ne contiennent par hypothèse pas de périodes de maintenance corrective, mais elles peuvent contenir des interventions de maintenance préventive corrélées à des perturbations observées au niveau des données SCADA par exemple. Ainsi, imposer un NII^F nul dans notre contexte industriel et méthodologique n'est pas réaliste. En conséquence, on considère que la solution de surveillance conditionnelle développée doit fournir un NII^F **inférieur ou égal à 1**.

La considération du nombre d'interventions inutiles dans le cahier des charges est important car dans la littérature, l'analyse des performances des méthodes de surveillance conditionnelle développées est généralement concentrée sur la bonne détection des défauts étudiés. Il est cependant nécessaire de vérifier que les résultats obtenus ne s'accompagnent pas d'une trop forte sensibilité aux perturbations en période de bon fonctionnement, se traduisant par un nombre conséquent d'interventions inutiles. Le cas échéant, les bénéfices occasionnés par la bonne détection précoce de la défaillance pourraient être annulés par les coûts de maintenance additionnels occasionnés par la surveillance conditionnelle d'autres composants.

L'intégration du nombre d'interventions inutiles à l'échelle du parc dans le processus d'évaluation des performances de l'outil de surveillance conditionnelle permet ainsi de faire un compromis entre détection et fausses alarmes : si le processus proposé permet une détection d'un défaut suffisamment en avance (au moins 7 jours), mais que en contrepartie cette performance occasionne en moyenne 3 interventions inutiles pour la modélisation des autres composants sains du parc considéré ($NII^F = 3$), alors on considère que la solution de surveillance conditionnelle proposée ne répond pas au cahier des charges, et n'est donc pas adaptée à une utilisation industrielle.

Disponibilité (dsp)

Il est nécessaire que la solution de surveillance conditionnelle proposée soit robuste à la contrainte de faible qualité de la donnée industrielle, qui peut se traduire par des périodes d'indisponibilité de la mesure SCADA. Ainsi, on considère qu'elle doit permettre une **disponibilité des indicateurs multi-turbine** (noté par la suite dsp) **supérieure ou égale à 50%**, peu importe la période considérée. L'évaluation du respect de cette contrainte permettra

par ailleurs une meilleure analyse des taux de positifs (TFP, TVP), calculés uniquement sur les dates de disponibilité des indicateurs multi-turbine.

Nous venons de définir le cahier des charges à intégrer à l'évaluation des performances du processus de surveillance conditionnelle proposé. Ce cahier des charges peut être défini par le système d'équation suivant :

$$\begin{cases} \text{NII}^F \leq 1 \\ \text{ADT} \geq 7 \\ \text{TVP} \geq 50\% \\ \text{dsp}_{H_1} \geq 50\% \end{cases} \quad (5.13)$$

5.4 Application sur la surveillance conditionnelle de quatre composants en défaut

Nous allons conclure ce chapitre en appliquant le processus de surveillance conditionnelle présenté sur les données associées à la modélisation des quatre composants en défaut présentés dans le chapitre 4 : le roulement arrière de la génératrice de E_f^{rg2} , le palier de l'arbre lent du parc M_f^{al} , le système de refroidissement de la multiplicatrice de M_f^{rm} et le système de refroidissement du convertisseur de N_f^{rc} .

Dans un premier temps, nous analyserons les résultats du processus de sélection de régresseurs LASSO-GFS. Nous réaliserons ensuite une analyse statistique des variations des indicateurs multi-turbine associés à la modélisation des quatre composants en défaut, afin de réaliser une première évaluation des capacités de détection de défaut du processus de surveillance conditionnelle proposé. Enfin, nous étudierons le respect du cahier des charges sur ces quatre cas de surveillance.

5.4.1 Processus de sélection de régresseurs LASSO-GFS

Régression pénalisée LASSO

Le Tableau 5.1 affiche les résultats de sélection de la régression pénalisée LASSO obtenus pour les cas de modélisations associés aux 4 défauts étudiés et présentés dans le chapitre 4 :

- La surchauffe de roulement arrière de génératrice d'une turbine du parc E_f^{rg2} .
- L'usure du roulement principal d'une turbine du parc M_f^{al} .
- L'échauffement anormal de la multiplicatrice d'une turbine du parc M_f^{rm} .

- Le défaut du système de ventilation des convertisseurs du parc N_f^{rc} , pour lequel on ne s'intéresse qu'à la turbine la plus impactée.

Ce tableau recense également les coefficients de régression médian (de toutes les turbines du parc éolien considéré) attribués aux variables de V_{Lasso} . Les variables sont classées par ordre décroissant de coefficient de régression médian : le régresseur v_1 est ainsi celui qui contribue le plus à la modélisation de y , et le régresseur v_{10} celui qui y contribue le moins. Pour faciliter l'analyse du tableau, les coefficients de régression inférieurs à 0.1 sont grisés. On appellera par la suite "régresseurs principaux" tous les régresseurs possédant un coefficient de régression médian supérieur ou égal à 0.1.

Concernant les régresseurs sélectionnés, on remarque que pour les quatre cas, l'application de la régression LASSO sélectionne à la fois des variables présentant des liens de similarité (températures) et de causalité (puissance, vitesses de rotation et de vent, variables électriques...) avec la variable de sortie à modéliser.

Pour rappel, comme on considère des modèles linéaires, les liens de causalité, généralement représentés par des relations non-linéaires dans le cas d'une éolienne, sont plus difficiles à modéliser que les relations de similarité. Il en résulte des erreurs de modélisations significatives, pouvant générer des fausses alarmes. Cependant, il est important de les intégrer dans les modèles de bon fonctionnement recherchés car elles illustrent certains comportements physiques des composants étudiés (par exemple le lien entre la puissance produite et la vitesse du vent, contenu dans l'équation 2.1, illustrant la production de l'énergie électrique d'origine éolienne), et rendent ainsi les modèles plus interprétables par un utilisateur expert.

En plus de cette diversité sur les liens observés entre les régresseurs sélectionnés par la régression LASSO et les variables de sortie, on observe une diversité dans les composants représentés. Par exemple, pour la modélisation du système de refroidissement de la multiplicatrice ($y = t_{\text{huileMulti}}$), en plus de variables illustrant des variations du composant modélisé (températures de roulement et du carter de la multiplicatrice, vitesse de rotation de l'arbre rapide, pression de l'huile), des variables appartenant au roulement principal (composant voisin de la multiplicatrice du côté des pales), ainsi qu'à la génératrice (localisée de l'autre côté de la multiplicatrice d'après le schéma de nacelle d'une éolienne du parc N_f^{rc} de la Figure 2.1, présentée au début du chapitre 2) sont également sélectionnées.

Une analyse des coefficients de régression médians permet d'émettre plusieurs observations. On remarque tout d'abord que pour chaque cas de modélisation, un ou plusieurs régresseur est associé à un coefficient de régression largement supérieur aux autres variables. Cette observation illustre la capacité de sélection de la régression LASSO. Par exemple, la modélisation de la température de l'huile de la multiplicatrice se fait prin-

	$t_{roulGene2}$		$t_{paliArbrLent}$		$t_{huilMulti}$		$t_{convertisseur}$	
	régresseur	coefficient	régresseur	coefficient	régresseur	coefficient	régresseur	coefficient
v_1	t_{stator}	0.468	$v_{ventNacelle}$	0.495	$t_{roulMulti2}$	0.850	$t_{convertisseur2}$	0.518
v_2	$t_{roulGene1}$	0.449	v_{vent2}	0.490	$v_{arbrRapide}$	0.158	$t_{transformateur}$	0.286
v_3	$t_{transformateur}$	0.057	t_{hub}	0.426	$t_{boitMulti}$	0.142	$t_{generatrice}$	0.158
v_4	$t_{interne}$	0.054	$t_{roulMulti1}$	0.136	$i_{produite}$	0.129	$v_{simpProduite}$	0.100
v_5	$t_{roulMulti2}$	0.040	$t_{exterieur}$	0.098	t_{stator}	0.123	$t_{huilMulti}$	0.060
v_6	$p_{apparente}$	0.024	$t_{convertisseur}$	0.069	$\rho_{huilMulti}$	0.063	$p_{activeConsigne}$	0.059
v_7	$d_{ventNacelle}$	0.009	$v_{simpProduite}$	0.061	$t_{paliArbrLent}$	0.055	$v_{arbrLent}$	0.059
v_8	$v_{simpProduite}$	0.005	$a_{reelPale1}$	0.053	$cos\phi_i$	0.037	$v_{ventNacelle}$	0.036
v_9	$t_{boitMulti}$	0.003	v_{vent1}	0.038	$p_{apparente}$	0.028	p_{rein}	0.030
v_{10}	$d_{nacelle}$	0.002	$\rho_{huilMulti}$	0.038	$t_{interne}$	0.025	$cos\phi_i$	0.018

Tableau 5.1: Résultat de sélection de la régression LASSO pour les modélisations associées aux 4 défauts présentés dans le chapitre 4

cipalement à partir de la température du roulement arrière de la multiplicatrice, avec un coefficient à 0.85, plus de cinq fois plus important que le coefficient du deuxième régresseur. Toutefois, la régression LASSO peut également réaliser une sélection moins forte, comme pour le cas du roulement principal, où la température du rotor (t_{hub}) ainsi que les vitesses de vent mesurées sont au même niveau de contribution à la modélisation de $t_{\text{palierArbreLent}}$ avec des coefficients de régression proches. Une interprétation des régresseurs principaux (coefficients de régression médian supérieurs ou égaux à 0.1) retenus par la régression LASSO permet d'évaluer les interactions du composant modélisé au sein du réseau de capteurs SCADA : on peut ainsi conclure que le roulement arrière de la génératrice est plus isolé que le roulement principal, car il n'est relié qu'avec des températures mesurées sur la génératrice alors que les variations thermiques du roulement principal peuvent être expliquées à l'aide des variations thermiques du rotor, de la multiplicatrice ainsi que des vitesses de vent mesurées.

Sélection automatique de régresseurs GFS

Le Tableau 5.2 présente les résultats de sélection de l'algorithme GFS appliqué sur les jeux de régresseurs V_{Lasso} présentés dans le Tableau 5.1, pour les 4 cas de modélisations associés aux 4 défauts présentés dans le chapitre 4.

Parc	Composant	y	x_1	x_2	x_3
E_f^{g2}	roulement arrière génératrice	$t_{\text{roulGene2}}$	$t_{\text{roulGene1}}$	t_{stator}	$p_{\text{apparente}}$
M_f^a	roulement principal	$t_{\text{paliArbrLent}}$	$t_{\text{convertisseur}}$	$t_{\text{roulMulti1}}$	$v_{\text{simpProduite}}$
M_f^m	refroidissement multiplicatrice	$t_{\text{huileMulti}}$	$t_{\text{roulMulti1}}$	$t_{\text{boitMulti}}$	COSphi
N_f^c	refroidissement convertisseur	$t_{\text{convertisseur}}$	$t_{\text{convertisseur2}}$	t_{transfo}	$t_{\text{huileMulti}}$

Tableau 5.2: Résultat de sélection de l'algorithme GFS pour les modélisations associées aux 4 défauts présentés dans le chapitre 4

On remarque que les régresseurs choisis par l'algorithme GFS ne sont pas les trois variables associées aux plus grands coefficients de régression LASSO du Tableau 5.1. Cela signifie que les variables jugées les plus pertinentes pour la modélisation de y par l'algorithme de régression LASSO ne fournissent pas un modèle optimal selon le critère de sélection défini dans l'algorithme GFS (la valeur médiane des MAE de toutes les turbines du parc).

La plupart des régresseurs sélectionnés par le processus LASSO-GFS sont des variables de température. Cela s'explique par la prédominance de cette catégorie de variable dans les bases de données SCADA, comme nous l'avons vu dans le chapitre 4, ainsi que par l'utilisation de modèles linéaires. En effet, les températures de la base de données SCADA présentent majoritairement des liens de similarité avec les variables de sortie (qui sont également des variables de température), qui se traduisent par des relations linéaires comme nous l'avons vu en section 4.4.2. Ainsi, intégrer ces liens

dans les modélisations évaluées par l'algorithme GFS permet une meilleure précision d'estimation, représentée par le MAE^F, que les liens de causalité, impactant généralement deux variables de nature différentes (par exemple la relation entre température et intensité du courant produit par la génératrice étudiée en section 4.4.2).

Toutefois, nous pouvons également observer dans le Tableau 5.2 la présence de liens de causalité, notamment par la sélection de la puissance apparente pour la modélisation du roulement arrière de la génératrice de E_f^{rg2} . Pour rappel, l'intégration de liens de similarité et de causalité assure l'**interprétabilité** des modèles de bon fonctionnement construits.

5.4.2 Analyses statistiques des indicateurs multi-turbine

Dans cette section, nous allons tout d'abord analyser les profils temporels des indicateurs multi-turbine associés aux modélisations des quatre composants en défaut. Puis nous évaluerons les capacités de détection à l'aide de courbes ROC.

Variations temporelles des indicateurs multi-turbine

La Figure 5.10 illustre les variations de l'indicateur multi-turbine pour la surveillance du roulement arrière de la génératrice du parc E_f^{rg2} sur la période sans défaut H_0 (Figure de gauche) et sur la période intégrant la surchauffe de génératrice H_1 (Figure de droite). Les zones colorées correspondent aux points de mesure associés à un résidu supérieur au seuil de détection calculé selon la méthode présentée en section 5.1.4.

Durant H_0 les valeurs de l'indicateur multi-turbine sont peu dispersées, et centrées autour de 0. Ainsi, les estimations de y sont proches des valeurs mesurées dans la base de donnée SCADA, ce qui signifie que le composant modélisé est dans un état de fonctionnement normal durant H_0 . Cependant des périodes de perturbation des valeurs de l'indicateur multi-turbine sont observables, notamment durant la fin de la période H_0 . Ces perturbations peuvent avoir différentes origines : approximation linéaire du modèle, présence de perturbations non filtrées par la phase de pré-traitement des données SCADA.

Malgré ce constat, le contraste de valeurs de l'indicateur multi-turbine est fort avec la période H_1 : elles atteignent presque 20 sur la fin de la période de défaut alors qu'elles ne dépassent pas 5 durant H_0 . Pour rappel, les valeurs des indicateurs multi-turbine sont standardisées à partir de leur moyenne et de leur écart-type, calculés sur les données d'apprentissage H_L . Ainsi, une valeur de 20 représente une déviation de 20 fois l'écart-type de l'indicateur multi-turbine en période de bon fonctionnement (qui est de moyenne nulle, et d'écart-type unitaire). Ces différences de valeur montrent que l'indicateur multi-turbine étudié permet de faire une distinction

forte entre une période de comportement normal et le développement d'un défaut au niveau du composant surveillé, tout en considérant uniquement des données d'apprentissage en bon fonctionnement, en suivant le principe NBM.

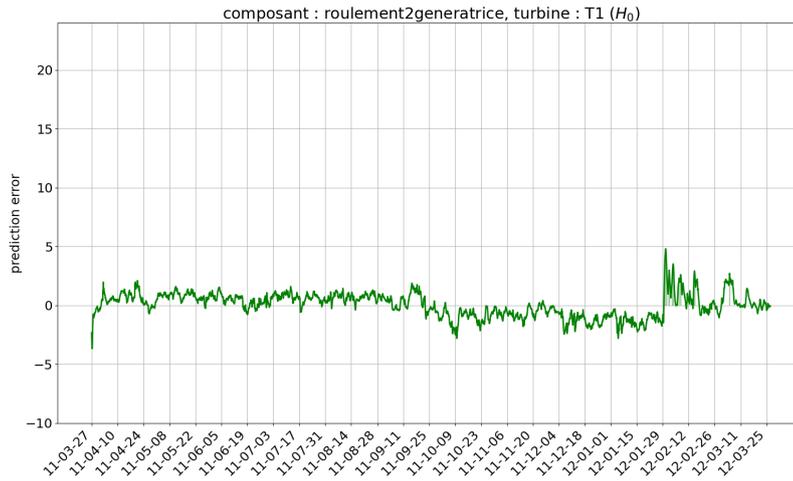
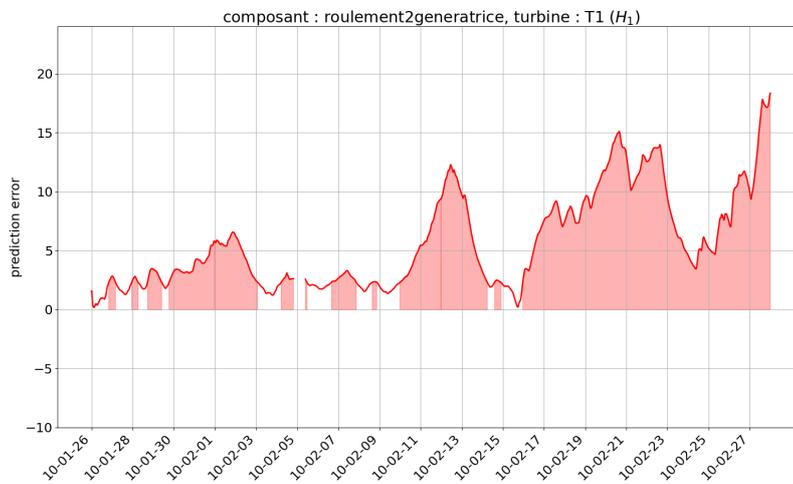
(a) H_0 (b) H_1

Figure 5.10: Variations temporelles de l'indicateur multi-turbine pour la surveillance du roulement arrière de la génératrice de E_f^{rg2} sur H_0 et H_1

La Figure 5.11 affiche les variations temporelles de l'indicateur multi-turbine associé à la surveillance du roulement principal en défaut du parc M_f^{al} .

Un premier constat est l'indisponibilité de l'indicateur multi-turbine sur

la première moitié de la période de développement du défaut H_1 . Ce résultat est lié aux choix de régresseurs pour la construction de l'indicateur multi-turbine : pour le cas de la modélisation du roulement principal de la turbine défaillante du parc M_f^{al} , le régresseur $t_{roulMulti1}$ (voir Tableau 5.2) est très peu disponible, comme on l'a observé sur la Figure 4.5 du chapitre 4.

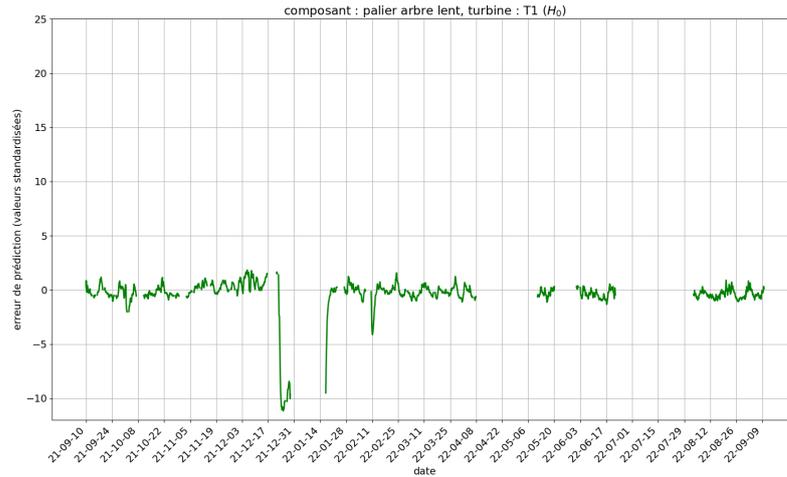
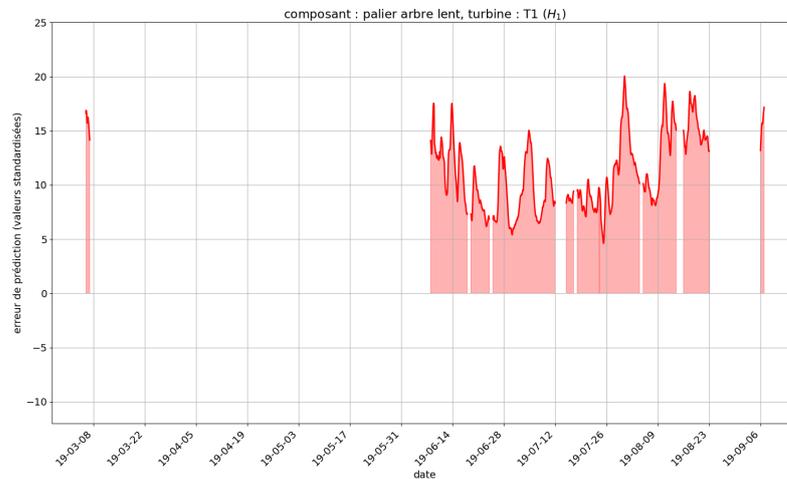
(a) H_0 (b) H_1

Figure 5.11: Variations temporelles des indicateurs multi-turbine pour la surveillance du roulement principal de M_f^{al} sur H_0 et H_1

On remarque que l'usure du roulement principal étudiée se développe sur une longue période, et que l'indicateur multi-turbine est très sensible au

défaut associé : alors qu'en période de fonctionnement normal les valeurs ne dépassent pas 2, elles atteignent 20 durant le développement du défaut d'usure.

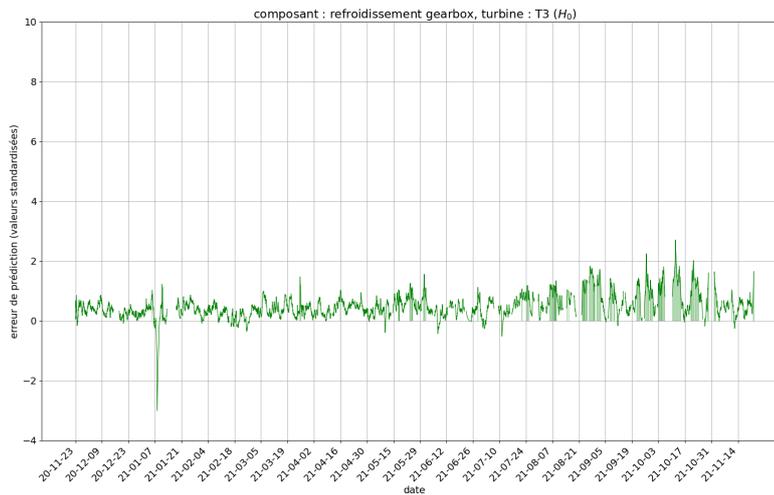
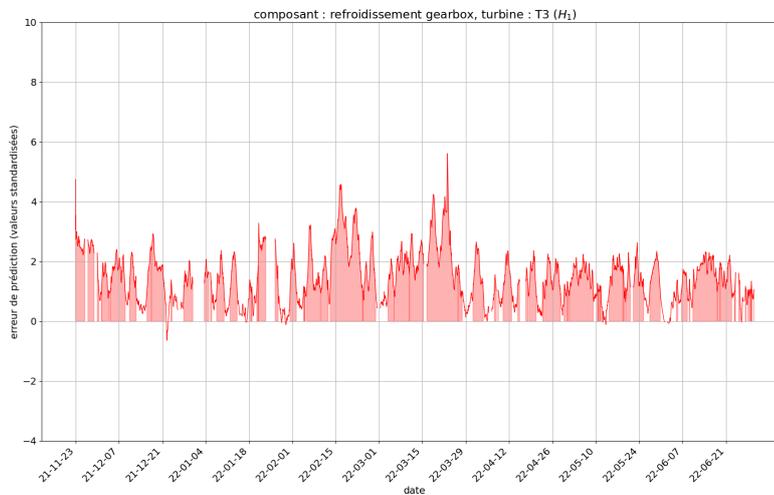
(a) H_0 (b) H_1

Figure 5.12: Variations temporelles des indicateurs multi-turbine pour la surveillance thermique de la multiplicatrice de M_f^{rm} sur H_0 et H_1

La Figure 5.12 affiche les variations temporelles de l'indicateur multi-turbine associé à la surveillance thermique de la multiplicatrice en défaut du parc M_f^{rm} .

L'indicateur multi-turbine est cette fois-ci disponible sur la quasi-totalité de la période H_1 . Cependant, le contraste de valeurs est beaucoup moins important que pour la surveillance du défaut de roulement d'arbre lent du parc M_f^{al} : les valeurs de l'indicateur multi-turbine sont à peine deux fois plus grandes en moyenne sur H_1 que sur H_0 . Ainsi, l'indicateur multi-turbine semble être peu sensible au défaut thermique impactant la multiplicatrice de M_f^{rm} .

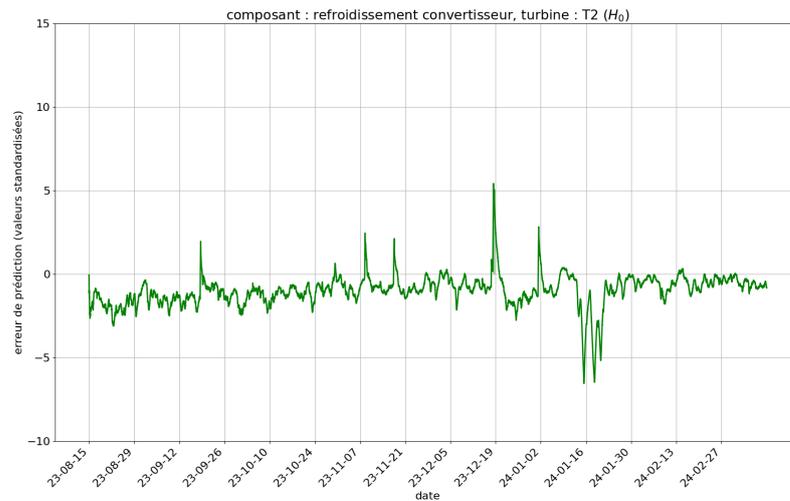
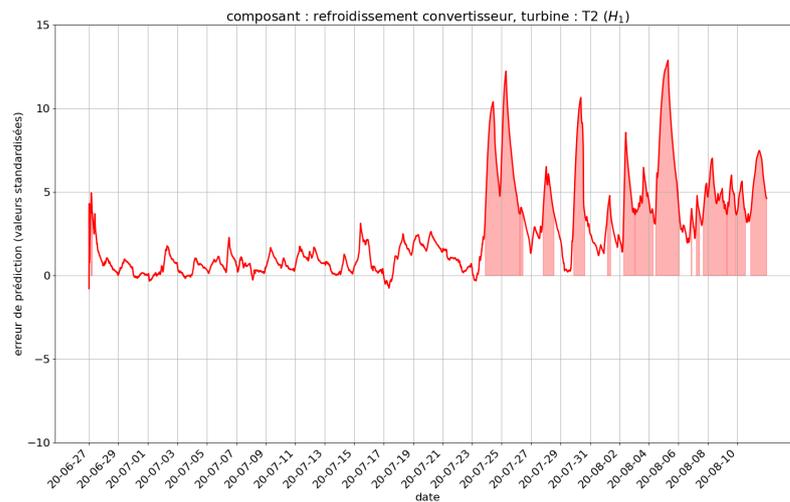
(a) H_0 (b) H_1

Figure 5.13: Variations des indicateurs multi-turbine pour la surveillance du refroidissement du convertisseur de N_f^{rc} sur H_0 et H_1

La Figure 5.13 affiche les variations temporelles de l'indicateur multi-turbine associé à la surveillance thermique du convertisseur en défaut du parc N_f^{rc} .

Le défaut de refroidissement du convertisseur de N_f^{rc} s'est développé sur une période bien plus courte que pour les deux cas de défaut précédents. On remarque que l'indicateur multi-turbine est disponible sur la totalité des dates considérées sur H_0 et H_1 , mais n'est sensible au défaut que durant les dernières semaines de la période H_1 .

Les indicateurs multi-turbine associés aux quatre cas de modélisation de composants en défaut considérés présentent des profils temporels très variés. Ils se distinguent notamment par leur disponibilité sur les périodes H_0 et H_1 , ainsi que leur sensibilité aux défauts étudiés, illustrée par le contraste de valeurs entre ces deux périodes. Nous allons maintenant analyser les courbes ROC associés à ces quatre cas d'étude, afin d'évaluer les capacités de détection des indicateurs multi-turbine.

Courbes ROC

Les courbes ROC TVP(TFP) associées aux données des indicateurs multi-turbine des modélisations des quatre composants en défaut est présentée dans la Figure 5.14.

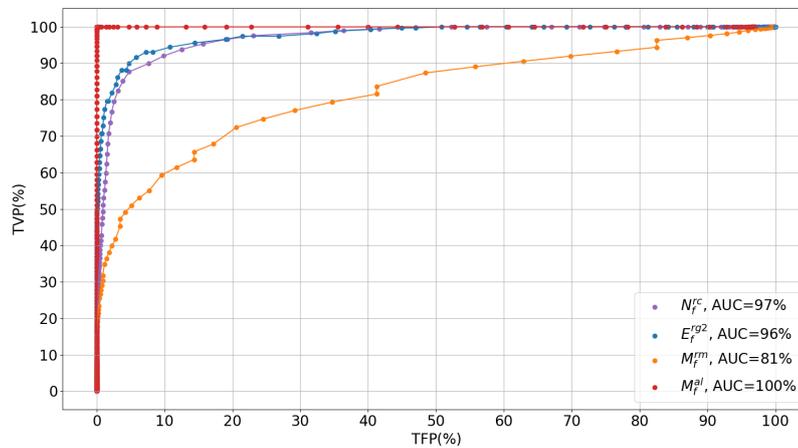


Figure 5.14: Courbes ROC des parcs E_f^{rg2} , M_f^{al} , M_f^{rm} , N_f^{rc}

Chaque cas de modélisation est représenté par une courbe de couleur propre : les métriques calculées à partir des données du parc E_f^{rg2} sont reportées sur la courbe bleue, celles du parc M_f^{rm} sur la courbe orange, celles du parc M_f^{al} sur la courbe rouge, et la courbe violette pour les données du parc N_f^{rc} .

Les courbes ROC construites à partir des données des parcs E_f^{rg2} et N_f^{rc} sont très similaires, et présentent des AUC respectives de 96% et 97%. Ce résultat signifie que les indicateurs multi-turbine étudiés permettent une excellente distinction des données en bon fonctionnement (H_0) des données en défaut (H_1).

Les performances de détection de la surveillance du roulement principal de M_f^{al} sont quant à elles optimales d'après cette courbe ROC (AUC = 100%), ce qui signifie que toutes les valeurs de l'indicateur multi-turbine sur H_1 sont supérieures à la valeur maximale sur H_0 (on le voit d'ailleurs sur la Figure 5.11).

Enfin, l'aire sous la courbe ROC de la surveillance thermique de la multiplicatrice de M_f^{rm} est plus faible, ce qui illustre une moins bonne sensibilité au défaut. Ce résultat est cohérent avec l'observation du plus faible contraste de résidus entre H_0 et H_1 réalisée précédemment sur la Figure 5.12.

Dans la section suivante, nous allons analyser les métriques de performances en détection définies dans la section 5.3. Cette analyse nous permettra d'évaluer le respect du cahier des charges défini en section 5.3.4, et de déterminer si la solution de surveillance conditionnelle proposée est adaptée à une application industrielle.

5.4.3 Évaluation du respect du cahier des charges

Composant	Parc	μ_{H_0}	var_{H_0}	μ_{H_1}	var_{H_1}	TFP(%)	dsp_{H_0} (%)	TVP(%)	dsp_{H_1} (%)	NII ^F	ADT (jours)
roulement arrière génératrice	E_f^{g2}	0.02	0.97	5.53	15.73	1.44	99.71	71.25	98.17	4	29
roulement principal	M_f^{al}	-0.39	2.46	11.19	12.01	0.0	60.53	99.56	24.75	3	90
refroidissement multiplicatrice	M_f^{rm}	0.55	0.83	2.26	2.99	9.57	86.62	59.29	90.9	3	219
refroidissement convertisseur	N_f^{rc}	-0.99	0.87	2.45	6.98	0.32	100	30.97	100	5	19

Tableau 5.3: Métriques de performances des indicateurs multi-turbine pour la modélisation des quatre composants en défaut

Le Tableau 5.3 contient l'ensemble des métriques de performance définies en section 5.3, calculées sur les données des périodes H_0 et H_1 des indicateurs multi-turbine pour chacune des modélisations de composant en défaut considérées, pour la valeur de seuil de détection calibrée selon le processus présenté en section 5.1.4 :

- Les taux de faux positifs (TFP) et taux de vrais positifs (TVP) calculés sur H_0 et H_1 .

Pour rappel, ces taux n'intègrent pas les dates où les indicateurs multi-turbine ne sont pas disponibles. Il est donc nécessaire d'interpréter leurs valeurs en complément de la disponibilité de l'indicateur multi-turbine : un indicateur présentant un taux de vrais positifs de 100% mais disponible seulement 40% de la période H_1 est en effet un choix

moins pertinent pour un industriel qu'un indicateur présentant un TVP de 80% mais disponible 100% de la période H_1 .

- La disponibilité des indicateurs multi-turbine sur H_0 et H_1 , notée respectivement dsp_{H_0} et dsp_{H_1} .
- Le nombre d'interventions inutiles à l'échelle du parc (NII^F). Pour rappel, cette métrique est calculée à partir des données des indicateurs multi-turbine de tous les composants de toutes les turbines du parc éolien considéré.
- L'avance à la détection (ADT).
- On considère également à titre informatif les moments statistiques (moyenne μ et variance var) calculés sur H_0 et H_1 , afin de quantifier les relatifs contrastes de valeurs entre les indicateurs multi-turbine en période de bon fonctionnement H_0 et en période de défaut H_1 .

Moyenne et écart type

On observe une différence entre les moments statistiques (moyenne, variance) calculés sur la période sans défaut H_0 et sur la période présentant le défaut H_1 pour les quatre modélisations. L'amplitude de cette différence dépend du cas considéré : elle est très forte pour la surveillance du roulement arrière de la génératrice de E_f^{rg2} et du roulement principal de M_f^{al} , mais moins importante pour le cas des parcs M_f^{rm} et N_f^{rc} .

Disponibilité

Pour rappel, le cahier des charges impose une disponibilité de l'indicateur multi-turbine sur H_0 et H_1 d'au moins 50%. À la lecture du Tableau 5.3, on peut remarquer que cette contrainte n'est pas respectée pour la surveillance conditionnelle du roulement principal en défaut du parc M_f^{al} : l'indicateur multi-turbine n'est disponible que 24.75% de la période H_1 .

Comme nous l'avons vu dans lors de l'analyse des variations des indicateurs multi-turbine pour les données du parc M_f^{al} (Figure 5.11), ce résultat est du à la présence de la température du roulement avant de la multiplicatrice parmi les régresseurs du modèle de bon fonctionnement utilisé, très peu disponible sur cette période.

Taux de positifs

Concernant les taux d'alarmes, le TFP est inférieur à 10% pour tous les cas de modélisation. On remarque notamment un résultat presque optimal pour le parc M_f^{al} (TFP = 0%, TVP = 99.56%). Cependant, ce résultat est à nuancer par la forte indisponibilité de l'indicateur associé.

Le taux de vrais positifs n'est à contrario pas aussi bon pour tous les cas : il est notamment très faible pour le parc N_f^{rc} (30.97%). Pour ce cas de modélisation, le cahier des charges (TVP $\geq 50\%$) n'est donc pas respecté.

ADT

On remarque que l'avance à la détection est très variable, et va de 19 jours pour la surveillance du défaut de refroidissement du convertisseur de N_f^{rc} à 219 jours pour le défaut de refroidissement de la multiplicatrice de M_f^{rm} . Cette variabilité s'explique par le fait que l'ADT est directement liée à la temporalité du défaut. Néanmoins, le cahier des charges reste respecté : l'avance à la détection reste toujours largement supérieure à 7 jours pour les quatre cas de surveillance de défaut étudié.

NII^F

Le nombre d'intervention inutiles calculé à l'échelle du parc NII^F est strictement supérieur à 1 pour tous les cas : le cahier des charges n'est ainsi pas respecté. Ce résultat signifie que les indicateurs multi-turbine sont trop perturbés sur des périodes sans défaut, et que l'utilisation du système de surveillance conditionnelle proposé implique un surcoût de maintenance conséquent.

5.5 Conclusion

Nous avons débuté ce chapitre en présentant le processus complet de surveillance conditionnelle proposé pour la détection de défaut en application industrielle à la filière éolienne. Son principe est fondé sur l'utilisation d'un unique modèle linéaire statique à trois variables d'entrée qui puisse expliciter l'évolution de la température d'un composant de toutes les turbines d'un même parc et générer des résidus multi-turbines.

Nous avons par la suite détaillé la méthode de sélection automatique de régresseurs LASSO-GFS qui est au coeur de notre solution. L'objectif de cette méthode est de sélectionner pour n'importe quel composant et n'importe quel parc d'une flotte d'éolienne les variables d'entrées les plus pertinentes pour prédire la température du composant, de manière à obtenir un modèle suffisamment précis pour générer des résidus d'amplitude plus faible en période de fonctionnement normal qu'en période de défaut.

Une réflexion a également été menée sur la stratégie de mesure de performances du processus de surveillance. Nous avons ainsi défini un ensemble varié d'outils pour réaliser cette évaluation des performances. Nous avons alors défini un cahier des charges pour l'analyse de ces éléments de mesure de performances.

Ce chapitre a été conclu par une application du processus de surveillance conditionnelle sur la surveillance de quatre composants en défaut. Nous avons pu voir que le processus de surveillance développé permettait en période de défaut l'adaptation du planning de maintenance pour la mise en place d'une intervention de maintenance préventive avec l'assurance d'une avance à la détection de 7 jours, respectant le cahier des charges défini. Toutefois, les analyses du respect du cahier des charges ont également pu mettre en évidence que le processus occasionnait en période de bon fonctionnement trop d'interventions inutiles et donc de déplacements coûteux de techniciens, et donc un surcoût de maintenance indésirable pour un industriel. En outre, nous avons pu observer en périodes de défaut un taux de vrais positifs trop faible pour le parc N_f^{rc} , symptôme d'une insensibilité au défaut étudié, ainsi qu'une trop forte indisponibilité de l'indicateur multi-turbine pour la surveillance thermique du roulement principal du parc M_f^{al} .

En conclusion, l'approche de surveillance conditionnelle présentée dans ce chapitre répond partiellement aux besoins et contraintes définis dans les Tableaux 2.1 et 2.2 : elle fournit des modèles **interprétables** et la méthodologie est **industrialisable**, mais peut fournir des indicateurs multi-turbine trop peu disponibles ($dsp < 50\%$), insensibles à certains défauts ($TVP < 50\%$), et occasionnant de trop nombreuses interventions inutiles en période de bon fonctionnement ($NII^F > 1$). Dans le chapitre suivant, nous proposerons ainsi une approche mettant à profit un ensemble de modèles générés automatiquement, et utilisés pour construire des résidus linéaires, diversifiés et complémentaires pour la surveillance d'un même composant. L'objectif est d'exploiter au mieux l'ensemble de l'information portée par les données SCADA, en conservant des modèles linéaires simples à 3 variables, et miser sur leurs diversités pour optimiser les performances du système de surveillance conditionnelle développé.

Surveillance conditionnelle multi-indicateur

À l'issue du chapitre 5, nous avons démontré que l'approche de surveillance conditionnelle fondée sur l'analyse d'un unique indicateur multi-turbine ne permettait pas de respecter le cahier des charges : elle occasionne un surcoût de maintenance trop important en période de bon fonctionnement, ainsi que des indicateurs potentiellement insensibles au défaut et trop peu disponibles. Afin de répondre à ces problèmes, nous allons présenter dans ce chapitre une approche de surveillance conditionnelle, inspirée des méthodes d'ensemble [10], basée sur la combinaison de plusieurs indicateurs construits à partir de modèles linéaires statiques simples, générés automatiquement.

Nous commencerons par analyser les origines des problèmes de l'approche de surveillance conditionnelle basée sur un unique indicateur multi-turbine construit à partir d'un modèle linéaire statique simple. Nous verrons alors qu'une approche "multi-indicateur" est pertinente pour la résolution de ces problèmes sous certaines hypothèses que nous préciserons. Nous détaillerons alors la méthode de génération de l'ensemble de modèles linéaires statiques simples, ainsi que les indicateurs associés. Cette méthode a fait l'objet d'une publication lors de la conférence ESREL (European Safety and Reliability Conference) de 2023 [52]. Afin d'analyser les valeurs de ces résidus pour la surveillance conditionnelle d'un composant, plusieurs stratégies seront présentées : deux fondées sur une fusion des valeurs des indicateurs générés, et deux réalisant une mise en commun des valeurs binaires des détecteurs associés à ces indicateurs. Ces stratégies seront analysées pour les cas de la surveillance des 4 composants en défaut étudiés au chapitre 5, et les métriques de performances définies dans le chapitre précédent seront comparées à celles d'autres stratégies de surveillance conditionnelle utilisant un unique indicateur. Nous pourrons alors déterminer la stratégie de détection la plus adaptée à une application industrielle, permettant de respecter au mieux le cahier des charges (équation 5.13).

6.1 Motivations de l'approche de surveillance conditionnelle à partir d'un ensemble de modèles

Nous allons dans cette section expliciter les causes des faiblesses de l'approche de détection "mono-indicateur" étudiée dans le chapitre 5 et expliquer sous quelles hypothèses une approche fondée sur un ensemble d'indicateurs linéaires pourrait les résoudre. Étant donné un composant à surveiller à l'aide de la variable de sortie y , nous appellerons X_0 les trois régresseurs retenus par application du processus LASSO-GFS.

6.1.1 Explication des limites de l'approche de surveillance conditionnelle "mono-indicateur"

Durant le chapitre 5, les résultats de l'application du processus de surveillance conditionnelle sur les cas de 4 composants en défaut ont montré que, si l'avance à la détection obtenue était suffisante (supérieure ou égale à 7 jours) et permettait la planification d'une intervention de maintenance pour les quatre parcs étudiés, les autres critères du cahier des charges n'étaient pas respectés : le nombre d'interventions inutiles à l'échelle du parc NII^F reste trop élevé pour les 4 cas (supérieur à 1), le taux de vrais positifs calculé sur H_1 est très faible pour le cas du défaut de refroidissement du convertisseur de N_f^{rc} , et l'indicateur multi-turbine est indisponible sur un peu plus des trois quarts de la période de défaut H_1 pour le cas du suivi du roulement principal de M_f^{al} .

Nombre d'interventions inutiles trop élevé

Un grand nombre d'interventions inutiles en période de bon fonctionnement peut être expliqué par différents facteurs. Comme on l'a vu dans les résultats d'application du processus de sélection de régresseurs présentés dans le chapitre 5, X_0 peut contenir des variables présentant des relations non linéaires avec y , qui sont mal représentées par les modèles linéaires. Elles peuvent donc entraîner la génération de fausses alarmes additionnelles, et donc potentiellement augmenter le nombre d'interventions inutiles recensées.

Une autre explication peut être obtenue en se référant à la méthode de sélection des périodes de bon fonctionnement. Pour rappel, une période est considérée comme présentant des conditions de bon fonctionnement pour l'ensemble des composants du parc considéré si elle ne contient pas d'interventions de maintenance corrective (remplacement de composants). Or, ces périodes peuvent contenir des interventions de maintenance préventive, faisant potentiellement suite à des perturbations de certaines variables SCADA (échauffement anormal de composant, bridage de la production, régulation thermique anormale, erreur de mesure de capteurs SCADA...). Toutes ces perturbations peuvent donc impacter des variables d'entrée des modèles con-

sidérés, et ainsi fausser l'estimation jusqu'à entraîner un résidu suffisamment grand pour générer des interventions en période de bon fonctionnement.

Taux de vrais positifs trop faible

Comme nous l'avons vu à travers l'analyse des performances du système de surveillance conditionnelle sur les 4 composants en défaut du chapitre 5 (Tableau 5.3), pour le cas du parc N_f^c , l'indicateur multi-turbine construit à partir de X_0 ne garantit pas une sensibilité aux défauts thermiques considérés ($TVP < 50\%$). Cela vient du fait que, selon le principe de modélisation NBM, le processus de sélection de régresseurs LASSO-GFS n'utilise pas de données en défaut dans sa phase d'évaluation des modèles (car son objectif est uniquement de bien prédire la variable de sortie en période de bon fonctionnement), et peut ainsi retenir des variables d'entrées elles aussi sensibles au défaut. Ceci a pour conséquence de rendre le résidu insensible au défaut.

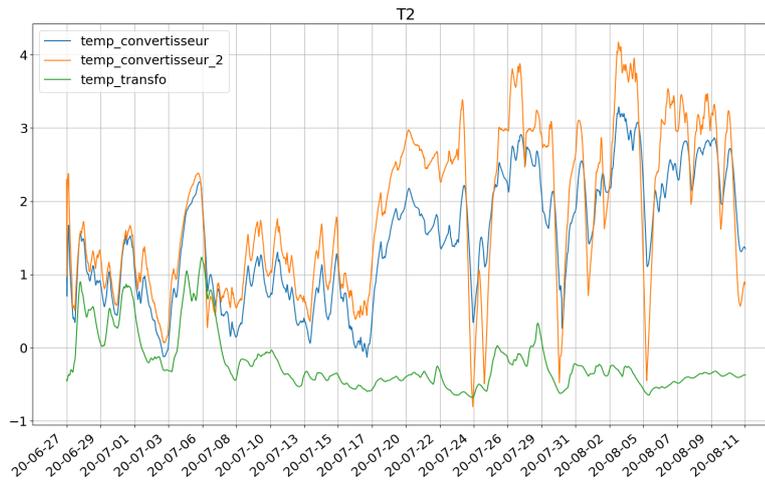


Figure 6.1: Illustration de la corrélation entre une variable à modéliser ($t_{\text{convertisseur}}$) et les régresseurs générés par le processus de sélection LASSO-GFS ($t_{\text{convertisseur2}}$, t_{transfo})

Pour illustrer ce risque, la Figure 6.1, représente les variations des régresseurs $t_{\text{convertisseur2}}$ et t_{transfo} , choisis par le processus LASSO-GFS pour la modélisation thermique du convertisseur ($y = t_{\text{convertisseur}}$) de la turbine en défaut de N_f^c durant H_1 . Le processus LASSO-GFS sélectionne $t_{\text{convertisseur2}}$, la température du fluide de refroidissement d'un convertisseur secondaire localisé à l'intérieur de la nacelle de la technologie de turbine N. Comme on le remarque sur cette figure, les variations de cette température sont très corrélées à celles de la variable modélisée, ce qui signifie que les deux

convertisseurs sont impactés par le défaut. De plus, on peut également observer une période de corrélation avec la température du transformateur en début de H_1 , ce qui peut expliquer la faible sensibilité au défaut observé durant les premières semaines de la H_1 dans l'analyse de performances du chapitre 5. Par conséquent, la valeur de température prédite est proche de la valeur mesurée, notamment en début de période de défaut où deux des régresseurs sont sensibles au défaut étudié, et le résidu garde une valeur proche de zéro malgré la présence du défaut. Le taux de vrais positifs est ainsi réduit.

Disponibilité trop faible

Enfin, le risque de faible disponibilité des indicateurs multi-turbine construits selon la méthodologie de détection mono-indicateur est directement relié aux variables d'entrée utilisées. En effet, comme nous l'avons vu dans le chapitre 4, les variables SCADA présentent de nombreuses perturbations de mesure, et peuvent être très peu disponibles, comme c'est le cas de la température des roulements de multiplicatrice de M_f^{al} (Figure 4.5). Un indicateur multi-turbine utilisant une de ces variables comme régresseur présentera ainsi la même indisponibilité sur la période associée.

6.1.2 Solution proposée et hypothèses d'étude

Pour corriger les limitations de l'approche de surveillance conditionnelle mono-indicateur, nous proposons dans ce chapitre de créer plusieurs modèles prédisant chacun la même variable de sortie à partir d'ensembles de régresseurs différents, et de combiner les valeurs des indicateurs construits à partir de ces différents modèles pour générer des alarmes alertant sur un défaut aux conséquences thermiques d'un des composants d'une éolienne. Ce choix permet de résoudre les limites de l'approche de surveillance conditionnelle mono-indicateur en faisant différentes hypothèses d'étude, explicitées ci-dessous.

Réduction du NII^F : indépendance des indicateurs en période de bon fonctionnement

On fait l'hypothèse que, pendant les périodes de bon fonctionnement H_0 , les différents indicateurs sont décorrélés (ou indépendants) et ne généreront pas d'alarmes simultanément si l'on utilise des modèles dont les régresseurs sont différents. En effet, nous avons expliqué dans la section précédente qu'une fausse alarme sur un indicateur était souvent causée par des perturbations impactant les variations d'une des variables d'entrée du modèle. En partant du principe que les modèles générés sont composés de régresseurs distincts, on suppose ainsi qu'une fausse alarme ne se produira que sur le résidu associé au modèle utilisant une variable d'entrée dont les variations

sont sensibles à la perturbation, et que cette fausse alarme pourra être filtrée par la combinaison des valeurs des autres résidus, non impactés par cette alarme car n'intégrant pas cette variable en entrée de leurs modèles.

Augmentation du TVP : non-unicité des indicateurs sensibles aux défauts thermiques

La deuxième hypothèse que nous émettons est que le résultat du processus de sélection automatique de régresseurs LASSO-GFS, présenté dans le chapitre précédent, ne se restreint pas à l'ensemble de régresseurs X_0 , obtenu par application de V_{Lasso} à l'algorithme GFS. Nous supposons ainsi qu'il est possible d'expliquer une variable de sortie sans dégradation de la précision d'estimation à l'aide de modèles utilisant d'autres ensembles de régresseurs. En utilisant des modèles dont les variables d'entrée sont différentes, nous augmentons la chance d'avoir des modèles générant des résidus sensibles aux défauts, et donc corrélés sur H_1 . Leur combinaison permettra de maximiser le taux de vrais positifs.

Augmentation de la disponibilité : diversité des ensembles de régresseurs

En approche de surveillance conditionnelle multi-indicateur, il suffit qu'un indicateur de l'ensemble utilisé soit disponible pour que la surveillance soit possible.

Il est évident que si pour une turbine T_i la variable de sortie y n'est pas disponible sur une période donnée, alors tous les indicateurs sont indisponibles sur cette même période. Toutefois, comme nous l'avons expliqué dans la section précédente, la disponibilité d'un indicateur dépend également des variables d'entrée choisies pour réaliser la prédiction de y .

Ainsi, en considérant des indicateurs construits à partir d'ensembles de régresseurs distincts, on peut faire l'hypothèse que pour une variable donnée indisponible à un instant t , il existe au moins un indicateur ne l'intégrant pas en variable d'entrée, et donc disponible à cette date. Sous cette hypothèse, la disponibilité de la solution de surveillance conditionnelle multi-indicateur est donc plus grande que la solution mono-indicateur étudiée dans le chapitre 5.

6.2 Génération des résidus et stratégies de détection de défaut

Nous proposons donc de générer plusieurs résidus pour la réalisation d'une approche de surveillance conditionnelle multi-indicateur. Nous allons dans cette section définir les modèles et variables d'entrées utilisées pour les construire. Nous expliquerons notamment comment générer plusieurs modèles

linéaires statiques uniques pour l'estimation d'une même variable en exploitant le processus de sélection LASSO-GFS et à partir de connaissances physiques. Nous présenterons alors les stratégies de combinaison de ces résidus pour améliorer la détection de défaut de l'approche de surveillance conditionnelle en développement.

6.2.1 Création des différents résidus

Indicateurs multi-turbine issus du processus LASSO-GFS

Notre objectif est de créer plusieurs modèles permettant d'expliquer la même variable de sortie y . Comme dans l'approche mono-indicateur présentée dans le chapitre précédent, on se restreint volontairement à des modèles linéaires statiques composés de 3 variables d'entrée afin de faciliter l'analyse des alarmes générées par le système et garantir ainsi l'**interprétabilité** du système de détection de défaut par un utilisateur expert. Afin de permettre le déploiement industriel de ce système, ces modèles doivent pouvoir être générés aisément pour chaque composant. Ainsi, une première solution est d'utiliser le processus de sélection de régresseurs LASSO-GFS présenté au chapitre 5 pour générer automatiquement ces nouveaux modèles.

Notons X_0 l'ensemble des trois régresseurs x_{01} , x_{02} , x_{03} obtenu par application du processus LASSO-GFS selon la méthodologie définie en section 5.2. La stratégie utilisée pour créer d'autres jeux de régresseurs que X_0 pour l'estimation de y est d'appliquer l'algorithme GFS en changeant le contenu de V_{Lasso} , afin de diriger le résultat vers un ensemble de régresseurs différent de X_0 . Pour rappel, dans la première phase du processus LASSO-GFS, l'algorithme de régression LASSO est utilisé pour générer V_{Lasso} , contenant les 10 variables les plus pertinentes pour expliquer les variations de y . Ainsi, en piochant parmi les variables de V_{Lasso} pour l'application de l'algorithme GFS, on s'assure d'avoir des modèles physiquement cohérents et relativement précis. Pour s'assurer d'obtenir un résultat de sélection différent de X_0 , une solution est de retirer certaines variables de X_0 de l'ensemble V_{Lasso} , et d'exécuter l'algorithme GFS à partir de cet ensemble modifié de V_{Lasso} .

Cette méthode est utilisée pour créer 4 nouveaux ensembles de trois régresseurs pour la modélisation de y :

- X_1 , obtenu en appliquant l'algorithme GFS à partir de V_{Lasso} privé de x_{02} et x_{03} , les deuxième et troisième variables de X_0
- X_2 , obtenu en appliquant l'algorithme GFS à partir de V_{Lasso} privé de x_{01} et x_{03} , les première et troisième variables de X_0
- X_3 , obtenu en appliquant l'algorithme GFS à partir de V_{Lasso} privé de x_{01} et x_{02} , les première et deuxième variables de X_0
- X_4 , obtenu en appliquant l'algorithme GFS à partir de V_{Lasso} privé de toutes les variables de X_0

La Figure 6.2 illustre le processus de génération de ces quatre jeux de régresseurs. Nous pourrions appliquer la stratégie présentée pour générer plus que 4 modèles, notamment en ne retirant qu'une variable de X_0 à l'ensemble V_{Lasso} . Cependant, on fait l'hypothèse que cela ne permet pas d'obtenir un résultat de sélection suffisamment distinct de X_0 pour qu'on puisse considérer que le modèle associé présente des estimations décorrélées de X_0 , selon les hypothèses d'étude énoncées en section 6.1.2.

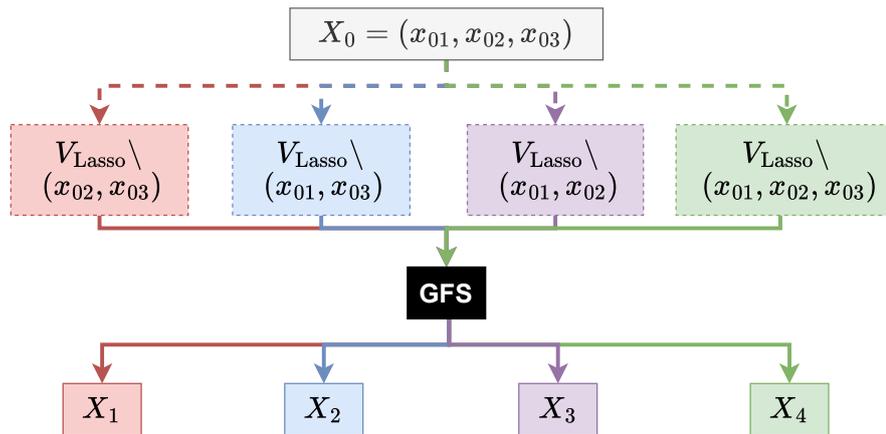


Figure 6.2: Processus de génération des jeux de régresseurs X_1 à X_4 par application de l'algorithme GFS

Cette stratégie permet d'obtenir des jeux de régresseurs physiquement cohérents parce qu'ils sont extraits de V_{Lasso} , retenant les dix variables contenant le plus d'information pour la modélisation de y . Par conséquent, on suppose que les performances de détection des indicateurs associés restent du même ordre de grandeur que celles de celui issu de X_0 . En outre, les modèles associés permettent des estimations de y différentes du modèle issu de X_0 puisqu'ils contiennent des variables d'entrée distinctes de X_0 .

Indicateur multi-turbine issu d'un modèle physique

D'autres modèles peuvent être construits sans appliquer le processus de sélection automatique de régresseur LASSO-GFS. Nous utilisons ainsi un modèle physique qui s'appuie sur des relations de cause à effet connues entre certaines variables opérationnelles et les températures.

Dans le cadre de la modélisation de températures, nous pouvons partir de la relation entre la température d'un composant d'une éolienne (variable de sortie du modèle), la vitesse du vent et la puissance électrique produite par l'éolienne apparente : la production électrique de l'éolienne est liée à la vitesse du vent par la courbe de puissance (Figure 2.2). Or, plus une

éolienne tourne vite (donc produit de l'électricité), plus les composants principaux de la nacelle sont sollicités, et donc plus leur température s'élève. La température des composants situés à l'intérieur de la nacelle d'une éolienne dépend aussi de la température extérieure, du fait de l'isolation thermique imparfaite de l'armature d'une nacelle d'éolienne.

On peut ainsi construire un modèle physique expliquant la variation de température de n'importe quel composant d'une éolienne en fonction de la vitesse du vent (v_{vent}), de la puissance apparente ($p_{\text{apparente}}$), définie à partir de la puissance active et de la puissance réactive selon l'équation 6.1, et de la température extérieure (t_{ext}). Nous désignerons cet ensemble de variables d'entrée X_r . On construit à partir de X_r un modèle linéaire statique simple par régression des moindres carrés (équation 5.1).

$$p_{\text{apparente}} = \sqrt{p_{\text{active}}^2 + p_{\text{réactive}}^2} \quad (6.1)$$

Puisqu'il est composé de variables insensibles à tout défaut de composant aux conséquences thermiques, mais représentant un lien avec toute température de composants situés à l'intérieur d'une nacelle d'éolienne, l'ensemble de régresseur X_r peut être utilisé pour expliquer n'importe quelle variable de sortie (Tableau 4.3). Le modèle résultant de X_r est alors "multi-cible" [40] puisqu'il est applicable à l'estimation de plusieurs variables de sortie.

À partir des six jeux de trois régresseurs que nous venons de définir, des indicateurs multi-turbine sont construits selon la formule détaillée dans l'équation 5.3.

Écart à la médiane

Nous venons de voir que la température d'un composant variait en fonction de la puissance produite, de la vitesse du vent et de la température extérieure de l'éolienne considérée. Or, dans un parc composé d'éoliennes appartenant à la même technologie, et où les éoliennes sont proches géographiquement, on peut considérer que les éoliennes sont soumises aux mêmes conditions environnementales (sous l'hypothèse que les effets de sillage¹ sont correctement gérés) : la température extérieure et le vitesse du vent sont les mêmes, et chaque éolienne produit à peu près la même puissance. Par conséquent, au sein d'un même parc éolien, la température d'un composant évolue de manière similaire d'une éolienne à l'autre en condition de bon fonctionnement. Un écart significatif de température d'un composant d'une éolienne

¹pour un ensemble d'éoliennes alignées, l'effet de sillage se manifeste par une diminution de la vitesse du vent à l'arrière de la première éolienne, impactant négativement la productivité des éoliennes localisées derrière

par rapport à la valeur moyenne ou médiane du parc peut ainsi être révélateur d'un état thermique anormal de ce composant et donc d'un défaut.

En plus de ces six indicateurs multi-turbine, on ajoute à l'approche multi-indicateur développée dans ce chapitre un septième résidu, noté e_{MD} , calculant la différence entre la valeur de la variable de sortie d'une éolienne donnée à l'instant t et la valeur médiane des N_T éoliennes à ce même instant, en suivant l'équation 6.2. Tout comme l'indicateur multi-turbine proposé par Lebranchu [27], e_{MD} mesure ainsi un écart avec une valeur médiane de référence calculée à l'échelle du parc. La valeur médiane des températures du parc est choisie comme valeur de référence plutôt que la valeur moyenne car elle est plus adaptée à un contexte d'utilisation de données SCADA industrielles car plus robuste aux perturbations de mesure.

$$e_{\text{MD}}^{\text{T}_i}(t) = y^{\text{T}_i}(t) - \text{médiane}_{j \in [1; N_T]}(y^{\text{T}_j}(t)) \quad (6.2)$$

Contrairement aux 6 autres indicateurs définis dans les sections précédentes, e_{MD} ne mesure pas l'écart entre une variable et son estimation par un modèle de bon fonctionnement. Il ne nécessite donc pas l'apprentissage de paramètres sur une période de bon fonctionnement, et peut être utilisé en-ligne sans délai. Pour mettre en avant cette distinction et éviter toute confusion avec les autres indicateurs multi-turbine, la notation "e" sera utilisée pour le désigner, tandis que les indicateurs multi-turbine intégrant une modélisation de y seront nommés " ε " et indexés par le nom de l'ensemble des variables du modèle utilisé. Par exemple, ε_{X_1} désigne l'indicateur multi-turbine construit à partir du jeu de régresseur X_1 . Toutefois, comme le reste des indicateurs, e_{MD} se voit attribuer un seuil de détection basé sur ses variations résiduelles sur H_L , suivant le processus détaillé dans la section 5.1.4.

L'écart à la médiane est un indicateur déjà utilisé par des entreprises O&M, dont VALEMO. C'est un signal intéressant pour un système de surveillance conditionnelle car, étant donné qu'il est composé d'une différence de températures sur des éoliennes géographiquement proches, il est supposé être très stable sur une période de bon fonctionnement (puisque la ressource en vent, impactant la température de fonctionnement des composants, est peu variable d'une turbine à l'autre) et présenter moins d'alarmes en période de bon fonctionnement que les indicateurs multi-turbine.

Cependant, c'est un indicateur qui utilise les données de la variable de sortie seulement, et assimile les écarts de valeurs de cette variable avec le reste des éoliennes du parc à des alarmes. Or, comme le rappelle Tchakoua [62], un échauffement significatif de température intervient généralement durant une phase de développement avancé d'un défaut, proche de la défaillance du composant considéré. Alors qu'un défaut se traduit antérieurement par des relations altérées entre la variable de sortie et un ensemble de variations (par exemple la puissance apparente et la vitesse du vent pour le cas

de X_r). Ainsi, en utilisant uniquement la variable de température du composant surveillé comme variable d'entrée d'un modèle pour le cas de e_{MD} , un défaut peut être détecté plus tardivement que par un indicateur multi-turbine considérant un ensemble de variables reliées à cette variable de sortie.

Entrées GFS	Régresseurs	Formule	Nom
V_{Lasso}	$X_0 = \{x_{01}, x_{02}, x_{03}\}$	Équation 5.3	ε_{X_0}
$V_{Lasso} \setminus \{x_{02}, x_{03}\}$	$X_1 = \{x_{11}, x_{12}, x_{13}\}$	Équation 5.3	ε_{X_1}
$V_{Lasso} \setminus \{x_{01}, x_{03}\}$	$X_2 = \{x_{21}, x_{22}, x_{23}\}$	Équation 5.3	ε_{X_2}
$V_{Lasso} \setminus \{x_{01}, x_{02}\}$	$X_3 = \{x_{31}, x_{32}, x_{33}\}$	Équation 5.3	ε_{X_3}
$V_{Lasso} \setminus \{x_{01}, x_{02}, x_{03}\}$	$X_4 = \{x_{41}, x_{42}, x_{43}\}$	Équation 5.3	ε_{X_4}
/	$X_r = \{p_{apparente}, v_{vent}, t_{exterieure}\}$	Équation 5.3	ε_{X_r}
/	y	Équation 6.2	e_{MD}

Tableau 6.1: Caractéristiques des sept indicateurs considérés dans l'approche de surveillance conditionnelle multi-indicateur

En conclusion, sept résidus uniques, dont les caractéristiques de construction sont rappelées dans le Tableau 6.1, sont considérés dans l'approche de surveillance conditionnelle multi-indicateur développée dans ce chapitre : les indicateurs multi-turbine $\varepsilon_{X_0}, \varepsilon_{X_1}, \varepsilon_{X_2}, \varepsilon_{X_3}, \varepsilon_{X_4}, \varepsilon_{X_r}$ et l'écart à la médiane e_{MD} .

6.2.2 Détection à partir d'un unique indicateur

L'objectif de ce chapitre est de comparer les performances de détection d'un processus de surveillance conditionnelle qui n'utilise qu'un unique indicateur (approche mono-indicateur), avec un processus de surveillance qui utilise simultanément les sept indicateurs définis dans la section précédente (approche multi-indicateur).

Nous allons évaluer les performances en détection de l'approche mono-indicateur en n'utilisant que 3 des 7 indicateurs présentés précédemment (Tableau 6.1). En effet, pour rappel, on fait l'hypothèse que tous les indicateurs multi-turbine composés de régresseurs sélectionnés par l'algorithme GFS ($\varepsilon_{X_0}, \varepsilon_{X_1}, \varepsilon_{X_2}, \varepsilon_{X_3}$ et ε_{X_4}) présentent des performances de détection similaires, parce qu'ils sont construits à partir de variables de V_{Lasso} , l'ensemble des régresseurs potentiels les plus pertinents pour l'estimation de y . Il n'est donc pas nécessaire de prendre en compte tous ces indicateurs dans l'analyse comparative des approches de surveillance conditionnelle mono-indicateur. Par conséquent, dans cette étude, ε_{X_0} est choisi arbitrairement comme représentant de ce groupe. On ajoute à l'analyse comparative de méthodes de détection l'indicateur multi-turbine construit à l'aide de connaissances physiques (ε_{X_r}), ainsi que l'écart à la médiane e_{MD} , qui ne prend en compte que les valeurs mesurées de la variable de sortie y de toutes les turbines du parc éolien considéré. Les détecteurs associés sont définis dans

les équations 6.3, 6.4 et 6.5.

$$D_{\varepsilon_{X_0}}(t) = \begin{cases} 1 & \text{si } \varepsilon_{X_0}(t) > \tau_{\varepsilon_{X_0}} \\ 0 & \text{sinon.} \end{cases} \quad (6.3)$$

$$D_{\varepsilon_{X_r}}(t) = \begin{cases} 1 & \text{si } \varepsilon_{X_r}(t) > \tau_{\varepsilon_{X_r}} \\ 0 & \text{sinon.} \end{cases} \quad (6.4)$$

$$D_{e_{\text{MD}}}(t) = \begin{cases} 1 & \text{si } e_{\text{MD}}(t) > \tau_{e_{\text{MD}}} \\ 0 & \text{sinon.} \end{cases} \quad (6.5)$$

6.2.3 Détection à partir de plusieurs indicateurs

Pour augmenter les capacités de détection des défaut d'un système de surveillance conditionnelle, il est pertinent de fusionner les informations fournies par ces indicateurs pour deux raisons.

Premièrement, rappelons que chacun de ces indicateurs est construit à partir de variables d'entrée distinctes. On peut donc imaginer que leurs variations pendant les périodes H_0 ne seront pas concomitantes. En effet, les modèles utilisés étant très simples et donc imparfaits, les fortes variations observées sur un indicateur, à l'origine de fausses alarmes, proviennent de fortes variations sur les variables d'entrée. Celles-ci étant différentes pour chaque modèle, et non liées les unes aux autres, leurs variations ne perturberont pas les résidus aux mêmes instants. On peut donc imaginer réduire le taux de fausses alarmes en combinant les indicateurs.

À l'inverse, si plusieurs indicateurs sont sensibles à un défaut, les variations engendrées par le défaut seront concomitantes durant la période H_1 . De plus, le fait d'utiliser plusieurs modèles distincts implique qu'un plus grand nombre de variables d'entrée est utilisé pour la détection. Par conséquent, la probabilité de sélectionner des variables d'entrées insensibles au défaut augmente, ce qui permet une meilleure détection du défaut, comme nous l'avons expliqué en section 6.1.2

Dans cette section, nous allons ainsi définir quatre méthodes de détection à partir des informations fournies par les sept indicateurs présentés dans la section 6.2.1. Ces méthodes sont fondées sur deux stratégies : la fusion des valeurs des sept indicateurs, et la fusion des valeurs binaires fournies par les détecteurs associés à ces sept indicateurs.

Fusion des valeurs des indicateurs

Une approche de surveillance conditionnelle à partir d'un ensemble d'indicateurs peut être fondée sur une fusion des valeurs fournies par cet ensemble. Selon cette stratégie, nous considérons ainsi deux approches de surveillance conditionnelle multi-indicateur, fondées sur les valeurs moyenne et médiane

des sept indicateurs définis en section 6.2.1. Ces valeurs sont représentées par les indicateurs respectifs $\bar{\varepsilon}$ et ε_{Me} , définis dans les équations 6.6 et 6.7.

$$\bar{\varepsilon}(t) = \frac{1}{7} * \left(\sum_{i=0}^4 \varepsilon_{X_i}(t) + \varepsilon_{X_r} + e_{MD}(t) \right) \quad (6.6)$$

$$\varepsilon_{Me}(t) = \text{médiane} \{ \{ \varepsilon_{X_i}(t), i \in \llbracket 0; 4 \rrbracket \}; \varepsilon_{X_r}(t); e_{MD}(t) \} \quad (6.7)$$

Une des hypothèses d'étude est que les indicateurs linéaires générés fournissent des performances de détection similaires. Par conséquent, il est possible de sommer leurs valeurs sans pondération dans le calcul des valeurs moyennes et médianes. Un seuil de détection est attribué à chacun de ces deux indicateurs selon la méthode décrite dans la section 5.1.4. On appellera alors $D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$, définis dans les équations 6.8 et 6.9, les détecteurs associés aux indicateurs $\bar{\varepsilon}$ et ε_{Me} .

$$D_{\bar{\varepsilon}}(t) = \begin{cases} 1 & \text{si } \bar{\varepsilon}(t) > \tau_{\bar{\varepsilon}} \\ 0 & \text{sinon.} \end{cases} \quad (6.8)$$

$$D_{\varepsilon_{Me}}(t) = \begin{cases} 1 & \text{si } \varepsilon_{Me}(t) > \tau_{\varepsilon_{Me}} \\ 0 & \text{sinon.} \end{cases} \quad (6.9)$$

Pour rappel, les seuils de détection $\tau_{\bar{\varepsilon}}$ et $\tau_{\varepsilon_{Me}}$ sont déterminés à partir des données d'apprentissage H_L du parc étudié, et permettent d'obtenir au plus une intervention inutile sur cette période.

Notion de vote

Nous venons de présenter deux approches de surveillance conditionnelle multi-indicateur fondée sur une combinaison des valeurs des sept indicateurs définis en section 6.2.1. Une autre approche de surveillance conditionnelle multi-indicateur est de combiner les valeurs binaires des détecteurs associés à ces sept indicateurs.

Pour rappel, les indicateurs construits sont par hypothèse indépendants (ils présentent des valeurs décorrélés) sur H_0 . Ainsi, une alarme levée par le traitement d'un de ces sept indicateurs (valeur binaire du détecteur associé égale à 1) n'est pas nécessairement représentative du groupe d'indicateurs, et doit par conséquent être comparée aux valeurs binaires des autres indicateurs pour pouvoir émettre une conclusion de détection.

On parlera par la suite de "vote" pour désigner une valeur binaire fournie par un détecteur. Dans la section suivante, nous présenterons deux méthodes de fusion de ces votes.

Fusion des votes des détecteurs mono-indicateur

Nous allons présenter dans cette section deux méthodes de surveillance conditionnelle multi-indicateur fondées sur une fusion des votes fournis par les détecteurs associés aux sept indicateurs définis en section 6.2.1. Nous appellerons par la suite détecteur multi-indicateur un détecteur fondé sur les valeurs de ces sept indicateurs.

La première méthode consiste à générer une alarme pour tout instant t où la **totalité** des votes disponibles (c'est-à-dire les détecteurs dont les résidus sont disponibles) sont égaux à 1. Le détecteur associé à cette méthode, qu'on désignera par D_{VU} , est donc fondé sur le **vote unanime** des détecteurs associés aux sept indicateurs. Pour cette approche, on fait l'hypothèse qu'un composant est en défaut à un instant t si tous les détecteurs associés aux sept indicateurs considèrent que le composant est en défaut. D_{VU} est défini dans l'équation 6.10 où $N_A(t)$ est le nombre d'indicateurs disponibles à l'instant t et Ω_ε l'ensemble des sept indicateurs définis en section 6.2.1.

$$D_{VU}(t) = \begin{cases} 1 & \text{si } \text{Card}_{\varepsilon \in \Omega_\varepsilon}(\varepsilon(t) > \tau_\varepsilon) = N_A(t) \\ 0 & \text{sinon.} \end{cases} \quad (6.10)$$

La deuxième méthode consiste à générer une alarme pour tout instant t où la **majorité** des votes disponibles sont égaux à 1. Le détecteur associé à cette méthode, qu'on désignera par D_{VM} , est donc fondé sur le **vote majoritaire** des détecteurs associés aux sept indicateurs définis en section 6.2.1. Pour cette approche, on fait l'hypothèse qu'un composant est en défaut à un instant t si la majorité des détecteurs associés aux sept indicateurs considèrent que le composant est en défaut. D_{VM} est défini dans l'équation 6.11.

$$D_{VM}(t) = \begin{cases} 1 & \text{si } \text{Card}_{\varepsilon \in \Omega_\varepsilon}(\varepsilon(t) > \tau_\varepsilon) > \lfloor \frac{N_A(t)}{2} \rfloor \\ 0 & \text{sinon.} \end{cases} \quad (6.11)$$

Ces deux détecteurs ne privilégient pas les mêmes performances en détection : D_{VU} permet une réduction forte des interventions inutiles sur H_0 car tous les indicateurs doivent être sensibles à une perturbation pour que le détecteur D_{VU} le soit. Mais pour la même raison, sa sensibilité aux défauts sur H_1 est moins forte, car conditionnée par la sensibilité de tous les indicateurs au défaut. À l'inverse, D_{VM} est plus robuste à la potentielle présence d'indicateurs non sensibles aux défauts, mais son potentiel de réduction d'interventions inutiles est moindre que celui de D_{VU} .

6.3 Application à la surveillance conditionnelle de quatre composants en défaut

Nous venons de définir plusieurs approches de surveillance conditionnelle multi-indicateur, fondées sur l'ensemble des sept indicateurs définis en section 6.2.1. Nous allons désormais analyser et comparer les performances de ces approches avec les approches de détection mono-indicateur définies en section 6.2.2. L'objectif est de déterminer si la combinaison d'indicateurs construits à partir de modèles linéaires statiques unique permet d'améliorer les performances du processus de surveillance conditionnelle défini dans le chapitre 5. Le cas échéant, nous retiendrons l'approche de surveillance conditionnelle multi-indicateur permettant de répondre au mieux aux limites de l'approche de surveillance conditionnelle mono-indicateur présentée dans le chapitre 5.

On s'intéressera tout d'abord à deux cas d'étude particuliers : la surveillance du roulement arrière de la génératrice de E_f^{rg2} , afin de déterminer l'éventuel gain occasionné par une approche multi-indicateur pour un cas où l'approche de surveillance mono-indicateur fourni déjà de très bonnes performances sur H_0 et H_1 (Tableau 5.3). Nous nous intéresserons ensuite à la surveillance thermique du convertisseur de N_f^{rc} , pour voir si la combinaison des informations fournies par l'ensemble d'indicateurs permet d'améliorer la sensibilité au défaut. On étudiera notamment le respect du critère du cahier des charges sur le taux de vrais positifs ($\geq 50\%$). Puis nous réaliserons une analyse comparative des métriques de performances définies en section 5.3 des approches de détection mono-indicateur et multi-indicateur, afin de déterminer celle permettant de respecter au mieux le cahier des charges. Nous conclurons cette section par une discussion sur l'analyse et l'interprétation du NII^F , en se basant sur une analyse critique de la méthode de sélection des périodes de bon fonctionnement.

6.3.1 Surveillance du roulement arrière de la génératrice de E_f^{rg2}

Dans cette section, ainsi que la section 6.3.2 pour les données du parc N_f^{rc} , nous allons analyser les approches de surveillance conditionnelle mono-indicateur et multi-indicateur définies en section 6.2.2 et 6.2.3 ainsi que les indicateurs et les modèles qui les composent. Dans un premier temps, nous étudierons les variables d'entrée générées par la méthode définie en section 6.2.1, ainsi que les variations des sept indicateurs associés. Après une analyse de courbes ROC, nous pourrions alors évaluer les hypothèses d'étude réalisées en début de chapitre concernant l'indépendance des estimations des modèles sur H_0 et l'existence d'indicateurs sensibles au défaut sur H_1 . Puis nous nous intéresserons aux détecteurs associés aux approches mono-indicateur et multi-indicateur, en analysant la répartition des interventions

inutiles sur H_0 et H_1 .

Modèles et variations des sept indicateurs

Le Tableau 6.2 présente les cinq ensembles de régresseurs obtenus par application du processus de sélection LASSO-GFS, selon la méthode définie en section 6.2.1, pour la modélisation du roulement arrière de la génératrice de E_f^{rg2} .

Identifiant	x_1	x_2	x_3
X_0	$t_{\text{roulGene1}}$	t_{stator}	$p_{\text{apparente}}$
X_1	$t_{\text{roulGene1}}$	$t_{\text{transformateur}}$	$t_{\text{roulMulti2}}$
X_2	t_{stator}	$t_{\text{transformateur}}$	$t_{\text{boitMulti}}$
X_3	$t_{\text{transformateur}}$	$t_{\text{roulMulti2}}$	t_{interne}
X_4	$t_{\text{transformateur}}$	$t_{\text{roulMulti2}}$	$t_{\text{boitMulti}}$

Tableau 6.2: Régresseurs générés par le processus de sélection LASSO-GFS pour l'approche multi-indicateur de la modélisation thermique du roulement arrière de la génératrice de E_f^{rg2}

Pour rappel, les ensembles X_1 , X_2 , X_3 et X_4 sont des nouveaux jeux de régresseurs pour la modélisation de y (ici $t_{\text{roulGene2}}$), obtenus en appliquant en entrée de l'algorithme GFS l'ensemble de variables V_{Lasso} amputé de certaines variables de X_0 . On remarque que les jeux de régresseurs obtenus se distinguent de X_0 , tout en restant physiquement cohérents : ils intègrent principalement des variables de température localisées sur la génératrice ($t_{\text{roulGene1}}$, t_{stator}), mais aussi la multiplicatrice ($t_{\text{roulMulti2}}$, $t_{\text{boitMulti}}$), qui est un composant voisin de la génératrice, ainsi que sur le transformateur ($t_{\text{transformateur}}$), également relié à la génératrice dans la chaîne de production de l'énergie électrique comme nous l'avons vu au début du chapitre 2.

La Figure 6.3 illustre les variations des sept indicateurs définis en section 6.2.1, pour la modélisation du roulement arrière de la génératrice en défaut de E_f^{rg2} .

En période de bon fonctionnement (Figure 6.3a), tous les indicateurs sont centrés autour de zéro avec un faible écart-type, et présentent des variations propres. On remarque notamment que tous les indicateurs ne sont pas sensibles à la perturbation du 29 Janvier au 12 Février 2012 (identifiée par les deux barres rouges verticales sur la Figure 6.3a), impactant ε_{X_0} , l'indicateur multi-turbine étudié dans le chapitre 5 : l'écart à la médiane e_{MD} ainsi que l'indicateur dont le modèle est fondé sur des connaissances physiques ε_{X_r} , semblent en effet conserver des valeurs autour de 0 sur ces périodes.

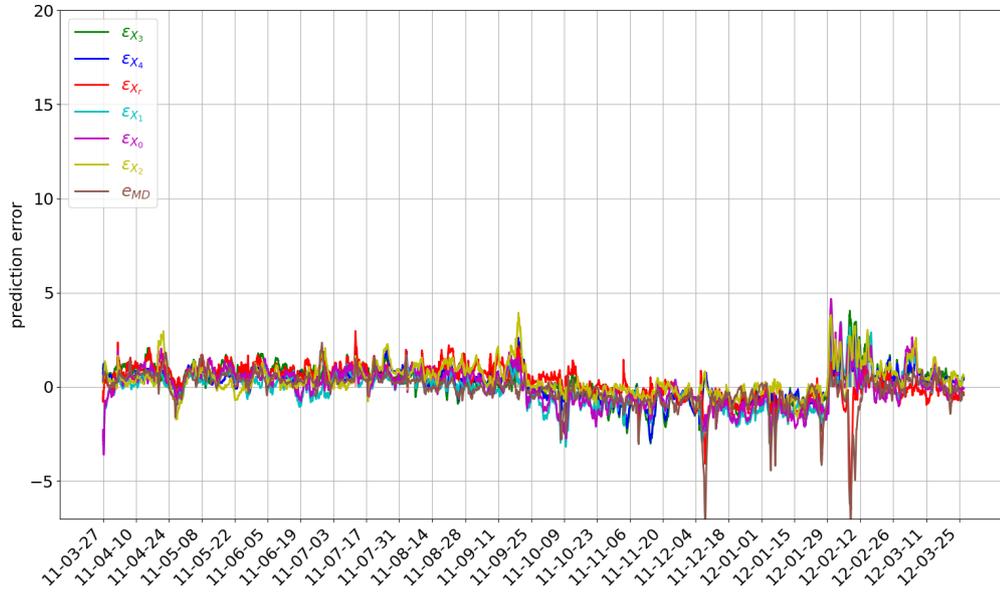
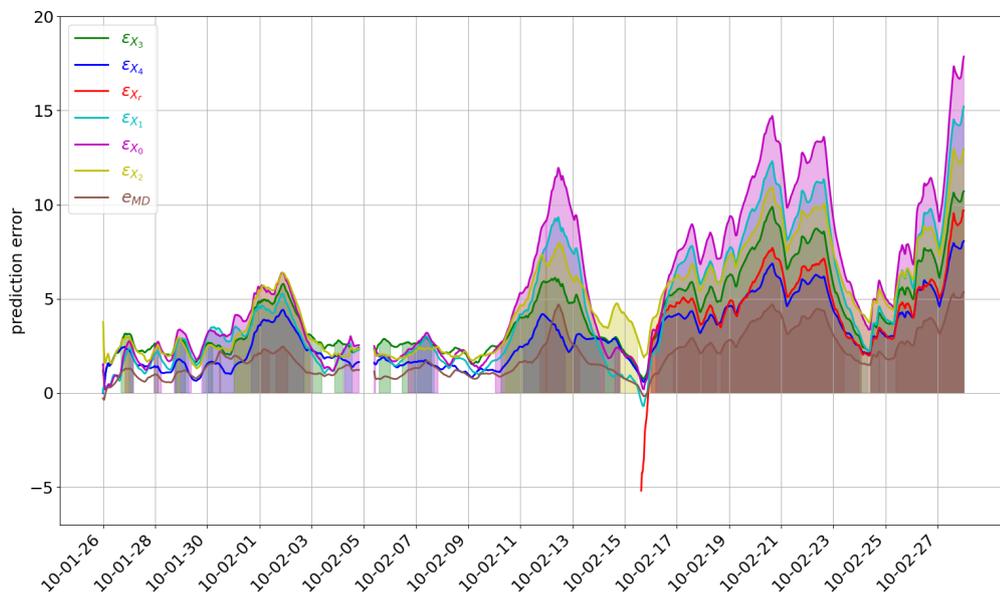
(a) H_0 (b) H_1

Figure 6.3: Variations des sept indicateurs définis en section 6.2.1 pour la modélisation du roulement arrière de la génératrice de E_f^{rg2}

En période de défaut H_1 (Figure 6.3b), tous les indicateurs sont corrélés, et présentent des amplitudes de croissance différentes : ε_{X_0} est l'indicateur le plus sensible au défaut, avec des valeurs dépassant 15 sur la fin de la période, alors que l'écart à la médiane dépasse à peine 5. Ces observations confirment les caractéristiques de l'indicateur e_{MD} spéculées en section 6.2.1 : l'écart à la médiane présente moins de fausses alarmes en période de bon fonctionnement, mais est moins sensible au défaut que les indicateurs multi-turbine, construits à partir des valeurs d'un ensemble de variables d'entrée reliées à y .

La Figure 6.4 représente les matrices de corrélation (en utilisant le coefficient de Pearson R^2) des sept indicateurs sur H_0 (6.4a) et H_1 (Figure 6.4b). Plus la case est verte, moins les valeurs du couple d'indicateurs considéré sont corrélées sur la période utilisée, et plus elle est rouge, plus elles sont corrélées.

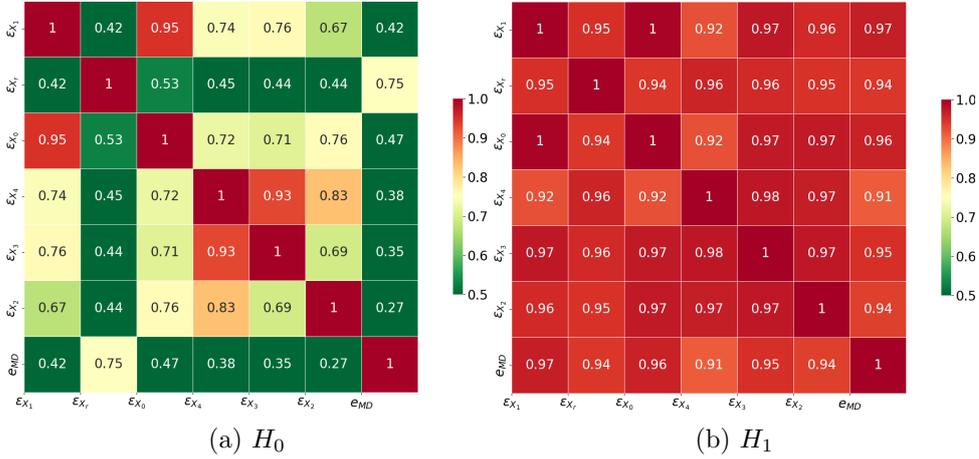


Figure 6.4: Matrices de corrélation des 7 indicateurs définis en section 6.2.1 pour la modélisation thermique du roulement arrière de la génératrice de $Erg2_f$

Le contraste observé par analyse des variations temporelles des indicateurs sur la Figure 6.3 est très visible sur ces matrices : les indicateurs sont relativement indépendants durant H_0 (majorité de cases en teinte verte, présentant des coefficients de corrélation inférieurs à 0.5), mais sont tous corrélés durant le défaut (totalité des cases en teinte rouge, présentant des coefficients de corrélation supérieurs à 0.9). On remarque toutefois que certains couple d'indicateurs ne présentent pas ce contraste. C'est le cas de $(\varepsilon_{X_0}, \varepsilon_{X_1})$, et $(\varepsilon_{X_3}, \varepsilon_{X_4})$, qui sont fortement corrélés sur H_0 . Pour comprendre cette observation, on peut se référer aux régresseurs associés à ces indicateurs (Tableau 6.2) : X_0 et X_1 contiennent $t_{roulGene1}$, qui possède un coefficient de régression très grand par rapport aux autres régresseurs d'après l'analyse réalisée en section 5.4.1. Quant à X_3 et X_4 , ils possèdent

deux variables en commun ($t_{\text{transformateur}}$, $t_{\text{roulMulti2}}$). Par conséquent, les estimations réalisées par ces groupes de modèles sont proches et les indicateurs multi-turbine associés sont corrélés.

La courbe de caractéristiques de performance (ou courbe ROC) présentée en Figure 6.5, permet de comparer les caractéristiques de performances des sept indicateurs, du point de vue des taux de positifs.

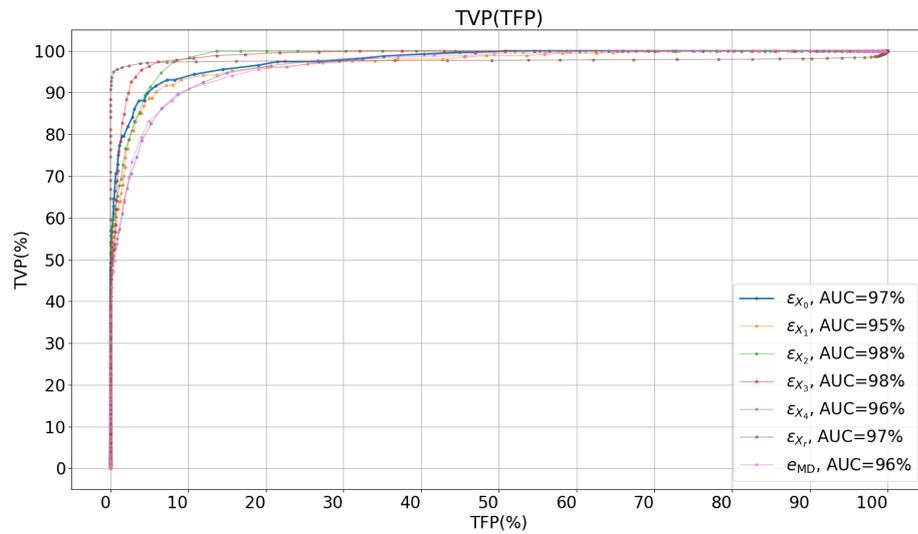


Figure 6.5: Courbe ROC du taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP) pour les sept indicateurs définis en section 6.2.1, appliquée aux données du parc E_f^{rg2}

En considérant l'aire sous la courbe (AUC) pour l'analyse comparative de ces courbes ROC, on observe que les performances de détection (du point de vue des taux de positifs) sont excellentes, et similaires à celles de l'indicateur multi-turbine ε_{X_0} . Ce résultat signifie qu'il est possible de générer plusieurs indicateurs linéaires pour la surveillance d'un même composant, sans dégradation des performances de détection. L'hypothèse réalisée en section 6.1.2 est ainsi confirmée par cet exemple.

Analyse des performances des détecteurs mono-indicateur et multi-indicateur

La Figure 6.6 représente la répartition des interventions inutiles par élément modélisé du parc E_f^{rg2} , représenté en axe des abscisses, pour les approches de détection mono-indicateur et multi-indicateur, distinguées par groupe de couleur.

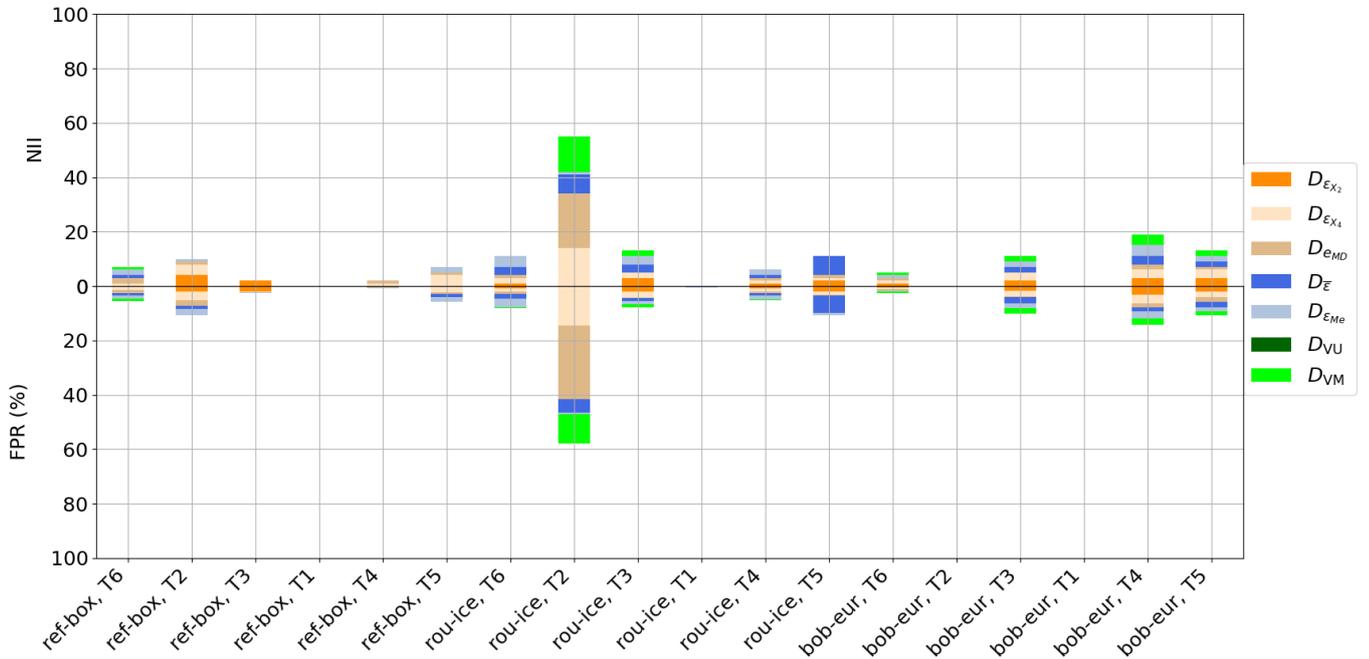


Figure 6.6: Répartition des interventions inutiles (NII) et des taux de faux positifs (TFP) par élément modélisé du parc E_f^{rg2} sur H_0 pour les approches de détection mono-indicateur et multi-indicateur

Les trois premières et dernières lettres des éléments modélisés sont utilisées pour les nommer par soucis d'ergonomie de visualisation. Le taux de faux positif (TFP) est également représenté en symétrie horizontale afin de mettre en avant la corrélation entre ces deux métriques. Les trois approches de détection mono-indicateur sont représentées par des nuances de couleur orangées, les deux approches multi-indicateur considérant une fusion des valeurs résiduelles ($D_{\bar{\epsilon}}$ et $D_{\epsilon_{Me}}$) par des nuances de couleur bleues, et les deux basées sur une fusion des votes des détecteurs associés aux sept indicateurs (D_{VU} et D_{VM}) des nuances de couleur vertes.

Ce graphique montre que les interventions inutiles ne sont pas uniformément réparties sur les 18 modélisations considérées : la modélisation du roulement arrière de la génératrice de la turbine T_2 occasionne ainsi la génération de presque 60 interventions inutiles sur l'année de donnée H_0 en cumulant les sept approches de surveillance considérées (soit un peu plus de 8 par approche), alors que le deuxième composant dont la surveillance occasionne le plus d'interventions inutiles est la bobine du transfor-

mateur de la turbine T_4 , avec un peu moins de 20 interventions. Toutefois, toutes les approches de surveillance conditionnelle ne contribuent pas à cet écart de manière équivalente : le détecteur $D_{e_{MD}}$ occasionne une vingtaine d'interventions inutiles pour la surveillance de ce composant, alors que le traitement du résidu issu du modèle physique (X_r) n'en occasionne aucune. On remarque notamment que si les approches multi-indicateur fondées sur les valeurs moyennes et médianes des sept indicateurs ($D_{\bar{\varepsilon}}$, $D_{\varepsilon_{Me}}$) semblent occasionner des bilans d'intervention du même ordre de grandeur que les approches mono-indicateur considérées, D_{VM} semble fournir tendanciellement moins d'interventions inutiles que les autres approches, et que D_{VU} n'en occasionne aucune sur toutes les 18 modélisations considérées, se présentant ainsi comme un détecteur optimal.

La figure 6.7 montre les valeurs des détecteurs (votes) mono-indicateur et multi-indicateur sur la période H_1 .

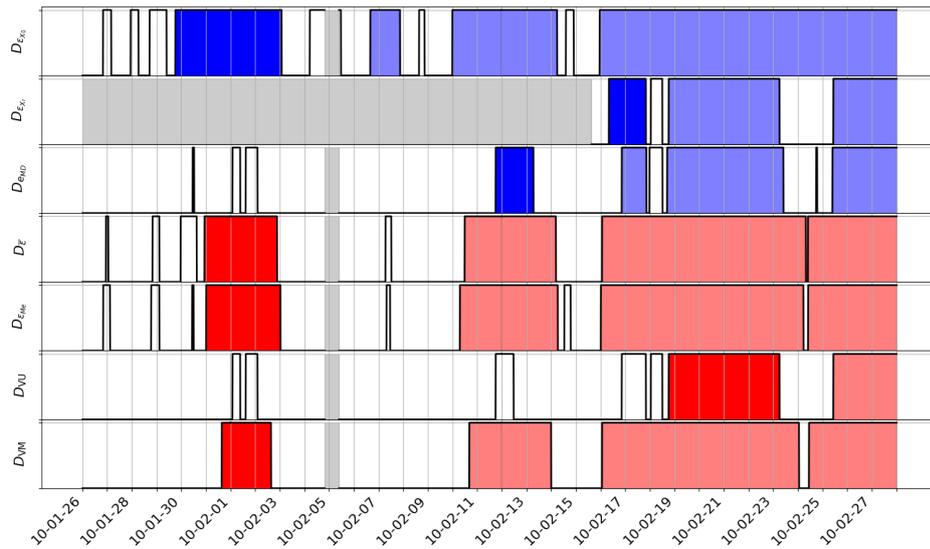


Figure 6.7: Visualisation des valeurs des détecteurs des trois approches de détection mono-indicateur (en bleu) et des quatre approches de détection multi-indicateur (en rouge) pour la modélisation du roulement arrière de la génératrice de E_f^{rg2} sur H_1

Pour rappel, les trois détecteurs mono-indicateur, ainsi que les deux détecteurs multi-indicateur $D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$, sont égaux à 1 pour toute date où les valeurs des indicateurs associés sont supérieurs au seuil de détection. Le détecteur D_{VU} est quant à lui égal à 1 à une date donnée si tous les détecteurs mono-indicateur disponibles sont égaux à 1 et D_{VM} est égal à 1 si la majorité des détecteurs mono-indicateur disponibles sont égaux à 1. Les détecteurs

associés aux approches mono-indicateur sont identifiés par une couleur bleue, tandis qu'une couleur rouge est utilisée pour les détecteurs multi-indicateur. Les zones grises correspondent à des périodes d'indisponibilité du détecteur considéré. Les zones colorées représentent les interventions (périodes où le détecteur considéré est égal à 1 sur au moins une journée), et la première intervention enregistrée par chaque approche de détection est représentée par une couleur plus vive que les interventions suivantes, afin de mettre en évidence l'avance à la détection (ADT).

De même que pour l'analyse de la répartition des interventions inutiles et des taux de faux positifs sur H_0 (Figure 6.6), on remarque que la proportion et la répartition des périodes de détection sont différentes d'une approche de surveillance conditionnelle à une autre (sauf pour les détecteurs issus des indicateurs moyens et médians $D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$, qui présentent des tendances de détection très similaires). On observe notamment que tout comme sur H_0 , D_{VU} est l'approche présentant le moins d'interventions sur H_1 . On remarque cependant que ce détecteur identifie le défaut bien plus tard que les autres approches, et que D_{VM} le détecte 19 jours plus tôt. Cela s'explique par le fait que $D_{e_{MD}}$ et $D_{\varepsilon_{X_r}}$ ne détectent le défaut avec persistance (c'est-à-dire pendant au moins une journée) qu'à partir de la seconde moitié du mois de Février, ce qui rend impossible la détection du défaut par D_{VU} les jours précédents de par sa définition (équation 6.10). Néanmoins, tout comme les 6 autres approches de détection, le détecteur D_{VU} permet une avance à la détection supérieure à 7 jours, et respecte ainsi le cahier des charges.

Sur la base de ce cas d'étude, il conviendrait donc de retenir D_{VU} comme meilleure approche de surveillance conditionnelle pour une application industrielle.

6.3.2 Surveillance thermique du convertisseur de N_f^{rc}

Modèles et variations des sept indicateurs

Le Tableau 6.3 présente les résultats de sélection de régresseurs du processus LASSO-GFS par application de la méthodologie de génération d'indicateurs définie en section 6.2.1 pour la modélisation thermique du convertisseur de N_f^{rc} .

Nous pouvons tout d'abord constater que la variable $t_{\text{convertisseur2}}$, responsable de la non-sensibilité de l'indicateur multi-turbine, est également retenue dans le modèle X_1 . Par conséquent, l'indicateur multi-turbine associé devrait présenter des performances de détection similaires sur H_1 .

Cependant, les autres applications du processus LASSO-GFS réalisées permettent d'exclure cette variable des régresseurs. À la place, on note un choix de variables présentant divers liens avec le convertisseur : la présence pour trois modèles de la tension produite par la génératrice ($v_{\text{simpleProduite}}$),

qui est un signal d'entrée du convertisseur, la vitesse du vent ($v_{\text{ventNacelle}}$), qui peut être utilisée pour prédire les variations thermiques des composants principaux de la nacelle comme nous l'avons expliqué en section 6.2.1, ainsi que la température de refroidissement de la génératrice $t_{\text{generatrice}}$, retenue pour 4 des 5 modèles générés. Le lien entre les variations thermiques du convertisseur et celles de la génératrice s'explique par effet Joule : la température de la génératrice augmente avec la puissance produite par l'éolienne. Ainsi, plus la génératrice s'échauffe, plus le courant électrique injecté dans le convertisseur est important, ce qui échauffe le composant. Enfin, nous remarquons que les modèles X_2 et X_4 sont identiques. Par conséquent, les indicateurs associés posséderont les mêmes variations, et donc les mêmes performances de détection sur H_0 et H_1 .

Identifiant	x_1	x_2	x_3
X_0	$t_{\text{convertisseur2}}$	t_{transfo}	$t_{\text{huileMulti}}$
X_1	$t_{\text{convertisseur2}}$	$t_{\text{generatrice}}$	$v_{\text{ventNacelle}}$
X_2	$t_{\text{generatrice}}$	$v_{\text{simplProduite}}$	$v_{\text{ventNacelle}}$
X_3	$t_{\text{generatrice}}$	$v_{\text{simplProduite}}$	$t_{\text{huileMulti}}$
X_4	$t_{\text{generatrice}}$	$v_{\text{simplProduite}}$	$v_{\text{ventNacelle}}$

Tableau 6.3: Régresseurs générés par le processus de sélection LASSO-GFS, considérés pour l'approche multi-indicateur de la modélisation thermique du convertisseur de N_f^{rc}

La Figure 6.8 illustre les variations des sept indicateurs définis en section 6.2.1 pour la modélisation thermique du convertisseur en défaut du parc N_f^{rc} .

Comme nous l'avons anticipé par analyse des régresseurs retenus pour l'approche multi-indicateur (Tableau 6.3), ε_{X_1} présente des variations très proches de ε_{X_0} sur H_1 (Figure 6.8b), et hérite ainsi de sa faible sensibilité au défaut. Toutefois, tous les autres indicateurs linéaires générés sont significativement plus sensibles au défaut, et possèdent des valeurs résiduelles plus grandes que ε_{X_0} et ε_{X_1} sur la totalité de la période H_1 . Nous pouvons enfin souligner que les valeurs des indicateurs présentant un gain de sensibilité sur H_1 sont du même ordre de grandeur que ε_{X_0} et ε_{X_1} sur les données de bon fonctionnement H_0 (Figure 6.8a). Par conséquent, la majorité des indicateurs générés selon la méthodologie détaillée en section 6.2.1 semblent fournir un gain de performances de détection considérable par rapport à l'approche mono-indicateur étudiée dans le chapitre 5.

La Figure 6.9 représente les matrices de corrélation (en utilisant le coefficient de Pearson R^2) des sept indicateurs sur H_0 (6.9a) et H_1 (Figure 6.9b).

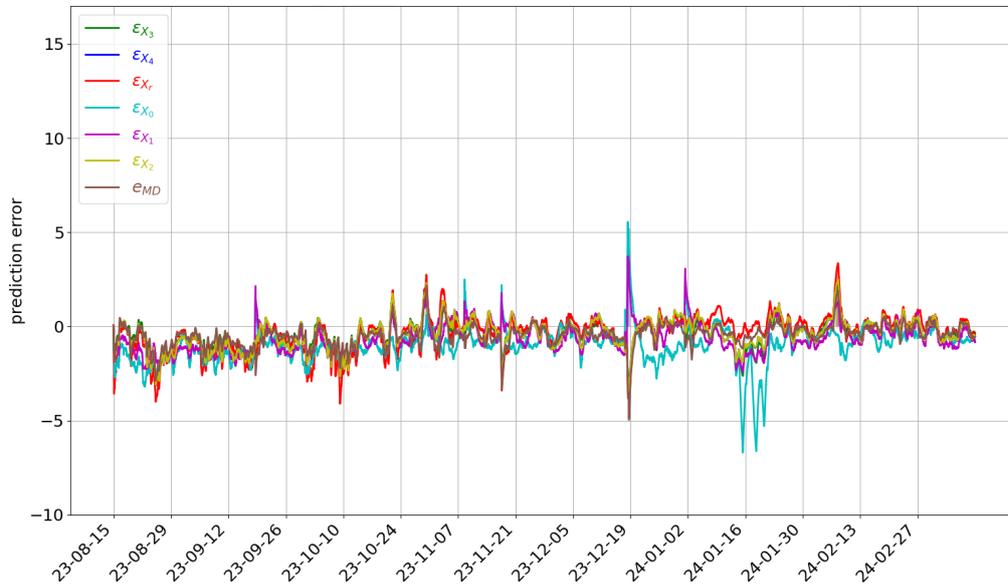
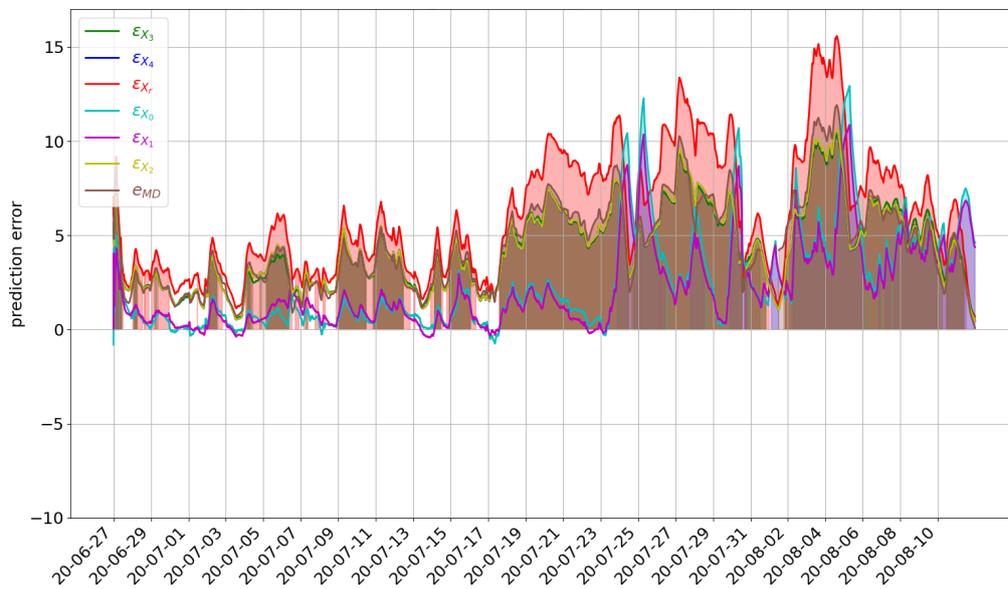
(a) H_0 (b) H_1

Figure 6.8: Variations des sept indicateurs définis en section 6.2.1 pour la modélisation thermique du convertisseur de N_f^{rc}

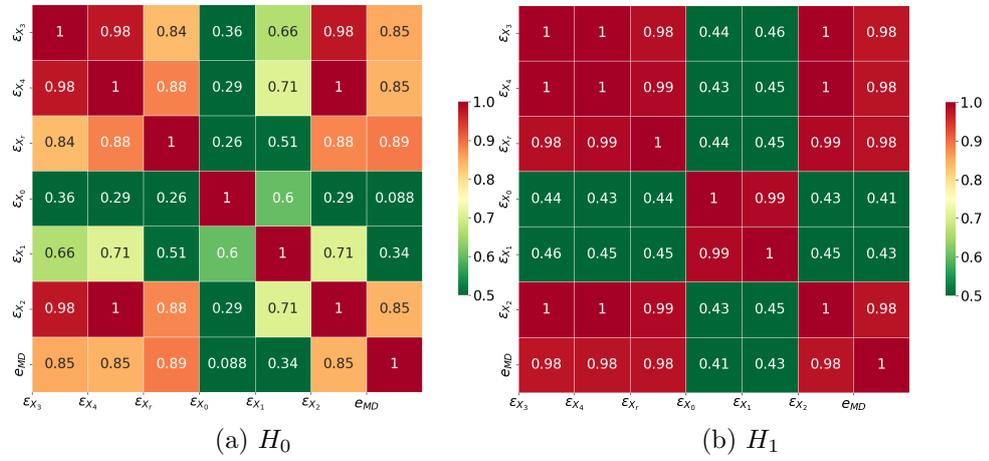


Figure 6.9: Matrices de corrélation des 7 indicateurs définis en section 6.2.1 pour la modélisation thermique du convertisseur de N_f^{rc}

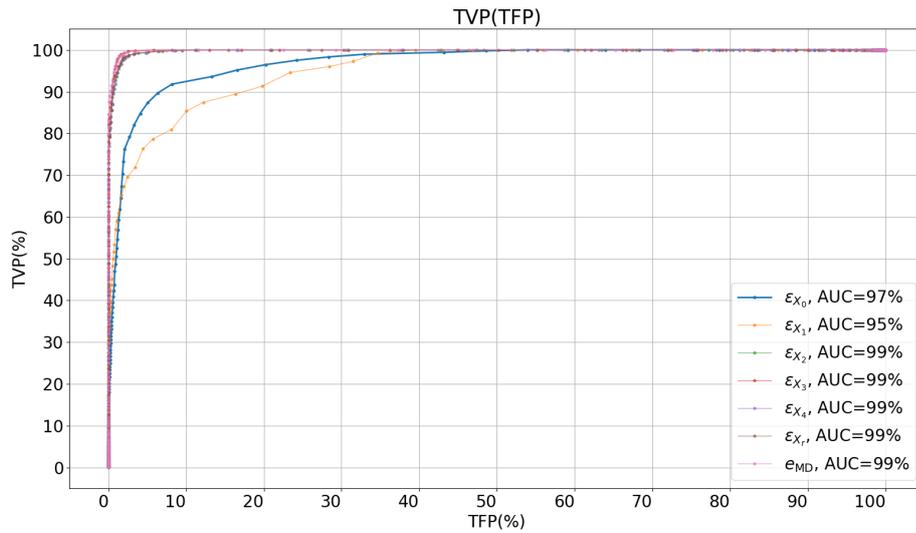


Figure 6.10: Courbe ROC du taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP) pour les sept indicateurs proposés, appliquée aux données du parc N_f^{rc}

La matrice de corrélation sur les données de la période H_1 (Figure 6.9b) permet de visualiser l'isolement des indicateurs ε_{X_0} et ε_{X_1} par rapport aux cinq autres indicateurs linéaires générés, principalement causé par la présence de $t_{\text{convertisseur}2}$ parmi leurs régresseurs. Sur les données de bon fonctionnement H_0 (Figure 6.9a), tout comme pour le cas du parc E_f^{rg2} , on

observe une indépendance des indicateurs très variable : les indicateurs sont dans l'ensemble moins corrélés entre eux sur H_0 que sur H_1 , mais certains restent très fortement corrélés, notamment ε_{X_3} et ε_{X_4} . Cette corrélation s'explique par le contenu des modèles utilisés, qui partagent deux régresseurs en commun ($t_{\text{generatrice}}$, $v_{\text{simplProduite}}$) d'après le Tableau 6.3.

La courbe de caractéristiques de performance (ou courbe ROC), présentée en Figure 6.10, permet de comparer les caractéristiques de performances des sept indicateurs, du point de vue des taux de positifs.

Ces courbes de caractéristiques de performance permettent de confirmer les observations que nous venons de réaliser : ε_{X_0} et ε_{X_1} présentent les moins bonnes courbes ROC (ε_{X_1} fait notamment baisser l'AUC de 2%). Tous les autres indicateurs réalisent une bien meilleure distinction des données en bon fonctionnement (H_0) des données en défaut (H_1), et leurs courbes ROC associées sont presque optimales.

Pour résumer les analyses réalisées, nous pouvons affirmer que la méthodologie développée en section 6.2.1 permet de générer une majorité d'indicateurs linéaires sensible au défaut étudié, tout en assurant une forte séparation des données résiduelles en bon fonctionnement (H_0) avec les données résiduelles en période de défaut (H_1). Nous allons voir dans la section suivante si ces affirmations s'accompagnent d'un meilleur respect du cahier des charges défini dans le chapitre 5, en analysant la répartition des interventions sur H_0 et sur H_1 pour les approches mono-indicateur et multi-indicateur étudiées.

Analyse des performances des détecteurs mono-indicateur et multi-indicateur

La Figure 6.11 représente la répartition des interventions inutiles par élément modélisé du parc N_f^c , représenté en axe des abscisses, pour les approches de détection mono-indicateur et multi-indicateur, distinguées par groupe de couleur.

Seuls les composants de 5 des 10 turbines du parc sont intégrés au graphique par soucis de clarté de lecture de l'axe des abscisses. De même que pour le cas du parc E_f^{rg2} (Figure 6.6), on remarque que la répartition des interventions inutiles n'est pas uniforme, et dépend de la modélisation considérée. Par exemple, la surveillance thermique du roulement arrière de la génératrice (désigné par "rou-ice") de la turbine T_4 occasionne la génération d'un peu plus de 100 interventions inutiles en faisant la somme des sept approches de détection considérées (soit plus de 14 par approche), alors que pour ce même composant pour la turbine T_3 , aucune intervention inutile n'est comptabilisée.

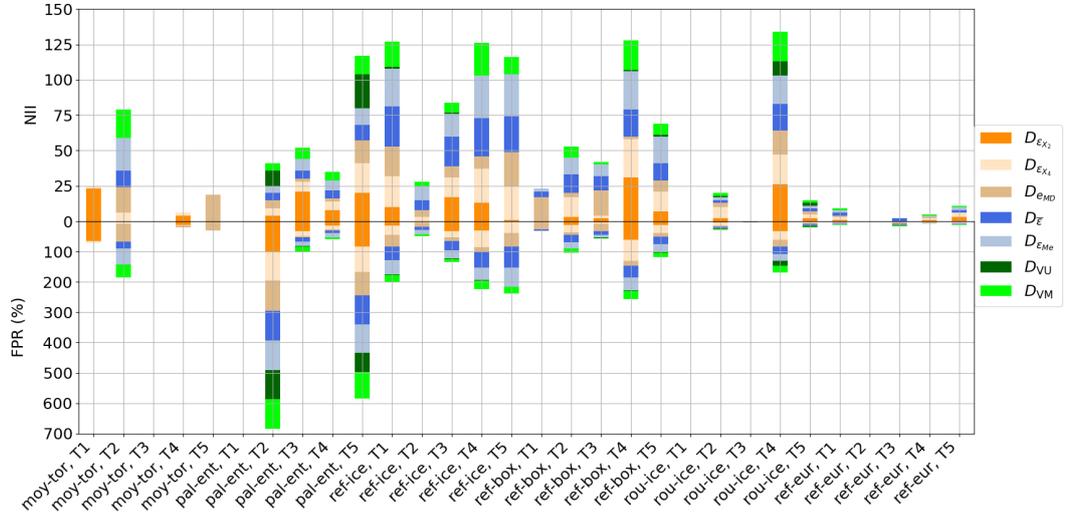


Figure 6.11: Répartition des interventions inutiles (NII) et des taux de faux positifs (TFP) par élément modélisé du parc N_f^{rc} sur H_0 pour les approches de détection mono-indicateur et multi-indicateur

En outre, pour une même modélisation, chaque approche de détection fournit un bilan d'interventions inutiles propre, et tout comme pour le cas du parc E_f^{rg2} , les détecteurs D_{VM} et D_{VU} semblent être ceux fournissant tendanciellement le moins d'interventions inutiles. On peut notamment relever à la lecture de cette figure que les approches de détection $D_{\bar{\epsilon}}$ et $D_{E_{Me}}$ ne semblent pas fournir un NII^F plus petit que les approches mono-indicateur. Enfin, la Figure 6.11 permet de voir que le TFP n'est pas toujours corrélé avec le NII, notamment pour le cas de la surveillance du palier de l'arbre lent ("pal-ent") de la turbine T_2 , où pour chaque méthode de détection évaluée, le TFP est proche de 100%, alors que les bilans d'intervention inutiles associés ne sont pas les plus conséquents de l'ensemble des modélisations étudiées.

Ces valeurs de TFP (qu'on peut également observer pour le cas de la surveillance du palier de l'arbre lent de T_5) s'expliquent par un saut de valeur moyenne des indicateurs multi-turbine entre H_L et H_0 , comme nous pouvons le voir sur la Figure 6.12, affichant les valeurs des indicateurs multi-turbine pour la surveillance de ce composant pour T_2 durant la période d'apprentissage et durant la période H_0 . Nous pouvons notamment remarquer que ce saut de moyenne concerne également e_{MD} , signifiant que durant H_0 , la température du palier de l'arbre lent est plus grande pour la turbine T_2 que pour les autres turbines du parc, contrairement à H_L où les températures sont proches d'après la Figure 6.12a (e_{MD} de moyenne nulle).

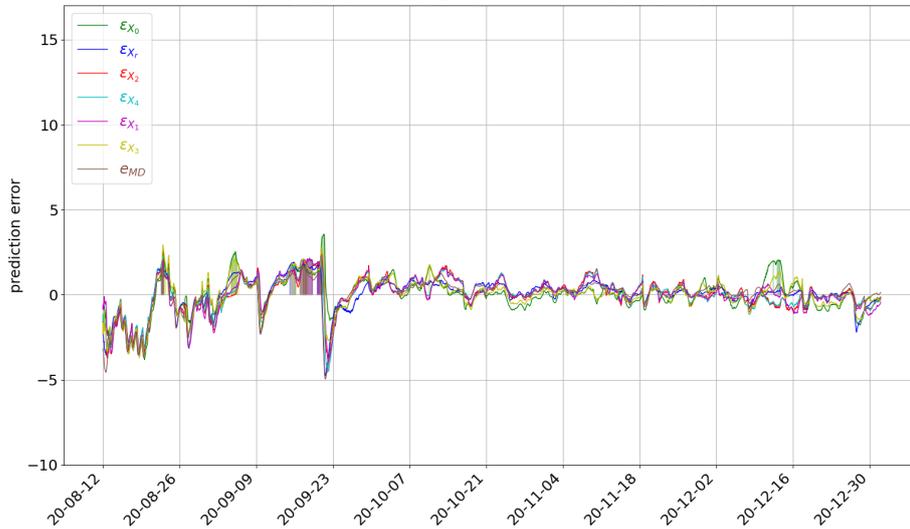
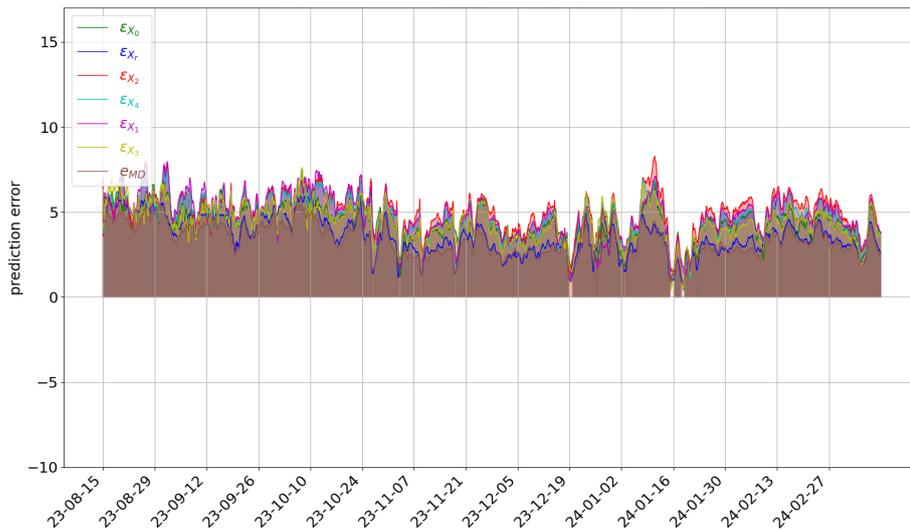
(a) H_L (b) H_0

Figure 6.12: Variations des sept indicateurs pour la modélisation thermique du palier de l'arbre lent de N_f^{rc} sur H_L et H_0

Nous pouvons donc relier ces sauts de moyennes des indicateurs multi-turbine à un échauffement progressif de la température du palier de l'arbre lent de T_2 (la variable de sortie y de tous les modèles) par rapport aux autres turbines du parc, démontrant une nécessité de réapprentissage des modèles

généérés afin de maintenir une précision d'estimation stable en période de bon fonctionnement (dans le cas où cet échauffement est jugé "naturel" par les opérateurs industriels).

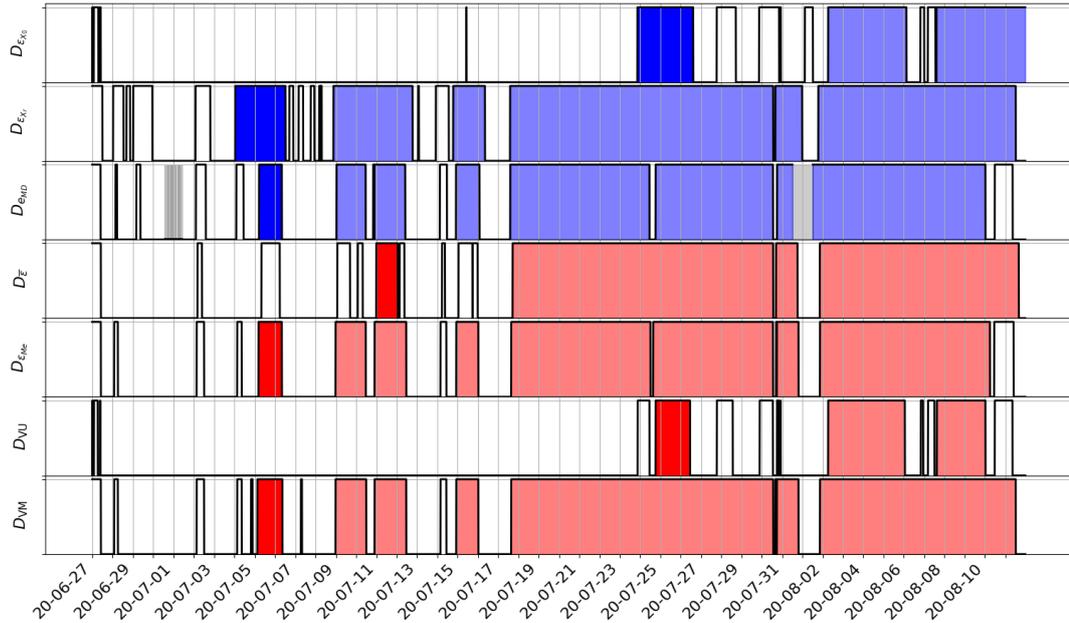


Figure 6.13: Visualisation des détecteurs des approches mono-indicateur (en bleu) et multi-indicateur (en rouge) pour la modélisation thermique du convertisseur de N_f^{rc} sur H_1

Comme nous l'avons observé sur la Figure 6.8, la majorité des indicateurs générés occasionnent un grand gain de sensibilité sur H_1 . Cette observation se confirme à la lecture du bilan de détection illustré par la Figure 6.13 : les deux approches mono-indicateur $D_{\varepsilon_{X_r}}$ et $D_{\varepsilon_{MD}}$ permettent de détecter le défaut plusieurs semaines avant $D_{\varepsilon_{X_0}}$, avec une forte persistance de cette détection jusqu'à la fin de la période H_1 . Cet accroissement des valeurs résiduelles en période de défaut de la majorité des indicateurs linéaires s'accompagne d'un gain des valeurs moyennes et médianes de ces sept indicateurs ($\bar{\varepsilon}$ et ε_{Me}), permettant d'assurer des bonnes performances de détection par les approches multi-indicateur associés $D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$. Il en va de même pour l'approche multi-indicateur fondée sur le vote majoritaire D_{VM} , qui présente un profil de détection similaire à celui de $D_{\varepsilon_{MD}}$. Cependant, à cause des mauvaises performances de détection des indicateurs $D_{\varepsilon_{X_0}}$, et $D_{\varepsilon_{X_1}}$, D_{VU} ne détecte le défaut que le 25 Juillet, date à laquelle $D_{\varepsilon_{X_1}}$ détecte le défaut. Néanmoins, le critère du cahier des charges reste respecté

pour toutes les approches de détection évaluées : l'avance à la détection mesurée est supérieure à 7 jours.

À travers l'exemple des données des parcs E_f^{rg2} et N_f^{rc} , nous avons vu par analyse des variations des sept indicateurs définis dans la section 6.2.1 que les hypothèses d'indépendance des valeurs des indicateurs sur H_0 réalisée en section 6.1.2 et d'existence de plusieurs indicateurs sensibles aux défauts durant H_1 étaient vérifiées. Nous avons également démontré que les approches de détection multi-indicateurs D_{VM} et D_{VU} semblaient être les plus adaptées à un objectif de respect complet du cahier des charges défini en section 5.3.4. En effet, ce sont les seules qui permettent une réduction drastique du nombre d'interventions inutiles sur H_0 tout en conservant une avance à la détection suffisante (supérieure à 7 jours) sur H_1 .

Toutefois, ces conclusions ne concernent que l'étude des défauts des parcs E_f^{rg2} et N_f^{rc} . Nous allons ainsi dans la section suivante procéder à une analyse comparative de toutes les métriques de mesure de performances définies en section 5.3 par approche de classification, pour les 4 parcs impactés par un défaut. L'objectif étant de déterminer la solution de surveillance conditionnelle la plus adaptée à un usage industriel, en considération du cahier des charges défini en section 5.3.4 et des besoins et contraintes identifiées en section 2.3.3.

6.3.3 Analyse comparative des métriques de performances

Les Tableaux 6.4 à 6.7 relèvent les métriques de performances des approches de surveillance conditionnelle mono-indicateur et multi-indicateur sur les données des parcs respectifs E_f^{rg2} , M_f^{al} , M_f^{rm} et N_f^{rc} . Ces métriques ont été définies dans le chapitre 5, et seront comparées pour l'utilisation des méthodes de détection mono-indicateur et multi-indicateur proposées afin de déterminer la ou les approches permettant de respecter au mieux le cahier des charges défini. Cette comparaison sera réalisée sur la base des résultats du détecteur $D_{\varepsilon_{X_0}}$, utilisé dans le chapitre 5 : on cherchera ainsi une méthode de détection permettant la réduction du NII^F pour les 4 cas de modélisation, une augmentation de la disponibilité pour le parc M_f^{al} et de la sensibilité au défaut (TVP) pour le parc N_f^{rc} .

Surveillance du roulement arrière de la génératrice de E_f^{rg2}

L'approche de surveillance mono-indicateur présentée dans le chapitre 5 occasionne un **nombre d'interventions inutiles** calculé à l'échelle du parc trop grand ($NII^F > 1$). On remarque sur le Tableau 6.4 que toutes les approches de détection multi-indicateur permettent de réduire cette métrique pour le cas de la surveillance du roulement arrière de la génératrice de E_f^{rg2} . Cependant, le NII^F reste trop grand pour les méthodes de détection multi-

indicateur $D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$, et seules les approches multi-indicateur D_{VU} et D_{VM} permettent de respecter le cahier des charges, en présentant un NII^F inférieur ou égal à 1, et nul pour le détecteur D_{VU} , comme nous l'avons observé au préalable en section 6.3.1, ce qui représente un résultat idéal pour un industriel car impliquant un surcoût de maintenance minimale.

Approche	détecteur	μ_{H_0}	var_{H_0}	μ_{H_1}	var_{H_1}	TFP(%)	dsp_{H_0} (%)	TVP(%)	dsp_{H_1} (%)	NII ^F	ADT
mono-indicateur	$D_{\varepsilon_{X_0}}$	0.02	0.97	5.53	15.73	1.44	99.71	71.25	98.17	4	29
	$D_{\varepsilon_{X_r}}$	0.25	0.80	4.83	4.0	0.0	99.76	71.05	37.49	4	11
	$D_{\varepsilon_{MD}}$	-0.16	0.86	1.97	1.39	0.05	99.76	35.48	98.17	3	16
multi-indicateur	$D_{\mathcal{F}}$	0.11	0.53	3.93	5.70	0.2	99.76	63.62	98.17	3	29
	$D_{\varepsilon_{Me}}$	0.13	0.56	4.02	5.60	0.3	99.76	64.6	98.17	2	29
	D_{VU}	non défini	non défini	non défini	non défini	0.0	99.76	30.9	98.17	0	9
	D_{VM}	non défini	non défini	non défini	non défini	0.0	99.76	51.95	98.17	1	27

Tableau 6.4: Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation du roulement arrière de la génératrice de E_f^{rg2}

Concernant la sensibilité au défaut, $D_{\varepsilon_{X_0}}$ présente le plus grand **taux de vrais positifs**, et les résultats sont très variables pour toutes les approches de détection considérées. On remarque notamment que le détecteur fondé sur le vote unanime (D_{VU}) ne respecte pas le critère du cahier des charges concernant le TVP, de même que l'écart à la médiane e_{MD} .

De plus, on peut constater que D_{VU} occasionne une **avance à la détection** très proche du cahier des charges (9 jours). Or, comme expliqué dans le chapitre 5 lors de la définition de ce critère du cahier des charges, l'objectif d'avance à la détection de 7 jours représente un contexte "normal" d'exploitation d'une flotte d'éolienne. Il peut en pratique être revu à la hausse dans des conditions de forte sollicitation de l'entreprise O&M où l'organisation spontanée d'une intervention de maintenance préventive sous une semaine n'est pas possible. Par conséquent, il est risqué de retenir en application industrielle l'approche de surveillance conditionnelle D_{VU} , fondée sur le vote unanime des détecteurs associés aux sept indicateurs.

En conclusion, sur la base des données du parc E_f^{rg2} , le détecteur multi-indicateur D_{VM} se positionne comme la meilleure approche de détection car permettant de respecter tous les éléments du cahier des charges, tout en fournissant une avance à la détection proche du mois, permettant la planification d'une maintenance préventive peu importe le contexte d'exploitation de l'entreprise.

Surveillance du palier de l'arbre lent de M_f^{al}

Comme on peut le voir sur le Tableau 6.5, le détecteur généré à partir du modèle physique X_r présente le même problème de disponibilité sur H_1 que $D_{\varepsilon_{X_0}}$. Cela est dû à la faible disponibilité de la température extérieure (Figure 4.5), régresseur de X_r . Toutefois, toutes les autres méthodes de détection

évaluées permettent un gain de presque 50% de disponibilité, accompagné d'un gain d'avance à la détection de 200 jours.

Approche	détecteur	μ_{H_0}	var_{H_0}	μ_{H_1}	var_{H_1}	TFP(%)	dsp $_{H_0}$ (%)	TVP(%)	dsp $_{H_1}$ (%)	NII ^F	ADT
mono-indicateur	$D_{\varepsilon_{X_0}}$	-0.39	2.47	11.19	12.01	0.0	60.53	99.50	24.75	3	90
	$D_{\varepsilon_{X_r}}$	0	2.77	4.01	4.06	2.25	60.53	100	24.75	2	90
	D_{EMD}	0.02	2.02	2.24	2.21	0.0	60.53	69.35	71.16	2	261
multi-indicateur	D_T	-0.14	1.96	3.74	6.67	0.41	60.53	72.9	71.33	2	280
	D_{Mc}	-0.19	2.38	3.88	7.07	0.14	60.53	66.26	71.33	2	280
	D_{VU}	non défini	non défini	non défini	non défini	0.0	60.53	66.01	71.33	1	261
	D_{VM}	non défini	non défini	non défini	non défini	0.09	60.53	69.8	71.33	1	278

Tableau 6.5: Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation du roulement principal de M_f^{al}

Cela signifie que les autres modèles générés pour l'approche multi-indicateur ne contiennent ni $t_{roulMulti1}$ (contenu dans le modèle X_0), ni $t_{exterieure}$, toutes deux très peu disponibles sur H_1 . Les détecteurs associés sont par conséquent bien plus disponibles que $D_{\varepsilon_{X_0}}$ et $D_{\varepsilon_{X_r}}$. Il en va de même pour les approches de détection multi-indicateur, qui n'ont besoin que d'un indicateur disponible pour être appliquées. Il est à noter que plus une intervention est réalisée à un stade de développement peu avancé du défaut, moins le coût de la maintenance est élevé. Ainsi, si du point de vue du respect du cahier des charges défini, l'analyse d'une avance à la détection de 90 jours ou 250 sera la même, en application industrielle, un utilisateur préférera la méthode de détection occasionnant le plus long délai de planification de maintenance préventive.

Enfin, on remarque que les détecteurs les plus disponibles sont également ceux présentant les taux de vrais positifs les plus bas (bien que supérieurs à 50%, respectant ainsi le cahier des charges). En effet, les taux de positifs de ces détecteurs étant mesurés à partir d'un plus grand nombre de valeurs d'indicateurs, ils sont plus précis. Or, les indicateurs linéaires construits sont "imparfaits" du fait des approximations de modélisation linéaire sur lesquelles ils se basent, et comme les périodes de bon fonctionnement sélectionnées peuvent contenir des perturbations impactant les valeurs résiduelles, augmenter le nombre de données résiduelles utilisées pour calculer les taux de positifs contribue à diminuer le TVP sur H_1 , car on peut intégrer des périodes de non-détection du défaut, et à augmenter le TFP sur H_0 , car on peut intégrer des périodes de perturbations en bon fonctionnement.

Tout comme pour le cas du parc E_f^{rg2} , il n'y a que les approches D_{VU} et D_{VM} qui permettent d'atteindre un NII^F inférieur ou égal à 1, et l'avance à la détection reste supérieure à 7 jours pour toutes les méthodes de détection considérées. Toutefois, on peut observer un grand gain d'avance à la détection de presque 200 jours pour une majorité des détecteurs.

Ainsi, pour ce cas d'étude, les deux approches de détection D_{VU} et D_{VM} sont les seules permettant de respecter tous les critères du cahier des charges.

Surveillance du système de refroidissement de la multiplicativeur de M_f^{rm}

Approche	détecteur	μ_{H_0}	var_{H_0}	μ_{H_1}	var_{H_1}	TFP(%)	dsP_{H_0} (%)	TVP(%)	dsP_{H_1} (%)	NII ^F	ADT
mono-indicateur	D_{X_0}	0.55	0.83	2.26	2.99	9.57	86.62	59.29	90.9	3	219
	D_{X_c}	0.99	3.38	4.36	9.53	9.22	86.62	63.34	90.9	3	219
	D_{MD}	0.81	2.41	3.42	6.68	6.19	99.42	52.35	97.99	3	219
	$D_{\bar{\epsilon}}$	0.82	2.20	3.51	6.23	5.75	99.45	53.26	98.23	5	219
multi-indicateur	D_{Me}	0.84	2.57	3.60	6.77	6.57	99.45	53.97	98.23	4	219
	D_{VU}	non défini	non défini	non défini	non défini	3.36	99.45	44.39	98.23	0	219
	D_{VM}	non défini	non défini	non défini	non défini	3.36	99.45	54.82	98.23	1	219

Tableau 6.6: Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation thermique de la multiplicativeur de M_f^{rm}

Concernant le **nombre d'intervention inutiles**, tout comme pour les cas précédents des parcs E_f^{rg2} et M_f^{al} , il n'y a que les approches D_{VU} et D_{VM} qui permettent d'atteindre un NII^F inférieur ou égal à 1. On remarque notamment que les détecteurs fondés sur la valeur moyenne ($D_{\bar{\epsilon}}$) et médiane ($D_{\epsilon_{Me}}$) des sept indicateurs fournissent un NII^F plus élevé que les approches mono-indicateur, et sont donc à exclure. Les sensibilités au défaut (TVP) restent assez basse (proches du cahier des charges) pour toutes les approches de détection évaluées. Cependant, le détecteur fondé sur le vote unanime (D_{VU}) fourni un **taux de vrais positifs** inférieur à 50%, démontrant une nouvelle fois sa plus faible sensibilité en période H_1 .

Ainsi, comme pour le cas d'étude du parc E_f^{rg2} , D_{VM} est ici la seule méthode de détection permettant de respecter entièrement le cahier des charges.

Surveillance du système de refroidissement du convertisseur de N_f^{rc}

Approche	détecteur	μ_{H_0}	var_{H_0}	μ_{H_1}	var_{H_1}	TFP(%)	dsP_{H_0} (%)	TVP(%)	dsP_{H_1} (%)	NII ^F	ADT
mono-indicateur	D_{X_0}	-0.99	0.87	2.45	6.98	0.32	100	30.97	100	5	19
	D_{X_c}	-0.34	0.87	6.07	10.73	0.19	100	79.18	100	5	37
	D_{MD}	-0.47	0.36	4.55	5.73	0.0	100	67.28	100	5	25
	$D_{\bar{\epsilon}}$	-0.52	0.35	4.04	4.49	0.0	100	61.91	100	6	24
multi-indicateur	D_{Me}	-0.45	0.39	4.31	4.63	0.0	100	66.77	100	6	24
	D_{VU}	non défini	non défini	non défini	non défini	0.0	100	25.15	100	1	18
	D_{VM}	non défini	non défini	non défini	non défini	0.0	100	66.78	100	4	25

Tableau 6.7: Métriques de performance des approches de détection mono-indicateur et multi-indicateur pour la modélisation thermique du convertisseur de N_f^{rc}

Sur le Tableau 6.7, nous pouvons remarquer que le taux de vrais positifs est fortement augmenté pour la majorité des approches de détection. Cela signifie qu'à partir d'une même base de données d'apprentissage, il est possible de générer plusieurs indicateurs sensibles à un même défaut. Cependant, toutes

les approches de détection multi-indicateur ne permettent pas d'assurer une sensibilité au défaut suffisante. On remarque notamment que, bien qu'étant la seule approche de détection permettant une réduction du NII^F suffisante pour respecter le cahier des charges (≤ 1), le détecteur multi-indicateur D_{VU} possède de mauvaises performances de détection sur H_1 , avec un très faible TVP (25.15%).

Au contraire, le vote majoritaire (D_{VM}) est bien plus sensible au défaut, et occasionne un taux de vrais positifs respectant le cahier des charges (66.78%), et permet de détecter le défaut une semaine plus tôt que D_{VU} . Cependant, D_{VM} occasionne 4 **interventions inutiles** en moyenne pour la surveillance du parc complet, ce qui est trop grand en considération du cahier des charges défini.

D'après l'analyse comparative des métriques de performances réalisée pour les 4 composants en défaut, il ressort qu'aucune des approches de surveillance conditionnelle mono-indicateur et multi-indicateur ne permet de respecter l'intégralité des critères du cahier des charges pour les 4 bases de données de parcs étudiés.

Toutefois, cette analyse comparative au cas par cas a permis de démontrer que le détecteur construit à partir du vote majoritaire D_{VM} présente le meilleur bilan de performance : il permet une amélioration considérable du cahier des charges par rapport à l'approche mono-indicateur présentée au chapitre 5, en présentant pour les 4 cas étudiés un taux de vrais positifs et une disponibilité toujours supérieures à 50% sur H_1 , ainsi qu'une avance à la détection largement supérieure à 7 jours. Toutefois, bien que ce détecteur permette une réduction du NII^F pour tous les cas, le critère du cahier des charges n'est pas respecté pour le cas du parc N_f^{rc} .

Cependant, après avoir expliqué les écarts de NII^F entre les approches de surveillance D_{VU} et D_{VM} , nous expliquerons dans la section suivante que, du fait des faibles connaissances du contenu des périodes de bon fonctionnement H_L et H_0 , l'analyse et l'interprétation du NII^F ne sont pas triviales. Nous verrons en effet qu'il est difficile d'assurer un nombre d'interventions inutiles à l'échelle du parc inférieur ou égal à 1 pour tous les parcs considérés, et qu'un tel résultat peut s'accompagner d'un filtrage des perturbations qui devraient être identifiées par le système de surveillance conditionnelle proposé.

6.3.4 Discussions autour de la notion d'"intervention inutile"

Cas particuliers des détecteurs D_{VM} et D_{VU}

Il a été observé à travers l'analyse comparative des performances des approches de surveillance conditionnelle mono-indicateur et multi-indicateur

que les approches basées sur les votes des détecteurs mono-indicateur (D_{VM} et D_{VU}) étaient les seules approches multi-indicateur permettant de réduire de manière significative le nombre d'interventions inutiles. En comparaison, les approches multi-indicateur fondées sur les valeurs moyenne et médiane des indicateurs résiduels ($D_{\bar{\varepsilon}}$ et $D_{\varepsilon_{Me}}$) fournissent un bilan d'intervention inutile du même ordre de grandeur que les approches mono-indicateur.

Pour comprendre ce constat, considérons une anomalie survenant en période de bon fonctionnement H_0 . Cette perturbation entraîne un écart de valeurs au niveau du ou des indicateurs qui y sont sensibles. L'amplitude de cet écart dépend de la nature de la perturbation, et peut donc être très grande. Considérons maintenant que la moitié des indicateurs disponibles soient sensibles à la perturbation sur cette période, et présentent des valeurs suffisamment hautes pour que la valeur médiane de l'indicateur ε_{Me} soit au-dessus de son seuil de détection : bien qu'une part non majoritaire d'indicateurs soit sensible à la perturbation, le détecteur multi-indicateur $D_{\varepsilon_{Me}}$ générera ainsi une alarme sur cette période. Concernant l'indicateur $\bar{\varepsilon}$, fondé sur la moyenne des sept indicateurs, il suffit qu'un de ces indicateurs soit sensible à la perturbation et présente des valeurs très hautes pour que la valeur moyenne soit impactée, au risque de générer une intervention inutile.

Si l'on considère les votes des détecteurs mono-indicateur, ceux qui sont associés à la part d'indicateurs sensibles à la perturbation auront une valeur binaire égale à 1 sur la période associée (valeurs supérieures au seuil de détection, sans prise en compte de l'amplitude de la déviation de valeurs), et les autres à 0 : les votes majoritaires (D_{VM}) et unanimes (D_{VU}) convergeront alors vers la non-présence d'un défaut (valeur binaire égale à 0).

Les approches de surveillance conditionnelle D_{VU} et D_{VM} permettent ainsi une réduction drastique du nombre d'interventions inutiles en période de bon fonctionnement, mais elles présentent des sensibilités très différentes en période de défaut H_1 . L'analyse comparative réalisée en section 6.3.3 a en effet mis en avant une très faible sensibilité au défaut (faible TVP et ADT très proche du cahier des charges) de l'approche D_{VU} en comparaison de celle fondée sur le vote majoritaire des détecteurs des sept indicateurs.

Ce résultat s'explique par la définition du détecteur multi-indicateur D_{VU} : pour identifier un défaut (valeur à 1), ce détecteur a besoin que tous les indicateurs disponibles soient sensibles à ce défaut. Cette définition n'est pas cohérente avec les motivations de mise en œuvre d'une approche multi-indicateurs : il a en effet été souligné au début de ce chapitre que le processus de surveillance conditionnelle présenté dans le chapitre 5 est susceptible de fournir des indicateurs insensibles aux défauts, avec l'exemple de la détection du défaut du système de refroidissement du convertisseur de N_f^{rc} .

En comparaison, le détecteur construit à partir du vote majoritaire D_{VM} présente l'avantage d'être robuste aux limitations du processus de surveillance mono-indicateur : la présence d'un indicateur insensible à un

défaut dans l'ensemble des sept indicateurs ne rend pas le détecteur D_{VM} systématiquement insensible au défaut. Par conséquent, l'utilisation du détecteur D_{VM} permet une plus grande persistance de la détection en période de défaut, ce qui explique l'écart de taux de vrais positifs et d'avance à la détection avec D_{VU} .

Calcul du NII^F pour l'ensemble de la flotte expérimentale

Les analyses de la section 6.3.3 ont montré que D_{VM} était l'approche de détection multi-indicateur permettant de respecter au mieux le cahier des charges défini en section 5.3.4 pour les quatre parcs considérés. Cependant, pour le cas de la surveillance du système de refroidissement de la turbine en défaut du parc N_f^{rc} , l'utilisation de ce détecteur occasionne un NII^F trop important, et ne respecte ainsi pas le cahier des charges.

Afin de déterminer si ce résultat est isolé ou si l'utilisation de D_{VM} pouvait générer trop d'interventions inutiles pour d'autres parcs de la flotte VALEMO, le Tableau 6.8 présente les NII^F obtenus pour les 11 parcs de la sous-flotte (Tableau 4.1) utilisée dans ces travaux de recherche, pour les approches mono-indicateur et multi-indicateur analysées dans ce chapitre, appliquées aux données de l'année 2023 (H_0). Les modèles, ainsi que les seuils de détection ont été appris à partir des données de l'année 2022 (H_L). La méthodologie de génération d'indicateurs linéaires présentée en section 6.2.1 a été appliquée sur H_L , fournissant pour chacun des parcs de la sous-flotte un ensemble de modèles uniques pour la surveillance des variables de sortie disponibles.

Approche	détecteur	E_f^{rg2}	E_1	M_f^{rm}	M_f^{al}	M_1	M_2	M_3	M_4	M_5	N_f^{rc}	N_1
mono-indicateur	$D_{\epsilon_{X_0}}$	4	5	3	3	4	3	4	4	3	5	1
	$D_{\epsilon_{X_r}}$	4	2	3	2	4	6	5	4	5	5	1
	$D_{e_{MD}}$	3	2	3	2	4	5	5	3	5	5	1
multi-indicateur	$D_{\bar{\epsilon}}$	3	3	5	2	3	5	5	4	5	6	1
	$D_{\epsilon_{Me}}$	2	3	4	2	3	5	4	4	5	6	1
	D_{VU}	0	0	0	1	2	0	1	2	1	1	0
	D_{VM}	1	1	1	1	2	4	4	3	3	4	0

Tableau 6.8: Comparaison du nombre d'interventions inutiles à l'échelle du parc NII^F par approche de surveillance conditionnelle, pour l'ensemble de la sous-flotte

On remarque que le cahier des charges ($NII^F \leq 1$) n'est pas respecté pour une majorité des cas. On peut notamment mentionner le cas des parcs M_1 et M_4 , ou même l'approche D_{VU} fournie un NII^F strictement supérieur à 1 sur l'année de données utilisée, alors que c'est la seule qui permet de respecter le cahier des charges pour les 4 parcs présentant un défaut étudiés d'après l'analyse réalisée en section 6.3.3. Ainsi, bien que les approches de surveillance fondées sur les votes des sept indicateurs (D_{VU} et D_{VM})

permettent tendanciellement de réduire le nombre d'interventions inutiles moyen par rapport aux autres approches pour tous les parcs de la sous-flotte, le constat est sans appel : aucun des processus de surveillance conditionnelle proposés ne permet de respecter le cahier des charges.

Difficulté d'interprétation du NII et revue du cahier des charges

Comme nous l'avons expliqué précédemment, les approches de détection D_{VM} et D_{VU} sont supposées robustes aux perturbations mineures (ne nécessitant pas la planification d'une intervention de maintenance préventive), et permettent de n'identifier que les perturbations majeures. Ainsi, si après application de ces détecteurs multi-indicateur sur une période de bon fonctionnement H_0 des interventions persistent, c'est que cette période ne représente pas intégralement une période de bon fonctionnement : elle intègre des perturbations impactant les variations résiduelles, et nécessitant la planification d'une intervention de maintenance préventive (pour rappel, une période est catégorisée comme "en bon fonctionnement" sur le seul critère de la non-présence d'intervention de maintenance corrective).

Ainsi, un NII^F de 4 ne signifie pas que l'outil de surveillance conditionnelle occasionnera le déplacement inutile de techniciens 4 fois par an et par élément du parc surveillé : il est possible que parmi ces 4 interventions, certaines soient nécessaires et relèvent d'une véritable anomalie en développement sur le composant modélisé.

Par conséquent, bien qu'utile pour fournir une information sur les tendances de bilan d'alarmes d'une approche de surveillance conditionnelle donnée, l'interprétation du NII^F n'est pas triviale : nous avons établi que le système de surveillance conditionnelle proposé devait occasionner un nombre d'interventions en période de bon fonctionnement à l'échelle du parc inférieur ou égal à 1 pour limiter les surcoûts de maintenance, mais cet objectif n'est pas toujours réalisable, ni même désirable à cause de la définition des périodes de bon fonctionnement proposée.

Enfin, il convient de rappeler qu'en utilisation en-ligne, sur des données dont on ne connaît pas la nature (bon fonctionnement H_0 ou défaut H_1), les "interventions" relevées par le système de surveillance conditionnelle développé sont analysées et interprétées par un utilisateur expert afin de déterminer la nécessité de mettre en place une intervention de maintenance préventive sur le composant surveillé. Par conséquent, la notion d'"intervention inutile" n'est pas déterminée par le système de surveillance conditionnelle, mais par l'utilisateur expert.

En conclusion, au vu de la stratégie de sélection des périodes de bon fonctionnement utilisée, l'analyse du nombre d'interventions inutiles ne permet pas d'évaluer les performances de détection sur H_0 . À moins de revoir la sélection de la période de bon fonctionnement H_0 de manière à ce qu'elle

ne contienne aucune perturbation nécessitant une maintenance préventive, le critère sur le NII^F doit donc être retiré du cahier des charges.

D'après cette conclusion, et au vu des analyses réalisées dans la Section 6.3, il convient ainsi de définir l'approche de surveillance conditionnelle fondée sur le vote majoritaire des détecteurs des sept indicateurs, D_{VM} comme l'approche la plus adaptée à une application industrielle.

Dans la suite du manuscrit, nous parlerons simplement de "nombre d'interventions" NI, et cette métrique sera analysée à titre informatif, sans objectif d'évaluation des performances de détection.

6.4 Conclusion

Nous avons débuté ce chapitre en rappelant les résultats du chapitre 5 : l'approche de surveillance conditionnelle mono-indicateur ne permet pas de respecter le cahier des charges, car elle occasionne un surcoût de maintenance trop important en période de bon fonctionnement, et peut générer des indicateurs potentiellement insensibles au défaut et trop peu disponibles durant cette période.

Afin de répondre à ces problèmes, nous avons ainsi proposé dans ce chapitre une approche de surveillance conditionnelle basée sur l'utilisation conjointe de plusieurs indicateurs construits à partir de différents modèles linéaires statiques, à trois variables d'entrées. Pour générer ces modèles, nous avons exploité le processus de sélection de régresseurs LASSO-GFS présenté au chapitre 5. Les modèles obtenus, aux côtés d'un modèle dont les variables d'entrée sont sélectionnées par leurs liens "physique" avec y , sont utilisés pour construire des indicateurs multi-turbine. Pour compléter cet ensemble d'indicateurs, nous avons également considéré l'écart à la médiane du parc des valeurs de la variable de sortie y . Plusieurs stratégies de surveillance conditionnelle multi-indicateur, fondées sur une combinaison des sept indicateurs obtenus, ont été étudiées : deux approches dont les détecteurs utilisent les valeurs moyenne et médiane de l'ensemble d'indicateurs construits, et deux approches dont les détecteurs utilisent le vote unanime et le vote majoritaire des détecteurs associés aux indicateurs disponibles.

Ces indicateurs, ainsi que les approches de surveillance conditionnelle multi-indicateur associées ont par la suite été analysées sur la base de deux cas de composant en défaut, afin d'illustrer les caractéristiques d'une surveillance conditionnelle fondée sur un ensemble de modèles, validant les hypothèses d'études réalisées : indépendance des indicateurs en période de bon fonctionnement, génération d'indicateurs sensibles et corrélés en période de défaut. Nous avons pu voir dans une analyse comparative de ces approches avec des méthodes de détection ne considérant qu'un unique indicateur, et sur 4 cas de composants en défaut, que de telles caractéristiques permettaient d'améliorer considérablement les performances et le respect du cahier

des charges, selon les métriques définis dans le chapitre 5.

Toutefois, au regard de ces analyses, il apparaît que le choix de l'approche de surveillance conditionnelle la plus adaptée à une application industrielle n'est pas trivial, car aucune ne permet de respecter pleinement le cahier des charges pour les quatre parcs étudiés. En effet, les approches de surveillance multi-indicateur D_{VU} , D_{VM} , fondées sur les votes des détecteurs des sept indicateurs se positionnent comme meilleurs choix ; néanmoins, l'approche basée sur le vote unanime des détecteurs, D_{VU} , est un choix risqué pour un déploiement industriel car elle peut occasionner une faible sensibilité aux défauts : taux de vrais positifs inférieur à 50% pour le cas du défaut du parc E_f^{rg2} , M_f^{rm} et N_f^{rc} , et avance à la détection très proche du cahier des charges (9 jours) pour les parcs E_f^{rg2} et N_f^{rc} . En comparaison, l'approche basée sur le vote majoritaire D_{VM} conserve une avance à la détection largement supérieure à 7 ainsi que taux de vrais positif et disponibilité supérieurs à 50% pour les quatre cas étudiés. Cependant, elle occasionne un trop grand nombre d'interventions inutiles dans le cas de l'utilisation des données du parc N_f^{rc} .

Néanmoins, nous avons vu qu'il n'était pas possible d'obtenir un NII^F respectant le cahier des charges pour tous les parcs éoliens de la flotte expérimentale, à cause de la méthode de sélection des périodes de bon fonctionnement, et qu'il était par conséquent préférable d'écarter cette métrique du cahier des charges. Par conséquent, l'approche de détection D_{VM} , basée sur le vote majoritaire des détecteurs mono-indicateur associés aux sept indicateurs est retenue, et intégrée à l'approche de surveillance conditionnelle multi-indicateur proposée dans ces travaux de thèse. La Figure 6.14 présente le processus complet de surveillance conditionnelle multi-indicateur retenu, de l'extraction des données SCADA à l'identification des interventions par l'utilisation du vote majoritaire.

Cependant, l'approche de détection multi-indicateur défendue dans ce chapitre implique une augmentation du nombre de modèles à générer en vue d'une application à une échelle plus grande, représentative de la flotte VALEMO. Si on considère de surcroît que de nouveaux modèles doivent être générés pour chaque parc éolien, alors le nombre de modèles à manipuler devient ingérable pour un exploitant industriel, ce qui va à l'encontre du besoin d'**ergonomie** (Tableau 2.1). Ainsi, dans le chapitre suivant, nous allons vérifier qu'il est possible d'appliquer les ensembles de régresseurs sélectionnés par le processus LASSO-GFS à partir d'un jeu de données provenant d'un parc d'une certaine technologie à toutes les turbines des parcs de même technologie, et ceci sans dégrader les performances de détection. Le cas échéant, le besoin d'**ergonomie** défini au chapitre 2 sera respecté : l'analyse des résidus est rendue plus simple pour les utilisateurs experts car les mêmes

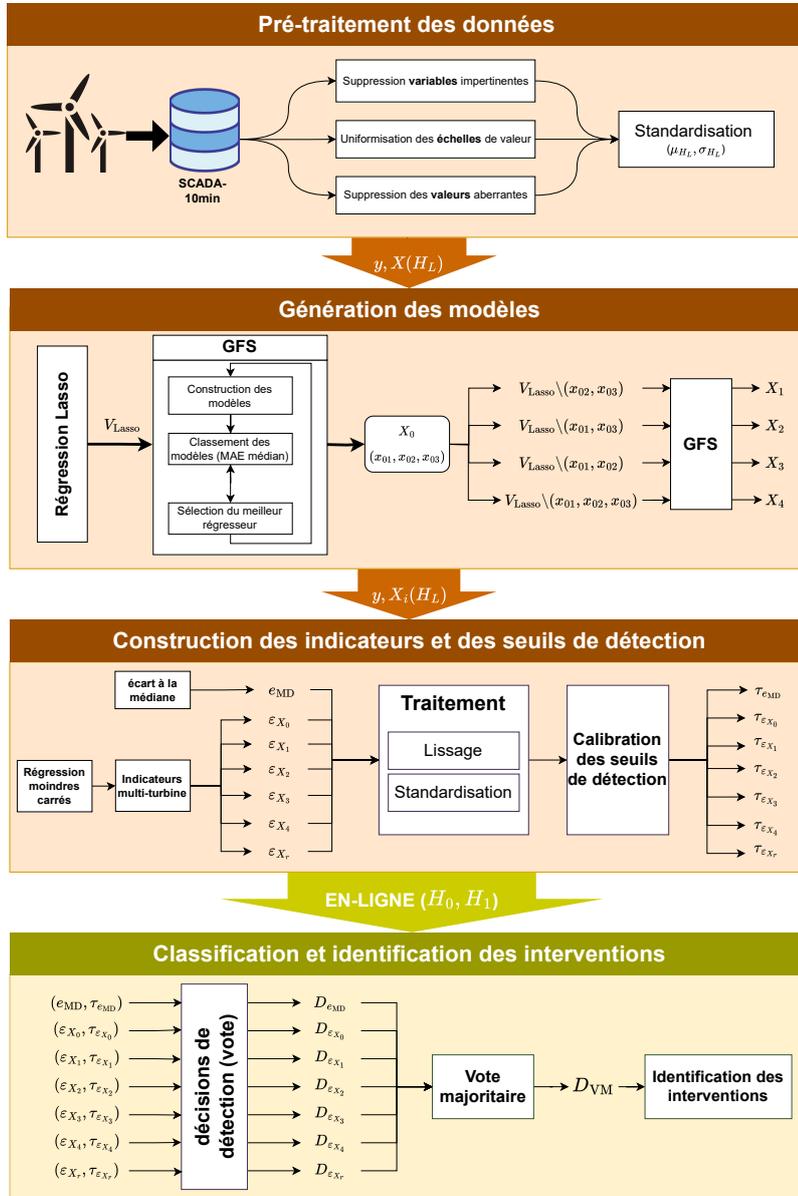


Figure 6.14: Schéma du processus complet de surveillance conditionnelle multi-indicateur proposé

ensembles de variables d'entrée peuvent ainsi s'appliquer sur plusieurs parcs provenant de la même technologie.

Étude de la généralisabilité des résultats de sélection du processus LASSO-GFS

Dans le chapitre précédent, nous avons développé une méthode de détection de défauts éoliens fondée sur la combinaison d'un ensemble de modèles linéaires statiques simples, dont la majorité est issue d'une application du processus de sélection de régresseurs LASSO-GFS. Nous allons voir dans ce chapitre que les variables d'entrée sélectionnées par l'algorithme dépendent du parc éolien considéré. Par conséquent, le processus de surveillance conditionnelle multi-indicateur doit utiliser un très grand nombre de variables d'entrées pour générer les résidus et surveiller l'ensemble de la flotte. Ainsi, l'objectif de ce chapitre est de voir s'il est possible d'améliorer l'**ergonomie** du système de surveillance conditionnelle proposé en s'assurant que les modèles générés par la sélection LASSO-GFS à partir de données provenant d'un parc éolien peuvent être appliqués sans perte de performance en détection à d'autres parcs composés d'éoliennes appartenant à la même technologie de turbine. Nous appellerons cette propriété "généralisabilité" des modèles aux parcs de même technologie. Dans ce chapitre, par souci de clarté, le terme "modèle" sera utilisé pour désigner la structure du modèle, c'est à dire les 3 variables d'entrée du modèle linéaire sélectionnées par l'algorithme LASSO-GFS. Ainsi, un modèle commun à plusieurs turbines signifie que la structure du modèle est la même, c'est-à-dire que les régresseurs sont les mêmes.. Cependant, rappelons que les paramètres du modèle (coefficients de régression estimés par méthode des moindres carrés) varient d'une turbine à l'autre. Pour répondre à cet objectif, nous nous baserons sur une analyse comparative de courbes ROC et des métriques de performances définies en section 5.3. On vérifiera notamment le maintien de l'état de satisfaction du cahier des charges défini en section 5.3.4 selon l'ensemble de variables d'entrée considéré.

7.1 Non reproductibilité du processus de sélection LASSO-GFS

7.1.1 Non-homogénéité des bases de données

Comme nous l'avons explicité dans le chapitre 4, les capteurs utilisés et leur localisation sur les éoliennes ne sont pas les mêmes en fonction des fabricants. Ainsi, du fait de cette **non-homogénéité** des bases de données SCADA (Tableau 2.2), le résultat de sélection de régresseurs du processus LASSO-GFS obtenu pour la modélisation d'un composant d'un parc d'éoliennes provenant d'un fabricant n'est pas applicable à la modélisation du même composant d'un parc dont les éoliennes proviennent d'un autre fabricant.

Dans le chapitre 4, nous avons défini une "technologie de turbine" comme une "gamme" d'éoliennes présentant des caractéristiques techniques similaires. Un fabricant d'éoliennes propose ainsi différentes "technologies de turbine", qui diffèrent principalement par leur efficacité énergétique. Les données SCADA provenant de deux turbines de deux technologies de turbine différentes sont elles aussi différentes. Le résultat de sélection de régresseurs du processus LASSO-GFS obtenu à partir de la base de données d'un parc d'éoliennes provenant d'une technologie de turbine n'est pas applicable à la base de données d'un parc dont les éoliennes proviennent d'une autre technologie de turbine, même si les turbines ont été construites par le même fabricant.

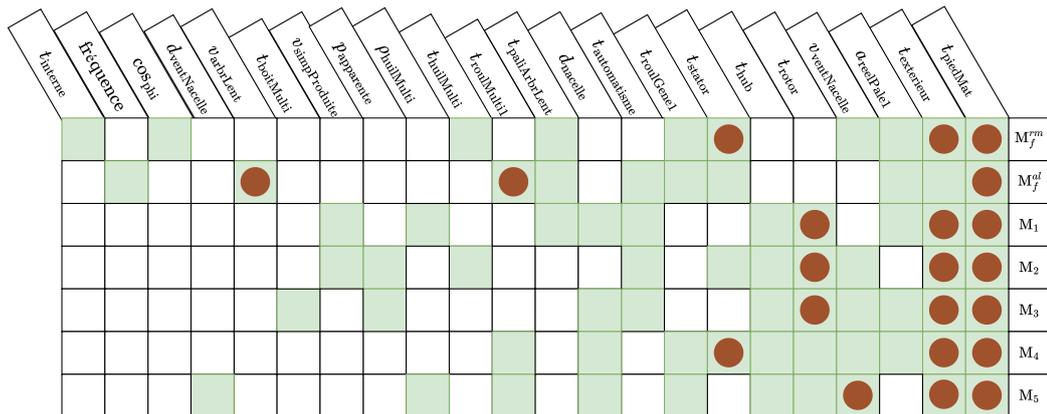


Figure 7.1: Visualisation de la sélection de variables réalisée par le processus LASSO-GFS pour la modélisation du convertisseur à partir de parcs provenant de la technologie de turbine M

7.1.2 Non reproductibilité à l'échelle de la technologie et enjeux de généralisation

La question de la reproductibilité des résultats obtenus par l'algorithme LASSO-GFS sur des parcs de même technologie se pose, quant à elle. En effet, les données SCADA disponibles sont les mêmes pour tous ces parcs. Afin d'apporter une réponse à cette question, la Figure 7.1 représente pour les 7 parcs de la technologie de turbine M de la flotte expérimentale considérée (Tableau 4.1) les variables retenues par le processus de sélection LASSO-GFS pour la modélisation du système de refroidissement du convertisseur. Sur chaque ligne (représentant un parc), les cases vertes représentent des variables retenues par la régression LASSO, et une pastille marron est attribuée aux variables également retenues par l'algorithme de sélection GFS.

Un premier constat est que le résultat de sélection du processus LASSO-GFS n'est pas reproductible à l'échelle d'une technologie de turbine : les variables retenues pour la modélisation du composant considéré dépendent de la base de données utilisée pour l'apprentissage, même quand les données proviennent de parcs appartenant à une même technologie de turbine.

Cette non reproductibilité du processus de sélection de régresseurs pose plusieurs problèmes dans un cadre applicatif industriel. Elle peut être mal perçue par un industriel : l'utilisation de modèles différents pour des parcs éoliens distincts mais constitués de turbines de la même technologie peut être une source de confusion et impacter par conséquent l'acceptabilité du système de surveillance conditionnelle résultant. En effet, dans un contexte applicatif industriel, un utilisateur expert a pour habitude d'utiliser des outils d'exploitation similaires sur plusieurs parcs appartenant à une même technologie (par exemple les systèmes de supervision de la technologie d'acquisition de données SCADA, dont le paramétrage est fixé pour une technologie de turbine donnée).

De plus, le fait que l'algorithme génère des modèles différents pour des parcs de même technologie nuit à l'**ergonomie** du système de surveillance conditionnelle : cela alourdit en effet très fortement la charge cognitive d'un opérateur de maintenance et complexifie la démarche diagnostique qu'il doit mettre en place à chaque levée d'alarmes. Ceci est d'autant plus vrai si on considère le cadre d'une surveillance conditionnelle multi-indicateur suivant le processus retenu à l'issue du chapitre précédent, car on peut s'attendre à obtenir des ensembles de régresseurs X_0 , X_1 , X_2 , X_3 et X_4 différents pour chaque composant et chaque parc éolien considéré dans une flotte industrielle. Pour illustration, la surveillance conditionnelle multi-indicateur de 4 parcs éoliens provenant d'une même technologie, dont 5 composants sont modélisables (c'est-à-dire dont les variables de sortie sont disponibles) nécessite l'utilisation de 100 ensembles de régresseurs potentiellement différents (5 composants fois 5 modèles uniques fois 4 parcs éoliens). Si un même modèle peut s'appliquer pour chaque composant à

toutes les turbines d'une même technologie, ce nombre redescend à 25.

7.1.3 Analyse des résultats de sélection des régresseurs

Les variables identifiées dans la Figure 7.1 sont ordonnées par ordre croissant d'occurrence de sélection par l'algorithme de régression LASSO : $t_{interne}$, située tout à gauche, n'est retenue que pour l'utilisation des données du parc M_f^{rm} alors que $t_{piedMat}$, située tout à droite, est retenue pour les 7 parcs considérées. On remarque ainsi que :

- Les températures dont les capteurs sont localisés à l'extérieur de l'éolienne ($t_{exterieur}$, $t_{piedMat}$) sont retenues pour les 7 bases de données.
- Les températures au niveau du moyeu du rotor (t_{rotor} , t_{hub}) sont retenues pour 5 des 7 bases de données.
- La température du roulement avant de la génératrice ($t_{roulGene1}$) est retenue pour 5 des 7 bases de données.
- La vitesse du vent mesurée ($v_{ventNacelle}$) est retenue par 5 des 7 bases de données.



Figure 7.2: Schéma en coupe d'une éolienne de la technologie de turbine M

Le convertisseur de la technologie de turbine M est localisé à l'extérieur de la nacelle au pied de l'éolienne. Il est couplé à la génératrice d'après la Figure 7.2, montrant la localisation des composants de la nacelle d'une éolienne de cette technologie, et l'électricité produite par cette machine est partiellement livrée par le convertisseur et par le rotor¹.

Ainsi, bien que différents, les choix de potentiels régresseurs pour l'estimation de y réalisés par V_{LASSO} sont physiquement cohérents pour les 7 bases de données considérées. Il en va ainsi de même pour la sélection de régresseurs réalisée par l'algorithme GFS, qui privilégie les mesures des capteurs localisés aux pieds des éoliennes. On remarque notamment que, à part pour l'utilisation des données du parc M_f^{al} , le processus de sélection

¹Annexe 8 : éolienne Repower MM92, étude d'impact sur l'environnement et la santé du projet éolien de la commune de Grand-Rozoy

LASSO-GFS retient des régresseurs communs à au moins 4 des 7 ensembles de variables V_{Lasso} considérés. Il est donc légitime de se demander si ces modèles sont si différents que ça, et quel serait l'impact d'un changement de régresseurs, parmi les sélections de la Figure 7.1, sur leur précision d'estimation.

7.1.4 Problématique

Pour assurer l'**ergonomie** du système de surveillance conditionnelle, il est nécessaire de n'utiliser qu'un seul modèle par composant et par technologie. Il faut donc vérifier qu'un modèle généré par l'algorithme pour un parc A peut être utilisé sur un autre parc B sans perdre en capacité de détection, même si l'algorithme LASSO-GFS entraîné sur le parc B aurait fourni un autre modèle.

À titre d'exemple, pour la température du système de refroidissement du convertisseur, on peut se demander si on peut utiliser n'importe lequel des ensembles de variables d'entrée de la Figure 7.1 pour modéliser cette variable de sortie, et obtenir de bonnes performances en détection sur n'importe quel parc de la flotte M.

Pour répondre à cette problématique, nous allons dans ce chapitre évaluer l'impact de ces différences entre régresseurs sur les performances de détection, selon une méthodologie présentée dans la section suivante. On nommera par la suite **généralisabilité** des modèles générés par le processus LASSO-GFS la capacité de reproductibilité des performances de détection d'un système de surveillance conditionnelle à l'aide de ces modèles à une certaine échelle d'application. L'objectif de ce chapitre est ainsi d'évaluer la **généralisabilité** des modèles à l'échelle de la technologie de turbine.

7.2 Stratégie d'évaluation de la généralisabilité des modèles

L'objectif de notre stratégie est de vérifier si un ensemble de régresseurs généré par le processus LASSO-GFS à partir de données provenant d'un parc peut être appliqué sur un autre parc provenant de la même technologie, sans perte significative de performances du système de surveillance conditionnelle. Nous préciserons également les hypothèses d'étude réalisées, permettant de généraliser les résultats de ce chapitre à toutes les technologies de turbine qui composent la flotte de VALEMO, et à tous les modèles issus du processus de sélection LASSO-GFS de l'approche de surveillance conditionnelle multi-indicateur développée au sein du chapitre 6.

7.2.1 Principe

Hors-ligne

La Figure 7.3 présente la partie hors-ligne de la stratégie d'évaluation de la généralisabilité du processus de sélection de régresseurs LASSO-GFS à l'échelle de la technologie de turbine proposée.

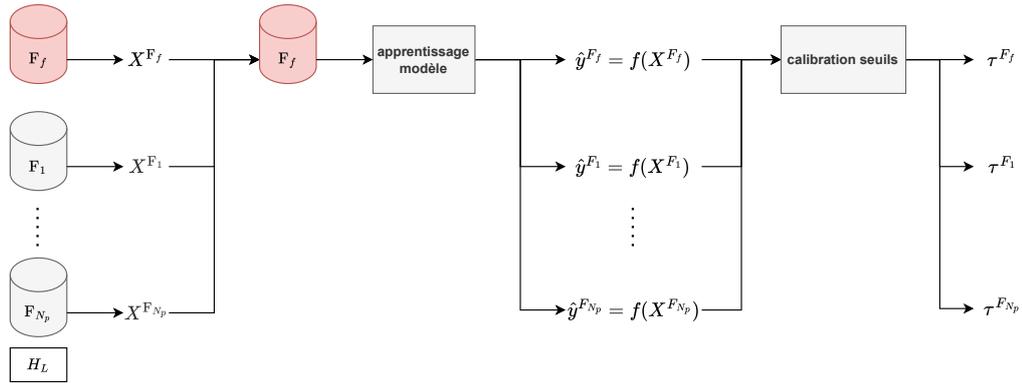


Figure 7.3: Partie hors-ligne de la stratégie d'évaluation de la généralisabilité des modèles générés par LASSO-GFS à l'échelle de la technologie de turbine

Le parc F_f présente un défaut sur un composant d'une de ses turbines, qu'on cherche à surveiller par modélisation de la variable de sortie y . Les parcs F_1 à F_{N_p} sont d'autres parcs provenant de la même technologie de turbine, ne présentant pas de défaut. La technologie de turbine étudiée est ainsi représentée par $N_p + 1$ parcs éoliens.

Les données d'apprentissage H_L de tous ces parcs sont utilisées pour générer $N_p + 1$ ensembles distincts de régresseurs par application du processus de sélection LASSO-GFS. Ces ensembles de régresseurs sont identifiés par le parc utilisé pour les générer. Ainsi, X^{F_i} désigne les trois variables d'entrée utilisées pour modéliser y , générées à partir des données du parc F_i .

Pour la suite de la stratégie d'évaluation de la généralisabilité des modèles LASSO-GFS, on utilise uniquement les données du parc F_f . Tout d'abord, on réalise l'apprentissage des paramètres des modèles construits à partir des $N_p + 1$ ensembles de régresseurs obtenus précédemment. On obtient alors $N_p + 1$ modélisations distinctes d'une variable de sortie y . Enfin, on procède à la calibration des seuils de détection associés à chacun de ces modèles, toujours en utilisant les données d'apprentissage H_L du parc F_f .

En-ligne

La Figure 7.3 présente la partie en-ligne de la stratégie d'évaluation de la généralisabilité du processus de sélection de régresseurs LASSO-GFS à l'échelle de la technologie de turbine proposée.

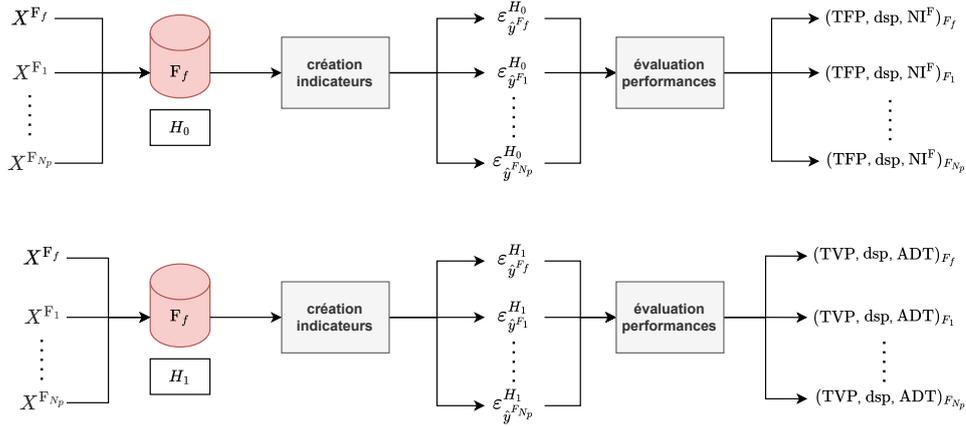


Figure 7.4: Partie en-ligne de la stratégie d'évaluation de la généralisabilité des modèles générés par LASSO-GFS à l'échelle de la technologie de turbine

On utilise les données du parc F_f , sur une période de bon fonctionnement H_0 et sur la période présentant le défaut H_1 .

Premièrement, $N_p + 1$ indicateurs multi-turbine sont construits sur les données H_0 et H_1 à partir des modèles générés durant la phase hors-ligne présentée dans la section précédente. Puis, les métriques de performances sont calculées à partir des données de chacun de ces indicateurs. Sur les données de bon fonctionnement H_0 , le taux de faux positifs TFP, la disponibilité de l'indicateur, et le nombre d'interventions inutiles à l'échelle du parc F_f , NI^F , sont mesurés. Sur la période de défaut H_1 , on mesure le taux de vrais positifs TVP, la disponibilité de l'indicateur, et l'avance à la détection ADT.

7.2.2 Analyse comparative de performances

Les taux de positifs (TVP, TFP) et les avances à la détection (ADT) sont calculées pour construire les courbes ROC TVP(TFP) et TVP(ADT) selon la méthode présentée en Section 5.3.2. Dans le cadre de l'étude de généralisabilité des résultats du processus de sélection de variables LASSO-GFS, les courbes ROC considérées sont un outil de visualisation des profils de sensibilité des indicateurs formés par les différents jeux de régresseurs

analysés, fournissant des éléments préliminaires de comparaison des performances.

Comme l'analyse de ces courbes de performances seules ne permet pas de conclure sur la généralisabilité de l'algorithme LASSO-GFS à l'échelle de la technologie de turbine, les métriques de performances définies durant le chapitre 5 (disponibilité des indicateurs sur H_0 et H_1 , taux de positifs TVP, TFP et avance à la détection ADT) sont mesurées et analysées. Nous vérifierons notamment le maintien du respect du cahier des charges pour tous les modèles considérés et concluons le cas échéant à la généralisabilité du processus de sélection de régresseurs LASSO-GFS à l'échelle de la technologie de turbine. Comme nous l'avons expliqué à l'issue du chapitre 6, il est difficile d'interpréter le NII^F à cause de la sélection des données H_0 . La notion d'"interventions inutiles" est ainsi exclue de l'étude réalisée dans ce chapitre, et nous parlerons simplement d'interventions. Nous analyserons ainsi le nombre d'interventions à l'échelle du parc NI^F , sans critère associé sur le cahier des charges.

Pour illustration, comme nous l'avons vu dans le chapitre 5, l'utilisation du jeu de régresseurs $X^{M_{el}}$ fournit un indicateur multi-turbine très peu disponible sur H_1 , mais respecte les autres critères du cahier des charge d'après le Tableau 5.3. Selon la stratégie d'évaluation de la généralisabilité du processus LASSO-GFS proposée, nous vérifierons ainsi que l'utilisation de jeux de régresseurs générés par utilisation des données de tous les autres parcs de la technologie de turbine M permette également de respecter le cahier des charges concernant le taux de vrais positifs ($\geq 50\%$), l'avance à la détection (≥ 7 jours) et la disponibilité des indicateurs multi-turbine sur H_0 .

7.2.3 Hypothèses d'étude

Extension de la validité des résultats à toute la flotte VALEMO

La méthodologie présentée d'évaluation de la généralisabilité du processus de sélection LASSO-GFS est appliquée sur les 11 parcs de la flotte expérimentale détaillée dans le Tableau 4.1. Cet échantillon de parcs éoliens contient 3 technologies de turbine : E, N, et M. Par manque de données historiques de défaillance, cette méthodologie ne peut pas être appliquée à l'intégralité de la flotte VALEMO, qui est composé de 24 technologies de turbine (4.1.1). par conséquent, les résultats de l'application de cette méthodologie ne sont théoriquement valides que pour une portion de la flotte VALEMO.

Cependant, il convient de rappeler que dans ce chapitre, nous évaluons la généralisabilité de la phase de sélection des régresseurs. Ces régresseurs sont sélectionnés à partir de données en bon fonctionnement, qui sont disponibles pour toutes les technologies de la flotte VALEMO. Nous pouvons donc

faire l'hypothèse que les résultats obtenus dans cette étude sont applicables à toutes les technologies de turbine, et que ce protocole permet donc bien d'évaluer la généralisabilité du résultat de sélection automatique de régresseurs LASSO-GFS à l'échelle de la technologie de turbine pour l'intégralité de la flotte VALEMO.

Extension de la validité des résultats à l'approche de surveillance conditionnelle multi-indicateur

Dans le cadre de l'évaluation de la généralisabilité du processus de sélection LASSO-GFS, nous considérons un processus de surveillance conditionnelle mono-indicateur, suivant la méthodologie présentée dans le chapitre 5. Ce choix est fait pour simplifier les analyses comparatives à réaliser, en ne considérant que le jeu de régresseur X_0 de l'approche de surveillance conditionnelle multi-indicateur (obtenu par application de l'algorithme GFS à partir de la totalité des variables de V_{Lasso} , comme expliqué en section 6.2.1).

Nous pouvons toutefois faire l'hypothèse que les résultats de l'étude de généralisabilité réalisée dans ce chapitre peuvent s'appliquer à l'approche de surveillance conditionnelle multi-indicateur. En effet, la méthode de détection étudiée dans le chapitre 6 est constituée de modèles dont les régresseurs sont générés à l'aide du processus de sélection LASSO-GFS, ainsi que de modèles aux variables d'entrée fixées (y , $p_{\text{apparente}}$, v_{vent} et $t_{\text{exterieur}}$ pour l'ensemble X_r), disponibles pour tous les parcs éoliens de la flotte VALEMO.

Analyses des taux de positifs et de la disponibilité

Pour rappel, le cahier des charges défini dans le chapitre 5 stipule qu'en vue d'une application industrielle, un système de surveillance conditionnelle doit proposer des détecteurs de défauts disponibles au moins 50% du temps, et présentant un taux de positifs sur H_1 d'au moins 50%.

Or nous avons vu dans le chapitre 6 que le processus de surveillance conditionnelle défini dans le chapitre 5 (fondé sur le processus de sélection de régresseurs LASSO-GFS) pouvait générer des indicateurs peu sensibles au défaut (car construits à partir de variables sensibles au défaut), et que leur disponibilité dépendait directement de la disponibilité des variables d'entrée considérées. par conséquent, puisque tous les jeux de régresseurs étudiés dans ce chapitre seront issus du même processus que celui utilisé dans le chapitre 5, ces cas de figure peuvent arriver : on sait que les contraintes du cahier des charges concernant le TVP et la disponibilité des indicateurs multi-turbine peuvent ne pas être respectées.

Toutefois, le chapitre 6 a également permis de démontrer que l'approche multi-indicateur (basée sur une fusion de modèles) considérant le vote majoritaire des détecteurs mono-indicateur (D_{VM}) permettait de régler ces

risques d'indisponibilité et de faible sensibilité au défaut. On considèrera ainsi dans l'analyse réalisée dans la section suivante qu'un non respect du cahier des charges illustré par une baisse de disponibilité ou de la sensibilité au défaut occasionnée par un changement de régresseurs ne permet pas de conclure à la non-généralisabilité du process LASSO-GFS, car ces faiblesses sont par hypothèse réglées par approche multi-indicateur.

Par conséquent, la décision quant à la généralisabilité du processus LASSO-GFS à l'échelle de la technologie de turbine sera basée sur deux éléments du cahier des charges: (1) le maintien du respect du cahier des charges sur l'avance à la détection ($ADT > 7$) et (2) la stabilité du nombre d'interventions à l'échelle du parc. En effet, le NI^F n'est pas intégré au cahier des charges, mais on considère qu'un changement de variables d'entrée ne doit pas pour autant impliquer une augmentation significative du nombre d'interventions.

7.3 Analyse des résultats

Dans un premier temps, nous allons dans cette section analyser les résultats de sélection du processus LASSO-GFS pour la modélisation des 4 composants en défaut à partir de bases de données provenant de plusieurs parcs appartenant à une même technologie de turbine. Puis nous appliquerons la méthode d'évaluation de la généralisabilité du processus de sélection LASSO-GFS que nous venons de présenter, en analysant les courbes ROC des indicateurs multi-turbine associés aux différents jeux de régresseurs considérés, ainsi que les métriques de performances et le maintien de l'état de respect du cahier des charges.

7.3.1 Résultats de sélection de régresseurs LASSO-GFS

Le Tableau 7.1 montre les résultats de l'application du processus LASSO-GFS pour la génération des régresseurs pour la modélisation des 4 composants en défauts étudiés dans les chapitres 5 et 6 :

- Les deux premières lignes montrent les variables d'entrée retenues par le processus de sélection LASSO-GFS pour la modélisation du roulement arrière de la génératrice du parc E_f^{rg2} , en utilisant les données de tous les parcs de la technologie de turbine E.
- Les sept lignes suivantes montrent les variables d'entrée retenues par le processus de sélection LASSO-GFS pour la modélisation du roulement principal du parc M_f^{al} , en utilisant les données de tous les parcs de la technologie de turbine M.
- Les sept lignes suivantes montrent les variables d'entrée retenues par le processus de sélection LASSO-GFS pour la modélisation du refroidisse-

ment de la multiplicatrice du parc M_f^{rm} , en utilisant les données de tous les parcs de la technologie de turbine M.

- Les deux dernières lignes montrent les variables d'entrée retenues par le processus de sélection LASSO-GFS pour la modélisation du refroidissement du convertisseur du parc N_f^{rc} , en utilisant les données de tous les parcs de la technologie de turbine N.

Parc	Sortie	Origine des régresseurs	x_1	x_2	x_3
E_f^{rg2}	$t_{roulGene2}$	E_f	$t_{roulGene1}$	t_{stator}	$p_{apparente}$
		E_1	$t_{roulGene1}$	t_{stator}	$t_{interne}$
M_f^{al}	$t_{paliArbrLent}$	M_f^{al}	$t_{roulMulti1}$	$t_{piedMat}$	$v_{simplProduite}$
		M_f^{rm}	$t_{piedMat}$	$v_{simplProduite}$	$t_{roulMulti1}$
		M_1	t_{rotor}	$v_{simplProduite}$	t_{hub}
		M_2	t_{hub}	$v_{simplProduite}$	$t_{roulMulti1}$
		M_3	$t_{convertisseur}$	$v_{simplProduite}$	t_{hub}
		M_4	$t_{piedMat}$	$v_{simplProduite}$	$t_{roulMulti1}$
M_f^{rm}	$t_{huilMulti}$	M_5	$t_{piedMat}$	$v_{simplProduite}$	$t_{roulMulti1}$
		M_f^{rm}	$t_{roulMulti2}$	$t_{boiteMulti}$	$\cos\phi$
		M_f^{al}	$t_{roulMulti2}$	$v_{arbrRapide}$	$t_{boitMulti}$
		M_1	$t_{roulMulti2}$	$t_{boiteMulti}$	$t_{roulGene2}$
		M_2	t_{stator}	$t_{boiteMulti}$	$t_{roulMulti1}$
		M_3	$t_{roulMulti2}$	$t_{boiteMulti}$	t_{stator}
N_f^{rc}	$t_{convertisseur}$	M_4	$t_{roulMulti2}$	$t_{boiteMulti}$	t_{stator}
		M_5	$t_{roulMulti2}$	$t_{boiteMulti}$	$t_{automatisme}$
N_f^{rc}	$t_{convertisseur}$	N_f	$t_{convertisseur2}$	$t_{transformateur}$	$t_{huilMulti}$
		N_1	$t_{roulMulti2}$	$t_{piedMat}$	ρ_{yaw}

Tableau 7.1: Résultats de sélection du processus LASSO-GFS selon le parc utilisé, par variable de sortie modélisée

Comme observé dans le cas de la modélisation du système de refroidissement du convertisseur pour la technologie N en début de chapitre, les ensembles de régresseurs sélectionnés pour modéliser la même sortie sont différents, mais partagent des variables en commun. On remarque notamment que, malgré ces différences, les choix de variables d'entrée réalisés par le processus LASSO-GFS peuvent s'expliquer en reprenant le fonctionnement interne d'une éolienne (section 2.1.1) :

- La modélisation du roulement arrière de la génératrice (parc E_f^{rg2}) se fait à l'aide de températures localisées sur la génératrice, ainsi qu'avec $p_{apparente}$, qui peut être utilisé pour réaliser l'estimation de n'importe quelle variable de sortie définie dans cette étude, comme nous l'avons vu en section 6.2.1.
- La tension produite par une éolienne $v_{simplProduite}$ présente une relation linéaire avec $p_{apparente}$. Il est donc pertinent de la retrouver dans tous les jeux de régresseurs générés, de même que la température du

roulement avant de la multiplicatrice, directement reliée à l'arbre lent. Le lien entre la température du palier de l'arbre lent et la température mesurée au pied du mat de l'éolienne s'explique par la faible vitesse de rotation de cet arbre, qui implique un faible échauffement du composant, par rapport à l'arbre rapide localisé en bout de multiplicatrice. par conséquent, les variations de température de l'arbre lent sont davantage expliquées par la température extérieure (similaire à la température mesurée en pied de mat) que par sa vitesse de rotation.

- Des variables appartenant principalement à la multiplicatrice et à la génératrice (qui y est couplée) sont utilisées pour la modélisation du système de refroidissement de la multiplicatrice (parc M_f^{rm})
- Le convertisseur est localisé au pied de l'éolienne de technologie N, aux côtés du transformateur, ce qui explique la présence des variables d'entrée $t_{\text{transformateur}}$ et t_{piedMat} . Ensuite, la multiplicatrice étant un élément central du fonctionnement d'une éolienne, fortement instrumenté, il est pertinent de l'identifier parmi les régresseurs de la modélisation du convertisseur de N_f^{rc} .

Ainsi, les choix de variables d'entrée réalisés par le processus LASSO-GFS, bien que dépendant du parc éolien considéré, semblent pertinents pour les modélisations associées. Nous allons désormais voir si les indicateurs multi-turbine construits à partir de ces jeux de régresseurs permettent d'assurer des performances de détection similaires à partir d'une analyse de courbes ROC.

7.3.2 Courbes ROC

Surveillance du roulement arrière de la génératrice de E_f^{rg2}

La Figure 7.5 montre les deux courbes ROC obtenues à partir des données des indicateurs multi-turbine utilisant les régresseurs générés par application du processus de sélection LASSO-GFS sur les données des parcs E_f et E_1 (Tableau 7.1), pour la surveillance du défaut de roulement arrière de la génératrice du parc E_f^{rg2} . La courbe ROC représentée en gras est celle calculée à partir du modèle issu de la base de données du défaut (ici le parc E_f^{rg2} , en rouge).

On constate que l'utilisation des régresseurs issus du parc E_1 permet un très léger gain de performances du point de vue de l'AUC. Cependant, au vu de la faible amplitude de ce gain (1%), on peut faire l'hypothèse que les deux indicateurs multi-turbine étudiés présentent des caractéristiques de performance similaires.

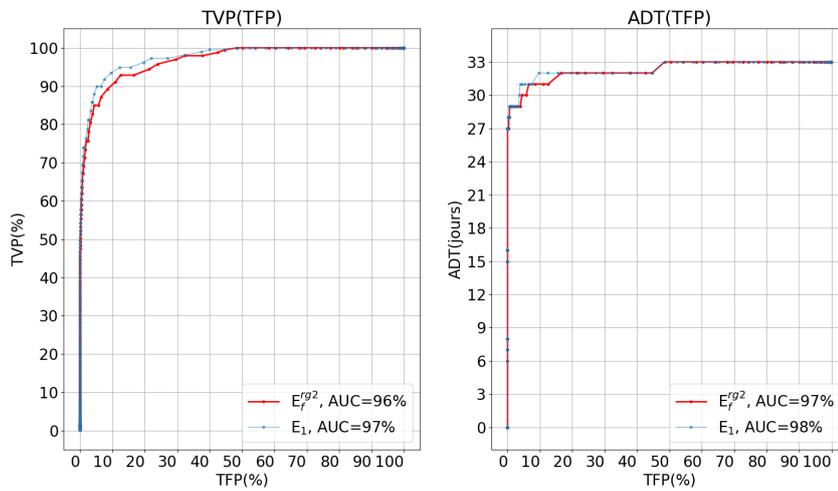


Figure 7.5: Courbes ROC de la modélisation du roulement arrière de la génératrice du parc E_f^{rg2} , en utilisant les jeux de régresseurs obtenus par application du processus LASSO-GFS sur les données des parcs de la technologie E

Surveillance du palier de l'arbre lent de M_f^{al}

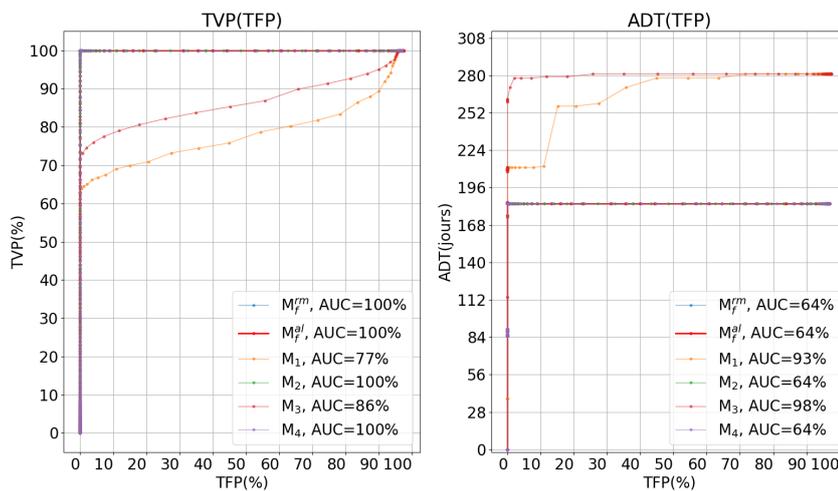


Figure 7.6: Courbes ROC de la modélisation du roulement principal à partir des données de M_f^{al} , en utilisant les modèles issus des parcs de la technologie M

Dans le cas de la surveillance du composant en défaut du parc M_f^{al} (Figure 7.6), deux profils de courbes ROC peuvent être identifiés.

L'utilisation des régresseurs générés à partir des parcs M_1 et M_3 , qui permettent une avance à la détection maximale de 280 jours ($AUC > 90\%$), mais présentent une aire sous la courbe ROC TVP(TFP) autour de 70%.

Au contraire, les courbes de performances TVP(TFP) obtenues par utilisation des régresseurs générés à partir des autres parcs sont optimales (aire sous la courbe de 100%), au détriment de l'avance à la détection, fortement réduite, à 184 jours ($AUC < 80\%$).

Comme nous l'avons vu dans le chapitre 6, cette observation est liée à la disponibilité des indicateurs multi-turbine, expliquée par le contenu en variables des modèles utilisés : d'après le Tableau 7.1, les régresseurs issus des données des parcs M_1 et M_3 pour la modélisation du palier de l'arbre lent de M_f^{al} sont en effet les seuls n'intégrant pas $t_{roulMulti1}$, fortement indisponible durant H_1 . Par conséquent, les indicateurs multi-turbine construits à partir de ces jeux de régresseurs sont bien plus disponibles, et il sont notamment définis sur la première moitié de la période H_1 , contrairement aux indicateurs utilisant $t_{roulMulti1}$ tels que ε_{X_0} comme nous l'avons sur la Figure 5.10b dans le chapitre 5.

Ainsi, l'ADT est fortement avancée et les taux de positifs mesurés sont alors plus précis car calculés à partir d'un plus grand nombre de valeurs d'indicateurs. Les indicateurs multi-turbine étant établis à partir d'approximations de modélisation linéaires, et la sélection des données de bon fonctionnement pouvant intégrer des perturbations impactant les valeurs résiduelles, augmenter le nombre de données résiduelles utilisées pour calculer les taux de positifs contribue à diminuer le TVP sur H_1 , et à augmenter le TFP sur H_0 .

Surveillance du système de refroidissement de la multiplicatrice de M_f^{rm}

La Figure 7.7 montre les courbes ROC pour la modélisation thermique de la multiplicatrice du parc M_f^{rm} .

On remarque que peu importe les variables d'entrée du modèle utilisées, les courbes ROC restent très proches, et présentent des AUC similaires, bien que légèrement supérieures (pour la courbe TVP(TFP)) à celle dont le modèle est généré par l'utilisation des données du parc M_f^{rm} .

Comme nous l'avons vu dans le chapitre 6, la forte variabilité de sensibilité au défaut des indicateurs multi-turbine (du point de vue du TVP) peut être expliquée par la présence de variables d'entrée sensibles au défaut (ici, les variables de mesure de températures localisées sur la multiplicatrice, d'après le Tableau 7.1). Or, par lecture du Tableau 7.1, on observe que tous les modèles générés intègrent au moins une variable de température localisée sur, ou à proximité de la multiplicatrice, sensible au défaut.

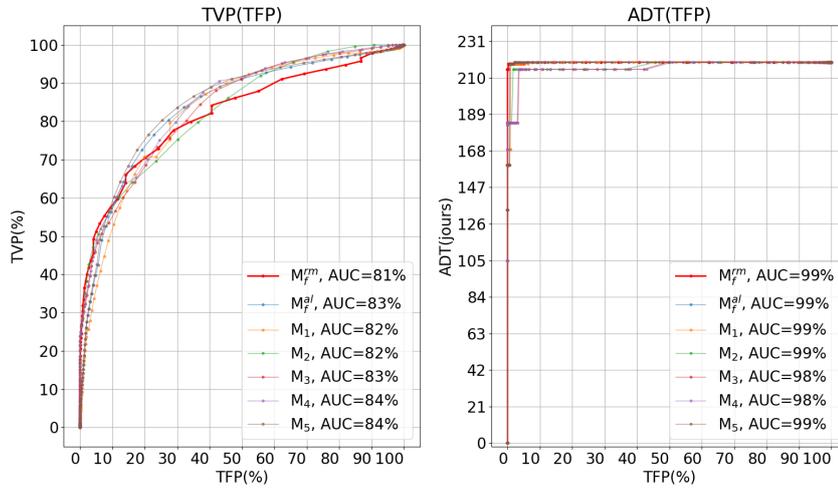


Figure 7.7: Courbes ROC de la modélisation de la multiplicatrice à partir des données de M_f^{rm} , en utilisant les modèles issus des parcs de la technologie M

Surveillance du système de refroidissement du convertisseur de N_f^{rc}

Enfin, en ce qui concerne le cas de la modélisation thermique du convertisseur du parc N_f^{rc} (Figure 7.8), on constate que l'utilisation des régresseurs sélectionnés à partir des données du parc N_1 dégrade significativement les courbes ROC (AUC réduite de 24% pour TVP(TFP) et 18% pour ADT(TFP)).

Ce résultat s'explique par une élévation de la moyenne des variations de la variable d'entrée ρ_{yaw} (sélectionnée par l'utilisation des données du parc N_1) durant la période de bon fonctionnement H_0 . Ce changement de distribution statistique du régresseur implique ainsi des taux de faux positifs très élevés sur cette période, et une distinction entre bon fonctionnement (H_0) et défaut (H_1) est dégradée.

Cet épisode d'élévation de moyenne n'ayant pas eu pour conséquence le remplacement d'un composant, il n'a pas été exclu des données de bon fonctionnement H_0 pour le parc N_f^{rc} . Toutefois, il est fort probable que l'application du système de surveillance conditionnelle à l'aide du modèle issu du parc N_1 pour la surveillance thermique du convertisseur de N_f^{rc} sur H_0 occasionne un nombre important d'interventions. Ce cas d'étude illustre la non-pertinence de la notion d'intervention inutiles dans un contexte d'application "en-ligne" de la méthode de détection, discutée à l'issue du chapitre 6.

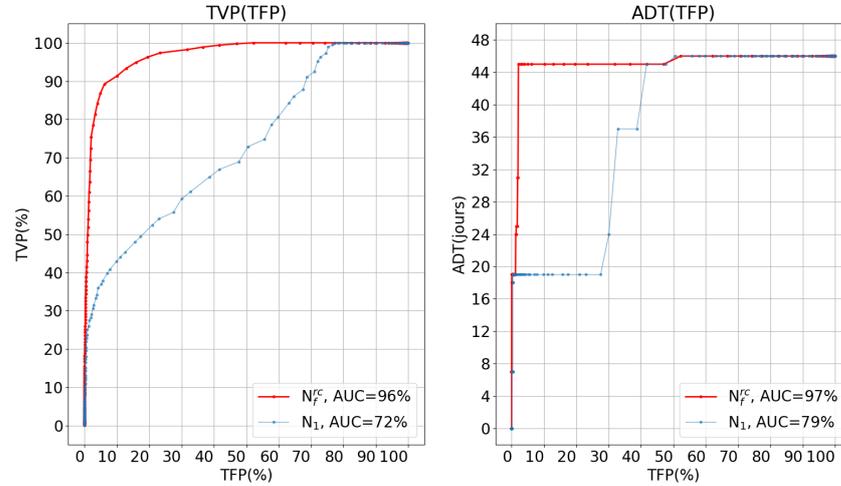


Figure 7.8: Courbes ROC de la modélisation du convertisseur à partir des données de N_f^{rc} , en utilisant les modèles issus des parcs de la technologie N

7.3.3 Métriques de performance

Les Tableaux 7.2, 7.3, 7.4 et 7.5 présentent les métriques de performance calculées pour les indicateurs multi-turbine associés à la surveillance des composants en défaut respectifs (roulement arrière de la génératrice du parc E_f^{rg2} , roulement principal du parc M_f^l , refroidissement de la multiplicative du parc M_f^m et refroidissement du convertisseur du parc N_f^c). Pour chaque modélisation considérée, les indicateurs multi-turbine sont construits à partir de différents jeux de régresseur obtenus par application du processus LASSO-GFS sur plusieurs parcs appartenant à la même technologie de turbine. Les résultats en gras représentent les performances obtenues par l'utilisation des données du parc intégrant le composant en défaut.

Commençons par analyser les variations de disponibilité et de sensibilité (TFP et TVP) des indicateurs multi-turbine. Pour rappel, l'hypothèse a été faite en Section 7.2 que le respect du cahier des charges relativement à ces métriques est assurée par l'application de la méthode de détection multi-indicateur défendue dans le chapitre précédent. Cette analyse est donc réalisée à titre informatif. On remarque tout d'abord par la lecture du Tableau 7.5 que l'utilisation du jeu de régresseurs issu du parc N_1 pour la modélisation du convertisseur de N_f^{rc} occasionne un fort taux de faux positifs, à cause de l'élévation de la valeur moyenne d'une des variables d'entrée du modèle, comme expliqué précédemment. Cependant, le taux de vrais positifs associé reste supérieur à 50% et respecte le cahier des charges.

Origine des régresseurs	TFP(%)	dsp _{H₀} (%)	TVP(%)	dsp _{H₁} (%)	NI ^F	ADT (jours)
E _f ^{rg2}	1.44	99.71	71.25	98.17	4	29
E ^I	1.87	99.76	76.39	98.17	4	29

Tableau 7.2: Bilan de performances des indicateurs multi-turbine pour la modélisation du roulement arrière de la génératrice de E_f^{rg2}

Origine des régresseurs	TFP(%)	dsp _{H₀} (%)	TVP(%)	dsp _{H₁} (%)	NI ^F	ADT (jours)
M _f ^{al}	0.0	60.53	99.56	24.75	3	90
M _f ^{rm}	0.0	60.53	99.56	24.66	3	90
M ₁	5.68	60.53	67.49	71.08	3	211
M ₂	5.69	60.53	100	24.66	3	90
M ₃	1.94	60.53	74.45	71.05	3	278
M ₄	0.0	60.53	99.57	24.66	3	90
M ₅	0.0	60.53	99.57	24.77	3	90

Tableau 7.3: Bilan de performances des indicateurs multi-turbine pour la modélisation du roulement principal de M_f^{al}

Origine des régresseurs	TFP(%)	dsp _{H₀} (%)	TVP(%)	dsp _{H₁} (%)	NI ^F	ADT (jours)
M _f ^{rm}	9.57	86.62	59.29	90.9	3	219
M _f ^{al}	3.7	86.62	34.97	90.9	3	219
M ₁	4.38	86.62	33.7	90.9	3	218
M ₂	1.28	23.59	31.59	17.39	3	184
M ₃	0.3	23.59	26.3	17.39	3	184
M ₄	0.26	23.9	24.89	17.39	3	184
M ₅	3.51	86.62	35.03	90.9	3	219

Tableau 7.4: Bilan de performances des indicateurs multi-turbine pour pour la modélisation du système de refroidissement de la multiplicatrice de M_f^{rm}

Origine des régresseurs	TFP(%)	dsp _{H₀} (%)	TVP(%)	dsp _{H₁} (%)	NI ^F	ADT (jours)
N _f ^{rc}	0.32	100	30.97	100	5	19
N ₁	31.62	100	66.4	100	5	37

Tableau 7.5: Bilan de performances des indicateurs multi-turbine pour la modélisation du système de refroidissement du convertisseur de N_f^{rc}

Les résultats sont très peu changeants pour les modèles générés par l'utilisation des données des parcs de la technologie E (Tableau 7.2), et on observe le lien entre disponibilité des données résiduelles sur H_1 et taux de vrais positifs pour les données issues de la modélisation du roulement principal (Tableau 7.3). Enfin, à la lecture des valeurs de métriques obtenues pour la surveillance du système de refroidissement de la multiplicatrice de M_f^{rm} (Tableau 7.4), on remarque que les légers gains d'AUC observés lors de l'analyse des courbes ROC s'expriment ici par des taux de faux positifs plus faibles (< 5%), mais au détriment du taux de vrais positifs, inférieur à 50%

pour l'utilisation de données provenant d'autres parcs pour la génération des variables d'entrée du modèle.

Pour chaque technologie de turbine, le NI^F reste inchangé peu importe l'ensemble de régresseurs utilisé pour construire l'indicateur multi-turbine.

Concernant l'avance à la détection, le critère du cahier des charges est respecté pour tous les modèles utilisés : l'ADT reste supérieure à 7 jours.

L'analyse comparative de métriques de performances réalisée permet d'affirmer que les modèles étudiés pour chaque cas de modélisation sont reproductibles à toute turbine provenant d'une même technologie de turbine : les indicateurs multi-turbine en découlant présentent des performances en détection similaires. Par hypothèse, la disponibilité des indicateurs, ainsi que le taux de vrais positifs, seront améliorés par application de la méthode de détection multi-indicateur développée durant le chapitre 6. D'après la méthodologie définie en section 7.2, nous pouvons donc conclure que le processus de sélection de régresseurs LASSO-GFS est généralisable à l'échelle de la technologie de turbine.

7.4 Conclusion

Dans ce chapitre, nous avons pu constater que pour plusieurs parcs appartenant à la même technologie de turbine, les résultats de sélection automatique de régresseurs du processus LASSO-GFS différaient selon le parc considéré, pour un même composant à surveiller. Dans un contexte d'application industrielle, ce résultat implique la manipulation d'un très grand nombre de modèles, ainsi qu'un risque de perte de confiance d'un utilisateur expert envers le processus de surveillance conditionnelle proposé. Cependant, en constatant que les ensembles de variables d'entrée générés par le processus LASSO-GFS, bien que différents, restaient interprétables par un utilisateur expert, et en considérant les hypothèses réalisées et validées dans le chapitre 6 sur l'existence de plusieurs modèles permettant l'estimation d'une variable de sortie avec des performances de détection similaires, nous nous sommes interrogé sur la possibilité de généralisation des résultats du processus LASSO-GFS à l'échelle d'une technologie de turbine. Cela permettrait en effet une réduction drastique du nombre de régresseurs à considérer pour une application industrielle du système de surveillance conditionnelle proposé.

Pour répondre à cette question, nous avons proposé une méthode d'évaluation des impacts du jeu de régresseurs considéré sur les performances en détection, en suivant le cahier des charges établi dans le chapitre 5, l'objectif étant de déterminer si les mêmes régresseurs pouvaient être utilisés pour surveiller un même composant sur tous les parcs d'une même technologie.

Plusieurs hypothèses d'étude ont été émises. Pour simplifier l'étude, nous nous sommes placés dans un cadre de surveillance conditionnelle mono-

indicateur (selon méthodologie du chapitre 5), en faisant l'hypothèse que si les résultats sont généralisables en mono indicateur, ils le seront en multi-indicateur. Nous avons également considéré que les faibles disponibilités en période H_1 et sensibilité au défaut des indicateurs multi-turbine seraient améliorées par l'application de l'approche de détection multi-indicateur défendue dans le chapitre 6.

La méthodologie consiste à comparer sur le parc en défaut les performances en détection des modèles utilisant des régresseurs sélectionnés à partir de données provenant d'autres parcs de même technologie. Elle a été appliquée pour quatre cas de modélisation de composants en défaut, à partir des données des 11 parcs éoliens de la flotte expérimentale présentée dans le chapitre 4. Nous avons pu montrer que les performances en détection restaient équivalentes, quelque soit le modèle utilisé.

Ce constat nous permet de conclure que le processus de sélection de régresseurs LASSO-GFS est généralisable à l'échelle de la technologie de turbine. Ce résultat est par hypothèse applicable à tout parc éolien de la flotte VALEMO, et permet ainsi d'assurer l'**ergonomie** du système de surveillance conditionnelle développé, notamment en vue de son implantation parmi les outils d'exploitation de l'entreprise VALEMO sur des données réelles.

Conclusion

8.1 Résumé général

L'optimisation de la maintenance de la filière éolienne est un levier majeur de la réduction du coût de production de cette énergie électrique renouvelable. Cet objectif peut être atteint par l'approche de maintenance préventive conditionnelle, fondée sur un système de surveillance conditionnelle, qui réalise un suivi en temps réel d'une représentation de l'état de santé des composants. Elle permet en effet de détecter un défaut à un stade de développement suffisamment peu avancé pour qu'une maintenance préventive puisse être organisée sans arrêt de la production. La planification des interventions est ainsi plus efficace, et permet de réduire la fréquence des remplacements de composant.

L'analyse des articles de la littérature traitant du développement d'un tel système de surveillance pour l'éolien a mis en évidence que la majorité des auteurs cherchaient à obtenir des modèles de bon fonctionnement les plus précis possibles sans s'interroger sur leur application dans un contexte industriel. Pourtant, nous avons relevé divers freins à leur utilisation dans ce contexte :

- Les modèles sont très complexes et donc non **interprétables** par utilisateurs experts. Ils sont par ailleurs difficiles à entraîner sur des données réelles d'exploitation de la filière éolienne, qui contiennent de nombreuses perturbations de mesure.
- "Modèle complexe" signifie souvent grand nombre de variables d'entrée, ce qui limite la capacité d'**industrialisation** du système de surveillance conditionnelle à l'échelle d'une flotte, en plus d'augmenter le risque d'**indisponibilité** des modèles.
- Enfin, ces études sont appliquées à un ensemble réduit de composants, et excluent certains composants dont la maintenance occasionne des pertes financières et énergétiques conséquentes.

Les travaux de recherche présentés dans ce manuscrit reposent ainsi sur l'élaboration d'un système de surveillance conditionnelle pour la détection de défaut applicable à l'intégralité d'une flotte d'éolienne industrielle, sur la base des données d'exploitation de l'entreprise VALEMO, au sein de laquelle les travaux ont été réalisés.

Ce système est fondé sur une analyse de résidus générés à partir des modèles de bon fonctionnement, prédisant l'évolution de la température des composants d'une éolienne. Ces modèles sont appris sur des données SCADA pendant des périodes sans défaut. La détection de défaut est alors réalisée par comparaison des valeurs résiduelles à un seuil de détection. Sur la base de l'état de l'art réalisé, notre contribution est de proposer des modèles simples, linéaires et statiques. Ce choix permet en effet de répondre au besoin d'**interprétabilité** (les variations des résidus associés sont facilement compréhensibles par les opérateurs), et les modèles résultants sont robustes à la contrainte de la **qualité de la donnée** (nous utilisons un nombre faible de variables d'entrée afin de limiter l'impact des perturbations de mesure sur la modélisation). Afin d'assurer l'**industrialisation** du processus de surveillance conditionnelle proposé, nous proposons un algorithme de sélection automatique de variables d'entrée de ces modèles, ou régresseurs, pouvant être appliqué à la modélisation de n'importe quel composant de n'importe quelle turbine de la flotte. Ainsi, le système de surveillance conditionnelle ne présente pas de limitation d'échelle d'application. Ce processus a été appliqué à la modélisation de quatre composants présentant un défaut. Par une analyse d'un ensemble de métriques de performance, nous avons pu démontrer que la méthode de détection de défaut proposée ne permettait pas de respecter entièrement le cahier des charges défini pour les performances en détection : le processus peut générer des modèles de bon fonctionnement très peu disponibles, ou dont les résidus sont insensibles aux défauts étudiés. De plus, elle engendre un surcoût de maintenance considérable (dû à des déplacements inutiles de techniciens) en période de bon fonctionnement.

Ces problèmes peuvent être expliqués par les variables d'entrées sélectionnées : elles peuvent être sensibles au défaut (rendant le résidu insensible) ou à des perturbations sur les périodes de bon fonctionnement utilisées (occasionnant ainsi la génération de fausses alarmes). Elles peuvent également présenter des disponibilités très faibles du fait de la faible qualité de la donnée industrielle. Les modèles de bon fonctionnement proposés étant très simples à générer, nous avons alors proposé une méthode de surveillance conditionnelle basée sur l'utilisation d'un groupe de modèles de bon fonctionnement pour l'estimation d'une même variable. Un ensemble de sept indicateurs linéaires construits à partir de ces modèles a été considéré. Plusieurs méthodes de détection "multi-indicateur" (c'est-à-dire basées sur la combinaison de ce groupe d'indicateurs) ont été définies et leurs performances ont été comparées à celles du processus de surveillance conditionnelle initialement conçu à partir d'un unique indicateur (mono-indicateur). Ces méthodes doivent répondre au triple objectif d'assurer la détection des défauts thermiques étudiés, de réduire le bilan d'interventions inutiles en période de bon fonctionnement et de maximiser la disponibilité de l'outil de surveillance conditionnelle. Les résultats ont montré que la méthodologie de

génération de modèles de bon fonctionnement pour l'estimation d'une même variable garantissait la présence d'indicateurs linéaires sensibles au défaut. De plus nous avons observé une augmentation conséquente de la disponibilité du système de surveillance conditionnelle pour toutes les méthodes de détection multi-indicateur considérées. Nous avons pu montrer qu'il était nécessaire de considérer une approche multi-indicateur fondée sur les décisions de détection (ou votes) des sept indicateurs linéaires pour détecter précocement un défaut en conservant un coût de maintenance raisonnable pendant les périodes de bon fonctionnement. Plus précisément, seule celle basée sur le vote majoritaire permet de répondre à l'intégralité des objectifs recherchés. Cette méthode de détection multi-indicateur est ainsi retenue et intégrée comme dernière étape du processus de surveillance conditionnelle développée.

Cette nouvelle approche de surveillance conditionnelle implique un nombre plus important de modèles (et donc de régresseurs) à intégrer au processus, ce qui va à l'encontre du besoin d'**ergonomie** de la solution recherchée. Nous avons donc conclu ce manuscrit par une étude sur la généralisabilité des résultats de l'algorithme de sélection automatique de régresseurs, sur la base de 11 parcs éoliens répartis sur 3 technologies de turbine. L'analyse des résultats a permis de mettre en évidence qu'un ensemble de régresseurs obtenu par notre algorithme de sélection de variables est applicable à toute turbine appartenant à la même technologie de turbine, sans diminution critique des performances. Ce résultat permet de réduire considérablement le nombre de modèles à considérer lors de l'implantation à l'échelle industrielle du système de surveillance conditionnelle développé.

En conclusion, les résultats de ces travaux de recherche sont prometteurs : le système de surveillance conditionnelle pour la détection de défaut développé est applicable sans investissement à n'importe quel élément de la flotte de VALEMO, à la seule condition de disposer de l'acquisition d'une variable de température d'intérêt pour sa modélisation, et permet des bonnes performances de détection en période de fonctionnement normal et en période de défaillances, respectant les besoins industriels et contraintes liées aux données industrielles.

8.2 Implantation expérimentale du processus de surveillance conditionnelle

Afin de confirmer l'applicabilité de la méthode de détection de défaut développée à une échelle industrielle, elle a été implantée durant les dernières semaines des travaux de thèse parmi les outils d'exploitation de l'entreprise VALEMO sur les données en temps réel d'une flotte de parcs éoliens plus grande et contenant plus de technologies de turbine que celle considérée dans

ce manuscrit.

Durant cette phase d'expérimentation, les alertes levées par le système de surveillance conditionnelle présenté dans ce manuscrit ont été analysées chaque semaine, et partagées avec les techniciens d'exploitation concernés, afin d'évaluer la fiabilité des détections générées par la méthode.

Cette implantation a permis de répondre à deux objectifs, pour approfondir l'évaluation des capacités d'industrialisation du système de surveillance conditionnelle développé dans ce manuscrit.

8.2.1 Extension et diversification de l'échantillon expérimental

La méthode de surveillance conditionnelle multi-indicateur a été étudiée dans ce manuscrit sur un échantillon de 11 parcs éoliens, présenté dans le tableau 4.1. Cet échantillon, bien que plus complet et représentatif d'une flotte industrielle d'éoliennes que les échantillons considérés dans la littérature, est un sous-ensemble de la flotte d'éoliennes exploitée par VALEMO, qui exclue une part significative de sa diversité. En effet, comme mentionné dans le chapitre 4, celle-ci se compose de 70 parcs éoliens de 24 technologies de turbine différentes.

Ainsi, le premier objectif de cette phase expérimentale de l'implantation des travaux de recherche réalisés a été de considérer un échantillon de parcs éoliens le plus représentatif possible de la flotte VALEMO, afin de valider les capacités d'industrialisation du processus de surveillance conditionnelle construit.

Nous avons ainsi utilisé une flotte expérimentale contenant un parc de chaque technologie de turbine exploitée par VALEMO, afin d'obtenir un ensemble le plus représentatif possible de la flotte exploitée par l'entreprise. Tous les parcs de cette flotte expérimentale sont exploités par au moins un technicien d'exploitation, avec qui les alertes ont été partagées.

8.2.2 Développement de la base de données de défauts

Cette expérimentation a été réalisée à partir de données d'exploitation en temps réel. Cela signifie que, contrairement aux études réalisées dans les chapitres 5, 6 et 7, on a aucune connaissance sur les potentiels défauts survenant durant cette phase d'expérimentation : on ne sait pas quand ils débiteront et finiront, comment ils évolueront, quels composants collatéraux ils affecteront, et quels seraient les coûts de maintenance associés à une non détection.

Le deuxième objectif de cette phase d'implantation expérimentale est ainsi d'étendre le spectre de profils de défaut détectables par l'outil de surveillance conditionnelle développé. Cela permettra par ailleurs d'évaluer l'apport de ce nouvel outil de surveillance pour l'amélioration du service d'exploitation de l'entreprise VALEMO, en comparant les détections réalisées

à celles des autres systèmes de surveillance conditionnelle (analyses de courbes de puissance, écarts à la médiane, supervision SCADA, analyse vibratoire).

Les résultats de cette période d'expérimentation de l'outil de surveillance conditionnelle sont plutôt positifs : plusieurs anomalies ont été détectées par l'outil, et ont été remontées aux fabricants pour suivi. Parmi celles-ci certaines sont connues des techniciens d'exploitation, comme les défauts du système de refroidissement des convertisseurs du parc N_f^{rc} , causés par le liquide utilisé, se cristallisant avec le temps, impliquant une mauvaise circulation du fluide. D'autres ne semblent avoir jamais été détectées par les outils d'exploitation jusqu'alors en place chez VALEMO, et illustrent ainsi des potentiels défauts dont on ne connaît pas encore les conséquences d'une non prise en charge. C'est le cas d'un échauffement anormalement rapide d'un roulement de la multiplicatrice d'un parc de la technologie de turbine G, non étudiée dans ce manuscrit.

Cette phase d'implantation a également permis d'identifier plusieurs recommandations pour les utilisateurs industriels, ainsi que quelques champs de recherche prospectifs pour améliorer les performances de l'outil, que nous allons développer dans la section suivante.

8.3 Perspectives

8.3.1 Recommandations pour les utilisateurs industriels

Extension et contrôle de la bases de données SCADA

Dans cette thèse, on a construit un processus complet de surveillance conditionnelle adapté aux contraintes et besoins industriels, mais dont le potentiel reste dépendant du contenu de la base de données SCADA.

En effet, comme on l'a vu sur la Figure 4.5 avec l'exemple du parc M_f^{al} , certains parcs présentent de gros problèmes de pertes de données (mise à jour logiciel, capteurs ou systèmes d'acquisition de données dysfonctionnels). De plus, un nombre significatif de parcs de la flotte VALEMO n'ont pas été intégrés à l'étude car l'ensemble de variables SCADA disponibles était insuffisant : aucune des variables de sortie considérées (Tableau 4.2) n'était disponible.

La relative qualité de données résultante peut fortement impacter l'efficacité du processus de surveillance conditionnelle développé sur différents aspects:

- En plus d'impacter la disponibilité de la solution de surveillance conditionnelle, les pertes de données réduisent la quantité d'information disponible en période d'apprentissage, augmentant ainsi le risque de générer et d'apprendre des modèles sous-optimaux.

- Outre la limitation du nombre de parcs intégrables à la solution de surveillance conditionnelle développée, l'amputation des variables disponibles dans les bases de données peut réduire le nombre de composant modélisables, et limiter le potentiel de l'approche de sélection de régresseurs LASSO-GFS : en effet, les possibilités de combinaison de variables à évaluer par le processus de sélection sont alors réduites, et on augmente le risque d'écarter des variables pertinentes pour la modélisation de y .

Redéfinir le contenu des bases de données SCADA qu'une entreprise O&M telle que VALEMO doit exploiter (en négociation avec le constructeur) permettrait ainsi d'assurer la modélisation et la surveillance conditionnelle de tous les composants d'une éolienne. Il pourrait être ainsi possible de modéliser d'autres composants que ceux considérés dans cette thèse : par exemple le système Yaw, contrôlant l'orientation de la nacelle pour maximiser la prise au vent, qui peut être modélisé par la température d'un de ses quatre moteurs. Cela permettrait également d'améliorer la précision des modèles considérés sans complexifier le système de surveillance conditionnelle. Enfin, les entreprises O&M doivent augmenter leurs efforts de contrôle de la disponibilité et de la qualité de la donnée exploitée pour maximiser la disponibilité de leurs outils de surveillance conditionnelle, et limiter les fausses alarmes occasionnées par des valeurs aberrantes, symptomatiques de défauts de capteurs.

Fusion des bases de données par technologie

À l'issue du chapitre 7, nous avons pu démontrer que les variables d'entrée générées par le processus de sélection LASSO-GFS appliqué sur un parc d'une technologie de turbine peuvent être appliquées à n'importe quelle turbine provenant de la même technologie. Par extension, et en considérant l'homogénéité des bases de données de parcs éoliens appartenant à la même technologie, on peut donc faire l'hypothèse que les bases de données de tous les parcs éoliens d'une même technologie peuvent être fusionnées et utilisées en entrée du processus LASSO-GFS pour générer un ensemble de modèles unique pour la surveillance de tous les composants de n'importe quelle turbine de cette technologie.

Nous avons pu observer lors de la phase d'expérimentation de la méthode de détection de défaut sur l'ensemble des technologies de turbine de VALEMO que les performances de détection des indicateurs multi-turbine étaient dégradées pour les "petits" parcs éoliens, contenant 3 ou 4 variables. Cela s'explique par le fait qu'une perturbation de mesure impactant le résidu d'une turbine d'un parc de 3 éoliennes a une influence sur le résidu médian bien plus grande que si ce parc contient 10 éoliennes par exemple.

Or, en réalisant cette fusion de bases de données, on considère un parc

éolien "virtuel" (d'un point de vue géographique) composé d'un très grand nombre d'éoliennes, égal à la somme des éoliennes de tous les parcs éoliens d'une technologie de turbine. Par conséquent, cette extension des bases de données d'étude permet aux étapes "multi-turbine" du processus de surveillance conditionnelle (estimation des coefficients de régression LASSO, calcul du critère de sélection de l'algorithme GFS, construction de l'indicateur multi-turbine) d'être plus robustes aux perturbations dans les mesures SCADA.

8.3.2 Champs de recherche prospectifs

Amélioration de la sélection des données d'apprentissage

Durant la période d'apprentissage H_L , les turbines d'un parc éolien sont supposées être en fonctionnement normal, ce qui signifie qu'elles ne présentent pas de maintenance corrective (remplacement de composant). Toutefois, elles peuvent présenter des interventions de maintenance préventive, faisant suite à des perturbations impactant les variations des données SCADA (décalage de la moyenne pour une mauvaise fixation ou calibration de capteur, "bond" de valeurs pour un choc externe...).

Par conséquent, cette simplification de la définition des données de bon fonctionnement ne permet pas d'écarter toutes les perturbations ayant un impact significatif sur la distribution statistique des variables utilisées pour construire les modèles, et donc sur la précision des modèles.

En vue d'assurer une meilleure précision de prédiction pour tous les modèles générés par le processus de surveillance conditionnelle proposé, il serait donc pertinent de développer une méthode de détection automatique de perturbations au niveau des données SCADA afin de construire de façon automatique des périodes d'apprentissage H_L plus représentatives d'un fonctionnement normal.

De telles méthodes de prétraitement de données avancées peuvent consister en des algorithmes de détection automatique de défaut de capteurs. Ces algorithmes peuvent être distingués en plusieurs approches, selon la revue d'Erhan [14] : tout d'abord les méthodes statistiques, permettant de détecter des changements statistiques. Ces algorithmes sont simple à implanter et peu coûteux, et on déjà fait leurs preuves en application à l'énergie éolienne [68]. Nous pouvons également mentionner les méthodes d'analyses temporelles [41], spectrales, utilisant notamment l'analyse de composants principaux [21], ou encore celles issues de la théorie de l'information, fondées sur l'entropie des signaux analysés [28].

Mise en place d'une routine de ré-apprentissage des modèles

La plupart des modèles considérés dans cette thèse ont été appris à partir de données proches des données de test H_0 , comme on peut le voir dans le

Tableau 4.1. Les modèles permettent une estimation suffisamment précise de y sur H_0 , se distanciant significativement de y en période de défaut. Cependant, le système éolienne est non-stationnaire, et cette non-stationnarité s'illustre notamment au niveau du réseau de capteur : les variations des composants mesurées dépendent de différents facteurs tels leur âge (dégradation "naturelle"), l'évolution des conditions météorologiques et climatiques (se traduisant par un changement des référentiels de température de fonctionnement normal), ou simplement un changement de technologie (de capteur ou de composants). La distribution statistique des variables SCADA n'est donc pas stable dans le temps, et peut être amenée à changer de façon permanente.

Ainsi, la conservation des performances des modèles sur H_0 et H_1 n'est pas assurée dans le temps, et il est fort probable que la paramétrisation des modèles devienne obsolète durant le cycle de fonctionnement d'une éolienne (autour de 15 ans). Il sera donc nécessaire durant ses premières années d'utilisation de suivre l'évolution des performances des indicateurs en période de fonctionnement normal, afin de déterminer les conditions de ré-apprentissage. Ce travail se traduit notamment par une analyse de la robustesse de la précision des modèles aux changements de technologie (à la suite d'un remplacement de composant par exemple), aux aléas climatiques, ou tout simplement à la durée de fonctionnement des composants.

Prise en compte de l'influence de la charge

Lebranchu [27] a mis en évidence une relation entre les variations résiduelles fournies par les indicateurs multi-turbine et la charge (c'est-à-dire la vitesse de vent mesurée) de l'éolienne. Comme nous l'avons vu dans le chapitre 6 pour la construction du modèle physique, les températures de tous les composants d'une éolienne sont en effet reliés à la puissance produite. Or, cette puissance présente des modes d'évolution différents selon la gamme de vitesse de vent, ou charge, mesurée (voir Figure 2.2).

Durant les périodes de fonctionnement normal, une éolienne est majoritairement en mode de production transitoire (phase II de la Figure 2.2). Ainsi, sur des données de test H_0 , lorsque l'éolienne fonctionne dans un autre mode de production (par exemple à puissance nominale), les relations apprises entre les températures modélisées et la charge ne sont plus valides, ce qui perturbe les valeurs résiduelles mesurées.

Ainsi, une perspective d'amélioration du processus de surveillance conditionnelle en vue de réduire le bilan de fausses alarmes serait de considérer plusieurs paramétrisations de modèles selon le mode de production. En pratique, cela se traduirait par une division des données d'apprentissage selon le mode de production considéré (déterminé à partir de la vitesse de vent mesurée), et l'application de la régression linéaire sur les différentes périodes d'apprentissage obtenues.

**Description des variables
SCADA de la base de
données**

182 ANNEXE A. DESCRIPTION DES VARIABLES SCADA DE LA BASE DE DONNÉES

Variable	Description	Unité
$v_{arbrLent}$	vitesse de rotation de l'arbre lent	tour/min
$v_{arbrRapide}$	vitesse de rotation de la génératrice	tour/min
$v_{generatrice}$	vitesse de rotation de la génératrice de référence	tour/min
v_{vent}	vitesse du vent mesurée par l'anémomètre actif	m/s
v_{vent1}	vitesse du vent mesurée par l'anémomètre primaire	m/s
v_{vent2}	vitesse du vent mesurée par l'anémomètre secondaire	m/s
d_{vent}	position du vent	°
$d_{nacelle}$	nacelle position	°
$t_{exterieure}$	température ambiante extérieure	°C
$t_{interne}$	température ambiante dans la nacelle	°C
$t_{paliArbrLent}$	température du palier de l'arbre lent (ou roulement principal)	°C
$t_{roulMulti1}$	température du roulement de la multiplicatrice, côté rotor	°C
$t_{roulMulti2}$	température du roulement de la multiplicatrice, côté génératrice	°C
$t_{roulGene1}$	température du roulement avant de la génératrice	°C
$t_{roulGene2}$	température du roulement arrière de la génératrice	°C
t_{stator}	température du stator de la génératrice	°C
$t_{generatrice}$	température du liquide de refroidissement de la génératrice	°C
$t_{transformateur}$	température de l'armoire contenant le transformateur	°C
$t_{transformateur1}$	température de la bobine du transformateur, phase 1	°C
$t_{transformateur2}$	température de la bobine du transformateur, phase 2	°C
$t_{transformateur3}$	température de la bobine du transformateur, phase 3	°C
$t_{convertisseur}$	température du liquide de refroidissement de l'onduleur	°C
$t_{convertisseur2}$	température de refroidissement du convertisseur	°C
$t_{huileMulti}$	température de l'huile de la multiplicatrice	°C
t_{rotor}	température au niveau du rotor	°C
t_{hub}	température au niveau du hub	°C
$t_{automatisme}$	température dans ou aux alentours de l'unité de contrôle "maître"	°C
$t_{piedMat}$	température au pied du mât de l'éolienne	°C
c_{angle1}	position de référence (système Pitch) la pale 1	°
c_{angle2}	position de référence (système Pitch) la pale 2	°
c_{angle3}	position de référence (système Pitch) la pale 3	°
$a_{reelPale1}$	position de la pale 1	°
$a_{reelPale2}$	position de la pale 2	°
$a_{reelPale3}$	position de la pale 3	°
$v_{simpProduite}$	tension simple produite, phase R	Volt
$i_{produit}$	courant produite, phase R	A
$p_{apparente}$	puissance apparente	W
ρ_{yaw}	pression de la station de pompage du système Yaw	bar
ρ_{frein}	pression des freins	bar
$\rho_{mecanique}$	pression mécanique dans la multiplicatrice	bar
$\rho_{huileMulti}$	pression électrique dans la multiplicatrice	bar

Tableau A.1: Description des variables SCADA disponibles

Bibliography

- [1] Tyler J Arsenault et al. “Development of a FBG based distributed strain sensor system for wind turbine structural health monitoring”. en. In: *Smart Mater. Struct.* (2013).
- [2] Henrique D M Azevedo. “Vibration monitoring, fault detection, and bearings replacement of a real wind turbine”. en. In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* (2017).
- [3] Hamed Badihi et al. “A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis”. en. In: *Proceedings of the IEEE* 110.6 (June 2022), pp. 754–806.
- [4] Yu Bie et al. “A review of the application of oil analysis in condition monitoring and life prediction of wind turbine gearboxes”. en. In: *Insight - Non-Destructive Testing and Condition Monitoring* 63.5 (May 2021), pp. 289–301.
- [5] Riccardo Bonanno, Francesca Viterbo, and Riva Giuseppe Maurizio. “Climate change impacts on wind power generation for the Italian peninsula”. en. In: *Regional Environmental Change* 23.1 (Dec. 2022), p. 15.
- [6] Philippe Cambron et al. “Bearing temperature monitoring of a Wind Turbine using physics-based model”. In: *Journal of Quality in Maintenance Engineering* 23.4 (Jan. 2017). Publisher: Emerald Publishing Limited, pp. 479–488.
- [7] Iñigo Capellán-Pérez. “Dynamic Energy Return on Energy Investment (EROI) and material requirements in scenarios of global transition to renewable energies”. en. In: *Energy Strategy Reviews* (2019).
- [8] James Carroll, Alasdair McDonald, and David McMillan. “Failure Rate, Repair Time and Unscheduled O&M Cost Analysis of Offshore Wind Turbines”. en. In: *wind energy* (2015).
- [9] Yue Cui, Jose Eduardo Urrea Cabus, and Lina Bertling Tjernberg. “A Fault Detection Approach Based on Autoencoders for Condition Monitoring of Wind Turbines”. en. In: *Women in Renewable Energy*. Ed. by Katherine T. Wang and Jill S. Tietjen. Cham: Springer International Publishing, 2023, pp. 193–211.
- [10] Thomas G. Dietterich. “Ensemble Methods in Machine Learning”. In: *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15. ISBN: 978-3-540-45014-6.

- [11] S. Djurovic et al. “Condition monitoring of wind turbine induction generators with rotor electrical asymmetry”. en. In: *IET Renewable Power Generation* 6.4 (July 2012). Publisher: IET Digital Library, pp. 207–216.
- [12] Franc Drobnič, Andrej Kos, and Matevž Pustišek. “On the Interpretability of Machine Learning Models and Experimental Feature Selection in Case of Multicollinear Data”. en. In: *Electronics* 9.5 (May 2020), p. 761.
- [13] Mian Du, Shichong Ma, and Qing He. “A SCADA data based anomaly detection method for wind turbines”. In: *2016 China International Conference on Electricity Distribution (CICED)*. ISSN: 2161-749X. Aug. 2016, pp. 1–6.
- [14] L. Erhan et al. “Smart Anomaly Detection in Sensor Systems: A Multi-Perspective Review”. In: *Information Fusion* 67 (Mar. 2021). arXiv:2010.14946 [cs], pp. 64–79.
- [15] Jian Fu et al. “Condition Monitoring of Wind Turbine Gearbox Bearing Based on Deep Learning Model”. en. In: *IEEE Access* 7 (2019).
- [16] Mari Cruz Garcia, Miguel A. Sanz-Bobi, and Javier del Pico. “SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox”. In: *Computers in Industry. E-maintenance Special Issue* 57.6 (Aug. 2006), pp. 552–568.
- [17] Elena Gonzalez et al. “Using high-frequency SCADA data for wind turbine performance monitoring: a sensitivity study”. en. In: *Renewable Energy* (2019).
- [18] Christopher J. Greenwood et al. “A comparison of penalised regression methods for informing the selection of predictive markers”. en. In: *PLOS ONE* 15.11 (Nov. 2020). Ed. by Yuka Kotozaki, e0242730.
- [19] Fouzi Harrou et al. “Uncovering sensor faults in wind turbines: An improved multivariate statistical approach for condition monitoring using SCADA data”. en. In: *Sustainable Energy* (2023).
- [20] Georg Helbing. “Deep Learning for fault detection in wind turbines”. en. In: *Renewable and Sustainable Energy Reviews* (2018).
- [21] Ling Huang et al. “In-Network PCA and Anomaly Detection”. In: *Advances in Neural Information Processing Systems*. Vol. 19. MIT Press, 2006.
- [22] Prince Waqas Khan, Chan Yeob Yeun, and Yung Cheol Byun. “Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning”. In: *Engineering Failure Analysis* 148 (June 2023), p. 107209.

- [23] Samir Khan et al. “A review of physics-based learning for system health management”. In: *Annual Reviews in Control* 57 (Jan. 2024), p. 100932.
- [24] Vipin Kumar and Sonajharia Minz. “Feature Selection: A literature Review”. en. In: *Smart Computing Review* 4.3 (2014).
- [25] Andrew Kusiak and Anoop Verma. “Analyzing bearing faults in wind turbines: A data-mining approach”. In: *Renewable Energy* 48 (Dec. 2012), pp. 110–116.
- [26] Alexis Lebranchu. “Analyse de données de surveillance et synthèse d’indicateurs de défauts et de dégradation pour l’aide à la maintenance prédictive de parcs de turbines éoliennes”. PhD thesis. École doctorale électronique, électrotechnique, automatique, traitement du signal, 2016.
- [27] Alexis Lebranchu et al. “A combined mono- and multi-turbine approach for fault indicator synthesis and wind turbine monitoring using SCADA data”. In: *ISA Transactions* 87 (Apr. 2019). Publisher: Elsevier, pp. 272–281.
- [28] Wenke Lee and Dong Xiang. “Information-Theoretic Measures for Anomaly Detection”. en. In: *Proceedings* (2001).
- [29] Simon Letzgus. “Change-point detection in wind turbine SCADA data for robust condition monitoring with normal behaviour models”. English. In: *Wind Energy Science* 5.4 (Oct. 2020). Publisher: Copernicus GmbH, pp. 1375–1397.
- [30] Jian Li et al. “Normal Behavior Models for the Condition Assessment of Wind Turbine Generator Systems”. In: *Electric Power Components and Systems* 42.11 (Aug. 2014), pp. 1201–1212. ISSN: 1532-5008.
- [31] L A M Lima, A Blatt, and J Fujise. “Wind Turbine Failure Prediction Using SCADA Data”. en. In: *Journal of Physics: Conference Series* 1618 (Sept. 2020), p. 022017. ISSN: 1742-6588, 1742-6596.
- [32] Yonggang Lin. “Fault analysis of wind turbines in China”. en. In: *Renewable and Sustainable Energy Reviews* 55 (2016), p. 9.
- [33] Jui-Hung Liu and Nelson T Corbita. “Wind Turbine Anomaly Detection Using Mahalanobis Distance and SCADA Alarm Data”. en. In: *Applied Science* 12 (2022).
- [34] Julie K Lundquist, Andrew Clifton, and Scott Dana. “Wind Energy Instrumentation Atlas”. en. In: *Renewable Energy* (2019).
- [35] M. Lydia et al. “A comprehensive review on wind turbine power curve modeling techniques”. en. In: *Renewable and Sustainable Energy Reviews* 30 (Feb. 2014), pp. 452–460.

- [36] Jorge Maldonado-Correa et al. “Using SCADA Data for Wind Turbine Condition Monitoring: A Systematic Literature Review”. en. In: *Energies* 13.12 (Jan. 2020). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 3132.
- [37] Pere Marti-Puig et al. “Effects of the pre-processing algorithms in fault diagnosis of wind turbines”. en. In: *Environmental Modelling & Software* 110 (Dec. 2018), pp. 119–128.
- [38] Pere Marti-Puig et al. “Feature Selection Algorithms for Wind Turbine Failure Prediction”. en. In: *Energies* 12 (2019).
- [39] A.P. Marugán et al. “A survey of artificial neural network in wind energy systems”. English. In: *Applied Energy* 228 (2018), pp. 1822–1836.
- [40] Angela Meyer. “Multi-target normal behaviour models for wind farm condition monitoring”. en. In: *Applied Energy* 300 (Oct. 2021), p. 117342.
- [41] Nawaz Mohamudally et al. “Building an anomaly detection engine (ADE) for IoT smart applications”. en. In: *Procedia Computer Science* (2018).
- [42] A.R. Nunes, H. Morais, and A. Sardinha. “Use of learning mechanisms to improve the condition monitoring of wind turbine generators: A review”. English. In: *Energies* 14.21 (2021).
- [43] Abdul Ghani Olabi et al. “A Review on Failure Modes of Wind Turbine Components”. en. In: *Energies* 14 (2021).
- [44] Ravi Kumar Pandit, Davide Astolfi, and Isidro Durazo Cardenas. “A Review of Predictive Techniques Used to Support Decision Making for Maintenance Operations of Wind Turbines”. en. In: *Energies* 16 (2023).
- [45] Seong Ho Park, Jin Mo Goo, and Chan-Hee Jo. “Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists”. en. In: *Korean Journal of Radiology* 5.1 (2004), p. 11.
- [46] H. Peng et al. “A Review of Research on Wind Turbine Bearings’ Failure Analysis and Fault Diagnosis”. English. In: *Lubricants* 11.1 (2023).
- [47] Han Peng et al. “Analysis of Wind Turbine Equipment Failure and Intelligent Operation and Maintenance Research”. en. In: *sustainability* 15 (2023).
- [48] Sebastian Pfaffel, Stefan Faulstich, and Dr Kurt Rohrig. “Performance and Reliability of Wind Turbines: A Review”. en. In: *Energies* 10 (2017).

- [49] J. Pohjankukka et al. “Comparison of estimators and feature selection procedures based on ALS and digital aerial imagery”. en. In: *Scandinavian Journal of Forest Research* 33.7 (Oct. 2018), pp. 681–694. ISSN: 0282-7581, 1651-1891.
- [50] J Ranstam and J A Cook. “LASSO regression:” en. In: *British Journal of Surgery* 105 (2018).
- [51] Théodore Raymond et al. “Data-driven Model Generation Process for Thermal Monitoring of Wind Farm Main Components through Residual Indicators Analysis”. en. In: *ESREL 2022 - 32nd European Safety and Reliability Conference (ESREL 2022), TU Dublin; ESRA, Aug 2022, Dublin, Ireland* 32 (Nov. 2022), p. 9.
- [52] Théodore Raymond et al. “On the Joint Use of An Ensemble of Linear Residuals to Improve Fault Detection in Wind Turbines”. In: *ESREL 2023 - 33rd European Safety and Reliability Conference*. Ed. by Mario P. Brito et al. Backup Publisher: University of Southampton and European Safety and Reliability Association. Southampton, United Kingdom: Research Publishing Services, Sept. 2023, pp. 1784–1791.
- [53] M Reder and J J Melero. “Modelling the Effects of Environmental Conditions on Wind Turbine Failures”. en. In: *Wind Energy* 21 ().
- [54] Meik Schlechtingen and Ilmar Ferreira Santos. “Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection”. en. In: *Mechanical Systems and Signal Processing* 25.5 (July 2011), pp. 1849–1875.
- [55] Meik Schlechtingen and Ilmar Ferreira Santos. “Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: Application”. en. In: *Applied Soft Computing* 14 (Jan. 2014), pp. 447–460. ISSN: 15684946.
- [56] Meik Schlechtingen, Ilmar Ferreira Santos, and Sofiane Achiche. “Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description”. en. In: *Applied Soft Computing* 13.1 (Jan. 2013), pp. 259–270. ISSN: 15684946.
- [57] Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. New York: Routledge, Oct. 2017.
- [58] Aurore Stéphant. “Transition énergétique : une nécessaire intégration des impacts environnementaux de l’industrie minière”. FR. In: *Revue internationale et stratégique* 128.4 (2022). Publisher: IRIS éditions, pp. 95–103.
- [59] Adrian Stetco et al. “Machine learning methods for wind turbine condition monitoring: A review”. en. In: *Renewable Energy* 133 (Apr. 2019), pp. 620–635.

- [60] Jialin Tang et al. “An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades”. In: *Renewable Energy* 99 (Dec. 2016), pp. 170–179.
- [61] J Tautz-Weinert and S J Watson. “Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection”. en. In: *Journal of Physics* (2016).
- [62] Pierre Tchakoua et al. “Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges”. en. In: *Energies* 7 (2014).
- [63] Wei Teng et al. “Vibration Analysis for Fault Detection of Wind Turbine Drivetrains—A Comprehensive Investigation”. en. In: *Sensors* 21 (2021).
- [64] Christian Tutiv’en et al. “Wind Turbine Main Bearing Condition Monitoring via Convolutional Autoencoder Neural Networks”. In: *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. Oct. 2021, pp. 01–06.
- [65] B. Venkatesh and J. Anuradha. “A Review of Feature Selection and Its Methods”. en. In: *Cybernetics and Information Technologies* 19.1 (Mar. 2019), pp. 3–26.
- [66] Lucas Vimpere et al. “Carbon isotope and biostratigraphic evidence for an expanded Paleocene–Eocene Thermal Maximum sedimentary record in the deep Gulf of Mexico”. In: *Geology* 51.4 (Feb. 2023), pp. 334–339.
- [67] Long Wang, Jia Xu, and Ruihua Liu. “Wind Turbine Gearbox Failure Identification with Deep Neural Networks”. en. In: *IEEE Transactions on Industrial Informatics* 13 (2017).
- [68] Xiukun Wei, Michel Verhaegen, and Tim van Engelen. “Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques”. en. In: *International Journal of Adaptive Control and Signal Processing* 24.8 (2010). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/acs.1162>, pp. 687–707.
- [69] Xiaocong Xiao et al. “Condition Monitoring of Wind Turbine Main Bearing Based on Multivariate Time Series Forecasting”. en. In: *Energies* 15.5 (Jan. 2022). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 1951.
- [70] Wenxian Yang, Richard Court, and Jiesheng Jiang. “Wind turbine condition monitoring by the approach of SCADA data analysis”. In: *Renewable Energy* 53 (May 2013), pp. 365–376.

- [71] Mao Ye et al. “Good Subnetworks Provably Exist: Pruning via Greedy Forward Selection”. In: *Proceedings of the 37th International Conference on Machine Learning*. Proceedings of Machine Learning Research 119 (July 2020). Ed. by Hal Daumé III and Aarti Singh, pp. 10820–10830.
- [72] Tongke Yuan, Zhifeng Sun, and Shihao Ma. “Gearbox Fault Prediction of Wind Turbines Based on a Stacking Model and Change-Point Detection”. en. In: *Energies* 12 (2019).
- [73] Samira Zare et al. “Convolutional neural networks for wind turbine gearbox health monitoring. — Energy Equipment & Systems — EBSCOhost”. fr. In: *Energy Equipment and Systems* (Mar. 2022).
- [74] Xiangjun Zeng, Ming Yang, and Senior Member. “A Generalized Wind Turbine Anomaly Detection Method Based on Combined Probability Estimation Model”. en. In: *Journal of Modern Power Systems and Clean Energy* 11 (2023).
- [75] Zhenyou Zhang. “Comparison of Data-driven and Model-based Methodologies of Wind Turbine Fault Detection with SCADA Data”. en. In: *European Wind Energy Association Annual Event* (2012).