



**HAL**  
open science

# 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

Alexis Lebranchu

## ► To cite this version:

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. Automatique / Robotique. Communauté Université Grenoble Alpes, 2016. Français. NNT: . tel-01503571v1

**HAL Id: tel-01503571**

**<https://theses.hal.science/tel-01503571v1>**

Submitted on 7 Apr 2017 (v1), last revised 25 Apr 2017 (v2)

**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 LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Automatique et Productique**

Arrêté ministériel : 25 mai 2016

Présentée par

**Alexis LEBRANCHU**

Thèse dirigée par **Sylvie CHARBONNIER** et  
codirigée par **Christophe BERENGUER** »

préparée au sein du **Grenoble Images Parole Signal  
Automatique (GIPSA-Lab)**  
dans l'**École Doctorale d'Électronique, Électrotechnique,  
Automatique et Traitement du Signal (EEATS)**

## ANALYSE DE DONNÉES DE SURVEILLANCE ET SYNTHÈSE D'INDICATEURS DE DÉFAUTS ET DE DÉGRADATION POUR L'AIDE À LA MAINTENANCE PRÉDICTIONNELLE DE PARCS DE TURBINES ÉOLIENNES

Thèse soutenue publiquement le **9 novembre 2016**  
devant le jury composé de :

**Pr, Vincent, COCQUEMPOT**

Professeur, Université Lille- Science et technologie, Président du jury

**Pr, Antoine, GRALL**

Professeur, Université de technologie de Troyes, Rapporteur

**Pr, Moamar, SAYED-MOUCHAWEH**

Professeur, École nationale supérieure des mines de Douai, Rapporteur

**Dr, Frédéric PREVOST**

Directeur, VALEMO SAS Bègles, Examineur

**Dr, Sylvie, CHARBONNIER**

Maitre de conférences, Université Grenoble Alpes, Directrice de thèse

**Pr, Christophe, BERENGUER**

Professeur, Grenoble-INP, Co-Directeur de thèse





## Remerciements

Premièrement, j'aimerais remercier profondément ma directrice et mon directeur de thèse, Sylvie CHARBONNIER et Christophe BERENGUER pour m'avoir accompagné durant ces années de thèse et pour l'intérêt sincère qu'ils ont eu sur tous les aspects de ce projet. Leurs enseignements, leur esprit critique et leurs soutiens pendant les moments les plus difficiles m'ont permis de me dépasser professionnellement et personnellement pour pouvoir mener cette thèse à son terme. Grâce à leurs nombreuses relectures et encore plus nombreux retours, j'ai pu tordre le cou à cette mystérieuse problématique et y apporter une réponse.

Pendant, cette thèse j'ai travaillé au sein de l'entreprise VALEMO qui a fourni les données et les informations nécessaires à l'élaboration de ce document. Je remercie chaleureusement son directeur et mon encadrant Frédéric PREVOST pour m'avoir fait confiance durant ces années de recherche et pour me faire confiance aujourd'hui pour poursuivre mes travaux et appliquer les méthodes développées. Je remercie aussi toute l'équipe de VALEMO pour m'avoir aussi bien accueilli. Cette bonne ambiance de travail qui leur est si chère m'a donné la chance de pouvoir travailler dans les meilleures conditions. Je les remercie aussi pour leur dynamisme, leur professionnalisme et pour avoir su partager avec moi un peu de leurs passions pour leurs parcs éoliens. Cela m'a tenu motivé et m'a tout de suite permis de percevoir l'intérêt de mes travaux dans leurs questionnements et aussi dans des applications concrètes. Pour m'avoir expliqué tout ce qu'il y a à savoir dans l'éolien et pour leur bonne humeur, je remercie mes premiers collègues de bureau Frank VEZIEN et Alban NOEL. VALEMO faisant partie d'un groupe plus important, VALOREM, je remercie tous les membres du groupe VALOREM avec qui j'ai partagé d'incroyables séminaires, pour leurs qualités humaines, leur tolérance et leur ouverture d'esprit. Je remercie aussi tout particulièrement ma chère amie Amandine ADANS qui m'a soutenu dans les moments les plus difficiles et avec qui j'ai pu partager tous les meilleurs moments de cette thèse.

Je remercie enfin toute ma famille pour leur aide et leur soutien, en particulier ma mère Edith et mon frère Guillaume sans qui je ne serais sans doute jamais arrivé jusque-là. Finalement, je voudrais avoir une pensée pour mon père Marc qui nous a quitté trop tôt et à qui je dédie cette thèse.



# Notations

## Acronymes

<b>ENR</b>	Énergie renouvelable
<b>SCADA</b>	Système de contrôle et d'acquisition de données de l'anglais Supervisory Control And Data Acquisition
<b>CMS</b>	Système de surveillance conditionnel, de l'anglais condition monitoring system
<b>SoES</b>	Service de l'Observation et des Statistiques
<b>WT</b>	Aérogénérateur ou éolienne, de l'anglais Wind Turbine
<b>HAWT</b>	Éolienne à axe horizontal, de l'anglais horizontal axis Wind turbine
<b>ROC</b>	Receiver operating characteristic
<b>SIMAP</b>	Intelligent System for Predictive Maintenance Application

## Unités

<b>W et kW</b>	Watt et kW
<b>min</b>	Minute
<b>m/s</b>	Mètre par seconde
<b>°C</b>	Degré Celsius

## Symboles

$N_A$	Nombre d'échantillons utilisés pour l'apprentissage des modèles
$N_E(t)$	Nombre d'éolienne en fonctionnement à l'instant $t$
$f$	Nombre d'échantillons utilisés pour calculer une moyenne mobile
$i$	Le numéro de l'éolienne
$j$	Le numéro du composant ou partie du composant
$N_T$	Le nombre d'éolienne du parc
$T_{rav}$	Température du roulement avant
$T_{rar}$	Température du roulement arrière
$T_{es}$	Température des enroulements statoriques

## Table des figures

Figure 2.1 Évolution de la taille des éoliennes (Edenhofer and IPCC, 2012).....	18
Figure 2.2 Évolution de la puissance éolienne en France d'après les chiffres du site du service de l'Observation et des Statistiques (SOeS) du ministère de l'Environnement français, de l'Énergie et de la Mer de 2015.....	18
Figure 2.3 Profil de courbe courant liant la vitesse du vent à la puissance générée d'une éolienne .....	21
Figure 2.4 Description de l'intérieur d'une nacelle éolienne HAWT .....	22
Figure 2.5 Modes de fonctionnement : Puissance moyenne, vitesse de rotation et angle de pale en fonction de la vitesse du vent. La puissance moyenne et la vitesse de rotation sont exprimées en pourcentage de leur valeur maximale .....	24
Figure 2.6 Évolution Vitesse du vent (m/s) et puissance active (MW) mesurées toutes les 10 minutes pendant 9 jours (source VALEMO) .....	25
Figure 3.1 Moyenne et écart type de la puissance générée mesurés en fonction de la vitesse du vent (source VALEMO) .....	37
Figure 3.2 Courbe de puissance réelle en rouge versus courbe garantie par le constructeur en noire (source (McLaughlin et al., 2009)fig.1) .....	38
Figure 3.3 Exemple de contrôle commande défaillant mis en évidence par la courbe de puissance (source (McLaughlin et al., 2009) fig.4) .....	39
Figure 3.4 Courbe de puissance normalisée mesurée pendant 1 an où la couleur est fonction de la densité (source (Butler et al., 2013) fig.4) .....	41
Figure 3.5 Courbe de puissance avec le découpage en familles (source (Elijorde et al., 2014)fig.2) .....	43
Figure 3.6 Équilibre énergétique d'un composant de la chaîne cinématique (source (Wilkinson et al., 2014) fig.5).....	45
Figure 3.7 Exemple de déviation du modèle pour la température du roulement principal, en gris les valeurs brutes et en noir la moyenne mobile (source (Wilkinson et al., 2014) fig.6).....	47
Figure 3.8 Évolution du résidu obtenu avec le modèle du tableau 2 en fonction du temps, détection d'un défaut de surchauffe d'un roulement 7 mois avant l'arrêt total (source (Wilkinson et al., 2014)fig.7) .....	47
Figure 3.9 Découpage en 9 zones (a) ; données réelles pour le couple vent vs puissance (source (Chen et al., 2013) fig.2).....	49
Figure 3.10 Base de données d'apprentissage de 6 défauts de pale, encadrées en violet des zones où les données sont trop peu nombreuses et où l'A Priori est nécessaire (source (Chen et al., 2013) fig.9) En vert les données de bon fonctionnement et en rouge les données représentant des défauts connus .....	49
Figure 3.11 Résidu des deux approches sur une période sans défaut (source (Zhang, 2014) fig.5) .....	51
Figure 3.12 Exemple de détection de défaut de roulement principal avec l'approche utilisant un réseau de neurones (source (Zhang, 2014) fig.6).....	52
Figure 3.13 Valeur de la température du roulement de la multiplicatrice lors de l'apparition d'un défaut (source (Garcia et al., 2006) fig.8).....	55
Figure 3.14 Exemple de règles pour SIMAP (source (Garcia et al., 2006)).....	56
Figure 3.15 Sélection des entrées et sorties des modèles (source (Schlechtingen et al., 2013) fig.6) .....	58

Figure 3.16 Organisation de la chaîne cinématique, localisation des capteurs (source (Butler, 2012) fig.7.3).....	62
Figure 3.17 Distribution de la température du roulement en fonction de la vitesse de rotation de l'arbre (source (Butler, 2012) fig 7.12).....	62
Figure 3.18 Utilisation du résidu filtré par basse vitesse de rotation pour la détection d'un défaut (source (Butler, 2012) fig.7.13).....	63
Figure 3.19 Lien de corrélation entre les températures du stator et du roulement génératrice (source (Schlechtingen and Santos, 2011) fig.14) .....	64
Figure 3.20 Prétraitement pour créer les modèles (source (Yang et al., 2013)fig.3).....	65
Figure 3.21 Projections des données suivant 2 composantes de la FDA, mise en évidence de familles de comportement (source (Kruger et al., 2013) fig.9) .....	67
Figure 3.22 Mise en évidence d'un défaut roulement génératrice (source (Astolfi et al., 2014)).....	71
Figure 3.23 Copie d'écran du logiciel de calcul de métrique (source(McLaughlin et al., 2009)) .....	73
Figure 3.24 Matrice de confusion obtenue sur un parc de 48 machines en utilisant des réseaux de neurones. (source (Papatheou et al., 2014)fig.2).....	75
Figure 4.1 Implantation du parc étudié de WT6 à WT11 .....	81
Figure 4.2 Répartition des occurrences de vitesse de vent mesurées sur le parc étudié pendant les 4 ans de l'étude .....	82
Figure 4.3 Position des capteurs utilisées.....	83
Figure 4.4 Évolution de la température du roulement avant et de la nacelle pendant 4 ans pour une éolienne du parc .....	84
Figure 4.5 Évolution de la température du roulement avant, arrière, de l'enroulement et de la nacelle pendant 1 mois pour une éolienne du parc.....	84
Figure 4.6 Histogramme de la vitesse du vent, de la puissance et de la vitesse de rotation de l'arbre rapide pour les éoliennes du parc pendant 4 ans .....	85
Figure 4.7 Photo de la défaillance du roulement arrière WT6 .....	91
Figure 4.8 Évolution de la température du roulement avant lors du défaut de roulement de l'éolienne WT9.....	92
Figure 4.9 Évolution de la température du roulement arrière lors du défaut d'enroulement de l'éolienne WT9.....	92
Figure 5.1 Schéma de principe de l'avance à la détection, avec un indicateur sur une plage de défaut .....	98
Figure 5.2 Schéma de principe d'une période de fausses alarmes avec un indicateur dans une période sans défaut .....	99
Figure 5.3 Schéma de principe de la courbe ROC adaptée .....	101
Figure 5.4 Impact de la persistance sur la courbe ROC .....	102
Figure 5.5 Schéma général des méthodes mono-turbines .....	103
Figure 5.6 Schéma de répartition des périodes d'apprentissages .....	107
Figure 5.7 Évolution de la température du roulement avant de l'éolienne 9 en défaut.....	110
Figure 5.8 Évolution de la température des roulements avant, des l'éolienne du parc en défaut, éolienne 9 en défaut. ....	111
Figure 5.9 Illustration des retards entre les turbines sur le parc .....	113
Figure 5.10 Périodicité annuelle et quotidienne des résidus mono-turbine basés sur la similarité roulement avant vs arrière.....	115
Figure 5.11 Schéma de principe de la comparaison multi-turbine utilisé en 2 dimensions .....	117
Figure 5.12 Schéma de la nouvelle méthode multi-turbine proposée, comparant les résidus provenant de modèle mono-turbine, entre les éoliennes .....	117

Figure 6.1 Schéma de principe de découpage en plages de bon fonctionnement et de défaut pour un défaut progressif.....	122
Figure 6.2 Schéma de principe de découpage en plages de bon fonctionnement et de défaut pour un défaut soudain .....	123
Figure 6.3 Température du roulement avant en fonction de la température du roulement arrière, coefficient de détermination $R^2 = 0.96$ .....	126
Figure 6.4 Température de l'enroulement stator en fonction de la température du roulement avant, coefficient de détermination $R^2 = 0.86$ .....	126
Figure 6.5 Température de l'enroulement stator en fonction de la température du roulement arrière, coefficient de détermination $R^2 = 0.91$ .....	127
Figure 6.6 Évolution de la température du roulement arrière de chaque éolienne .....	128
Figure 6.7 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant les températures brutes, approche mono-turbine .....	129
Figure 6.8 Avance à la détection vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche mono-turbine .....	130
Figure 6.9 Persistance vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche mono-turbine .....	130
Figure 6.10 Évolution de la température du roulement arrière de chaque éolienne, pendant le défaut de roulement arrière WT6.....	131
Figure 6.11 Évolution de la distance entre la température du roulement arrière de chaque éolienne et la référence parc, approche multi-turbine.....	132
Figure 6.12 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant les températures brutes, approche multi-turbine.....	132
Figure 6.13 Avance à la détection vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche multi-turbine .....	133
Figure 6.14 Persistance vs nombre d'interventions inutiles des défauts, en utilisant les températures brutes, approche multi-turbine.....	133
Figure 6.15 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases en fonctionnement du roulement arrière de chaque éolienne, approche mono-turbine ...	136
Figure 6.16 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent du roulement arrière de chaque éolienne, approche mono-turbine filtré .....	137
Figure 6.17 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine .....	138
Figure 6.18 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine .....	138
Figure 6.19 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine .....	139
Figure 6.20 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent avec plusieurs taille de fenêtre, approche mono-turbine.....	139
Figure 6.21 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine .....	140
Figure 6.22 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine .....	140

Figure 6.23 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine .....	141
Figure 6.24 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche mono-turbine.....	142
Figure 6.25 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche mono-turbine, pendant le défaut de roulement arrière WT6 .....	143
Figure 6.26 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine .....	144
Figure 6.27 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine.....	144
Figure 6.28 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine .....	145
Figure 6.29 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche multi-turbine .....	145
Figure 6.30 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine .....	146
Figure 6.31 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine .....	146
Figure 6.32 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine .....	147
Figure 6.33 Avance à la détection vs nombre d'interventions inutiles pour les défauts critiques, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine .....	147
Figure 6.34 Avance à la détection vs nombre d'interventions inutiles pour les défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine .....	148
Figure 6.35 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine .....	148
Figure 6.36 Évolution de l'indicateur similarité du roulement arrière de chaque éolienne, approche mono-turbine.....	149
Figure 6.37 Avance à la détection vs nombre d'interventions inutiles pour les défauts critiques, en utilisant un modèle de similarité, approche mono-turbine.....	150
Figure 6.38 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, approche mono-turbine .....	150
Figure 6.39 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, approche mono-turbine .....	151
Figure 6.40 Évolution de l'indicateur similarité avec plusieurs tailles de fenêtre $f$ sur la période de défaut du roulement avant WT9, approche mono-turbine.....	152
Figure 6.41 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle similarité, approche multi-turbine.....	153
Figure 6.42 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, approche multi-turbine .....	153
Figure 6.43 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, approche multi-turbine.....	154
Figure 6.44 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine .....	154
Figure 6.45 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine.....	155

---

Figure 6.46 Persistance vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine .....	155
Figure 6.47 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, comparaison globale .....	156
Figure 6.48 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, comparaison globale .....	157
Figure 6.49 Persistance vs nombre d'interventions inutiles des défauts, comparaison globale....	157

<b>REMERCIEMENTS .....</b>	<b>3</b>
<b>NOTATIONS.....</b>	<b>4</b>
ACRONYMES.....	4
UNITES .....	4
SYMBOLES .....	4
<b>TABLE DES FIGURES.....</b>	<b>5</b>
<b>TABLE DES MATIERES .....</b>	<b>10</b>
<b>1. INTRODUCTION.....</b>	<b>14</b>
<b>2. CONTEXTE ET OBJECTIFS .....</b>	<b>17</b>
2.1. LA PRODUCTION D'ENERGIE ELECTRIQUE EOLIENNE.....	17
2.1.1. Situation actuelle en France et en Europe .....	17
2.1.2. Description d'une éolienne .....	21
2.2. MAINTENANCE D'UN PARC EOLIEN.....	25
2.2.1. Organisation actuelle de la maintenance .....	26
2.2.2. Vers un développement de la maintenance conditionnelle .....	27
2.3. SURVEILLANCE DES EOLIENNES.....	28
2.3.1. Condition Monitoring System (CMS) .....	28
2.3.2. Supervisory Control and Data Acquisition (SCADA) .....	30
2.4. OBJECTIFS DE LA THESE.....	31
<b>3. ÉTAT DE L'ART SUR LA CREATION D'INDICATEURS DE DEFAUT POUR LES EOLIENNES EN UTILISANT DES DONNEES SCADA.....</b>	<b>34</b>
3.1. LES METHODES MONO-TURBINES.....	35
3.1.1. Principe général des méthodes mono-turbines.....	35
3.1.2. Méthodes utilisant les liens de causalité .....	36
3.1.3. Méthodes utilisant la similarité entre composants.....	60
3.1.4. Conclusion sur les méthodes mono-turbines .....	68
3.2. LES METHODES MULTI-TURBINES.....	69
3.2.1. Comparaison directe de courbes empiriques.....	70
3.2.2. Comparaison à des modèles de référence constructeur.....	72
3.2.3. Comparaison de résidus issus de modèles appris sur différentes turbines .....	74
3.2.4. Conclusion sur les méthodes multi-turbine .....	75
3.3. CONCLUSION GENERALE ET PERSPECTIVES DE RECHERCHE .....	76
<b>4. PRESENTATION DE LA BASE DE DONNEES VALEMO.....</b>	<b>80</b>
4.1. PRESENTATION DU PARC EOLIEN .....	80
4.2. COMPORTEMENT DES VARIABLES SCADA.....	83
4.3. LES DEFAUTS TYPES D'UNE EOLIENNE .....	86
4.3.1. Le rotor .....	86
4.3.2. La multiplicatrice .....	87
4.3.3. La génératrice.....	89

4.3.4.	Le convertisseur et le transformateur .....	89
4.3.5.	La station météorologique .....	90
4.4.	DEFAUTS PRESENTS DURANT LA PERIODE ETUDIEE.....	90
4.5.	CONCLUSION .....	94
<b>5.</b>	<b>METHODES PROPOSEES POUR LA CREATION ET L'EVALUATION</b>	
<b>D'INDICATEURS .....</b>	<b>96</b>	
5.1.	INTRODUCTION.....	96
5.2.	CRITERES PROPOSES POUR EVALUER LES PERFORMANCES .....	96
5.2.1.	L'avance à la détection .....	97
5.2.2.	Le nombre d'interventions inutiles .....	98
5.2.3.	La disponibilité .....	99
5.2.4.	La persistance .....	100
5.3.	CREATION DES SEUILS .....	100
5.3.1.	Interventions inutiles versus avance à la détection .....	100
5.3.2.	Nombre d'interventions inutiles fixe .....	102
5.4.	CREATION D'INDICATEURS .....	103
5.4.1.	Méthodes mono-turbines proposées .....	103
5.4.2.	Méthodes multi-turbines proposées .....	108
5.5.	CONCLUSION .....	118
<b>6.</b>	<b>ÉVALUATION DES PERFORMANCES DES METHODES PROPOSEES</b>	
<b>SUR LA BASE DE DONNEES VALEMO .....</b>	<b>121</b>	
6.1.	REGLAGES PRELIMINAIRES .....	121
6.1.1.	Définitions des plages de défauts et de bon fonctionnement.....	121
6.1.2.	Définition des paramètres <i>NA</i> , <i>NE<sub>min</sub></i> , <i>f</i> et <i>seuil production</i> .....	124
6.1.3.	Choix des températures utilisées comme indicateurs .....	125
6.2.	PERFORMANCES DES INDICATEURS UTILISANT DES VARIABLES SCADA BRUTES .....	128
6.2.1.	Utilisation des variables brutes, approche mono-turbine .....	128
6.2.2.	Utilisation des variables brutes, approche multi-turbine.....	131
6.3.	INDICATEURS UTILISANT LES RESIDUS ISSUS DE MODELES UTILISANT LES LIENS DE	
CAUSALITE .....	135	
6.3.1.	Résidus issus de modèles prenant en compte la dynamique de la température....	135
6.3.2.	Résidus issus de modèles statiques .....	142
6.3.3.	Comparaison globale des méthodes utilisant les liens de causalité mono-turbine et	
multi-turbine .....	147	
6.4.	INDICATEURS UTILISANT DES RESIDUS ISSUS DE MODELES UTILISANT LA SIMILARITE ....	149
6.4.1.	Approche mono-turbine .....	149
6.4.2.	Approche multi-turbine.....	152
6.4.3.	Comparaison globale des méthodes utilisant des résidus issus de modèles utilisant	
la similarité	154	
6.5.	ÉVALUATION GLOBALE DES DIFFERENTS INDICATEURS .....	156
6.5.1.	Analyse avec un seuil variable .....	156
6.5.2.	Comparaison des performances pour un taux d'interventions inutiles donné .....	159
6.6.	CONCLUSION .....	161
<b>7.</b>	<b>CONCLUSION GENERALE.....</b>	<b>164</b>



---

7.1. LES DIFFERENCES ENTRE METHODES UTILISANT LES LIENS DE CAUSALITE ET LA SIMILARITE.....	165
7.2. L'APPRENTISSAGE .....	166
7.3. LE SEUIL ET L'EVALUATION DES PERFORMANCES .....	167
7.4. INFLUENCE DE LA CHARGE SUR L'INDICATEUR .....	167
7.5. DIAGNOSTIC.....	168
7.6. GENERALISATION A D'AUTRES PARCS.....	168
7.7. UTILISATION DE DONNEES EXTERIEURES ET FUSION DES INDICATEURS .....	169
<b>8. BIBLIOGRAPHIE.....</b>	<b>172</b>



# 1. Introduction

Le secteur de l'énergie se transforme de manière rapide, en France et dans le monde, afin de faire face aux nouvelles problématiques liées au réchauffement climatique et aux engagements pris par les États pour réduire leurs émissions de gaz à effet de serre. Ces nouvelles contraintes sur la production d'énergie font apparaître les énergies renouvelables comme une solution efficace. Par conséquent, le développement de nouvelles centrales d'énergie renouvelables s'est multiplié, en particulier dans le domaine éolien où la puissance électrique installée a été multipliée par dix durant les dix dernières années.

Cette augmentation du nombre d'éoliennes et l'évolution du marché de l'énergie obligent à repenser les activités de maintenance et d'exploitation. Les centrales d'énergies renouvelables comme les éoliennes possèdent la particularité de produire de manière variable tout au long de l'année en suivant les variations de la source d'énergie. Afin de pouvoir garantir la disponibilité des installations, pour pouvoir avoir un réseau le plus stable possible, il est maintenant indispensable que les acteurs de l'éolien s'équipent d'outils avancés pour la surveillance de leurs installations. Ces outils doivent permettre d'informer sur l'état de santé des centrales et aider dans les décisions liées à la maintenance.

Un changement radical est nécessaire dans les stratégies de maintenance. Pour le moment, les entreprises chargées de la maintenance s'organisent en intervenant de manière périodique ou interviennent lorsqu'une panne apparaît. Mais afin de diminuer les pertes de production, pour garantir une meilleure disponibilité et une meilleure rentabilité des installations, ces stratégies de maintenance doivent être maintenant basées sur la surveillance de la dégradation des machines. En effet, et c'est l'un des objectifs de cette thèse, de nouvelles méthodes de surveillance doivent être développées afin de permettre de prévoir l'apparition de défaut pour permettre des ajustements ou des réparations au plus tôt.

Plusieurs pistes sont actuellement explorées pour permettre la création de ces outils de surveillance et il est possible de les classer en deux catégories. L'une consiste à rajouter des capteurs spécialisés dans la détection de tel ou tel phénomène lié à l'apparition d'un défaut, et l'autre consiste à extraire des informations sur l'apparition de défaut à partir des capteurs déjà installés dans les éoliennes pour le système de contrôle. Les travaux présentés par la suite explorent cette deuxième possibilité, qui a l'avantage d'être plus abordable pour les industriels.

La surveillance des éoliennes possède des spécificités propres par rapport à la surveillance d'autres installations. En effet, les éoliennes sont des centrales électriques qui produisent de façon variable en suivant les changements aléatoires du comportement du vent. Les éoliennes évoluent dans un environnement qui n'est pas contrôlé et qui a une influence sur le comportement des grandeurs mesurées pour la surveillance. En plus des variations de l'environnement, les variations incessantes de la production ont aussi un impact important sur ces mêmes grandeurs. Par conséquent, les grandeurs mesurées sur les éoliennes varient fortement, même lorsqu'il n'y a pas de défaut. Tout cela rend la tâche de détection des défauts en avance particulièrement difficile.

Cette thèse présente deux principales avancées. Tout d'abord, l'étude des méthodes de surveillance dans le domaine de l'éolien a mis en évidence un manque de méthodes de l'évaluation des performances. Cette thèse propose des critères d'évaluation qui permettent de comparer les méthodes proposées en fonction des contraintes des industriels. En effet, la plupart des méthodes proposées dans

la littérature se contentent de montrer leur capacité à réagir et détecter des défauts en avance. Cependant, les méthodes proposées n'évaluent pas leur comportement lors des périodes de fonctionnement normal. Il est pourtant crucial que les outils développés n'incitent pas les industriels à intervenir alors qu'il n'y a aucune raison, en générant trop de fausses alarmes.

Dans un deuxième temps, cette thèse propose une solution originale pour détecter les défauts suffisamment en avance tout en garantissant de minimiser les interventions inutiles. L'étude de méthodes existantes présentées dans la littérature a permis de mettre en évidence l'existence de deux approches distinctes, dénommées dans cette thèse l'approche mono-turbine et l'approche multi-turbine. Ces deux approches possèdent des avantages propres et complémentaires. En effet, l'approche mono-turbine qui consiste à utiliser les capteurs venant d'une seule turbine pour détecter les défauts permet de s'affranchir des variations des conditions de fonctionnement, et l'approche multi-turbine qui consiste à comparer les éoliennes entre elles permet de s'affranchir des variations des conditions environnementales. Cette thèse propose une méthode originale permettant de profiter des avantages des deux approches en les combinant. L'objectif est ainsi de diminuer l'impact de l'environnement et du fonctionnement machine, et ainsi permettre de faire plus facilement la différence entre un défaut et le comportement normal de l'éolienne. Au final, l'objectif est de diminuer au maximum le nombre d'interventions inutiles tout en permettant de détecter un maximum de défaut et le plus tôt possible.

Ces travaux proposent d'évaluer et de comparer la nouvelle solution proposée avec les autres solutions proposées dans la littérature. Pour cela, une base de données a été construite à partir de données venant d'un parc éolien en activité collectées pendant 4 ans. Durant cette période, plusieurs défauts majeurs ont eu lieu et ont engendré des pertes de productions significatives. Utiliser cette base de données permet de se mettre dans des conditions réelles d'utilisation et voir l'apport de ces méthodes pour un industriel.

Dans une première partie, nous présentons le contexte industriel et les enjeux de cette thèse. Dans cette partie, le comportement des éoliennes et les problématiques de maintenance sont décrites.

Ensuite, une étude bibliographique des derniers ouvrages parus entre 2001 et 2014 est réalisée. La problématique de surveillance des éoliennes date de plusieurs dizaines d'années, mais ce n'est que récemment que le nombre de propositions venant de la communauté scientifique a augmenté. Cette partie détaille les deux approches mono-turbines et multi-turbines et montre leurs avantages et inconvénients respectifs.

Puis, la base de données utilisée pour évaluer et comparer les méthodes est présentée. Cette base de données a été construite durant ces travaux avec des données provenant d'un parc éolien récent de 6 éoliennes d'une puissance de 2 MW.

Dans la partie suivante, les nouvelles méthodes proposées seront présentées. Tout d'abord, nous détaillerons les critères d'évaluation et leur intérêt pour mesurer les performances des méthodes. Puis nous détaillerons les méthodes que nous proposons pour réaliser la surveillance des éoliennes. Une adaptation des méthodes mono-turbines existantes est présentée puis une méthode originale qui combine les approches mono-turbines en une approche multi-turbine est proposée. Toutes ces méthodes et approches sont comparées et évaluées dans la dernière partie.



## 2. Contexte et objectifs

### 2.1. La production d'énergie électrique éolienne

#### 2.1.1. Situation actuelle en France et en Europe

La production d'énergie éolienne en France et en Europe a évolué significativement depuis les 20 dernières années. Ces changements sont générés par les progrès technologiques, par l'évolution du marché de l'énergie et surtout par les modifications dans la régulation de l'énergie faites par les différents gouvernements européens. En effet, les recherches scientifiques ont permis de mettre en évidence et d'apporter des preuves des modifications qui s'opèrent dans le climat planétaire. Les émissions de gaz à effet de serre sont maintenant mises en cause directement dans l'augmentation des températures mondiales comme le montrent (Groupe d'experts intergouvernemental sur l'évolution du climat et al., 2015) et (Edenhofer and IPCC, 2012). La prise de conscience des populations a conduit les gouvernements à favoriser les énergies dites « vertes » ou renouvelables par rapport aux énergies fossiles responsables de la majorité des émissions de gaz à effet de serre.

De nouvelles technologies permettant d'optimiser la production de centrales d'énergie renouvelable (ENR) ont été développées. Les principales sources d'énergie renouvelable sont regroupées suivant 4 catégories de technologie : l'hydraulique, l'éolien, le photovoltaïque, les biomasses et les nouvelles technologies. L'énergie hydraulique provient de barrages placés le long des cours d'eau, ces barrages permettent de récupérer l'énergie potentielle de l'eau entre deux points à différentes altitudes. L'énergie hydraulique provient en majeure partie de quelques installations massives, mais est aussi générée par un grand nombre d'installations modestes le long des cours d'eau, positionnée à la place des anciens moulins à eau. L'énergie photovoltaïque est l'énergie venant du rayonnement solaire. En France, les panneaux photovoltaïques sont les plus utilisés pour capter cette énergie, mais d'autres technologies émergent dans le monde comme les systèmes solaires à concentration. Les nouvelles technologies sont nombreuses et variées. Citons par exemple : les centrales hydroliennes marémotrices ou osmotiques qui utilisent les différences de salinité de l'eau pour générer de l'électricité. Elles représentent des centrales qui, en 2016, ne produisent pas significativement par rapport aux autres centrales ENR.

L'énergie éolienne représente l'énergie extraite du vent. Cette énergie est captée par des éoliennes et transformée en énergie électrique. Durant les 40 dernières années, comme le montre la Figure 2.1, la taille des éoliennes a été multipliée par 7 permettant maintenant à une éolienne d'avoir une puissance nominale supérieure à 2MW pour l'éolien terrestre et plus de 5MW pour l'éolien en mer. Grâce à ces évolutions technologiques, les coûts de production de l'énergie éolienne sont devenus compétitifs par rapport aux sources d'énergie fossiles ce qui permet d'envisager désormais un déploiement massif des énergies renouvelables et envisager une part prépondérante dans le mix énergétique.

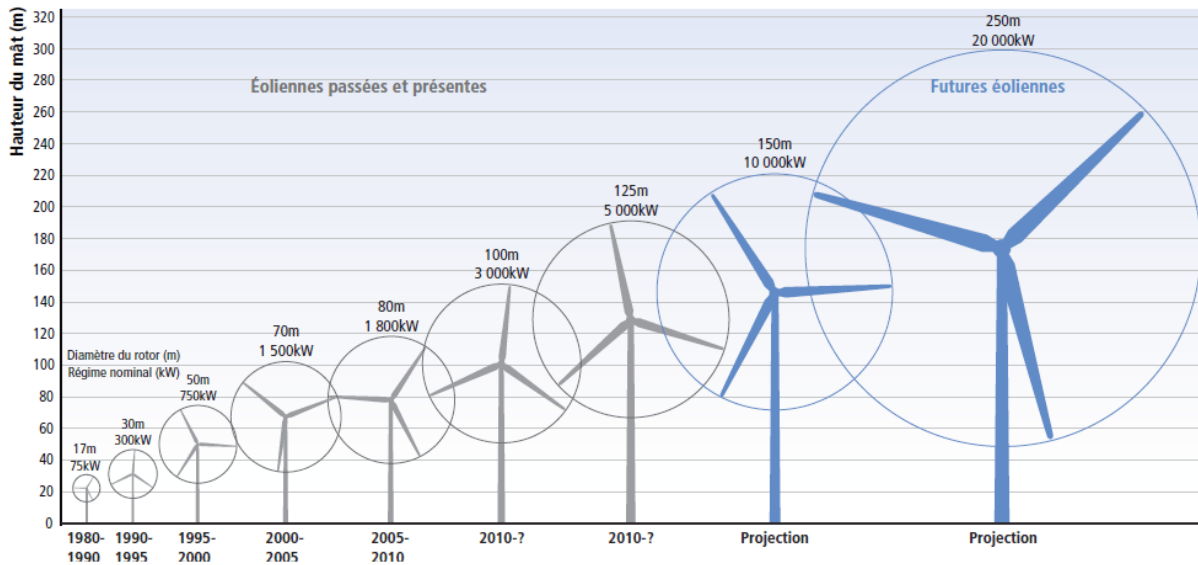


Figure 2.1 Évolution de la taille des éoliennes (Edenhofer and IPCC, 2012)

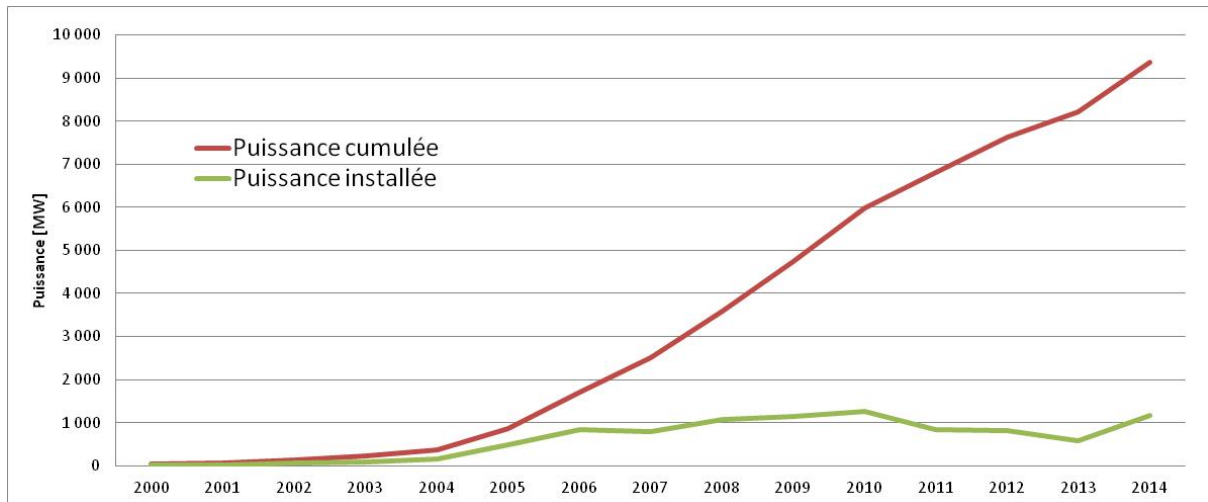







Figure 2.2 Évolution de la puissance éolienne en France d'après les chiffres du site du service de l'Observation et des Statistiques (SOeS) du ministère de l'Environnement français, de l'Énergie et de la Mer de 2015

La prise de conscience du public et des gouvernements successifs lié à baisse des coûts de production de l'énergie éolienne et le développement de nouvelles technologies ont permis à l'éolien de se développer très fortement en France. La puissance nominale installée est passée de 1GW à 10GW entre 2005 et 2015 comme le montre la Figure 2.2. Cette forte évolution de l'éolien en France est cependant à mettre en contraste avec les autres pays européens. Comme le montre le Tableau 2.1, la puissance installée en France est très inférieure à celle de ses partenaires européens. La principale raison provient de la forte croissance du secteur nucléaire français. La France est l'un des leaders mondiaux en matière d'énergie nucléaire et possède le plus grand nombre de centrales nucléaires par habitant, et est le deuxième producteur d'énergie nucléaire mondial juste après les États unis. Pour cette raison, l'énergie éolienne n'a pas été tout de suite nécessaire pour limiter les émissions de gaz à effet de serre. Cependant, la production d'énergie nucléaire est liée à 2 des plus grandes catastrophes écologiques de ces 50 dernières années, Tchernobyl et Fukushima. Durant ces événements, les réacteurs ont fui et déversé une partie de leur contenu radioactif dans l'environnement, tuant ou blessant

une partie de la population avoisinante et détruisant la faune et la flore locale. À cause de ces catastrophes, l'opinion publique mondiale s'est éloignée peu à peu de l'énergie nucléaire. Des pays comme l'Allemagne ont fait le choix de sortir complètement du nucléaire, ce qui explique pourquoi l'énergie éolienne y est la plus développée en Europe.

Tableau 2.1 Puissance éolienne installée dans l'Union Européenne (MW) (European Wind Map, 2008),(Wind in power, European statistics 2011, 2012)(EUROBSERVER, 2016)

Rang (fin 2014)	Pays	fin 2008	fin 2009	fin 2010	fin 2011	fin 2012	fin 2013	fin 2014
1	 Allemagne	23 903	25 777	27 214	29 060	31 308	34 660	40 456
2	 Espagne	16 740	19 160	20 676	21 674	22 796	22 959	22 987
3	 Royaume-Uni	3 241	4 245	5 204	6 540	8 649	11 209	12 475
4	 France	3 404	4 574	5 660	6 800	7 623	8 243	9 285
5	 Italie	3 736	4 849	5 797	6 747	8 144	8 557	8 662

En France, le gouvernement a décidé de réduire à 50% la part de l'électricité produite par le nucléaire. En 2014, la part du nucléaire en France était de 77% (cf Tableau 2.2). Bien que l'hydraulique soit la deuxième source d'énergie en France, les possibilités d'augmentation de production sont limitées. L'énergie photovoltaïque reste encore peu rentable par rapport à l'éolien qui représente la solution la plus probable pour arriver à cet objectif tout en continuant à produire autant.

Tableau 2.2 Part des sources d'énergie dans la production électrique en France en 2014 (RTE, 2014)

Énergie produite	TWh	Variation 2014/2013	Part de la production
Production nette	540,6	-1,8%	100%
Nucléaire	415,9	3,0%	77%
Thermique à combustible	27,0	-39,6%	5,0%
Hydraulique	68,2	-9,7%	12,6%
Éolien	17,0	+6,7%	3,1%
Photovoltaïque	5,9	+27,2%	1,1%
Autres Sources d'énergies	6,6	+6,7	1,2%

Afin d'atteindre l'objectif de 50% de l'électricité produite par le nucléaire, le nombre d'éoliennes et la puissance éolienne installée vont devoir augmenter énormément dans les prochaines années. Le Tableau 2.3 montre la puissance installée pour chaque source d'énergie pour l'année 2014 en France. Afin de bien comprendre ce tableau, il faut prendre en compte ce que l'on appelle le facteur de charge. Le facteur de charge représente le ratio entre la puissance nominale indiquée de la centrale et la puissance fournie en moyenne. Les facteurs de charge des ENR sont nettement moins élevés que les centrales à énergie fossile ou nucléaire (cf Tableau 2.4). Cette différence s'explique par le fait que les centrales ENR produisent de l'électricité en fonction de la ressource naturelle associée. Pour l'éolien, une turbine éolienne ne peut pas produire plus que ce que l'énergie cinétique du vent lui permet. Cette énergie dépend de la vitesse du vent, qui varie fortement au cours du temps. Les éoliennes sont



dimensionnées pour fonctionner sur la plus large plage de vitesses de vent possibles, mais par conséquent ne fonctionnent pas en permanence à leur puissance nominale. Pour simplifier grandement, on peut dire que 1GW d'énergie nucléaire installé va produire trois fois plus que 1GW d'éolien installé. Le facteur de charge des énergies prend aussi en compte la disponibilité de celle-ci, et donc les arrêts pour maintenance, c'est pour cela que le facteur de charge des énergies non renouvelables n'est pas égal à 100%. Il est toutefois important de ne pas tomber dans l'exagération, une éolienne n'est pas 24% du temps en marche et 76% du temps à l'arrêt. Les éoliennes fonctionnent en permanence, mais ne produisent pas tout le temps à leur puissance maximale.

Tableau 2.3 Puissance installée au 31/12/2014 ((RTE, 2014))

Énergie produite	TWh	Variation 2014/2013 [%]	Variation 2014/2013 [MW]	Part de la production
Total	128943	+0,5%	+662	100%
Nucléaire	63130	+0,0%	+0	48,9%
Thermique à combustible	24411	-5,0%	-1296	18,9%
Hydraulique	25411	-0,1%	-23	19,7%
Éolien	9120	+11,8%	+963	7,2%
Photovoltaïque	5292	+21,2%	+926	4,1%
Autres Sources d'énergies	1579	+6,2%	+92	1,2%

Tableau 2.4 Facteur de charge annuel type (Statistique Canada, 2009) (RTE, 2014)

Type d'énergie	Valeur du facteur de charge
Éolienne	24%
Nucléaire	75%
Solaire photovoltaïque	13%
Hydroélectrique	28%
Centrale thermique	82%
Cycle combiné	43%

En prenant en compte le facteur de charge, on se rend compte de l'importance de la puissance éolienne à installer en France pour atteindre les objectifs du gouvernement. La croissance du secteur éolien est donc garantie pour la première fois pour plusieurs années puisque l'objectif annoncé des industriels est de passer de 10GW de puissance éolienne installée en 2015 à 20GW en 2020.

Cette croissance forte ne pourra se faire uniquement sur la terre. Bien que les parcs éoliens les plus vieux puissent être améliorés et profiter de technologies produisant plus sur la même surface au sol, le nombre de parcs possible à développer sur le sol ne suffira pas pour produire autant que désiré. De plus, l'augmentation du nombre d'éoliennes sur terre commence à engendrer des réticences vis-à-vis de la population. L'éolien en mer apparait ici comme la solution. Le potentiel d'éolien en mer est estimé en 2014 par le syndicat FEE (France énergie éolienne) à 200GW avec les technologies actuelles. Ce potentiel pourra largement permettre à la France de diminuer la part de l'énergie nucléaire à 50%. Le gouvernement français a déjà soutenu la construction de 6GW d'ici à 2020 et soutiendra la construction

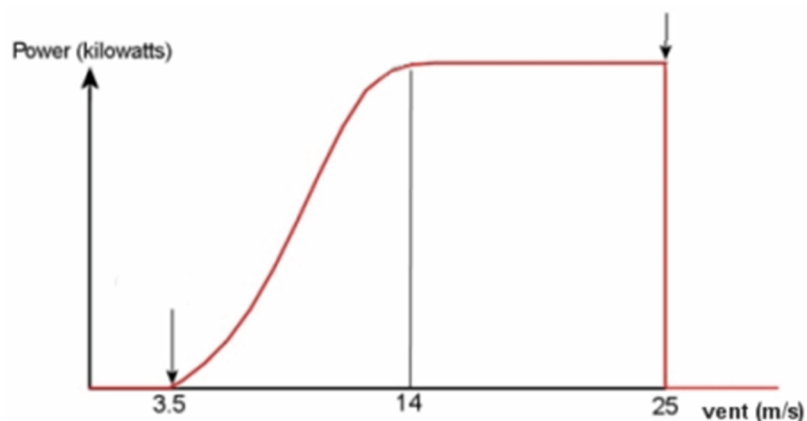
de 20GW en mer à 2030. La production d'énergie éolienne va donc quadrupler dans les quinze prochaines années.

Selon ADEME, la France pourrait passer à une production 100% ENR, en gardant un prix de l'électricité similaire. La principale condition à de tels changements, 50% nucléaire ou 100% ENR, réside dans la mutation du réseau électrique actuel. Pour le moment, la majorité de l'électricité est produite par le nucléaire, il s'agit d'une production très centralisée et complètement contrôlée. Les centrales ENR sont au contraire réparties sur tout le territoire. Ces centrales produisent selon la disponibilité de la ressource. Il faudra donc transformer le réseau pour s'adapter à cette énergie intermittente, grâce à des systèmes de stockage et en combinant des ENR complémentaires comme le photovoltaïque et l'éolien. Avec la mutation physique du réseau vient aussi la libéralisation du marché de l'énergie. Ces deux éléments vont obliger les producteurs d'énergie à investir dans la maintenance de leurs centrales, afin de garantir une disponibilité maximum lors de pics de consommation.

En conclusion, on peut voir que l'énergie éolienne en France et en Europe possède une forte croissance qui va persister durant au moins 15 ans, voir plus. Cette augmentation du nombre des éoliennes et la mutation du réseau vont avoir un impact important sur le métier d'exploitation et de maintenance des parcs éoliens, qui va devoir s'adapter en développant de nouvelles techniques de conduite.

### 2.1.2. Description d'une éolienne

Les éoliennes permettent de transformer l'énergie cinétique du vent en énergie mécanique puis électrique pour l'injecter dans les réseaux électriques. La vitesse du vent est la variable la plus importante pour les éoliennes. Une éolienne est souvent caractérisée par la relation qui lie cette vitesse et la puissance qu'elle est capable de délivrer. Cette relation est représentée dans la Figure 2.3, appelée courbe de puissance. Cette courbe montre que les éoliennes sont construites pour fonctionner à plusieurs vitesses de vent. Une vitesse minimale est nécessaire pour produire de l'énergie. Une fois cette valeur atteinte, la production d'énergie augmente avec la vitesse de vent, jusqu'à atteindre une puissance maximale qui sera maintenue si la vitesse du vent reste dans une certaine gamme. Si le vent devient trop fort, la machine est arrêtée afin d'éviter qu'elle s'endommage.



**Courbe typique de la puissance délivrée par une éolienne par vent stable**

*Figure 2.3 Profil de courbe courant liant la vitesse du vent à la puissance générée d'une éolienne*

La taille des pales et leurs nombres ont évolué depuis la première éolienne (voir Figure 2.1), mais le modèle HAWT (horizontal axis wind turbine) à 3 pales représente la majeure partie des éoliennes terrestres et marines construites et installées. Les composants d'éoliennes à axe vertical sont ceux décrits dans la Figure 2.4 dans la majorité des cas. Les pales qui tournent sur elles-mêmes permettent d'ajuster la prise au vent et l'énergie captée par l'éolienne. Ces pales entraînent le rotor (ou hub) en rotation qui va à son tour entraîner l'arbre principal en rotation. L'arbre principal tourne à des vitesses lentes de l'ordre de 10 à 20 tr/min. Ces vitesses de rotation nécessitent des génératrices particulières pour produire de l'électricité. Dans le cas d'éoliennes à entraînement direct, ces génératrices sont directement liées à l'arbre lent. Ces machines sont de plus en plus répandues notamment dans l'éolien en mer, mais en règle générale, la vitesse de rotation de l'arbre principal ou lent est adaptée aux vitesses de fonctionnement des génératrices classiques, par l'intermédiaire de la boîte à engrenages appelée multiplicatrice. Les multiplicatrices classiques augmentent la vitesse de rotation d'un facteur proche de 100. La génératrice va produire l'électricité triphasée qui est ensuite envoyée dans le transformateur pour l'adapter au réseau inter-éolien puis au réseau électrique.

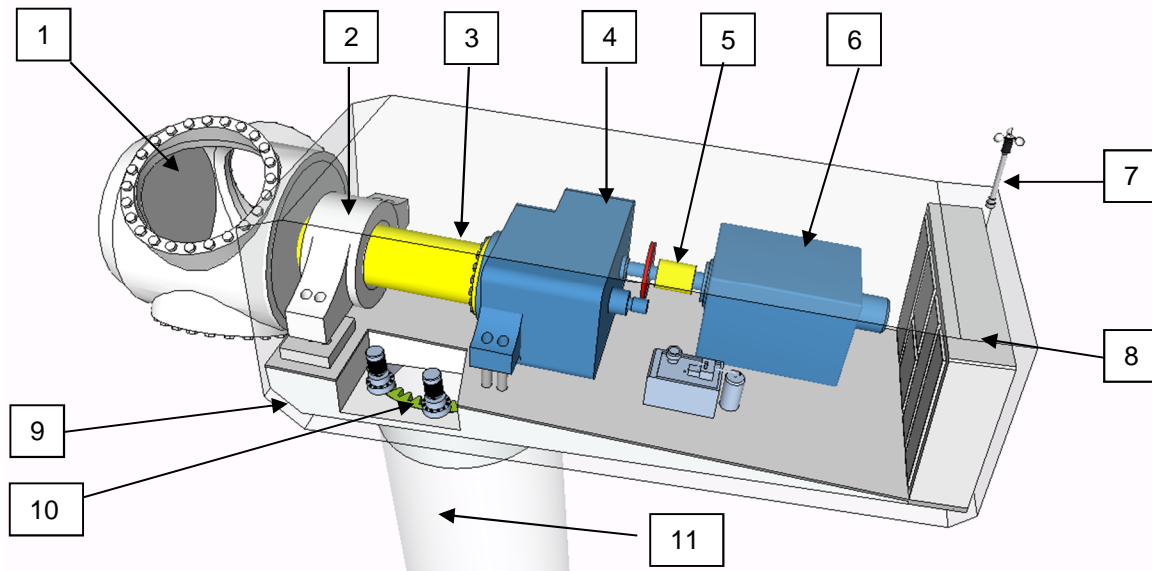


Figure 2.4 Description de l'intérieur d'une nacelle éolienne HAWT

1. Rotor ou hub
2. Roulement principal
3. Arbre principal ou arbre lent
4. Boîte à engrenages ou multiplicatrice
5. Arbre rapide
6. Génératrice
7. Station météorologique
8. Transformateur
9. Nacelle
10. Roulement nacelle
11. Mât

La nacelle comprend la majeure partie de ces composants et est située en haut du mat qui peut faire plus de 100 mètres de haut. Cette nacelle peut pivoter sur elle-même pour suivre la direction du vent. La direction, la vitesse du vent et la température extérieure sont mesurées dans la station

météorologique à l'arrière de la nacelle. Contrairement à ce que l'on pourrait croire, les éoliennes ne sont pas asservies par la mesure de la vitesse de vent. L'anémomètre sert uniquement à lancer le démarrage de la machine ou à l'arrêter en cas de vitesse de vent trop importante. Les angles de pales sont asservies grâce à la puissance et à la vitesse de rotation, l'éolienne cherche à produire au maximum.

La Figure 2.5 décrit les modes de fonctionnements de la machine en fonction du vent. Ces modes correspondent aux principales phases de fonctionnement dites « normales » communes aux turbines 2MW HAWT. Ces phases sont décrites ci-dessous :

- Phase 1. Dans cette phase, la machine est en attente d'une vitesse de vent suffisante pour démarrer. Les pales de l'éolienne ont une prise au vent maximale, et le rotor tourne de manière libre. Afin d'éviter d'actionner les moteurs des pales ou les systèmes de commande à cause d'une brève rafale, l'automate attend de mesurer une vitesse suffisante pendant quelques instants. La vitesse de vent nécessaire est comprise entre 3 et 4 m/s.
- Phase 2. Une fois l'éolienne activée, l'automate laisse la vitesse de rotation de l'arbre rapide atteindre une vitesse minimum. Une fois cette vitesse atteinte, la génératrice est couplée au réseau et commence à produire.
- Phase 3. La puissance produite et la vitesse de rotation augmentent en fonction de la vitesse du vent. La vitesse de rotation atteint sa valeur maximale avant la puissance. Dans cette phase, l'angle des pales est constant.
- Phase 4. La vitesse de rotation reste à valeur maximale tandis que la puissance continue d'augmenter. Afin de réguler correctement la puissance produite, l'angle des pales (et donc la prise au vent) est adapté pour ne pas dépasser les valeurs maximales de la machine.
- Phase 5. La puissance et la vitesse de rotation sont à leurs valeurs maximales, c'est uniquement l'angle des pales qui varie.
- Phase 6. La vitesse de vent maximale autorisée est atteinte. La machine entame donc un arrêt de sa prise au vent en continuant d'orienter les pales. Cela permet d'arrêter complètement la machine pour la protéger durant des périodes de fort vent.

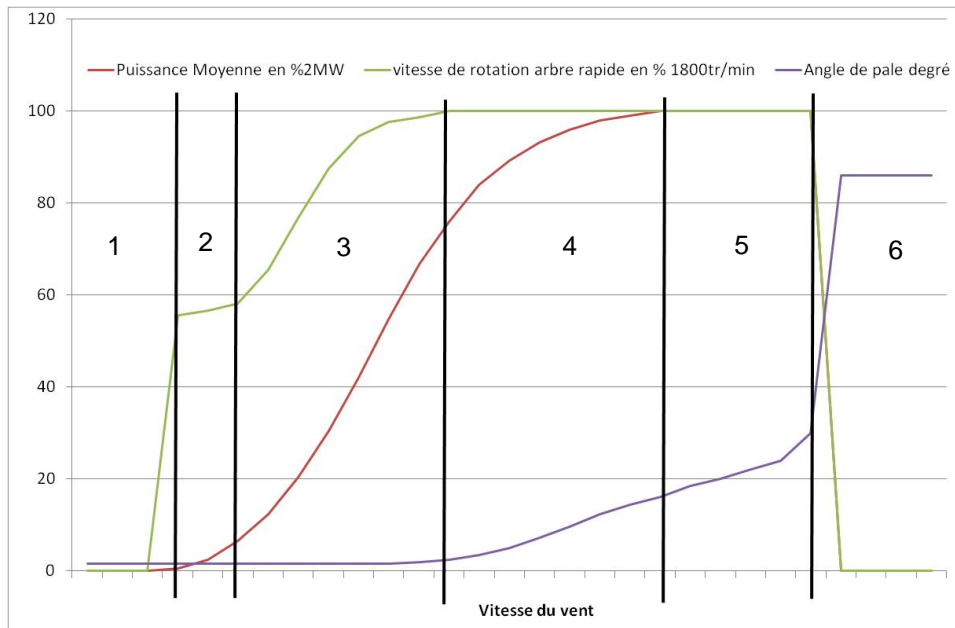


Figure 2.5 Modes de fonctionnement : Puissance moyenne, vitesse de rotation et angle de pale en fonction de la vitesse du vent. La puissance moyenne et la vitesse de rotation sont exprimées en pourcentage de leur valeur maximale

Il est important de noter que ces phases ne représentent pas tous les modes de fonctionnement de la machine. En effet, il existe d'autres opérations mineures qui vont se réaliser durant ces phases.

Les éoliennes sont conçues pour fonctionner sur une plage de vitesses de vent et avec une puissance maximale. Cette valeur de puissance est la valeur nominale de la génératrice, mais elle est rarement atteinte. Le fonctionnement de l'éolienne dépend de la vitesse du vent. Cette vitesse de vent va constamment varier comme le montre la Figure 2.6, qui représente l'évolution de la vitesse du vent et de la production énergétique pendant une durée de 9 jours. Cette figure permet de constater que le vent va varier sur toute la gamme de vitesses, pendant des périodes qui sont courtes au regard de la durée des opérations d'exploitation et de maintenance. Le vent peut passer d'une vitesse permettant d'atteindre la puissance nominale à un arrêt complet en une journée, par exemple. Une éolienne n'évolue donc presque jamais dans des conditions de fonctionnement stationnaires, mais produit la majorité de son énergie en alternant entre les phases 2, 3 et 4. La non-stationnarité des éoliennes est l'une des plus grandes difficultés de la surveillance de ces machines puisque les variables qui sont mesurées vont être influencées par ces variations de modes de fonctionnement.

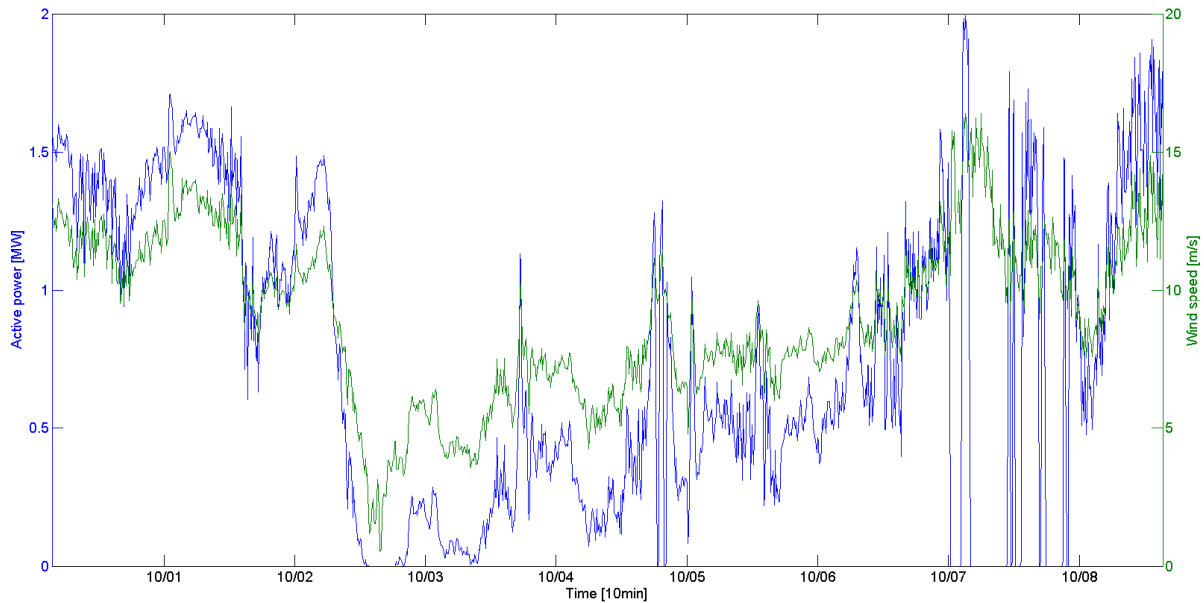


Figure 2.6 Évolution Vitesse du vent (m/s) et puissance active (MW) mesurées toutes les 10 minutes pendant 9 jours (source VALEMO)

De plus, comme les éoliennes évoluent en extérieur dans un environnement non contrôlé, les variables de surveillance sont soumises à d'autres influences que la puissance produite. Les conditions environnementales interviennent elles aussi et peuvent être difficiles à évaluer. La notion de vent ne peut pas être définie simplement par la seule mesure de vitesse anémométrique réalisée sur les éoliennes. La densité de l'air et la nature turbulente ou laminaire sont d'autres caractéristiques des masses d'air qui peuvent avoir une influence sur le comportement des éoliennes. Ces caractéristiques sont difficilement mesurables mais peuvent faire changer le comportement de l'éolienne. Par exemple, si le vent est turbulent, des tourbillons vont se former derrière les pales, ce qui peut créer des vibrations et va modifier les efforts mécaniques sur les pales et sur tous les composants mécaniques.

Pour conclure sur cette partie, il est important de garder en mémoire que les éoliennes évoluent dans un environnement fortement non stationnaire. Cette particularité rend la surveillance de ces machines difficile. En effet, les variables utilisables pour cette tâche sont influencées par les conditions de fonctionnement et les conditions environnementales.

## 2.2. Maintenance d'un parc éolien

Le monde de l'énergie doit relever les défis posés par le passage d'énergies fossiles et nucléaires à des énergies renouvelables. La production d'énergie devient moins stable et centralisée qu'elle ne l'était. L'objectif principal des industriels est de lisser au maximum les fluctuations sur le réseau électrique afin de faire coïncider l'offre et la demande à chaque instant. Si le réseau perdait son équilibre, cela pourrait entraîner des coupures d'électricité généralisées dans toute la France et l'Europe. Les producteurs d'électricité doivent donc garantir de plus en plus une production stable, lissée et contrôlée. La maintenance doit donc changer ses modèles d'organisation et s'adapter à cette nouvelle façon de

produire l'électricité afin de garantir une plus grande disponibilité des installations. On estime aujourd'hui qu'une turbine aura entre 4 et 5 défaillances non-prévues par an (Takoutsing et al., 2014). Ce chiffre peut paraître faible mais il faut le mettre en relief avec le nombre important de machines à gérer, un parc éolien pouvant aller de 5 à 30 machines pour l'éolien terrestre et jusqu'à 100 machines pour l'éolien en mer.

La maintenance éolienne est pour le moment principalement organisée en maintenance corrective et en maintenance préventive systématique. Afin de pouvoir garantir des disponibilités plus importantes et aussi diminuer les coûts, la maintenance doit évoluer vers la maintenance conditionnelle et prévisionnelle.

### **2.2.1. Organisation actuelle de la maintenance**

L'industrie éolienne utilise plusieurs techniques de maintenance simultanément afin d'optimiser les coûts en profitant des avantages de ces différentes techniques. L'organisation de la maintenance éolienne est corrective et préventive (Tavner, 2012).

La maintenance corrective éolienne consiste à intervenir lorsque l'un des composants de l'éolienne est défaillant et l'empêche de produire. Ce type de maintenance permet d'avoir des coûts de mise en place réduits, mais engendre des coûts d'indisponibilité élevés. En effet, l'implémentation d'une telle stratégie corrective ne coûte rien puisqu'il faut simplement attendre la défaillance. En revanche, le montant de l'arrêt de production peut être très élevé. L'éolienne tombera plus fréquemment en panne dans des phases de production d'électricité, c'est-à-dire quand les conditions météorologiques sont favorables, ce qui va engendrer des pertes de productions importantes. En France, le prix moyen de l'électricité pour des parcs éoliens avec un contrat d'obligation d'achat est de l'ordre de 0.80 €/kWh. L'heure d'arrêt pour une éolienne de 2MW est donc d'environ 1600€. Ce chiffre peut paraître faible, mais une défaillance va souvent nécessiter une intervention sur place. Ainsi, si un composant facilement remplaçable tombe en panne, l'arrêt peut d'être rendu plus long à cause du temps de déplacement des équipes. Dans le cas de remplacement de composants importants, comme la génératrice, des travaux supplémentaires d'aménagement peuvent engendrer des surcoûts élevés. Finalement, le coût de remplacement des composants est aussi à considérer. Par exemple, si la multiplicatrice est défaillante, il faut en racheter une neuve (environ 200 000€), la remplacer en déployant des équipes techniques et en installant des engins de levage (environ 100 000€ pour l'éolien sur terre). Tout cela peut prendre d'une semaine à plusieurs mois en fonction de la disponibilité des équipes, des pièces, du matériel et de l'accessibilité du site (les pertes de production vont dépendre de la vitesse du vent et de la durée de l'arrêt).

La maintenance préventive éolienne est d'abord une maintenance systématique. Les interventions sont planifiées pour vérifier l'état de l'éolienne avant l'apparition de défaillances. Cette stratégie a un coût de mise en place qui peut être difficile à quantifier. En effet, la planification de cette maintenance prend parfois du temps, car elle doit être optimisée. Puisque les interventions sont prévues, elles seront souvent regroupées avec d'autres afin d'optimiser les coûts de déplacements. De plus, le coût d'arrêt de production peut lui aussi être optimisé. Contrairement à la maintenance corrective, la période d'intervention peut être choisie lorsque la vitesse de vent est minimale. Par conséquent, les coûts d'arrêts de production sont le plus souvent négligeables. En revanche, le vent n'est pour le moment prédit que sur des horizons courts de l'ordre de la semaine. L'organisation de la maintenance préventive peut donc changer d'une semaine à l'autre, voire d'un jour à l'autre. Il n'est pas rare qu'une intervention soit annulée le jour même alors que le vent s'est soudainement levé. Les coûts de mise en place de

cette stratégie vont donc varier fortement et sont parfois impossibles à calculer sans des outils de modélisation et d'évaluation adaptés. Cette stratégie de maintenance préventive permet aussi de limiter les dommages causés sur certains composants très coûteux. En remplaçant certaines pièces ou en réalisant des opérations de graissage, par exemple, il est possible d'empêcher la casse de composants plus importants et plus coûteux.

Il est important de rappeler que ce type de stratégie de maintenance ne permet pas d'empêcher la totalité des défaillances. Une maintenance corrective résiduelle sera toujours nécessaire en parallèle.

### **2.2.2. Vers un développement de la maintenance conditionnelle**

Une autre solution pour diminuer les coûts des défaillances est la maintenance conditionnelle. La définition de la maintenance conditionnelle donnée par la norme NF EN 13306 X 60-319 (AFNOR, 2001) est la suivante : « 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 surveillance du fonctionnement et des paramètres peut être exécutée selon un calendrier, ou à la demande, ou de façon continue dans l'éolien. Ces techniques en ligne ou ponctuelles seront détaillées dans la partie suivante 2.3.1.

La maintenance conditionnelle est une maintenance préventive, mais non systématique. Les interventions ne sont planifiées que lorsqu'un comportement anormal est détecté par la surveillance. La détection de ces anomalies dans des stades précoces doit pouvoir permettre de prévoir les défaillances et d'adapter la maintenance. Cela pourrait permettre de diminuer le coût d'intervention de la maintenance préventive si la surveillance du fonctionnement est efficace. En effet, certaines interventions d'inspection pourraient être supprimées si la surveillance permet d'avoir une idée précise de la santé de l'éolienne. La maintenance conditionnelle permettrait donc, en théorie, de supprimer toutes les interventions inutiles et ainsi de diminuer fortement les coûts de maintenance. D'autre part, puisque les défaillances peuvent être prévues, les pertes de production dues au remplacement peuvent aussi être diminuées fortement.

En réalité, dans l'éolien comme dans d'autres industries, la maintenance conditionnelle est plutôt utilisée en complément des autres stratégies de maintenance. Toutes les défaillances ne peuvent pas être détectées grâce à la surveillance, et ce principalement à cause d'un manque d'informations disponibles. Cela peut être un manque d'exhaustivité dans les données utilisables ou encore un manque de connaissance sur la machine. La généralisation de ces méthodes est récente dans l'industrie et par conséquent, la confiance que l'on y accorde est encore faible. En revanche, certaines opérations de maintenance conditionnelle ponctuelles sont habituellement mutualisées avec des opérations de maintenance systématique afin de limiter les coûts. La maintenance conditionnelle permet aussi de surveiller des composants qui ne sont pas toujours inspectés dans les maintenances systématiques. Les maintenances systématiques sont faites tous les 6 ou 12 mois et ces maintenances peuvent être complètes ou partielles. Entre deux maintenances, une défaillance peut apparaître et la surveillance des installations peut permettre de détecter la formation de ces défaillances si elle est faite de manière continue.

Pour surveiller les éoliennes, il faut pouvoir créer des indicateurs de dégradation. Ces indicateurs sont créés à partir de données quantitatives ou qualitatives. La partie qui suit présente différentes techniques pour obtenir et traiter des données quantitatives pertinentes pour la surveillance et la maintenance conditionnelle.



## 2.3. Surveillance des éoliennes

Les éoliennes sont des machines qui fonctionnent de manière indépendante et souvent dans des régions éloignées et difficiles d'accès. Ces conditions particulières font de la maintenance une source de coût important que l'on cherche à diminuer. La partie 2.2.2 montre que l'une des solutions pour diminuer ces coûts est de passer d'une maintenance systématique et corrective à une maintenance conditionnelle.

La mise en place d'une stratégie de maintenance conditionnelle nécessite des indicateurs de dégradation, rendant compte de l'état de la machine à maintenir. Ces indicateurs peuvent être quantitatifs ou qualitatifs. Ils peuvent venir de signaux mesurés et traités ou de rapports de maintenance. Le but de ces indicateurs est de refléter l'état de santé de la machine, ils sont construits pour évoluer avec la dégradation. Lorsque ces indicateurs atteignent une certaine valeur seuil, une action de maintenance conditionnelle est déclenchée.

Grâce à l'interprétation de ces indicateurs, il est possible d'effectuer un diagnostic des machines. L'objectif est dans un premier temps de connaître l'état général de la machine. Si une défaillance est détectée alors une inspection humaine est dans la plupart des cas engagée, afin de connaître avec précision le ou les composants impliqués et de s'assurer qu'il ne s'agit pas d'une fausse alarme.

En effet, ces systèmes de surveillance et de diagnostic ne sont pas parfaits et génèrent parfois des interventions inutiles. Dans certains cas extrêmes, une maintenance conditionnelle peut être plus coûteuse qu'une maintenance corrective. En effet, si l'interprétation des indicateurs est erronée alors cela engendre des interventions de maintenance inutiles. Ces interventions ont un coût et, dans le cas où les indicateurs ou leurs interprétations sont complètement faux, cela peut engendrer un coût d'interventions inutiles plus important que les coûts des défaillances non planifiées. Afin d'optimiser encore les coûts de maintenance, la localisation et l'évaluation de la gravité de l'anomalie doivent pouvoir être précises. En effet, en fonction du type de défaillance, les opérations à prévoir et leurs coûts peuvent être très différents. Les indicateurs que l'on cherche à obtenir doivent pouvoir faire la différence entre certains défauts de l'éolienne. Par exemple, la détection d'une défaillance d'un roulement de la génératrice amènera à des préparatifs différents de ceux liés à la détection de la défaillance du système de refroidissement. Dans le premier cas, des aménagements longs et la commande de pièces coûteuses seront à prévoir, dans l'autre cas, il s'agira de remplacer des pièces de taille réduite qui sont souvent en stock.

Il existe actuellement plusieurs outils et approches pour créer des indicateurs de surveillance qui possèdent leurs défauts et avantages propres (Tavner, 2012).

### 2.3.1. Condition Monitoring System (CMS)

Une des solutions disponibles dans le commerce consiste en un système dédié appelé CMS (condition monitoring system, ou système de surveillance conditionnelle). Cette appellation générique est d'ailleurs souvent source de confusion, car il existe de nombreux CMS différents qui sont listés dans (Yang et al., 2014). Ces systèmes se composent d'une série de capteurs supplémentaires installés dans l'éolienne permettant de surveiller un type de composants à la fois. Ils permettent de générer des indicateurs liés à des grandeurs physiques sélectionnées pour leurs relations avec l'origine des défaillances potentielles.

Tableau 2.5 Liste des CMS disponibles sur le marché (source (Feng Yanhui et al., 2014)Tab.1)

Techniques CMS	Composants surveillés
Thermocouple	Roulement génératrice, Convertisseurs, Transformateurs
Compteur de particule dans l'huile	Roulement multiplicatrice
Analyse vibratoire	Arbre principale, roulement principal génératrice, multiplicatrice, structure de la nacelle, tour et fondation
Test ultrason	Structure de la tour et des pales
Analyse électrique (tension et courant, charge et décharge)	Enroulement génératrice
Qualité de l'huile	Roulement multiplicatrice
Émissions acoustiques	Pale, roulement principal, génératrice et multiplicatrice, structure de la tour
Vibration angulaire	Arbre principal et multiplicatrice
Jauge de contrainte fibre optique	Pale
Thermographie	Pale, arbre principal, roulement principal, multiplicatrice, génératrice convertisseur, structure nacelle, transformateurs
Mesure de couple	Pale, roulement principal, arbre principal
Shock pulse method (SPM)	Roulement et engrenage

Dans le Tableau 2.5 sont listées la majorité des techniques CMS existantes (dans le domaine éolien), ainsi que les composants qu'elles surveillent. Ces techniques de CMS se concentrent sur les roulements et les arbres qui composent la chaîne cinématique. Cependant, des techniques comme la thermographie permettent de surveiller d'autres composants, mais elles ne sont pas aussi répandues dans la surveillance éolienne. Les performances en détection et en diagnostic des systèmes CMS et principalement de l'analyse vibratoire de la chaîne cinématique ne sont plus à démontrer, comme le rappellent les articles (McMillan and Ault, 2007; Yang et al., 2014). Ces techniques permettent d'identifier et de diagnostiquer très précisément des défauts dans les roulements, allant jusqu'à localiser le défaut dans une bille de roulement. La majorité de ces solutions se concentrent uniquement sur les composants critiques de la chaîne cinématique. La chaîne cinématique comprend le roulement principal, les roulements de la boîte à engrenages et ceux de la génératrice. Ces composants mécaniques sont les plus critiques en termes de coût de remplacement. Ils ne sont pas les composants qui tombent en panne le plus fréquemment, mais leur remplacement est rendu difficile et long par leur taille imposante et les difficultés d'approvisionnement. Les coûts de remplacement de la chaîne cinématique peuvent donc être grandement diminués si une maintenance conditionnelle y est appliquée. C'est pour cela que la recherche en la matière s'est grandement développée autour de ces composants en particulier.

Il est important de noter que le régime non stationnaire des éoliennes est toujours source de fausses alarmes pour ces systèmes. En effet, dans le cas des CMS vibratoires par exemple, les variations de la sollicitation de la machine vont, comme l'apparition d'un défaut, faire augmenter le niveau vibratoire des composants. Si ces variations de charge ne sont pas prises en compte, de fausses alarmes peuvent être générées. Ces fausses alarmes vont alors engendrer du travail d'analyse supplémentaire ou des interventions inutiles. Pour résoudre ce problème, la vitesse de rotation est toujours mesurée ainsi que la puissance produite, dans certains cas. Ces informations sur les conditions de fonctionnement de la machine vont permettre d'adapter le traitement qui est fait sur les signaux vibratoires et d'adapter la génération d'alarmes.

Malgré les nouvelles techniques actuelles, la prise en compte des variations de conditions de fonctionnement reste un problème pour ces CMS. Il n'est pas rare alors que des experts soient dédiés à l'analyse des signaux en détail, ce qui fait augmenter grandement les coûts de mise en place de stratégies de maintenance conditionnelle utilisant ces CMS.

Pour pouvoir surveiller toute l'éolienne, il faut ajouter plusieurs systèmes indépendants. Le prix du CMS peut devenir un problème pour des installations avec des marges faibles. La rentabilité des CMS pour l'éolien en mer est importante et permet de justifier de telles installations, mais dans l'éolien terrestre, cette rentabilité est encore discutée. La complexité de ces systèmes nécessite un expert pour l'analyse des résultats. Cela diminue la confiance portée à de tels systèmes par des personnes « non expertes ». Par conséquent, en pratique, une inspection visuelle sera toujours effectuée, augmentant encore les coûts de fonctionnement. Même s'ils sont performants, les CMS sont encore peu répandus dans l'éolien terrestre, car il nécessite un investissement lourd en matériel et personnel expert.

### 2.3.2. Supervisory Control and Data Acquisition (SCADA)

Le SCADA (Supervisory Control and Data Acquisition, ou système de contrôle et d'acquisition de données) est un système de télégestion implanté sur un automate installé dans la nacelle en parallèle de l'automate de contrôle. Ce système analyse les données de capteurs déjà présents dans l'éolienne et utilisés pour sa commande. La machine est contrôlée principalement grâce à sa puissance, à la vitesse de rotation des arbres, et aux angles des pales. En plus de ces mesures liées au fonctionnement principal de la machine, il existe aussi des mesures de température pour la surveillance des systèmes de refroidissement et la sécurité de la machine. Le SCADA a deux fonctions, l'acquisition des données et la supervision. L'acquisition permet d'avoir accès aux valeurs des capteurs en temps réel et aussi de stocker automatiquement ces valeurs dans une base de données. La supervision est réalisée grâce à un système d'alarmes automatisées qui donne des informations aux opérateurs en direct. Ces alarmes sont liées à des procédures de mise en sécurité qui, en fonction de la situation, vont stopper la machine.

Ils existent en moyenne 40 variables mesurées dans les SCADA modernes. Ces variables peuvent changer en fonction de la marque du SCADA ou du modèle de l'éolienne. Voici une liste fournie dans (Yang et al., 2013) des variables stockées au minimum dans les éoliennes :

- Puissance active, moyenne et écart type
- Puissance réactive
- Facteur de puissance
- Courant et tensions génératrices
- Vitesse anémomètre moyenne et écart type
- Vitesses arbre lent et rapide
- Température de roulement multiplicatrice (si turbine équipée)
- Température de l'huile multiplicatrice (si turbine équipée)
- Température d'enroulement statorique génératrice
- Température de roulement génératrice
- Température de la nacelle

Dans la liste ci-dessus, on peut voir apparaître les variables de contrôle de la fonction principale de génération d'électricité et aussi des variables liées à des fonctions plus secondaires comme le refroidissement. Le SCADA contient des données hétérogènes qui permettent de surveiller l'ensemble des composants de la nacelle. Afin de pouvoir stocker ces données, l'échantillonnage choisi est de 10

minutes. La moyenne, l'écart type, le minimum et le maximum des dernières 10 minutes sont stockés à chaque pas de temps. Ce choix d'une fréquence d'échantillonnage basse a été fait afin de limiter la taille de la base de données et de limiter les volumes de données qui transitent.

La surveillance est faite par la création d'alarmes et la mise en place de programmes automatisés de mise en sécurité. Ces alarmes sont aussi stockées dans la base de données et accessibles par les opérateurs. Chacune des alarmes permet d'avoir un descriptif succinct de l'état de l'éolienne. Ces alarmes sont aussi associées à une procédure d'arrêt et de relance. Une machine arrêtée peut être relancée automatiquement par le programme, relancée à distance, ou encore nécessiter la relance sur site par un opérateur. Ces trois procédures de relance sont liées à la gravité de la situation détectée par le SCADA.

La deuxième utilisation du SCADA est liée à l'acquisition des données. La création automatique des bases de données permet de surveiller de manière régulière la performance de la machine. La lecture de la puissance générée en fonction du vent permet de connaître la production de chaque turbine sur une période. La vérification de la performance de la machine est liée au métier d'exploitant. Le but est de pouvoir suivre la production au cours du temps et valider les prévisions de production établies lors de la création du parc. Il faut pouvoir assurer au propriétaire du parc éolien que le parc produit l'énergie pour lequel il a été créé. Le SCADA permet enfin d'analyser les causes d'une panne. Les données sont stockées en continu et permettent de connaître l'état de la machine avant une panne. Ces informations sont utiles pour pouvoir faire une analyse des causes. En effet, en lisant la liste des alarmes apparues avant une panne et en analysant l'évolution des variables mesurées, on peut connaître précisément l'état de la machine dans les instants qui ont précédé ce défaut. Le type d'alarmes générées avant l'arrêt permet de connaître les raisons de celui-ci et d'organiser l'intervention curative nécessaire.

La très grande quantité d'informations du SCADA est utilisée essentiellement pour des analyses a posteriori. De plus, les décisions de maintenance conditionnelle basées sur les données SCADA restent assez simples puisque les indicateurs ne sont basés que sur une seule grandeur à la fois et que les variations de conditions de fonctionnement ne sont pas prises en compte. Bien que les informations ainsi récoltées soient cruciales pour l'analyse des pannes, elles ne permettent pas d'améliorer la planification de la maintenance sans être prétraitées, ni analysées plus finement (Wenxian and Jiesheng, 2011).

## **2.4. Objectifs de la thèse**

Cette thèse s'inscrit dans le cadre d'une collaboration de type CIFRE entre le laboratoire GIPSA-lab et l'entreprise VALEMO qui propose des services d'exploitation et de maintenance pour des propriétaires de parcs éoliens. Le but de la thèse est de créer des indicateurs de santé de composants d'éoliennes à partir des données provenant du SCADA qui seront ensuite utilisés pour la maintenance conditionnelle et prédictive. VALEMO, en tant que représentant du propriétaire des parcs éoliens dans sa mission d'exploitant, a accès à toutes les données provenant du SCADA. Malheureusement, jusqu'à présent, ces données étaient très peu, voire pas, utilisées. En se munissant de ces indicateurs, VALEMO cherche à améliorer ses outils de surveillance pour pouvoir proposer de nouveaux services de suivi et de maintenance à ses clients. Grâce à la maintenance conditionnelle et prédictive, VALEMO veut pouvoir garantir plus de disponibilité à ces clients et donc proposer de nouveaux contrats plus rentables.

Les techniques de CMS utilisées par VALEMO jusqu'à présent sont jugées trop coûteuses pour les éoliennes terrestres de l'entreprise et difficilement utilisables au quotidien puisqu'elles génèrent trop d'alarmes qui doivent être interprétées longuement. Fort de ce constat, VALEMO cherche à développer un système abordable financièrement, générant un minimum de fausses alarmes et apportant suffisamment d'informations sur les défaillances détectées pour entreprendre une action de maintenance. Les SCADA possèdent l'avantage d'être complètement financièrement amortis (Wenxian and Jiesheng, 2011) et possèdent suffisamment de capteurs pour couvrir les composants les plus importants de l'éolien. Néanmoins, ces données, contrairement aux données CMS, ne sont pas conçues pour cette tâche d'élaboration d'indicateurs pour la maintenance conditionnelle, elles n'ont qu'une faible fréquence d'échantillonnage et les capteurs ne sont pas d'une grande précision.

Depuis 2001, plusieurs méthodes utilisant les données SCADA pour la création d'indicateurs ont cependant vu le jour et été publiées. Ces méthodes permettent de détecter les défauts naissants suffisamment tôt, mais elles n'ont pas réussi à proposer des systèmes automatisés générant un nombre limité de fausses détections. Un expert est donc toujours requis à plein temps pour interpréter toutes ces alarmes, ce qui est trop coûteux pour une entreprise comme VALEMO. Cette thèse rentre donc dans un contexte applicatif qui a pour but de créer des indicateurs utilisables directement par les opérateurs. Pour cela, l'entreprise VALEMO a mis à disposition l'intégralité de sa base de données pour valider les méthodes sur des situations réelles. De manière générale, la création d'indicateurs de santé est une problématique difficile. Elle l'est encore plus pour les éoliennes qui évoluent dans un milieu non stationnaire que pour machines qui évoluent à vitesse et couple fixes, comme les papèteries par exemple. Il faut pouvoir prendre en compte les variations des conditions de fonctionnement et de l'environnement.

L'objectif de ces travaux de thèse est donc de concevoir de nouvelles méthodes pour créer des indicateurs de surveillance éolienne à partir des données SCADA. Ces indicateurs doivent permettre de détecter les défauts les plus critiques en générant suffisamment peu de fausses alarmes pour être implémentés chez VALEMO.



### **3. État de l'art sur la création d'indicateurs de défaut pour les éoliennes en utilisant des données SCADA**

L'objectif de cette thèse est l'élaboration d'indicateurs de dégradation à l'aide de données SCADA 10 minutes. Ces indicateurs seront utilisés pour la mise en place d'une maintenance conditionnelle. L'indicateur doit permettre de détecter un comportement anormal de l'éolienne, associé à un défaut en cours d'apparition. Un défaut correspond à un fonctionnement dégradé d'un composant. Une machine en défaut peut continuer à produire, mais à terme, le défaut évoluera vers une défaillance, qui provoquera l'arrêt de la machine. L'indicateur doit donc détecter le défaut au plus tôt. Idéalement, il doit aussi localiser et isoler l'origine du défaut, c'est-à-dire déterminer le composant en défaut et évaluer son niveau de gravité. On parle alors de diagnostic sur la machine. Seule cette information précise permettra une maintenance conditionnelle efficace.

Depuis que les SCADA sont intégrés dans tous les parcs éoliens, les données s'accumulent dans les bases de données des industriels. Néanmoins, ce n'est qu'en 2001 que les premiers travaux de recherche ont été publiés pour fournir des solutions au problème de surveillance en utilisant des données SCADA 10 min dans (Li et al., 2001). Puis il faudra attendre les années 2010 pour que les travaux se multiplient pour désormais devenir un domaine en émergence avec (Wei and Liu, 2010), (Gray et al., 2011) ou encore (Farkas, 2011). Le faible coût et la grande quantité de ces données ont motivé le développement de nouvelles techniques de surveillance pour l'éolien.

Traditionnellement, on trouve deux types d'approches en diagnostic de système : les méthodes utilisant des modèles de connaissance et les méthodes utilisant les données.

Dans les méthodes utilisant des modèles de connaissance, le fonctionnement de la machine est décrit de manière précise à l'aide d'un modèle mathématique, généralement à base d'équations différentielles. Les écarts entre le modèle et la machine servent d'indicateurs de défauts. Ces méthodes supposent des modèles fiables et précis, qui sont impossibles à obtenir dans le domaine éolien. Les modèles de connaissance y sont rares pour deux raisons : il y a un manque de connaissances sur les phénomènes liés à la création de défauts qui sont très complexes et il n'y a pas actuellement de modèles capables de décrire précisément tous les phénomènes physiques en jeu lors de périodes de fonctionnement sans défaut. Enfin, les données disponibles actuellement sont des données avec une faible précision et une fréquence d'échantillonnage basse, mal adaptées aux méthodes à base de modèles. Celles-ci obligeraient à rajouter des capteurs et des systèmes d'acquisitions coûteux.

L'approche basée sur les données consiste à utiliser les données pour construire des indicateurs de dégradation. Le comportement de la machine à un instant donné est comparé à son comportement lorsqu'elle fonctionnait sans défaut. Le bon fonctionnement d'une machine est décrit à l'aide de modèles empiriques, appris sur des jeux de données sans défauts. Les modèles peuvent être des expressions mathématiques décrivant des liens entre différentes variables. Dans ce cas, les indicateurs de défauts mesurent l'écart entre la variable prédite par le modèle et sa valeur enregistrée en ligne, ou encore l'écart entre le modèle appris avec celui identifié en ligne. Les modèles peuvent aussi représenter des classes localisées dans un espace de caractéristiques. Les indicateurs de défauts mesurent alors l'écart des données enregistrées en ligne à ces classes de bon fonctionnement. Ces dernières méthodes permettent la détection d'une évolution du système, mais ne permettent pas de localiser un défaut. Pour

cela, il est nécessaire d'apprendre aussi des classes de défaut, ce qui nécessite un nombre de points enregistrés suffisants lors de l'occurrence des défauts, qui ne sont pas toujours disponibles.

Le choix de travailler avec des modèles créés à partir des données est motivé par la grande quantité de données disponibles, ce qui permet un apprentissage fin. De plus, le temps de calcul n'est pas un problème puisque l'apprentissage est fait hors ligne et que les ordinateurs récents sont tout à fait capables de gérer ce genre de bases de données. La prise de décision, réalisée en ligne, nécessite, quant à elle, souvent moins de calcul.

Dans la suite de ce mémoire, nous nous concentrerons sur les méthodes publiées dans la littérature, développées pour la surveillance d'éoliennes et utilisant des données SCADA basse fréquence. Nous avons regroupé ces méthodes en deux catégories : les méthodes dites « mono-turbines » où les indicateurs de défauts sont construits à partir de variables enregistrées sur une seule turbine et les méthodes « multi-turbines » où les indicateurs sont construits à partir de variables enregistrées sur différentes turbines d'un même parc. Enfin, notre analyse de l'état de l'art nous amènera à proposer une nouvelle approche, que nous introduirons dans le dernier paragraphe de ce chapitre.

### **3.1. Les méthodes mono-turbines**

#### **3.1.1. Principe général des méthodes mono-turbines**

Les méthodes mono-turbines modélisent l'évolution d'une variable SCADA d'une turbine en fonction d'autres variables de la même turbine. Un modèle mathématique liant ces variables entre elles est construit à partir de données sélectionnées parmi de nombreuses données collectées et enregistrées dans la base SCADA. On peut ainsi comparer le comportement actuel de la turbine avec son comportement passé.

Dans cette partie, deux approches différentes pour créer ces modèles sont présentées. Ces approches se différencient par la nature des relations qu'elles cherchent à décrire : relation de cause à effet due à la physique des phénomènes mis en jeu, ou relation de similarité de comportement entre des composants au sein d'une même machine. Ainsi, dans la première approche, les données utilisées pour modéliser l'évolution d'une variable sont sélectionnées à partir du comportement physique de la machine. On choisit les variables dont on sait qu'elles ont une action sur la variable dont on cherche à prédire l'évolution. Dans la deuxième approche, les données sont sélectionnées en fonction de la similarité existant entre deux composants. Cette approche compare des grandeurs similaires mesurées sur des composants différents en modélisant la relation qui les lie. Pour les deux approches, on vérifie que la variable modélisée conserve le même comportement dans des conditions identiques de fonctionnement de la machine (c'est-à-dire celles pour lesquelles le modèle a été construit), en analysant l'écart entre la valeur mesurée et la valeur prédite.

Pour détecter des anomalies de comportement, les méthodes présentées vont soit créer un résidu entre la prédiction du modèle et la valeur mesurée, soit comparer deux modèles construits sur des périodes différentes. Le choix des variables d'entrée et la construction du résidu ont une influence à la fois sur la complexité du modèle et sur ses performances.



### 3.1.2. Méthodes utilisant les liens de causalité

Les méthodes utilisant les liens de causalité permettent de lier les variables par des liens causaux. On cherche à expliquer l'évolution d'une variable vue comme la sortie d'un système, qui a pour entrée d'autres variables de la même turbine. Le choix du système ainsi décrit est crucial, car il est lié au défaut que l'on pourra ou non détecter. En effet, pour qu'un défaut soit détecté, il faut qu'il ait un impact significatif sur le système que l'on observe en particulier. Les méthodes utilisant les liens de causalité tentent de modéliser les causes physiques de l'évolution d'une variable suivant l'équation ci-dessous :

$$y_j^i(t) = f_j^i(y_1^i(t), y_2^i(t), \dots, y_k^i(t)) \quad (3.1)$$

où

$i$  correspond au numéro de la turbine

$j$  correspond au numéro de la variable

$y_k^i(t)$  correspond à la  $k^{\text{ème}}$  variable, mesurée à l'instant  $t$  sur la turbine  $i$

#### 3.1.2.1. Courbes de puissance

L'approche « vent-puissance » décrit le fonctionnement de l'éolienne dans son ensemble. On cherche à vérifier que la machine remplit sa fonction première : générer de l'énergie. La sortie du système est donc la puissance électrique générée, la ou les entrées sont liées aux conditions environnementales mesurables. Ce rapport entre la puissance et l'environnement est appelé courbe de puissance électrique. Il s'agit de l'outil le plus connu et le plus utilisé de nos jours pour la surveillance des performances des éoliennes. Les courbes les plus simples représentent la puissance électrique générée à un instant donné par rapport à la vitesse du vent mesurée au même instant, des courbes plus évoluées prennent aussi en compte d'autres mesures de l'environnement, comme la densité de l'air, la nature du vent ...

Dans les méthodes présentées dans la suite de ce paragraphe, la variable  $y_j^i(t)$  sera toujours la puissance,  $y_k^i(t)$  comprendra soit le vent venant de la turbine  $i$  seul, soit d'autres variables explicatives supplémentaires.

##### 3.1.2.1.1. Présentation de la courbe de puissance

La courbe de puissance permet d'évaluer le fonctionnement global de la turbine. Elle fournit une indication sur le rendement théorique attendu entre la puissance mécanique fournie par le vent (mesurée de manière indirecte par la vitesse du vent) et la puissance électrique délivrée par la machine. Cette courbe devrait permettre de détecter la présence de pannes naissantes. En effet, si un défaut apparaît, le rendement de la machine diminue. Cette diminution engendre des pertes de performance liées à la gravité du défaut. La courbe de puissance présentée en bleu dans la Figure 3.1 est le lien le plus simple que l'on puisse établir entre l'environnement et la puissance générée. La vitesse du vent est placée en abscisse et la puissance générée en ordonnée. Cette représentation permet de voir qu'il existe plusieurs phases dans la génération de puissance, une phase de démarrage, une phase de montée en puissance et une phase de saturation.

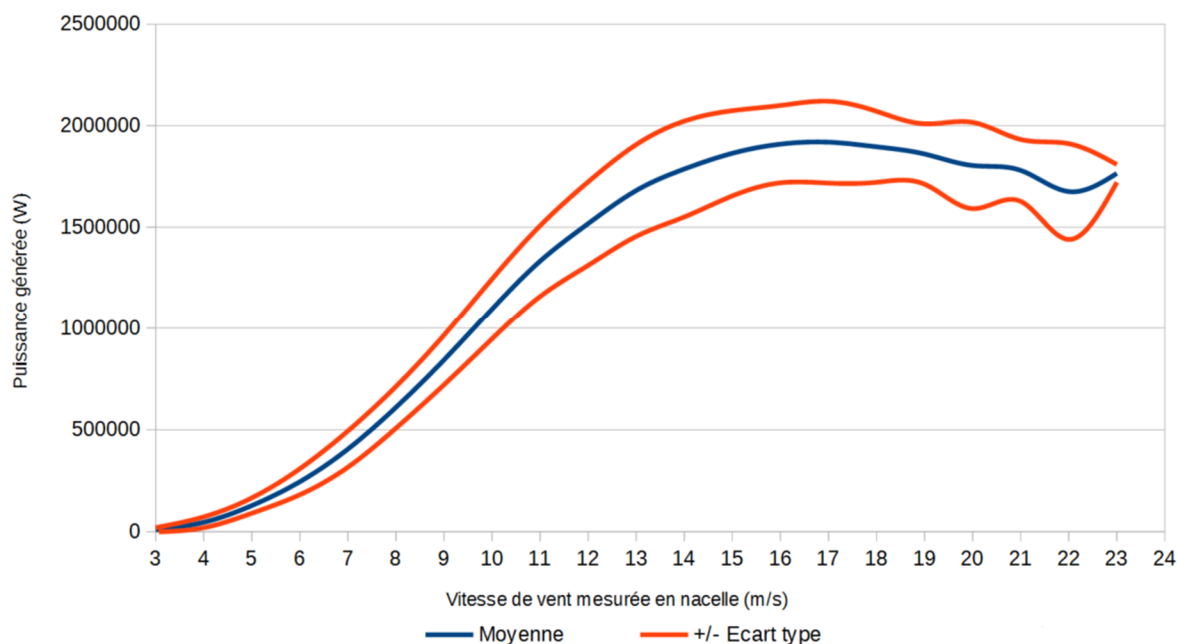


Figure 3.1 Moyenne et écart type de la puissance générée mesurés en fonction de la vitesse du vent (source VALEMO)

La difficulté dans l'utilisation de la courbe de puissance est sa dispersion. La Figure 3.1 permet de voir l'écart type de la puissance en fonction de la vitesse du vent. Ainsi, pour une mesure de vent donnée, en fonctionnement normal, la mesure de puissance admet une gamme de valeurs acceptables qui augmente avec la puissance et peut atteindre jusqu'à plus ou moins 200 kW par rapport à la valeur théorique. Cette dispersion est due à de nombreux facteurs environnementaux comme la nature du vent (turbulent, laminaire), la densité de l'air ou l'angle d'inclinaison des pales qui influent sur la quantité d'énergie mécanique transmise à la turbine. La détection ne peut se faire que lorsque le défaut est suffisamment important pour engendrer une perte de performance significative. La puissance nominale des turbines actuelles est de l'ordre de 2MW. Même avec les meilleurs composants, le rendement d'une turbine n'est jamais de 100 %. Une turbine qui possède un rendement de 99 % (que l'on peut considérer comme excellent) perd une puissance de 20kW dans la partie saturée de la courbe de puissance (cf Figure 3.1, vitesse de vent supérieur à 15m/s). Un défaut ne peut être détecté sur cette turbine que si les pertes qu'il engendre sont supérieures à 20kW. Cette quantité d'énergie est élevée et on s'attend à détecter des défauts déjà importants.

### 3.1.2.1.2. Modélisation de la courbe de puissance

La première partie de (McLaughlin et al., 2009) décrit la technique actuelle utilisée par les exploitants pour analyser la courbe de puissance. Afin de savoir si la machine fonctionne correctement, les exploitants comparent les mesures SCADA à la courbe de puissance de référence donnée par le constructeur. Cette courbe est construite à partir de mesures de vent et de puissance enregistrées lors de la mise en service. La production d'énergie est garantie par le constructeur à  $\pm 5$  %. La méthode proposée par l'auteur consiste à collecter les données de puissance et de vent d'une turbine pendant une période de temps suffisamment longue afin de reconstruire en temps réel la courbe de puissance de la turbine. La méthode présentée est graphique. Les données mesurées sont superposées à la

courbe de référence comme montrée sur la Figure 3.2. En général, la longueur de la période de mesure choisie est d'une semaine (environ 1000 points).

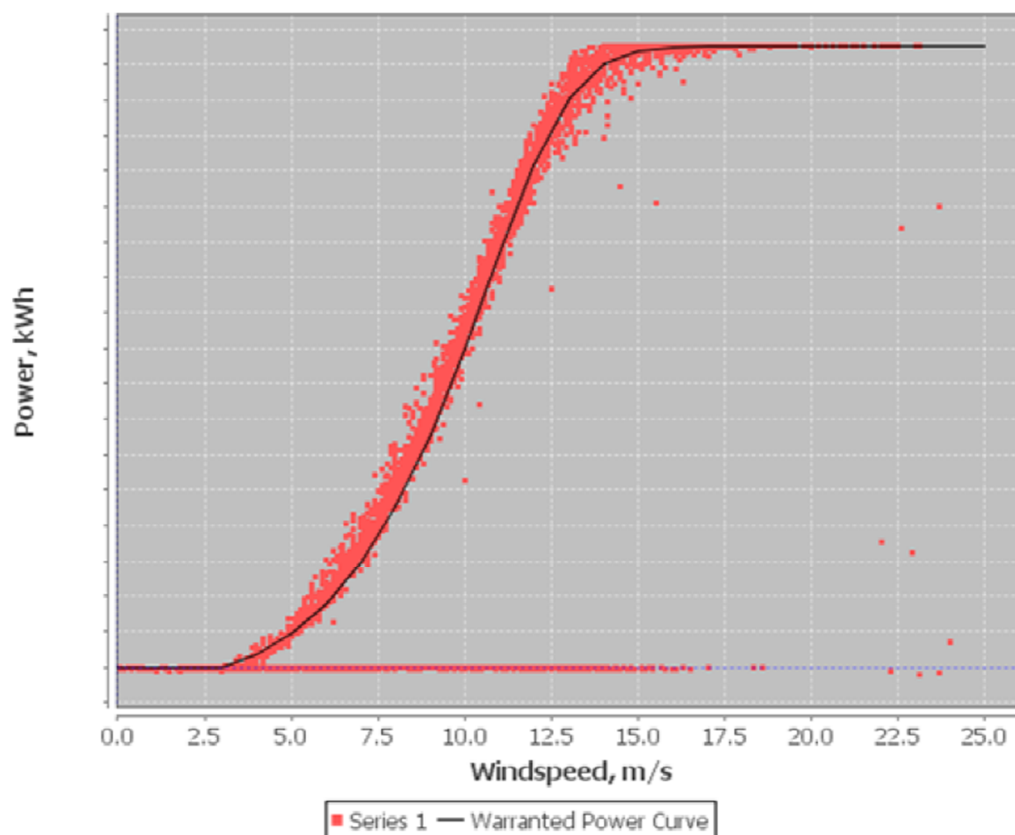


Figure 3.2 Courbe de puissance réelle en rouge versus courbe garantie par le constructeur en noire (source (McLaughlin et al., 2009)fig.1)

La courbe construite à partir des données mesurées sur la semaine observée est alors comparée à celle du constructeur. Tout écart anormal entre les deux permet de mettre en évidence une perte de production liée à un défaut. Il n'y a pas de méthodes automatiques pour évaluer l'importance de l'écart entre les courbes. La criticité de la perte de production et donc du défaut est évaluée grâce à l'expérience de l'utilisateur. Dans certains cas, la nature du défaut peut être connue en fonction de l'allure de la courbe observée.

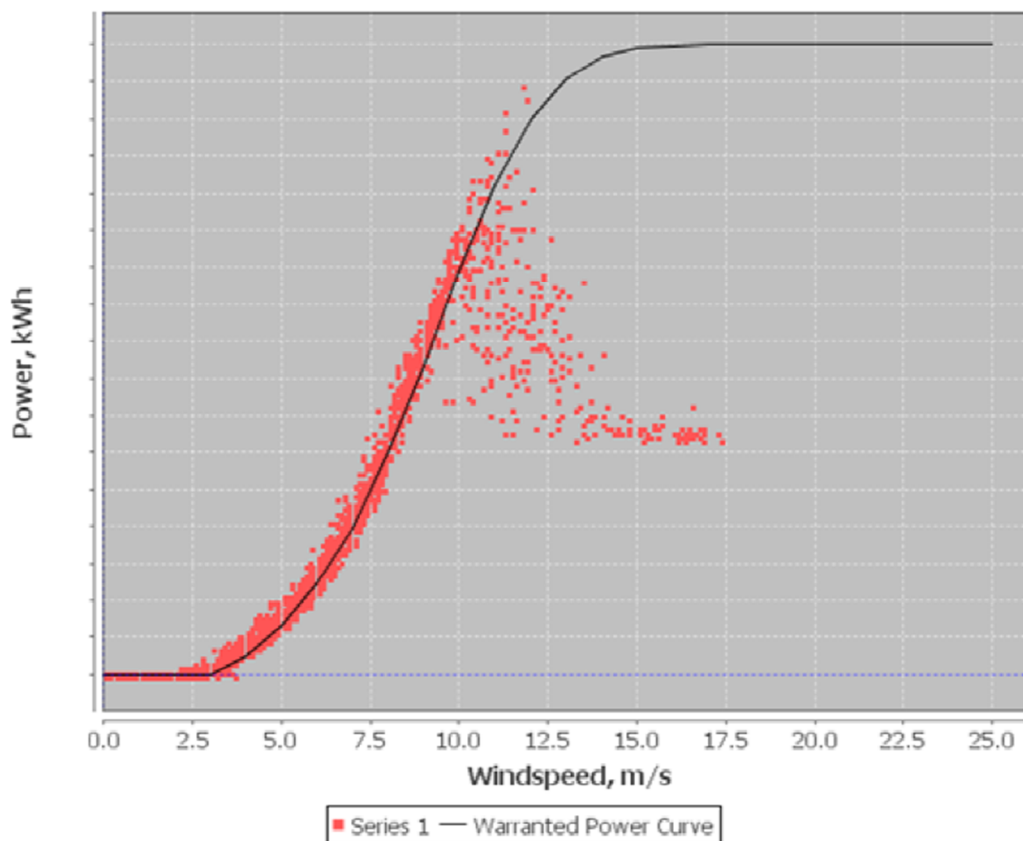


Figure 3.3 Exemple de contrôle commande défaillant mis en évidence par la courbe de puissance (source (McLaughlin et al., 2009) fig.4)

Par exemple, on peut voir sur la Figure 3.3, qui représente la courbe de puissance d'une éolienne dont le contrôle/commande est défaillant, que la puissance n'atteint pas celle attendue pour des vitesses de vent supérieures à 10 m/s. Pour ce type d'éoliennes, les pales commencent à s'orienter (cf Figure 2.5) aux alentours de cette vitesse. La chute de la courbe de puissance est donc due à une mauvaise orientation des pales. Cette « signature » peut être associée à un défaut de contrôle commande.

La méthode proposée dans l'article permet aussi de mettre en évidence des défauts d'anémomètres ou alors de détecter la présence de givre. Ces 3 défauts diminuent suffisamment le rendement pour être détectés. Cependant, notons que, même si la méthode permet la détection de ces défauts, cela ne peut se faire de manière suffisamment précoce. Il faut en effet un nombre important de points pour qu'une différence entre la courbe pratique et théorique devienne visible. Enfin, la courbe constructeur n'est pas forcément la référence la mieux adaptée. Les exploitants de parcs éoliens l'utilisent, car elle permet de vérifier que le contrat passé avec le constructeur est bien respecté. Cependant, elle est légèrement sous dimensionnée. On peut voir sur les Figure 3.2 et la Figure 3.3 que les échantillons en rouge, représentant les mesures en production, sont pour la majorité bien au-dessus de la courbe de puissance garantie par le constructeur, tracée en noir, ce qui montre que cette courbe garantie est sous dimensionnée. L'écart entre les deux courbes est ainsi biaisé. Un défaut générant une perte de performances sera plus difficile à déceler.

Enfin, rappelons que (McLaughlin et al., 2009) ne propose pas de méthodes automatiques pour détecter des défauts connus ou toute autre anomalie. La décision dépend donc de l'appréciation de

l'expert, qui peut ne pas connaître toutes les spécificités du parc. Certains défauts peuvent donc être détectés tard, voire ne pas être détectés du tout. La détection est d'autant plus difficile que la dispersion autour de la courbe de référence constructeur est grande. Cette dispersion reste le problème majeur de cette méthode.

---

Dans (Li et al., 2001), les auteurs proposent une méthode automatique pour créer un indicateur de performance. Il est paru en 2001 et est l'un des premiers articles sur l'utilisation de données SCADA pour la surveillance éolienne. Cet article propose de construire un modèle de la puissance en fonction de la vitesse et de la direction du vent. Ces variables proviennent de deux mâts de mesure installés sur le parc, ce qui permet d'avoir des valeurs de vitesse et de direction du vent plus précises que celles fournies par les anémomètres installés sur les turbines. Cette méthode a pour but de diminuer la dispersion en prenant en compte un facteur environnemental supplémentaire, la direction du vent, et en utilisant des capteurs plus précis. L'article montre que la direction du vent est un facteur important sur la puissance générée. En fonction de la direction d'origine du vent, celui-ci peut être plus ou moins turbulent en fonction des obstacles qu'il rencontre avant d'atteindre l'éolienne. Un vent turbulent a environ la même vitesse moyenne qu'un vent laminaire, mais ne transmettra pas l'énergie aux pales de la même manière. Les turbulences diminuent non seulement l'énergie transmise, mais génèrent aussi des vibrations dans les pales. L'article suppose que les obstacles sont immobiles et invariants au cours du temps. Leur impact sur le vent ne dépend donc que de la direction du vent.

Pour pouvoir modéliser le rapport vent puissance, un réseau de neurones est utilisé avec une base d'apprentissage constituée de données enregistrées pendant une période sans défaut d'un mois. Le résidu est généré en comparant la valeur prédite par le réseau de neurones et la valeur de puissance réelle mesurée par le SCADA, à chaque période d'échantillonnage. Cet article montre que l'augmentation significative du résidu permet la détection d'un défaut, mais pas son identification.

Dans cet article, (Li et al., 2001) compare aussi les valeurs de puissance prédites par le réseau de neurones avec les valeurs fournies par la courbe de référence du turbinier. La différence entre la valeur mesurée de puissance et la puissance prédite par le réseau de neurones au cours de la période d'apprentissage et sur un mois de données de validation est plus faible que l'écart observé entre la mesure de puissance et la puissance théorique fournie par la courbe de référence. D'après l'auteur, le réseau de neurones a de meilleurs résultats puisqu'il est appris sur un mois de données alors que la courbe de référence est construite à partir de mesures enregistrées quelques jours au début du fonctionnement de la machine. La méthode proposée dans l'article prend aussi en compte la direction du vent, ce que la courbe de référence ne fait pas. La dispersion est nettement diminuée grâce à cela. En revanche, cette méthode présente certaines limites. La variation du résidu est beaucoup plus élevée lors du mois de validation que lors du mois d'apprentissage alors qu'il n'y a pas de défaut. Cette augmentation de la variation du résidu peut être due à la modification d'autres facteurs environnementaux non pris en compte.

Pour conclure, cette méthode permet de se rendre compte que la courbe de puissance doit être enrichie pour être utilisée efficacement. Par exemple, l'utilisation de la direction du vent permet de diminuer la dispersion de la courbe. Notons que, même si l'article propose un indicateur automatique,

celui-ci ne permet pas d'identifier le défaut. Il faut revenir à l'analyse de l'allure de la courbe de puissance pour déterminer s'il s'agit d'un défaut connu.

Dans (Farkas, 2011), l'utilisation de la densité de l'air dans la modélisation de la courbe de puissance a permis de diminuer l'erreur de prédiction jusqu'à 16 %. En reprenant, ces résultats (Butler et al., 2013) construisent un modèle utilisant le vent et la densité en entrée et la puissance en sortie. La densité est mesurée grâce à un mât météorologique présent sur place. Notons qu'il est rare qu'une mesure de la densité soit faite par le SCADA. Le modèle est un modèle de régression gaussienne appris à l'aide d'un jeu de données sans défaut sur une durée de un an. Afin de faciliter l'apprentissage, les données sont d'abord sélectionnées afin de ne garder que les points correspondant à la partie centrale de la courbe de puissance. En supprimant la partie de saturation et de démarrage, la relation entre vent et puissance se rapproche d'une relation linéaire. Cela permet d'améliorer les performances de prédiction en ne sacrifiant qu'un nombre limité de points, soit environ 30 %. En effet comme montré dans la section 2.1.2 et la Figure 2.5 décrivant les modes de fonctionnement de l'éolienne, la majorité du temps l'éolienne produit dans la partie centrale de la courbe de puissance.

Un résidu est créé en prenant la différence entre la valeur mesurée et la valeur prédite par la régression. L'article montre que, comme prévu, l'utilisation de la densité a permis de diminuer l'erreur de prédiction suffisamment pour permettre la détection d'un défaut de roulement principal. La Figure 3.4 permet de voir le lien entre courbe de puissance et densité. La puissance augmente avec la densité.

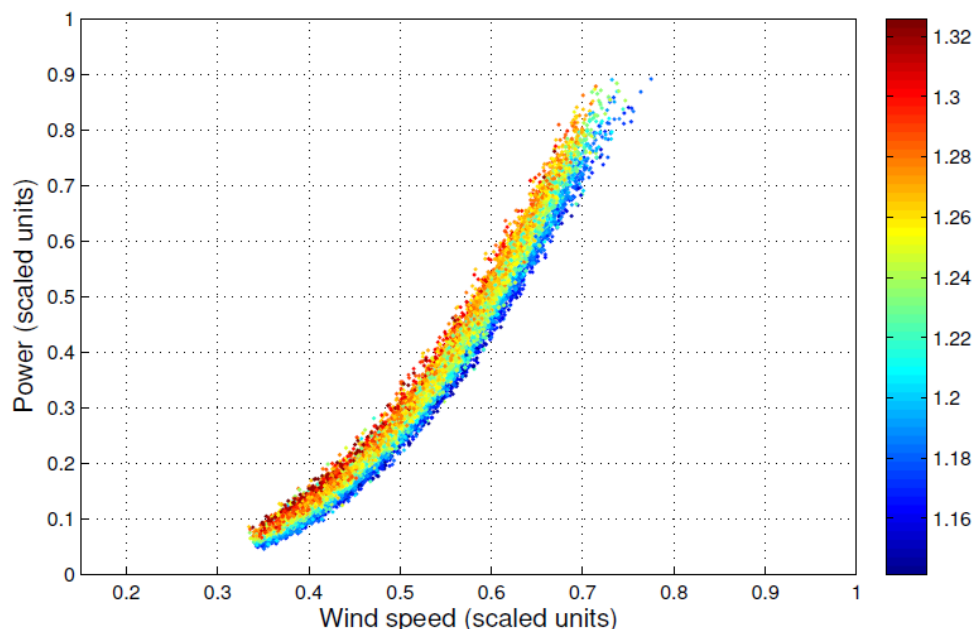


Figure 3.4 Courbe de puissance normalisée mesurée pendant 1 an où la couleur est fonction de la densité (source (Butler et al., 2013) fig.4)

(Butler et al., 2013) montre encore une fois l'influence importante de l'environnement sur la courbe de puissance. Cette influence est en partie prise en compte en augmentant fortement la taille de la base

d'apprentissage et en prenant la densité en entrée du modèle. Notons que la taille de la base sera un problème pour les nouvelles turbines, puisqu'il faudra un an avant de pouvoir détecter un défaut. De plus, la densité, elle, n'est que rarement mesurée sur les nacelles, ce qui peut être un frein à la méthode. Il faut la présence d'un mât météorologique pour acquérir sa valeur. Il est cependant possible de l'estimer grâce à la température extérieure, qui, elle, est toujours mesurée. Il est intéressant de noter que la sélection des données pourrait aussi être réalisée grâce à la vitesse de rotation de l'arbre. En effet, sur la partie centrale de la courbe de puissance, là où le rapport vent puissant est presque linéaire (cf Figure 2.5), soit la vitesse de rotation varie soit elle reste constante. On peut imaginer créer deux modèles différents pour ces deux plages de fonctionnement ou alors prendre la vitesse de rotation en entrée du modèle. Prendre en compte la vitesse de rotation peut diminuer la dispersion due à la différence dans la répartition des efforts entre les modes de fonctionnement identifiés (cf Figure 2.5).

Dans (Butler et al., 2013), même si le défaut a été détecté, l'identification de celui-ci n'a été faite qu'a posteriori, la méthode proposée ne permettant pas de le faire de manière automatique. Cependant, il est important de remarquer que la diminution de la dispersion de la courbe de puissance a permis de détecter un défaut de multiplicatrice. Les méthodes précédentes détectent uniquement des défauts venant du contrôle commande. En continuant de prendre en compte d'autres facteurs environnementaux, on peut espérer détecter d'autres défauts ou les détecter plus tôt.

---

L'article (Elijorde et al., 2014) propose de diviser la courbe de puissance en plusieurs zones de fonctionnement et de traiter chacune des périodes indépendamment, afin de diminuer la dispersion. La méthode consiste dans un premier temps à regrouper en familles les données d'une base de données de bon fonctionnement. Pour cela, l'auteur utilise l'algorithme des K-moyennes (MacQueen, 1967) en fixant le nombre de familles à 5. Les données sont regroupées en utilisant la distance euclidienne entre vecteurs formés de 4 variables : la puissance générée, la vitesse du vent, la vitesse de rotation de l'arbre rapide et l'angle d'une des pales. La Figure 3.5 représente la courbe de puissance avec les familles ainsi construites, il s'agit d'une projection utilisant uniquement la puissance et la vitesse du vent. On peut remarquer que les familles construites se rapprochent des modes de fonctionnement mis en évidence dans la Figure 2.5.

Une fois les familles identifiées, l'auteur propose d'utiliser l'algorithme FP-Growth pour extraire des règles de type si-alors expliquant les liens que l'on observe entre certaines variables quand le système est en bon fonctionnement, au sein de chaque famille.

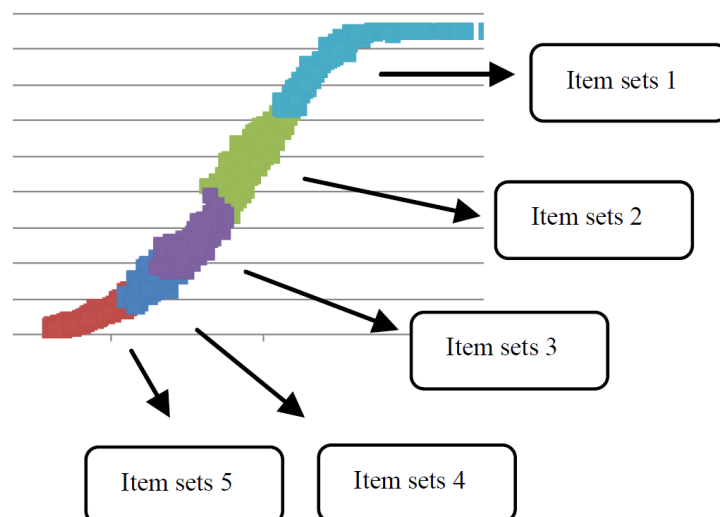


Figure 3.5 Courbe de puissance avec le découpage en familles (source (Elijorde et al., 2014)fig.2)

Afin de détecter un défaut, chaque nouvelle observation est associée à une des familles créées et on vérifie que les règles de la famille sont bien respectées. L'auteur de (Elijorde et al., 2014) propose des indicateurs qui vont compter le nombre de fois où les règles ne sont pas respectées par variable. Si une variable est responsable d'un non-respect de règles, on incrémente son compteur. Cela permet de limiter les fausses alarmes. En effet, l'auteur montre que les compteurs peuvent s'incrémenter alors qu'il n'y a pas défaut. C'est l'évolution des indicateurs qui permettra de détecter la présence d'un défaut réel. Si le compteur augmente de manière jugée anormale, alors un défaut est considéré comme présent. Malheureusement, l'évaluation de l'évolution des compteurs reste à l'appréciation de l'utilisateur. En revanche, puisqu'il y a en fait un compteur par variable, on peut avoir un diagnostic simple du défaut. Si le compteur correspondant à « l'angle de pale » augmente, il peut s'agir d'un défaut du système de commande, si c'est celui de la vitesse du vent qui augmente, il peut s'agir d'un défaut d'anémomètre.

Cette méthode montre comment, en utilisant d'autres variables que la vitesse du vent et la puissance, on peut diminuer la dispersion de la courbe de puissance. La vitesse de rotation et l'angle des pales permettent de définir des familles de données ayant des caractéristiques plus communes. Ces familles sont en fait les modes de fonctionnement de la machine. La densité de l'air n'est pas prise en compte car elle n'est probablement pas disponible. Prendre en compte l'environnement permettrait sans doute d'avoir des règles plus précises et de diminuer encore la dispersion. Un des désavantages de cette méthode provient du fait que le nombre de règles créées est très important, plusieurs centaines. Elles sont donc difficilement lisibles et utilisables par les opérateurs pour le diagnostic, qui reste donc limité.

### 3.1.2.1.3. Conclusion sur l'utilisation de la courbe de puissance pour la surveillance

Les méthodes présentées permettent, grâce à l'analyse de la puissance produite par l'éolienne, d'associer une perte de performance à l'apparition d'un défaut. Néanmoins, les défauts détectés sont pour la plupart associés au système de contrôle, car ils engendrent une forte perte de puissance. Un défaut même mineur sur le système de contrôle entraîne une perte de rendement significative. Pour pouvoir détecter d'autres défauts, il faut impérativement prendre en compte l'environnement comme la densité et la direction du vent afin de réduire la dispersion des points autour de la courbe. La densité



étant rarement présente dans le SCADA, celle-ci doit être estimée pour être utilisée. La température extérieure mesurée par le SCADA peut offrir une bonne image de la densité. Même réduite, la dispersion de la courbe de puissance reste un problème. En effet, elle ne peut pas être inférieure aux variations propres à un fonctionnement normal de la machine. Enfin, la courbe de puissance traduit le fonctionnement de l'ensemble de la turbine qui intègre trop de composants pour que les défauts puissent être isolés. Afin de pallier ce problème d'isolation des défauts, une solution est de décomposer l'éolienne en sous-systèmes plus petits et plus simples, pouvant être distingués les uns des autres.

### 3.1.2.2. Décomposition en sous-systèmes

Dans la section précédente, on peut voir que l'analyse de la courbe de puissance montre certaines limites pour la surveillance éolienne. La première difficulté est d'identifier et de localiser la panne. La courbe de puissance permet d'avoir une information sur la performance du système tout entier. L'apport de chaque composant dans la génération d'énergie ne peut être calculé. Les solutions proposées maintenant dans cette section consistent à diviser l'éolienne en sous-systèmes. Ce découpage a pour but d'isoler un composant ou un groupe de composants et de comparer son comportement courant par rapport à son comportement lors d'une période de temps sans défaut à l'aide d'un modèle. Un résidu formé de la différence entre l'observation mesurée et celle prédite par le modèle permet la détection de défauts, dont l'origine se situe au niveau du système analysé. Si le découpage en sous-systèmes est suffisamment fin, le défaut peut être localisé géographiquement dans l'éolienne et associé à un composant.

#### 3.1.2.2.1. Notion de sous-système et principe de décomposition

Un sous-système est défini comme un ensemble de composants liés géographiquement auxquels sont associés des mesures SCADA. Un sous-système possède un ensemble de variables de sortie dont l'évolution peut être expliquée par un ensemble de mesures issues du même sous-système. La variable de sortie traduit le fonctionnement des composants.

L'équation d'évolution de la variable de sortie est conforme à l'équation (3.1), rappelée ci-dessous :

$$y_j^i(t) = f_j^i(y_1^i(t), y_2^i(t), \dots, y_k^i(t)) \quad (3.2)$$

où

$i$  correspond au numéro de la turbine,

$j$  correspond au numéro de la variable et

$y_k^i(t)$  correspond à la  $k^{\text{ème}}$  variable, mesurée à l'instant  $t$  sur la turbine  $i$

La sortie n'est plus nécessairement la puissance de la turbine  $i$ . Les  $y_k^i(t)$  en entrée sont principalement des variables liées aux modes de fonctionnement, comme la vitesse de rotation, la puissance générée ...

Afin de localiser simplement un défaut, l'idéal est de considérer chaque pièce de la machine comme un sous-système. Or, toutes les grandeurs physiques ne sont pas forcément mesurées dans le SCADA. Certains composants ne possèdent aucune variable associée. La décomposition en sous-systèmes ne peut donc être limitée à une pièce, et un sous-système contiendra parfois plusieurs composants.

En règle générale, les sous-systèmes sont définis de manière à correspondre aux groupes de composants décrits dans la Figure 2.4, c'est-à-dire des composants qui sont reliés physiquement. Cela présente deux avantages : les composants forment un groupe physique que l'on peut isoler des autres,

ce qui permet de localiser le défaut ; un sous-système correspond à un groupe de composants dont le remplacement est fait en même temps. Si un défaut est avéré sur un des composants de la multiplicatrice, il est fort probable que toute la multiplicatrice soit changée. Il en est de même pour la génératrice.

Dans la suite, nous présenterons plusieurs exemples de surveillance de sous-systèmes. Dans un premier temps, les méthodes qui suivent cherchent à détecter des défauts dans les sous-systèmes de la chaîne cinématique. Dans un deuxième temps, ces méthodes sont généralisées au reste de la turbine.

### 3.1.2.2.2. Découpage de la chaîne cinématique en modèles statiques

Les sous-systèmes définis dans cette partie appartiennent à la chaîne cinématique. Rappelons que ses composants sont les plus critiques en termes de prix et de temps de réparation. C'est pourquoi les systèmes CMS d'analyse vibratoire se sont d'abord concentrés sur ces composants. (Wilkinson et al., 2014) présente une méthode pour étudier les variations de comportement des composants de la chaîne cinématique à l'aide de variables SCADA. Plusieurs sous-systèmes sont étudiés : ici le rotor, la multiplicatrice et la génératrice. La performance énergétique de ces sous-systèmes est difficile, voire impossible à modéliser complètement, car il n'existe pas de mesure précise de l'énergie en entrée et de l'énergie en sortie. La méthode présentée se concentre sur la modélisation des pertes dues à l'échauffement thermique. Pour cela, des modèles sont construits pour prédire la température d'un ou de plusieurs composants des sous-systèmes.

Le nombre de modèles associés au sous-système « génératrice » est en général supérieur aux autres. Le choix du nombre de modèles est uniquement guidé par le nombre de variables disponibles. Certains SCADA mesurent la température de tous les roulements de la multiplicatrice, d'autres ne mesurent que la température de l'huile.

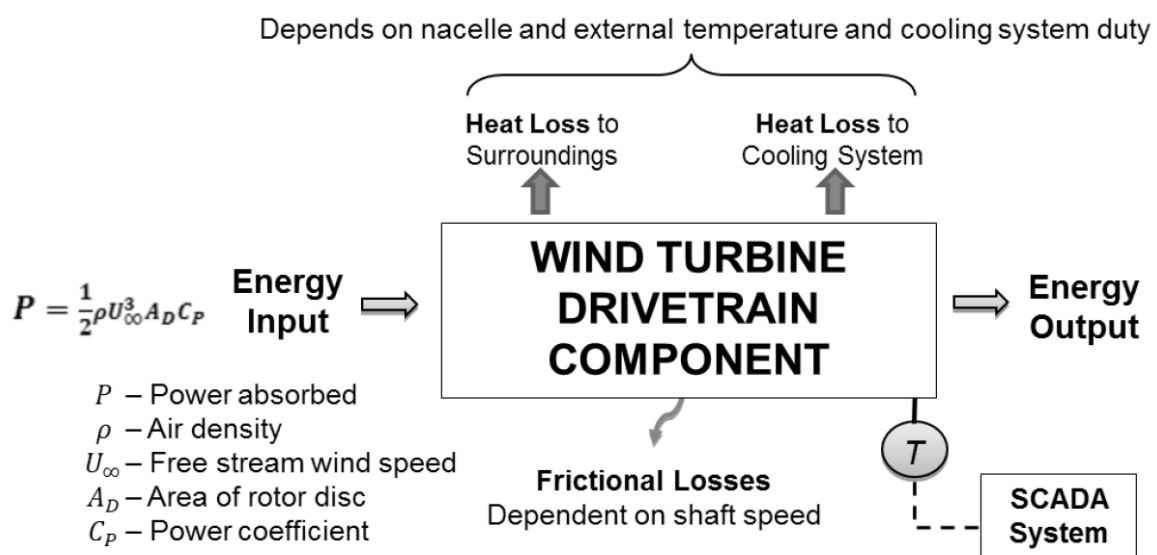


Figure 3.6 Équilibre énergétique d'un composant de la chaîne cinématique (source (Wilkinson et al., 2014) fig.5)

Dans la Figure 3.6, l'équilibre énergétique au sein d'un composant de la chaîne cinématique (en anglais drivetrain) est représenté et permet d'expliquer le choix des variables. On peut voir que trois

sources de pertes énergétiques sont représentées, les pertes liées aux frottements, les pertes liées à l'environnement, les pertes liées au système de refroidissement. Chaque perte est liée à une variable SCADA utilisée en entrée du modèle. La température du composant surveillé permet de suivre l'évolution de ces pertes d'énergie et donc informe sur l'état du composant. Elle est donc choisie pour construire l'indicateur de défaut.

Ainsi, la température du composant que l'on surveille est modélisée comme une fonction de ces trois variables : la puissance électrique générée, la vitesse de rotation de l'arbre et la température ambiante. L'article précise que la température est estimée en utilisant les corrélations entre les variables d'entrée et la température à estimer sur un ensemble de points défini comme sans défaut. L'algorithme utilisé est décrit dans l'équation (3.3) ci-dessous. Il consiste en une somme pondérée de fonctions de corrélation entre la température du composant que l'on cherche à surveiller et les variables décrivant le contexte.

$$T_j^i = f_j^i(y_1^i(t), y_1^i(t), \dots, y_k^i(t)) = \sum_k \frac{g_j^i(y_k^i(t))}{R} \quad (3.3)$$

où chaque fonction  $g$  est définie comme une régression linéaire de la variable à prédire  $j$  par une variable explicative  $k$ .

Avec

$$g_j^i(y_k^i(t)) = R_k^2 \cdot (m_k \cdot y_k^i + c_k) \quad (3.4)$$

et

$$R = \sum_k R_k^2 \quad 3.5$$

où

$R^2$  est le coefficient de corrélation entre l'entrée et la sortie, élevé au carré, c'est à dire le coefficient de détermination

$m_k$  est le coefficient directeur

$c_k$  est l'ordonnée à l'origine de la droite de régression obtenue entre la variable  $j$  à prédire et la variable  $k$ .

L'indicateur utilisé pour détecter un défaut est un résidu : c'est la différence entre la valeur prédite et la valeur de température actuellement mesurée par le SCADA. Une moyenne glissante est utilisée sur le résidu pour le lisser. La Figure 3.7 montre un exemple de résidu obtenu en modélisant la température d'un roulement principal, tracé en fonction du temps. Dans cette illustration, on peut voir l'utilité de la moyenne mobile qui filtre le résidu. Le résidu non filtré contient des pics qui ne sont pas des défauts, mais correspondent à des erreurs de modélisation. Ces erreurs de modélisation peuvent être liées à l'inertie du système, à des phénomènes d'hystérésis non pris en compte dans le modèle ou à des variations non expliquées par les variables d'entrée. La méthode proposée se concentre sur une augmentation du résidu filtré plutôt que du résidu instantané.

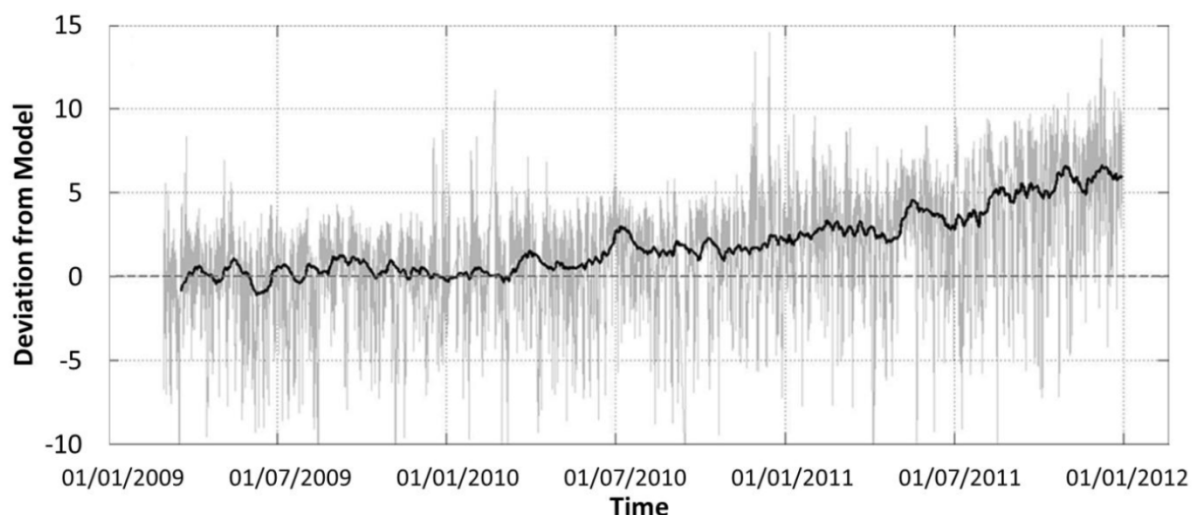


Figure 3.7 Exemple de déviation du modèle pour la température du roulement principal, en gris les valeurs brutes et en noir la moyenne mobile (source (Wilkinson et al., 2014) fig.6)

La méthode proposée permet la détection de défauts plusieurs mois à l'avance sur les composants critiques de l'éolienne. La Figure 3.8 montre la détection d'un défaut de roulement génératrice 7 mois à l'avance faite à partir du résidu obtenu par le modèle décrit dans le Tableau 3.1. Pendant ces 7 mois, une surchauffe est survenue sur le roulement qui, vraisemblablement, n'a pas été détectée par le SCADA, puisque celui-ci aurait mis la machine à l'arrêt.

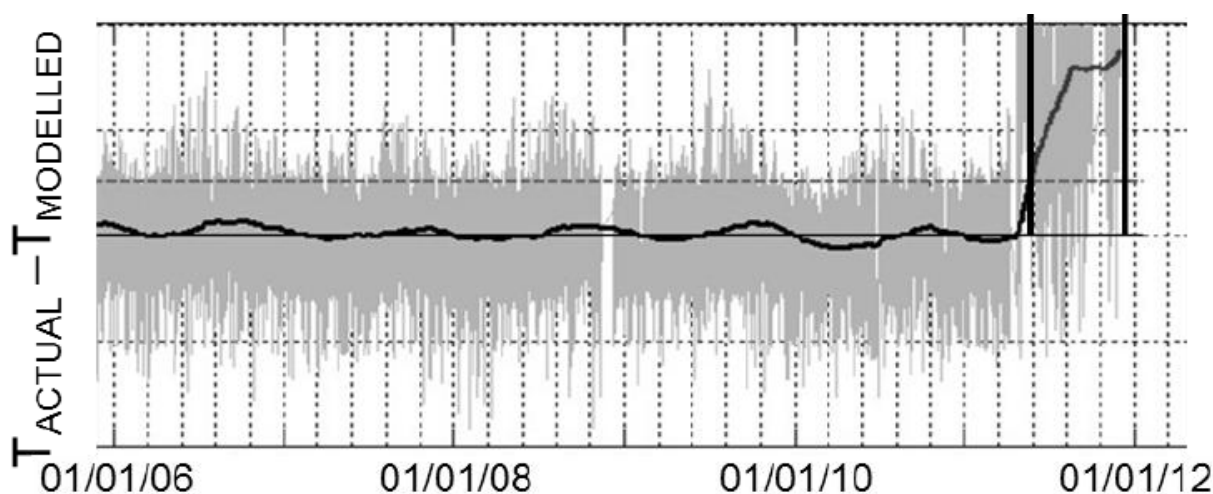


Figure 3.8 Évolution du résidu obtenu avec le modèle du tableau 2 en fonction du temps, détection d'un défaut de surchauffe d'un roulement 7 mois avant l'arrêt total (source (Wilkinson et al., 2014)fig.7)

Tableau 3.1 Description des entrées et sortie d'un modèle de prédiction de la température roulement avant de génératrice (source (Wilkinson et al., 2014) fig.7)

Sortie du modèle $y_j^i(t)$	Entrée du modèle $y_k^i(t)$
Température roulement avant de génératrice (t)	Puissance active (t)
	Température nacelle (t)
	Vitesse arbre rapide (t)

L'article (Wilkinson et al., 2014) montre comment créer un modèle pour les composants de la chaîne cinématique. Ce modèle permet de détecter une évolution de la température moyenne qui est supposée liée à un défaut. Cependant, le résidu du modèle varie entre  $\pm 10$  °C, ce qui est très élevé pour ces composants. Les températures des roulements évoluent entre 15 et 90°C, une erreur de 10°C correspond à une erreur de prédiction de plus de 10 %, ce qui n'est pas négligeable. Cela montre qu'il existe de nombreux instants de mesures où le modèle décrit mal la mesure. Cela vient probablement du fait que les modes de fonctionnement de l'éolienne ne sont pas pris en compte par cette approche, malgré l'utilisation de la puissance et de la vitesse de rotation. Le modèle suppose une relation linéaire entre les variables, ce qui n'est probablement pas vérifié sur l'ensemble de la plage de fonctionnement de l'éolienne.

Même si l'article de (Wilkinson et al., 2014) propose un filtrage par moyenne mobile pour diminuer la dispersion, la taille et la nature du filtre ne sont pas discutées alors qu'elles peuvent avoir une importance. En effet, certaines conditions de fonctionnement et conditions environnementales ne sont pas prises en compte par les variables en entrée, il est fort probable qu'elles varient d'une semaine ou d'un mois à l'autre. Enfin, on peut remarquer que le résidu, après filtrage, oscille autour de zéro, et qu'il possède une saisonnalité marquée (période proche des saisons). Chaque fois que le résidu atteint son maximum, cela correspond à une période entre juillet et août, ce qui correspond à l'été. Il est possible que la température nacelle utilisée ne soit pas suffisamment précise, soit parce que le point de mesure est trop éloigné du roulement, soit parce que l'air utilisé pour le refroidissement vient de l'extérieur de la nacelle.

---

Dans l'article de (Chen et al., 2013), le sous-système étudié est le rotor afin de détecter des défauts dans le contrôle commande ainsi que dans les moteurs orientant les pales. Le but ici est de créer des modèles de bon fonctionnement liant 2 variables entre elles. L'auteur va utiliser la vitesse du vent, la vitesse de rotation de l'arbre, l'angle des pales, la puissance ainsi le couple pour pouvoir créer de multiples modèles.

L'algorithme APK-Anfis est utilisé pour créer des modèles à partir du lien entre 4 couples de variables : vitesse du vent vs couple, vitesse du vent vs angle de pale, vitesse du vent vs vitesse de l'arbre et vitesse du vent vs puissance. La méthode permet de diviser l'espace créé par les deux variables que l'on considère en zones et ainsi permet d'avoir un modèle par zones. Cela permet d'apprendre sur des données plus semblables et de diminuer la dispersion des modèles. Les zones correspondent au découpage fait par la logique floue. L'article propose de décrire chaque variable avec 3 gammes en fonction de leurs valeurs allant ainsi de bas, moyen à haut. Il y a donc 9 zones créées lorsque l'on étudie le lien entre 2 variables (cf Figure 3.9).

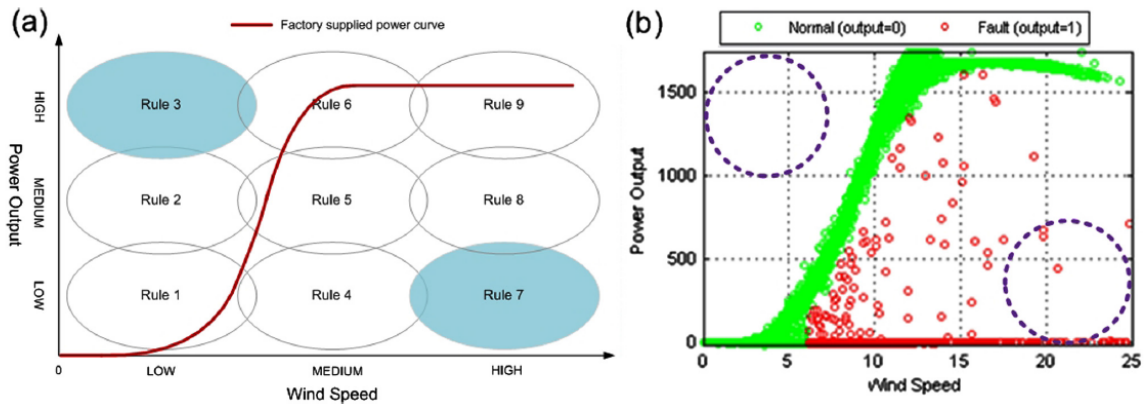


Figure 3.9 Découpage en 9 zones (a) ; données réelles pour le couple vent vs puissance (source (Chen et al., 2013) fig.2)

L'article montre que certaines zones ne sont pas suffisamment représentées par les données pour pouvoir étudier leurs liens. Ces zones numéros 3 et 7 correspondent en fait à des zones en défaut. Le module APK permet au système de prendre en compte cette connaissance a priori dans l'apprentissage du modèle.

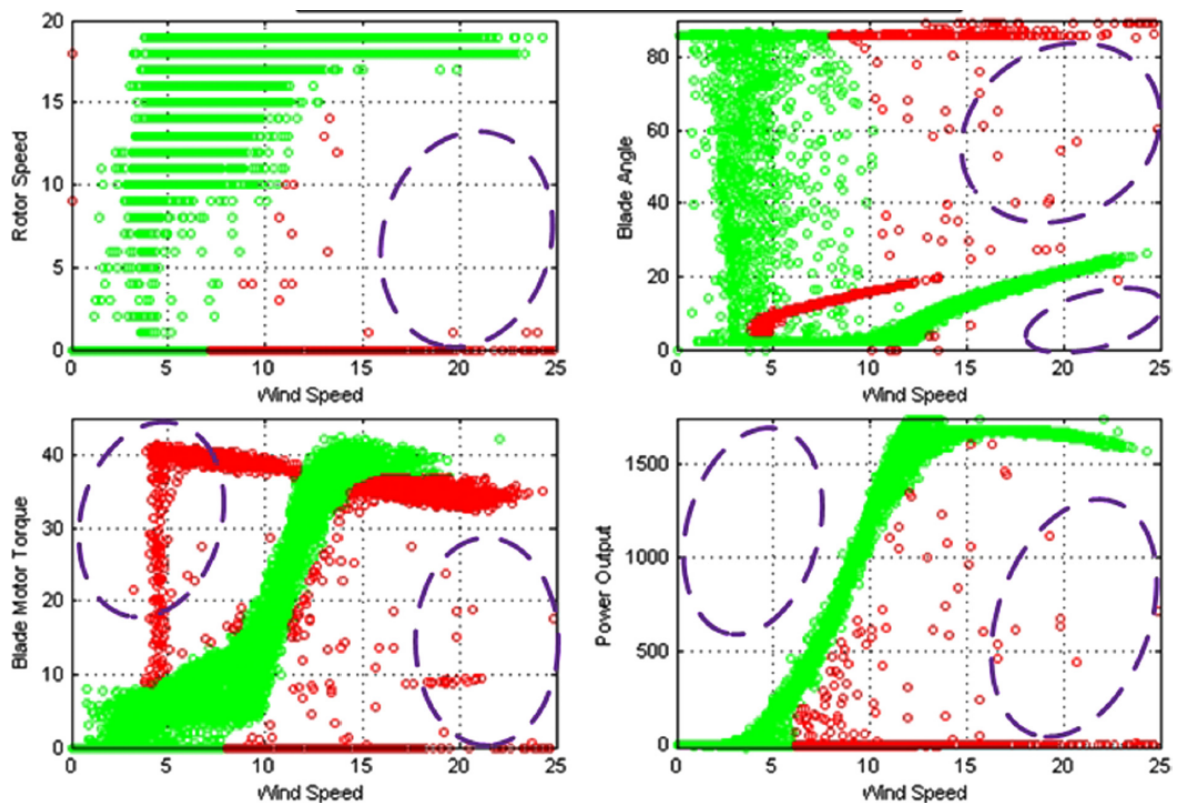


Figure 3.10 Base de données d'apprentissage de 6 défauts de pale, encadrées en violet des zones où les données sont trop peu nombreuses et où l'A Priori est nécessaire (source (Chen et al., 2013) fig.9) En vert les données de bon fonctionnement et en rouge les données représentant des défauts connus

La Figure 3.10 montre comment 6 défauts de pale apparaissent dans l'espace de données. 5 des 6 défauts sont utilisés pour la validation des données et pour choisir automatiquement le découpage en zone le plus pertinent. L'article montre que ces défauts peuvent être détectés avec 21 jours d'avance.

La méthode proposée dans (Chen et al., 2013) possède de nombreux avantages. Le premier est qu'elle permet de prendre en compte les connaissances a priori des experts grâce à des règles de logique floue. Les connaissances des experts permettent de définir des zones où il n'y a pas suffisamment de données, indiquées en violet dans la Figure 3.10. Cette méthode montre aussi comment, en utilisant plus que la puissance et le vent, on peut mieux aborder les modes de fonctionnement de la machine pour la surveillance. Certains défauts ne sont pas visibles sur la courbe de puissance, mais visibles sur une représentation vent versus angle de pale. En revanche, pour être réellement efficace, cette méthode doit prendre en compte une liste exhaustive des défauts possibles de la machine, qui doivent être présents dans la base d'apprentissage. Cette méthode est basée sur de la reconnaissance de défauts connus dans une base plutôt que sur la création de modèles de bon fonctionnement. Enfin, le découpage en zones proposé reste large, ce qui fait que les défauts doivent avoir un impact important sur le fonctionnement de la machine pour pouvoir être détectés. Ce type de défaut est déjà très avancé et l'objectif est de les détecter avant d'en arriver là.

### 3.1.2.2.3. Découpage de la chaîne cinématique en modèles dynamiques

Cette partie est consacrée à la présentation de l'utilisation de la dynamique dans les modèles. Comme vu précédemment, les modèles statiques décrivent de manière imprécise l'évolution d'une température. Utiliser un modèle dynamique permet de prendre en compte l'inertie thermique, phénomène négligé dans les modèles statiques.

L'article (Zhang, 2014) présente un exemple de modèle utilisant la température du composant surveillé 10 minutes auparavant. Ici, le sous-système étudié est le rotor, plus particulièrement le roulement principal de l'éolienne. On cherche à détecter des défauts d'usure prématurée entraînant des échauffements dans le roulement. Les entrées et sorties du modèle sont listées dans la table 3. Elles sont identiques au modèle présenté dans (Wilkinson et al., 2014), l'innovation vient de l'introduction de la température du roulement à l'instant précédent.

Tableau 3.2 Description du modèle dynamique pour la température du roulement principale (source (Zhang, 2014))

Sortie du modèle $y_j^i(t)$	Entrée du modèle $y_j^i(t - lag_k)$
Température roulement (t)	Température roulement (t-1)
	Puissance active (t)
	Température nacelle (t)
	Vitesse arbre rapide (t)

La température du roulement est modélisée via deux modèles distincts : un réseau de neurones multicouche, avec une seule couche cachée de 10 neurones et une régression linéaire (équation (3.6)) qui sont développés séparément et ensuite comparé appelé respectivement le modèle ANN et le modèle mathématique.

$$\hat{T}_{roulement}^i(t) = a.T_{roulement}^i(t-1) + b.n_{rapide}^i(t) + c.P_{génératrice}^i(t) + d.T_{nacelle}^i(t) \quad (3.6)$$

où

$\hat{T}_{roulement}^i(t)$  est la température du roulement prédite à l'instant  $t$

$T_{roulement}^i(t-1)$  est la température du roulement que l'on cherche à prédire à l'instant  $t-1$

$n_{rapide}^i(t)$  est la vitesse de l'arbre rapide à l'instant  $t$

$P_{génératrice}^i(t)$  est la puissance produite par la génératrice à l'instant  $t$

$T_{nacelle}^i(t)$  est la température de la nacelle à l'instant  $t$

et  $a, b, c$  et  $d$  sont des paramètres appris à partir des données.

Les paramètres de l'équation (3.6) ou les poids du réseau de neurones sont appris en utilisant une base de données pour l'apprentissage de 3 mois pendant laquelle aucun défaut n'est présent. L'article propose d'utiliser un réseau de neurones qui doit permettre de prendre en compte les non-linéarités dues aux modes de fonctionnement. Les réseaux de neurones permettent aussi une régression d'ordre supérieur comparativement à l'autre méthode utilisant l'équation (3.6). En théorie, le réseau de neurones devrait obtenir de meilleurs résultats. Les deux modèles obtiennent cependant des résultats semblables. L'erreur quadratique moyenne entre la température prédite et la température réelle est de  $0.24^{\circ}\text{C}$  pour le réseau de neurones et de  $0.78^{\circ}\text{C}$  pour la régression linéaire.

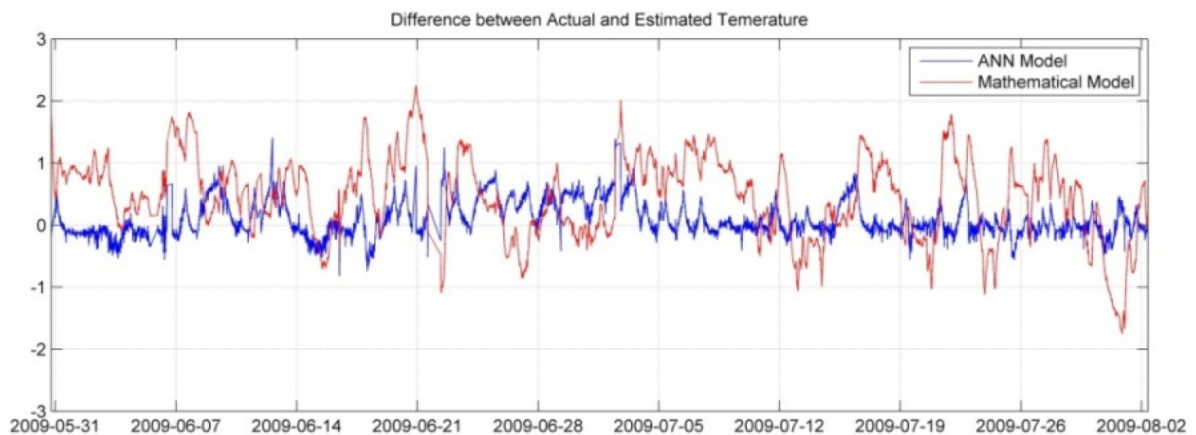


Figure 3.11 Résidu des deux approches sur une période sans défaut (source (Zhang, 2014) fig.5)



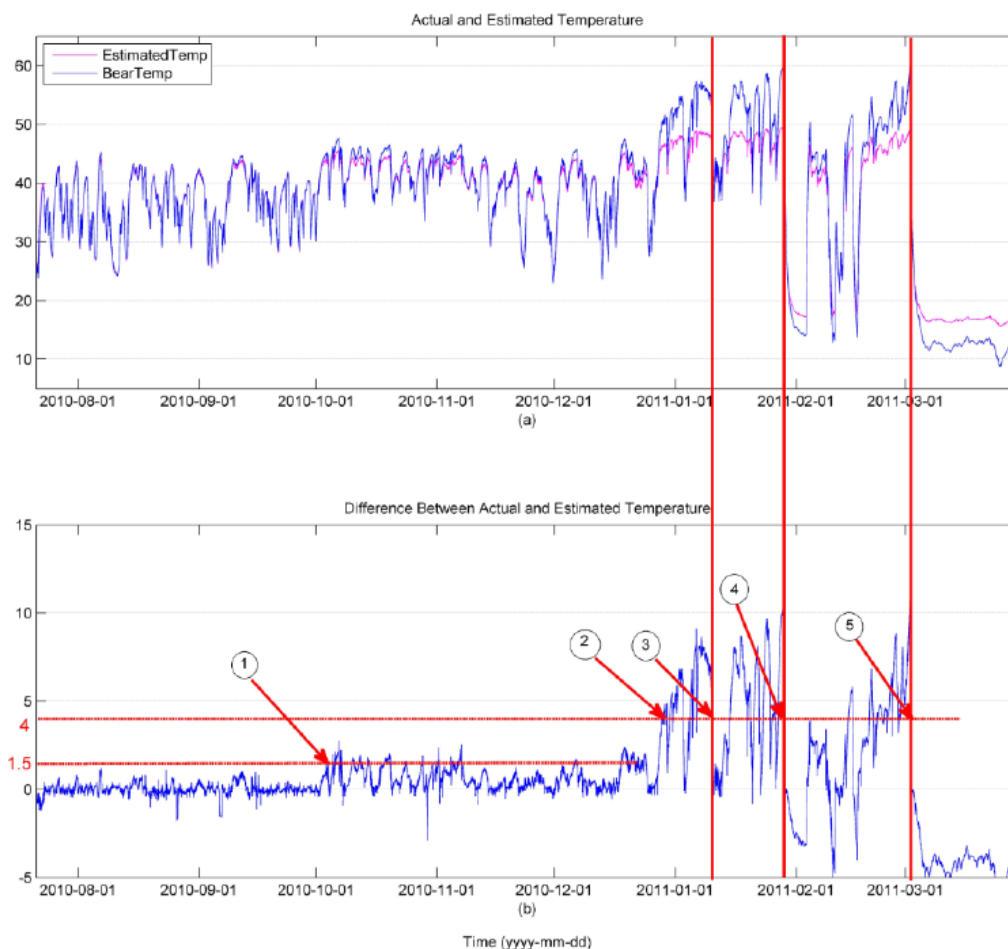


Figure 3.12 Exemple de détection de défaut de roulement principal avec l'approche utilisant un réseau de neurones (source (Zhang, 2014) fig.6)

Sur la Figure 3.11, on peut voir le résidu obtenu par les deux méthodes sur une période sans défaut. Grâce à cette méthode, un défaut avéré du roulement principal est détecté plusieurs mois avant que la turbine ne soit stoppée complètement. La Figure 3.12 permet de voir l'apparition de la surchauffe au travers de l'évolution du résidu obtenu par réseau de neurones (la méthode par régression n'est pas représentée, mais l'évolution du résidu est sensiblement la même). En 1, on peut voir une augmentation de l'écart type du résidu, jusqu'à ce que la température moyenne augmente en 2. En 3 et 4, les opérateurs ont tenté de réparer la machine. En 5, la machine a dû être stoppée totalement et le roulement sera remplacé par la suite. La Figure 3.12 montre comment le résidu peut permettre de détecter un défaut à l'avance de manière automatique. Les résidus obtenus par les deux modèles sont sensiblement identiques. En réduisant la taille de la base d'apprentissage de 3 mois à 10 jours, l'article met en évidence que l'écart type du résidu du réseau de neurones augmente lorsque la base de données diminue trop alors que la régression linéaire garde des résultats semblables. L'article conclut que la régression linéaire est moins dépendante de la taille de la base d'apprentissage.

Dans cet article, (Zhang, 2014) propose une première approche de l'utilisation de la dynamique en utilisant la température à t-10 minutes, il propose de comparer les performances d'un réseau de neurones et d'une régression linéaire simple. L'article conclut que les deux méthodes ont des résultats similaires et que, en utilisant une équation linéaire, on obtient des résultats exploitables pour la

détection, même avec quelques jours de données pour l'apprentissage. De plus, on peut voir sur la Figure 3.11 que l'erreur est de  $\pm 2^{\circ}\text{C}$ , ce qui paraît faible par rapport aux températures de fonctionnement des roulements de  $40^{\circ}\text{C}$ . Cette faible erreur est probablement due principalement au fait que les modèles ont eu tendance à recopier la température à  $t-10$  pour la prédiction, la température évoluant au maximum de quelques degrés d'une période d'échantillonnage à l'autre. Malgré tout, la détection n'a pas été impactée, car un roulement défaillant génère en général une forte augmentation de la vitesse de montée en température. Ici, c'est une vitesse de chauffe anormale qui est détectée, plus qu'une surchauffe entre la température prédite et la température réelle.

#### **3.1.2.2.4. Découpage de la chaîne cinématique et fusion de modèles**

Dans cette partie, deux articles ayant la particularité d'utiliser plusieurs modèles d'un même sous-système en même temps seront présentés. Contrairement aux articles de la partie 3.1.2.2.2 qui proposent une méthode pour modéliser les composants des sous-systèmes, ces deux articles proposent, en plus, de combiner les résidus de chacun des modèles pour établir un premier diagnostic. Dans la chaîne cinématique, le rotor n'est souvent surveillé que par l'intermédiaire du roulement principal. Il n'a qu'un modèle associé, mais la génératrice et la multiplicatrice peuvent en posséder plusieurs. La multiplicatrice, par exemple, possède plusieurs roulements et un circuit de refroidissement. La température de ces composants est souvent présente dans le SCADA, on peut donc créer un modèle pour chacun d'eux. En multipliant les modèles pour un même sous-système, il est possible de localiser encore plus précisément le défaut. Par exemple, si un roulement est défaillant, alors lui seul verra sa température monter, mais, si c'est le système de refroidissement qui est défaillant, alors tous les modèles de la multiplicatrice verront leurs résidus augmenter. Les deux articles proposent à la fois une méthode de modélisation des composants et une méthode pour interpréter automatiquement les résidus obtenus. La différence entre les deux articles vient des sous-systèmes et des composants étudiés et de la manière d'interpréter les résultats.

L'article (Garcia et al., 2006) propose de surveiller la multiplicatrice en la découpant en 3 modèles indépendants : un modèle pour les surchauffes d'un roulement, un pour la surchauffe au niveau de l'huile et un pour surveiller le circuit de refroidissement. Ce découpage est basé sur les variables SCADA disponibles. Plus le nombre de capteurs de température est important, plus le nombre de modèles possibles est important. Les modèles créés utilisent en entrée la puissance, la vitesse de rotation, la température nacelle et l'information de marche et d'arrêt du système de refroidissement comme décrit dans le Tableau 3.3.

Tableau 3.3 Description des 3 modèles permettant la surveillance de la multiplicatrice (source (Garcia et al., 2006))

Sortie du modèle $y_j^i(t)$	Entrée du modèle $y_j^i(t - lag_k)$
Température du roulement multiplicatrice (t)	Température du roulement multiplicatrice (t-1)
	Température du roulement multiplicatrice (t-2)
	Puissance active (t-3)
	Température nacelle (t)
	Vitesse arbre rapide (t-2)
	Status vitesse basse du ventilateur (t-2)
	Status vitesse haute du ventilateur (t-2)
Différence de température dans le circuit de refroidissement multiplicatrice (t)	Différence de température dans le circuit de refroidissement multiplicatrice (t -1)
	Puissance active (t-2)
	Température nacelle (t)
	Vitesse arbre rapide (t-2)
	Status vitesse basse du ventilateur (t-2)
Température de l'huile de refroidissement multiplicatrice (t)	Status vitesse haute du ventilateur (t-2)
	Température de l'huile de refroidissement multiplicatrice (t-1)
	Puissance active (t-2)
	Température nacelle (t)
	Vitesse arbre rapide (t-2)

À l'instar de l'article (Wilkinson et al., 2014) de la partie 3.1.2.2.2, les modèles présentés par (Garcia et al., 2006) prennent en compte l'évolution dynamique de la température. L'intervalle de temps pris en compte pour prédire la température est élargi, car les températures à l'instant  $t - 1$  et  $t - 2$ , soit 20 minutes avant l'instant de prédiction, sont utilisées pour prédire la température à l'instant  $t$ . Un retard est aussi introduit dans les autres variables. Plutôt que de prendre la puissance à l'instant de la prédiction, on prend la puissance à  $t - 3$  ce qui correspond à la puissance prise 30 minutes avant. La valeur du retard a été choisie en étudiant les corrélations entre les variables d'entrée et la variable de sortie. Comme précédemment, c'est un réseau de neurones qui est utilisé pour prédire les différentes températures du sous-système. La taille de la base d'apprentissage n'est cependant pas précisée.

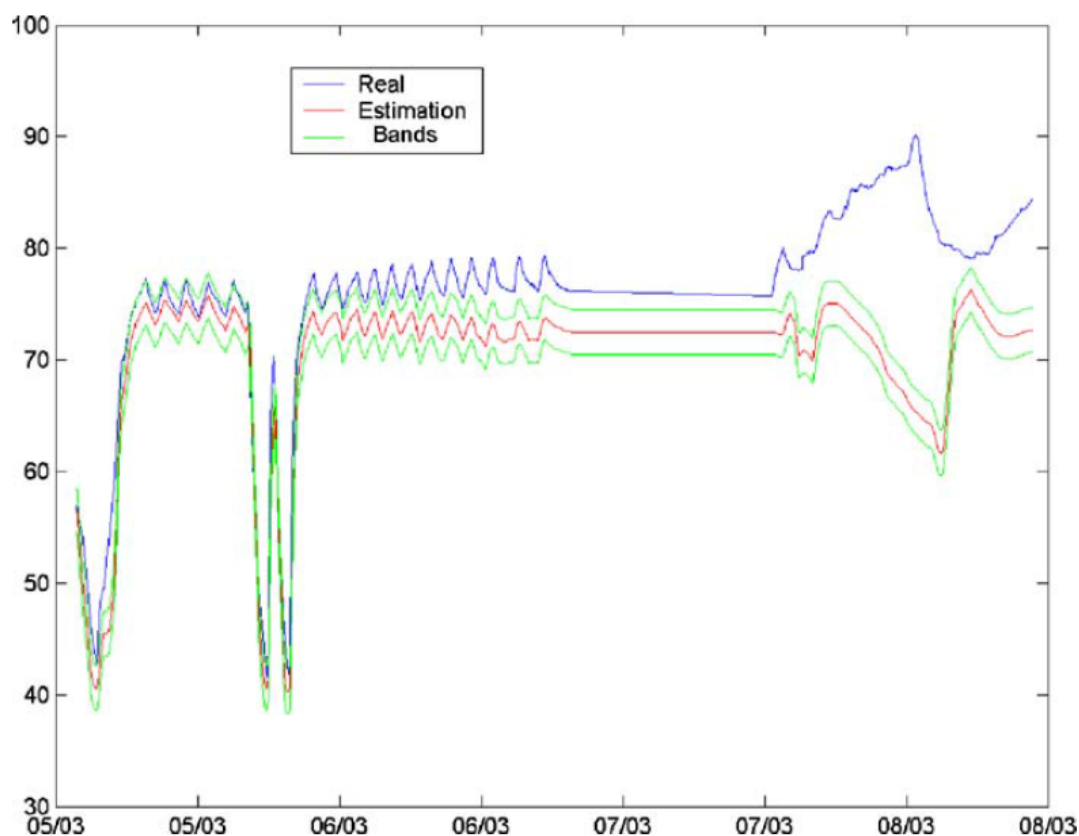


Figure 3.13 Valeur de la température du roulement de la multiplicatrice lors de l'apparition d'un défaut (source (Garcia et al., 2006) fig.8)

La Figure 3.13 présente l'évolution de la température réelle d'un des roulements de la multiplicatrice par rapport à la température prédite pendant un défaut de la multiplicatrice. En vert est représentée une zone de confiance de 95 % autour de la température prédite. Cet intervalle représente probablement la valeur prédite  $\pm 2$  écarts type calculés sur une période sans défaut qui a permis la validation, cela n'est pas complètement décrit dans l'article. On peut voir qu'il existe deux phases dans l'apparition du défaut : d'abord, une augmentation progressive de la température le premier jour, suivi par un dérèglement catastrophique de la machine. Notons que seul l'écart au modèle instantané est étudié et non sa dynamique. L'allure du résidu pourrait informer sur la criticité du défaut.

Les auteurs utilisent plusieurs modèles pour prédire l'évolution de différentes températures de la multiplicatrice. Ils proposent une méthode automatisée pour interpréter les informations contenues dans les différents résidus à l'aide d'un ensemble de règles issues de la connaissance des hommes de métier. Ces règles sont décrites par logique floue. Un exemple de règle est présenté dans la Figure 3.14.

<p><b>Rule name</b> <i>r_m1</i></p> <p><b>IF</b></p> <p>Gearbox main bearing temperature is HIGH <b>AND</b> Gearbox thermal difference is HIGH <b>AND</b> Cooling oil temperature is NORMAL</p> <p><b>THEN</b></p> <p><b>Diagnostic MF1:</b> Failure in the gearbox main bearing is <b>certain</b> <b>Diagnostic MF2:</b> Unbalance in the gearbox main shaft is <b>quite certain</b></p> <p><i>fuzzy implication rule: ra</i></p>
<p><b>Rule name</b> <i>r_m2</i></p> <p><b>IF</b></p> <p>Cooling oil temperature of the gearbox is HIGH <b>AND</b></p> <p><b>THEN</b></p> <p><b>Diagnostic MF3:</b> Failure in the gearbox cooling circuit is <b>certain</b></p> <p><i>fuzzy implication rule: ra</i></p>

Figure 3.14 Exemple de règles pour SIMAP (source (Garcia et al., 2006))

Les résidus permettent de détecter la présence d'un défaut, la base de règle permet sa localisation et son isolation. Ainsi l'article propose une méthode complètement automatique appelée SIMAP (Intelligent System for Predictive Maintenance Application) pour la surveillance de la multiplicatrice. Cet article montre comment interpréter les résultats de plusieurs modèles de manière automatique. Les variables SCADA sont nombreuses et la modélisation de chacune des températures augmente encore le nombre d'informations à suivre par les opérateurs. Ici, l'article montre comment les sorties des modèles peuvent être synthétisées dans une information claire et rapidement interprétable par les opérateurs, afin de permettre un premier diagnostic en plus de la localisation du défaut.

---

Le deuxième article (Zaher et al., 2009) propose une approche similaire, mais ne se restreint pas à la multiplicatrice. L'auteur montre comment la méthode de (Garcia et al., 2006) peut être généralisée. La génératrice et la courbe de puissance sont surveillées en plus de la multiplicatrice. La génératrice est surveillée par l'intermédiaire de la température de ses enroulements. Le modèle utilise la température des enroulements en sortie et en entrée la puissance, la vitesse de rotation, la température nacelle et la température passée des enroulements. Le modèle pour la courbe de puissance n'est pas décrit.

Tableau 3.4 Description des modèles dynamiques simplifié (source (Zaher et al., 2009))

Sortie du modèle $y_j^i(t)$	Entrée du modèle $y_j^i(t - lag_k)$
Température du roulement multiplicatrice (t)	Température du roulement multiplicatrice (t-1)
	Température du roulement multiplicatrice (t-2)
	Puissance active (t-3)
	Température nacelle (t)
Température circuit de refroidissement multiplicatrice (t)	Température circuit de refroidissement multiplicatrice (t-1)
	Température circuit de refroidissement multiplicatrice (t-2)
	Puissance active (t-2)
	Température nacelle (t)
Température enroulement stator (t)	Température enroulement stator (t-1)
	Température enroulement stator (t-2)
	Puissance active (t-3)
	Température nacelle (t)

L'article utilise donc les mêmes modèles que (Garcia et al., 2006), mais cette fois sans utiliser l'information sur la vitesse du ventilateur comme décrit dans le Tableau 3.4. L'information sur la différence de température dans le circuit de refroidissement n'est pas présente. L'article ne modélise que deux températures pour le sous-système multiplicatrice.

Afin de vérifier la pertinence de chacune des variables d'entrée choisies dans (Garcia et al., 2006), l'article (Zaher et al., 2009) présente des résultats obtenus sur la même période de validation en omettant certaines variables en entrée. Au final, toutes les variables mises en évidence avec SIMAP sont utiles et sont gardées. En effet, l'erreur quadratique est minimale quand toutes les variables sont utilisées. La dynamique permet donc de diminuer la dispersion du résidu dans ce cas. L'article propose 4 modèles : 1 pour le système et 3 sous-systèmes. Les trois sous-systèmes possèdent le même type d'entrée, c'est simplement la température que l'on cherche à prédire qui va varier. Contrairement à SIMAP, ici aucune méthode automatique n'est proposée pour interpréter les résidus des modèles. Les résidus sont présentés simplement à l'expert qui décidera de la marche à suivre. Un modèle de courbe de puissance est utilisé, car il permet d'avoir une information supplémentaire sur la criticité du défaut. Les défauts sont détectés par sous-systèmes et ensuite la perte d'énergie associée à ce défaut est donnée par le modèle de la courbe de puissance. Plus la baisse de performance est importante, plus le défaut est considéré comme avancé.

L'article (Zaher et al., 2009) permet de voir que les modèles proposés dans SIMAP peuvent être transposés à d'autres composants de la chaîne cinématique comme la génératrice. On peut supposer que cela est aussi possible pour le roulement principal, mais cela n'est pas montré dans l'article. Même si les modèles proposés n'utilisent pas la vitesse ventilateur, car non présente, il est précisé que celle-ci permettrait d'améliorer les résultats.

### 3.1.2.2.5. Autres systèmes que la chaîne cinématique

On a pu voir que des modèles peuvent être créés pour tous les composants de la chaîne cinématique. Les articles qui vont suivre présentent des méthodes pour surveiller les autres composants de l'éolienne. Ils suivent la même logique, c'est-à-dire développer des modèles représentant des

relations de causes à effets entre variables et générer des résidus mesurant la différence entre la grandeur physique mesurée et la même grandeur prédite par le modèle.

Les articles (Schlechtingen et al., 2013) et (Schlechtingen and Santos, 2014) proposent d'étudier les liens causaux existant entre un ensemble de variables SCADA mesurées sur l'ensemble de l'éolienne. Pour cela, ils utilisent un algorithme génétique, qui n'est pas décrit entièrement dans leurs articles, capable de reconnaître ces liens. Certaines variables sont identifiées comme des entrées, d'autres comme des sorties. La Figure 3.15 montre les variables sélectionnées et les liens qui ont été trouvés. Ceux-ci sont représentés par des traits bleus. Les auteurs précisent que l'algorithme sélectionne en priorité des entrées qui possèdent des liens de causes à effets avec la sortie. Dans la suite, chaque lien est traduit sous la forme d'un modèle mathématique. Notons que l'algorithme a parfois relié des variables mesurant des grandeurs identiques. Ces liens génèrent des modèles indiqués « cross » qui possèdent une particularité qui n'est pas mise en évidence dans cet article, mais qui sera décrite dans la partie 3.1.3 suivante.

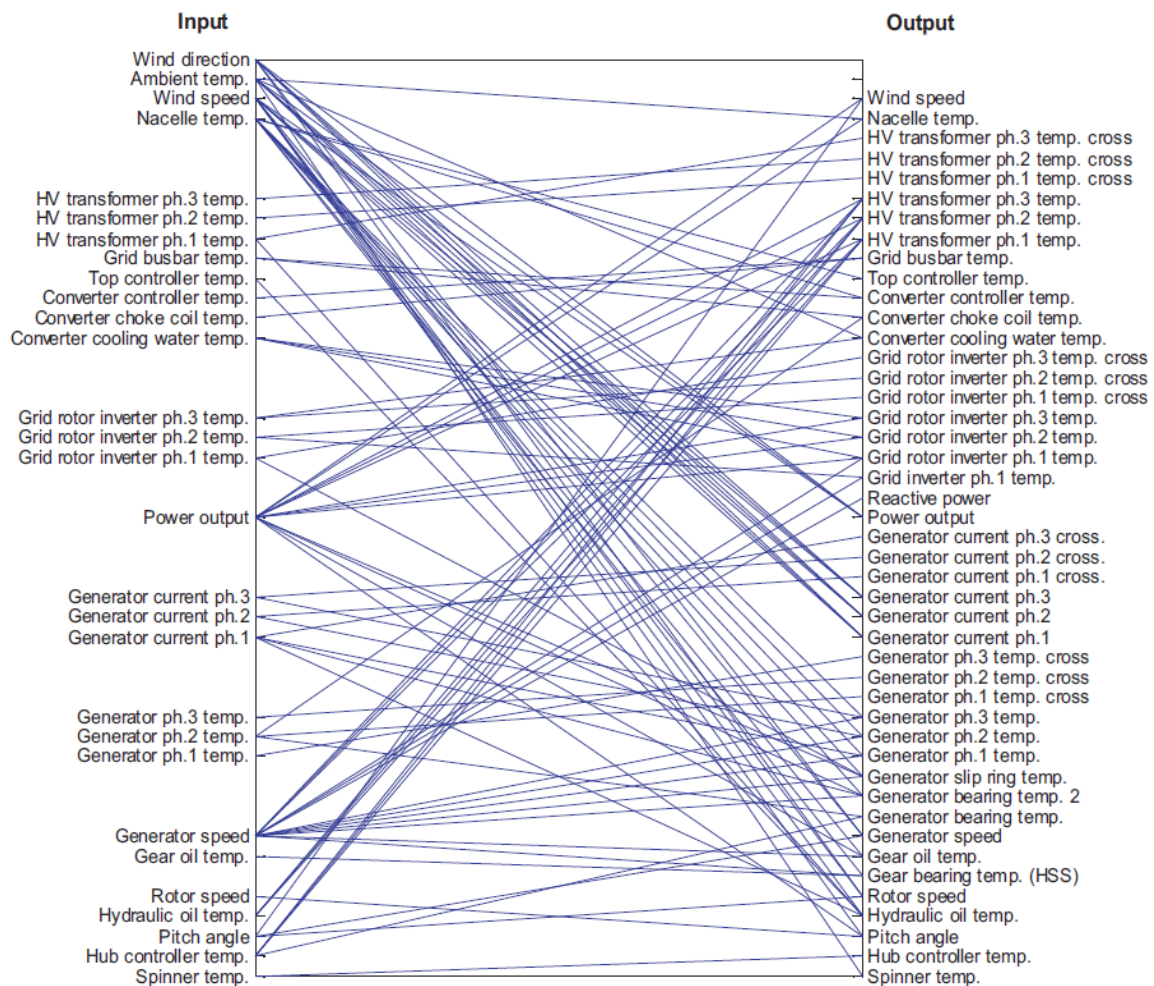


Figure 3.15 Sélection des entrées et sorties des modèles (source (Schlechtingen et al., 2013) fig.6)

Un réseau de neurones est créé pour modéliser chacun des liens mis en évidence dans la Figure 3.15, à partir de données enregistrées en bon fonctionnement. Un total de 45 modèles est créé. Des modèles décrivant le fonctionnement de sous-systèmes de la chaîne cinématique ont pu être créés,

mais aussi des modèles pour d'autres sous-systèmes. Le convertisseur, le transformateur, la mesure de vent et de température se voient maintenant attribuer un modèle de bon comportement. On obtient un nombre de résidus important et l'article propose une manière automatique de les synthétiser et de les interpréter. Un module utilisant de la logique floue permet de générer des règles pour interpréter les nombreux résidus obtenus. Ces règles sont générées automatiquement à partir d'une base de données sans défauts, mais peuvent aussi être définies par les experts. L'article propose à l'utilisateur une interface simple pour visualiser l'évolution des indicateurs ainsi qu'un diagnostic sur les défauts détectés. Pour détecter les défauts, des seuils autorisant 1% de fausses alarmes sont créés à partir de la distribution des résidus. L'article revient sur l'intérêt d'un filtre moyen sur les résidus obtenus. Les résidus sont filtrés par une moyenne glissante de 1 jour, permettant de diminuer les fausses alarmes. Malheureusement, l'article ne montre pas l'effet de la taille de la moyenne mobile sur les performances en détection.

La force de ces articles, (Schlechtingen et al., 2013) et (Schlechtingen and Santos, 2014), ne tient pas dans les performances individuelles de chacun des modèles. Cet article montre surtout qu'il est possible de détecter des défauts sur d'autres parties que la chaîne cinématique. Un exemple de détection d'un défaut de température convertisseur est présenté. Le convertisseur est resté sans système de refroidissement pendant plusieurs mois sans que cela n'ait été pris en compte par les opérateurs. Le SCADA n'avait pas détecté la défaillance, alors que la méthode proposée détecte ce problème. Même si cela ne correspond pas à un défaut majeur, ce genre de problème de refroidissement peut diminuer la durée de vie du convertisseur. Être capable de détecter ce genre de surchauffe peut éviter qu'un problème de convertisseur apparaisse et s'aggrave. De plus, l'article montre que l'on peut localiser précisément les différents défauts détectés grâce aux SCADA modernes. Un premier diagnostic peut même être fait grâce à des règles établies automatiquement ou par les experts.

### **3.1.2.3. Conclusion sur les méthodes utilisant les liens de causalité**

Les méthodes présentées proposent de relier des variables SCADA à l'aide d'un modèle mathématique. Les modèles décrivent des relations de cause à effet, qui correspondent à des lois de la physique. Le modèle le plus général est celui décrivant la relation entre la vitesse du vent et la puissance électrique générée. On a vu que ces modèles basés sur la courbe de puissance ne permettent de détecter que des défauts avancés. Comme la courbe de puissance englobe tout le fonctionnement de la machine, ces modèles ne permettent pas de localiser les défauts. Pour ce faire, une solution est de découper l'éolienne en sous-systèmes de taille plus petite. Plus le nombre de variables SCADA est important, plus l'éolienne pourra être découpée finement en sous-systèmes, améliorant la détection et le diagnostic. Plus le découpage est fin, plus la détection est précise, car les composants sont isolés les uns des autres. Ces méthodes ne sont pas encore généralisées sur le marché de l'éolien, mais pourraient déjà, à elles seules, permettre d'améliorer grandement la maintenance.

Très souvent, les modèles utilisés, en dehors de la courbe de puissance, ne prennent pas en compte l'environnement c'est-à-dire les conditions météorologiques. Ils ne prennent pas non plus en compte l'ensemble des modes de fonctionnement. En effet, une éolienne est une machine complexe, évoluant rarement en régime permanent et alternant fréquemment entre arrêt et redémarrage. Elle fonctionne dans des conditions environnementales très variables suivant les saisons. Son comportement est fortement non linéaire et non stationnaire. Les modèles proposés, souvent linéaires et restreints à quelques variables, ne peuvent donc pas décrire très précisément l'évolution des différentes variables



et on observe donc une forte dispersion des résidus. La réduction de la dispersion des résidus nécessiterait d'intégrer un grand nombre de variables dans le modèle qui ne sont pas nécessairement disponibles dans le SCADA. De plus, les modèles sont appris sur un jeu de données où des conditions environnementales particulières sont présentes lors de l'apprentissage (par exemple l'apprentissage est réalisé pendant les mois d'été). L'environnement pouvant varier au cours du temps, le résidu peut augmenter, car le modèle n'est plus adapté aux nouvelles conditions environnementales. Les méthodes décrites dans ce paragraphe font l'hypothèse que si la plage d'apprentissage est suffisamment longue alors toutes les conditions environnementales possibles seront apprises. Mais cela est en fait impossible. En prenant l'exemple du vent seul, on peut voir que le gisement en vent diffère d'une année sur l'autre. Il faudrait donc au minimum plusieurs années pour prendre en compte ce phénomène. De manière plus générale, on peut voir que pour pouvoir réellement prendre en compte les conditions environnementales et intégrer leur influence dans le modèle, il faudrait avoir accès à des informations sur la densité de l'air, la pression atmosphérique et sur la nature du front d'air. Or, toutes ces informations ne sont pas mesurées. Il y aura donc toujours une perturbation minimum qui va augmenter la dispersion et parfois même engendrer des fausses alarmes.

Il est possible d'améliorer les modèles en tenant compte de l'inertie des composants ou en utilisant des techniques d'apprentissage plus performantes. Les structures des réseaux de neurones peuvent toujours être optimisées, mais il y restera quand même des phénomènes physiques non pris en compte qui perturberont la détection. La partie suivante est dédiée à des méthodes qui permettent de s'affranchir de certaines variations de l'environnement pour diminuer encore la dispersion des modèles.

### **3.1.3. Méthodes utilisant la similarité entre composants**

Les variables SCADA ne mesurent pas toutes les variations de l'environnement et parfois ne mesurent pas suffisamment de grandeurs liées aux modes de fonctionnement. Ce manque d'information en entrée des modèles engendre une forte dispersion des résidus, ce qui augmente le retard à la détection d'un défaut. Le but des méthodes présentées par la suite est d'aller chercher autrement l'information contextuelle non mesurée. Certains chercheurs ont remarqué que des composants au sein d'une éolienne se comportent de manière similaire. Plutôt que de créer un modèle qui va prédire une variable à partir des grandeurs physiques liées aux modes de fonctionnement ou aux conditions environnementales, ces nouvelles approches vont prédire la grandeur physique mesurée sur un composant à partir d'une autre variable qui n'a aucun lien de cause à effet. L'idée est de créer un modèle qui permet de vérifier que les deux variables évoluent ensemble.

L'hypothèse utilisée par ces méthodes est que les composants synchronisés sont soumis aux mêmes variations, qu'elles soient générées par des changements de modes de fonctionnement ou par l'environnement. Par conséquent, ces changements, qui sont parfois non mesurables, seront pris en compte directement dans la variable associée au composant de référence. Toute la difficulté réside dans le fait de trouver un composant suffisamment « semblable » à un autre pour que l'on puisse prédire le comportement de l'un en connaissant l'autre.

Ici on cherche à comparer des composants entre eux en les intégrant dans la modélisation.

$$y_j^i(t) = g_{j,ref}^i(y_{ref}^i(t)) \quad (3.7)$$

$$y_j^i(t) = h_{j,ref}^i(y_{ref}^i, y_1^i(t), y_2^i(t), \dots, y_k^i(t)) \quad (3.8)$$

où  $y_{ref}^i$  correspond à une variable mesurée sur un composant similaire au composant  $j$ .

Dans l'équation (3.7), la variable est modélisée uniquement avec la variable de référence. Dans l'équation (3.8), la variable est modélisée en tenant compte du contexte, c'est-à-dire en intégrant des variables décrivant les modes de fonctionnement ou l'environnement, car la similarité entre les deux variables peut être influencée par le contexte. Plusieurs méthodes sont proposées pour tirer parti de la similarité entre les composants, il est proposé dans la littérature de comparer visuellement les modèles entre eux ou alors de calculer l'écart la prédiction du modèle avec la valeur réelle mesurée.

### 3.1.3.1. Modèle liant la température du frein hydraulique et celle du roulement principal

(Butler, 2012) présente un modèle de régression pour prédire la température du roulement principal de l'éolienne. Le modèle utilisé est représenté par l'équation suivante :

$$\hat{T}_{roulement}^i(t) = h_{roulement, frein}^i(T_{frein}^i(t), Pr_{frein}^i(t), \theta_{pale}^i(t), n_{lent}^i(t)) \quad (3.9)$$

où

$T_{roulement}^i(t)$  représente la température du roulement à prédire

$T_{frein}^i(t)$  représente la température du frein hydraulique qui sert ici de référence  $y_{ref}^i(t)$

$Pr_{frein}^i(t)$  la pression du frein hydraulique,  $\theta_{pale}^i(t)$  l'angle de pale et  $n_{lent}^i(t)$  la vitesse de l'arbre lent qui représentent les variables  $y_k^i(t)$  décrivant le contexte des conditions de fonctionnement

La vitesse de l'arbre, la position de l'angle de pale et la pression du frein permettent d'avoir les informations contextuelles habituelles sur les modes de fonctionnement. Ici, la pression du frein hydraulique est utilisée car l'auteur explique qu'il est possible que le frein soit testé et que, lors de ces tests, il exerce une pression qui augmente l'échauffement. En prenant en compte la pression, cela permet de prendre en compte les tests. L'angle de pale permet d'avoir une information liée à la puissance. La température ambiante ou extérieure n'est pas présente ici car elle est utilisée en amont de la modélisation. L'auteur montre que la relation entre la température extérieure et la température du roulement est linéaire et choisit de soustraire celle-ci à la température du roulement à prédire. L'auteur explique que cela permet de simplifier le modèle en supprimant une entrée.

La température du frein hydraulique est une nouvelle variable utilisée. L'auteur explique que cette température est très fortement corrélée avec le roulement principal. Dans la Figure 3.16, on peut voir que le frein hydraulique est placé loin du roulement principal. Le frein n'a aucune influence sur celui-ci. La température du frein a pourtant un comportement similaire à la température du roulement pour 2 raisons. Tout d'abord, il s'agit d'une température, les deux capteurs mesurent une grandeur physique de même nature. Deuxièmement, l'évolution de ces températures dépend des mêmes conditions externes : la température extérieure, la vitesse de rotation et la puissance.

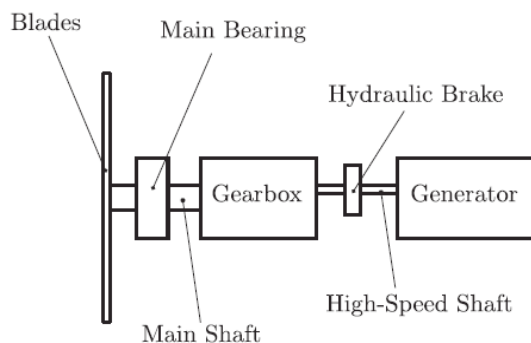


Figure 3.16 Organisation de la chaîne cinématique, localisation des capteurs (source (Butler, 2012) fig.7.3)

Afin d'apprendre ce modèle, l'auteur utilise un SBL « Sparse Bayesian Learning » pour faire de la régression entre les variables. Une fois le modèle appris, celui-ci est testé sur un exemple de roulement défectueux. Dans l'article, il est précisé que le résidu ainsi obtenu dépend encore des modes de fonctionnement. La Figure 3.17 montre que la distribution de la température du roulement principal dépend de la vitesse de rotation de l'arbre. L'auteur suggère aussi que, en fonction du vent et de la vitesse de rotation, la manière dont l'échauffement est généré dans le roulement est modifiée. Au final, pour remédier à ce problème il est proposé de trier le résidu en trois classes de vitesses de rotation : « low » « medium » et « high ». Il est surprenant de constater que, même en intégrant la vitesse de rotation en entrée du modèle, le résidu soit encore dépendant de celle-ci.

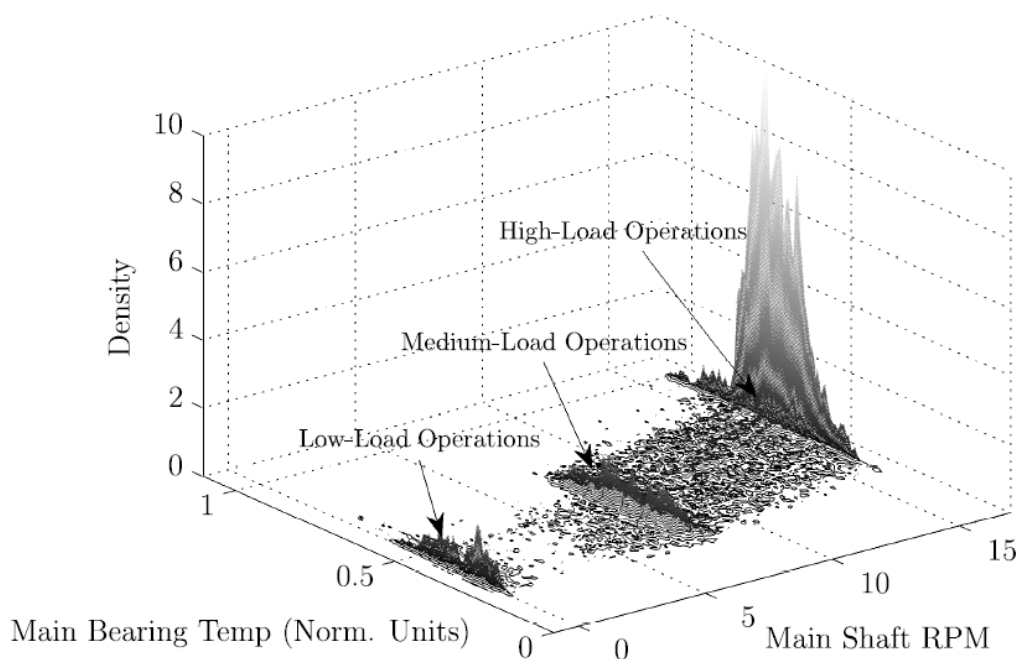


Figure 3.17 Distribution de la température du roulement en fonction de la vitesse de rotation de l'arbre (source (Butler, 2012) fig 7.12)

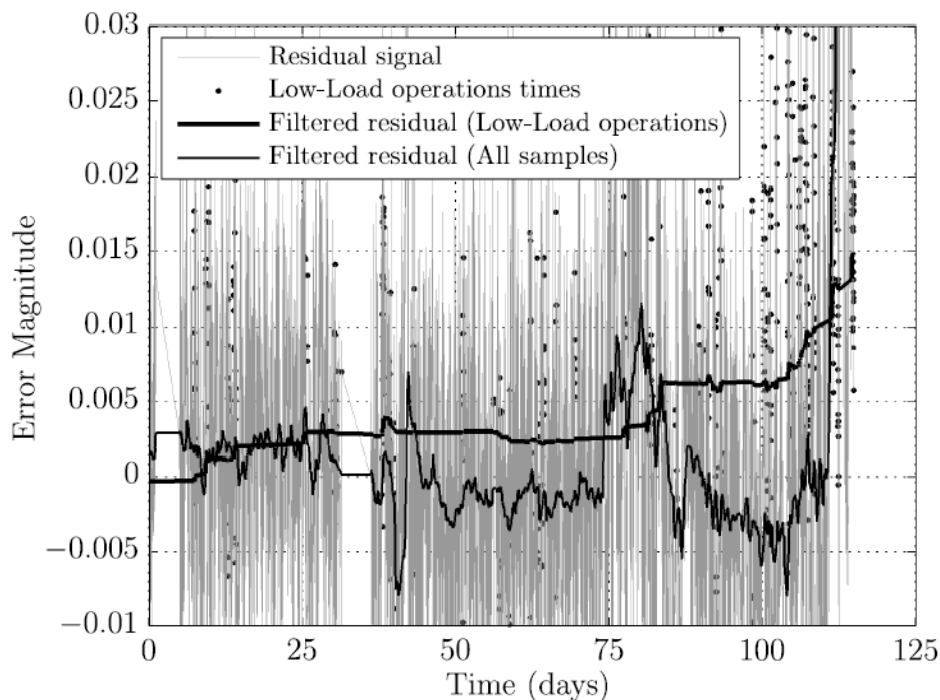


Figure 3.18 Utilisation du résidu filtré par basse vitesse de rotation pour la détection d'un défaut (source (Butler, 2012) fig.7.13)

La Figure 3.18 montre l'évolution du résidu en fonction du temps pendant un défaut de l'arbre principal, cette figure montre le comportement du résidu global et le comportement du résidu pris uniquement pendant les périodes de faible vitesse de rotation dites « low load ». La Figure 3.18 permet de voir comment le résidu est modifié une fois trié par rapport aux classes de vitesse. L'évolution du résidu est plus pertinente et augmente jusqu'à la casse le 110ème jour.

(Butler, 2012) montre qu'utiliser dans un modèle des variables issues de composants différents, mais se comportant de manière similaire, peut être pertinent. Toute la difficulté est d'arriver à déterminer quelles variables se comporteront de manière suffisamment proche pour que le modèle décrivant leur similarité soit plus simple qu'un modèle issu de la physique.

### 3.1.3.2. Modèle liant la température de l'enroulement du stator et celle du roulement de la génératrice

L'article suivant (Schlechtingen and Santos, 2011) propose un modèle permettant d'expliquer la température des roulements de la génératrice en fonction de la température des enroulements stator. Dans ces articles, plusieurs méthodes utilisant soit une régression linéaire soit un réseau de neurones avec ou sans dynamique sont comparées. Les auteurs montrent que ces approches offrent des résultats similaires.

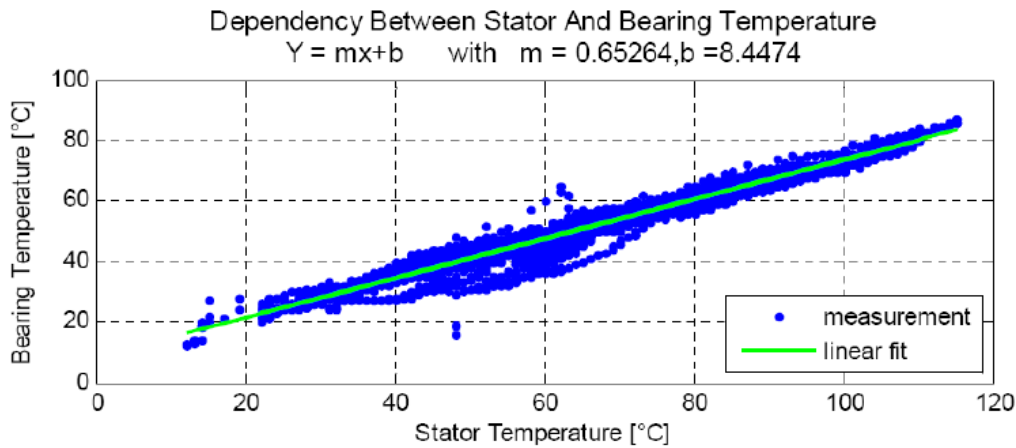


Figure 3.19 Lien de corrélation entre les températures du stator et du roulement génératrice (source (Schlechtingen and Santos, 2011) fig.14)

L'auteur propose plusieurs modèles et étudie plusieurs couples de composants similaires issus de la chaîne cinématique. Seul le couple stator/roulement génératrice sera présenté ici, mais tous les modèles sont créés de la même façon. Dans la Figure 3.19, on peut voir la forte corrélation entre la température du stator et celle du roulement génératrice.

Cette similarité est exploitée pour prédire la température du roulement à un instant donné. Afin de diminuer encore la dispersion du résidu, la température de la nacelle, la puissance et la vitesse de rotation ainsi que la température du roulement, mesurée à des instants précédents sont inclus dans le modèle. Les entrées du modèle (réseau de neurones ou régression linéaire) sont présentées dans le Tableau 3.5. La variable que l'on surveille est une température. L'inertie thermique introduit non seulement un hystérésis, mais aussi un retard entre le moment où l'énergie est transmise et où la température varie effectivement. La durée du retard associé à chaque variable d'entrée est calculée en étudiant les corrélations croisées entre les variables d'entrée et la température du roulement à prédire.

Tableau 3.5 Entrées du modèle température roulement génératrice versus stator(source (Schlechtingen and Santos, 2011))

Sortie du modèle $y_j^i(t)$	Entrée du modèle $y_j^i(t - lag_k)$	
Température du roulement génératrice (t)	Température enroulement stator (t-t <sub>1</sub> )	$y_{ref}^i$
	Température du roulement génératrice (t-t <sub>2</sub> )	$y_k^i$
	Puissance active (t-t <sub>3</sub> )	
	Vitesse arbre rapide (t-t <sub>4</sub> )	
	Température nacelle (t-t <sub>5</sub> )	

avec  $t_1, t_2, t_3, t_4$  et  $t_5$  les différents retards associés à chaque variable.

L'auteur précise que les variables habituelles  $y_k^i$  décrivant les modes de fonctionnement et l'environnement (puissance, vitesse, température nacelle et température passée) ne sont utilisées ici que pour diminuer la dispersion autour de la droite présentée en Figure 3.19.

### 3.1.3.3. Analyse des liens entre variables, comparaison de courbes empiriques

L'article (Yang et al., 2013) compare des courbes reliant des variables, créées pendant des périodes de bon fonctionnement avec les mêmes courbes construites à partir de données récentes. (Yang et al., 2013) propose d'analyser les liens les variables associés aux composants, sans se préoccuper de savoir si ces liens proviennent d'une causalité, et propose des modèles utilisant la similarité entre les composants. Afin de réaliser l'analyse, les données sont d'abord prétraitées puis une courbe reliant les variables est créée à partir d'une base d'apprentissage considérée comme sans défaut. Ensuite la même courbe est recréée en ligne, à partir de données récentes. L'auteur propose de calculer une métrique qui correspond à la surface entre les deux courbes. Plus cette surface est grande, plus la distance au modèle est importante, plus le défaut est critique.

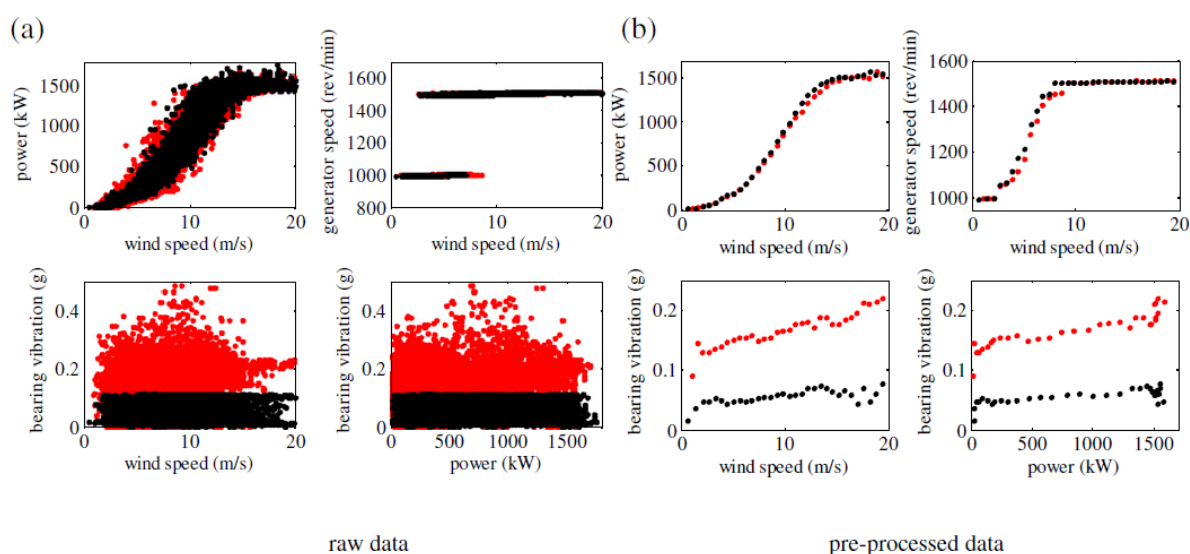


Figure 3.20 Prétraitement pour créer les modèles (source (Yang et al., 2013)fig.3)

La Figure 3.20 permet de voir l'influence du prétraitement. Les données sont regroupées par gamme en fonction d'une variable prise pour référence. Par exemple, pour les deux variables vent et puissance, les données peuvent être découpées par gamme de vitesses de rotation. Le but de cette variable de référence est de mieux prendre en compte les différents modes de fonctionnement. Une fois les données regroupées par gamme, une moyenne est faite pour ne garder qu'un point par gamme. Cela permet de diminuer la dispersion et de n'avoir qu'une seule courbe et non plus un nuage de points. Cette courbe sera utilisée comme une référence. Pour détecter un défaut, la méthode consiste à mesurer la distance entre une courbe reconstruite à partir de données récentes et la référence créée en amont.

Cette méthode montre que des liens peuvent être trouvés sur la plupart des variables SCADA. L'auteur propose aussi d'associer un type de défaut à chaque écart à la courbe détecté. Autrement dit, en fonction du couple de variables qui permet la détection, on peut savoir quel est le défaut. Cela revient à découper en sous-systèmes. Cependant, l'inconvénient de la méthode est son retard à la détection. La méthode proposée ne parvient pas à détecter des défauts plus de 3 jours avant la défaillance, temps nécessaire à l'élaboration de la courbe reliant les variables en ligne. Ceci n'est pas suffisant.

Notons que toutes les courbes obtenues sont considérées avoir la même pertinence. L'article ne fait pas la différence entre une relation due à la physique comme un modèle vent/puissance et une relation entre températures n'ayant aucune influence l'une sur l'autre. Pourtant, la relation de similarité entre les températures de la multiplicatrice permettrait d'apporter plus d'informations sur l'apparition d'un défaut que la relation entre le vent et la puissance. Enfin, la moyenne effectuée par gamme semble une bonne idée pour réduire la dispersion, mais engendre un retard à la détection mais il est possible qu'elle ne soit pas assez précise et qu'elle rende le système moins sensible au défaut.

#### 3.1.3.4. Modélisation par analyse en composantes principales et analyse discriminante

(Kruger et al., 2013) utilise les variables venant de tous les composants de la chaîne cinématique. Ce sont des variables de même grandeur : les températures de chacun des roulements et de chacun des systèmes de refroidissement. De plus, la méthode intègre aussi les vitesses de rotations mesurées sur l'arbre rapide, l'arbre multiplicatrice et l'arbre lent, pour avoir une information sur les modes de fonctionnement mais aussi parce que ces variables sont corrélées avec les températures. Les variables utilisées sont listées dans le Tableau 3.6. L'idée de base de cette méthode est d'appliquer une transformation linéaire afin d'analyser les données dans un nouvel espace. Les deux méthodes utilisées sont l'Analyse en Composantes principales et l'Analyse Discriminante de Fisher (FDA) (Duda et al., 2001). L'ACP exploite les corrélations entre les variables (ici, des températures mesurées à différents endroits dans le système) pour projeter les données dans un nouvel espace aux dimensions réduites. Les données ainsi représentées dans un espace de dimension réduite, où la variance des données projetées est maximale, peuvent être interprétées plus facilement par un utilisateur.

Tableau 3.6 Liste des variables utilisées pour la ACP (source (Kruger et al., 2013))

Numéro de Variable	Nom	Composant	Unité
1	Température roulement génératrice 1	Génératrice	°C
2	Température roulement génératrice 2	Génératrice	°C
3	Température enroulement stator	Génératrice	°C
4	Température roulement multiplicatrice 1	Multiplicatrice	°C
5	Température roulement multiplicatrice 2	Multiplicatrice	°C
6	Température carter multiplicatrice 2	Multiplicatrice	°C
7	Température Huile multiplicatrice	Multiplicatrice	°C
8	Vitesse intermédiaire	Multiplicatrice	tr/min
9	Vitesse arbre rapide	Génératrice	tr/min
10	Vitesse arbre lent	Rotor	tr/min

La détection de défaut par ACP utilise 2 indicateurs pour détecter les défauts. Un premier indicateur (le test T<sup>2</sup>) utilise les données projetées dans l'espace des premières composantes principales, le deuxième les données reconstruites dans l'espace initial après projection sur les premiers axes principaux (le test SPE).

La deuxième méthode, la FDA, calcule une transformation linéaire telle que les classes de fonctionnement observables lorsque la machine est sans défaut soient séparées au mieux dans ce

nouvel espace. De même, deux indicateurs identiques à la SPE et le T2 sont calculés à l'aide de ce nouveau changement de repère.

Les méthodes sont utilisées sur deux défauts, un défaut de refroidissement et un défaut de lubrification. Les résidus issus de l'ACP permettent de mettre en évidence un changement de comportement du système. Le SPE est plus particulièrement sensible aux défauts, bien que le taux de bonne détection n'atteigne que 71% pour 0 % de fausses alarmes. On obtient de meilleures performances avec les résidus issus de la FDA avec 99.5% de bonnes détections pour 4.3% de fausses alarmes.

Garder 3 à 4 % de fausses alarmes reste un problème qui peut sûrement être résolu comme pour les autres méthodes en lissant les résultats avec une moyenne mobile. Ces fausses alarmes viennent sûrement du fait que la méthode proposée ne prend pas en compte suffisamment de variables de fonctionnement comme la puissance. De même, la méthode ne prend pas en compte la dynamique des variables.

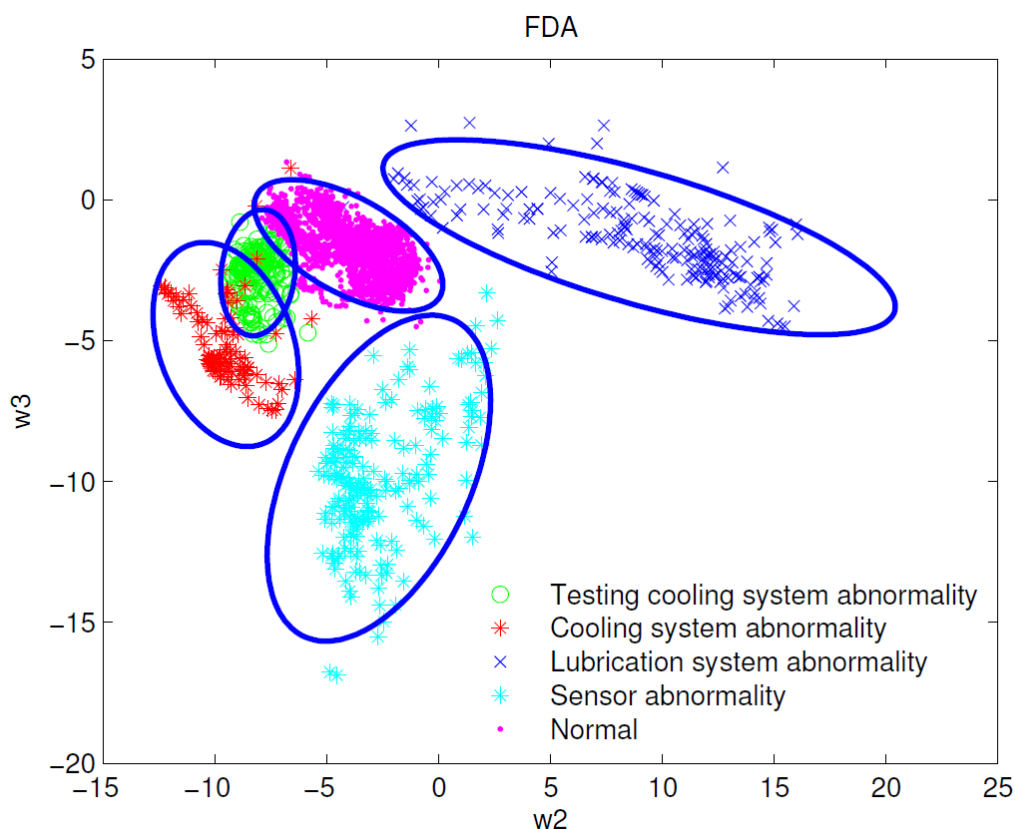


Figure 3.21 Projections des données suivant 2 composantes de la FDA, mise en évidence de familles de comportement (source (Kruger et al., 2013) fig.9)

Finalement, cette méthode, présentée par (Kruger et al., 2013), permet de faire une première analyse diagnostique. La projection des données suivant les deux premiers axes de la FDA permet de voir apparaître des groupements de données, associés à des défauts. L'article conclut malheureusement que la classification n'est pas encore possible et qu'elle doit être améliorée. Mais on peut voir que certains défauts sont clairement identifiables. Les ellipses bleues correspondent à un intervalle de confiance de 95 % construit uniquement avec des données d'un type de défaut particulier.



La base de données possède 4 familles de défauts, les sphères sont calculées sur les données correspondant à une famille de défaut. Certaines sphères sont nettement séparées des autres, et d'autres s'entrecroisent rendant la détection automatique impossible en l'état, sans avoir un fort taux d'erreur.

### **3.1.3.5. Conclusion sur les méthodes utilisant la similarité**

En conclusion, les méthodes utilisant la similarité entre les composants permettent d'obtenir des modèles plus simples que ceux obtenus par les méthodes utilisant les liens de causalité, en réduisant le nombre de variables d'entrées. Ces méthodes permettent de prendre en compte les variations de modes de fonctionnement et de l'environnement, puisqu'on choisit de comparer des composants qui réagissent de manière identique à ces variations. Ainsi, il n'est pas forcément nécessaire d'introduire dans le modèle des variables mesurant ces changements de modes de fonctionnement et d'environnement. Les modèles utilisant les liens de causalité sont généralement complexes, avec des non-linéarités et une influence mal modélisée des modes de fonctionnement et de l'environnement. Dans les modèles par similarité, si les composants sont bien choisis, la principale relation à modéliser peut être une simple droite. La dispersion restante est beaucoup moins importante que pour les modèles utilisant les liens de causalité.

Cette approche par similarité n'est pas parfaite pour autant. Il existe plusieurs difficultés à résoudre. Tout d'abord, elle ne peut fonctionner que si un seul des composants est en défaut à la fois. Si le stator et le roulement génératrice sont tous les deux en surchauffe, il est possible que la surchauffe de l'un masque celle de l'autre. Pour résoudre ce problème, il peut être intéressant de trouver plusieurs composants similaires à un même composant, afin de créer plusieurs résidus indépendants. Par exemple, on pourrait expliquer la température d'un roulement par un autre roulement avec un premier modèle et expliquer la température du même roulement par le stator. La probabilité pour que trois composants soient en panne en même temps est bien plus faible que pour deux composants seulement. On peut aussi rechercher des composants éloignés géographiquement, qui ont moins de chance d'être en panne simultanément. La principale difficulté de cette méthode est de trouver des composants suffisamment similaires pour que la relation entre eux soit plus simple à modéliser qu'un modèle utilisant les liens de causalité classique. Heureusement, la chaîne cinématique regroupant les composants les plus critiques se prête facilement à cet exercice, car les composants sont parfois physiquement similaires et leurs sollicitations sont synchronisées dans le temps. Il existe plusieurs roulements de taille proche dans la chaîne. L'origine de l'échauffement thermique que l'on surveille vient des mêmes arbres de rotation. Sortir de la chaîne cinématique pour trouver un composant similaire peut s'avérer plus difficile. Par exemple, il peut être difficile de trouver une température évoluant de la même façon que celle du convertisseur qui est un composant isolé des autres.

Ce type de méthode est plus efficace si le lien de similarité entre deux composants est plus simple à modéliser que le modèle physique. Les composants de référence devraient être choisis en fonction de leur nature, dans un premier temps. Ainsi, une température de roulement sera la meilleure référence pour une autre température de roulement.

### **3.1.4. Conclusion sur les méthodes mono-turbines**

Les méthodes mono-turbines sont découpées en 2 approches, les méthodes utilisant les liens de causalité, et les méthodes utilisant la similarité entre composants. La première approche a permis, à travers plusieurs publications, de montrer que les données SCADA peuvent être utilisées pour la

détection de défauts. En découpant l'éolienne en sous-systèmes, ces méthodes montrent qu'une localisation du défaut est aussi possible. Si le découpage est assez fin et va jusqu'au composant, il est possible d'avoir une idée de la gravité de la situation. Localiser le défaut sur un composant permet de trouver plus facilement la panne pendant une inspection ou encore de prévoir directement le remplacement de celui-ci. La gravité du défaut peut quant à elle être estimée par les experts puisque les résidus obtenus restent des grandeurs physiques connues par les opérateurs éoliens. Cependant, cette première approche a certaines limites, principalement dues à la difficulté de prise en compte des conditions environnementales et des modes de fonctionnement. Les méthodes utilisant les liens de causalité cherchent à expliquer l'évolution d'une variable par d'autres variables. Or, toutes les causes d'une variation ne sont pas toujours mesurables par les données SCADA, même les plus récentes. Les variations de l'environnement et certains modes de fonctionnement ne sont pas mesurés. De plus, même avec les variables disponibles, certaines non-linéarités restent difficiles à modéliser. Plusieurs articles montrent que les résultats entre une méthode linéaire et non linéaire sont semblables. Malheureusement, aucune explication n'a encore été donnée. Il est possible que cela vienne du fait que tous les modes ne soient pas représentés par le même nombre d'occurrences. Comme vu précédemment, les éoliennes fonctionnent la majorité du temps au début de la courbe de puissance ou dans la partie centrale. Aucune approche locale n'a été proposée pour permettre de trier les données d'apprentissage en fonction du mode de fonctionnement.

Les approches utilisant la similarité permettent d'améliorer les résultats de détection. Celles-ci permettent d'avoir une référence de départ qui prend déjà en compte les variations des modes de fonctionnement et de l'environnement. Cela permet de simplifier le modèle, car contrairement au modèle utilisant les liens de causalité, le lien entre les composants est lui « majoritairement » linéaire. On a pu voir que le principal lien entre les composants peut être expliqué par une droite ou une courbe simple avec une dispersion réduite. Il reste cependant toujours une dispersion qui, elle, peut être expliquée par les variables des modes de fonctionnement.

Finalement, il faut remarquer que plusieurs questions restent en suspens. La première est la validité des méthodes au cours du temps. Les modèles doivent-ils être réappris d'année en année ? Les méthodes utilisant les liens de causalité doivent normalement être utilisées uniquement sur des plages de conditions environnementales identiques à celle de l'apprentissage, ce qui n'est pas possible en pratique. La dynamique de la machine n'a pas non plus été suffisamment étudiée. Les entrées des modèles proposés sont généralement limitées aux 30 dernières minutes de fonctionnement, alors que la température d'un roulement peut mettre plusieurs heures pour atteindre un régime permanent. Une étude avec des horizons plus longs pourrait permettre d'améliorer encore les résultats.

### **3.2. Les méthodes multi-turbines**

La classe des « méthodes multi-turbines » regroupe les méthodes dont le principe repose sur la comparaison de variables enregistrées sur différentes turbines d'un même parc. En pratique, les exploitants de parcs éoliens utilisent couramment la comparaison turbine à turbine de façon empirique, mais les méthodes de comparaison utilisées ne reposent pas toujours sur des bases bien formalisées. Remarquons d'abord que la comparaison entre turbines d'un même parc pour la mise en évidence de défauts s'appuie sur deux hypothèses fortes :

- Les éoliennes d'un parc sont homogènes. Un parc éolien est construit en une seule fois, la plupart du temps. Les éoliennes sont donc choisies à l'avance pour être toutes du même type.
- Toutes les éoliennes d'un parc sont soumises à des conditions de vent comparables. La nature du vent est supposée semblable en tout point du parc. Cette hypothèse semble raisonnable puisque les parcs éoliens se concentrent sur de petites surfaces par rapport aux principaux flux d'air. C'est pour cette raison que les exploitants supposent que les machines évoluent ensemble dans le temps puisqu'elles sont soumises à des sollicitations semblables.

Si l'on considère 2 modèles  $f_j^1$  et  $f_j^2$  décrivant les variations d'une variable  $j$  en fonction des autres variables de la même turbine :

$$y_j^1 = f_j^1(y_1^1(t), y_2^1(t), \dots, y_k^1(t)) \text{ pour la turbine 1}$$

$$y_j^2 = f_j^2(y_1^2(t), y_2^2(t), \dots, y_k^2(t)) \text{ pour la turbine 2}$$

Afin de détecter des comportements anormaux, les méthodes multi-turbines peuvent être formalisées de cette manière et consistent à comparer les modèles  $f_j^1$  et  $f_j^2$  liant les variables entre elles. Cependant, dans les travaux publiés sur les comparaisons inter-turbines, cette comparaison de modèles est rarement formalisée en tant que telle, et les modèles mathématiques  $f_j^1$  et  $f_j^2$  ne sont pas établis de façon explicite. Comme on va le voir dans la présentation des méthodes qui suit, la comparaison inter-turbines consiste donc souvent en pratique à comparer des réponses de modèle, en se limitant même parfois à une comparaison visuelle graphique des courbes de réponses des modèles représentant différentes turbines.

### 3.2.1. Comparaison directe de courbes empiriques

Une méthode de comparaison directe de modèles a été présentée dans (Astolfi et al., 2014). Elle consiste à comparer les différents modèles liant la puissance et la température d'un roulement établis pour plusieurs turbines, à partir de données enregistrées sur la même période. L'hypothèse sous-jacente est que le lien puissance/température devrait être le même pour des turbines identiques, installées sur le même parc. Comme vu dans la section 3.1.2, la température d'un roulement ne dépend pas uniquement de la puissance, mais aussi des conditions environnementales, qui sont identiques pour l'ensemble des turbines.

#### 3.2.1.1. Création de la courbe puissance versus température

La comparaison de courbes liant la puissance et la température peut se voir ici comme une forme de comparaison de modèles utilisant les liens de causalité. La puissance étant la variable décrivant le mieux les échanges énergétiques et les modes de fonctionnement, la température est ici expliquée uniquement par la puissance. Afin de diminuer une partie de la dispersion autour de la courbe, l'auteur propose de moyenniser les mesures de température en fonction de gammes de puissance, qui peuvent être assimilées à des modes de fonctionnement. L'auteur crée ainsi 10 gammes de puissance, une toutes les 10 % de la puissance nominale. Les données de température sont donc triées et moyennées de sorte d'avoir un point représentatif par gamme de puissance. L'auteur précise que le nombre de gammes a peu d'impact sur le résultat final.

L'article met en évidence que certains effets topologiques ne sont pas négligeables. Une forêt ou une colline peut modifier le profil du vent. Les éoliennes entre elles peuvent jouer le rôle d'éléments perturbateurs. Les parcs éoliens sont construits en ligne de manière à ce que le vent dominant ne soit pas perturbé par des éléments du décor. L'auteur choisit de trier les données pour que seul le vent soufflant de manière orthogonale à la ligne du parc soit pris en compte. Cela permet à la fois d'éviter le sillage que peut créer une éolienne sur une autre et aussi d'avoir le minimum d'éléments perturbateurs.

### 3.2.1.2. Détection de défauts

La détection de défaut se fait en analysant la similarité entre deux courbes. Une fois les données filtrées et triées par gamme de puissance, la courbe liant la puissance à la température peut être utilisée. L'auteur construit ces courbes sur des périodes allant d'une semaine à un mois. La taille de la période d'analyse est choisie pour avoir suffisamment de points pour que les comparaisons soient significatives. Une semaine correspond à 1000 points de mesure SCADA, ce qui est suffisant pour être significatif.

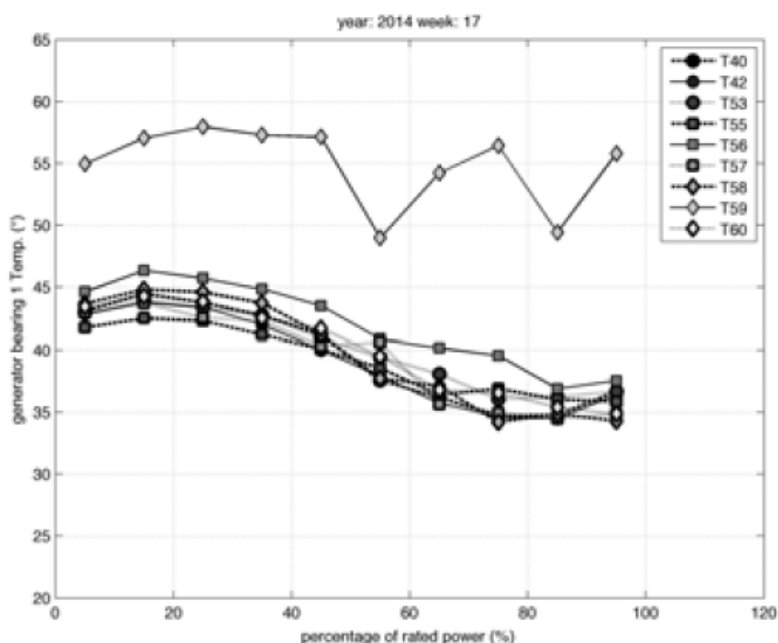


Figure 3.22 Mise en évidence d'un défaut roulement génératrice (source (Astolfi et al., 2014))

Dans la Figure 3.22, on peut voir un exemple de comparaison des courbes liant puissance et température sur les turbines d'un même parc. Les courbes ont été construites une semaine avant l'arrêt dû à un défaut de température roulement génératrice. Une surchauffe de la turbine 59 en défaut est clairement visible sur cette figure. L'écart entre la turbine 59 et les autres turbines est de 10 °C pour chacune des gammes de puissance, alors que les turbines sans défauts sont elles proches de quelques degrés.

### 3.2.1.3. Avantages/limites de l'approche

#### 3.2.1.3.1. Création de courbes simplifiées

Cette première approche présente plusieurs avantages. Elle est tout d'abord facile à interpréter puisqu'elle reprend la philosophie de comparaison utilisée à l'origine avec le SCADA. Les courbes sont

aussi très simples à construire puisqu'ici le calcul ne comprend qu'une moyenne sur une période réduite d'une semaine, contrairement aux régressions précédentes utilisant des bases d'apprentissage de plusieurs mois. Les modèles prennent en compte implicitement une partie des modes de fonctionnement, ce qui permet de diminuer la dispersion de la température. De plus, on peut considérer que les conditions environnementales sont les mêmes sur toutes les turbines du parc. Contrairement aux méthodes précédentes qui utilisent un modèle appris avec des conditions environnementales passées, ici les courbes comparées sont construites dans les mêmes conditions. Cela permet de diminuer la dispersion due à l'environnement

#### **3.2.1.3.2. Méthode compatible avec d'autres variables**

L'article ne présente que des comparaisons de températures de la chaîne cinématique, mais on sait, grâce aux méthodes internes (ou mono-turbines), que l'on peut créer des modèles sur les autres variables. Il paraît donc vraisemblable que d'autres modèles utilisant les liens de causalité ou de similarité puissent être comparés entre les turbines.

#### **3.2.1.3.3. Méthode encore sensible aux modes opératoires**

Même si la puissance a été prise en compte, cela n'empêche pas la méthode d'être influencée par les modes de fonctionnements car la température dépend d'autres variables. De plus, on peut remarquer que l'écart entre les turbines semble dépendre de la puissance. Un système automatisé pour créer un seuil de détection devra le prendre en compte. Pour le moment, seule l'allure globale des courbes est appréciée. Certains défauts peuvent n'apparaître que sur une gamme de puissance par exemple. Un écart de même amplitude aura plus d'importance sur une gamme de puissance faible que sur une gamme proche de la puissance nominale.

#### **3.2.1.4. Dispersion inter-turbine**

On peut voir finalement qu'il existe une dispersion entre les différentes turbines. Cette dispersion dépend de l'orientation de la nacelle. En fonction des secteurs de vent, la similarité inter-turbine varie. L'article ne propose pas en revanche de vérifier que le temps ou d'autres variables SCADA ont une influence sur la dispersion inter-turbine. Cette dispersion observée montre que le parc n'est pas parfaitement homogène. Les éoliennes peuvent avoir de légères différences de fonctionnement et surtout être des éléments perturbateurs qui modifient le vent qu'elles perçoivent. Malgré cette dispersion, cette méthode est capable de détecter des défauts.

### **3.2.2. Comparaison à des modèles de référence constructeur**

#### **3.2.2.1. Utilisation de la courbe de puissance référence constructeur**

(McLaughlin et al., 2009) propose aussi de comparer des courbes entre elles. Cette fois-ci c'est la courbe de puissance en fonction du vent qui est reprise. En revanche, plutôt que de comparer les courbes de puissance des turbines entre elles, l'article propose de comparer les turbines à leur courbe de référence constructeur. Ce sont les distances de chaque turbine à sa courbe de référence qui sont ensuite comparées. Les turbines n'ont pas toutes les mêmes courbes de référence, mais on suppose qu'elles en sont toutes à la même distance en fonctionnement normal.

### 3.2.2.2. Calcul de métriques

L'article montre les résultats du calcul de plusieurs métriques entre la courbe de puissance de référence et celle de la turbine sur une période choisie. La courbe de puissance de référence est construite en prenant des mesures de puissance et de vent au début de la vie de la machine. Il est d'usage de prendre un point de mesure tous les 0,5 m/s afin de construire la courbe. Tout comme dans la méthode de la partie 3.2.1, les valeurs de la variable de sortie (ici la puissance) sont comparées par gamme de la variable d'entrée (ici le vent tous les 0,5 m/s). Cela permet de diminuer la dispersion en effectuant une moyenne pour chaque gamme. Les métriques ne sont pas présentées précisément dans l'article. On peut supposer qu'elles sont construites afin de détecter un écart global à la courbe de référence ou bien des écarts correspondant à des défauts connus. Un écart positif ou négatif par rapport à la courbe n'a pas la même signification. En effet, un écart négatif va représenter une perte de production par rapport à la production attendue, ce qui peut être révélateur d'un défaut. En revanche, un écart positif va être synonyme d'une surproduction, ce qui montre plus un problème au niveau de la création du modèle. Un écart sur la partie centrale ou saturée peut être lié à des défauts différents. L'étude de la courbe de puissance permet notamment d'identifier certains défauts principalement liés au contrôle/commande comme on a pu le voir dans la section 3.1.2.1. On peut supposer que les métriques sont construites de manière à identifier ces défauts.

### 3.2.2.3. Comparaison visuelle

Dans la Figure 3.23, les turbines sont représentées par 2 métriques qui ont été calculées. Chaque point du graphique synthétise une semaine de données pour une turbine. La principale information de ce graphique est de montrer que les turbines en bon fonctionnement sont regroupées le long d'une ligne. Cette ligne devient donc la nouvelle référence commune en remplacement des différentes courbes constructeur.

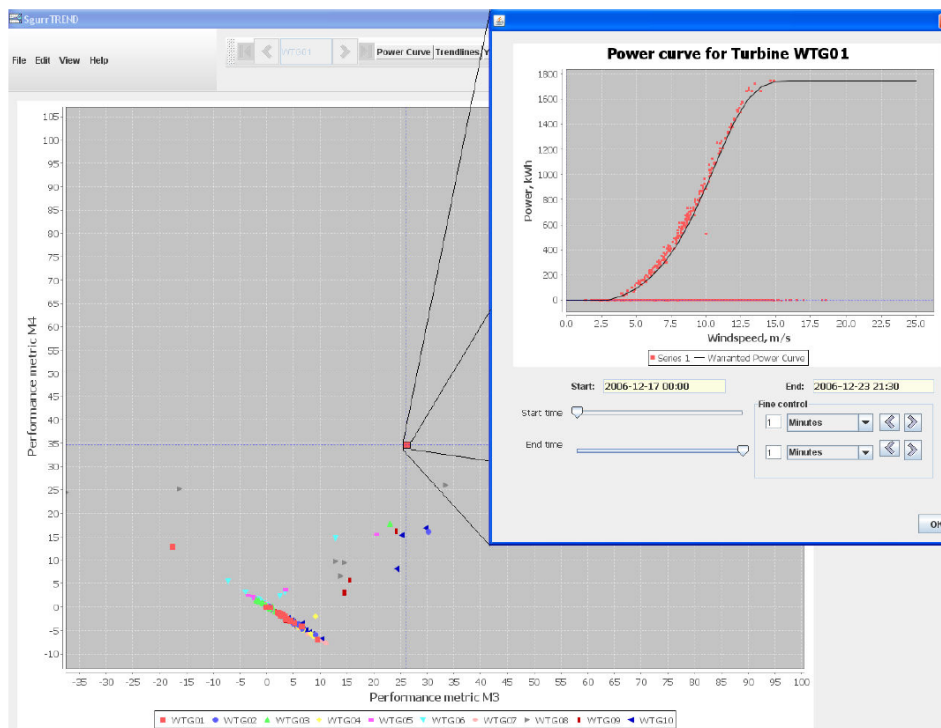


Figure 3.23 Copie d'écran du logiciel de calcul de métrique (source(McLaughlin et al., 2009))

#### **3.2.2.4. Avantages/limites de la méthode**

Cette méthode présente l'avantage de permettre une bonne distinction visuelle entre les turbines en défaut et les turbines en fonctionnement normal. Les métriques définies dans la méthode présentée dans (McLaughlin et al., 2009) permettent d'identifier un défaut sur une turbine en fonction de la position des données de cette turbine par rapport à la ligne de référence. En effet, ces métriques sont basées sur la courbe de puissance qui permet par observation d'identifier certains types de défaut, comme les défauts du système de contrôle des pales (cf partie 3.1.2.1.2).

Cette méthode ne prend en compte que la courbe de puissance, la manière dont les métriques sont calculées n'est pas décrite. On peut supposer cependant que d'autres métriques doivent pouvoir être créées pour d'autres courbes comme une courbe puissance versus température, par exemple. Cette méthode n'utilise que la vitesse du vent et la puissance pour la courbe. Le sillage et la densité de l'air ont un impact sur la courbe de puissance, ce qui entraîne des baisses de performance. La densité, la direction du vent et les conditions environnementales en général ne seront pas les mêmes lors du calcul des métriques que lors de la création de la courbe de référence. S'ils ne sont pas pris en compte, ils peuvent créer des fausses alarmes importantes. Il serait intéressant de voir comment la densité ou la direction du vent peuvent être intégrées dans le calcul des métriques.

#### **3.2.3. Comparaison de résidus issus de modèles appris sur différentes turbines**

##### **3.2.3.1. Apprentissage d'un modèle sur une turbine et utilisation sur une autre turbine**

L'article (Papatheou et al., 2014) étudie les résultats de modèles basés sur la courbe de puissance. Cet article propose de comparer l'erreur de prédiction entre des modèles appris sur différentes turbines. La méthode est la suivante. Pour les  $n$  turbines d'un même parc, on apprend  $n$  modèles liant la puissance générée à la vitesse du vent à partir de données venant de la turbine correspondante. Dans un second temps, on applique ce modèle sur les données d'une base de test. Chaque modèle va chercher à faire la prédiction de la puissance générée pour chacune des  $n$  turbines sur la période de test. On obtient ainsi  $n$  erreurs de prédiction pour chaque modèle. Les résultats sont synthétisés dans une matrice  $n \times n$  comprenant une évaluation de l'erreur de prédiction obtenue par chaque modèle pour chaque turbine.

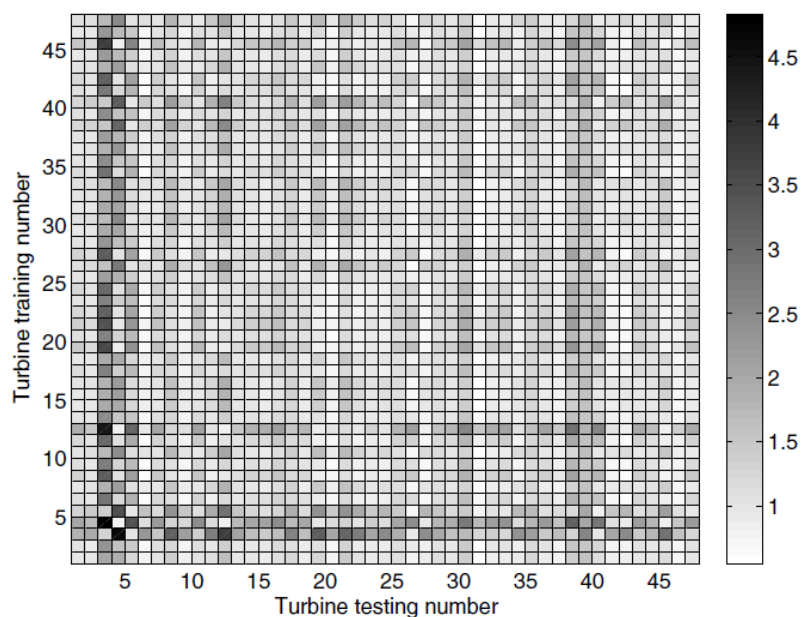


Figure 3.24 Matrice de confusion obtenue sur un parc de 48 machines en utilisant des réseaux de neurones. (source (Papatheou et al., 2014)fig.2)

La Figure 3.24 représente cette matrice n par n. Le numéro de la turbine qui a servi pour l'apprentissage est indiqué en ordonnée et le numéro de la turbine qui a servi pour le test est indiqué en abscisse. L'échelle utilisée correspond à la somme de l'erreur quadratique (SEQ) normalisée sur la période de test. On peut voir sur le graphique que les résultats obtenus sont en dessous de 5, ce qui est considéré par l'auteur comme un bon résultat de prédiction. L'article montre donc qu'un modèle appris sur une turbine peut être utilisé pour une autre. Cela permet de vérifier que le système que l'on cherche à modéliser (ici la relation vent/puissance) est similaire sur toutes les turbines du parc. Les turbines fonctionnent de la même manière.

### 3.2.3.2. Mise en évidence des effets de site

La Figure 3.24 permet de voir que presque toutes turbines ont une SEQ inférieure à 3. Les turbines ayant la plus grande erreur sont les turbines 3 et 4 et on peut voir que les modèles appris sur ces turbines donnent aussi les plus mauvais résultats. Il est expliqué que ces 2 turbines sont en fait situées côte à côte sur le bord du parc. Même si l'origine de cette différence n'est pas identifiée complètement dans l'article, elle est cependant associée à un effet de site. Il faut donc prendre en compte certaines particularités des éoliennes avant de les comparer. Pour prendre en compte ces effets de site, on pourrait commencer par utiliser la direction du vent pour créer les modèles. Les effets de site peuvent être plus ou moins influents en fonction de la direction du vent. En adaptant le modèle aux différents effets du terrain associé à une certaine direction du vent, on devrait pouvoir diminuer la dispersion.

### 3.2.4. Conclusion sur les méthodes multi-turbine

Les méthodes multi-turbines de la littérature utilisent l'effet parc. Le comportement de chaque turbine d'un même parc est supposé être similaire aux autres si les turbines sont sans défaut. Les méthodes utilisent principalement la relation vent/puissance qui permet de comparer les performances énergétiques des turbines ou la relation puissance/températures qui permet de détecter d'éventuelles



surchauffes. Les courbes, construites de manière empirique sur des périodes de temps récentes, sont comparées entre elles ou encore les modèles appris sur une turbine sont utilisés sur d'autres turbines. Les résultats obtenus par (Astolfi et al., 2014) montre une dispersion entre les éoliennes lorsque l'on essaye de comparer des courbes de température. Néanmoins, le deuxième article (McLaughlin et al., 2009) montre que cette dispersion et ces différences propres aux machines peuvent être dissociées de la dispersion qui serait due à un défaut, en incluant d'autres variables disponibles pour créer les modèles.

Le sillage d'une machine sur une autre ou l'influence de spécificités topologiques peut modifier la nature du vent perçu par une turbine, en créant des turbulences, par exemple, ce qui va influencer la performance de la machine de manière significative. Cette modification n'influence pas forcément la vitesse de vent mesurée par le SCADA. En effet, la vitesse mesurée dans le SCADA représente une unique mesure de la vitesse du vent en un point situé sur la nacelle. Cette mesure ne suffit pas pour caractériser la totalité du front d'air. Prendre en compte ces spécificités est crucial si l'on veut diminuer la dispersion des courbes, pour des sites éoliens particulièrement soumis aux effets de sillage.

L'hypothèse de base de la comparaison des turbines est de considérer qu'elles sont toutes semblables. Étant du même modèle et étant sollicitées par un vent similaire, on s'attend à ce qu'elles évoluent ensemble dans le temps et donc que des variables de même nature mesurées sur différentes turbines évoluent de manière identique. Or, on peut remarquer qu'il existe peu de méthodes exploitant ce lien de corrélation entre les variables des différentes turbines.

En pratique, les spécificités du site et des éoliennes se traduisent par un décalage dans les phases démarrage et les arrêts des éoliennes. Cette désynchronisation des machines peut générer une dispersion des indicateurs entraînant des fausses alarmes. Afin de limiter cet effet, les éoliennes sont comparées sur une période et non sur un seul échantillon. En effet, le décalage entre les machines n'est jamais de plus de quelques échantillons 10 minutes. Par conséquent les comparaisons sur une semaine entre les turbines qui sont proposées dans les articles précédents sont peu sensibles à ce phénomène de désynchronisation. En revanche, la comparaison sur une période induit un retard à la détection qui est lié à la taille de la période de comparaison. Il y a un compromis à faire entre la diminution de l'impact de la désynchronisation et le retard à la détection.

Les méthodes proposées dans la littérature se concentrent sur la comparaison de modèles liant le vent à la puissance entre les turbines. Ce type de comparaison rend difficile la localisation des défauts dans l'éolienne. Cependant, des premiers modèles proposés dans (Astolfi et al., 2014) lie la température de roulement de génératrice avec la puissance ce qui permet d'isoler le défaut plus précisément. Il faut maintenant étudier la possibilité de généraliser ces modèles sur les autres composants de l'éolienne.

### **3.3. Conclusion générale et perspectives de recherche**

Le SCADA permet de stocker un grand nombre de données qui, jusqu'il y a peu, n'étaient pas utilisées réellement. Le nombre de variables et le nombre de turbines à surveiller rendent leur analyse par des opérateurs difficile sans un outil adapté. Les opérateurs se concentrent sur les variables directement liées à la performance ou analysent les variables a posteriori, suite à l'occurrence d'un défaut. Plusieurs méthodes ont été publiées récemment dans la littérature pour pallier ce manque.

Parmi les méthodes proposées, une première famille rassemble les méthodes mono-turbines. Les méthodes mono-turbines sont des méthodes « myopes ». Elles ne prennent en compte que des variables mesurées sur la même turbine, sans s'intéresser au comportement des autres turbines du même parc. Certains travaux se concentrent sur l'utilisation de la courbe de puissance qui permet d'analyser la performance énergétique globale de la machine. Malheureusement, la dispersion des points de mesures autour de la courbe de référence limite son utilisation à l'apparition de défauts sévères, induisant une perte de rendement significative. De plus, ces méthodes ne permettent pas de localiser le défaut. D'autres travaux se concentrent sur l'élaboration d'un modèle mathématique reliant des variables du SCADA, appris à partir d'une base d'apprentissage sans défaut. L'indicateur de dégradation est la différence entre la variable prédite par le modèle et la variable enregistrée par le SCADA. Les méthodes utilisant les liens de causalité traduisent une relation de cause à effet entre les variables tandis que les méthodes utilisant la similarité modélisent une similarité de comportement. Ces méthodes permettent de découper la machine en sous-systèmes, ce qui facilite la localisation des défauts. Malheureusement, les données SCADA sont des données moyennées toutes les 10 minutes, évaluant un comportement général des variables. Les modèles proposés sont donc relativement grossiers et génèrent des résidus dont la dispersion est importante. L'inconvénient de ces méthodes est donc le nombre de fausses alarmes qu'elles génèrent. Plus particulièrement, ces modèles reliant des variables SCADA sont appris sur des périodes de temps limitées à quelques mois. Ils ne peuvent pas prendre en compte l'ensemble de la variabilité des conditions environnementales, qui fluctuent sur une période de un an voire plus. Ainsi, des écarts de température ou de conditions de vent par rapport à celles observées lors de la période d'apprentissage génèrent des fluctuations importantes sur les résidus et donc des fausses alarmes. Enfin, notons qu'aucun article dans la littérature ne mentionne le fait que le comportement de la machine peut évoluer au cours du temps et n'évoque la nécessité d'avoir des modèles dont les paramètres s'adaptent au cours du temps. Pour conclure sur ces méthodes, aucun auteur n'évalue sur une longue période les performances des indicateurs proposés. Ils se contentent la plupart du temps de vérifier la détection d'un défaut sur une période de temps précédent la détection du défaut par le SCADA. Il est donc difficile d'évaluer la capacité de ces indicateurs à répondre à la demande des exploitants de parcs éoliens : être capable de détecter un défaut au plus tôt tout en limitant le nombre de fausses alarmes, qui génèrent un surcoût important pour les équipes de maintenance.

Un deuxième type d'approche est l'approche multi-turbine. Pour s'affranchir des conditions environnementales mal prises en compte par les méthodes mono-turbines, ces méthodes proposent de comparer des éoliennes d'un même parc entre elles, sur une période de temps donnée. Les principales méthodes proposées utilisent la courbe de puissance élaborée à partir de données enregistrées sur une période allant d'une semaine à un mois. Les machines étant soumises aux mêmes conditions de vent et étant de même marque, leurs performances énergétiques, estimées par les courbes de puissance, devraient être similaires. Les méthodes proposées sont principalement visuelles. Elles proposent un outil de contrôle à l'opérateur qui doit détecter seul l'occurrence d'un défaut au vu de la position des courbes des turbines les unes par rapport aux autres. La dispersion des points autour de la courbe est diminuée par des méthodes de filtrage adaptées. Par contre, ces méthodes souffrent des mêmes défauts que les méthodes mono-turbines utilisant la courbe de puissance : elles ne permettent pas de localiser le défaut. De plus, elles peuvent générer un retard à la détection important, compte tenu du nombre de points nécessaires à l'élaboration de la courbe. On ne trouve pas dans la littérature de méthodes proposant de relier les variables SCADA 10 minutes de différentes turbines par un modèle.

Au vu de cette analyse bibliographique, on constate que plusieurs axes n'ont pas été explorés par les différents auteurs. Le premier consiste à proposer une méthode permettant d'évaluer de manière

performante les méthodes proposées. Il s'agit ici de proposer des indicateurs de performances permettant d'évaluer le gain et le surcoût générés par les méthodes pour les opérateurs de maintenance. En plus de proposer des indicateurs, cela suppose de mener une campagne de validation sur une période de temps conséquente et sur plusieurs turbines d'un même parc à la fois. Les différentes méthodes proposées dans la littérature pourraient ainsi être comparées et leur utilisation dans le cadre de la surveillance d'un parc éolien en cours de fonctionnement évaluées en prenant en compte les contraintes industrielles. Dans un deuxième temps, il serait intéressant de proposer une méthode permettant de coupler les méthodes mono-turbines et multi-turbines. En effet, l'analyse bibliographique effectuée a mis en évidence les avantages et les inconvénients de chacune : la sensibilité aux variations environnementales pour les méthodes mono-turbines, la difficulté de localisation du défaut avec les méthodes multi-turbines. Il semble donc pertinent de combiner les deux méthodes afin de profiter de la force de chacune d'elle.

Dans le cadre de cette thèse, ce sont donc sur ces deux axes que nous avons plus précisément focaliser nos recherches.



## 4. Présentation de la base de données VALEMO

Afin d'évaluer les méthodes proposées dans la littérature et de valider les nouvelles méthodes proposées dans cette thèse, une base de données a été créée avec des données appartenant à VALEMO. Rappelons que si cette thèse vise à proposer des contributions méthodologiques innovantes, elle a dans le même temps l'ambition d'évaluer leurs performances du point de vue de l'industriel et cherche à se placer au plus proche des conditions opérationnelles. Pour cela, les méthodes étudiées et proposées seront testées sur des données réelles qui permettront de valider la pertinence des indicateurs pour un industriel de l'éolien. Ainsi, une base de données a été construite grâce à des données provenant d'un parc en activité et ayant connu des défauts majeurs. Cette base de données est loin d'être exhaustive, mais elle permet de se rendre compte de l'efficacité des indicateurs en termes de détection, en les étudiant sur les périodes en défaut, mais aussi en terme de fausses alarmes, en les étudiant sur les périodes sans défaut.

### 4.1. Présentation du parc éolien

La construction de la base de données a été réalisée pour qu'elle soit la plus représentative possible de cas réels. La base de données contient des données des 6 machines identiques de 2MW à axe vertical. Elles font parties d'un même parc éolien et sont numérotées de WT6 à WT11, leur localisation géographique est représentée sur la Figure 4.1. Les données du parc voisin ne sont pas disponibles pour cette étude, car elles n'appartiennent pas à VALEMO, c'est pour cela que l'étude se concentre uniquement sur les 6 machines les plus au sud. Ce type de puissance et de machine est le plus courant actuellement. De plus, bien qu'il existe des parcs de plus grande taille dans le monde, en France, les parcs comprenant moins de 10 machines sont les plus courants.

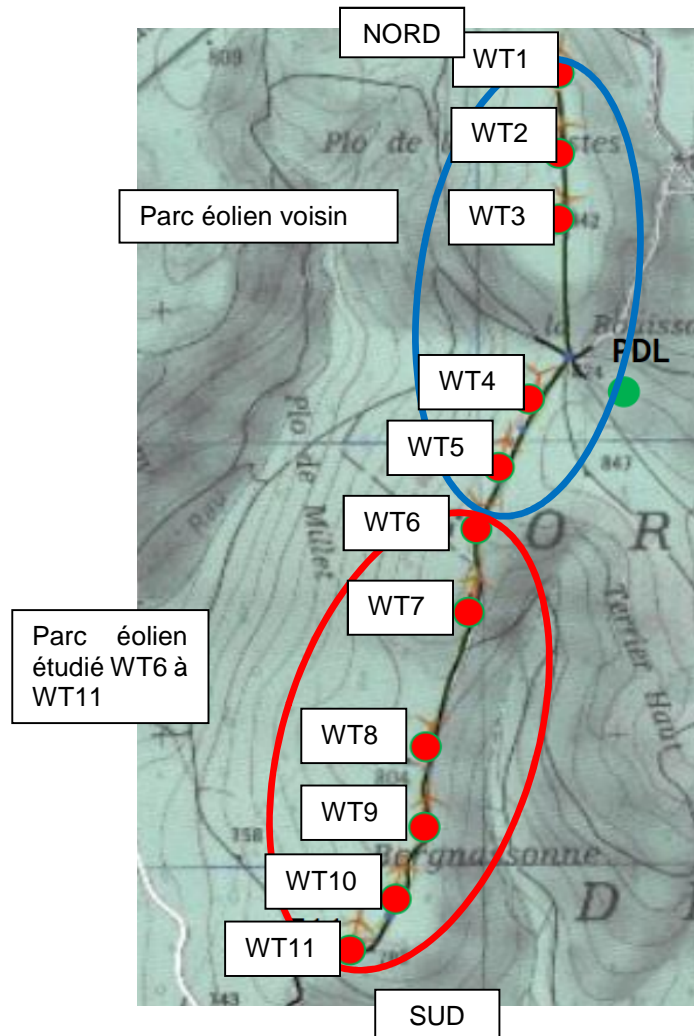


Figure 4.1 Implantation du parc étudié de WT6 à WT11

Ce parc est organisé en une seule ligne en raison de sa taille, comme le montre la Figure 4.1. Cette répartition en une unique ligne est faite pour diminuer les effets de sillage. Le sillage est l'influence qu'une éolienne peut avoir sur une autre lorsque le vent souffle dans la direction de leur alignement. Cet effet de sillage peut avoir une influence importante sur les performances de la machine, et par conséquent sur les indicateurs de dégradation (cf partie 3.2.4). En règle générale, les constructeurs de parcs éoliens cherchent à diminuer au maximum ces effets en plaçant les parcs sur des zones où une direction de vent est prédominante et en orientant le parc en fonction de cette direction. De cette manière, une éolienne est rarement dans le sillage d'une autre.

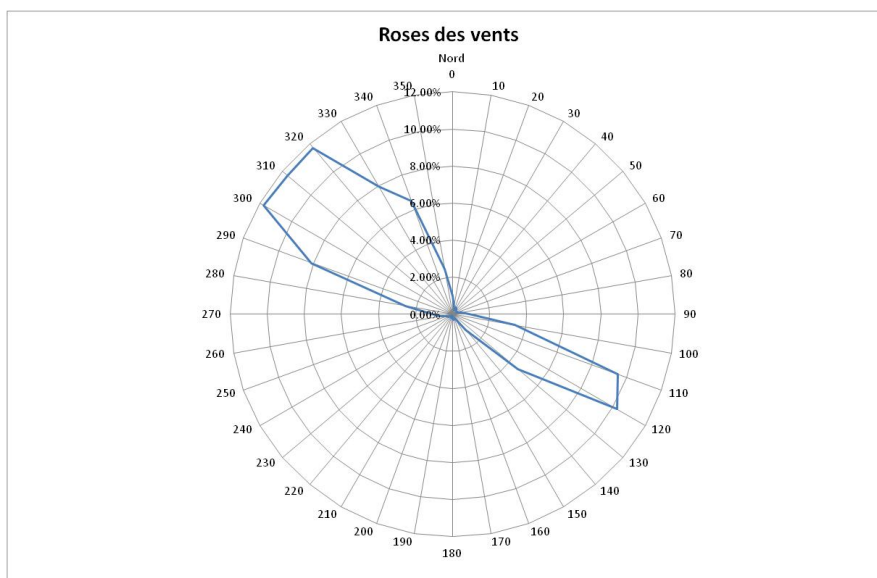


Figure 4.2 Répartition des occurrences de vitesse de vent mesurées sur le parc étudié pendant les 4 ans de l'étude

La Figure 4.2 montre la répartition des directions du vent qui ont été mesurées sur le parc. Cela permet de constater qu'il existe une direction majoritairement représentée du Sud Est au Nord Ouest. Par conséquent, les effets de sillages sont très peu impactants sur ce site.

La base de données comprend des données pendant les 4 premières années de la vie du parc, du 1er novembre 2009 au 31 décembre 2013. Durant ces 4 années, pour chaque éolienne, 28 variables provenant du SCADA ont été enregistrées ainsi qu'une variable venant d'un mat de mesure proche. Le Tableau 4.1 liste toutes ces variables.

Tableau 4.1 Liste des variables disponibles dans la base de données, en gras les variables utilisées pour les modèles testés sur cette base de données.

Date (jj/mm/aaaa hh:mm)	Consigne d'angle de pale 2 (degré)
<b>Puissance active (W)</b>	Consigne d'angle de pale 3 (degré)
Puissance réactive (W)	Angle réel de pale 1 (degré)
Vitesse du vent nacelle (m/s)	Angle réel de pale 2 (degré)
Direction du vent (degré)	Angle réel de pale 3 (degré)
Température extérieur nacelle (°C)	Fréquence d'alimentation (Hz)
<b>Température intérieur nacelle (°C)</b>	Tension produite par la génératrice (V)
Température du roulement multiplicatrice 1 (°C)	Intensité produite par la génératrice (A)
Température du roulement multiplicatrice 2 (°C)	Direction de la nacelle par rapport (degré)
<b>Température enroulement statorique (°C)</b>	Température de l'intérieur de la multiplicatrice (°C)
Température du transformateur (°C)	Couple de la génératrice (N.m)
Température de l'huile multiplicatrice (°C)	Consigne de couple de la multiplicatrice (N.m)
Vitesse de rotation de l'arbre lent (tr/min)	<b>Température roulement avant génératrice (°C)</b>
<b>Vitesse de rotation de l'arbre rapide (tr/min)</b>	<b>Température roulement arrière génératrice (C)</b>
Consigne d'angle de pale 1 (degré)	Température extérieur mât de mesure (°C)

Les données sont toutes mesurées et enregistrées au dixième de l'unité de mesure, à une fréquence de 10 minutes. On peut voir qu'une partie de la chaîne cinématique est représentée. Plusieurs températures sont disponibles pour la génératrice et pour la multiplicatrice. Malheureusement, aucune température n'est mesurée au niveau du roulement principal et des roulements de pales. Ces variables sont parfois disponibles en fonction des modèles de machines, cependant la norme n'impose que peu de variables et la liste du Tableau 4.1 est considérée comme étendue et contient un large nombre de variables par rapport aux variables présentées dans (Yang et al., 2013). Les variables que nous allons utiliser pour créer nos indicateurs de surveillance pour cette base de données sont représentées en gras.

La Figure 4.3 montre la position des différents capteurs mesurant les variables du Tableau 4.1.

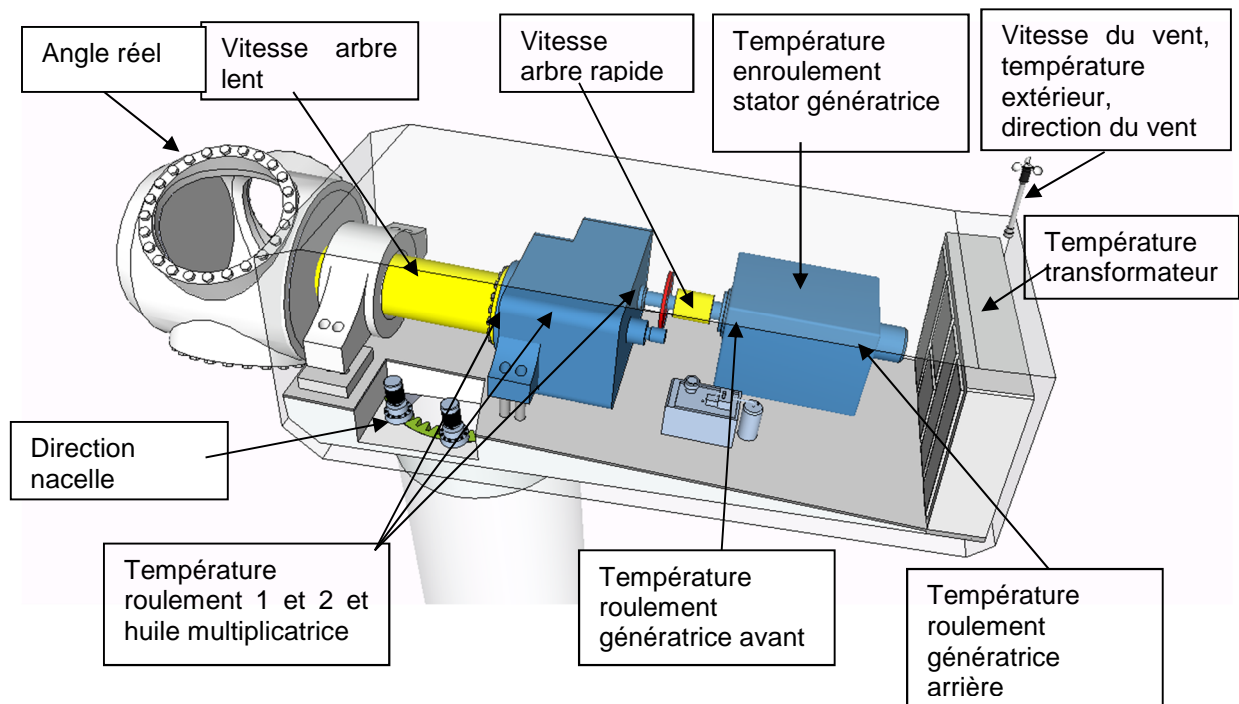


Figure 4.3 Position des capteurs utilisées

Toutes les variables disponibles dans les SCADA des différents parcs ne sont pas rapatriées par VALEMO, pour des raisons logistiques principalement. Le stockage des variables a un coût qui n'était pas justifié avant le début de cette thèse puisque peu d'analyses de ces données étaient faites. Ce parc a été choisi car un grand nombre de ces variables sont déjà rapatriées, ce qui permet d'avoir un historique de données suffisant.

## 4.2. Comportement des variables SCADA

Parmi les variables disponibles dans le SCADA (cf Tableau 4.3), les températures des composants sont utilisées pour la surveillance. Ces températures doivent permettre de détecter des surchauffes et de mettre la machine à l'arrêt automatiquement. Ce sont ces températures qui seront traitées par la suite et utilisées pour la création des indicateurs de surveillance avancée. En effet, ces températures permettent de suivre l'évolution des défauts mécaniques en donnant des informations sur les surchauffes qu'ils engendrent.



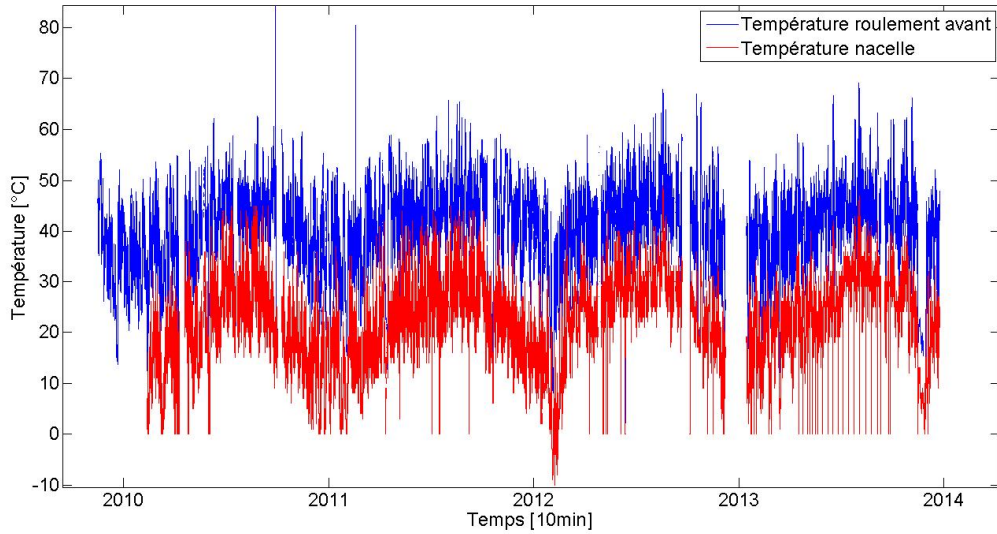


Figure 4.4 Évolution de la température du roulement avant et de la nacelle pendant 4 ans pour une éolienne du parc

La Figure 4.4 montre l'évolution de la température du roulement avant et de la température ambiante de la nacelle d'une éolienne du parc au cours des 4 ans de la base de données. Cette figure permet de constater que la température du roulement varie fortement dans le temps. Une grande partie de cette variation provient de l'impact des saisons au cours de l'année. Il est possible de voir sur cette même figure que la température du roulement et celle de la nacelle sont fortement corrélées. Cette grande variation due à l'influence de l'environnement rend difficile la tâche de détection des défauts. En effet, lors des périodes de faible température extérieure, les défauts risquent d'être masqués par la variation saisonnière normale.

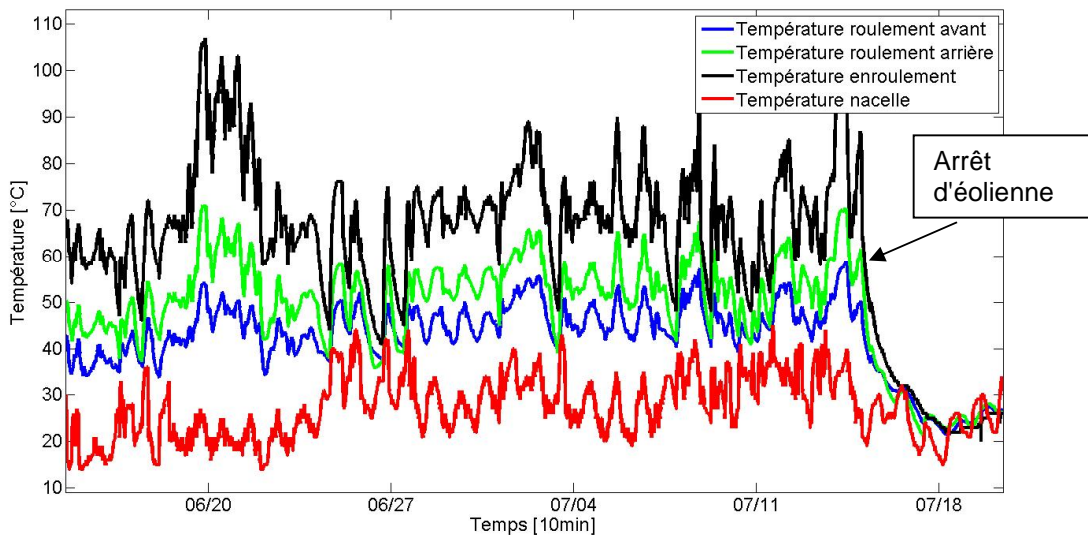


Figure 4.5 Évolution de la température du roulement avant, arrière, de l'enroulement et de la nacelle pendant 1 mois pour une éolienne du parc.

La Figure 4.5 montre l'évolution des températures des roulements avant et arrière, de l'enroulement statorique et de la nacelle pendant un mois. Une fois encore, cette figure permet de constater que toutes ces températures sont fortement corrélées entre elles. De plus, on peut voir que les températures des parties de la génératrice dépendent des conditions de fonctionnement de l'éolienne. On peut voir des périodes de refroidissement dues à des arrêts (07/15) ou des périodes de forte montée en température comme le (06/20). La variation due au fonctionnement rend encore plus difficile la détection puisque cette variation est directement liée à l'évolution stochastique de l'environnement. Certes on peut observer des phénomènes périodiques annuels et quotidiens, mais il y a aussi des variations dues au vent par exemple, qui pour le moment restent impossibles à prévoir.

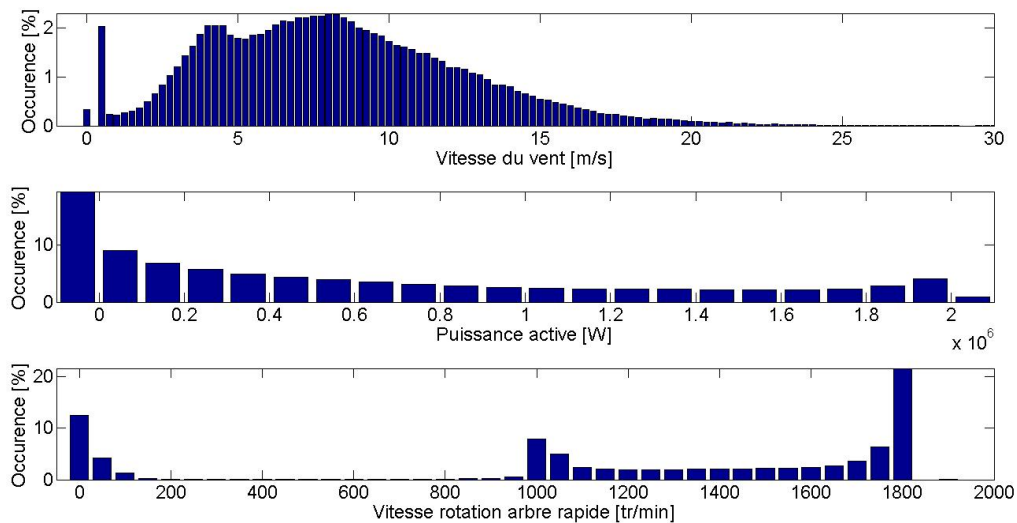


Figure 4.6 Histogramme de la vitesse du vent, de la puissance et de la vitesse de rotation de l'arbre rapide pour les éoliennes du parc pendant 4 ans

La Figure 4.6 décrit la répartition des vitesses du vent, des valeurs de puissance et de la vitesse de rotation de l'arbre rapide durant les 4 ans de la base de données. En ce qui concerne la vitesse du vent, le premier pic proche de 0 m/s correspond à la valeur d'offset de l'anémomètre. Ensuite, la vitesse du vent se répartit autour d'une valeur moyenne de 8 m/s sur ce parc, ce qui en fait un parc très venté.

Sur cette même figure, on peut comparer la répartition de la puissance active et de la vitesse de rotation de l'arbre rapide, qui sont très différentes de celle du vent. On peut constater que les éoliennes génèrent de l'énergie de manière très variable et que, bien que la puissance nominale des éoliennes du parc soit de 2MW, cette valeur est en fait très peu souvent atteinte. En revanche, on peut voir que la vitesse de rotation prend majoritairement des valeurs proches de la vitesse nominale 1800tr/min. Ensuite, on peut voir apparaître les phases 2 et 3 (cf Figure 2.5 de la partie 2.1.2) où la vitesse de rotation varie entre 1000 et 1800 tr/min. Les valeurs au-dessus de 100 et inférieur à 1000 tr/min sont très peu représentées car elles correspondent aux phases de démarrage.

Ce parc possède un taux de charge supérieur à 28% sur ces 4 années. Cela en fait un des parcs les plus productifs de France. En revanche, il s'agit d'un parc très venté et fortement turbulent, avec environ 100 arrêts par machine et par an. Il est intéressant de remarquer que les éoliennes ne sont à l'arrêt qu'environ 20% du temps comme le montre la Figure 4.6. Pendant le reste du temps, la puissance des éoliennes va varier en fonction du vent. Rappelons que les variations des conditions de fonctionnement

et de l'environnement représentent la difficulté majeure de la surveillance du domaine éolien. Elles sont responsables de la grande variabilité des variables utilisées pour créer les indicateurs de surveillances même lors de phase de bon fonctionnement.

### **4.3. Les défauts types d'une éolienne**

Durant la vie d'une éolienne, une multitude de défauts différents peut se produire. Faire une liste exhaustive de ces défauts est presque impossible, puisque toutes les pièces peuvent être source de défaillance. Néanmoins, certains défauts sont similaires et des familles de défauts peuvent émerger. De plus, dans le cas de la surveillance éolienne, connaître précisément l'origine du défaut n'est pas une priorité. Les opérations de maintenance curative se font en général en remplaçant les pièces défectueuses par bloc ou groupe de composants. Dans le cas d'un roulement défaillant, c'est tout le roulement qui est remplacé et non simplement la bille ou le rouleau défectueux.

Dans cette partie, une organisation des défauts types par composant avec, à chaque fois, une présentation des conséquences du défaut sur la machine et les variables qui peuvent être influencées par le défaut est présentée. Tous les défauts ne seront pas listés, seulement ceux qui semblent pouvoir avoir une influence sur les variables SCADA.

Les composants de l'éolienne peuvent être organisés en 6 groupes. Les groupes proposés sont définis, premièrement, par la position des éléments les uns par rapport aux autres. Ensuite, les composants sont groupés avec les autres composants que l'on remplace habituellement ensemble (dépendance structurelle). Finalement, la fonction des composants est prise en compte (dépendance fonctionnelle). L'objectif de cette organisation est de montrer qu'il est possible d'isoler les composants en groupes relativement indépendants.

#### **4.3.1. Le rotor**

Le rotor est composé des 3 pales actionnées et supportées par trois groupes moteurs et 3 roulements fixés au moyeu qui est lui-même supporté par le roulement principal.

Il existe de nombreux capteurs surveillant le comportement du rotor. En fonction des types de SCADA, les signaux rapatriés peuvent être différents et certains défauts ne pourront donc pas être détectés. Le Tableau 4.2 liste des parties du rotor ainsi que les défauts les plus fréquents et les variables mesurées utilisables pour la surveillance de chacune de ces parties.

Tableau 4.2 Défaut fréquent du rotor

Pièce	Défauts fréquents	Variables utilisables
Pale	Fissure, défauts Structurels	
Roulement	Fissure, défauts Structurels	
Actionneur (moteur électrique ou groupe hydrauliques)	Désalignement des pales, surchauffes des actionneurs, système de contrôle de l'angle des pales défaillant	Angles des 3 pales Tension et courant moteurs Températures huile ou moteur Vitesse du vent Puissance active produite Vitesse de rotation arbre lent
Moyeu	Fissure, défauts Structurels	
Roulement principal et arbre lent	Fissure, défauts Structurels	

Pour les pales, le moyeu ou les roulements de pale, la plupart des défauts sont structurels. Aucun capteur du SCADA ne permet d'avoir des informations sur ces défauts. Pour les surveiller, les industriels rajoutent des systèmes CMS dédiés. Ces défauts mécaniques sont souvent rares cependant. En revanche, les défauts du système de commande sont plus fréquents et heureusement moins critiques. Ils impactent la performance globale de la machine, mais leurs réparations ne nécessitent pas des temps d'arrêt importants, contrairement au remplacement d'une pale ou d'un roulement principal. Le système de contrôle permet d'orienter les pales sur elles-mêmes en fonction de la vitesse du vent. C'est ce système qui permet la régulation de la puissance. De plus, en cas d'arrêt d'urgence ou pour cause de grand vent, ce système va orienter les pales pour qu'elles n'aient plus aucune prise au vent.

Un défaut fréquent est la désorientation d'une pale par rapport aux autres. Si cela arrive, cela crée un déséquilibre sur le moyeu, augmente les efforts dans toutes les pièces du rotor, diminue la durée de vie de celles-ci et atteint directement les performances de la machine. La mesure individuelle de l'angle des pales est une variable indispensable pour détecter ce type de problème. Le contrôle des pales peut aussi être atteint. Dans ce cas précis, la courbe de puissance où le lien entre vitesse de rotation de l'arbre lent, vitesse du vent et angle de pale, permettra de surveiller l'apparition de ce défaut. Enfin, si les actionneurs sont impactés, la température du moteur, la tension et le courant semblent pouvoir fournir suffisamment d'informations pour détecter des anomalies.

### 4.3.2. La multiplicatrice

Certaines éoliennes sont équipées de multiplicatrice (ou boîte à engrenage) qui permettent d'augmenter la vitesse de rotation des arbres, pour être plus proche de la vitesse de rotation nominale des génératrices courantes. La multiplicatrice permet de multiplier la vitesse de rotation par un facteur d'environ 100. Elle est composée de plusieurs étages d'engrenages parallèles ou épicycloïdaux.

Le Tableau 4.3 liste des parties du rotor ainsi que les défauts les plus fréquents et les variables mesurées utilisables pour la surveillance de chacune de ces parties.

Tableau 4.3 Défaut fréquent de la multiplicatrice

Pièce	Défauts fréquents	Variables utilisables
Roulement	Fissure, défauts Structurels	Températures roulements Température de l'huile Puissance
Engrenage	Fissure, défauts Structurels	Températures roulements Température de l'huile Puissance
Système de lubrification refroidissement	Défaut de pompe, manque d'huile	Températures roulements Température de l'huile Puissance
Arbre d'entraînement	Désalignement des arbres, balourd, excentricité	Puissance

La multiplicatrice est composée de plusieurs roulements qui soutiennent les arbres d'entraînement. La température de certains de ces roulements est surveillée par des capteurs dédiés. Dans les modèles récents de multiplicatrice, tous les roulements sont équipés d'un capteur. Cette température va permettre de mettre en évidence les échauffements anormaux du roulement. Les défauts mécaniques créent des frottements supplémentaires qui viennent ajouter des échauffements. Ces surchauffes sont plus ou moins importantes en fonction du niveau de dégradation. Si les températures de roulements ne sont pas disponibles alors la température de l'huile peut donner des informations sur la surchauffe diffusée dans toute la multiplicatrice. Finalement, si aucune température n'est disponible, une analyse des baisses de performance de la machine en utilisant la puissance peut permettre de détecter une anomalie. Cependant, pour que le défaut fasse augmenter significativement l'huile ou baisser les performances, il faut qu'il soit très avancé et c'est donc souvent trop tard pour intervenir.

Les engrenages, tout comme les roulements, peuvent être soumis à des usures prématurées et des défaillances. Cependant, la surveillance est ici plus compliquée puisqu'il n'y a pas de capteurs de température installés sur ces engrenages. Par ailleurs, il paraît difficile de mesurer la température des engrenages puisque ce sont des parties mobiles et que la surface échauffée n'est pas accessible. C'est donc la température des composants ou de l'huile qui va permettre d'avoir une information sur l'échauffement, ou encore une perte de puissance. Encore une fois, si la surchauffe se transmet sur les autres composants ou impacte la production, c'est que la défaillance est déjà très avancée.

Le système de refroidissement ou le système de lubrification vont impacter plusieurs composants lorsqu'ils sont défaillants. Les capteurs de température des roulements et de l'huile vont permettre de détecter ce type de défaillance. Dans certains cas, le système de distribution est équipé d'un capteur de pression et de température, mais cela est très rare. Évidemment, si ces capteurs sont présents, ils fourniront des informations précieuses qui aideront à la détection de défaillance.

Finalement, les arbres de la multiplicatrice peuvent aussi être affectés. En revanche, le seul moyen pour détecter ce genre de défaut avec les données SCADA est de surveiller les pertes de production. Encore une fois, si de telles pertes sont détectées, cela signifie que le défaut est déjà très important.

### 4.3.3. La génératrice

La génératrice est reliée directement à l'arbre lent dans le cas d'éolienne à entraînement direct et pour les autres machines, elle est reliée à l'arbre rapide de sortie de la multiplicatrice. L'arbre d'entraînement est soutenu par 2 roulements de chaque côté de la génératrice. L'ensemble est refroidi par un bloc de ventilation qui fait circuler l'air dans toute la génératrice.

Le Tableau 4.4 liste des parties de la génératrice ainsi que les défauts les plus fréquents et les variables mesurées utilisables pour la surveillance de chacune de ces parties.

Tableau 4.4 Défaut fréquent de la génératrice

Pièce	Défauts fréquents	Variables utilisables
Roulement	Fissure, défauts Structurels	Températures roulements Puissance
Enroulement statorique et rotorique	Court-circuit, déséquilibre entre phase,	Températures roulements Température enroulement statorique Puissance
Système de refroidissement	Défaut de ventilateurs ou d'extraction d'air	Températures roulements Température de l'huile Puissance
Arbre d'entraînement	Désalignement des arbres, balourd, excentricité	Puissance

Comme pour la multiplicatrice, la température des roulements est mesurée par des capteurs. Ces capteurs vont permettre de détecter les surchauffes provenant d'une usure des roulements. La puissance peut aussi être utilisée si les roulements sont dans un état de détérioration avancé.

Les enroulements statoriques et rotoriques peuvent être déséquilibrés, ce qui peut affecter les performances de l'éolienne. Dans certains cas, les isolants peuvent être détériorés, ce qui peut causer des courts-circuits et mettre la génératrice hors d'usage. La durée de vie des isolants est très affectée par les surchauffes, les températures des enroulements et des roulements peuvent permettre de mettre en évidence des échauffements anormaux et prévenir l'usure prématurée des isolants.

La défaillance du système de refroidissement impacte la température de toutes les parties de la génératrice. La température des roulements et enroulements permettra donc de mettre en évidence une défaillance du refroidissement.

Finalement, comme pour la multiplicatrice, un défaut sur les arbres peut être détecté si celui-ci est suffisamment important pour impacter la production de manière significative.

### 4.3.4. Le convertisseur et le transformateur

Le convertisseur et le transformateur sont des composants supplémentaires qui permettent la régulation de la production électrique avant de l'injecter sur le réseau. Néanmoins, il s'agit de composants électriques et ils sont très peu surveillés par le SCADA. Dans certains cas, toutefois, des SCADA proposent le rapatriement de mesures de température de chaque phase du transformateur ou de la température à l'intérieur du convertisseur. Certaines valeurs de tensions et courants sont relevées, mais l'échantillonnage de 10 minutes rend impossible le suivi du comportement électrique.

La température du convertisseur ou de chacune des phases du transformateur permet de surveiller que leurs systèmes de refroidissement respectifs fonctionnent correctement. Il paraît très difficile de surveiller des défauts électriques grâce aux températures.

#### **4.3.5. La station météorologique**

Finalement, la station météorologique mesurant la vitesse du vent, la direction du vent et de la nacelle ainsi que la température extérieure peuvent aussi être surveillées. Dans un premier temps, la courbe de puissance peut permettre de constater des mesures de vent erronées. En effet, si une augmentation des performances est détectée, cela peut être synonyme d'une erreur de mesure sur la vitesse du vent. Ensuite, la direction du vent et de la nacelle peut permettre de détecter des erreurs d'alignements, ce qui est très important pour un industriel, puisqu'un désalignement de quelques degrés peut diminuer de quelques pourcents les performances machines.

### **4.4. Défauts présents durant la période étudiée**

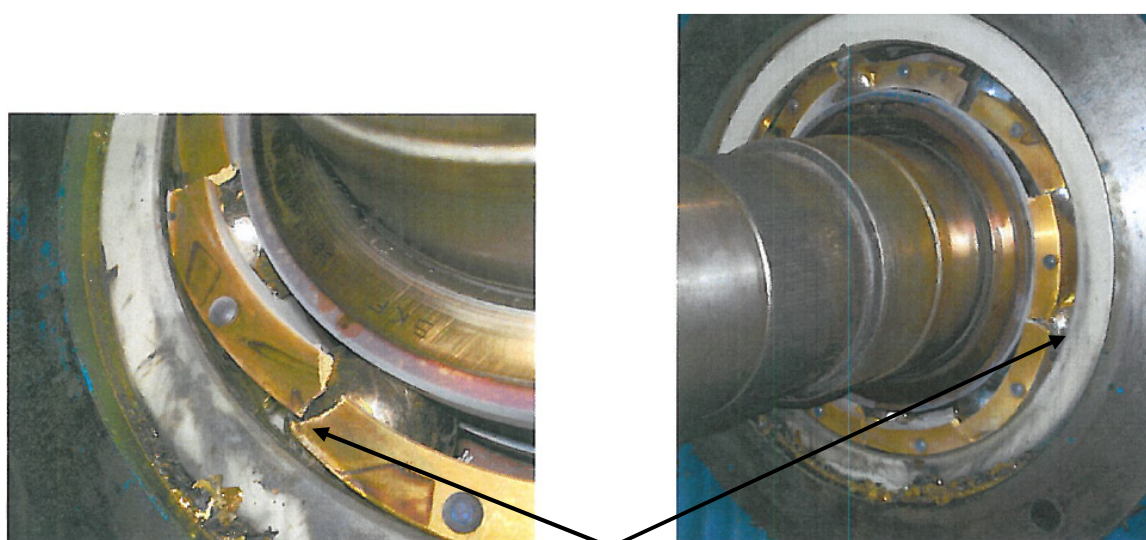
Durant les 4 ans représentés dans la base de données, une multitude de défaillances se sont produites. Isoler les défauts ayant conduit à ces défaillances est obligatoire pour comprendre l'origine de ces défaillances et surtout étudier l'impact des défauts sur les variables et donc les indicateurs de surveillance. Malheureusement, le nombre de défaillances est tellement important qu'il est très difficile de créer des plages de bon fonctionnement si l'on considère la totalité de l'éolienne. Par conséquent, dans nos travaux, nous faisons le choix d'isoler la génératrice et d'étudier exclusivement ce composant. Seuls les défauts impactant la génératrice seront donc étudiés.

On considère par la suite que la génératrice est un système isolé qui n'est pas impacté par les défauts des autres composants. Cette hypothèse est justifiée par le fait que les composants sont suffisamment éloignés les uns des autres et qu'il n'y a pas eu défaut sur les arbres reliant la génératrice à la multiplicatrice, ni de défaut sur le système de commande de la génératrice. Par conséquent, aucun composant n'a eu d'influence sur la génératrice.

Ce choix est motivé par le fait que la génératrice est un des composants les plus surveillés. En effet, il y a un capteur de température pour chaque élément la composant. De plus, les variables décrivant les conditions de fonctionnement comme la puissance générée ou la vitesse de rotation des arbres, permettent de décrire tout le comportement de la génératrice. Finalement, ce choix est aussi motivé par le fait qu'il y a eu 8 défauts durant les 4 ans de données impactant directement la génératrice. Même si ce nombre semble faible, il existe peu de bases de données comprenant autant de défauts sur un même parc. De plus, 4 de ces défauts sont considérés comme majeurs puisqu'ils ont donné lieu au remplacement de la génératrice. Ces opérations de remplacement étant critiques, l'objectif de cette étude sur cette base de données est de déterminer si ces défauts peuvent être détectés et d'évaluer l'impact que la surveillance pourrait avoir sur les opérations de maintenance de VALEMO.

Ainsi, il s'est produit 8 défauts affectant la génératrice pendant les 4 années formant la base de données : 2 défauts roulements, 2 défauts d'enroulement et 4 défauts du système de refroidissement.

- Défauts de roulement** : Pendant la période étudiée, 2 roulements de génératrice sont tombés en panne. D'abord le roulement arrière sur l'éolienne WT6 puis plusieurs mois après, le roulement avant sur l'éolienne WT9. Ces deux défaillances sont considérées comme critiques puisque les opérations de remplacements peuvent être longues et le prix de ces composants est très élevé, supérieur à 50k€. De plus, ces événements ont eu pour conséquence la casse totale de tout le roulement. Dans le cas de l'éolienne WT6, c'est la génératrice complète qu'il a fallu changer puisque le logement du roulement était tellement endommagé que celui-ci n'a pas pu être séparé de la génératrice. La Figure 4.8 montre une photo du roulement arrière WT6 prise pendant les opérations de réparation. On peut voir sur cette figure que la bague extérieure du roulement est complètement détruite et qu'il n'est pas réparable. Il a fallu un mois pour la remplacer à cause de délai d'approvisionnement et de disponibilité des équipes. En ce qui concerne le roulement avant de l'éolienne WT9, seul le roulement a été changé au bout de 2 semaines.



Bague extérieur Brisée

Figure 4.7 Photo de la défaillance du roulement arrière WT6

Ces défauts de roulements sont dus, dans les deux cas, à un problème de lubrification qui a engendré une usure prématurée des surfaces. En règle générale, les défauts de roulements sont lents et progressifs, ils s'étendent sur plusieurs semaines. L'usure de la surface des pièces du roulement va créer des frottements supplémentaires qui vont engendrer une élévation de la température due aux frictions. Le défaut étant progressif, on s'attend à ce que la surchauffe du roulement augmente avec le temps et l'usure. La surchauffe reflète donc l'usure jusqu'à la casse du roulement.

La génératrice est équipée de capteurs de température sur les roulements avant et arrière. L'étude de ces températures montre une élévation nette et un changement de comportement quelques heures avant le défaut comme le montre la Figure 4.8 qui représente l'évolution de la température du roulement avant pendant le défaut de la turbine WT9. La présence visible de l'influence de la défaillance sur les données semble indiquer que ces défauts de roulements sont détectables dans des stades moins avancés. Après discussion avec les experts de VALEMO, on peut conclure que ces



défauts doivent être détectés au moins une semaine à l'avance pour qu'un industriel puisse modifier sa stratégie de maintenance.

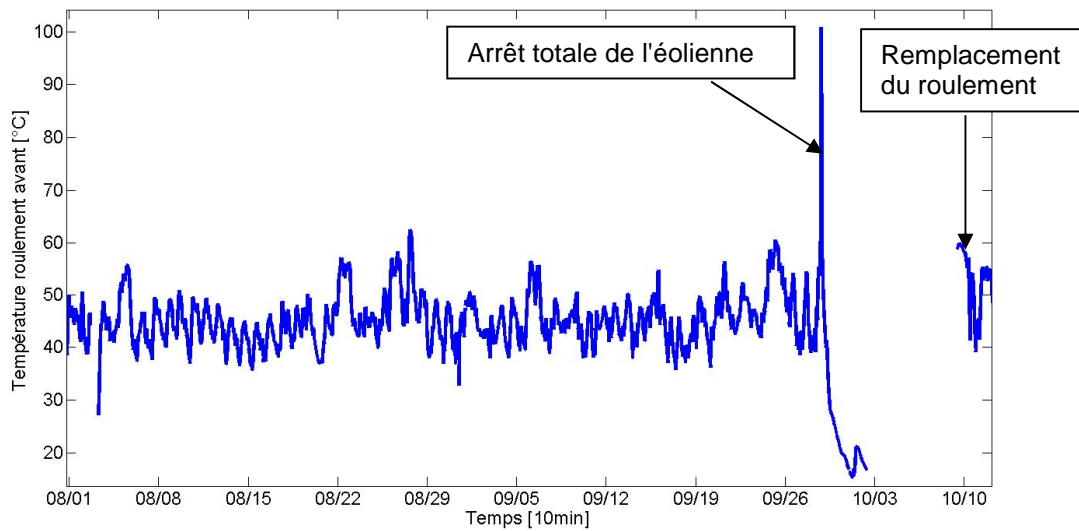


Figure 4.8 Évolution de la température du roulement avant lors du défaut de roulement de l'éolienne WT9

- Défauts d'enroulement :** Deux autres défauts ont eu lieu sur les éoliennes WT11 puis WT9 à quelques mois d'intervalle. Ici, il s'agit d'une défaillance au niveau des enroulements rotoriques de la génératrice. Un court-circuit entre 2 enroulements a détruit ces derniers et rendu la génératrice inutilisable. Comme pour les roulements, ces défaillances sont critiques puisque la génératrice complète doit être changée. Encore une fois, le prix élevé de 250k€ de la génératrice et la durée du remplacement peuvent avoir un impact considérable sur la rentabilité du parc. Ici aussi les nouvelles génératrices ont été remplacées en deux semaines comme le montre la Figure 4.9 qui représente l'évolution de la température du roulement arrière pendant le défaut d'enroulement de la turbine WT9. Contrairement au défaut de roulement, aucun signe avant-coureur n'est visible sur les températures.

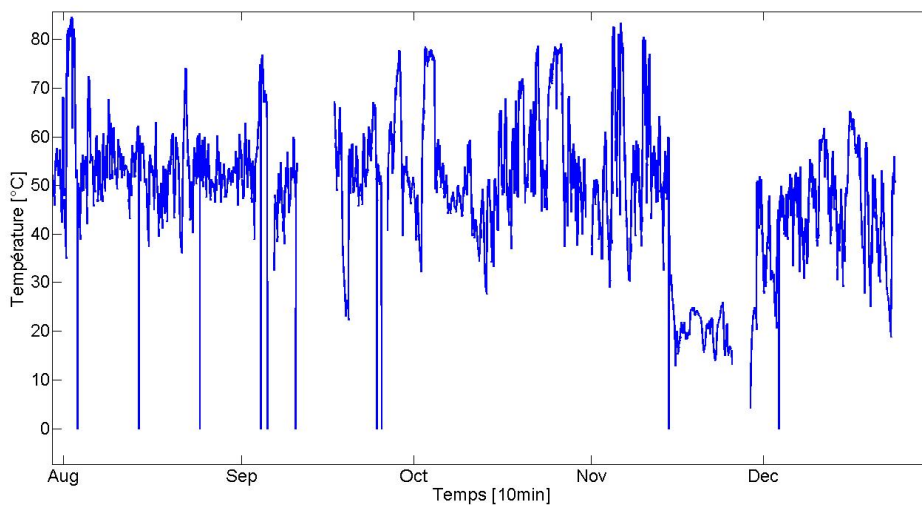


Figure 4.9 Évolution de la température du roulement arrière lors du défaut d'enroulement de l'éolienne WT9

Ces défauts ayant eu lieu de manière rapprochée sur deux éoliennes différentes, il est légitime de penser qu'ils ont la même origine. Les rapports d'expertises n'ont pas conclu de manière précise sur l'origine des défauts. En revanche, VALEMO suppose fortement que cela proviendrait des surchauffes à répétitions sur les éoliennes. Ces surchauffes ont pu diminuer fortement la qualité de l'isolant entre les enroulements jusqu'à créer un court-circuit.

Une première étude des tensions et des courants de la génératrice a été réalisée, mais aucun signe avant-coureur (ne serait-ce même qu'une heure avant le défaut) ne semble apparaître dans ces signaux. En revanche, une surchauffe apparaît bien au niveau de la mesure de température des enroulements statoriques. Il n'y a pas de capteur de température au niveau du rotor, le capteur de température statorique est le plus près du défaut. Bien que la température du stator augmente quelques heures avant le défaut, les experts semblent beaucoup plus sceptiques à l'idée d'être capables de détecter ces défauts plus tôt. Ce défaut est considéré comme un court-circuit électrique qui n'est donc pas progressif et apparaît brutalement. Il semble donc difficile de le surveiller pour le détecter en avance. Néanmoins, ici ces courts-circuits sont dus à une destruction prématurée des isolants. C'est l'apparition d'une surchauffe que l'on va chercher à surveiller et non l'apparition d'un court-circuit.

- **Défaut du système de refroidissement** : Finalement, la base de données contient aussi des défauts du système de refroidissement. Ces défauts sont évidemment moins critiques que les autres, puisqu'ils n'arrêtent pas la machine et leur réparation est faite rapidement. Cependant, ils engendrent des surchauffes qui, à long terme, diminuent la durée de vie des composants. Même si l'urgence à intervenir est moindre, il faut limiter ce type de configuration. De plus, la surchauffe engendrée par un défaut de refroidissement va empêcher toute surveillance des composants impactés puisque les valeurs de températures mesurées seront faussées.

La base contient 4 défauts de ventilateur identiques sur 4 éoliennes dans l'ordre d'apparition, la WT9, WT7, WT6 et WT10. Dans chacun des cas, il s'agit d'une vis de fixation qui s'est desserrée jusqu'à ce que l'entraînement du ventilateur ne se fasse plus. Le ventilateur ne refroidissant plus les composants, la température augmente. Les ventilateurs influençant les températures des différentes parties de la génératrice, la surchauffe doit pouvoir être visible sur plusieurs capteurs de température en même temps. Ces défauts ne sont pas progressifs, une fois la vis desserrée, le ventilateur ne fonctionne plus du tout.

Les températures mesurées au niveau des génératrices des éoliennes WT9, WT7 et WT10 augmentent fortement avant la réparation du défaut. Seul le défaut sur WT6 semble être plus difficile à détecter. L'objectif ici n'est pas de détecter les défaillances avant qu'elles n'atteignent un niveau avancé qui arrêterait l'éolienne, mais de détecter le défaut au plus tôt après que la vis se soit desserrée. Ces défauts de ventilateur peuvent parfois passer inaperçus avec la surveillance SCADA classique. Il s'agit ici d'éviter les surchauffes au maximum pour protéger la durée de vie des composants sur le long terme.

## 4.5. Conclusion

La base de données étudiée contient 8 défauts impactant la génératrice, dont 4 majeurs ayant nécessités des interventions lourdes. Ces 4 défauts, les défauts de roulements et d'enroulement sont appelés « défauts critiques » par la suite. Même si cette base de données ne contient pas tous les types de défauts possibles, elle permet de valider les méthodes pour la surveillance d'un composant critique de l'éolienne. La base de données s'étend sur 4 ans, ce qui permet de valider la pertinence des indicateurs en termes de bonnes détections et de fausses alarmes. Cette base de données a cependant plusieurs limites, car elle ne se concentre que sur la génératrice et que sur un seul parc. Cette base de données ne permettra pas à elle seule de généraliser les conclusions sur les méthodes étudiées à d'autres composants, ni à d'autres modèles d'éoliennes ou d'autres lieux géographiques soumis à des conditions environnementales différentes.



## **5. Méthodes proposées pour la création et l'évaluation d'indicateurs**

### **5.1. Introduction**

Dans cette partie sont présentées les différentes méthodes de création d'indicateurs retenues dans cette thèse. Il est proposé de combiner les avantages des approches mono-turbines peu sensibles aux variations des conditions de fonctionnement et les approches multi-turbines peu sensibles aux variations des conditions environnementales. Plusieurs approches combinées sont proposées et comparées avec des méthodes mono-turbines et multi-turbines plus classiques.

Ces nouvelles solutions ont été pensées afin de mieux répondre aux besoins des industriels. Par conséquent, la comparaison de celles-ci est faite sur la base de données réelle présentée dans le chapitre précédent. De plus, afin de montrer les limites des précédentes méthodes et de montrer l'intérêt de combiner les approches, une méthode d'évaluation s'appuyant sur 4 critères a aussi été développée et validée par VALEMO.

Ces critères ont été choisis ou construits pour prendre en compte au mieux les contraintes industrielles. Deux critères classiques permettent de mesurer la capacité de détection des indicateurs : le temps de détection avant la panne et le taux d'interventions inutiles calculé à partir du taux de fausses alarmes. Ensuite, deux autres critères plus proches de l'utilisateur s'intéressent à la disponibilité de l'indicateur, et à sa persistance.

Le plan de ce chapitre est le suivant. Dans un premier temps, nous présenterons les critères que nous avons retenus, nous permettant d'avoir une évaluation des performances des méthodes prenant en compte les besoins des industriels. Dans un deuxième temps, nous présenterons la démarche utilisée pour fixer les seuils de détection sur chaque indicateur puis nous présenterons les différents indicateurs que nous souhaitons évaluer : les indicateurs mono-turbines, inspirés de l'étude bibliographique menée sur le sujet, et la fusion de ces indicateurs dans une approche multi-turbines, qui constitue notre contribution au domaine de recherche.

### **5.2. Critères proposés pour évaluer les performances**

Les industriels comme VALEMO ont besoin de méthodes de surveillance automatisées capables de générer des alarmes lors de défaillance. Cette détection doit se faire le plus tôt possible pour pouvoir adapter la stratégie de maintenance et éviter des coûts liés à la perte de production ou aux changements d'organisation de la maintenance. La création d'un seuil de détection est une tâche délicate puisque les indicateurs ne sont pas parfaits et varient même lorsqu'il n'y a pas de défaut. Un compromis doit donc être fait entre la détection de défauts réels et le nombre de fausses alarmes que le système va générer. Peu d'articles montrent comment faire ce compromis dans la pratique, compte tenu du comportement fortement non stationnaire des éoliennes. Les articles de la littérature se contentent de vérifier la capacité d'un indicateur à détecter la présence d'un défaut avéré, sans s'intéresser à l'évolution de l'indicateur sur des périodes sans défaut. Ainsi, dans cette thèse, en plus de vérifier la capacité d'un

indicateur à détecter un défaut avéré et de mesurer *l'avance à la détection*, nous proposons trois autres critères d'évaluation :

- Le nombre de fausses alarmes, que nous convertissons en *nombre d'interventions inutiles*, est indispensable pour mettre en perspective les capacités de détection.
- Le concept de *disponibilité* d'un indicateur est aussi important pour des industriels. En effet, ils doivent pouvoir surveiller leurs équipements dans n'importe quelle situation toute l'année. La mise en œuvre de certaines approches oblige à supprimer des échantillons de mesures qui sont considérés comme inutilisables pour les besoins de la méthode. Ces points supprimés peuvent parfois être nombreux. La disponibilité évalue le temps pendant lequel l'indicateur peut effectivement être calculé.
- Le dernier critère est la *persistance* de l'indicateur. Il traduit la capacité du système à continuer d'indiquer la présence d'un défaut jusqu'à la réparation, à partir du moment où il a commencé à indiquer ce défaut. Il mesure le temps pendant lequel l'indicateur reste au-dessus du seuil de détection pendant une période en défaut. Cette persistance permet aux opérateurs d'avoir plus confiance dans le système. Si le système ne fait que changer d'avis, les opérateurs peuvent douter de sa pertinence.

Le calcul de ces différents critères nécessite la définition de plages de bon fonctionnement et de plages de défaut. Une plage de bon fonctionnement est une période de temps continue pendant laquelle aucun défaut n'est apparu sur l'éolienne. Une plage de défaut est une période de temps continue pendant laquelle un défaut était présent sur l'éolienne. La fin de la plage de défaut correspond à l'instant de panne, le moment où l'impact du défaut est tellement important que la machine ne peut plus fonctionner. L'éolienne est alors à l'arrêt, en attente d'être réparée puis redémarrée.

### 5.2.1. L'avance à la détection

L'avance à la détection mesure le temps qui sépare l'instant où le défaut a été détecté par une méthode de détection de l'instant de panne. L'instant où le défaut a été détecté par une méthode est le moment où l'indicateur franchit le seuil de détection pour la première fois au cours de la période de défaut. L'avance à la détection s'exprime en périodes d'échantillonnage. La Figure 5.1 illustre le principe d'avance à la détection. Cette figure représente un indicateur pendant une période de défaut avec un seuil de détection. La durée entre le premier dépassement et la fin de la période correspondant à la panne totale de l'éolienne est l'avance à la détection.

Notons que ce critère évalue la marge de manœuvre qu'aurait eue un industriel comme VALEMO pour intervenir avant la panne. Après concertation avec des experts, il est estimé qu'il faut détecter le défaut au moins une semaine en avance pour pouvoir modifier la stratégie de maintenance. Cette limite d'une semaine correspond à la limite minimum pour avoir le temps d'organiser une inspection et mettre la turbine en sécurité. Cela peut être, par exemple, mettre la turbine à l'arrêt ou limiter sa production.

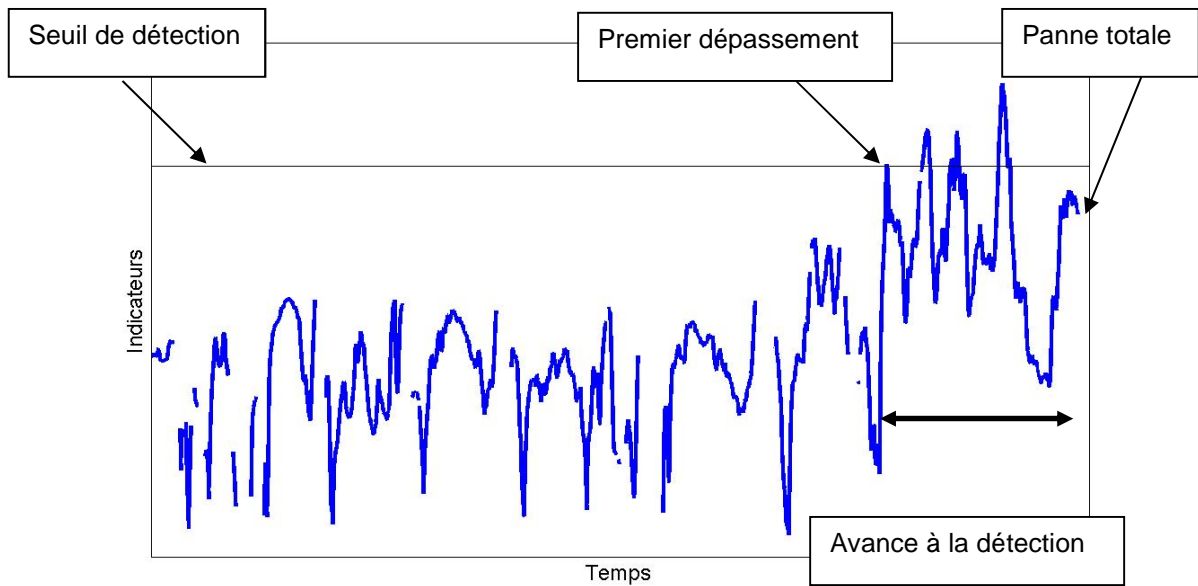


Figure 5.1 Schéma de principe de l'avance à la détection, avec un indicateur sur une plage de défaut

### 5.2.2. Le nombre d'interventions inutiles

Une fausse alarme correspond à un dépassement de seuil pendant une plage de bon fonctionnement. Elle indique aux opérateurs la présence d'un défaut alors qu'il n'y en a pas. Afin de se rapprocher de l'utilisation réelle d'un système de décision et de mieux mesurer le coût pour les opérateurs de surveillance, le taux de fausses alarmes est converti en nombre d'interventions inutiles. On définit comme jour de fausses alarmes une journée d'une période de bon fonctionnement pendant laquelle l'indicateur a franchi au moins une fois le seuil de détection c'est-à-dire qu'au moins une fausse alarme a été déclenchée. On définit comme période de fausses alarmes une période de temps composé de jours de fausses alarmes consécutifs. Une période de fausses alarmes a une durée minimale de un jour. On suppose qu'une équipe de maintenance se déplace une fois à chaque période de fausses alarmes. Le nombre d'interventions inutiles est donc le nombre de périodes de fausses alarmes. La Figure 5.2 illustre la définition des périodes de fausses alarmes. Cette figure représente un indicateur pendant une période sans défaut avec 2 exemples de seuil de détection. Avec le seuil A, seuls les jours 1 et 3 sont des jours de fausses alarmes, car au moins un échantillon 10 minutes est au-dessus du seuil pendant ces journées. Comme le jour 2 n'est pas un jour de fausses alarmes, le seuil A crée deux périodes de fausses alarmes séparées et donc 2 interventions inutiles. En revanche, avec le seuil B, les jours 1, 2 et 3 sont des jours de fausses alarmes donc il génère une période de fausse alarme continue allant de la journée 1 à 3 et donc une seule intervention inutile.



Figure 5.2 Schéma de principe d'une période de fausses alarmes avec un indicateur dans une période sans défaut

Le nombre d'interventions inutiles doit être minimum. Il est évident qu'un système parfait sans fausses alarmes, mais qui permet de détecter tous les défauts est un impossible. Après concertation avec VALEMO, un nombre maximum d'interventions inutiles a pu être défini à 1 par mois et par parc. Ce nombre peut paraître faible, mais au regard du nombre croissant d'éoliennes dans un parc, il peut devenir très important. VALEMO possède plus de 140 éoliennes en exploitation répartie sur 27 parcs allant de 3 à 8 machines. Cela va donc représenter un peu plus d'une alarme par jour ( $27 \text{ parcs} / 30 \text{ jours} = 0.90$ ), ce qui peut mobiliser une personne à plein temps, compte tenu du temps nécessaire aux vérifications nécessaires et générer un surcoût important. La limite d'une intervention inutile par mois pour un parc est une limite maximale à ne pas dépasser. Dans cette thèse, le critère de nombre d'interventions inutiles est donc un élément très important pour l'évaluation des différentes méthodes.

### 5.2.3. La disponibilité

Certaines méthodes nécessitent l'élaboration de modèles dont l'utilisation n'est valide que sous certaines conditions de fonctionnement, L'indicateur ne peut pas toujours être calculé. De plus, si ces conditions de validité sont trop contraignantes alors il est possible de ne pas pouvoir calculer l'indicateur et donc ne pas pouvoir surveiller les éoliennes pendant de trop longue période de temps. Il est donc important de savoir si l'indicateur définit théoriquement peut en pratique être implémenté avec des données de terrain et s'il peut effectivement permettre la surveillance en continue des installations.

On définit la disponibilité d'un indicateur comme le rapport du nombre de périodes d'échantillonnage pendant lesquelles cet indicateur peut être calculé sur le nombre de périodes d'échantillonnage total.

Une autre manière d'évaluer la disponibilité est d'analyser la distribution des périodes d'indisponibilité. Ne pas avoir de réponse du système pendant dix minutes ou une heure n'est pas important par rapport à une journée, une semaine ou plus.

En règle générale, les conditions de production où le modèle est valide sont plus difficiles à réunir lors de périodes de forte production où les conditions de fonctionnement sont particulièrement variables. Par exemple, en hiver, le vent est très turbulent, mais porteur d'une grande quantité d'énergie. Une forte



production est synonyme de forte probabilité de défaut. Les pré-traitement à effectuer sur les données permettant de garantir la validité des modèles risquent de supprimer plus d'échantillons lors de ces périodes alors que c'est précisément dans ces périodes que les systèmes de surveillances sont nécessaires. Ainsi, cette notion de disponibilité est importante pour un industriel. À taux de nombre d'interventions inutiles égales, on préférera un système capable de surveiller le système pendant des périodes plus longues.

#### **5.2.4. La persistance**

La persistance évalue la durée pendant laquelle l'indicateur reste au-dessus du seuil de détection pendant une période de défaut. Elle est mesurée par le rapport entre les échantillons au-dessus du seuil et la totalité des échantillons d'une période de défaut.

La persistance de l'indicateur est un critère moins classique que les 3 autres, il est conçu pour refléter la confiance qu'un opérateur peut avoir dans l'indicateur une fois qu'il s'est déclenché. L'analyse des indicateurs se fait de manière quotidienne. Si une alarme est générée mais disparaît rapidement, l'opérateur peut croire à une fausse alarme. Cependant, si l'indicateur a tendance à persister au-dessus du seuil pendant une période longue alors l'opérateur y suspectera plus facilement une anomalie et déclenchera les analyses supplémentaires ou directement une intervention. Ainsi, ce critère permet de différencier les indicateurs très fluctuants d'autres plus stables. De plus, les variations des conditions de fonctionnement ou environnementales de la machine ont un fort impact sur l'influence du défaut sur les indicateurs. Ces sollicitations variant énormément, ce critère va permettre de différencier des indicateurs influencés par ces phénomènes non-stationnaires des autres.

### **5.3. Création des seuils**

Tous les critères proposés dans le paragraphe précédent seront calculés pour chaque indicateur. Ces critères dépendent d'un seuil de détection, qu'il faut fixer pour chaque indicateur. Dans cette partie, nous présentons les méthodes retenues pour fixer un seuil de détection et ainsi évaluer et comparer leurs performances de manière aussi objective que possible. Notre objectif est d'avoir une évaluation des performances de chaque indicateur au niveau du parc. Ainsi, par exemple, le critère d'interventions inutiles pour un indicateur et un seuil donné est le nombre d'interventions sur l'ensemble du parc, donc pour toutes les machines, pendant la durée de la base de données soit 4 ans. Cela nécessite de fixer un seuil de détection unique par indicateur, dont la valeur est la même pour toutes les éoliennes du parc. Ce choix peut se justifier par le fait que l'impact d'un défaut sur un indicateur dépend du défaut lui-même et est identique sur chaque machine du parc.

#### **5.3.1. Interventions inutiles versus avance à la détection**

Les méthodes peuvent être comparées sur la base de deux critères simultanés, le nombre d'interventions inutiles, et l'avance à la détection par le moyen d'une courbe ROC adaptée (de l'anglais Receiver Operating Characteristic (Swets, 1996; Van Trees, 2001)). En effet, ces 2 critères sont influencés par le seuil de détection. L'augmentation du seuil de détection génère une diminution des fausses alarmes ainsi qu'une diminution de l'avance à la détection. La courbe ROC permet d'explorer les deux critères simultanément pour un ensemble de valeurs de seuils. Une courbe ROC adaptée est construite comme décrite dans la Figure 5.3.

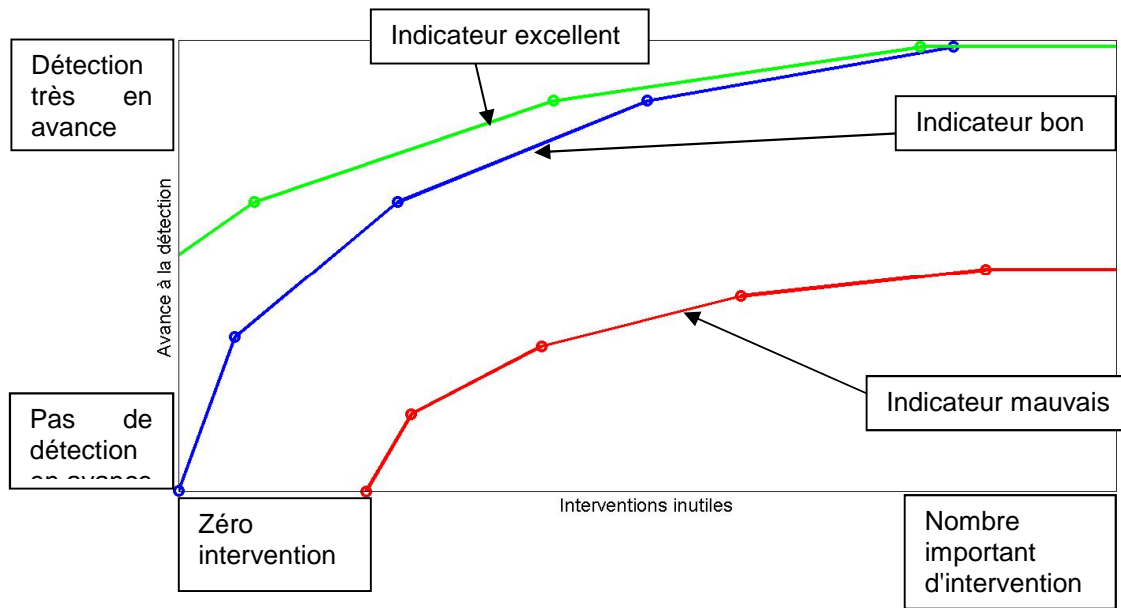


Figure 5.3 Schéma de principe de la courbe ROC adaptée

Pour un défaut donné et pour une valeur de seuil, le nombre d'interventions inutiles et l'avance à la détection sont calculées. Cela permet de construire chaque point de la courbe. Un ensemble de seuils est calculé avec des valeurs qui varient entre la valeur minimum et maximum de l'indicateur, mesurées sur la base de données que l'on cherche à évaluer. Le pas est défini pour avoir une résolution acceptable pour l'analyse en fonction des indicateurs. Attention, contrairement à une courbe ROC traditionnelle, l'abscisse n'est pas un taux de fausses alarmes et l'ordonnée n'est pas un pourcentage de détection. La position par rapport à une diagonale ne peut pas être interprétée comme une indication sur la performance de la méthode. La diagonale ne correspond plus à une méthode qui prendrait une décision au hasard. Sur la Figure 5.3, on peut voir que 3 courbes sont représentées. Chaque point d'une courbe correspond à la performance obtenue pour un seuil parmi l'ensemble des seuils calculés. La courbe bleue peut correspondre à un bon indicateur, la courbe verte à un indicateur excellent et la courbe rouge à un indicateur ne fournissant pas de résultats suffisants. L'objectif est d'avoir un indicateur générant une courbe ROC la plus haute et la plus à gauche possible pour avoir le minimum d'interventions inutiles tout en ayant le maximum d'avance à la détection.

De la même manière, le critère de persistance est calculé pour l'ensemble des valeurs de seuil et la courbe reliant la persistance et le nombre d'interventions inutiles peut être construite. En effet, la persistance dépend de la valeur du seuil. Un seuil très élevé risque de ne détecter que de fortes augmentations de l'indicateur sur des périodes très courtes, ce qui se traduit par une persistance très faible. Un seuil trop bas, voire nul, entraîne une persistance très forte. L'objectif de ce critère de persistance est de faire la différence entre une détection brève et une détection durable. Il est important de remarquer que la courbe ROC est construite avec l'avance à la détection qui correspond au premier moment où l'indicateur dépasse le seuil pendant la période de défaut. Il est possible que ce premier moment soit aussi le seul avant la panne totale. La persistance permettra de détecter de tels comportements, en plus d'une étude graphique des indicateurs pendant les périodes de défaut. La Figure 5.4 montre deux indicateurs ayant la même avance à la détection. On peut voir que l'indicateur supérieur reste au-dessus du seuil pendant toute la durée de la période et a une persistance élevée

alors que l'indicateur inférieur diminue et remonte au-dessus du seuil ensuite et à donc une persistance plus faible. Les deux indicateurs détectent le défaut au même moment et auront donc la même avance à la détection. De plus, dans le cas où ces deux indicateurs génèrent un nombre d'interventions inutiles similaire alors il sera très difficile de les différencier sur la courbe ROC puisqu'ils auront les mêmes valeurs en abscisse et en ordonnées. Cependant, comme on peut le voir ces deux indicateurs ont des comportements très différents et on préférera l'indicateur supérieur avec une plus grande persistance. La courbe ROC seule n'est donc pas suffisante, il faut aussi regarder la persistance et aussi le comportement de l'indicateur graphiquement.

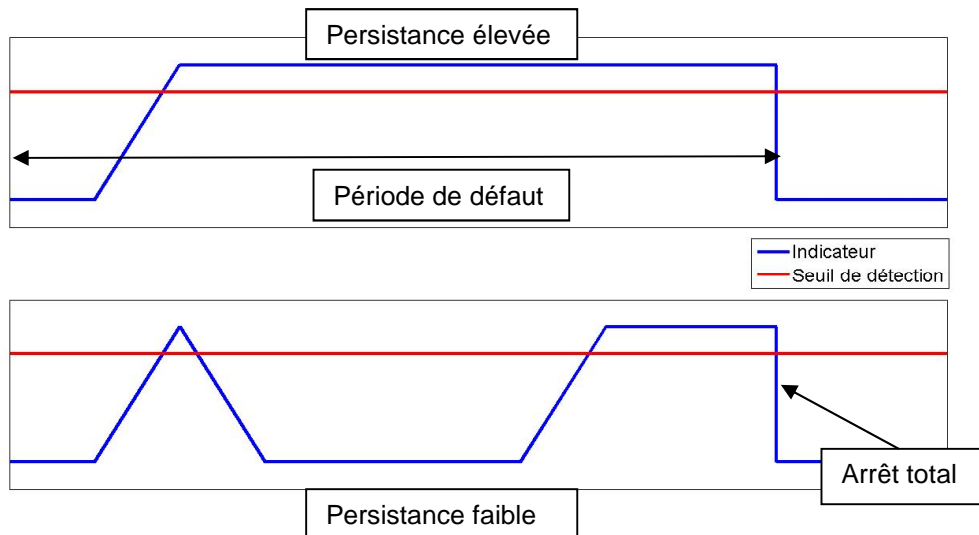


Figure 5.4 Impact de la persistance sur la courbe ROC

La courbe ROC proposée est construite pour comparer les méthodes sur certaines plages de valeurs considérées comme pertinentes. En effet, les courbes ROC proposées ont en ordonnée des valeurs d'avance à la détection qui vont de 0 au maximum de la période de défaut. Par exemple, si la période de défaut dure 2 semaines ou 2000 échantillons alors l'ordonnée ira de 0 à 2000. De cette manière, si une courbe atteint le haut du graphique alors l'indicateur associé est capable de détecter le défaut dès le début de la période de défaut. Il est important aussi que les méthodes proposées puissent détecter les défauts plus de 1000 échantillons en avance. Si une courbe ne dépasse jamais cette valeur, on peut considérer que l'indicateur associé ne permet pas de détection suffisamment tôt pour adapter la stratégie de maintenance.

### 5.3.2. Nombre d'interventions inutiles fixe

Une autre manière de comparer les méthodes est aussi utilisée par la suite, plus particulièrement pour les défauts qui apparaissent soudainement et pour lesquels l'avance à la détection n'est pas un critère pertinent. Pour ces défauts, il suffit de vérifier si le défaut est détecté. En revanche, la persistance de l'indicateur prend, elle, plus d'importance pour ces défauts soudains.

Dans ce cas, le seuil de détection est fixé de manière à assurer le même nombre d'interventions inutiles pour tous les indicateurs. Les indicateurs sont alors comparés en vérifiant que le défaut est bien détecté et en analysant la persistance.

## 5.4. Création d'indicateurs

L'étude bibliographique a permis de montrer que les méthodes d'élaboration d'indicateurs de défauts à partir des données SCADA peuvent être classées en deux approches : les méthodes mono-turbines et les méthodes multi-turbines. Ces deux approches possèdent des avantages et inconvénients complémentaires. En effet, les méthodes mono-turbines sont moins sensibles aux variations des conditions de fonctionnement et plus sensibles aux variations environnementales, alors que c'est l'inverse pour les méthodes multi-turbines. L'idée défendue dans cette thèse est qu'en combinant les 2 approches, on peut réduire significativement l'influence de l'environnement et du fonctionnement machine. Ainsi, le nombre de fausses alarmes devrait être diminué tout en conservant de bonnes capacités de détection. Afin de mettre en évidence l'apport de chaque approche et l'intérêt de les combiner, différentes méthodes seront testées sur la base de données.

La nouvelle méthode proposée consiste à comparer les résidus de méthodes mono-turbines entre les éoliennes. Afin de comprendre les choix qui nous ont amenés à une telle solution, une première partie est dédiée à la présentation des méthodes mono-turbines. La seconde partie est dédiée aux méthodes multi-turbines et aux différentes manières de comparer les éoliennes.

### 5.4.1. Méthodes mono-turbines proposées

L'approche mono-turbine consiste à utiliser des variables mesurées sur une turbine comme entrées des modèles de bon comportement qui permettront la surveillance de cette même turbine. Les modèles de bon comportement (Isermann, 2006) permettent de définir une valeur prédite de sortie du modèle et c'est la différence entre la valeur prédite et la valeur surveillée à un instant donné qui constitue l'indicateur comme illustré dans la Figure 5.5. Nous proposons de post-traiter le résidu en utilisant une moyenne mobile de taille  $f$ . C'est le résidu moyenné qui sert ensuite d'indicateur.

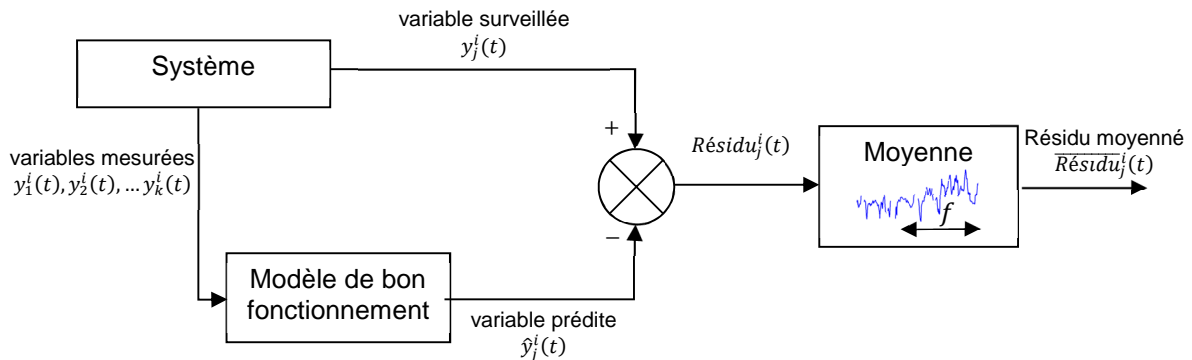


Figure 5.5 Schéma général des méthodes mono-turbines

D'après l'étude bibliographique, nous avons classé les modèles mono-turbines en deux types : ceux utilisant le lien de causalité physique et ceux utilisant la similarité de comportement entre les variables.

### 5.4.1.1. Modèles utilisant les liens de causalité

#### 5.4.1.1.1. Modèle dynamique appliqué sur toutes les phases en fonctionnement

Le modèle utilisé est issu des travaux de (Zhang, 2014) qui propose un modèle utilisant les liens de causalité pour surveiller le roulement principal d'une éolienne. Ici, il a été adapté pour les roulements et enrroulements générateurs. Il inclut une notion temporelle en prenant en compte des valeurs à l'instant  $t$  et l'instant  $t - 1$ .

Le principe de ce modèle est de décrire l'évolution de la température de la pièce en utilisant des variables fournies par le SCADA. La température du composant évolue en fonction de plusieurs sources de chaleur. Les données SCADA permettent d'avoir une information partielle sur ces différentes sources. Dans le modèle utilisé, 4 sources sont identifiées : la puissance transmise dans le composant, la friction due à la rotation dans les roulements, l'inertie thermique du composant et les échanges de températures avec l'environnement ambiant. Quatre variables SCADA sont donc choisies comme variables d'entrée des modèles : la puissance active produite, la vitesse de rotation de l'arbre de la génératrice, la température passée du composant et la température dans la nacelle.

Le modèle linéaire générique utilisé est le suivant :

$$\hat{T}_j^i(t) = a_j^i \cdot T_j^i(t-1) + b_j^i \cdot P^i(t) + c_j^i \cdot V^i(t) + d_j^i \cdot T_{nacelle}^i(t) + e_j^i \quad (5.1)$$

où

$\hat{T}_j^i(t)$  est la température de la partie  $j$  de la génératrice prédite, de la turbine  $i$ , à l'instant  $t$

$T_j^i(t-1)$  est la température de la partie  $j$  de la génératrice, de la turbine  $i$ , à l'instant  $t-1$

$P^i(t)$  est la puissance active générée par la turbine  $i$ , à l'instant  $t$

$V^i(t)$  est la vitesse de l'arbre rapide, de la turbine  $i$ , à l'instant  $t$

et  $a, b, c, d$  et  $e$  sont des paramètres appris à partir des données

$t$  décrit un instant d'échantillonnage,  $t-1$  l'instant précédent, mesuré 10 min avant.

Le résidu s'exprime comme :

$$Résidu_j^i(t) = T_j^i(t) - \hat{T}_j^i(t) \quad (5.2)$$

Ensuite un post-traitement est appliqué sous la forme d'une moyenne mobile de taille  $f$  :

On applique une moyenne mobile sur les températures  $Résidu_j^i(t)$  entre  $t$  et  $t-f$

$$\overline{Résidu}_j^i(t) = \frac{1}{f} \sum_{l=t-f}^t Résidu_j^i(l) \quad (5.3)$$

Si le nombre d'échantillons  $Résidu_j^i(t)$  valides entre  $t$  et  $t-f$  est supérieur à  $f/2$  alors  $\overline{Résidu}_j^i(t)$  est considéré comme valide pour cette éolienne à l'instant  $t$ .

L'indicateur de défaut associé à la turbine  $i$  est le résidu moyenné,  $\overline{Résidu}_j^i(t)$ .

Ce modèle linéaire n'est valide que lorsque la turbine est en production. La température prédite n'est calculée que si la valeur de la puissance active de l'éolienne  $i$  à l'instant  $t$   $P^i(t)$  est supérieure à un seuil de production c'est-à-dire :

$$P^i(t) > \text{seuil}_{\text{production}}^1 \quad (5.4)$$

#### 5.4.1.1.2. Modèle dynamique appliqué uniquement pendant les phases de régime permanent

Cette approche prenant en compte la dynamique peut nécessiter une adaptation. En effet la validité du modèle peut être remise en question lors des phases de démarrage de la machine, durant lesquelles la dynamique des échanges thermiques est plus rapide que lors des phases de fonctionnement de longue durée. Il faut attendre environ 3 heures (soit 18 échantillons 10 minutes) pour atteindre le régime permanent et retrouver la validité du modèle linéaire proposé.

Ainsi le modèle linéaire n'est valide que lorsque la turbine est en production depuis au moins 3 heures, la température prédite n'est calculée que si  $P^i(t)$ , la valeur de la puissance active de l'éolienne  $i$ , est supérieure au seuil de production entre les instants  $t$  et  $t - 18$

#### 5.4.1.1.3. Modèle statique appliqué sur toutes les phases de fonctionnement

La prise en compte de la dynamique est une tâche difficile et une autre méthode utilisant les liens de causalité est proposé en adaptant les travaux (Zhang, 2014) et (Wilkinson et al., 2014).

Le modèle linéaire proposé ne contient plus la variable de température à  $t - 1$ ,  $T_j^i(t - 1)$ , pour devenir statique, comme décrit ci-dessous :

$$\hat{T}_j^i(t) = a_j^i \cdot P^i(t) + b_j^i \cdot V^i(t) + c_j^i \cdot T_{nacelle}^i(t) + d_j^i \quad (5.5)$$

où

$\hat{T}_j^i$  est la température de la partie  $j$  de la génératrice prédite, de la turbine  $i$ , à l'instant  $t$

$P^i(t)$  est la puissance active générée par la turbine  $i$ , à l'instant  $t$

$V^i(t)$  est la vitesse de l'arbre rapide, de la turbine  $i$ , à l'instant  $t$

et  $a$ ,  $b$ ,  $c$ ,  $d$  et  $e$  sont des paramètres appris à partir des données  $t$  décrit un instant d'échantillonnage.

Le résidu s'exprime comme :

$$\text{Résidu}_j^i(t) = T_j^i(t) - \hat{T}_j^i(t) \quad (5.6)$$

Ensuite un post-traitement est appliqué sous la forme d'une moyenne mobile de taille  $f$  :

On applique une moyenne mobile sur les températures  $\text{Résidu}_j^i(t)$  entre  $t$  et  $t - f$

$$\overline{\text{Résidu}}_j^i(t) = \frac{1}{f} \sum_{l=t-f}^t \text{Résidu}_j^i(l) \quad (5.7)$$

Si le nombre d'échantillons  $Résidu_j^i(t)$  valides entre  $t$  et  $t - f$  est supérieur à  $f/2$  alors  $\overline{Résidu}_j^i(t)$  est considéré comme valide pour cette éolienne à l'instant  $t$ .

L'indicateur de défaut associé à la turbine  $i$  est le résidu moyenné,  $\overline{Résidu}_j^i(t)$ .

Ce modèle linéaire n'est valide que lorsque la turbine est en production. La température prédite n'est calculée que si  $P^i(t)$  la valeur de la puissance active de l'éolienne  $i$  à l'instant  $t$  est supérieure à un seuil de production c'est-à-dire :

$$P^i(t) > seuil_{production}^1 \quad (5.8)$$

Trois méthodes utilisant la causalité sont donc proposées et comparées dans cette thèse : une méthode statique, une méthode dynamique et une méthode dynamique utilisée uniquement sur les périodes de régime permanent.

#### 5.4.1.2. Modèles utilisant la similarité

Certaines parties des composants sont soumises à des sollicitations semblables et fortement corrélées. Il en résulte que l'évolution de certaines grandeurs physiques est similaire entre ces composants. Ces ressemblances sont exploitées par (Schlechtingen and Santos, 2011) afin de créer un modèle de bon comportement. La variable d'entrée du modèle est choisie en fonction de sa ressemblance de comportement avec la variable à prédire.

Dans l'article de (Schlechtingen and Santos, 2011, figure 22), plusieurs modèles sont testés avec d'autres variables en entrée comme la puissance ou la vitesse de rotation, et des modèles linéaires et non linéaires sont considérés. La comparaison de ces différents modèles montre que les résultats sont très proches. Les modèles utilisés ici sont donc des régressions linéaires prenant en entrée uniquement les données de température des composants.

$$\hat{T}_j^i(t) = a_j^i \cdot T_{refj}^i(t) + b_j^i \quad (5.9)$$

où

$\hat{T}_j^i$  est la température prédite de la partie  $j$  de la génératrice de la turbine  $i$  à surveiller

$T_{refj}^i$  est la température de la partie de la génératrice de la turbine  $i$  utilisée comme référence pour la partie  $j$

et  $a$  et  $b$  sont des paramètres appris à partir des données.

Le résidu est décrit comme :

$$Résidu_j^i(t) = T_j^i(t) - \hat{T}_j^i(t) \quad (5.10)$$

Ensuite un post-traitement est appliqué sous la forme d'une moyenne mobile de taille  $f$  :

On applique une moyenne mobile sur les températures  $RésiduT_j^i(t)$  entre  $t$  et  $t - f$

$$\overline{Résidu}_j^i(t) = \frac{1}{f} \sum_{l=t-f}^t Résidu_j^i(l) \quad (5.11)$$

Si le nombre d'échantillons  $Résidu_j^i(t)$  valide entre  $t$  et  $t - f$  est supérieur à  $f/2$  alors  $\overline{Résidu}_j^i(t)$  est considéré comme valide pour cette éolienne à l'instant  $t$ .

L'indicateur de défaut associé à la turbine  $i$  est le résidu moyenné,  $\overline{Résidu}_j^i(t)$ .

Tout comme les modèles utilisant les liens de causalité, ces modèles ne sont valides que lorsque l'éolienne est en fonctionnement.

La sortie du modèle ne peut être calculée que si :

$$P^i(t) > seuil_{production2} \quad (5.12)$$

Contrairement aux modèles basés sur la causalité, ces modèles sont valides sur une plus grande gamme de conditions opératoires. En effet, la température des deux composants est affectée par des effets d'inertie et de non-linéarités semblables. Néanmoins, il est nécessaire de s'assurer que les machines soient en production.

Le seuil de détection  $seuil_{production2}$  peut donc être réglé à une valeur plus faible que  $seuil_{production1}$ . La période d'échantillonnage des données est de 10 minutes, et la variation de production peut être importante sur cette période. Le seuil doit être fixé au-dessus de 0W pour être sûr que la machine est majoritairement en fonctionnement pendant 10 minutes.

#### 5.4.1.3. Apprentissage des paramètres du modèle

Les modèles qui sont développés utilisent des paramètres qu'il faut estimer à partir des données SCADA sur des périodes de bon fonctionnement. Ces paramètres sont calculés à partir de données d'apprentissage issues de périodes de bon fonctionnement précédant les défauts avérés.

Il est important de noter qu'après chaque défaut les paramètres des modèles doivent être appris de nouveau puisque les composants sont remplacés ou modifiés. Le comportement est donc modifié. La base de données est découpée en alternant les périodes de bon comportement et de défaut.

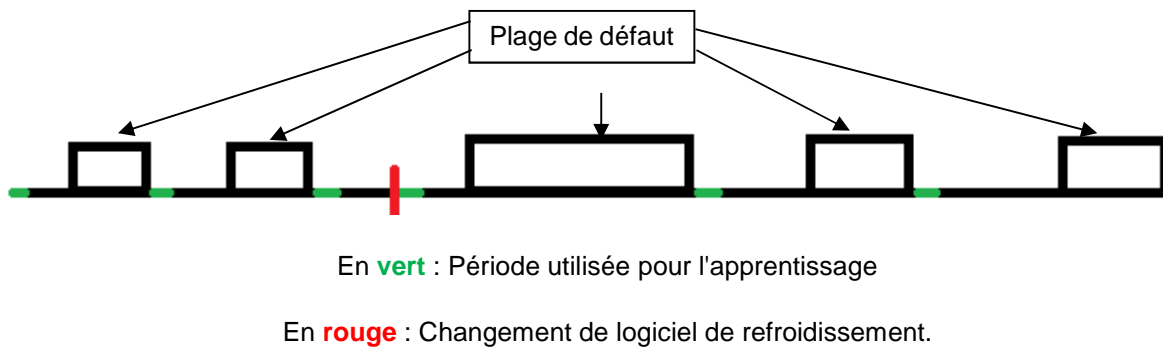


Figure 5.6 Schéma de répartition des périodes d'apprentissages

La Figure 5.6 montre comment les périodes d'apprentissage des paramètres sont réparties dans la base de données. Après chaque défaut avéré, une période de bon fonctionnement permet de réapprendre les paramètres des modèles. De plus, après un changement important dans la calibration des capteurs ou dans le fonctionnement de la régulation de la génératrice, une nouvelle période de bon fonctionnement permet de réapprendre les paramètres des modèles.



Notons que toute la période de bon fonctionnement précédant un défaut n'est pas utilisée pour l'apprentissage. L'un des objectifs de nos travaux est en effet de respecter au mieux les contraintes industrielles. Ainsi, un nombre minimal d'échantillons  $N_A$  est défini pour réaliser l'apprentissage. Les  $N_A$  premiers échantillons valides suivant la réparation sont utilisés pour estimer les paramètres.

Notons qu'après chaque défaut, sur la période de durée  $N_A$  nécessaire au réapprentissage, la surveillance des machines n'est plus possible puisque les modèles de bon fonctionnement en cours d'apprentissage ne sont pas disponibles. Afin de respecter les contraintes industrielles, la durée de cette période doit être la plus courte possible.

## 5.4.2. Méthodes multi-turbines proposées

L'approche multi-turbine est basée sur l'hypothèse que toutes les éoliennes d'un parc sont soumises aux mêmes conditions environnementales et subissent donc la même sollicitation. Dans la mesure où ces éoliennes sont du même modèle, leur comportement doit être semblable. Sur le principe, c'est cette ressemblance qui est utilisée pour créer les indicateurs de défauts en évaluant des écarts entre les machines. Peu de travaux ont été publiés sur la comparaison des éoliennes pour créer des indicateurs par rapport à l'approche mono-turbine. Il existe plusieurs manières pour combiner les informations de différentes éoliennes, chacune ayant des avantages et inconvénients. Cette partie présente différentes méthodes qui seront comparées par la suite dans la section 6, consacrée à l'évaluation des performances sur la base de données VALEMO. Ces méthodes ont été élaborées dans le cadre de cette thèse à partir de l'étude bibliographique du chapitre 3. Elles ont pour objectif, dans un premier temps, de prendre en compte la variation de l'environnement en comparant les éoliennes entre elles. Ensuite, dans un deuxième temps, elles permettent de s'affranchir des variations de fonctionnement de la machine en les combinant avec des méthodes mono-turbines.

### 5.4.2.1. Principes des méthodes multi-turbines.

Le principe des méthodes multi-turbines consiste à comparer des variables mesurées ou calculées sur chaque éolienne du parc à une référence parc, qui décrit le comportement médian ou moyen de la variable au niveau du parc. Les différences entre les méthodes dépendent des variables utilisées pour la comparaison et de la plage temporelle pendant laquelle la comparaison est effectuée. Ainsi, les variables à comparer peuvent être :

- des mesures provenant du SCADA, comme des mesures de températures
- des résidus issus des modèles développés dans l'approche mono-turbine

La comparaison peut se faire

- à un instant d'échantillonnage  $t$  (10 min)
- sur une période temporelle comprenant un nombre périodes d'échantillonnage égal à  $f$

L'approche multi-turbine que nous proposons repose donc sur un principe original, se démarquant nettement des approches identifiées dans l'état de l'art et présentées dans la partie 3.2: elle s'appuie sur des modèles mono-turbines, mais ne cherche pas à comparer ces modèles entre eux. L'approche proposée consiste plutôt à étudier les écarts entre « résidus mono-turbines » (erreurs de prédiction des modèles mono-turbines) pour identifier des défauts

Afin de pouvoir comparer les éoliennes entre elles, une référence du parc est d'abord construite. Chaque éolienne est ensuite comparée à cette référence en mesurant la différence entre la variable et

la référence. La référence doit être la moins sensible possible aux comportements atypiques qui vont apparaître quand une des turbines est en défaut. De plus, la référence doit rester robuste même si une turbine du parc est à l'arrêt. Ceci peut arriver fréquemment, car les machines sont souvent arrêtées pour des raisons de maintenance ou d'inspection. Ainsi, dans la suite, la référence sera toujours construite comme la valeur médiane de la variable à comparer sur l'ensemble du parc. Afin de garantir une certaine robustesse, la référence parc ne sera calculée à l'instant  $t$  que si le nombre d'éoliennes en fonctionnement  $N_E(t)$  est supérieur ou égal à  $N_{Emin}$ . La valeur de  $N_{Emin}$  peut varier en fonction de la taille du parc, notée  $N_T$  par la suite. Fixer  $N_{Emin}$  à la moitié de la taille du parc semble être une bonne valeur. Une éolienne est considérée en fonctionnement si sa puissance est supérieur à  $seuil_{production}$ <sup>3</sup>. L'utilisation d'une simple médiane permet d'avoir des résultats satisfaisants sans avoir à définir des paramètres supplémentaires par apprentissage.

Le tableau suivant résume les 4 différentes stratégies qui peuvent être utilisées pour comparer les turbines et détaille leurs avantages et inconvénients.

Tableau 5.1 Avantages et inconvénient des différentes méthodes multi-turbines proposées.

	Variable brute	Résidu mono-turbine
Comparaison Instantanée	<p>++ peu sensible aux variations de l'environnement</p> <p>-- sensible aux différences entre éoliennes</p> <p>--sensible aux variations de conditions de fonctionnement</p> <p>++ bon contraste entre défaut et bon comportement</p>	<p>++ peu sensible aux variations de l'environnement</p> <p>++ peu sensible aux variations de conditions de fonctionnement</p> <p>-- sensible aux erreurs de modélisation des méthodes mono-turbine</p> <p>++ bon contraste entre défaut et bon comportement</p>
Comparaison sur une période	<p>++ peu sensible aux variations de l'environnement</p> <p>++ peu sensible aux différences entre éoliennes</p> <p>--sensible aux variations de conditions de fonctionnement</p> <p>-- contraste entre défaut et bon comportement diminué fortement</p>	<p>++ peu sensible aux variations de l'environnement</p> <p>++ peu sensible aux variations de conditions de fonctionnement</p> <p>- moins sensible aux erreurs de modélisation des méthodes mono-turbine</p> <p>+ contraste entre défaut et bon comportement peu diminué</p>

L'idée principale est que les températures enregistrées via le SCADA ou les résidus mono-turbines sont sensibles aux variations de la température extérieure ou aux conditions de vent. Les fluctuations générées sur les indicateurs par ces changements de conditions expérimentales peuvent être éliminées en analysant les turbines entre elles, car ces variations devraient être concomitantes. Ainsi, chacune des stratégies a comme avantage le fait d'être peu sensible aux variations environnementales.

La comparaison des variables à chaque période d'échantillonnage souffre du fait que les machines n'ont pas un comportement synchrone et les phases d'arrêt et de redémarrage sont nombreuses. Ainsi, même si les conditions expérimentales sont les mêmes à un instant donné, certaines machines peuvent être en phase transitoire de démarrage ou d'arrêt, alors que d'autres sont en régime permanent. Ces méthodes restent donc sensibles aux différences entre turbines, et ceci d'autant plus si on compare les variables brutes comme les températures. La comparaison des variables sur une période de temps permet de s'affranchir de ce problème, car l'impact des périodes de transitoire est réduit. Les modèles

mono-turbines permettent de prendre en compte l'effet des conditions de fonctionnement et d'obtenir des résidus insensibles à ces dernières. Comparer les résidus sur une période de temps devrait permettre de réduire la sensibilité aux conditions de fonctionnement.

Dans ce qui suit, nous allons présenter plus en détail les 4 méthodes que nous proposons.

#### 5.4.2.2. Méthodes comparant les variables brutes

L'utilisation des variables brutes seules ne permet pas une détection suffisamment tôt des défauts comme le montre la Figure 5.7. En effet, la température du roulement évolue fortement au cours du temps et il est difficile de voir le défaut avant qu'il n'atteigne un stade de dégradation avancé et que la surchauffe dépasse les 65°C. La détection à ce stade est peu utile pour un industriel. En revanche, en superposant les variables brutes, on peut observer plus facilement la différence avec les autres turbines du parc lorsque celle-ci est en défaut. Grâce à cela, on peut voir que de premiers signes d'anomalie (écart de comportement par rapport au reste du parc) sont présents environ 2 semaines avant la panne totale (2 semaines  $\approx$  2000 points 10min).

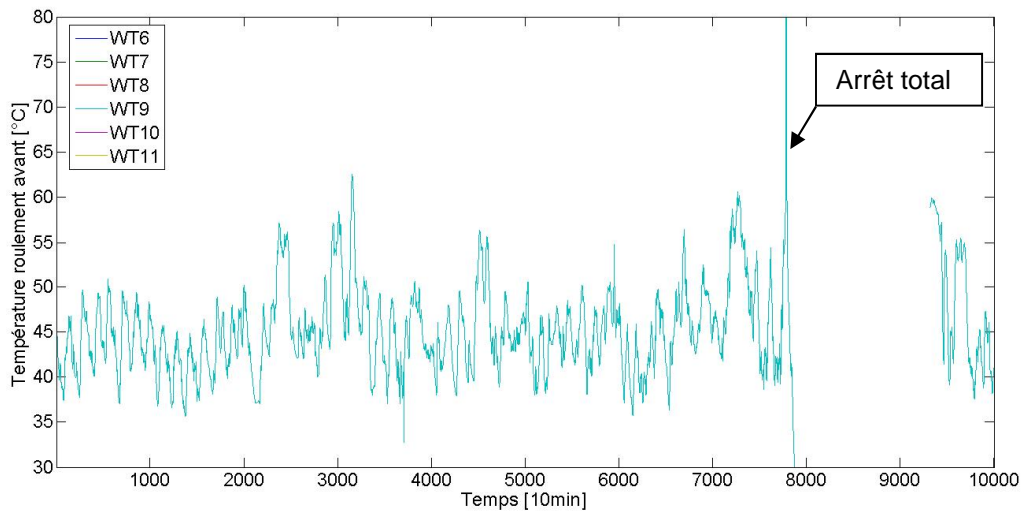


Figure 5.7 Évolution de la température du roulement avant de l'éolienne 9 en défaut.

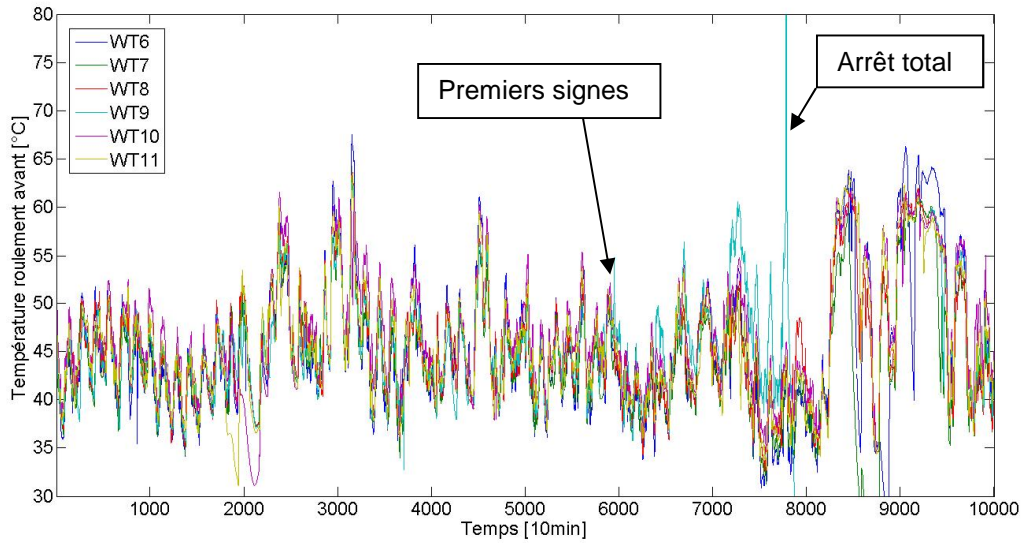


Figure 5.8 Évolution de la température des roulements avant, des éoliennes du parc en défaut, éolienne 9 en défaut.

#### 5.4.2.2.1. Comparaison à chaque période d'échantillonnage

La première solution que nous proposons est de comparer les données brutes de chaque éolienne à chaque instant. Cette solution a été présentée dans (Lebranchu et al., 2016) au cours de cette thèse. La comparaison se fait en utilisant une seule mesure par turbine, prise au même moment.

La référence parc est calculée en prenant la médiane des données brutes à l'instant de la comparaison. L'indicateur utilisé pour la détection de défaut est la différence calculée entre la valeur de la variable brute de la turbine surveillée et la référence parc.

La variable est une température mesurée sur une partie  $j$  de la génératrice de la turbine :  $T_j^i(t)$

Une première sélection est appliquée aux données pour ne considérer les éoliennes que pendant des périodes de production. Pour un parc comprenant  $N_T$  éoliennes :

Pour chaque instant  $t$

$$N_E(t) = 0$$

1/ On sélectionne les variables d'entrée

Pour chaque turbine  $i$  de 1 à  $N_T$

Si  $P^i(t) > \text{seuil}_{\text{production}}$  avec  $P^i(t)$  la valeur de la puissance active de l'éolienne  $i$  à l'instant  $t$

Alors la mesure  $T_j^i(t)$  est conservée.

$$N_E(t) = N_E(t) + 1$$

(5.13)

2/ On calcule la référence du parc

S'il y a suffisamment d'éoliennes valides, c'est à dire  $N_E(t) > N_{Emin}$

Alors on construit la référence parc :

$$T_{med_j}(t) = \text{median}(T_j^k(t))_{k=1 \text{ à } N_E(t)}$$

3/ On calcule la différence entre la température de la turbine  $i$  et la référence :

Pour chaque turbine  $i$  de 1 à  $N_E(t)$

$$\text{Différence}_j^i(t) = T_j^i(t) - T_{med_j}(t)$$

La différence ainsi obtenue peut servir d'indicateur de santé de l'éolienne. Notons que l'indicateur est une différence et non pas une distance. En effet, on ne s'intéresse ici qu'aux surchauffes. Une différence négative ne peut donc pas être associée à un défaut.

#### 5.4.2.2.2. Comparaison sur une période de temps

Même si les turbines ont des comportements similaires, elles n'ont pas nécessairement un fonctionnement synchrone. Dans certains cas, des différences peuvent être observées entre les turbines. Ces différences sont dues à des événements propres à une turbine en particulier, comme des arrêts de la machine ou des effets topologiques. La Figure 5.9 permet de voir deux types de différences :

- Une différence liée à un arrêt de la turbine : lorsque la turbine s'arrête brièvement pendant une période de production, elle refroidit et prend du temps pour arriver de nouveau à son régime permanent et donc rejoindre le comportement du parc.
- On peut voir aussi en période de très faible production et donc de faible vent que certaines éoliennes démarrent et d'autres non, en raison soit d'effets de site soit d'effets de calibration anémométrique. Par conséquent, certaines turbines auront « préchauffées » et seront en avance par rapport à d'autres qui n'auront pas du tout commencé à produire. Ces différences dynamiques de température sont vite corrigées lorsque le vent se lève et que la production démarre vraiment. Au bout de quelques heures (environ 3 heures = 18 points 10min) toutes les éoliennes se suivent de nouveau.

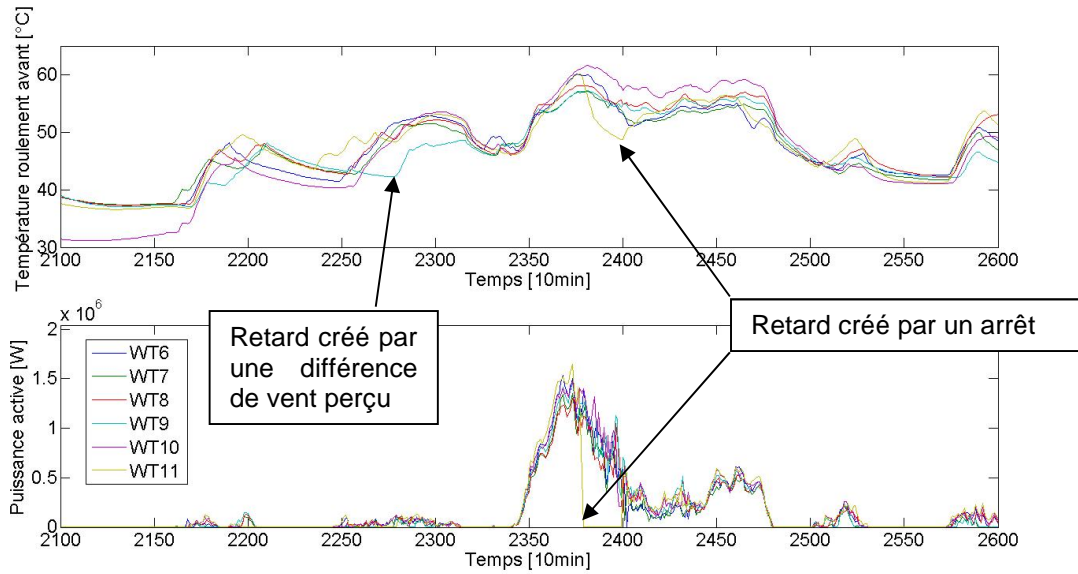


Figure 5.9 Illustration des retards entre les turbines sur le parc

Une solution à ce problème serait d'appliquer un filtre pour exclure les périodes de démarrage et attendre que toutes les turbines aient atteint leur régime permanent. Néanmoins, le nombre d'arrêts et de redémarrages est important, ce qui génère l'exclusion d'un trop grand nombre d'échantillons et diminue fortement la disponibilité du système de surveillance.

L'autre solution que nous proposons est de comparer les éoliennes sur une période plutôt que sur un seul instant  $t$ . La comparaison sur une période se traduit par l'utilisation d'une moyenne mobile de taille de fenêtre  $f$ . Puisque les démarrages et les arrêts sont des événements courts, en choisissant une fenêtre suffisamment grande, on peut diminuer l'impact des différences entre turbines. En revanche, plus la fenêtre est grande, plus la détection est tardive.

La solution devient donc :

Pour chaque instant  $t$

$$N_E(t) = 0$$

1/ On sélectionne les variables d'entrée

Pour chaque turbine  $i$  de 1 à  $N_T$ ,

Si  $P^i(t) > \text{seuil}_{\text{production}}$  avec  $P^i(t)$  la valeur de la puissance active de l'éolienne  $i$  à l'instant  $t$

Alors la mesure  $T_j^i(t)$  est conservée.

2/ On applique une moyenne mobile sur les températures  $T_j^i(t)$  entre  $t$  et  $t - f$

$$\bar{T}_j^i(t) = \frac{1}{f} \sum_{l=t-f}^t T_j^i(l)$$

Si le nombre d'échantillons  $T_j^i(t)$  entre  $t$  et  $t - f$  est supérieur à  $f/2$ , (5.14)

Alors  $\bar{T}_j^i(t)$  est considéré comme valide

$$N_E(t) = N_E(t) + 1$$

3/ On calcule la référence du parc

S'il y a suffisamment d'éoliennes valides, c'est à dire  $N_E(t) > N_{Emin}$

Alors on construit la référence parc :

$$T_{medj}(t) = \text{median}(\bar{T}_j^k(t))_{k=1 \text{ à } N_E(t)}$$

4/ On calcule la différence entre la température de la turbine  $i$  et la référence :

Pour chaque turbine  $i$  de 1 à  $N_E(t)$

$$\text{Différence}_j^i(t) = \bar{T}_j^i(t) - T_{medj}(t)$$

La différence ainsi obtenue,  $\text{Différence}_j^i(t)$ , est utilisée comme indicateur de santé de l'éolienne.

#### 5.4.2.3. Méthodes comparant des résidus

Ce dernier type de méthodes correspond à la contribution principale de cette thèse. L'objectif est de construire un indicateur de santé de l'éolienne qui s'affranchit des variations des conditions de fonctionnement et des conditions environnementales. Pour cela, la solution proposée consiste à combiner les résidus mono-turbines en une approche multi-turbine.

La comparaison des températures brutes des éoliennes permet de s'affranchir des variations environnementales, mais souffre des différences entre les turbines qui sont dues aux différences de conditions de fonctionnement. Les méthodes mono-turbines sont quant à elles capables de fournir un résidu faiblement influencé par ces variations, mais fortement influencé par l'environnement. L'idée est donc de comparer les résidus entre les turbines plutôt que de comparer directement les variables brutes venant du SCADA.

Pour cela, une première solution est de comparer à un instant  $t$  les résidus des turbines provenant d'une méthode mono-turbine surveillant un composant.

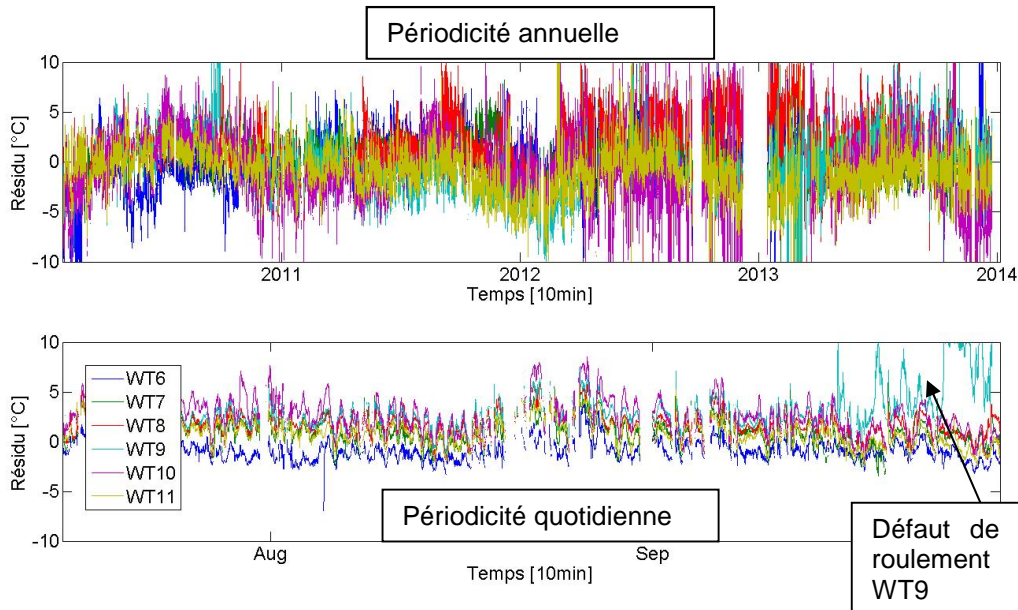


Figure 5.10 Périodicité annuelle et quotidienne des résidus mono-turbine basés sur la similarité roulement avant vs arrière

La Figure 5.10 montre l'évolution des résidus d'une méthode mono-turbine basée sur la similarité entre le roulement avant et arrière. Il existe des variations très basse fréquence qui correspondent à l'évolution de la température au cours des saisons de l'année. Les résidus semblent évoluer conjointement au cours de l'année. De plus, lorsque l'on s'intéresse aux variations à l'échelle d'une semaine, on peut voir qu'une fluctuation quotidienne apparaît, qui correspond aux cycles de lever et coucher du soleil. Finalement, il existe des phénomènes non saisonniers dus à la nature stochastique du vent qui font varier les résidus conjointement. Par conséquent, en comparant les résidus des éoliennes entre elles, il semble possible de diminuer ces variations pour que les défauts soient plus facilement détectables, ce qui permettra de diminuer le nombre de fausses alarmes ou de détecter les défauts plus tôt.

#### 5.4.2.3.1. Comparaison à chaque période d'échantillonnage

Comme pour la comparaison des variables brutes, une première solution est de comparer les résidus entre les turbines à chaque période d'échantillonnage  $t$ . Cette comparaison se fait à chaque instant en ne prenant qu'une seule valeur du résidu par turbine au même moment.

La référence parc est calculée en prenant la médiane des résidus mono-turbines à l'instant de la comparaison. L'indicateur utilisé pour la détection de défaut est la différence calculée entre la valeur du résidu de l'éolienne surveillée et la référence parc, comme décrit dans la partie 5.4.1. Le résidu est décrit comme suit :

$$\text{Résidu}_j^i(t) = T_j^i(t) - \hat{T}_j^i(t) \quad (5.15)$$



Une première sélection est appliquée aux données lors du calcul du résidu en fonction de la méthode mono-turbine choisie. Elle a pour but de ne considérer les éoliennes que pendant des périodes de production.

Pour chaque instant  $t$

$$N_E(t) = 0$$

1/ On calcul les résidus de chaque éolienne grâce à une méthode mono-turbine

Pour chaque turbine  $i$  de 1 à  $N_T$ ,

$$\text{Résidu}_j^i(t) = T_j^i(t) - \hat{T}_j^i(t)$$

Si  $\text{Résidu}_j^i(t)$  est valide

Alors

$$N_E(t) = N_E(t) + 1 \tag{5.16}$$

2/ On calcule la référence du parc

S'il y a suffisamment d'éoliennes valides, c'est à dire  $N_E(t) > N_{Emin}$

Alors on construit la référence parc :

$$\text{Résidu}_{medj}(t) = \text{median}(\text{Résidu}_j^k(t))_{k=1 \text{ à } N_E(t)}$$

3/ On calcule la différence, entre la température de la turbine  $i$  et la référence :

Pour chaque turbine  $i$  de 1 à  $N_E(t)$

$$\text{Différence}_j^i(t) = \text{Résidu}_j^i(t) - \text{Résidu}_{medj}(t)$$

La différence ainsi obtenue peut servir d'indicateur de santé de l'éolienne.

#### 5.4.2.3.2. Comparaison sur une période de temps

La comparaison à un instant  $t$  suppose que les modèles mono-turbines fournissent un résidu suffisamment indépendant des variations de condition de fonctionnement pour que les différences instantanées entre les turbines soient négligeables. Or, en réalité, les modèles ne sont pas parfaits et ne capturent pas tous les aspects du fonctionnement d'une machine. Dans certains cas, le résidu mono-turbine est toujours influencé par le fonctionnement de la machine. Afin de diminuer ce problème, la comparaison des résidus sur une période de temps semble plus efficace que la comparaison instantanée.

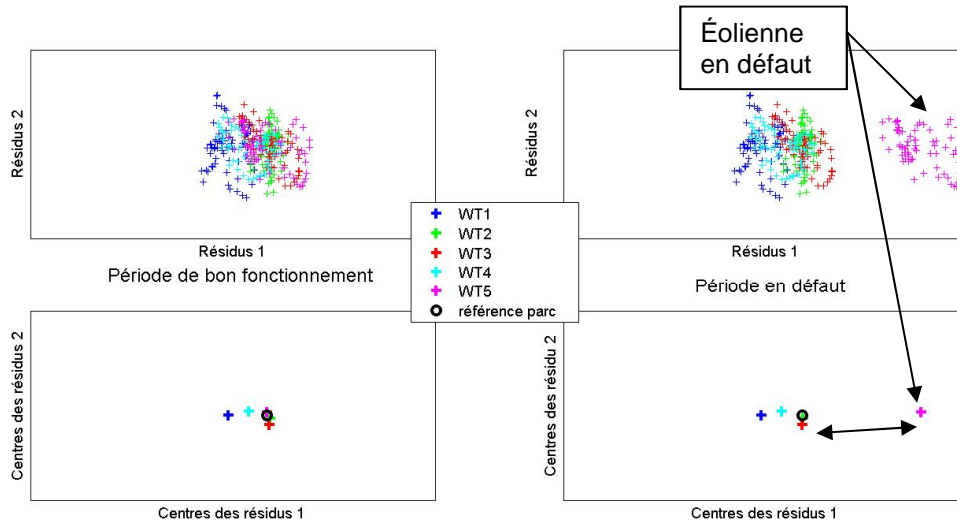


Figure 5.11 Schéma de principe de la comparaison multi-turbine utilisée en 2 dimensions

La Figure 5.11 décrit le principe de la comparaison des résidus entre plusieurs éoliennes. Les deux graphiques supérieurs montrent les échantillons sur une plage de temps égale à  $f$ . Il y a donc un maximum de  $f$  échantillons par turbine. Pour plus de lisibilité, cet exemple est construit en 2 dimensions en montrant la comparaison sur 2 résidus en même temps. Dans la suite, les méthodes proposées sont en une dimension seulement. Les deux graphiques inférieurs montrent la position des centres de chaque éolienne, construite comme la moyenne des  $f$  échantillons. Le cercle noir représente la référence parc construit comme la médiane des autres cercles. La Figure 5.11 montre qu'en principe les éoliennes évoluent en groupe lors de période de bon fonctionnement. Lorsqu'un défaut apparaît, le centre de l'éolienne défaillante s'écarte progressivement du groupe. Plus l'éolienne est éloignée de la référence parc, plus le défaut est marqué.

La solution proposée dans cette thèse se présente comme un assemblage des méthodes précédentes. Elle est décrite ci-dessous :

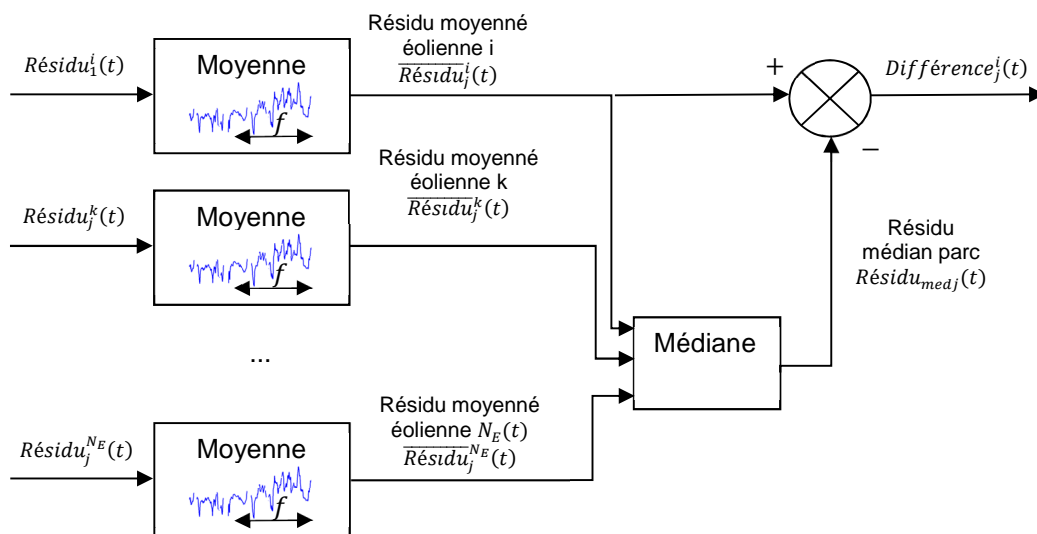


Figure 5.12 Schéma de la nouvelle méthode multi-turbine proposée, comparant les résidus provenant de modèle mono-turbine, entre les éoliennes

Pour chaque instant  $t$

$$N_E(t) = 0$$

1/ On calcule les résidus de chaque éolienne grâce à une méthode mono-turbine

Pour chaque turbine  $i$  de 1 à  $N_T$ ,

$$\text{Résidu}_j^i(t) = T_j^i(t) - \hat{T}_j^i(t)$$

Si  $\text{Résidu}_j^i(t)$  est valide alors cette valeur est conservée

2/ On applique une moyenne mobile sur les températures  $\text{Résidu}_j^i(t)$  entre  $t$  et  $t - f$

$$\overline{\text{Résidu}}_j^i(t) = \frac{1}{f} \sum_{l=t-f}^t \text{Résidu}_j^i(l)$$

Si le nombre d'échantillons  $\overline{\text{Résidu}}_j^i(t)$  entre  $t$  et  $t - f$  est supérieur à  $f/2$ ,

Alors  $\overline{\text{Résidu}}_j^i(t)$  est considéré comme valide

$$N_E(t) = N_E(t) + 1$$

3/ On calcule la référence du parc

S'il y a suffisamment d'éoliennes valides, c'est à dire  $N_E(t) > N_{Emin}$

Alors on construit la référence parc :

$$\text{Résidu}_{medj}(t) = \text{median}(\overline{\text{Résidu}}_j^k(t))_{k=1 \text{ à } N_E(t)}$$

4/ On calcule la différence, entre la température de la turbine  $i$  et la référence :

Pour chaque turbine  $i$  de 1 à  $N_E(t)$

$$\text{Différence}_j^i(t) = \text{Résidu}_j^i(t) - \text{Résidu}_{medj}(t)$$

La différence ainsi obtenue peut servir d'indicateur de santé de l'éolienne.

La Figure 5.12 montre le principe de la nouvelle méthode multi-turbine proposée. Les résidus mono-turbines sont obtenus par les méthodes décrites dans la partie 5.4.1 (cf Figure 5.5). L'indicateur est défini comme la différence entre la médiane des résidus moyennés de chaque éolienne et le résidu moyenné de l'éolienne que l'on cherche à surveiller.

## 5.5. Conclusion

Dans ce chapitre, nous avons présenté les contributions méthodologiques développées dans cette thèse. Nous avons proposé des contributions selon deux axes :

Le premier axe de contribution concerne l'évaluation des performances des indicateurs de dégradation et de la détection de défauts. Nous avons proposé 4 critères permettant de déterminer la

capacité de chaque indicateur à détecter des défauts, mais aussi à évaluer sa capacité à être mise en œuvre dans le cadre d'une maintenance réaliste, dans des conditions de terrain. Ainsi, nous avons proposé trois nouveaux critères : le nombre d'interventions inutiles, la disponibilité et la persistance des indicateurs. Ces critères vont permettre de déterminer dans quelle mesure les indicateurs proposés peuvent répondre à un besoin industriel. En effet, ils permettent d'estimer si ces indicateurs sont acceptables pour des opérateurs de maintenance et les surcoûts qu'ils peuvent engendrer. La persistance va pouvoir être liée à la confiance qu'un opérateur pourra avoir dans une alarme et la disponibilité permet de s'assurer que les méthodes fonctionneront en continu. Enfin, le nombre d'interventions inutiles permet de mesurer le surcoût engendré.

Le deuxième axe de contribution porte sur la proposition d'algorithmes d'élaboration d'indicateurs de défauts à partir de données SCADA disponibles sur un parc de machines. Nous avons d'abord décrit les méthodes mono-turbines, où un modèle de bon fonctionnement est appris pour chaque turbine sur une période de temps sans défaut et utilisé pour générer un résidu. Ces méthodes ont déjà été proposées dans la littérature, mais n'ont jamais été évaluées à l'échelle d'un parc sur une période de temps de plusieurs années. Les méthodes dites multi-turbines ont été assez peu présentées dans la littérature. Dans cette famille, nous avons proposé une approche originale permettant de combiner les résidus des méthodes mono-turbines sur une période de temps.

Cette nouvelle méthode et les autres sont évaluées dans le chapitre suivant sur les bases des 4 critères proposés en utilisant la base de données décrite au chapitre 4. Le but de cette évaluation sur données réelles exclusivement est de montrer l'apport de la combinaison des approches mono-turbines et multi-turbines ainsi que les limites de ces méthodes dans un contexte industriel.



## 6. Évaluation des performances des méthodes proposées sur la base de données VALEMO

Dans ce chapitre, nous évaluons les performances des différents indicateurs de défauts présentés au chapitre précédent sur la base de données VALEMO. Comme nous l'avons vu précédemment, chaque indicateur mono-turbine peut être combiné dans une approche multi-turbine. L'objectif de ce chapitre est de montrer l'intérêt de la fusion multi-turbine. Les résultats sont présentés de la manière suivante. Chaque type de méthodes (utilisation des données brutes, indicateurs obtenus par modèles utilisant les liens de causalité, indicateurs obtenus par modèles utilisant la similarité) est présenté dans un paragraphe distinct. Au sein de chaque paragraphe, les résultats utilisant l'approche mono-turbine sont d'abord présentés suivis par ceux obtenus par l'approche multi-turbine. L'influence de l'intervalle de temps utilisé pour comparer les indicateurs est aussi analysée. Enfin, un dernier paragraphe compare l'ensemble des méthodes entre elles et conclut ce chapitre. Les performances des méthodes sont analysées en utilisant les critères de performances définis au chapitre précédent. Rappelons qu'un seuil de détection est fixé pour chaque méthode et que ce seuil est le même pour toutes les turbines du parc. Le critère d'interventions inutiles s'entend donc au niveau parc. Pour un seuil donné, ce critère mesure le nombre d'interventions inutiles sur l'ensemble des turbines du parc pendant la durée d'observation, c'est-à-dire 4 ans.

### 6.1. Réglages préliminaires

#### 6.1.1. Définitions des plages de défauts et de bon fonctionnement

Afin de pouvoir calculer les critères d'évaluation des performances de détection de défauts sur la base de données étudiée, celle-ci doit d'abord être découpée en plage de bon fonctionnement et en plage de défaut. Ce découpage est nécessaire pour pouvoir choisir les plages sur lesquelles les paramètres des modèles sont calculés et comparer les comportements des indicateurs entre périodes de défaut et de bon fonctionnement.

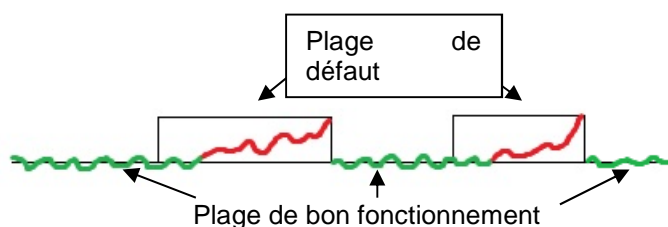
Rappelons que 8 défauts sont présents dans la base : 2 défauts de roulement et 2 défauts d'enroulement dont l'apparition est progressive et 4 défauts du système de refroidissement (ventilateur) dont l'apparition est soudaine. Les 2 défauts de roulement et les 2 défauts d'enroulement sont considérés comme critiques. Ces défauts ont eu un impact sur la rentabilité du parc puisqu'ils ont généré de fortes pertes de production, à cause des interventions de remplacement qu'ils ont générées. Il est très important de les détecter au plus tôt. Ainsi, dans la suite de ce chapitre, ces défauts seront dénommés « défauts critiques ». Les 4 autres défauts sont nommés « défauts ventilateur ». Les défauts de ventilateur sont moins graves, car ils ne présentent pas de danger pour la machine sur le court terme. Par contre, si ces défauts de ventilateur perduraient, ils diminueraient la durée de vie des isolants dans la génératrice. Il faut donc les détecter.

Pour chacun de ces 8 défauts, une période pendant laquelle ce défaut est considéré comme présent, appelée « plage de défaut », est définie. Durant cette période, le comportement de la machine est dit anormal. Les échantillons de ces périodes sont exclus des phases d'apprentissage des modèles de « bon fonctionnement ».

Les 8 défauts de la génératrice sont considérés comme les seuls événements qui ont eu un impact sur l'état de dégradation des parties de la génératrice. La base de données, en se concentrant sur ces défauts, est construite de manière à isoler la génératrice du reste des composants. Cela permet d'être sûr qu'une variation significative observée dans les indicateurs sera bien due à un des 8 défauts.

Les 4 défauts ventilateur sont traités différemment des 4 autres défauts critiques en raison des différences dans l'apparition et de la manifestation du défaut sur le fonctionnement de l'éolienne. En effet, la défaillance des ventilateurs est apparue de manière soudaine et a duré pendant une longue période jusqu'à la réparation. Ces défauts n'ont pas entraîné l'arrêt de l'éolienne. Pour ces défauts, la date de fin de la période de défaut est l'instant de réparation. Pour les défauts de roulements et d'enroulements, la date de fin de la période de défaut est la date de défaillance totale où l'éolienne n'a pas pu être redémarrée sans remplacer la pièce.

La date de début des plages de défaut est définie avec l'aide des experts de VALEMO, en fonction des informations récoltées sur les défauts. Cette date est approximative. Afin de s'assurer que la totalité du défaut est contenue dans la plage défaut, une marge de sécurité est toujours prise. On suppose ainsi qu'une partie de la plage de défaut est composée d'échantillons en bon fonctionnement. Cela ne pose pas de problème par la suite puisque toutes les méthodes sont comparées de manière relative sur les mêmes plages. Ce découpage est décrit dans la Figure 6.1. En vert est représenté le comportement des indicateurs en temps normal. Ces indicateurs oscillent autour de zéro. Lorsque le défaut apparaît, l'amplitude de l'indicateur a tendance à augmenter, car nous nous intéressons uniquement ici à détecter des surchauffes. Cette tendance persiste jusqu'à la panne totale de la machine. Ensuite, une fois le composant remplacé, l'indicateur revient osciller autour de zéro.

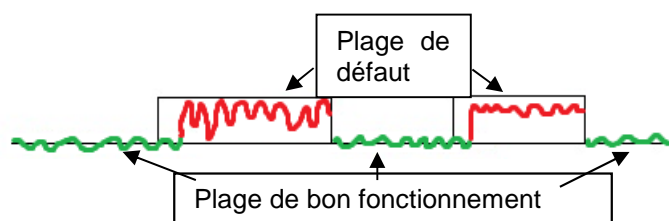


En vert : l'éolienne est en fonctionnement normal

En rouge : l'éolienne à un défaut

Figure 6.1 Schéma de principe de découpage en plages de bon fonctionnement et de défaut pour un défaut progressif

Pour les défauts de ventilateur, la date de fin de la période de défaut est le moment où la réparation a eu lieu. Ces défauts n'ont pas mis la machine à l'arrêt. Ils sont arrivés soudainement et ont continué sans que le SCADA détecte le dysfonctionnement. Seule une étude approfondie par des experts des variables et indicateurs a pu permettre de situer le début du défaut et de fixer une date approximative de début. Afin de s'assurer que le défaut est bien détecté au plus tôt par toutes les méthodes, une semaine supplémentaire a été ajoutée à cette première date de début, comme le décrit la Figure 6.2.



En vert : l'éolienne est en fonctionnement normal

En rouge : l'éolienne à un défaut

Figure 6.2 Schéma de principe de découpage en plages de bon fonctionnement et de défaut pour un défaut soudain

Finalement, les échantillons en dehors de ces périodes sont considérés comme des échantillons de bon fonctionnement. Ainsi, la base de données est découpée en alternant périodes de défaut et périodes de bon fonctionnement, définies dans le Tableau 6.1.

Tableau 6.1 Définitions des différentes périodes de défaut et de bon fonctionnement

Description	Durée (en nombre d'échantillons)	Début	Fin
A	10641	01/11/2009	13/01/2010
Défaut Roulement arrière WT6	6641	13/01/2010	01/03/2010
B	27720	01/03/2010	09/09/2010
Défaut Roulement avant WT9	2785	09/09/2010	28/09/2010
Défaut Refroidissement WT9	3238	29/09/2010	22/10/2010
C1	25273	22/10/2010	15/04/2011
C2	25273	15/04/2011	08/10/2011
Défaut Refroidissement WT7	24913	08/10/2011	29/03/2012
D	27681	29/03/2012	07/10/2012
Défaut Refroidissement WT6	12641	07/10/2012	03/01/2013
Défaut Enroulement WT11	8785	03/01/2013	05/03/2013
E	25201	05/03/2013	27/08/2013
Défaut Refroidissement WT10	11377	27/08/2013	14/11/2013
Défaut Enroulement WT9	15552	15/09/2013	01/01/2014

Le franchissement d'un seuil par l'indicateur est considéré comme une bonne détection si il se produit dans une plage de défaut, sinon, il s'agit d'une fausse alarme.



### 6.1.2. Définition des paramètres $N_A$ , $N_{Emin}$ , $f$ et $seuil_{production}$

Comme nous l'avons vu dans le chapitre précédent, il existe plusieurs paramètres à régler en fonction des méthodes utilisées.

Le paramètre  $seuil_{production}$  permet d'exclure les échantillons obtenus quand la machine n'est pas en production. L'éolienne est considérée en production quand la puissance fournie est au-dessus de  $seuil_{production} = 50kW$ . De plus, les indicateurs réalisés à partir de modèles dynamiques utilisant les liens de causalité sont calculés uniquement pendant les phases de régime permanent. On considère que l'éolienne est dans une phase de régime permanent si elle produit une puissance active supérieure à 50kW depuis au moins 3 heures.

Dans les approches multi-turbines, il est nécessaire de fixer un nombre minimal de turbines valides,  $N_{Emin}$ , pour pouvoir calculer la médiane des résidus (ou des variables brutes), qui va servir de référence parc. Le nombre de turbines du parc étudié vaut  $N_T = 6$ . Par conséquent, nous avons choisi de fixer la valeur de  $N_{Emin}$  à 3. Il faut qu'au moins la moitié des éoliennes du parc aient des variables ou résidus valides pour que la médiane soit calculée.

$N_A$  définit la taille de la fenêtre d'apprentissage. C'est le nombre d'échantillons nécessaires au réapprentissage du modèle, après une période de défaut. Pour définir sa valeur, plusieurs essais ont été effectués. Les indicateurs et critères de performances ont été calculés pour des valeurs de  $N_A$  en nombre d'échantillons égales à: 144, 500, 1000, 2000, 3000, ce qui correspond à un jour, une demi-semaine, une semaine, deux semaines, trois semaines. Les meilleurs résultats ont été obtenus pour  $N_A = 2000$  échantillons soit 2 semaines. Cette fenêtre est suffisamment longue pour pouvoir prendre en compte la périodicité journalière comprise dans les signaux et couvrir la majorité des conditions de fonctionnement. Tous les résultats présentés par la suite sont obtenus avec cette fenêtre d'apprentissage.

Nous avons fait varier la taille de la fenêtre de moyenne mobile  $f$  utilisée comme post traitement des méthodes mono-turbine. Cette fenêtre  $f$  est utilisée aussi dans les méthodes multi-turbines pour comparer les éoliennes entre elles sur une période de temps plutôt que sur un seul échantillon. En effet, cette comparaison sur une période prend la forme de la comparaison de moyennes mobiles de fenêtre  $f$  (cf partie 5.4.2.2.2 et 5.4.2.3.2) Afin d'étudier l'impact de la taille de cette fenêtre sur performances des indicateurs, 5 valeurs de  $f$  sont testées : 1, 36, 72, 144, 500 et 1000 échantillons.

Les résultats de l'évaluation des performances sont présentés dans la suite pour chacune de ces valeurs, qui sont rappelées dans le Tableau 6.2.

Tableau 6.2 Rappel des valeurs des paramètres de réglages

Nom du paramètre	Valeur
$N_A$	2000 échantillons
$N_{Emin}$	3 éoliennes
$N_T$	6 éoliennes
$f$	1, 36, 72, 144, 500 et 1000 échantillons sont testés
$seuil_{production}$ 1, 2 et 3	50 kW

### 6.1.3. Choix des températures utilisées comme indicateurs

La génératrice est composée de plusieurs sous-parties. Les variables disponibles dans la base de données permettent de surveiller les températures du roulement avant, du roulement arrière et des enroulements rotoriques. Les 8 défauts de la base ont un impact différent sur les variables mesurées sur ces 3 parties. Cette différence permet d'isoler le défaut détecté et d'entamer un diagnostic. Il convient donc de construire un indicateur pour chacune de ces parties si l'on veut avoir une surveillance complète de la génératrice. Pour chaque méthode présentée, 3 indicateurs associés à chacune des 3 parties surveillées sont calculés à l'aide des méthodes décrites dans le chapitre 5, chacun utilisant soit :

- La température du roulement avant, notée  $T_{rav}$
- La température du roulement arrière, notée  $T_{rar}$
- La température de l'enroulement statorique, notée  $T_{es}$

Chacune de ces 3 températures et donc les indicateurs qui leur sont associés sont sensibles à certains types de défauts. Pour plus de lisibilité, seul l'indicateur le plus sensible est présenté dans les courbes ROC et la persistance, pour l'évaluation des performances des méthodes.

Les courbes ROC et la persistance sont calculés pour chaque défaut. Il y a 8 courbes ROC et 8 courbes représentant la persistance pour évaluer la performance d'un modèle sur la base de données VALEMO. Pour chaque courbe, seul l'indicateur le plus sensible est représenté. De plus, le nombre d'interventions inutiles présenté est le nombre d'interventions générés uniquement par l'indicateur le plus sensible.

#### 6.1.3.1. Indicateurs issus de modèles utilisant les liens de causalité

Les températures des 3 parties de la génératrice à surveiller sont estimées en utilisant les 3 équations :

$$\hat{T}_{rav}^i(t) = a_1^i \cdot T_{rav}^i(t-1) + b_1^i \cdot P^i(t) + c_1^i \cdot V^i(t) + d_1^i \cdot T_{nacelle}^i(t) + e_1^i \quad (6.1)$$

$$\hat{T}_{rar}^i(t) = a_2^i \cdot T_{rar}^i(t-1) + b_2^i \cdot P^i(t) + c_2^i \cdot V^i(t) + d_2^i \cdot T_{nacelle}^i(t) + e_2^i \quad (6.2)$$

$$\hat{T}_{es}^i(t) = a_3^i \cdot T_{es}^i(t-1) + b_3^i \cdot P^i(t) + c_3^i \cdot V^i(t) + d_3^i \cdot T_{nacelle}^i(t) + e_3^i \quad (6.3)$$

où

$T_{rav}^i$  est la température du roulement avant de la turbine  $i$

$T_{rar}^i$  est la température du roulement arrière de la turbine  $i$

$T_{es}^i$  est la température de l'enroulement stator de la turbine  $i$

$T_{nacelle}^i$  est la température de la nacelle de la turbine  $i$

$P^i$  est la puissance active générée par la turbine  $i$

$V^i(t)$  est la vitesse de l'arbre rapide de la turbine  $i$

et  $a$ ,  $b$ ,  $c$ ,  $d$  et  $e$  sont des scalaires.

#### 6.1.3.2. Indicateurs issus de modèles utilisant la similarité

Les méthodes utilisant la similarité nécessitent le choix d'une variable de référence. La génératrice possède des roulements avant et arrière de même modèle. Ainsi, il semble naturel d'utiliser la température du roulement opposé comme variable d'entrée du modèle permettant la surveillance d'un

roulement. En revanche, pour la surveillance du stator, c'est la température du roulement arrière qui est utilisée, car il n'y a pas d'autres pièces de la même nature. Le choix du roulement arrière est justifié par le fait que ce roulement est le plus proche des stators.

Afin de valider ce choix des composants de référence, une étude sur la corrélation entre les différents couples de variables a été faite. Le choix des composants de référence pour les modèles est réalisé grâce à l'étude des coefficients de détermination qui exprime le taux de variance expliqué par le modèle linéaire, comme le montrent les figures Figure 6.3, Figure 6.4 et Figure 6.5 suivantes:

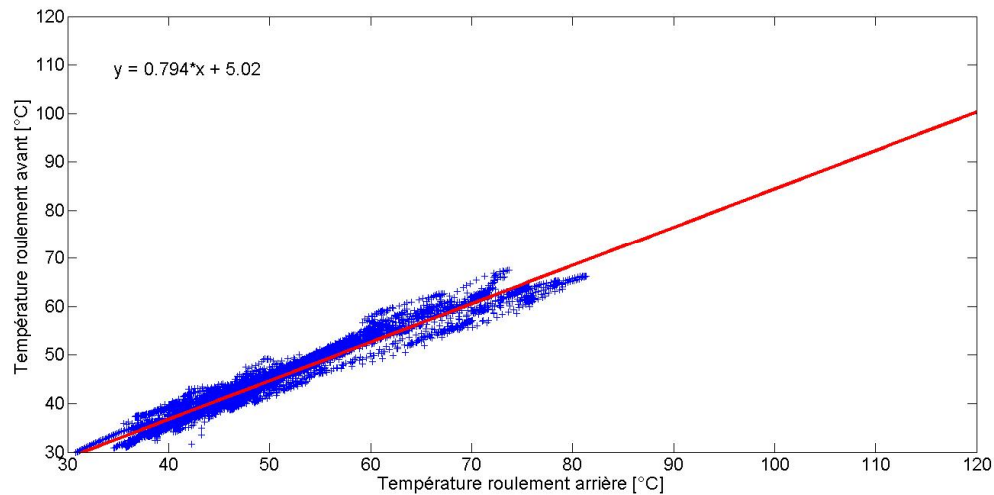


Figure 6.3 Température du roulement avant en fonction de la température du roulement arrière, coefficient de détermination  $R^2 = 0.96$

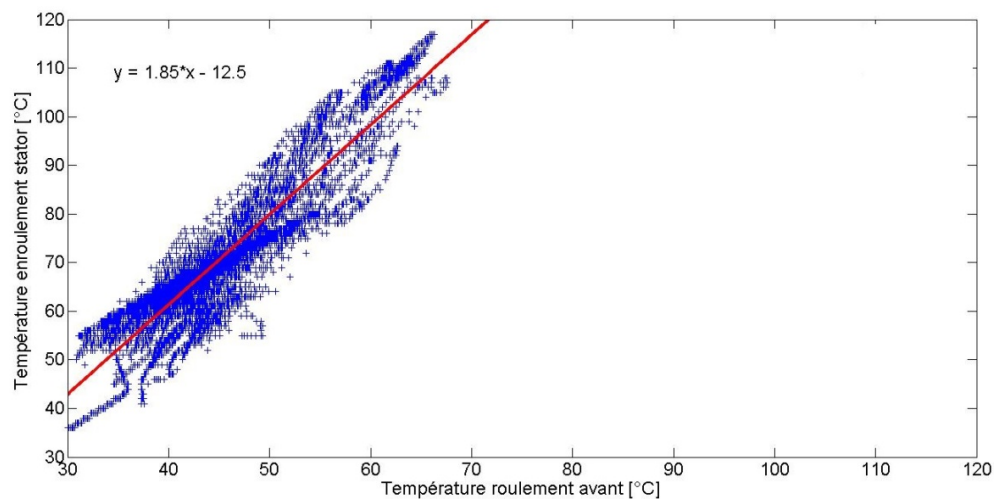


Figure 6.4 Température de l'enroulement stator en fonction de la température du roulement avant, coefficient de détermination  $R^2 = 0.86$

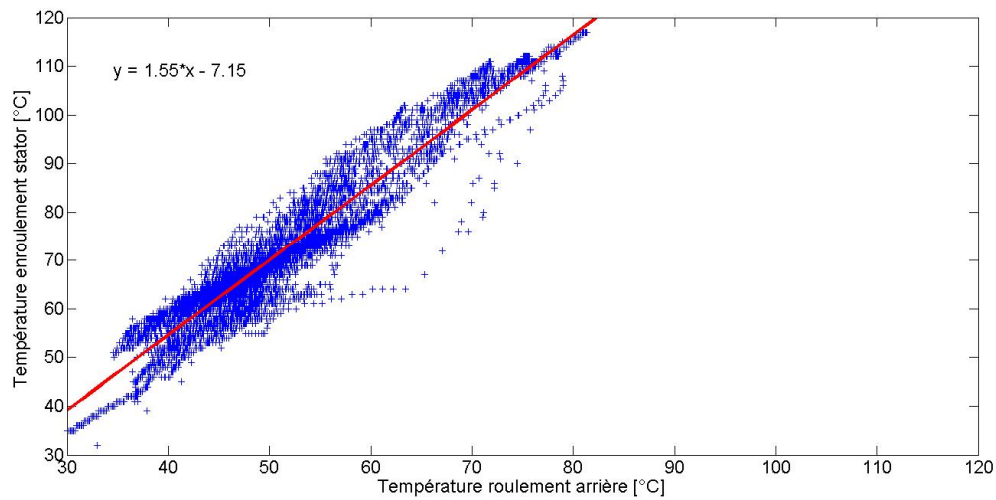


Figure 6.5 Température de l'enroulement stator en fonction de la température du roulement arrière, coefficient de détermination  $R^2 = 0.91$

Le coefficient de détermination est plus élevé pour le couple de température stator vs roulement arrière  $R^2 = 0.91$  plutôt que stator roulement vs roulement avant  $R^2 = 0.86$ . Par conséquent, on choisit la température du roulement arrière ( $T_{rar}^i$ ) comme variable de référence pour estimer la température des enroulements stator.

Le Tableau 6.3 récapitule la variable de référence utilisé en entrée des modèles utilisant la similarité pour estimer les trois températures de la génératrice.

Tableau 6.3 Choix des variables de références pour la méthodes mono-turbine utilisant la similarité

Partie de la génératrice $j$ à surveiller	Partie de référence ( $ref_j$ )
Roulement avant 1	Roulement arrière 2
Roulement arrière 2	Roulement avant 1
Enroulement stator 3	Roulement arrière 2

Pour cette évaluation de performances, nous utiliserons 3 modèles pour créer les indicateurs de surveillance des méthodes utilisant la similarité :

$$\hat{T}_{rav}^i(t) = a_1^i \cdot T_{rar}^i(t) + b_1^i \quad (6.4)$$

$$\hat{T}_{rar}^i(t) = a_2^i \cdot T_{rav}^i(t) + b_2^i \quad (6.5)$$

$$\hat{T}_{es}^i(t) = a_3^i \cdot T_{rar}^i(t) + b_3^i \quad (6.6)$$

où

$T_{rav}^i$  est la température du roulement avant de la turbine  $i$ , à l'instant  $t$

$T_{rar}^i$  est la température du roulement arrière de la turbine  $i$ , à l'instant  $t$

$T_{es}^i$  est la température de l'enroulement du stator de la turbine  $i$ , à l'instant  $t$

et  $a$  et  $b$  sont des paramètres appris à partir des données.

## 6.2. Performances des indicateurs utilisant des variables SCADA brutes

Cette partie se concentre sur la présentation de l'évaluation des performances des nouvelles méthodes présentées dans le chapitre 5, utilisant les variables brutes. Les performances obtenues lorsque les températures des parties de la génératrice sont utilisées directement comme indicateurs de surveillance, comme cela est fait dans le SCADA, sont présentées. Ensuite, les performances obtenues en comparant les températures entre éoliennes seront présentées.

### 6.2.1. Utilisation des variables brutes, approche mono-turbine

Les variables venant du SCADA peuvent être utilisées directement comme indicateurs de surveillance. La Figure 6.6 montre la température du roulement arrière pour chacune des 6 turbines pendant les 4 ans de la base de données. Le défaut de roulement sur la turbine WT6, le défaut de ventilateur de la turbine WT7 et le défaut de ventilateur de la turbine WT10 sont mis en évidence par des flèches. On observe que les températures varient suivant le cycle des saisons, cela met en évidence l'effet de l'environnement sur les données SCADA. De plus, des pics de température importants sont aussi visibles. Ils correspondent aux périodes de forte production, ce qui montre l'influence du fonctionnement de la machine sur la température des pièces qui la compose.

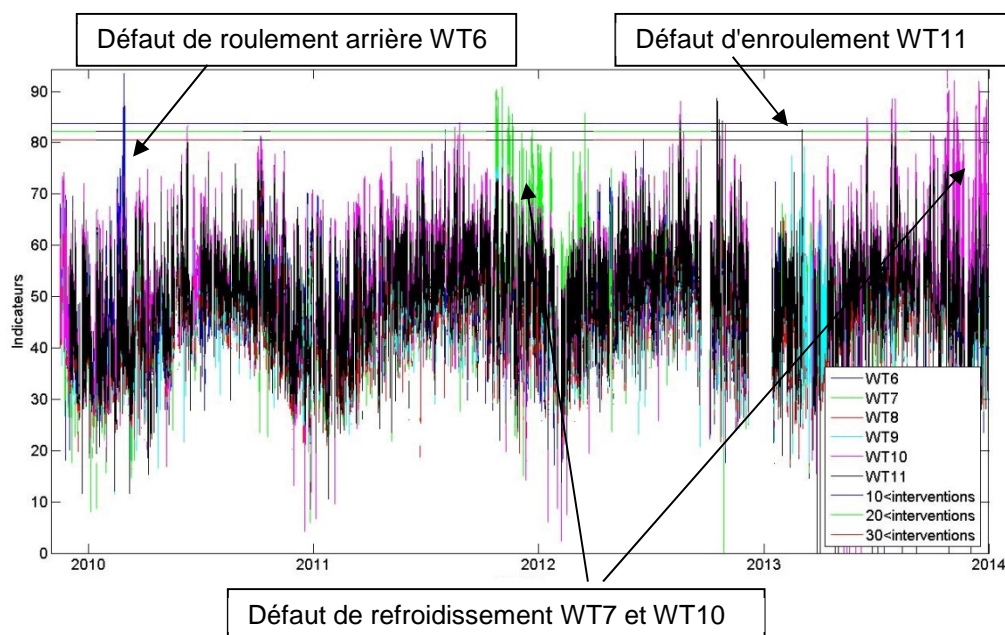


Figure 6.6 Évolution de la température du roulement arrière de chaque éolienne

Néanmoins, malgré ces fortes variations, il est possible d'observer les trois défauts apparaissant début 2010, fin 2011 et fin 2014. 5 seuils de détections sont symbolisés par un trait horizontal sur la Figure 6.6. Ces seuils sont définis pour avoir moins de 10, 20 ou 30 interventions inutiles avec cet indicateur sur la totalité de la base, soit 4 ans. L'analyse visuelle de l'indicateur nous montre une vraie détection, mais dans les faits, si l'on souhaite placer un seuil, les défauts critiques de roulement et d'enroulement ne peuvent être détectés que très tardivement, si l'on veut un nombre d'interventions inutiles acceptable.

Le lien entre le nombre d'interventions inutiles et l'avance à la détection est crucial pour un industriel. En effet, le but est d'avoir une méthode capable de détecter le défaut à temps (soit au plus tard 1000

échantillons avant l'arrêt machine, c'est-à-dire une semaine) et avec un nombre d'interventions inutiles raisonnable. Les Figure 6.7 et Figure 6.8 présentent les courbes ROC décrites dans la partie 5.3.1, montrant l'avance à la détection en fonction du nombre d'interventions inutiles générées sur toutes les plages de bon fonctionnement durant les 4 ans de la base de données VALEMO. Ces courbes sont tracées en faisant varier la valeur du seuil de détection. Plusieurs courbes sont présentées, elles sont obtenues pour différentes tailles de la fenêtre de lissage par moyenne mobile  $f$ . L'axe des ordonnées représente l'avance à la détection allant de zéro à la taille maximum de la période de défaut étudiée. Ces courbes sont construites en prenant l'indicateur le plus sensible pour chaque défaut entre chaque partie de la génératrice. Parmi les trois indicateurs surveillant les 3 parties de la génératrice, seul l'indicateur ayant la meilleure persistance sur ce défaut est représenté. La Figure 6.7 présente les courbes ROC pour les défauts critiques (défaut de roulement avant WT6, défaut de roulement arrière WT9 et les défauts d'enroulement rotorique WT11 et WT9).

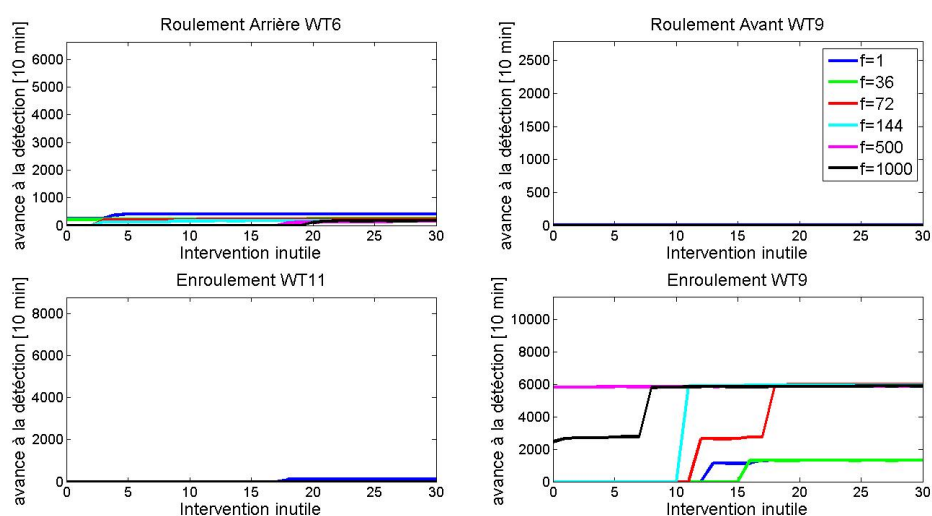


Figure 6.7 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant les températures brutes, approche mono-turbine

La Figure 6.7 permet de voir que, parmi les défauts critiques, seul le défaut d'enroulement est détecté suffisamment tôt avant la panne complète de la turbine. Le défaut de roulement arrière est détecté, mais, comme le montre la Figure 6.6 et la Figure 6.7, cette détection se fait alors que le défaut est très avancé, seulement 411 échantillons 10 minutes en avance soit 3 jours environ seulement avant la panne totale.

La Figure 6.8 présente les courbes ROC pour les défauts de ventilateurs des turbines WT9, WT7, WT6 et WT10. Cette figure montre, en revanche, que 3 des défauts de ventilateurs peuvent être détectés avec une avance à la détection de plus de 1000 échantillons et un nombre d'interventions inutiles acceptable (inférieur à 30 sur 4 ans).

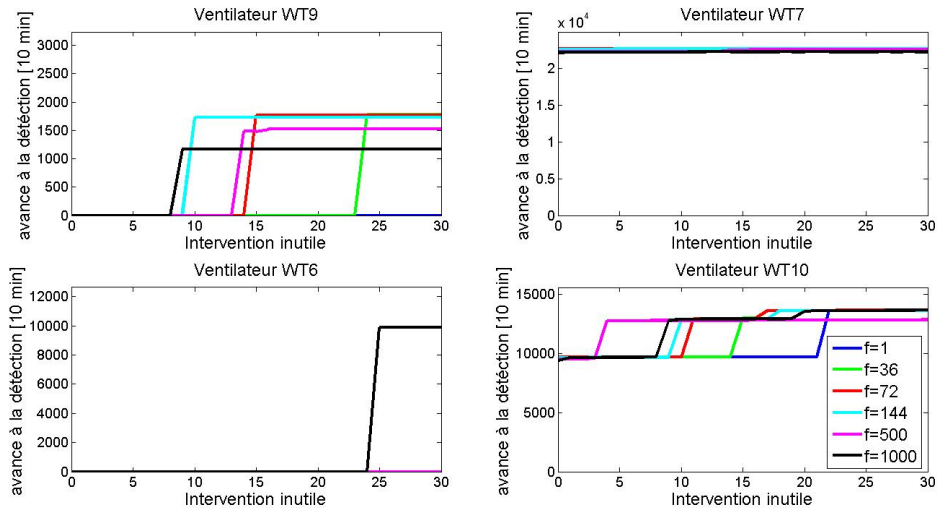


Figure 6.8 Avance à la détection vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche mono-turbine

La Figure 6.9 montre la persistance des indicateurs utilisant les variables brutes pour chacun des 8 défauts. Plusieurs tailles de fenêtre  $f$  pour le calcul de la moyenne mobile sont testées. La persistance est le rapport entre le nombre d'échantillons au-dessus du seuil et le nombre d'échantillons total de la période en défaut. Cette figure montre que la persistance de l'indicateur n'est pas suffisamment élevée pour dire qu'il s'agit d'une réelle détection et non pas de simples pics.

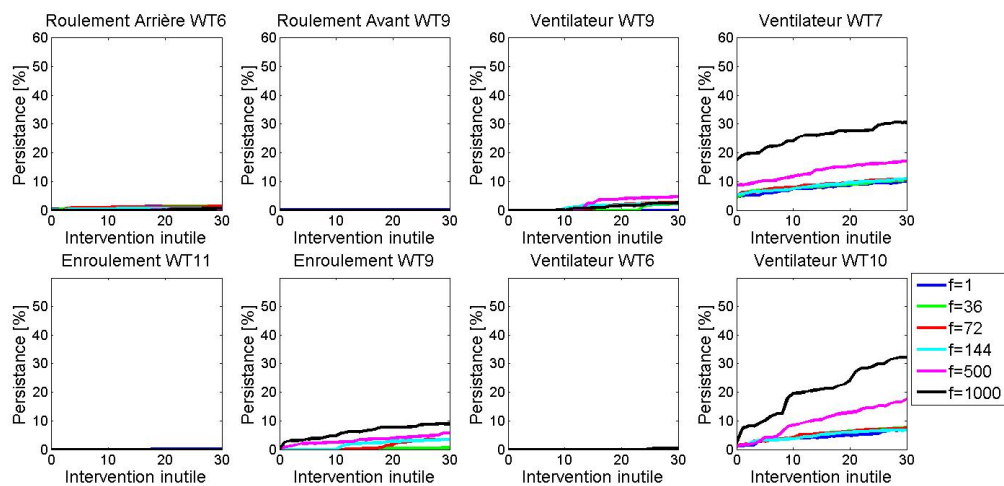


Figure 6.9 Persistance vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche mono-turbine

On peut voir sur la Figure 6.9 qu'en utilisant une fenêtre de 1000 points, il est possible d'avoir les meilleurs résultats sur les ventilateurs, tout en gardant de bons résultats sur l'enroulement. Cette taille de fenêtre permet aussi d'améliorer au mieux la persistance.

Ces résultats ne sont pas suffisants pour permettre une réelle surveillance de la génératrice. En effet, les défauts critiques de la base ne sont pas tous détectés en générant moins de 30 interventions inutiles. Néanmoins, ils permettent de fixer les performances minimales à dépasser pour les autres



méthodes. En effet, l'utilisation des variables brutes correspond à la surveillance actuellement en place sur le SCADA, où l'opérateur est alerté quand la température franchit un seuil. Cela permet déjà, en utilisant une moyenne mobile, de détecter les défauts de ventilateurs et un défaut d'enroulement. Il est important de voir que ces défauts peuvent être détectés sans interventions inutiles. Néanmoins, c'est la capacité à détecter les défauts critiques qui sont les défauts de roulement et d'enroulement qui est importante pour la surveillance.

### 6.2.2. Utilisation des variables brutes, approche multi-turbine

La Figure 6.10 montre l'évolution de la température du roulement arrière de chaque éolienne pendant la période du défaut de roulement arrière de l'éolienne WT6. On peut voir que les températures de roulement évoluent de manière simultanée entre les turbines. Seule la turbine WT6 a une température de plus en plus élevée par rapport aux autres. La comparaison entre les turbines vise à s'appuyer sur ces comportements pour générer des indicateurs de défaut.

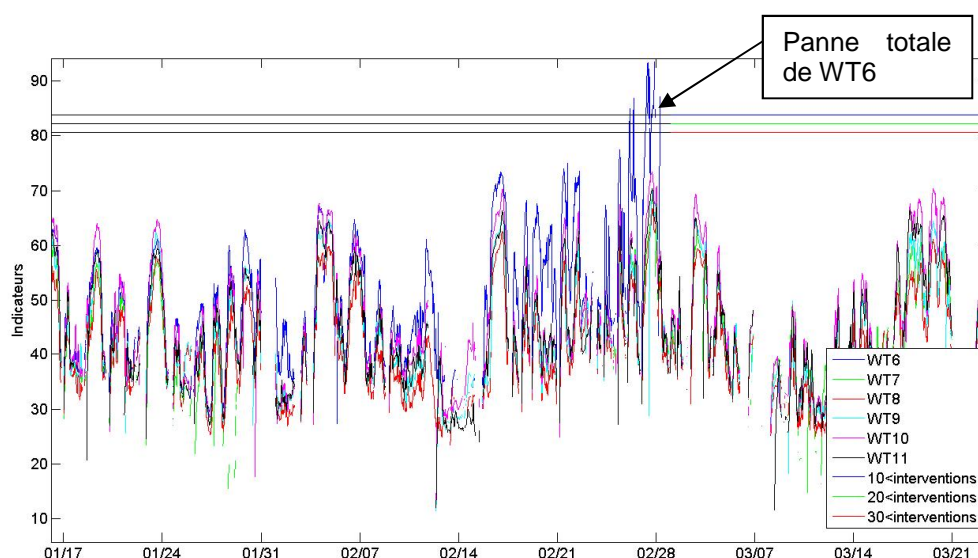


Figure 6.10 Évolution de la température du roulement arrière de chaque éolienne, pendant le défaut de roulement arrière WT6

La Figure 6.11 montre l'évolution de l'indicateur créé avec une approche multi-turbine utilisant les variables de température brutes décrite dans la partie 5.4.2, sur les 4 ans de la base. Cette figure permet de visualiser distinctement l'effet du défaut de roulement avant WT6, du défaut d'enroulement WT11 et des défauts de ventilateurs WT7 et WT10. En effet, en comparant les turbines entre elles, l'indicateur s'affranchit de la saisonnalité des variables d'origine. Cette saisonnalité n'est pas le seul effet de l'environnement qui a pu être diminué, ce qui permet de voir apparaître très clairement de nouveaux défauts critiques. En plus de pouvoir observer les défauts, un seuil peut maintenant être fixé pour les détecter avec moins de 10 interventions inutiles.



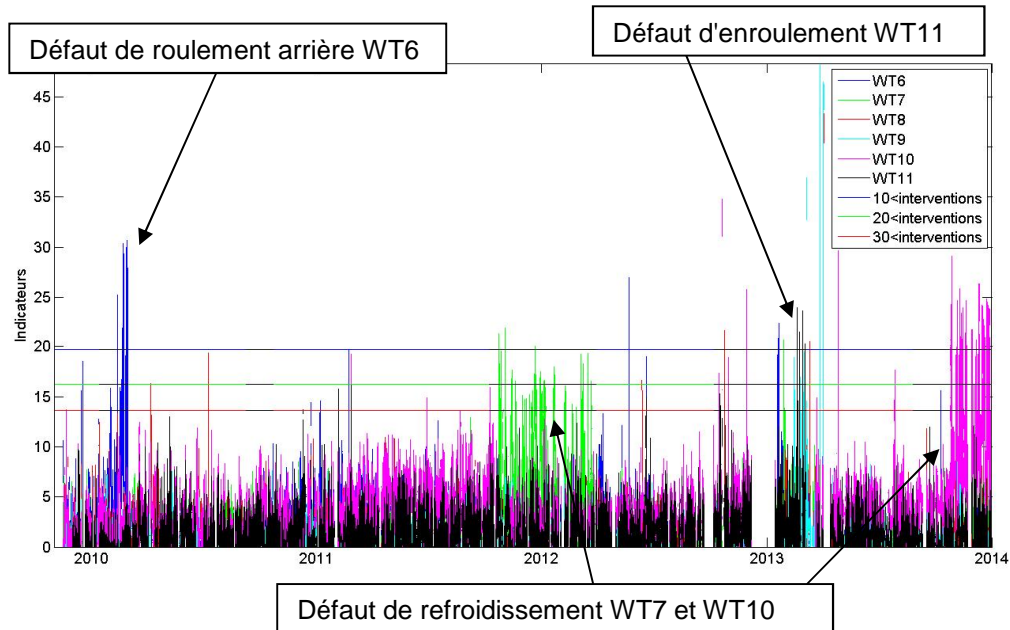


Figure 6.11 Évolution de la distance entre la température du roulement arrière de chaque éolienne et la référence parc, approche multi-turbine

Comme le montrent les figures Figure 6.12 et Figure 6.13 qui présentent les courbes ROC d'avance à la détection en fonction du nombre d'interventions inutiles pour chaque défaut, l'utilisation d'une approche multi-turbine avec des variables brutes permet de détecter certains défauts critiques plus de 1000 échantillons en avance, en conservant un nombre d'interventions inutiles faible (inférieur à 10). Elles montrent que la détection est devenue possible sur les défauts critiques, néanmoins le défaut de ventilateur de la turbine WT9 n'est plus visible.

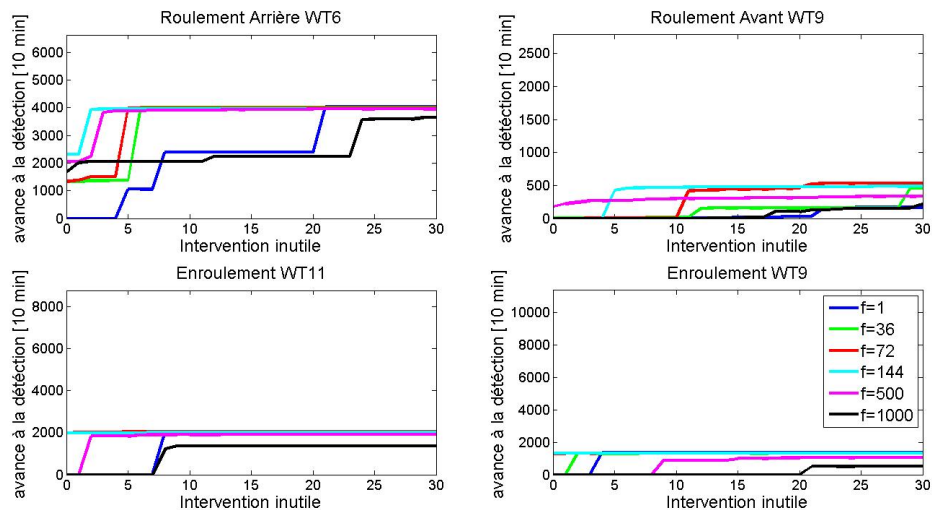


Figure 6.12 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant les températures brutes, approche multi-turbine

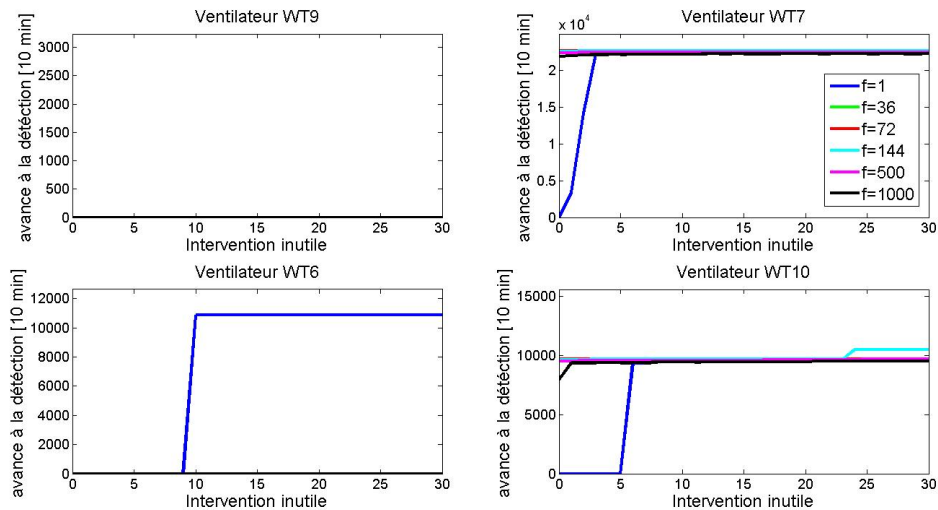


Figure 6.13 Avance à la détection vs nombre d'interventions inutiles des défauts de ventilateur, en utilisant les températures brutes, approche multi-turbine

La Figure 6.14 montre la persistance de l'approche multi-turbine utilisant les températures brutes pour chaque défaut. Elle confirme que la détection n'est pas simplement faite de pics isolés. La persistance calculée pour les défauts critiques d'enroulement montre que ceux-ci ne sont visibles que brièvement toutefois par rapport au défaut de roulement. Cela est peut-être dû à la nature du défaut qui peut avoir moins duré dans le temps et être moins progressif.

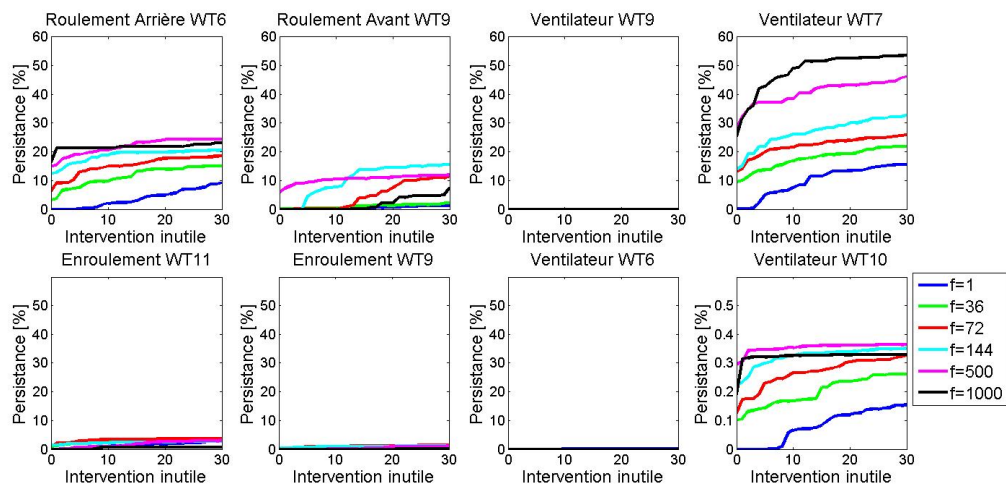


Figure 6.14 Persistance vs nombre d'interventions inutiles des défauts, en utilisant les températures brutes, approche multi-turbine

Il est aussi important de noter que la comparaison des températures brutes sur une plage de temps permet d'augmenter les performances tant que la taille la période de comparaison  $f$  reste inférieure à 1000. À partir de cette taille, l'effet de moyenne sur les  $f$  a un impact négatif par rapport à la comparaison instantanée. Pour rappel, la fenêtre  $f$  correspond à la fois à la taille de la fenêtre de la moyenne mobile utilisée comme post traitement des méthodes mono-turbines et à la taille de la période utilisé pour la comparaison dans l'approche multi-turbine.

Tableau 6.4 Disponibilité des indicateurs construits à partir des méthodes utilisant les variables brutes  
 (pour 1 315 008 échantillons dans la base)

		Mono		Multi	
		Nombre d'échantillons valides	Disponibilité [%]	Nombre d'échantillons valides	Disponibilité [%]
Roul. Avant	f=1	873182	66.40%	815486	62.01%
	f=36	818963	62.28%	761551	57.91%
	f=72	806951	61.36%	746427	56.76%
	f=144	800636	60.88%	736451	56.00%
	f=500	795252	60.48%	725828	55.20%
	f=1000	789231	60.02%	718457	54.64%
Roul. Arrière	f=1	874920	66.53%	817197	62.14%
	f=36	820697	62.41%	763141	58.03%
	f=72	808605	61.49%	747882	56.87%
	f=144	802379	61.02%	737995	56.12%
	f=500	799075	60.77%	729481	55.47%
	f=1000	791605	60.20%	721826	54.89%
Enroul. Stator	f=1	879273	66.86%	821630	62.48%
	f=36	825140	62.75%	767609	58.37%
	f=72	813405	61.86%	752539	57.23%
	f=144	808817	61.51%	744286	56.60%
	f=500	805039	61.22%	735549	55.93%
	f=1000	796363	60.56%	726771	55.27%

Le Tableau 6.4 montre la disponibilité des indicateurs utilisant les variables brutes pour des approches mono-turbines et multi-turbines et pour chacune des 3 parties de la génératrice. Ce tableau indique le nombre d'échantillons valides sur les 13150008 échantillons de la base de données pour différentes tailles de fenêtre  $f$ . Il permet de voir dans un premier temps que la disponibilité des variables brutes est sensiblement identique entre les 3 parties de la génératrice. Il est important de rappeler que toutes les méthodes ne sont valides que si la valeur de puissance active dépasse un seuil de production. Le même seuil est choisi pour chacune des méthodes, soit 50kW. Les éoliennes ont une production variable et ne produisent pas en continu (cf Figure 4.6 de la partie 4.2). Plus de 30% du temps, les éoliennes produisent moins de 50kW. À cela vient s'ajouter des pertes de communication ou des défauts de capteurs, ce qui au final donne une disponibilité maximale de 66% pour les variables brutes. Cette disponibilité peut paraître faible mais l'analyse des durées des périodes d'indisponibilité montre que l'indicateur que l'on peut créer avec les variables brutes n'est jamais indisponible plus d'une journée.

On peut voir que l'utilisation d'une moyenne mobile pour l'approche mono-turbine ou d'une période de comparaison pour l'approche multi-turbine diminue la disponibilité. En effet, le calcul de moyenne n'est considéré comme valide que si au moins la moitié des échantillons utilisés est valide (cf parties 5.4.1, 5.4.2.2.2 et 5.4.2.3.2). Plus la taille de la fenêtre est importante et plus la disponibilité est faible.

En effet, si  $f$  est choisie égale à 1000 échantillons alors la disponibilité passe de 66% à 60%. Finalement, ce tableau montre l'impact d'une approche multi-turbine sur la disponibilité. La méthode proposée cherche à comparer les éoliennes entre elles et, pour cela, doit construire une référence parc. Cette référence parc n'est valide que si au moins 3 éoliennes sont en fonctionnement. Par conséquent, étant donné que les turbines peuvent avoir du retard les unes par rapport aux autres (cf partie 5.4.2.2.2) et qu'elles peuvent avoir des défaillances en même temps, le nombre d'échantillons valides a tendance à diminuer en perdant environ 5% pour chacune des différentes tailles de fenêtre  $f$ . Cependant, la disponibilité est toujours suffisante pour assurer une surveillance quotidienne.

En conclusion sur l'utilisation des variables brutes, les défauts critiques ne peuvent pas être détectés avec une approche mono-turbine. En revanche, l'approche multi-turbine permet d'avoir des résultats relativement intéressants. Cette solution permet à moindre coût de détecter les défauts critiques avec un niveau d'interventions inutiles acceptable. En effet, il est important de rappeler que cette solution ne nécessite aucun apprentissage, elle peut donc être mise en place très facilement et apporter une première solution à la surveillance des composants des éoliennes. De plus, cette solution simple ne nécessite aucune modélisation préalable. Peu de travaux ont été réalisés sur l'approche multi-turbine, et malgré sa simplicité et ses bonnes performances, la comparaison de variables brutes n'a jamais été mise en œuvre en pratique ou servie de référence aux méthodes présentées dans la littérature.

Les méthodes présentées ensuite vont permettre de voir l'apport de l'utilisation des résidus dans la comparaison entre les turbines. Deux types de méthodes sont présentés par la suite pour construire le modèle et construire les résidus dans une approche mono-turbine : les modèles utilisant les liens de causalité et les modèles basés sur la similarité.

### 6.3. Indicateurs utilisant les résidus issus de modèles utilisant les liens de causalité

Cette partie se concentre sur la présentation de l'évaluation des performances des nouvelles méthodes présentées dans le chapitre 5, utilisant les liens de causalité. Les performances obtenues lorsque l'écart entre la valeur estimée et la valeur réelle est utilisé comme indicateur sont présentées. Ensuite, les performances obtenues en comparant les résidus des modèles entre les éoliennes sont présentées.

#### 6.3.1. Résidus issus de modèles prenant en compte la dynamique de la température

##### 6.3.1.1. Approche mono-turbine

Le modèle décrit dans la partie 5.4.1.1.1 avec l'équation n°(5.1) est maintenant utilisé pour produire un résidu.

$$\hat{T}_j^i(t) = a_j^i \cdot T_j^i(t-1) + b_j^i \cdot P^i(t) + c_j^i \cdot V^i(t) + d_j^i \cdot T_{nacelle}^i(t) + e_j^i \quad (6.7)$$

$$Résidu_j^i(t) = T_j^i(t) - \hat{T}_j^i(t) \quad (6.8)$$

Cette équation permet de prendre en compte la dynamique du système avec la variable  $T_j^i(t - 1)$

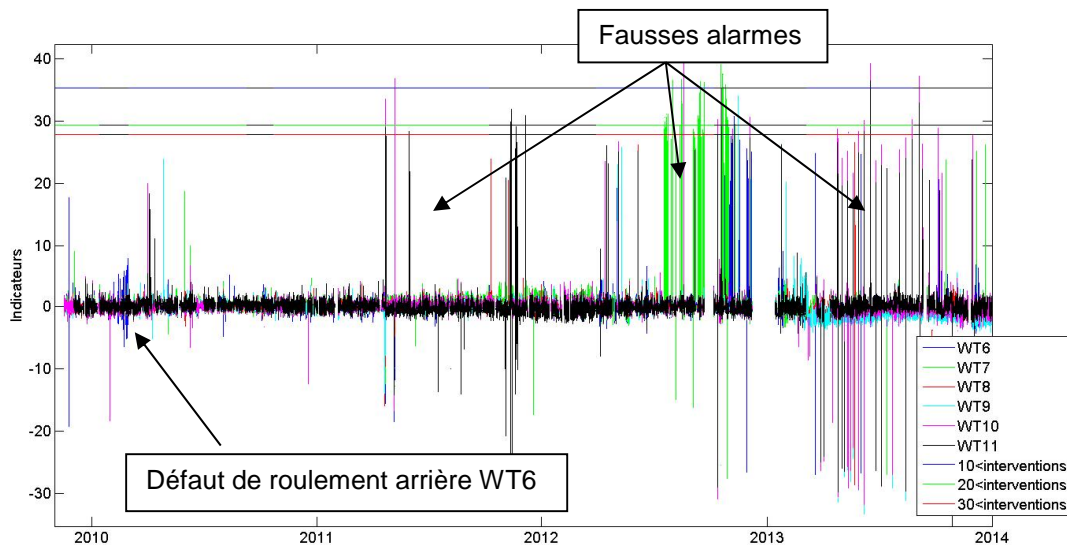


Figure 6.15 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases en fonctionnement du roulement arrière de chaque éolienne, approche mono-turbine

La Figure 6.15 montre l'évolution du résidu associé à la température du roulement arrière, obtenu à partir d'une méthode utilisant les liens de causalité dynamique appliquée sur toutes les phases en fonctionnements sur les 4 ans de la base de données. C'est ce résidu qui sert d'indicateur de défaut. La présence de nombreux pics tout au long de l'année montre que cet indicateur ne peut être utilisé tel quel pour la surveillance. En effet, bien que le défaut de roulement arrière WT6 soit visible au début de l'année 2010, les seuils de détection les plus bas (fixés en fonction du nombre maximum d'interventions inutiles admissibles) restent bien au-dessus du défaut.

L'origine de ces pics (qui nécessitent de remonter les seuils pour diminuer les fausses alarmes) provient des phases de démarrage des machines. En effet, les échanges thermiques semblent ne pas être les mêmes à chacune des phases de fonctionnement de la machine. Le roulement peut doubler sa température en moins de 10 minutes, passant de la température ambiante au repos à la température de fonctionnement, pendant le démarrage. Afin de supprimer ces pics, une solution a été proposée en filtrant les phases de démarrage et en ne gardant que les phases de régime permanent (cf partie 5.4.1.1.2).

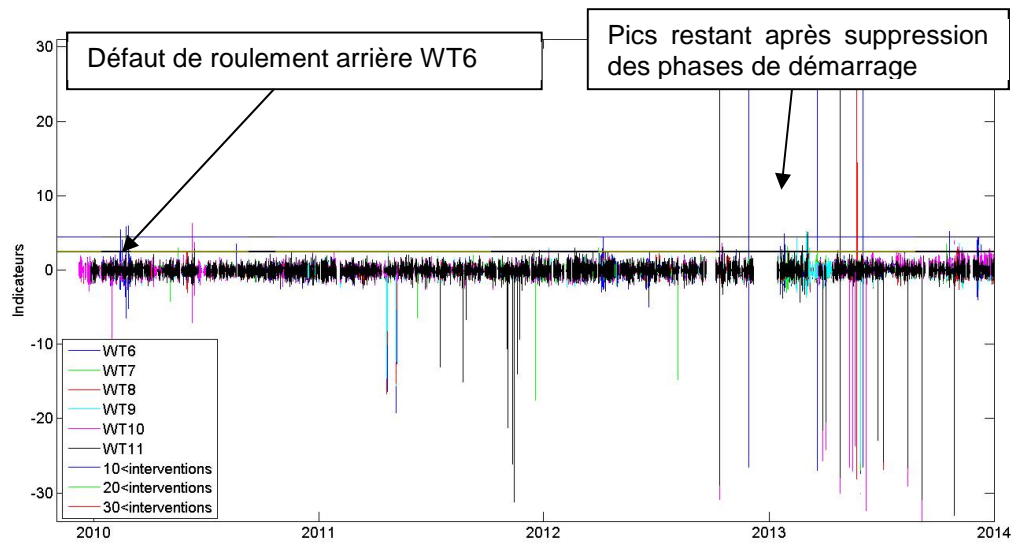


Figure 6.16 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent du roulement arrière de chaque éolienne, approche mono-turbine filtré

L'étude des données et du fonctionnement machine a permis de déterminer qu'il faut au moins 3 heures (3h équivaut à 18 échantillons 10 minutes) pour que les températures des composants atteignent leur régime permanent. La Figure 6.16 montre l'évolution du résidu calculé uniquement sur les périodes de régime permanent. En supprimant les échantillons des phases de régime transitoire, on constate qu'il est possible de réduire drastiquement le nombre de pics.

Les Figure 6.17 et Figure 6.18 montrent les courbes ROC pour chacun des défauts de la méthode utilisant les liens de causalité dynamique appliquée sur les phases de régime permanent. Elles permettent de voir que cette méthode ne peut pas détecter tous les défauts critiques. En filtrant les phases de régime transitoire, il est possible de détecter certains des défauts avec un nombre d'interventions inutiles acceptable. Mais ce filtre fait baisser la disponibilité de l'indicateur de 10% passant de 65% à 54% de disponibilité. Cette baisse reste cependant raisonnable puisqu'elle permet de continuer la surveillance au quotidien des machines. Les défauts de ventilateurs semblent être mieux détectés lorsque l'on regarde l'avance à la détection uniquement.

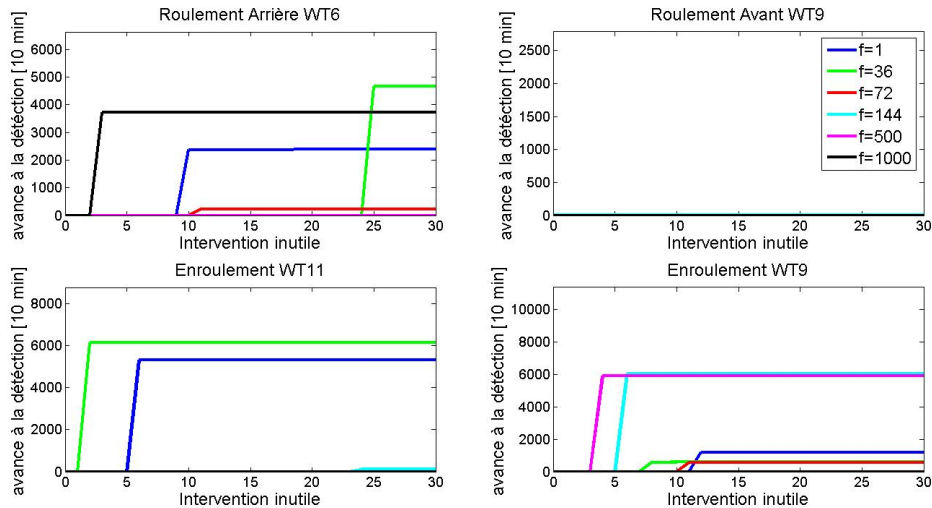


Figure 6.17 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine

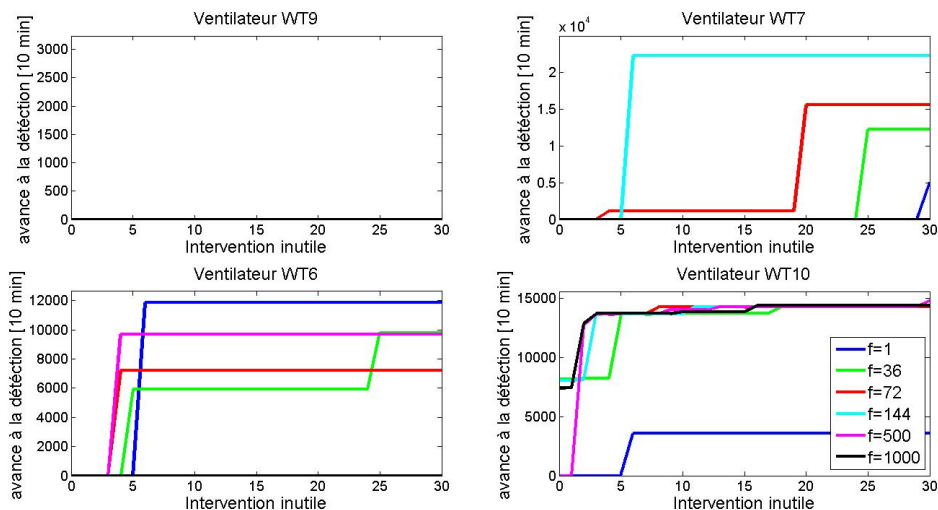


Figure 6.18 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine

La Figure 6.19 montre la persistance obtenue pour les différentes tailles de fenêtre  $f$  pour chaque défaut pour une méthode mono-turbine utilisant des résidus calculés uniquement pendant les phases de régime permanent. La persistance que l'on peut observer sur cette figure montre que seul le défaut de ventilateur de la turbine WT10 est réellement détecté, et qu'il ne s'agit pas de pics isolés. La persistance faible sur les autres défauts montre que l'indicateur ne dépasse le seuil de détection que pendant des périodes transitoires de courte durée.

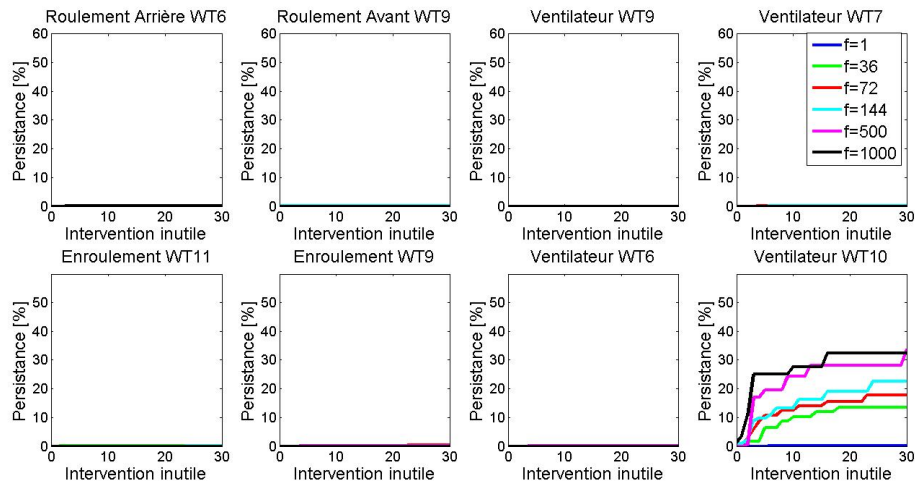


Figure 6.19 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent, approche mono-turbine

La Figure 6.20 montre le comportement de l'indicateur pendant le défaut de roulement arrière de la turbine 6 au début de l'année 2010. La courbe en bleu représente l'indicateur sans moyenne mobile. Il ne génère que quelques pics très courts (un pic par semaine pendant le mois de février qui a précédé la panne totale). Il est difficile, voire impossible, de prendre la décision d'intervenir avec ces informations.

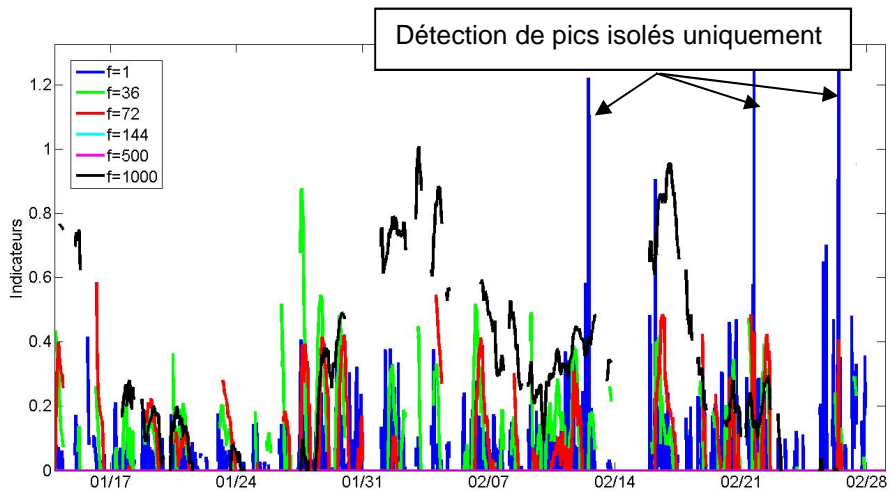


Figure 6.20 Évolution de l'indicateur utilisant les liens de causalité dynamique appliqué sur toutes les phases de régime permanent avec plusieurs taille de fenêtre, approche mono-turbine

Au regard de ces résultats, il est possible de conclure que l'utilisation d'une moyenne mobile n'améliore pas significativement le résultat en terme d'augmentation de l'avance à la détection ou de diminution du nombre d'interventions inutiles. On préfère utiliser cet indicateur sans post traitement.

### 6.3.1.2. Approche multi-turbine

L'approche multi-turbine utilisant les résidus des modèles utilisant les liens de causalité dynamique ne semble pas améliorer significativement les résultats comme le montrent les figures Figure 6.21,



Figure 6.22 et Figure 6.23. La persistance est toujours aussi faible, seuls des pics isolés permettent la détection des défauts, sauf pour le défaut de ventilateur de la turbine WT10.

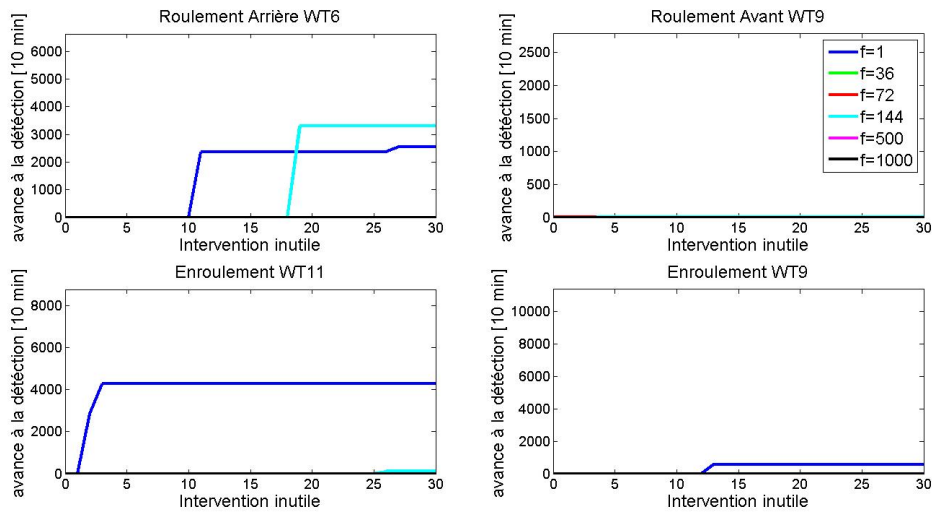


Figure 6.21 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine

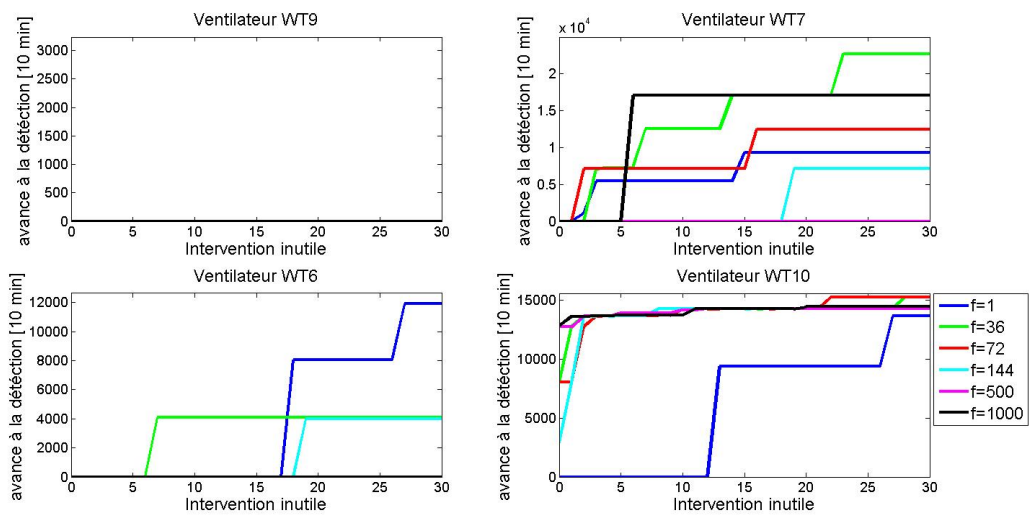


Figure 6.22 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine

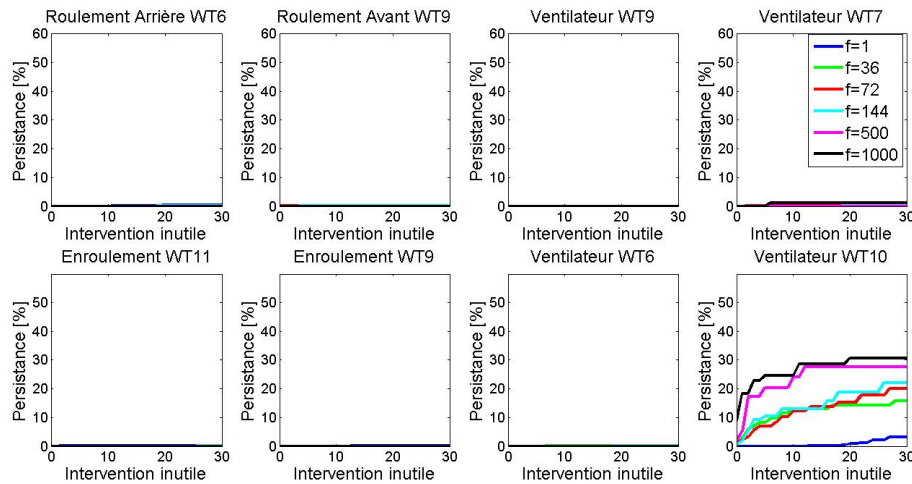


Figure 6.23 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité dynamique, approche multi-turbine

Le Tableau 6.5 montre la disponibilité de chaque indicateur, pour chacune des méthodes utilisant les liens de causalité dynamique et pour chacune des parties de la génératrice. Ce tableau permet de voir le nombre d'échantillons valides et la disponibilité pour les méthodes utilisant les liens de causalité dynamique avec une approche mono-turbine ou multi-turbine. Il permet de constater que le filtre permettant de ne garder que les phases de régime permanent diminue de 10% la disponibilité et que l'utilisation de l'approche multi-turbine la diminue encore de 4% environ. La disponibilité passe de 65% à 54% puis 50%; cette diminution n'est pas problématique, car la surveillance peut encore se faire tous les jours. Mais on peut voir que la moitié de la base de données est inutile.

Tableau 6.5 Comparaison de la disponibilité entre les méthodes utilisant les liens de causalité dynamique (pour les 1315008 échantillons dans la base)

	Nom de l'indicateur	Nombre d'échantillons valides	Disponibilité [%]
<b>Toutes les phases en fonctionnement</b>	Mono T°Roul. Avant Phy. Dyn non-filtré	850537	64.68%
	Mono T°Roul. Arrière Phy. Dyn non-filtré	858573	65.29%
	Mono T°Enroul. Stator Phy. Dyn non-filtré	862722	65.61%
<b>Régime permanent Seulement</b>	Mono T°Roul. Avant Phy. Dyn filtré	707688	53.82%
	Mono T°Roul. Arrière Phy. Dyn filtré	709335	53.94%
	Mono T°Enroul. Stator Phy. Dyn filtré	712726	54.20%
	Multi T°Roul. Avant Phy. Dyn filtré	650551	49.47%
	Multi T°Roul. Arrière Phy. Dyn filtré	652007	49.58%
	Multi T°Enroul. Stator Phy. Dyn filtré	655344	49.84%

En conclusion sur ces méthodes utilisant les liens de causalité dynamique, on peut dire que le modèle dynamique proposé ne peut pas être aussi facilement généralisé à d'autres composants. En

effet, il a été conçu pour un roulement d'arbre lent (Zhang, 2014) qui a une dynamique différente de l'arbre rapide de la génératrice. Le modèle utilise la température de la partie de la génératrice surveillée (roulement ou enroulement) une période d'échantillonnage avant la température prédite (cf 5.4.1.1.1 avec l'équation (5.1)). Il est possible que le choix de ce retard d'une période d'échantillonnage puisse influencer les résultats de la méthode et d'autres retards devraient être testés. De même, un retard pourrait être introduit dans les autres variables d'entrée. Il est aussi tout à fait possible que la dynamique du système que l'on cherche à surveiller soit trop rapide pour les données 10 minutes du SCADA.

Cette méthode utilisant les liens de causalité dynamique est donc trop limitée pour la surveillance complète de la génératrice. Une autre méthode est proposée en supprimant la composante dynamique, les résultats présentés précédemment vont pouvoir servir de référence.

### 6.3.2. Résidus issus de modèles statiques

#### 6.3.2.1. Approche mono-turbine

Les modèles utilisant les liens de causalité dynamique cherchent à prendre en compte toutes les origines de l'échauffement dans les parties de la génératrice au travers de leurs variables d'entrées tirées du SCADA. Néanmoins, l'inertie thermique qui était représentée par la température à  $t - 1$  dans le modèle précédent semble être une source d'erreur importante. D'autres modèles sont proposés dans la littérature qui ne prennent pas en compte la dynamique. Cette partie est consacrée à l'évaluation des performances de la méthode statique décrite par cette équation (cf partie 5.4.1.1.3 équation (5.5)) :

$$T_j^i(t) = a_j^i \cdot P^i(t) + b_j^i \cdot V^i(t) + c_j^i \cdot T_{nacelle}^i(t) + d_j^i \quad (6.9)$$

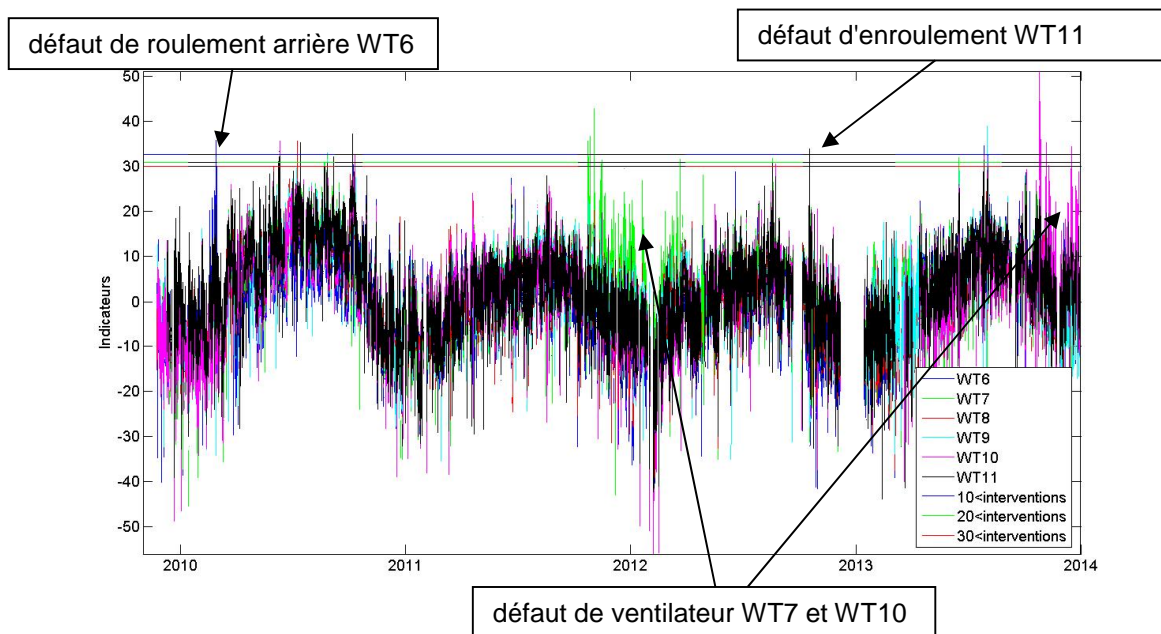


Figure 6.24 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche mono-turbine

La Figure 6.24 montre l'évolution de l'indicateur obtenu à partir du modèle statique utilisant les liens de causalité utilisé pour la température du roulement arrière en fonction du temps, pour les 6 turbines. Contrairement au modèle utilisant les liens de causalité dynamique, une forte saisonnalité peut être

observée. Comme pour les variables brutes, il s'agit de l'effet des variations de l'environnement comme la température extérieure. L'effet le plus visible est celui des saisons, mais ce n'est pas le seul phénomène. Bien que le défaut de roulement arrière soit visible sur cette courbe au début de l'année 2010, le seuil de détection automatique générant moins de 30 interventions inutiles ne détecte ce défaut qu'au dernier moment, à cause de ces fluctuations saisonnières. En revanche, les défauts apparaissant dans les périodes où la température est élevée, comme en été, pourront être eux plus facilement détectés. C'est le cas du défaut de roulement arrière de la turbine WT9 apparaissant fin 2011.

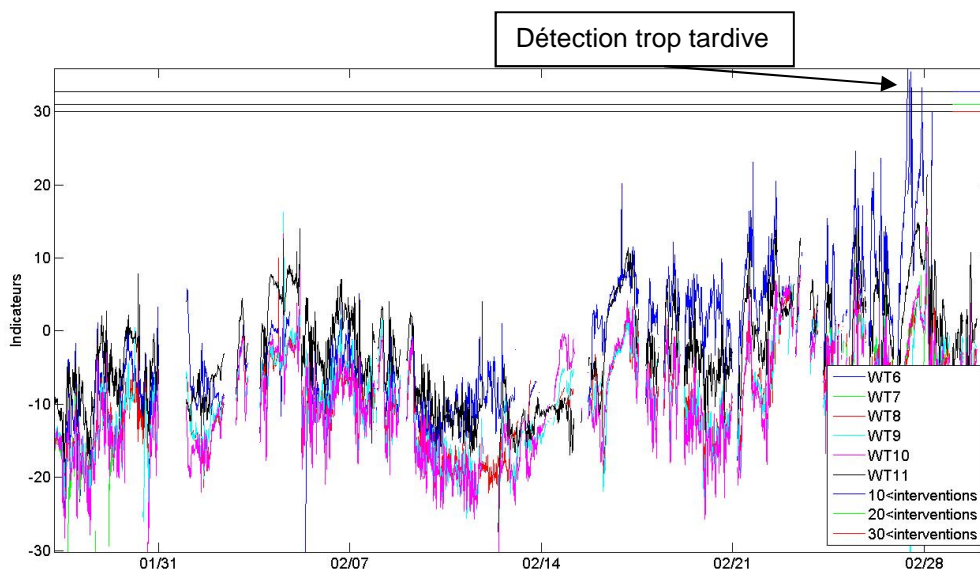


Figure 6.25 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche mono-turbine, pendant le défaut de roulement arrière WT6

Ces variations de l'environnement sont difficiles à identifier de manière exhaustive, mais on peut citer, par exemple, les variations de pression ou d'humidité. Ces variables ne sont pas prises en compte dans le modèle utilisant les liens de causalité statique, car elles ne sont pas disponibles dans le SCADA. La Figure 6.25 est un agrandissement de la zone de défaut du roulement arrière de la turbine WT6. En affichant les autres turbines sur le même graphique, on se rend compte qu'il y a d'autres influences extérieures que les saisons qui touchent toutes les turbines en même temps.

Les figures Figure 6.26 et Figure 6.27 montrent les courbes ROC obtenues pour chacun des défauts et pour différentes tailles de fenêtre pour une approche mono-turbine utilisant les résidus d'une méthode utilisant les liens de causalité statique. La Figure 6.28 montre elle, la persistance obtenue. Ces figures montrent que l'utilisation d'un modèle utilisant les liens de causalité statique permet la détection d'un défaut critique avec un nombre d'interventions inutiles acceptable, inférieur à 30. Les indicateurs fournis par ce type de modèle permettent aussi de détecter les défauts de ventilateur de la turbine WT7, WT9 et WT10, avec une persistance suffisamment élevée pour indiquer qu'il ne s'agit pas que de simples pics isolés.

L'utilisation d'une moyenne glissante permet d'améliorer l'avance à la détection du défaut critique de roulement arrière WT9 lorsque la fenêtre est fixée à 144 échantillons.

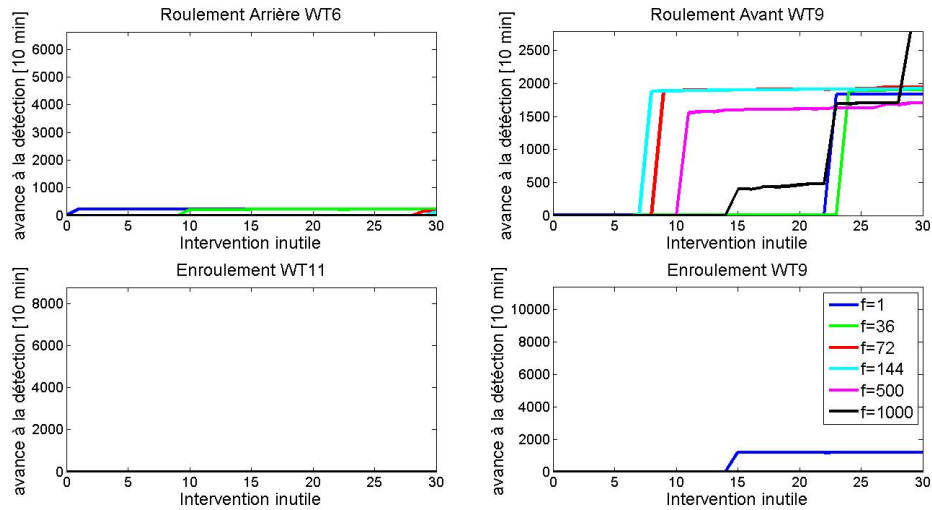


Figure 6.26 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine

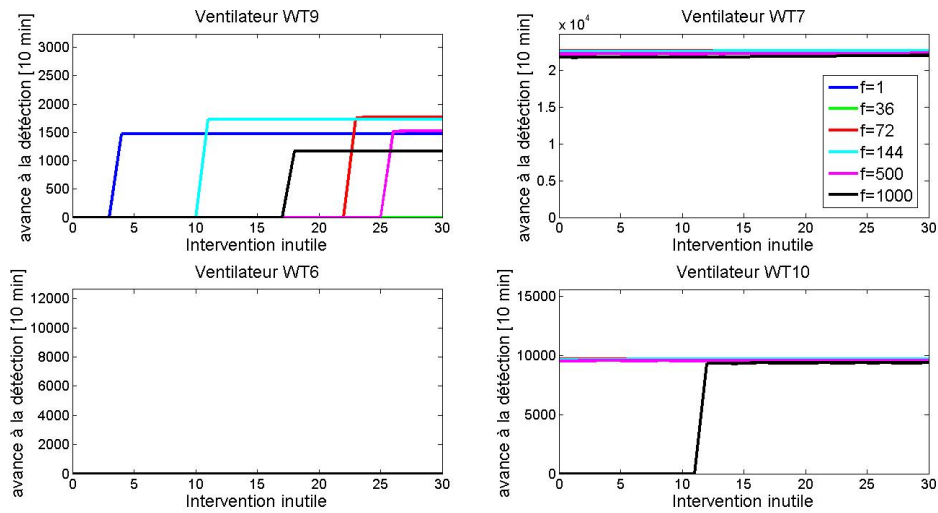


Figure 6.27 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine

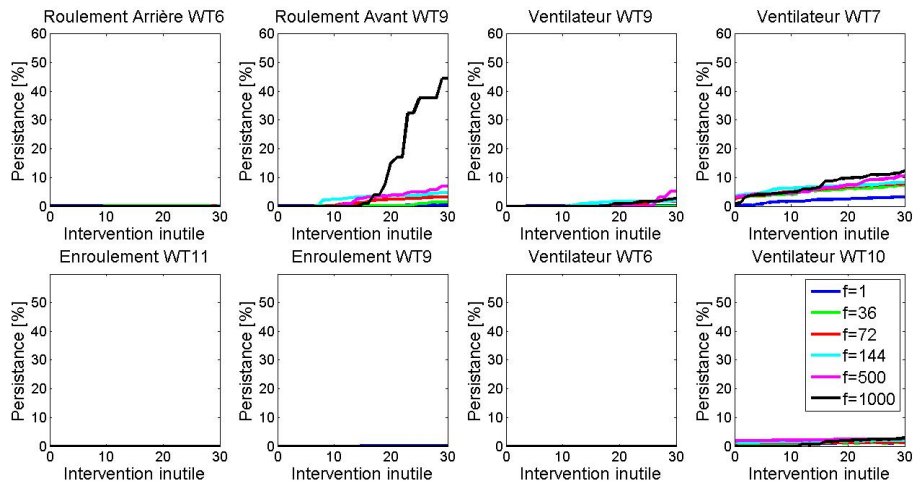


Figure 6.28 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité statique, approche mono-turbine

### 6.3.2.2. Approche multi-turbine

La Figure 6.29 montre l'évolution de l'indicateur obtenue avec une approche multi-turbine en utilisant les résidus d'une méthode mono-turbine utilisant les liens de causalité statique appliquée sur le roulement arrière. Cette figure permet de constater que l'approche multi-turbine permet de s'affranchir d'une partie de l'influence des variations de l'environnement. Les variations annuelles ont disparu, ce qui permet de fixer un seuil sur le défaut de roulement arrière WT6 début 2010. Il est aussi intéressant de voir que le défaut d'enroulement début 2013 est clairement visible maintenant.

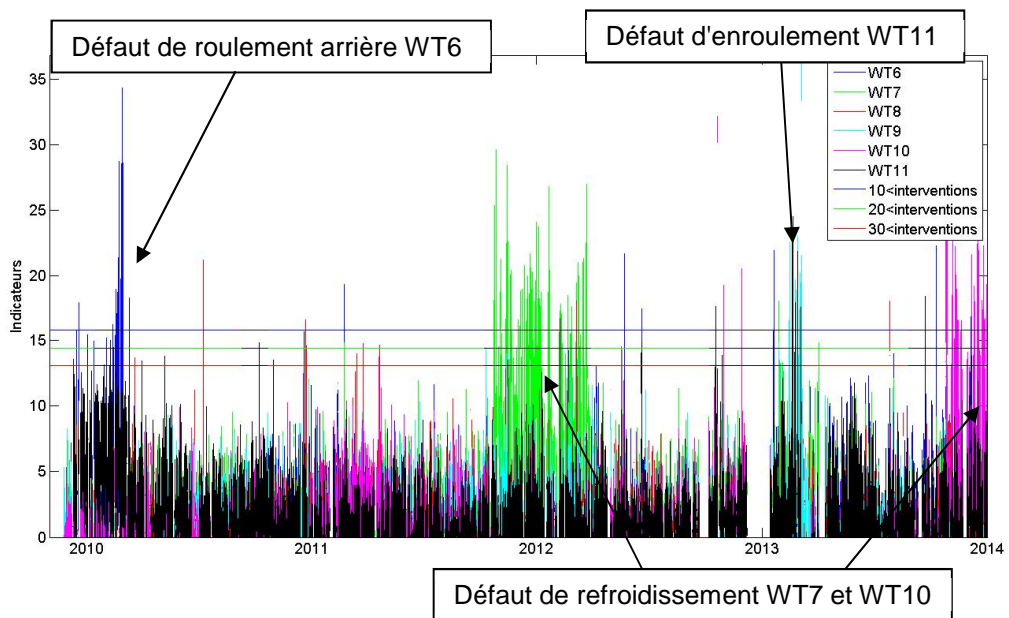


Figure 6.29 Évolution de l'indicateur utilisant les liens de causalité statique du roulement arrière de chaque éolienne, approche multi-turbine

La Figure 6.30 et la Figure 6.31 montrent les courbes ROC obtenues pour chaque défaut en utilisant une approche multi-turbine avec des résidus d'une méthode utilisant les liens de causalité statique pour



différentes valeurs de fenêtre. Ces figures montrent l'efficacité de l'approche combinée entre utilisant les liens de causalité statique et comparaison entre éoliennes. Elle permet de détecter les 4 défauts critiques de la base de données avec un très faible nombre d'interventions inutiles. Les défauts de ventilateurs sont aussi détectés grâce à cette méthode.

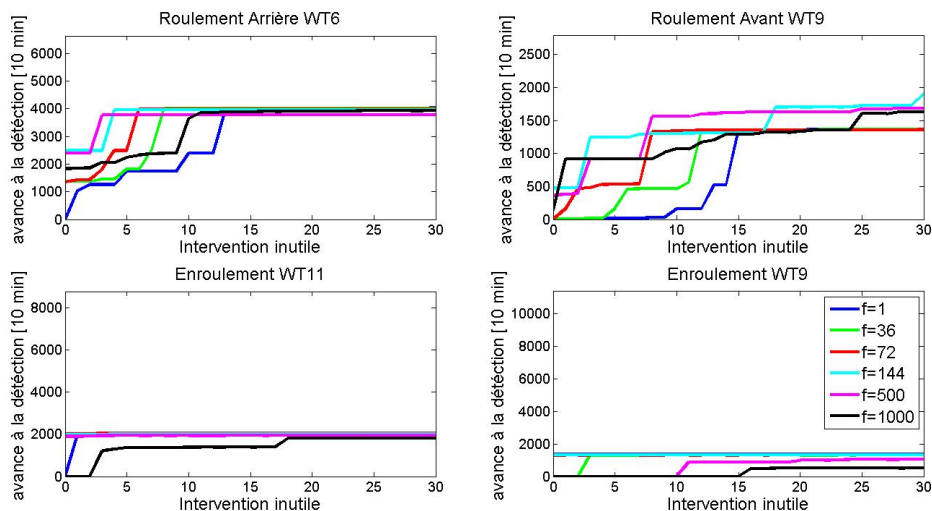


Figure 6.30 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine

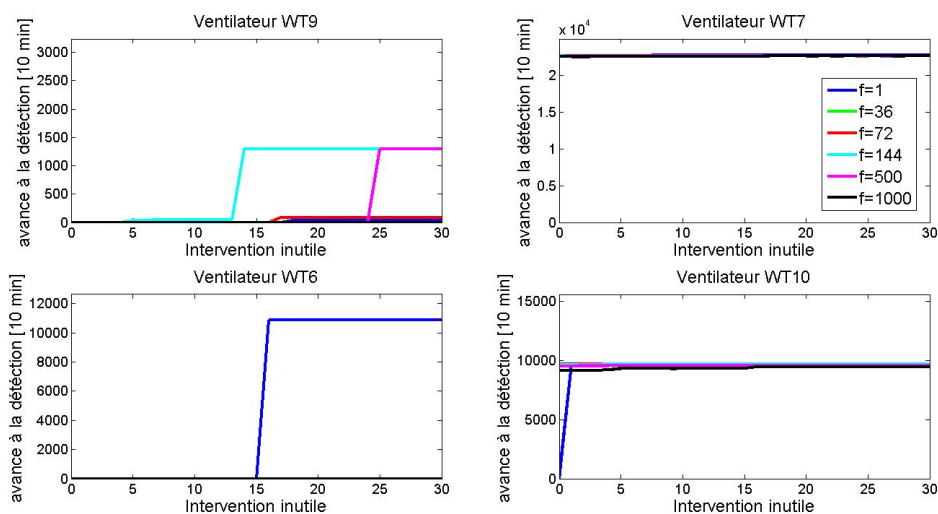


Figure 6.31 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine

La Figure 6.32 montre que la persistance est particulièrement élevée avec cette approche. Il est plus facile d'avoir confiance dans les détections et donc de programmer une intervention sur site avec les informations provenant ces indicateurs.

Ces figures permettent aussi de constater que la comparaison sur une plage de temps est plus pertinente qu'une comparaison instantanée, même pour des résidus. En effet, les résidus fournis par une approche mono-turbine utilisant les liens de causalité statique ne sont pas complètement

indépendants des variations de fonctionnement et des décalages temporels entre les turbines peuvent encore exister. L'utilisation d'une plage de 144 échantillons pour la comparaison permet d'avoir les meilleurs résultats sur les défauts critiques en gardant un niveau de persistance très élevé.

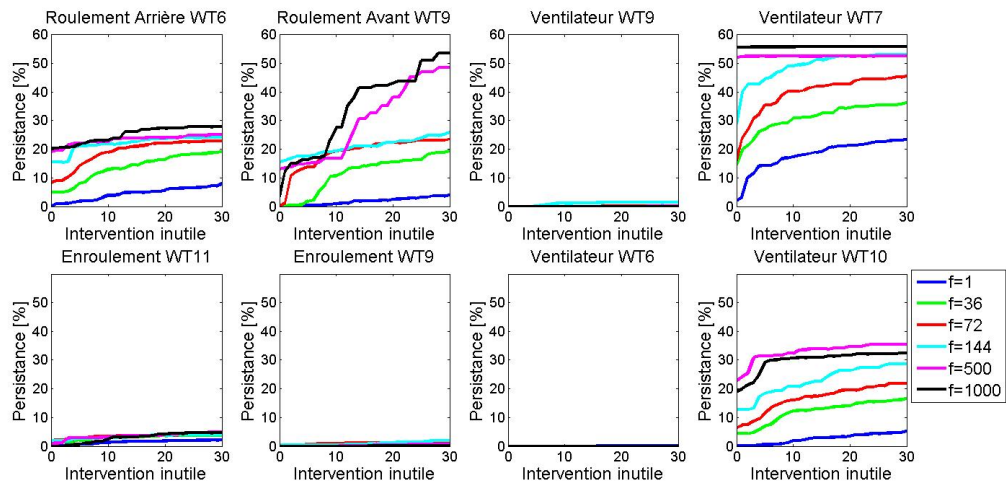


Figure 6.32 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité statique, approche multi-turbine

### 6.3.3. Comparaison globale des méthodes utilisant les liens de causalité mono-turbine et multi-turbine

Les Figure 6.33 et Figure 6.34 montrent les courbes ROC des approches mono-turbines et multi-turbines utilisant des résidus provenant de méthodes utilisant les liens de causalité statique et dynamiques utilisées uniquement sur les phases en fonctionnement. Seules les tailles de fenêtre  $f$  ayant donné les meilleurs résultats en termes d'avance à la détection, de nombre d'interventions inutiles et de persistance sont représentées.

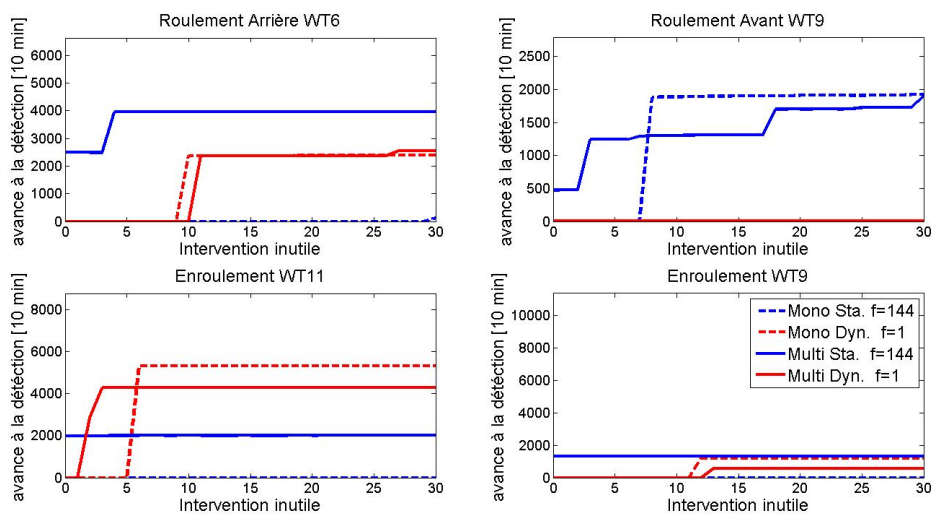


Figure 6.33 Avance à la détection vs nombre d'interventions inutiles pour les défauts critiques, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine



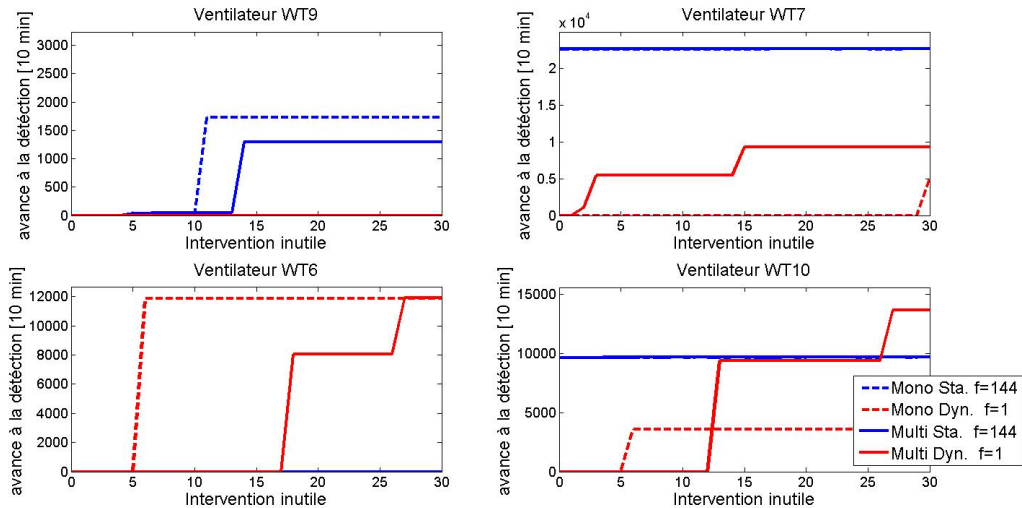


Figure 6.34 Avance à la détection vs nombre d'interventions inutiles pour les défauts ventilateurs, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine

Pour conclure sur les méthodes utilisant les liens de causalité, on peut dire que la prise en compte de la dynamique dans les modèles reste difficile. Elle semble diminuer les performances des indicateurs. En effet, lorsque l'on compare les méthodes mono-turbines dynamiques et statiques sur les Figure 6.33 et Figure 6.34, on observe que les modèles statiques permettent une réelle surveillance des éoliennes du parc. Les défauts critiques et une partie des défauts de ventilateurs sont détectés par les modèles statiques alors que les modèles dynamiques ne voient que des pics isolés pour un défaut critique. La Figure 6.35 montrant la persistance de ces approches confirme que l'utilisation d'une approche multi-turbine permet d'augmenter grandement la persistance, ce qui permet aux opérateurs d'avoir plus confiance en cet indicateur. La prise en compte de la dynamique est donc à améliorer si l'on veut la généraliser sur les composants de la génératrice.

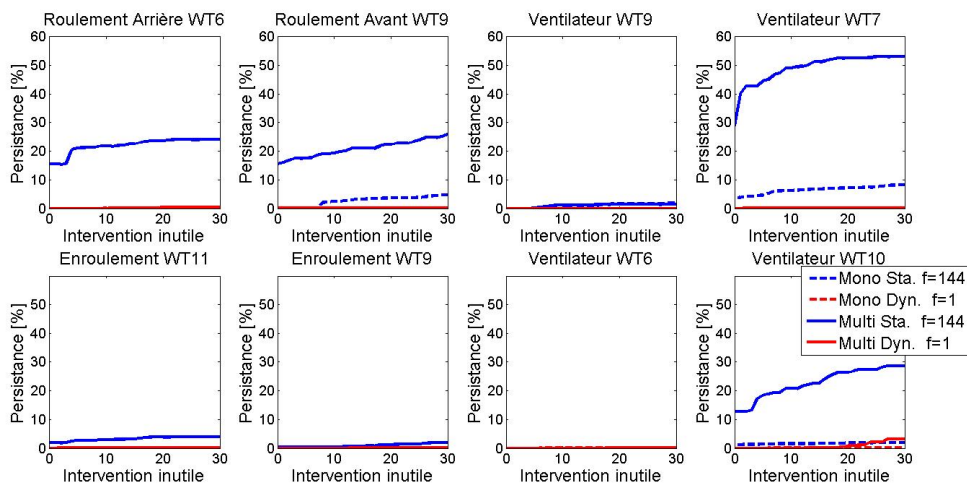


Figure 6.35 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle utilisant les liens de causalité, comparaison mono-turbine vs multi-turbine

## 6.4. Indicateurs utilisant des résidus issus de modèles utilisant la similarité

Cette partie se concentre sur la présentation de l'évaluation des performances des nouvelles méthodes présentées dans le chapitre 5, utilisant la similarité. Les performances obtenues lorsque l'écart entre la valeur estimée et la valeur réelle est utilisé comme indicateur, sont présentées. Ensuite, les performances obtenues en comparant les résidus des modèles entre les éoliennes sont présentées.

### 6.4.1. Approche mono-turbine

La Figure 6.37 représente l'indicateur obtenu en utilisant un modèle de similarité sur le roulement arrière par rapport au roulement avant. Il suit l'équation (cf partie 5.4.1.2 équation (5.9)) :

$$T_j^i(t) = a_j^i \cdot T_{ref_j}^i(t) + b_j^i \quad (6.10)$$

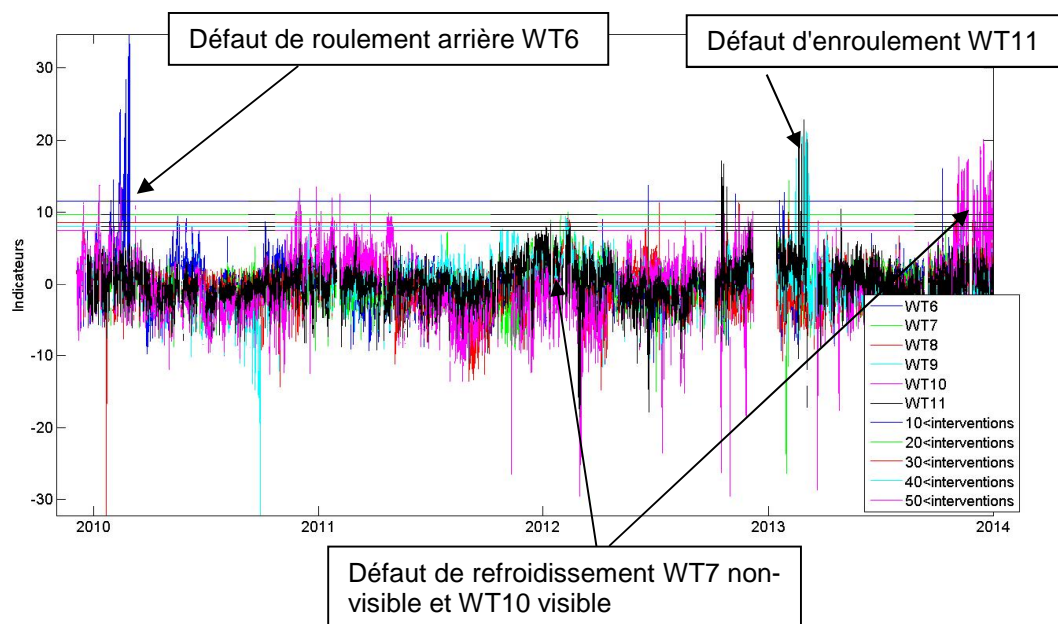


Figure 6.36 Évolution de l'indicateur similarité du roulement arrière de chaque éolienne, approche mono-turbine

La Figure 6.36 permet de constater, dans un premier temps, que les variations saisonnières sont atténuées sur cet indicateur. Les défauts critiques sont visibles et un seuil automatique peut être fixé pour les détecter, début 2010 avec le défaut de roulement arrière, début 2013 avec le défaut d'enroulement WT11 et le défaut de ventilateur WT10 à la fin 2014.

La Figure 6.37 et la Figure 6.38 montrent les courbes ROC obtenues pour chaque défaut avec une approche mono-turbine utilisant une méthode utilisant la similarité pour différentes tailles de fenêtre  $f$ . La Figure 6.37 montre que cette méthode obtient d'excellents résultats sur les défauts critiques. Elle permet de détecter ces défauts avec un nombre d'interventions inutiles très faible voir nul.

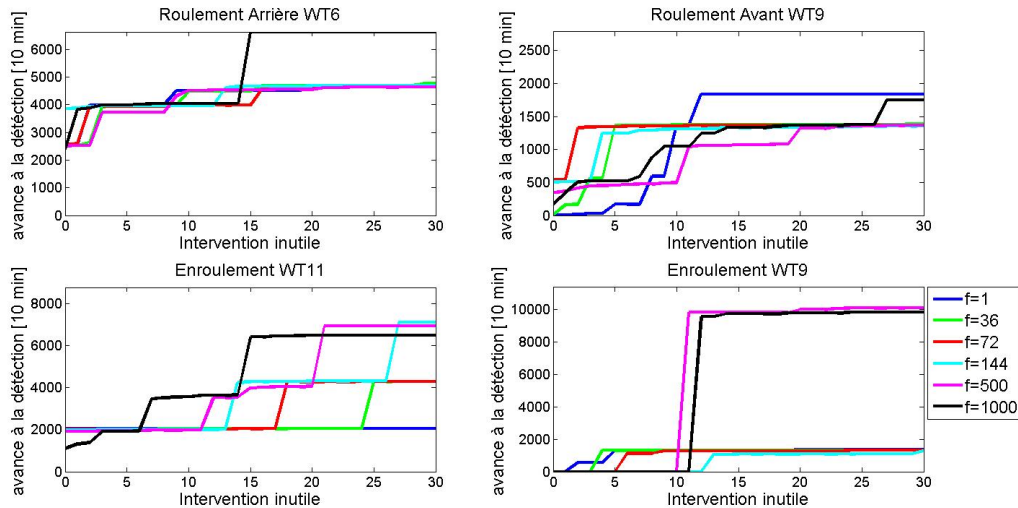


Figure 6.37 Avance à la détection vs nombre d'interventions inutiles pour les défauts critiques, en utilisant un modèle de similarité, approche mono-turbine

La Figure 6.38 permet de constater que les défauts de ventilateur peuvent être détectés sans intervention inutile. L'efficacité des méthodes utilisant la similarité peut être fortement diminuée si le défaut impacte la variable de référence et la variable prédite (cf partie 3.1.3.5). Le défaut de ventilateur WT7 n'a pas eu exactement la même influence sur toutes les parties de la génératrice, il est donc possible de trouver un couple de variables (ici température stator par rapport à température roulement arrière) impacté suffisamment différemment par le défaut. Mais cela peut ne pas toujours être le cas, ce qui perturbe une détection précoce du défaut.

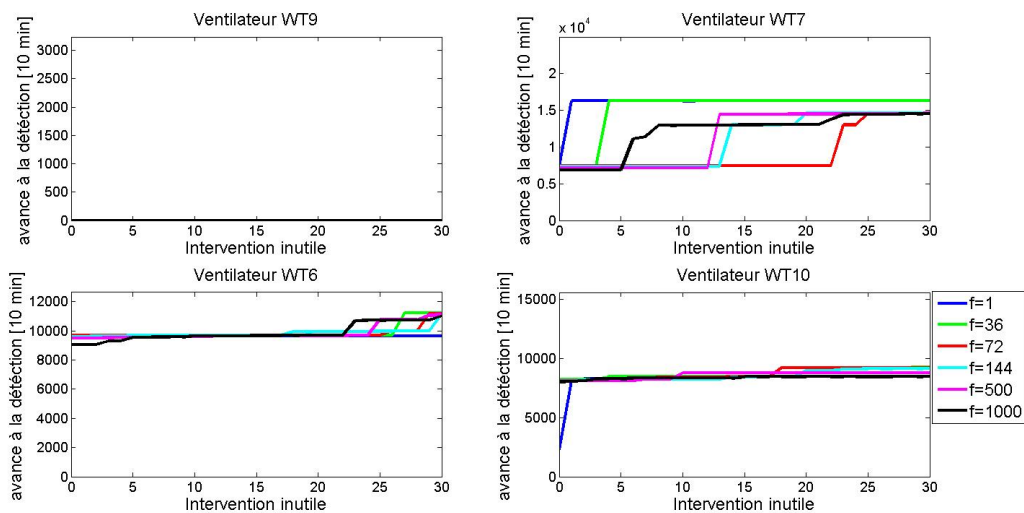


Figure 6.38 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, approche mono-turbine

Il est important de noter que le défaut de ventilateur WT7 apparaissant normalement fin 2011 est ici très peu visible dans la Figure 6.36. Cela est dû à la nature de ce défaut. En effet, le défaut de ventilateur a eu impact sur les trois parties de la génératrice. La méthode par similarité utilise une autre partie de la génératrice comme référence pour créer la température théorique. Comme toutes les températures

ont augmenté, la différence entre la température théorique et réelle est restée proche de zéro. Il s'agit ici d'une des faiblesses de la méthode puisque si un défaut a un impact similaire sur la température de référence et sur la température que l'on cherche à surveiller alors le défaut ne sera pas ou peu visible.

La Figure 6.39 montre que la persistance de l'indicateur est assez élevée, ce qui indique qu'il ne s'agit pas de simples pics isolés. On peut voir, grâce à cette figure, que tous les défauts sont détectés, à l'exception du défaut de ventilateur WT9.

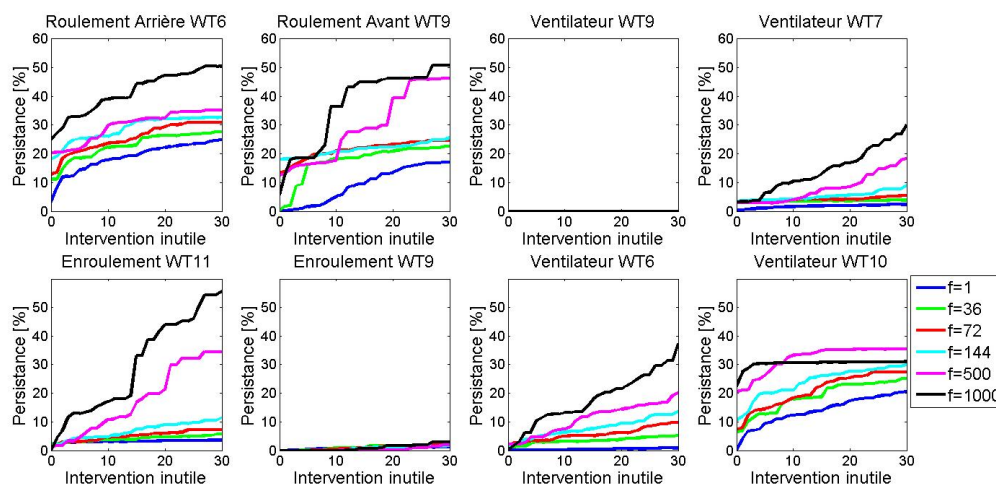


Figure 6.39 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, approche mono-turbine

L'utilisation d'une moyenne mobile permet d'améliorer grandement les performances. On observe qu'en choisissant une fenêtre de 144 échantillons, on peut détecter tous les défauts et fournir un seuil où le défaut de roulement arrière est détecté un mois en avance sans aucune intervention inutile. En revanche, cette taille de fenêtre diminue fortement les résultats sur la détection du défaut d'enroulement WT9.

On peut voir ici que le choix de la taille de la fenêtre  $f$  peut avoir un impact important sur la détection des défauts de cette base de données en particulier. La meilleure avance à la détection s'obtient en choisissant entre  $f = 72$  et  $f = 144$ . Le choix de  $f$  a un impact en fonction du défaut,  $f = 144$  permet d'avoir de meilleurs résultats pour le roulement arrière WT6 et  $f = 72$  permet d'avoir de meilleurs résultats pour le roulement avant WT9. Il est difficile de conclure sur l'impact des fenêtres sur la détection de certains types de défaut, et nous choisirons par la suite de conserver les 2 tailles de fenêtre pour les comparaisons. On peut cependant conclure que l'utilisation d'une fenêtre mobile améliore les résultats sur cette base de données à partir du moment où celle-ci reste inférieure à 500 échantillons. Ensuite, le retard à la détection est trop important par rapport à la diminution du nombre d'interventions inutiles.

Il est intéressant de remarquer aussi que, si on n'applique pas de moyenne mobile, il est possible de détecter le défaut de roulement avant plus tôt, mais avec un nombre d'interventions inutiles plus important. La Figure 6.40 montre l'évolution des indicateurs mono-turbine utilisant la similarité avec plusieurs tailles de fenêtre  $f$  sur la période de défaut du roulement avant de la turbine WT9. On peut voir que le premier signe du défaut est un pic très court aux alentours de 12h le 11 septembre. Ce pic ne dure qu'une quarantaine de minutes et disparaît si on utilise une moyenne mobile.

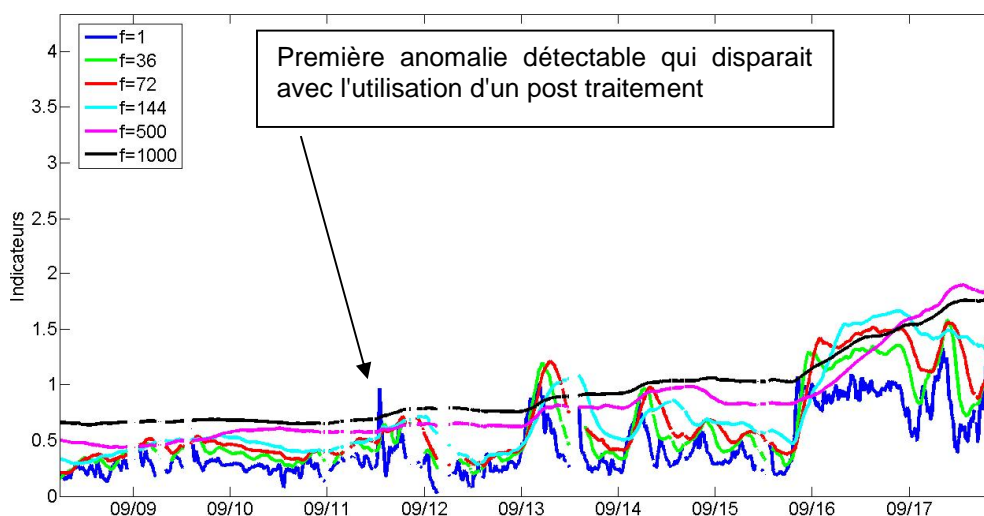


Figure 6.40 Évolution de l'indicateur similarité avec plusieurs tailles de fenêtre  $f$  sur la période de défaut du roulement avant WT9, approche mono-turbine

La détection au plus tôt est très importante, mais ce pic de première détection est isolé du reste. Il est difficile pour la personne qui effectue une surveillance d'y voir un défaut. Ce genre de pic ne permet pas de lancer une intervention et le temps gagné en détection n'aurait sûrement pas pu être mis à profit.

#### 6.4.2. Approche multi-turbine

La méthode basée sur la similarité produit un nouveau type d'indicateur mono-turbine. Cette partie est consacrée à l'utilisation de ces indicateurs avec une approche multi-turbine.

La Figure 6.41 montre les performances en détection de l'approche multi-turbine basée sur les résidus des méthodes mono-turbine utilisant la similarité sur les défauts critiques. On peut voir que cette approche est performante et permet la détection des défauts critiques. De plus, l'utilisation d'une plage d'échantillons pour la comparaison est meilleure que la comparaison instantanée puisqu'elle permet d'augmenter l'avance à la détection pour un même nombre d'interventions inutiles et cela pour 3 défauts critiques. L'utilisation d'une plage d'échantillons pour la comparaison est équivalente pour les autres défauts si la plage ne dépasse pas 144 échantillons. Au-delà de 144 échantillons, les performances diminuent pour le défaut de roulement arrière. De plus, l'approche multi-turbine fournit avec une fenêtre  $f = 72$ , une configuration où la détection du défaut de roulement arrière est possible sans fausse alarme, 2000 échantillons en avant en avance.

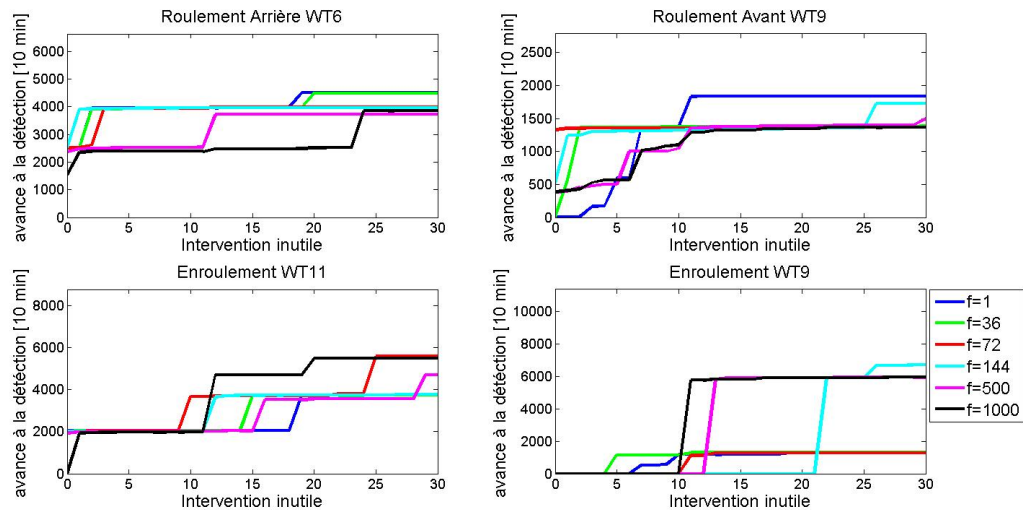


Figure 6.41 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle similarité, approche multi-turbine

La Figure 6.42 montre les courbes ROC sur les défauts de ventilateurs. Cette figure montre que cette approche peut détecter les défauts de ventilateur à l'exception de celui de la turbine WT9. On peut voir, de plus, avec les courbes de persistance présentées dans la Figure 6.43, que les indicateurs restent au-dessus des seuils de manière durable. La persistance confirme que le défaut d'enroulement WT9 n'est pas détecté.

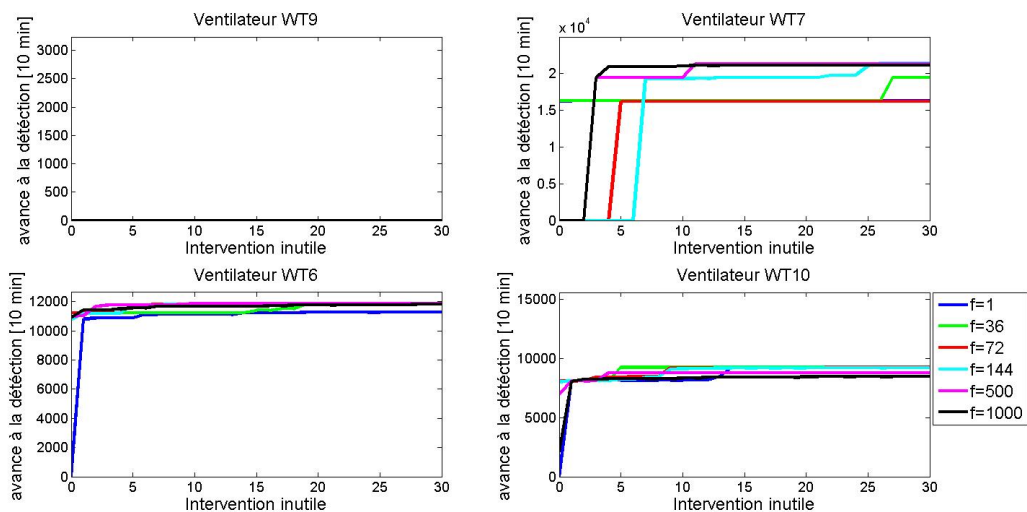


Figure 6.42 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, approche multi-turbine



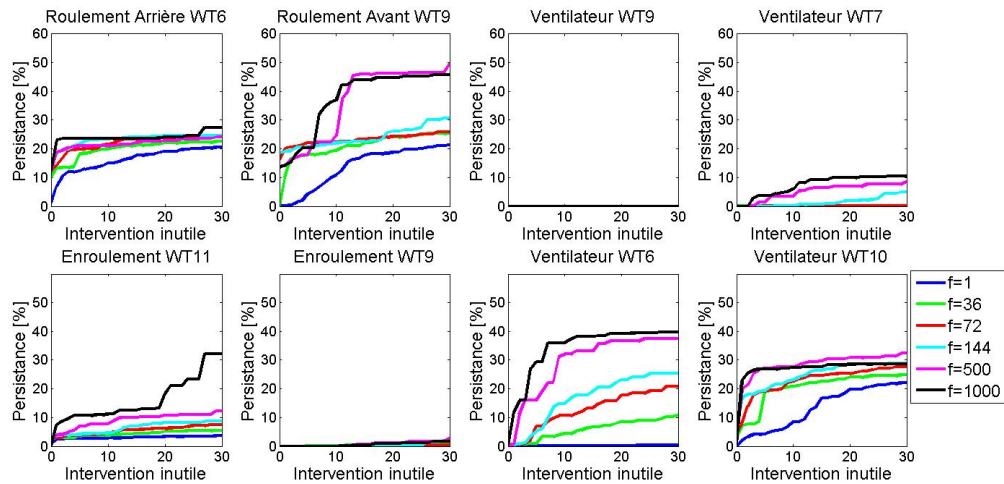


Figure 6.43 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, approche multi-turbine

### 6.4.3. Comparaison globale des méthodes utilisant des résidus issus de modèles utilisant la similarité

Les Figure 6.44, Figure 6.45 et Figure 6.46 permettent de comparer les approches mono-turbines et multi-turbines basées sur la similarité. Les Figure 6.44, Figure 6.45 montrent les courbes ROC pour chaque défaut pour les approches mono-turbines et multi-turbines. La Figure 6.46 représente les persistances entre ces approches. Pour des raisons de clarté, seules les 2 meilleures configurations,  $f = 72$  et  $f = 144$ , sont présentées.

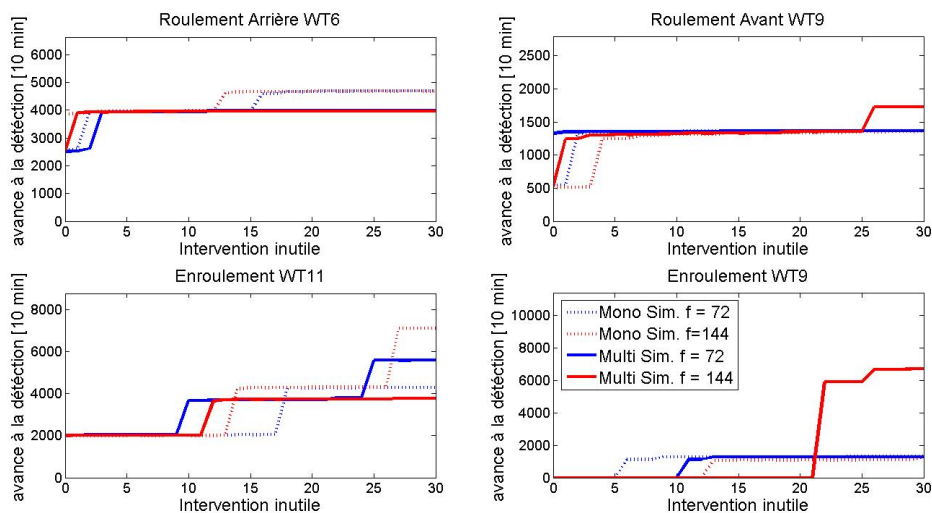


Figure 6.44 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine

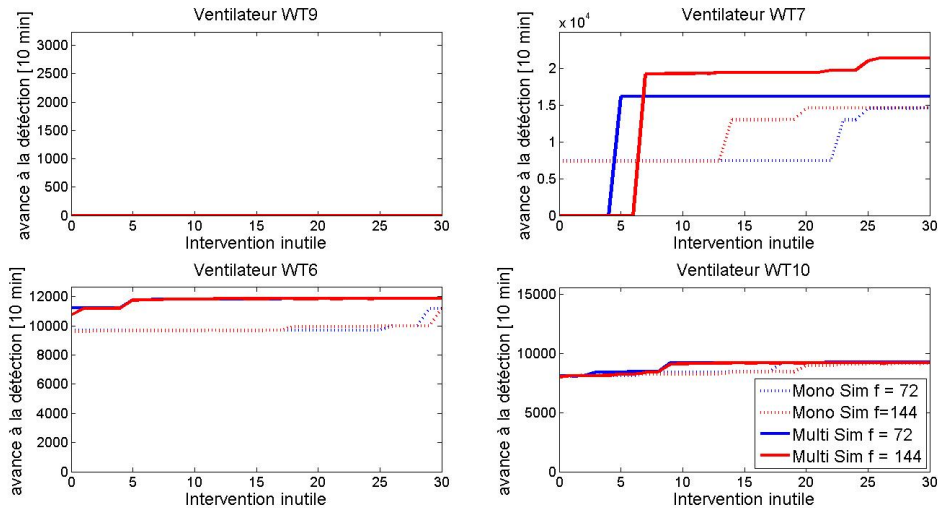


Figure 6.45 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine

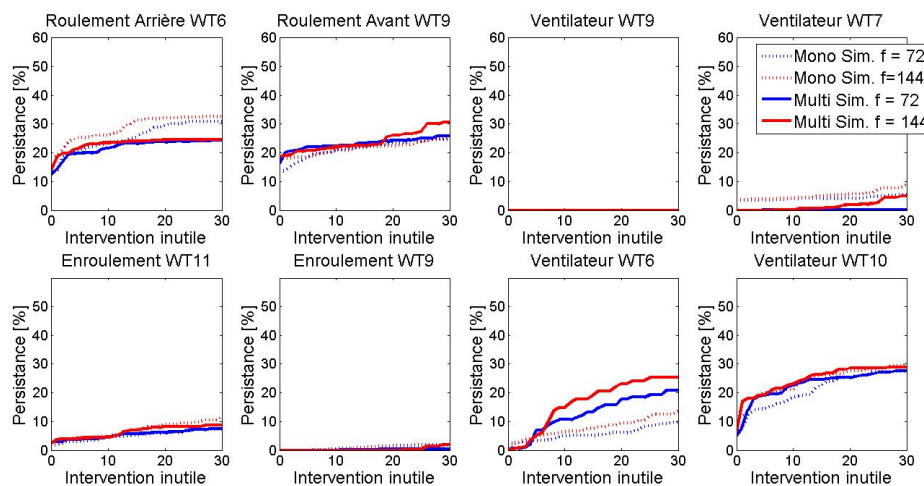


Figure 6.46 Persistence vs nombre d'interventions inutiles des défauts, en utilisant un modèle similarité, comparaison mono-turbine vs multi-turbine

L'approche similarité mono-turbine montre des résultats qui varient en fonction de la taille de la fenêtre. L'approche multi-turbine permet d'améliorer les résultats de 2 défauts critiques et diminue les résultats des 2 autres, le défaut de roulement arrière WT6 et le défaut d'enroulement WT9.

La méthode mono-turbine permet de détecter le défaut de roulement arrière WT6, avec un nombre d'interventions inutiles égales à zéro, une semaine avant à l'approche multi-turbine. Pour le défaut d'enroulement WT9, c'est le nombre d'interventions inutiles qui est augmenté. L'approche multi-turbine est influencée par le nombre de turbines sur le calcul de la référence du parc qui est ici la médiane des résidus similarité entre les turbines. En effet, sur ces deux exemples, le nombre de turbines est diminué à cause d'arrêts, ce qui fait changer la valeur de la médiane de manière importante.

Durant le premier défaut, la turbine WT7 était à l'arrêt, par conséquent la médiane ne se calcule plus que sur 5 turbines au lieu de 6. De plus, cette turbine a un résidu moyen plus faible que les autres



habituellement. L'absence de cette turbine a donc fait monter artificiellement la référence du parc. La détection de la surchauffe pendant les premiers instants du défaut est donc plus difficile et il faut rabaisser le seuil, ce qui génère des interventions inutiles.

Dans le deuxième cas, une situation similaire apparaît puisqu'il y a deux défauts en même temps sur le parc. En effet, le défaut d'enroulement de la turbine WT9 apparaît alors que le défaut de ventilateur de la turbine W10 est déjà en cours. Les défauts de ventilateurs sont bien plus marqués que les défauts critiques, ce qui a élevé la valeur de la médiane.

On peut apercevoir ici une des limites de la comparaison entre les turbines proposée. Le calcul de la médiane est trop sensible au nombre de turbines lorsque celles-ci sont peu nombreuses dans le parc. De plus, si deux défauts apparaissent au même moment, même si cela est très rare, la comparaison peut être impactée.

## 6.5. Évaluation globale des différents indicateurs

Cette partie propose une comparaison globale des indicateurs présentés précédemment. La fenêtre d'apprentissage est fixée à 2000 échantillons et les meilleures configurations pour la fenêtre de moyenne mobile de chaque méthode sont sélectionnées pour la comparaison.

### 6.5.1. Analyse avec un seuil variable

La Figure 6.47 et la Figure 6.48 montrent l'instant de détection de chaque méthode pour les défauts en fonction du nombre d'interventions inutiles, grâce aux courbes ROC. Rappelons qu'en dessous d'une avance à la détection de 1000 échantillons, il n'est pas possible de mettre en place une stratégie de maintenance pour remédier au problème avant la panne.

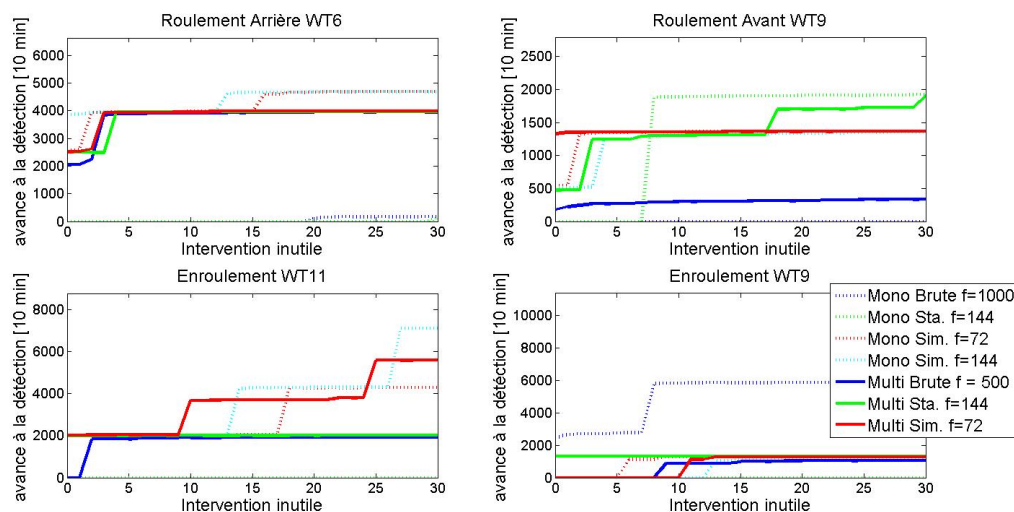


Figure 6.47 Avance à la détection vs nombre d'interventions inutiles des défauts critiques, comparaison globale

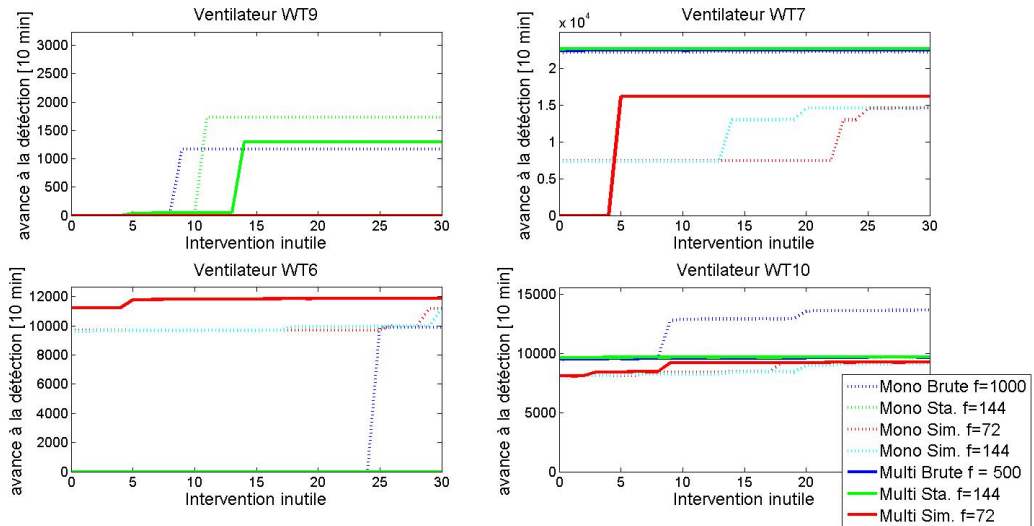


Figure 6.48 Avance à la détection vs nombre d'interventions inutiles des défauts ventilateurs, comparaison globale

Sur ces figures, on peut voir que les résultats des méthodes sont différents en fonction du défaut. Certaines méthodes semblent meilleures pour certains types de défauts. On voit par exemple que les défauts de roulement sont très bien détectés par les méthodes par similarité alors que le défaut d'enroulement WT9 est mieux détecté par une méthode utilisant les liens de causalité. En ce qui concerne les défauts de ventilateur, les méthodes utilisant les liens de causalité montrent de meilleurs résultats sauf pour le défaut de ventilateur de la turbine WT6. Ce dernier défaut n'est détecté que par les méthodes utilisant la similarité. Lorsque l'on regarde les défauts d'enroulement, c'est l'approche utilisant les liens de causalité multi-turbine qui permet de mieux les détecter. Au regard de ces résultats, on peut conclure que ces méthodes sont complémentaires, chacune obtenant les meilleures performances sur certains types de défauts et qu'il serait pertinent de mettre en place et d'analyser les indicateurs en parallèle.

L'objectif de cette étude est surtout de conclure sur l'intérêt de l'approche multi-turbine.

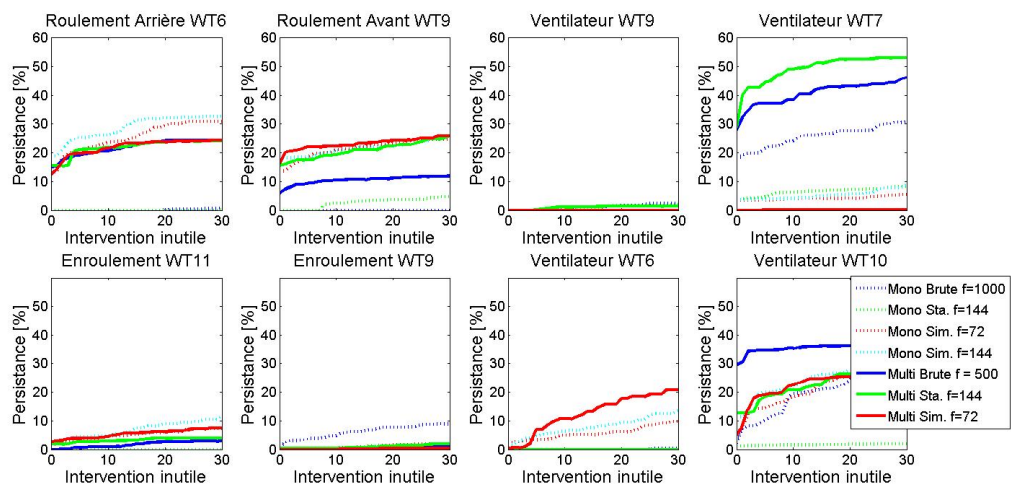


Figure 6.49 Persistence vs nombre d'interventions inutiles des défauts, comparaison globale

---

La Figure 6.49 montre la persistance pour chacune des méthodes et chacun des défauts. On peut voir que la persistance est améliorée avec l'utilisation d'une approche multi-turbine pour la plupart des défauts. Les défauts de ventilateur en particulier voient leur persistance améliorée. En revanche, pour le défaut de roulement arrière, la persistance est diminuée, mais reste suffisamment élevée pour que l'opérateur puisse prendre une décision. Les deux défauts d'enroulements sont difficiles à détecter pour toutes les méthodes et leurs persistances sont très basses, ce qui pourrait amener à douter de l'indicateur.

Dans la partie précédente, il a été montré que l'utilisation d'une plage de temps pour la comparaison permet d'avoir de meilleurs résultats. Ici, la Figure 6.47, la Figure 6.48 et Figure 6.49 montrent que l'utilisation de résidus permet d'améliorer les résultats par rapport à la comparaison de variables brutes, pour la méthode multi-turbine. En effet, la méthode multi-turbine utilisant les variables brutes permet certes la détection des défauts critiques, mais offre des performances moins élevées en termes d'interventions inutiles, d'avance à la détection et de persistance qu'une méthode multi-turbine utilisant des résidus issus de modèles utilisant les liens de causalité ou utilisant la similarité. L'utilisation de variables brutes pour l'approche multi-turbine est meilleure en persistance uniquement pour le défaut de ventilateur WT10. Pour les autres défauts, elle est équivalente ou moins performante.

Les méthodes multi-turbines ne sont pas toujours les meilleures sur la détection de défauts, en particulier sur le défaut de roulement arrière de la turbine WT6 et sur le défaut d'enroulement de la turbine WT9. Comme expliqué dans la partie 6.4.3, ce manque de performances est dû à l'influence du nombre de turbines dans la comparaison, le nombre de turbines dans le parc étant faible. Ce problème devrait être moins fréquent dans des parcs de taille plus importante comme les parcs en mer.

Dans le cas où une approche multi-turbine ne serait pas applicable par manque de données d'autres turbines ou parce que le parc est composé d'un nombre trop faible d'éoliennes, on peut voir que la méthode utilisant la similarité en approche mono-turbine fournit les meilleurs résultats. En revanche, si aucun modèle mono-turbine ne peut être développé, une comparaison entre les turbines des variables brutes fournit des résultats pertinents en termes de surveillance.

Tableau 6.6 Comparaison globale de la disponibilité des méthodes (pour les 1315008 échantillons dans la base)

Approches	Nom de l'indicateur	Nombre d'échantillons valides	Disponibilité [%]
Mono-turbine	Mono T°Roul. Avant Brute	789231	60.02%
	Mono T°Roul. Avant Phy.Sta.	790461	60.11%
	Mono T°Roul. Avant Sim. $f=72$	795656	60.51%
	Mono T°Roul. Avant Sim. $f=144$	789574	60.04%
	Mono T°Roul. Arrière Brute	791605	60.20%
	Mono T°Roul. Arrière Phy.Sta.	791252	60.17%
	Mono T°Roul. Arrière Sim. $f=72$	795656	60.51%
	Mono T°Roul. Arrière Sim. $f=144$	789574	60.04%
	Mono T°Enroul. Stator Brute	796363	60.56%
	Mono T°Enroul. Stator Phy.Sta.	798641	60.73%
	Mono T°Enroul. Stator Sim. $f=72$	797298	60.63%
	Mono T°Enroul. Stator Sim. $f=144$	791252	60.17%
Multi-turbine	Multi T°Roul. Avant Brute	725828	55.20%
	Multi T°Roul. Avant Phy.Sta.	725597	55.18%
	Multi T°Roul. Avant Sim.	734430	55.85%
	Multi T°Roul. Arrière Brute	729481	55.47%
	Multi T°Roul. Arrière Phy.Sta.	725933	55.20%
	Multi T°Roul. Arrière Sim.	734430	55.85%
	Multi T°Enroul. Stator Brute	735549	55.93%
	Multi T°Enroul. Stator Phy.Sta.	733089	55.75%
	Multi T°Enroul. Stator Sim.	735389	55.92%

Le Tableau 6.6 montre la disponibilité des méthodes pour chacune des parties de la génératrice. Seules les meilleures tailles de fenêtre  $f$  sont représentées. Ce tableau permet de constater que la disponibilité des indicateurs est très similaire entre toutes les méthodes. Cette disponibilité permet dans chaque cas de surveiller les éoliennes quotidiennement. On peut toutefois noter que l'approche multi-turbine fait baisser de 5% la disponibilité puisque la construction de la référence parc nécessite de réunir les conditions de validation sur au moins 3 éoliennes en même temps. Cette baisse est due au petit nombre d'éoliennes dans le parc. Si le parc était plus important, cette baisse serait plus faible, voire inexistante.

### 6.5.2. Comparaison des performances pour un taux d'interventions inutiles donné

La comparaison avec des seuils variables nous permet d'avoir une vue globale des performances des méthodes. Cependant, dans un contexte industriel, le seuil est fixé et n'est modifié que rarement. En se plaçant dans ce contexte de mise en œuvre, on peut définir 2 seuils de détection particulièrement représentatifs d'une utilisation par un industriel : un seuil générant 10 interventions inutiles et un seuil ne générant aucune intervention inutile.

Au début de la thèse, les discussions avec les experts au sein de VALEMO suggéraient que le nombre d'interventions inutiles ne devait pas dépasser 1 par mois sur les 4 ans, ce qui correspond à 48 sur les 4 ans de la base de données. Lorsque les premiers résultats ont été générés, les méthodes mono-turbine et multi-turbine ont montré qu'il est possible de détecter les défauts avec un nombre d'interventions inutiles bien plus faible que ce qui était attendu. Dix interventions inutiles sur les 4 ans paraissent être un bon compromis au regard des résultats obtenus avec un seuil variable.

Tableau 6.7 Comparaison de l'avance à la détection pour 2 seuils fixés pour avoir 0 et 10 interventions inutiles, en vert les défauts détectés au moins une semaine à l'avance et en rouge les défauts non détectés ou trop tard

	pour 10 interventions inutiles				pour 0 interventions inutiles			
	Roulement arrière E6	Roulement Avant E9	Enroulement E11	Enroulement E9	Roulement arrière E6	Roulement Avant E9	Enroulement E11	Enroulement E9
Mono. brute f= 1000	● 0	● 0	● 0	● 5836	● 0	● 0	● 0	● 2452
Mono. Phy. Sta. f= 144	● 0	● 1889	● 0	● 0	● 0	● 0	● 0	● 0
Mono. Sim. f=72	● 3983	● 1359	● 2031	● 1287	● 2546	● 541	● 2014	● 0
Mono. Sim. f=144	● 3958	● 1315	● 2010	● 0	● 3868	● 506	● 1992	● 0
Multi. brute f= 500	● 3916	● 297	● 1903	● 880	● 2053	● 178	● 0	● 0
Multi. Phy. Sta. f= 144	● 3957	● 1303	● 2007	● 1298	● 2492	● 476	● 1993	● 1298
Multi. Sim. f=72	● 3949	● 1358	● 3672	● 0	● 2516	● 1329	● 2030	● 0

Le Tableau 6.7 montre l'avance à la détection obtenue pour tous les défauts et pour toutes les méthodes. Deux seuils de détection sont représentés, l'un permettant d'avoir 10 interventions inutiles par partie de la génératrice et l'autre permettant de n'avoir aucune intervention inutile sur les 4 ans. En vert sont représentés les défauts détectés au moins une semaine à l'avant (1000 échantillons 10 minutes). Cette indication permet de voir si l'avance à la détection permettrait à un industriel comme VALEMO de modifier sa stratégie de maintenance et de limiter les pertes. En rouge, correspond toutes les détections réalisées moins de 1000 échantillons en avance, pour le nombre d'interventions inutiles fixé.

On peut voir que les défauts critiques peuvent être détectés plusieurs semaines voire un mois à l'avance. Les résidus issus d'un modèle utilisant les liens de causalité en approche multi-turbine et les résidus utilisant un modèle de similarité en approche mono-turbine permettent d'avoir des résultats similaires. Malheureusement, du fait de la présence simultanée de deux défauts sur le parc, l'approche similarité multi-turbine ne permet pas de détecter le défaut d'enroulement WT9.

Le défaut de roulement arrière de la turbine WT6 peut être détecté environ 4000 échantillons en avance, ce qui correspond à 4 semaines. Avec ce type d'information, les pertes de production auraient pu être réduites fortement. En effet, il a fallu un mois pour remplacer la génératrice de cette machine quand une à deux journées suffisent en général. Dans ce cas, c'est l'approvisionnement qui a duré le plus de temps. De plus, le défaut de roulement était tellement avancé qu'il a détruit son logement dans la génératrice. Si le défaut avait été détecté en avance, on peut penser que la génératrice n'aurait pas eu à être changée complètement, le changement de roulement seul aurait suffi. En détectant la dégradation, le coût du remplacement aurait pu être diminué de plusieurs dizaines de milliers d'euros.

Le roulement de la turbine WT9 aurait pu être détecté plus de 1300 échantillons en avance, soit plus de 9 jours. Cette fois-ci, seul le roulement avant a dû être changé, ce qui a nécessité 8 jours d'arrêt. En mettant en place des indicateurs de surveillance, les pertes de productions dues au délai d'approvisionnement auraient pu être diminuées significativement.

Les deux défauts d'enroulement WT9 et WT11 peuvent être détectés une et deux semaines en avance. Cette fois-ci, toute la génératrice a dû être changée et seules les pertes de production auraient pu être fortement diminuées. Il faut cependant rappeler que ces défauts d'enroulement sont causés par une surchauffe de la génératrice qui a détérioré prématurément les isolants entre phases. D'autres défauts d'enroulement sont causés par des problèmes de nature électrique et n'auraient pas pu être détectés de cette manière.

L'utilisation d'indicateurs permet maintenant de sensibiliser fortement les équipes sur l'impact des surchauffes sur la durée de vie des roulements. Grâce à l'indicateur, on peut voir assez facilement les défauts de ventilateur et s'assurer que ces derniers durent le moins de temps possible.

En fixant maintenant le nombre d'interventions inutiles à zéro, on remarque que seule une combinaison des méthodes multi-turbines utilisant des résidus provenant de modèles basés sur la similarité et de modèles utilisant les liens de causalité permet la détection des défauts critiques. En effet, le défaut d'enroulement WT9 est détecté par le modèle utilisant les liens de causalité multi-turbine et le défaut de roulement avant WT6 est détecté par le modèle similarité multi-turbine. La détection se fait moins en avance, mais reste supérieure à 1000 échantillons, ce qui permet d'avoir le temps d'agir et de diminuer significativement les pertes de production. Pour le moment, aucune méthode automatique n'a été proposée pour fixer le seuil, mais on peut voir qu'il existe une configuration où le nombre d'interventions inutiles est nul.

La diminution du nombre d'interventions est extrêmement importante pour un industriel. En effet, pour le moment, les indicateurs proposés ne surveillent que la génératrice et un parc de 6 machines. L'objectif est de généraliser ce type d'indicateurs sur d'autres composants critiques comme la multiplicatrice et le roulement principal. Ce type de généralisation va donc multiplier le nombre d'interventions inutiles par autant de fois le nombre de composants que l'on cherche à surveiller. De plus, VALEMO possède 30 parcs en exploitation, une seule intervention inutile par composant et par parc sur 4 ans représente  $(30\text{parcs} \times 3 \text{ composants}) / (4 \text{ ans} \times 12 \text{ mois}) \approx 2$  interventions inutiles par mois. Il faut bien comprendre que pour chaque alarme que remontera un système de surveillance, du temps sera passé pour la réception de l'information, son analyse et l'intervention. Deux interventions par mois restent raisonnables, mais on voit que ce chiffre peut très vite augmenter si l'on cherche à détecter plus tôt les défauts et donc mobiliser des personnes à temps plein, ce qui diminuerait la rentabilité de la surveillance.

## 6.6. Conclusion

Différentes méthodes de la littérature ont été adaptées et comparées avec une nouvelle approche sur une base de données réelle. Les indicateurs ont été calculés de manière à prendre en compte les contraintes d'implémentation industrielle et permettre d'étudier leurs impacts dans le quotidien d'une entreprise d'exploitation et de maintenance éolienne comme VALEMO.

Afin de mesurer cet impact, différents critères ont été proposés et calculés pour rendre compte au mieux des contraintes des industriels. L'avance à la détection est un critère classique, mais ne suffit pas. En effet, le nombre d'interventions inutiles que génèrent les indicateurs est crucial pour qualifier un système de détection automatique. Le principal objectif de ces méthodes est de pouvoir détecter des défauts de manière automatisée pour pouvoir faire face au nombre grandissant d'éoliennes en fonctionnement. De plus, deux autres critères ont été proposés comme la disponibilité et la persistance qui permettent de rendre compte de la stabilité des indicateurs calculés. En effet, ces indicateurs doivent pouvoir surveiller les éoliennes en continu et lorsqu'un défaut est détecté, celui-ci doit persister à indiquer que l'éolienne a un comportement anormal jusqu'à réparation. Cela permet d'avoir une plus grande confiance dans l'indicateur.

Parmi les méthodes testées, les approches multi-turbines et l'approche mono-turbine utilisant la similarité sont compatibles avec les besoins des industriels. Elles permettent de détecter les défauts au moins une semaine en avance, ce qui permet d'avoir le temps de réagir et de diminuer significativement les pertes de productions liées aux défaillances. De plus, ces détections se font avec un nombre d'interventions inutiles inférieur à 10 sur les quatre années de la base de données. Ce nombre d'interventions est particulièrement bas et tout à fait accepté par VALEMO.

Il est important de rappeler ici que ce nombre d'intervention inutile ne représente que le nombre d'intervention généré par l'indicateur le plus sensible pour chaque défaut. Or, il est nécessaire d'avoir plusieurs indicateurs pour pouvoir détecter tous les défauts de la base. Le nombre de 10 interventions inutiles est donc à multiplier par le nombre d'indicateur nécessaire pour la surveillance des éoliennes. Dans le cas de la surveillance de la génératrice de cette base de données 3 indicateurs sont nécessaires, qui génère donc 30 interventions inutiles pendant 4 ans, ce nombre reste tout à fait acceptable pour une entreprise comme VALEMO.

Cette étude permet de voir que les approches multi-turbines sont dans l'ensemble plus persistantes que les approches mono-turbine, ce qui augmente la confiance des opérateurs dans les indicateurs. Un indicateur qui a tendance à changer d'avis peut être mal interprété et ne pas donner lieu à l'intervention nécessaire. On peut aussi voir que la disponibilité de toutes les approches est suffisamment élevée pour permettre une surveillance continue des installations. Ce critère n'est pas particulièrement discriminant et ne permet pas de dégager une méthode meilleure qu'une autre sur cette base de données. Mais il est important de montrer que cette disponibilité est suffisamment élevée lorsque l'on présente des méthodes. La comparaison de plusieurs méthodes pour obtenir des résidus mono-turbine permet de mettre en évidence des différences de comportement en termes de détection des défauts. En effet, ces méthodes apparaissent comme complémentaires. Au vu des niveaux faibles d'interventions inutiles générées, il est possible d'imaginer d'utiliser ces deux sources d'informations en parallèle. On peut aussi imaginer qu'une fusion est possible pour profiter des avantages des méthodes utilisant les liens de causalité et par la similarité.

La combinaison de l'approche mono-turbine et multi-turbine permet d'améliorer les résultats en diminuant le nombre d'interventions inutiles déjà bas et en conservant la même avance à la détection. Chaque fausse intervention a un coût, lié à l'intervention elle-même ou à l'analyse des indicateurs. Un industriel cherche à diminuer au maximum ces coûts qui sont complètement improductifs.





## 7. Conclusion générale

La surveillance éolienne doit permettre de détecter l'apparition de défauts suffisamment tôt pour pouvoir adapter la stratégie de maintenance et ainsi diminuer les pertes de production. Plusieurs solutions ont été envisagées pour réaliser cette surveillance. L'une d'entre elle, les CMS, proposent de rajouter des capteurs spécialisés pour détecter les défauts. Ces systèmes ont montré leur efficacité, mais ils sont encore considérés comme trop coûteux par les industriels pour être généralisés sur toutes les éoliennes. En plus du coût des capteurs et de l'acquisition, ces systèmes nécessitent des experts pour pouvoir interpréter correctement les informations qu'ils génèrent. Par conséquent, l'utilisation de nouvelles approches est envisagée, en particulier l'utilisation des données SCADA. L'objectif de cette thèse était d'évaluer la faisabilité de création d'indicateurs de dégradation à partir de données SCADA, dans un objectif d'utilisation industrielle.

Une étude des travaux présentés dans la littérature a permis de mettre en évidence des différences entre les méthodes proposées et ainsi dégager deux familles de méthodes appelées mono-turbines et multi-turbines, ayant des avantages et des inconvénients distincts et complémentaires. Les méthodes mono-turbines identifiées proposent d'utiliser des variables venant d'une même éolienne pour construire des modèles de bon fonctionnement et générer des résidus. Ces méthodes permettent de s'affranchir d'une partie des variations des conditions de fonctionnement, mais restent très sensibles aux variations des conditions environnementales. À l'inverse, les méthodes multi-turbines, qui consistent à comparer les variables des éoliennes d'un même parc entre elles, sont sensibles aux variations des conditions de fonctionnement, mais permettent de s'affranchir d'une partie des variations de l'environnement. Une contribution de ces travaux a été de profiter des avantages de ces deux approches en les combinant dans une nouvelle méthode originale qui, en comparant les résidus provenant des méthodes mono-turbines entre les éoliennes, permet de diminuer le nombre de fausses alarmes et d'interventions inutiles.

L'étude bibliographique a montré que les données SCADA permettent de développer des solutions pour la création d'indicateurs pour la maintenance conditionnelle. En effet, les méthodes déjà proposées dans la littérature montrent une capacité à détecter les défauts de composants critiques, puisque certaines d'entre elles permettent de détecter certains défauts plus d'une semaine à l'avance. Cependant, ces méthodes n'ont pas été évaluées pleinement en termes d'interventions inutiles générées. Ainsi, des critères d'évaluation ont été proposés dans ces travaux afin d'évaluer les méthodes en fonction des besoins et des contraintes des industriels. En plus de mesurer l'avance à la détection d'un indicateur et le taux de fausses alarmes, converti en nombre d'interventions inutiles, des notions comme la disponibilité (pourcentage de temps pendant lequel l'indicateur peut être calculé) et la persistance, qui mesure le temps pendant lequel l'indicateur génère une alarme pendant une période de défaut, ont été proposées. Ces critères ont été calculés sur une base de données construite avec VALEMO pour montrer les performances des méthodes SCADA proposées dans la littérature et des méthodes originales proposées dans cette thèse.

Les critères de performances obtenus sur la base de données composée d'enregistrements de variables SCADA durant 4 ans sur un parc composé de 6 éoliennes ont permis de mettre en évidence l'apport de la combinaison entre méthodes mono-turbines et multi-turbines. En effet, l'utilisation d'une approche multi-turbine utilisant des résidus permet de diminuer les interventions inutiles. Cette diminution d'interventions inutiles permet de limiter le surcoût de la mise en place d'une maintenance conditionnelle par rapport à une maintenance périodique ou corrective. De plus, cela permet à

l'opérateur d'avoir plus confiance dans les indicateurs, ce qui permettra par la suite de réagir plus rapidement.

La base de données testée provient de données réelles et contient des défauts apparus par le passé. L'évaluation sur cette base a montré que non seulement le nombre d'interventions inutiles peut être réduit, mais qu'il est possible qu'il s'approche de zéro en utilisant la méthode proposée combinant une approche multi-turbine et mono-turbine. Cette possibilité d'avoir un nombre d'interventions inutiles proche de zéro, tout en étant capable de détecter les défauts suffisamment tôt ; peut permettre d'implémenter cette solution à plus grande échelle. En effet, pour le moment, cette technique a été testée sur un parc de 6 machines, mais VALEMO gère une flotte d'une centaine de machines, ce qui peut rendre la gestion des interventions inutiles coûteuses si elles sont trop importantes.

Les tests sur cette base de données ont aussi montré l'efficacité des méthodes utilisant des données SCADA. On peut voir qu'en affinant les modèles et en combinant les approches, il est possible de détecter les défauts présents dans la base suffisamment tôt, avec un nombre d'interventions inutiles raisonnable et un indicateur disponible tout au long de l'année. Ces détections auraient pu permettre de diminuer significativement les pertes de productions, voir, de limiter les dégâts collatéraux occasionnés par certain défaut.

Cette première thèse et ces résultats positifs ont permis de convaincre la société VALEMO de continuer d'investir dans ces techniques de surveillance utilisant les données SCADA. Ces techniques vont maintenant être implémentées progressivement sur la flotte entière. Ces premiers travaux de recherche entre le GIPSA-lab et VALEMO ont permis de dégager de nouvelles perspectives qui sont détaillées par la suite. Ces nouvelles perspectives feront l'objet d'une nouvelle collaboration entre VALEMO et le GIPSA-lab dans le cadre d'une nouvelle thèse CIFRE.

Ces perspectives d'amélioration sont présentées ci-dessous.

### **7.1. Les différences entre méthodes utilisant les liens de causalité et la similarité**

L'étude des modèles utilisant les liens de causalité a montré que la dynamique des composants est difficile à prendre en compte, mais qu'elle pourrait être améliorée. Ainsi, les délais sur les variables d'entrées doivent être étudiés en fonction de la nature du composant que l'on cherche à surveiller et des connaissances disponibles sur la machine. Cependant, en multipliant les délais et les entrées du modèle, il deviendra difficile pour un modèle linéaire d'appréhender les relations entre les variables. D'autres types de modèles plus complexes peuvent être envisagés comme des réseaux de neurones.

Notons que la dynamique des échanges thermiques peut aussi être plus rapide que l'échantillonnage 10 minutes fourni avec les données SCADA actuellement, ce qui est un frein pour ces modèles. A l'avenir, les nouveaux SCADA permettront d'avoir des données échantillonnées toutes les secondes, ce qui permettra sûrement de créer des modèles utilisant les liens de causalité prenant encore mieux en compte la dynamique et donc le fonctionnement machine en général.

Les méthodes par Similarité peuvent aussi être améliorées, en ajoutant d'autres variables contextuelles par exemple. De précédents auteurs avaient montré que ces modèles plus « complexes » n'apportaient qu'une faible amélioration dans (Schlechtingen and Santos, 2011). Une étude de

---

performance sur un parc plus important et avec des critères plus proches de l'industrie permettra peut-être de conclure autrement.

La comparaison sur cette base de données montre que les méthodes utilisant les liens de causalité ou la similarité sont complémentaires, il est possible d'imaginer une combinaison des deux techniques pour créer des modèles plus efficaces. En effet, les défauts semblent influencer différemment les indicateurs mono-turbines. Les défauts de ventilateur sont plus facilement détectés par les méthodes utilisant les liens de causalité et, au contraire, sont moins détectés par les méthodes utilisant la similarité. À l'inverse, les méthodes utilisant la similarité semblent être capables de détecter les défauts critiques plus tôt. Ces deux méthodes semblent donc complémentaires, leur utilisation en même temps doivent pouvoir améliorer encore les performances de détection.

## 7.2. L'apprentissage

Il est important de garder à l'esprit que des modèles plus complexes doivent être conçus en respectant les contraintes de mise en service des industriels. La taille de la base d'apprentissage pour les modèles linéaires présentés ne dépasse pas 2 semaines. Un ANN avec de nombreuses entrées devra sûrement apprendre sur un voire plusieurs mois, ce qui n'est pas envisageable dans un contexte industriel. L'effet des opérations de maintenance sur les éoliennes impose des réapprentissage réguliers. On ne peut pas perdre la surveillance pour des opérations mineures (changement de soft) pendant plusieurs mois. L'utilisation d'apprentissage adaptatif et d'une configuration par défaut de démarrage peut limiter ces effets, cependant, mais doit être étudiée. Une première proposition de modèle adaptatif a été faite dans (Bangalore and Tjernberg, 2014), mais les performances n'ont pas été testées. L'article montre que l'utilisation d'un modèle adaptatif permet de diminuer l'erreur de prédiction des variables SCADA, mais comme cette méthode n'a pas été testée sur des défauts réels, il est pour le moment impossible de conclure sur l'augmentation des performances. En effet, l'une des difficultés d'un modèle adaptatif est de trouver un compromis entre s'adapter aux nouvelles conditions de fonctionnement et ne pas apprendre le comportement pendant un défaut. Or, on peut voir que certains défauts sont très lents comme le défaut de roulement arrière qui a progressé sur plusieurs mois. Il peut y avoir une incompatibilité entre vouloir s'adapter aux périodicités des données SCADA, qui sont quotidiennes et saisonnières, et détecter au plus tôt un défaut qui progresse très lentement.

Le modèle adaptatif ne résoudrait pas complètement le problème de l'apprentissage. En effet, pour être complètement opérationnel sur un nouveau parc, il faut quand même un nombre minimal d'échantillons. Ce nombre  $N_A$  définit dans la thèse est fixé pour les tests à 2000 échantillons, ce qui correspond à 2 semaines. Un industriel peut raisonnablement attendre deux semaines dans le cadre de la mise en service d'un nouveau parc, mais les modèles plus complexes peuvent avoir besoin de plus d'échantillons. Une solution à ce problème pourrait être d'établir des paramètres par défaut pour les modèles. Ces paramètres par défaut pourraient être définis tout d'abord par expérience. Par exemple, on peut imaginer que les paramètres des modèles entre les éoliennes d'un même constructeur soient similaires. Mais ces paramètres ne dépendent pas que des conditions de fonctionnement, mais aussi de l'environnement. Les spécificités géographiques du site éolien peuvent avoir une influence importante. Une perspective intéressante serait d'étudier l'influence du site et des modèles sur les paramètres pour pouvoir définir des paramètres par défaut à utiliser lors de la mise en service pour permettre à la surveillance d'être mise en place au plus tôt.

La surveillance proposée est conçue pour détecter des défauts lors du fonctionnement de l'éolienne tout au long de son cycle de vie. Si cette surveillance peut être mise en place pendant la mise en service, elle pourrait par exemple aider à détecter des problèmes d'installation, ce qui accélérerait la mise en service du parc éolien.

### **7.3. Le seuil et l'évaluation des performances**

Les critères d'évaluations des performances proposées permettent de prendre en compte les contraintes industrielles. La simple étude de la courbe Avance à la Détection/Interventions inutiles ne suffit pas. On voit que la première détection peut être due à un dépassement de seuil isolé et d'une durée très faible. Le critère de persistance est donc essentiel à l'analyse des performances.

La persistance est un critère qui permet de faire une distinction entre les différents indicateurs. Cela permet de différencier une détection qui serait faite de pics isolés difficiles à interpréter pour un opérateur, d'un changement de tendance plus persistant qui, lui, serait directement associé à un défaut.

Cependant, la prise de décision par simple franchissement de seuil est limitée. Une perspective intéressante de ce travail est de réfléchir à une méthode de fusion des indicateurs et de prise de décision plus élaborée que le franchissement de seuil.

L'une des difficultés rencontrées lors de cette thèse est que les connaissances du terrain peuvent être limitées sur la présence de certains défauts. En effet, durant ces travaux, le calcul des premiers résidus et indicateurs a permis de mettre en évidence la présence de défauts jusqu'à lors inconnus. En enquêtant dans les rapports de maintenance, il est souvent possible de trouver une opération de réparation qui a permis de faire disparaître l'anomalie détectée, mais il est très difficile d'estimer le début des périodes de défauts. Dès lors, il peut être délicat de définir des périodes où le fonctionnement est considéré comme normal ou des périodes de défaut.

### **7.4. Influence de la charge sur l'indicateur**

L'évolution des indicateurs pendant les défauts semble influencée par la charge de la machine. On peut le voir dans la Figure 6.40 qui montre l'évolution du défaut de roulement avant la panne. Prendre en compte la charge de la machine pour adapter le seuil permettrait peut-être d'améliorer les résultats. Cela a été proposé par (Butler, 2012), les indicateurs fournis par une méthode par similarité peuvent être ensuite séparés en fonction des gammes de vitesse de rotation de l'arbre rapide. Même si le modèle est censé prendre en compte le fonctionnement machine, celui-ci ne reflète que l'influence de la charge sur la température dans des conditions normales. Le modèle ne prend pas en compte les variations de l'influence du défaut en fonction de la charge sur la température. En étant capable de prendre ces variations en compte, on peut espérer augmenter l'avance à la détection.

## 7.5. Diagnostic

Les méthodes présentées dans cette thèse permettent de détecter des anomalies grâce à des modèles de bon fonctionnement. Il semble possible d'isoler les défauts en fonction de l'indicateur qu'il l'a détecté. Le roulement avant n'est détecté que par des indicateurs surveillant la température de ce roulement, alors que le défaut de roulement arrière n'est détecté que par les indicateurs surveillant sa température.

Localiser le défaut de la sorte permettra d'aider les équipes de maintenances. Mais, à l'avenir, pour des parcs difficiles d'accès ou pour l'Offshore, il est crucial de connaître la criticité du défaut. Les 8 défauts de la base nous permettent de constater des différences de comportement dans les indicateurs en fonction du type de défaut. Les défauts de roulements semblent progressifs alors que les défauts de ventilateur et d'enroulement sont plus soudains. Il semble donc possible en étudiant l'évolution de l'indicateur pendant une détection, d'isoler la nature du défaut. Néanmoins, ce type de recherche nécessite une base de données de défaut bien plus importante que les 8 défauts présentés.

## 7.6. Généralisation à d'autres parcs

Il est crucial d'étudier l'impact des différences entre les modèles de turbine sur les méthodes proposées. En plus des différences entre les modèles de turbine viendront s'ajouter les différences entre les topologies des parcs. Le parc étudié dans cette thèse est de taille moyenne comprenant 6 machines et n'est soumis qu'à très peu de sillage. D'autres parcs, comme les parcs en mer, sont composés de plusieurs dizaines de turbines, ce qui peut nécessiter une adaptation des méthodes lors de la comparaison entre turbines. L'hypothèse que le vent est similaire sur toutes les turbines peut ne plus être respectée si le parc est étalé sur une grande surface. Il est alors possible d'imaginer que des groupes de turbines fonctionneront ensemble et que la comparaison devra se faire par groupe.

De plus, les modèles proposés ne se concentrent pour le moment que sur la génératrice et en particulier sur les échanges thermiques. Il existe d'autres mesures de température sur la multiplicatrice et sur les roulements principaux qui pourrait permettre d'étendre la surveillance sur les autres composants de la machine. D'autres modèles doivent donc être conçus, en particulier les modèles utilisant la similarité puisqu'il faut trouver la meilleure température de référence. Les températures des composants critiques permettront leur surveillance, mais d'autres composants comme les pales ou le convertisseur ne sont pas équipés de capteurs de température. Pour le moment, l'échantillonnage 10 minutes imposait de se concentrer sur des défauts mécaniques ayant une dynamique lente. En revanche, avec l'arrivée des données secondes, des modèles plus proches du contrôle des composants pourront être conçus. En effet, le SCADA fournira des données de courant et tension ou encore les angles de pales et direction de la nacelle toutes les secondes, ce qui pourra permettre de vérifier l'efficacité et l'invariance du système de contrôle des différents composants de la turbine. De plus, la performance de la machine en terme de génération d'électricité pourra être aussi être mieux surveiller qu'avec des données 10 minutes.

À la fin de cette thèse, des travaux ont été engagés par Monsieur Usama AZIZ, dans le cadre d'un stage de master chez VALEMO, pour montrer la capacité de généralisation de ces méthodes sur d'autres types de parcs et de composants. Ces travaux ont été réalisés sur un parc de 18 machines dans une plaine de la région champagne. L'intérêt de ce parc est sa différence avec celui testé

précédemment. En effet, ce parc a 3 fois plus de machines qui proviennent d'un autre constructeur. Ensuite, ce parc n'est pas situé en montagne, mais dans une plaine avec un profil de vent beaucoup moins turbulent. Finalement, ce parc est organisé en 3 lignes de 6 machines collées les unes aux autres pour former un rectangle de 3 par 6. Cette différence d'organisation engendre des effets de sillage. Les machines percevant le vent en premier créent des sillons et augmentent sa turbulence pour les machines suivantes. Ce phénomène de sillage était presque inexistant dans le parc précédent et nous cherchons maintenant à évaluer l'impact du sillage sur les nouvelles méthodes proposées.

Ce stage a permis de montrer que ces méthodes peuvent être implémentées sur d'autres modèles d'éoliennes. Malheureusement, le nombre de défauts détectables n'était pas suffisant dans l'historique des données pour pouvoir conclure sur les performances en appliquant la méthode d'évaluation proposée dans cette thèse. Cependant, ce stage a permis de conclure sur plusieurs points.

Dans un premier temps, nous avons pu constater que la surveillance éolienne souffre pour le moment d'un manque de connaissance de terrain. La base de données de ce nouveau parc a été construite à partir des rapports d'intervention et d'exploitation. L'objectif est de définir des périodes de bon fonctionnement et des périodes de défauts que l'on va chercher à détecter. Cependant, lors du calcul des premiers indicateurs, un grand nombre de variations a pu être observé. Ces observations ont pu ensuite être liées à des événements. Ces événements n'avaient pas été pris en compte au début, car ils étaient noyés dans un grand nombre d'autres interventions et semblaient totalement anodins. Il est donc important de gagner progressivement en expérience pour mieux comprendre quelles interventions ou défaillances aura ou non un impact sur les données SCADA. Il est aussi très important de changer la manière dont les rapports sont écrits. En effet, pour le moment, les rapports de maintenance se concentraient uniquement sur les activités de maintenance usuelle, et le niveau de détail est insuffisant pour faire de la surveillance à grande échelle et faire du diagnostic.

Dans un deuxième temps, nous avons pu voir que le nombre d'éoliennes et les effets de sillage ne semblent pas avoir un impact significatif sur les indicateurs calculés. Nous nous attendions à observer que les indicateurs de certaines éoliennes évoluent en groupe, liées entre elles par une position géographique par exemple, mais cela n'est pas apparu de manière claire dans les résultats. On peut toutefois imaginer que ces groupes existent et que de futurs travaux de recherche puissent en tirer parti.

## **7.7. Utilisation de données extérieures et fusion des indicateurs**

Enfin, en plus des données SCADA 10 minutes, il existe d'autres données et d'autres indicateurs qui permettraient d'améliorer les performances de surveillance. En particulier, le SCADA fournit des alarmes qui peuvent fournir des compléments d'information. Dans (Qiu et al., 2015, 2011) les auteurs montrent que ces alarmes peuvent être retraitées pour apporter des compléments d'information et aider au diagnostic une fois le défaut apparu. Ensuite, dans (Godwin et al., 2013; Godwin and Matthews, 2013; Kusiak and Li, 2011), les alarmes sont combinées aux données SCADA pour aider au diagnostic. En plus de ces données qualitatives, les indicateurs construits avec les données haute fréquence des CMS peuvent aussi être utilisés en parallèle, comme proposé dans (Feng et al., 2011) où les indicateurs créés à partir des données vibratoires de la chaîne cinématique sont associés à des indicateurs créés à partir des données SCADA. Malheureusement, ces méthodes n'ont pas été évaluées en terme de fausses alarmes. De futurs travaux de recherche permettraient de développer des méthodes de fusion de ces différents indicateurs avec les indicateurs SCADA.

Finalement, nous avons pu constater que la surveillance des composants des éoliennes nécessite l'utilisation de plusieurs indicateurs provenant des données SCADA, en parallèle. De plus, d'autres techniques comme le CMS ou les alarmes doivent permettre d'améliorer les performances en détection. En revanche, la multiplication des indicateurs peut engendrer une multiplication du nombre de fausses alarmes et donc du nombre d'interventions inutiles. Il est donc crucial de développer des techniques pour fusionner ces indicateurs et éviter que les opérateurs soient « noyés » par trop d'information. La redondance d'information entre les systèmes de surveillance doit permettre de diminuer les fausses alarmes et aussi d'aider à établir un diagnostic précis. L'objectif pour des entreprises comme VALEMO est de pouvoir automatiquement être informé de l'apparition de défaut. Il est important de se rendre compte que, pour le moment, aucune action de maintenance ne se fait sans avoir au préalable été sur site pour vérifier visuellement la présence d'un défaut. C'est encore possible de travailler de cette manière pour l'éolien en terre, mais pour l'éolien en mer, les industriels auront besoin d'outils performants pour permettre la détection et l'identification des défauts. La planification d'une opération en mer est un problème technique en lui-même et il est impossible financièrement d'effectuer des interventions inutiles.





## 8. Bibliographie

- AFNOR, 2001. NF EN 13306 X 60-319 Terminologie de la maintenance.
- Astolfi, D., Castelani, F., Terzi, L., 2014. Fault prevention and diagnosis through scada temperature data analysis of an onshore wind farm. *Diagnostyka* 15.
- Bangalore, P., Tjernberg, L.B., 2014. Self evolving neural network based algorithm for fault prognosis in wind turbines: A case study. Presented at the Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, Durham, England, pp. 1–6. doi:10.1109/PMAPS.2014.6960603
- Butler, S., 2012. Prognostic Algorithms for Condition Monitoring and Remaining Useful Life Estimation. National University of Ireland, Maynooth.
- Butler, S., Ringwood, J., O'Connor, F., 2013. Exploiting scada system data for wind turbine performance monitoring. Presented at the Control and Fault-Tolerant Systems (SysTol), IEEE, Nice, France, pp. 389–394.
- Chen, B., Matthews, P.C., Tavner, P.J., 2013. Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS. *Expert Syst. Appl.* 40, 6863–6876. doi:10.1016/j.eswa.2013.06.018
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern classification, 2. ed. ed, A Wiley-Interscience publication. Wiley, New York, NY.
- Edenhofer, O., IPCC (Eds.), 2012. Renewable energy sources and climate change mitigation: special report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, Cambridge.
- Elijorde, F., Kim, S., Lee, J., 2014. A Wind Turbine Fault Detection Approach Based on Cluster Analysis and Frequent Pattern Mining. *KSII Trans. Internet Inf. Syst.* TIIS 8, 664–677.
- EUROSERV'ER, 2016. Baromètre ÉOLIEN.
- European Wind Map, 2008. . Presented at the European Wind Energy Association (EWEA).
- Farkas, Z., 2011. Considering air density in wind power production. *ArXiv Prepr.* ArXiv11032198.
- Feng Yanhui, Yang Wenxian, Cao Mengnan, Wang Hao, Sun Juan, Infield, D., Qiu Yingning, 2014. Model Based Wind Turbine Gearbox Fault Detection on SCADA Data. *Institution of Engineering and Technology*, pp. 1.1.3–1.1.3. doi:10.1049/cp.2014.0820
- Feng, Y., Qiu, Y., Crabtree, C.J., Long, H., Tavner, P.J., 2011. Use of SCADA and CMS signals for failure detection and diagnosis of a wind turbine gearbox. Presented at the European Wind Energy Association (EWEA), Brussels, Belgium.
- Garcia, M.C., Sanz-Bobi, M.A., del Pico, J., 2006. SIMAP: Intelligent System for Predictive Maintenance. *Comput. Ind.* 57, 552–568. doi:10.1016/j.compind.2006.02.011
- Godwin, J.L., Matthews, P.C., 2013. Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis. *IJPHM Spec. Issue Wind Turbine PHM Color* 4, 90.
- Godwin, J.L., Matthews, P.C., Watson, C., 2013. Classification and detection of electrical control system faults through scada data analysis. *Chem. Eng. Trans.* 33, 985–990. doi:10.3303/CET1333165
- Gray, C., Langmayr, F., Haselgruber, N., Watson, S.J., 2011. A Practical Approach to the Use of SCADA Data for Optimized Wind Turbine Condition Based Maintenance, in: EWEA. Presented at the EWEA, Amsterdam.
- Groupe d'experts intergouvernemental sur l'évolution du climat, Pachauri, R.K., Meyer, L.A., 2015. Changements climatiques 2014: rapport de synthèse : contribution des Groupes de travail I, II et III au cinquième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat. GIEC, Genève (Suisse).
- Isermann, R., 2006. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer, Berlin ; New York.

- Kruger, M., Ding, S.X., Haghani, A., Engel, P., Jeinsch, T., 2013. A data-driven approach for fault diagnosis in gearbox of wind energy conversion system. Presented at the Control and Fault-Tolerant Systems (SysTol), IEEE, pp. 359–364.
- Kusiak, A., Li, W., 2011. The prediction and diagnosis of wind turbine faults. *Renew. Energy* 36, 16–23. doi:10.1016/j.renene.2010.05.014
- Lebranchu, A., Charbonnier, S., Berenguer, C., Prevost, F., 2016. Using SCADA Data for Fault Detection in Wind Turbines: Local Internal Model Versus Distance to a Wind Farm Reference, in: Chaari, F., Zimroz, R., Bartelmus, W., Haddar, M. (Eds.), *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*. Springer International Publishing, Cham, pp. 225–236.
- Li, S., Wunsch, D.C., O'Hair, E.A., Giesselmann, M.G., 2001. Using neural networks to estimate wind turbine power generation. *IEEE Trans. Energy Convers.* 16, 276–282. doi:10.1109/60.937208
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Calif., pp. 281–297.
- McLaughlin, D., Clive, P.J.M., McKenzie, J.H., 2009. Wind farm performance assessment in the real world. Presented at the European Wind Energy Conference (EWEC), Marseille, France.
- McMillan, D., Ault, G.W., 2007. Quantification of Condition Monitoring Benefit for Offshore Wind Turbines. *Wind Eng.* 31, 267–285. doi:10.1260/030952407783123060
- Papatheou, E., Dervilis, N., Maguire, E., Worden, K., 2014. Wind turbine structural health monitoring: a short investigation based on SCADA data, in: *EWSHM-7th European Workshop on Structural Health Monitoring*. Nantes, France.
- Qiu, Y., Richardson, P., Feng, Y., Tavner, P.J., Erdos, G., 2011. Scada alarm analysis for improving wind turbine reliability. Presented at the European Wind Energy Association (EWEA), Brussels, Belgium.
- Qiu, Y., Zhang, W., Cao, M., Feng, Y., Infield, D., 2015. An Electro-Thermal Analysis of a Variable-Speed Doubly-Fed Induction Generator in a Wind Turbine. *Energies* 8, 3386–3402. doi:10.3390/en8053386
- RTE, 2014. bilan électrique.
- Schlechtingen, M., Santos, I.F., 2014. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: Application examples. *Appl. Soft Comput.* 14, 447–460. doi:10.1016/j.asoc.2013.09.016
- Schlechtingen, M., Santos, I.F., 2011. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech. Syst. Signal Process.* 25, 1849–1875.
- Schlechtingen, M., Santos, I.F., Achiche, S., 2013. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description. *Appl. Soft Comput.* 13, 259–270. doi:10.1016/j.asoc.2012.08.033
- Statistique Canada, 2009. Production, transport et distribution d'électricité (au catalogue: 57-202-X). *Statistique Canada* 9–12.
- Swets, J.A., 1996. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers, Scientific psychology series. L. Erlbaum Associates, Mahwah, N.J.
- Takoutsing, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T., Ekemb, G., 2014. Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges. *Energies* 7, 2595–2630. doi:10.3390/en7042595
- Tavner, P.J., 2012. *Offshore wind turbines: reliability, availability and maintenance, IET power and energy series*. Institution of Engineering and Technology, London, U.K.

- 
- Van Trees, H.L., 2001. Detection, estimation, and modulation theory: Part 1: Detection, estimation, and linear modulation theory. Wiley, New York Chichester.
- Wei, X., Liu, L., 2010. Fault detection of large scale wind turbine systems, in: Computer Science and Education (ICCSE), 2010 5th International Conference on. IEEE, pp. 1299–1304.
- Wenxian, Y., Jiasheng, J., 2011. Wind Turbine Condition Monitoring and Reliability Analysis by SCADA Information, in: 2011 Second International Conference on Mechanic Automation and Control Engineering July 15-17, 2011, Inner Mongolia, China: Proceedings. IEEE, Piscataway, N.J., pp. 1872–1875.
- Wilkinson, M., Harman, K., van Delft, T., Darnell, B., 2014. Comparison of methods for wind turbine condition monitoring with SCADA data. IET Renew. Power Gener. 8, 390–397. doi:10.1049/iet-rpg.2013.0318
- Wind in power, European statistics 2011, 2012. . Presented at the European Wind Energy Association (EWEA).
- Yang, W., Court, R., Jiang, J., 2013. Wind turbine condition monitoring by the approach of SCADA data analysis. Renew. Energy 53, 365–376. doi:10.1016/j.renene.2012.11.030
- Yang, W., Tavner, P.J., Crabtree, C.J., Feng, Y., Qiu, Y., 2014. Wind turbine condition monitoring: technical and commercial challenges: Wind turbine condition monitoring: technical and commercial challenges. Wind Energy 17, 673–693. doi:10.1002/we.1508
- Zaher, A., McArthur, S.D.J., Infield, D.G., Patel, Y., 2009. Online wind turbine fault detection through automated SCADA data analysis. Wind Energy 12, 574–593. doi:10.1002/we.319
- Zhang, Z.-Y., 2014. Comparison of data-driven and model-based methodologies of wind turbine fault detection with scada data. Presented at the European Wind Energy Association (EWEA), Barcelona, Spain.

## **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**

Le secteur de l'énergie éolienne est en pleine expansion depuis les 10 dernières années. Le nombre et la taille des éoliennes ont été multipliés, ce qui accroît la difficulté et la criticité de la maintenance, et impose aux industriels de passer d'une maintenance corrective et systématique à une maintenance conditionnelle et prédictive. L'objectif de ces travaux est de développer des indicateurs de dégradation, en utilisant les données numériques fournies par le SCADA, disponibles à faible coût mais enregistrées à une faible fréquence d'échantillonnage (10 min) dans un objectif de suivi de production.

Une analyse bibliographique approfondie des travaux réalisés sur la surveillance de parcs éoliens à partir de données SCADA montre que deux types d'approches sont généralement proposés. Les méthodes dites mono-turbines où un modèle de comportement d'une turbine est appris sur des périodes de bon fonctionnement, à partir duquel il est possible de créer des résidus mesurant la différence entre la valeur prédite par le modèle et la mesure en ligne, qui servent d'indicateurs de défaut. Les modèles mono-turbines ont la particularité d'utiliser des variables provenant de la même turbine du parc. Les deuxièmes méthodes, dites multi-turbines, sont des méthodes où l'effet parc et la similarité entre machines sont utilisés. Là où les recherches les plus récentes proposent principalement de créer des courbes de performances pour chaque machine du parc pendant une période de temps et de comparer ces courbes entre elles, nous faisons la proposition originale de combiner les deux approches et de comparer les résidus mono-turbines à une référence parc traduisant le comportement majoritaire des turbines du parc.

Nous validons de manière extensive ces indicateurs en analysant leurs performances sur une base de données composée d'enregistrements de variables SCADA d'une durée de 4 ans sur un parc de 6 machines. Nous proposons aussi des critères de performances pertinents permettant d'évaluer de manière réaliste les gains et éventuels surcoûts que généreraient ces indicateurs s'ils étaient intégrés dans un outil de maintenance. Ainsi, nous montrons que le taux d'interventions inutiles associées à des fausses alarmes produites par les indicateurs de défaut, et qui provoquent un surcoût très important pour l'entreprise, peut être fortement diminué par la fusion d'indicateurs mono-turbines que nous proposons, tout en conservant une avance à la détection suffisante pour planifier la mise en place d'interventions par les équipes de maintenance.

**Mots-clés :** analyse de données, surveillance & diagnostic, pronostic, maintenance prédictive, fiabilité, éolienne

**Monitoring data analysis and synthesis of deterioration & failure indicators for predictive maintenance decision-making regarding wind turbines.**

The wind energy sector has rapidly grown in the last 10 years. The number and the size of wind turbines have multiplied, which increases the difficulty and the criticality of the maintenance, and forces the wind turbine industry to change from a corrective and systematic maintenance to a conditional and predictive maintenance. The objective of this research is to develop failure indicators using numerical SCADA data, available at a low price but with a very low sampling frequency (10 min), in order to make online monitoring.

A thorough bibliographical analysis on the surveillance of wind farms using SCADA data shows that two types of approaches are usually suggested. The first approach, called mono-turbine, where a good behaviour model of a turbine is learnt over unfaulty periods. With this approach, it is possible to create residuals measuring the difference between the predicted value by the model and the on-line measure, which serves as failure indicators. The mono-turbine models have the peculiarity in that they use variables coming from the same turbine as the farms. The second approach, called multi-turbine, are methods where the similarity between machines is used. Where the most recent researches mostly suggest creating performance curves for every machine on the farm during a period of time and comparing these curves between each other, we make the original proposal to combine both approaches and compare mono-turbine residuals with a farm reference representing the behaviour of the turbines of the farm.

We validate in an extensive way those failure indicators by analysing their performances on a database made up of SCADA variable recordings of a duration of 4 years on a windfarm of 6 machines. We also propose relevant performance criteria allowing an estimation in a realistic way of the gains and possible additional costs which would generate these indicators if they were integrated into a tool of maintenance. Therefore, we show that the rate of useless interventions associated with false alarms produced by the failure indicators, which cause a heavy additional cost for the company, can be strongly decreased by the mono-turbines indicators merging that we propose, while preserving a sufficient detection time for the maintenance teams to plan interventions.

**Keywords :** data analysis, monitoring & diagnosis, prognosis, predictive maintenance, reliability, wind turbine