

Contribution au diagnostic et au pronostic par mesure indirecte
des actifs d'énergies renouvelables à partir des données SCADA
avec quantification d'incertitude

par

Abderraouf Ramdane BENABDESSELAM

THÈSE PRÉSENTÉE À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION
DU DOCTORAT EN GÉNIE
Ph.D.

MONTREAL, LE 28 AVRIL 2026

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Abderraouf Ramdane Benabdesselam, 2026



Cette licence Creative Commons signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette oeuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'oeuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE:

M. Antoine Tahan, directeur de thèse
Département de génie mécanique - École de Technologie Supérieure

M. Martin Gagnon, codirecteur
Institut de Recherche d'Hydro-Québec

M. Roland Maranzana, président du jury
Département de génie des systèmes - École de Technologie Supérieure

M. Éric David, membre du jury
Département de génie mécanique - École de Technologie Supérieure

M. Adrian Ilinca, membre du jury
Département de génie mécanique - École de Technologie Supérieure

M. André Bégin-Drolet, examinateur externe indépendant
Département de génie mécanique et de génie industriel - Université Laval

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 15 AVRIL 2026

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens tout d’abord à exprimer mes plus sincères remerciements à Antoine Tahan, mon directeur de thèse, pour son encadrement attentif, rigoureux, encourageant et stimulant tout au long de mon doctorat. Je lui suis profondément reconnaissant de m’avoir offert l’opportunité de réaliser une thèse au sein de son équipe de recherche et d’avoir cru en moi lorsque, parfois, moi-même j’en doutais. Sa grande expertise, sa disponibilité, sa reconnaissance de mon travail, ainsi que sa bienveillance et son humanité ont contribué à instaurer un environnement de recherche bienveillant et fructueux.

Mes remerciements s’adressent également à Martin Gagnon, mon codirecteur de thèse et chercheur à l’IREQ, pour son accompagnement scientifique rigoureux, sa bienveillance et sa grande expertise, qui ont grandement contribué à la réussite de mon projet. C’est à ses côtés que j’ai pleinement compris le véritable sens des mots rigueur scientifique et précision. Je ne le remercierai jamais assez de m’avoir initié au monde de l’intelligence artificielle et de la quantification de l’incertitude. Avec Martin, j’apprends chaque jour un peu plus.

Je souhaite également remercier Monelle Comeau, experte principale en science des données chez Power Factors, de m’avoir offert l’opportunité de découvrir et d’explorer le domaine de l’éolien. Sa méticulosité, sa bienveillance et sa très grande disponibilité ont été plus que précieuses tout au long de cette aventure. Son accompagnement attentif et ses conseils avisés ont largement contribué à enrichir mon parcours de recherche.

Mes remerciements s’adressent également à ceux qui étaient présents à mes débuts, lorsque mes idées portaient dans toutes les directions, bien qu’elles se dispersent encore, en particulier à Quentin Dollon, ainsi qu’à Quang Hung Pham et Ryad Zemouri, tous chercheurs à l’IREQ. J’exprime aussi ma reconnaissance envers l’ensemble de l’équipe de l’IREQ pour sa bienveillance. Je remercie également notre équipe de recherche à l’ÉTS, en particulier Adaiton Oliveira-Filho et James Cave pour tous nos échanges et collaborations.

Je tiens également à exprimer ma gratitude envers l'ÉTS, Hydro-Québec, Power Factors, Mitacs, le CRSNG pour la confiance et le soutien financier apportés à nos travaux de recherche. J'adresse également mes sincères remerciements aux membres du jury pour le temps consacré à la lecture et à l'évaluation de cette thèse.

Je conclus avec des remerciements personnels. À ma mère, et à Yemma : quand on a trop de mots, on ne sait plus parler simplement. Moi qui sais d'ordinaire écrire mais ne sais jamais parler, cette fois-ci je ne sais même plus écrire. Simplement, je ne vous remercierai jamais assez. À la mémoire de mon père : « Même si tant d'étoiles se sont anéanties, le ciel ne s'anéantira jamais ». À toute ma famille, en particulier à mon frère, à ma sœur et à mes oncles, pour votre soutien et votre amour : merci pour tout. À mes acolytes de Polytechnique dispersés un peu partout dans le monde, après une vie de militaire qui a fait ressortir le meilleur de nous, mais aussi le pire ! À cette nouvelle vie loin de ma Méditerranée, dépourvue de chaleur mais riche en bienveillance, là où le mal se fait très mal, mais où l'on finit par retrouver le nord. À toute personne ayant rendu cette aventure plus agréable, ne serait-ce que par un simple sourire. Et enfin, c'est pour me rappeler que le succès n'est pas final, que l'échec n'est pas fatal, mais que c'est le courage de continuer qui compte.

Contribution au diagnostic et au pronostic par mesure indirecte des actifs d'énergies renouvelables à partir des données SCADA avec quantification d'incertitude

Abderraouf Ramdane BENABDESSELAM

RÉSUMÉ

La transition énergétique mondiale repose sur une intégration croissante des énergies renouvelables, notamment l'éolien et l'hydroélectricité, dont la durabilité dépend de la fiabilité et de la disponibilité des actifs. Soumis à des conditions d'exploitation sévères, ces systèmes sont exposés à des mécanismes de dégradation tels que la cavitation ou la fatigue de composants critiques. Dans ce contexte, le pronostic et la gestion de la santé des systèmes (PHM) constituent un levier essentiel pour anticiper les dégradations et réduire les arrêts non planifiés.

Les approches de PHM reposent sur la disponibilité de données fiables. Les systèmes de suivi des conditions (CMS) offrent une forte sensibilité aux dégradations mais restent peu déployés en raison de contraintes économiques et opérationnelles. À l'inverse, les données SCADA sont disponibles sur l'ensemble des installations, mais présentent une résolution temporelle limitée et une observabilité partielle des mécanismes de dégradation.

En collaboration avec Hydro-Québec et Power Factors, cette thèse propose un cadre de PHM fondé exclusivement sur les données SCADA, visant, lorsque possible, à remplacer les CMS par des approches de mesure indirecte afin d'assurer une surveillance continue et déployable à l'échelle de flottes d'actifs. Les contributions portent sur le développement d'approches de mesures indirectes basées sur les données, leur exploitation pour le diagnostic et le pronostic malgré les limitations des données SCADA, ainsi que l'intégration d'une quantification d'incertitude valide et correctement calibrée. Deux cas industriels sont étudiés. Le premier concerne la surveillance et le diagnostic de la cavitation érosive dans des turbines hydrauliques Francis, où une approche de mesure indirecte fondée sur l'intelligence artificielle permet d'estimer la perte de masse due à la cavitation érosive à partir des données SCADA, en remplaçant les CMS, tout en assurant la transférabilité entre turbines. Le second porte sur le pronostic d'environ 200 éoliennes commerciales à partir de données SCADA uniquement, via la classification pronostique des anomalies thermiques puis la prédiction de la durée de vie résiduelle à l'aide d'indicateurs de santé adaptés aux contraintes SCADA. L'ensemble des approches intègre des méthodes de quantification de l'incertitude, notamment avec les méthodes conformes.

Les résultats montrent que, pour certains mécanismes de dégradation observables, les données SCADA, associées à des approches de mesure indirecte, constituent une alternative crédible aux CMS pour la surveillance, le diagnostic et le pronostic, ouvrant la voie à un déploiement industriel à grande échelle.

Mots-clés: mesures indirectes, diagnostic, pronostic, durée de vie résiduelle, quantification d'incertitude, énergies renouvelables

Contribution to the diagnosis and prognosis of renewable energy assets through indirect measurements based on SCADA data with uncertainty quantification

Abderraouf Ramdane BENABDESSELAM

ABSTRACT

The global energy transition relies on the increasing integration of renewable energy sources, notably wind and hydropower, whose sustainability depends on the reliability and availability of production assets. Subject to harsh operating conditions, these systems are exposed to degradation mechanisms such as cavitation or fatigue of critical components. In this context, Prognostics and Health Management (PHM) constitutes a key lever for anticipating degradation and reducing unplanned outages.

PHM approaches rely on the availability of reliable data. Condition Monitoring Systems (CMS) offer high sensitivity to degradation but remain sparsely deployed due to economic and operational constraints. Conversely, SCADA data are available across all installations but suffer from limited temporal resolution and partial observability of degradation mechanisms.

In collaboration with Hydro-Québec and Power Factors, this thesis proposes a PHM framework exclusively based on SCADA data, aiming, when possible, to replace CMS with indirect measurement approaches in order to ensure continuous and fleet-scale deployable monitoring. The contributions focus on the development of data-driven indirect measurements, their exploitation for diagnosis and prognosis despite the limitations of SCADA data, and the integration of valid and well-calibrated uncertainty quantification. Two industrial case studies are investigated. The first addresses the monitoring and diagnosis of erosive cavitation in Francis hydraulic turbines, where an artificial intelligence-based indirect measurement approach enables the estimation of mass loss due to erosive cavitation from SCADA data, replacing CMS while ensuring transferability between turbines. The second focuses on the prognosis of approximately 200 commercial wind turbines using SCADA data only, through prognostic classification of thermal anomalies followed by remaining useful life prediction using health indicators adapted to SCADA constraints. All approaches integrate uncertainty quantification methods, in particular conformal methods.

The results show that, for certain observable degradation mechanisms, SCADA data, combined with indirect measurement approaches, constitute a credible alternative to CMS for monitoring, diagnosis, and prognosis, paving the way for large-scale industrial deployment.

Keywords: indirect measurements, diagnosis, prognosis, remaining useful life estimation, uncertainty quantification, renewable energies

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
0.1 Bilan de l'introduction	7
0.2 Organisation de la thèse	7
 CHAPITRE 1 REVUE DE LITTÉRATURE	 11
1.1 Mesures indirectes	11
1.1.1 Les approches basées sur les données	14
1.1.1.1 Modèles autorégressifs	14
1.1.1.2 Modèles d'apprentissage automatique	14
1.1.2 Les approches basées sur la physique	16
1.1.3 Les approches hybrides	16
1.1.4 Discussion	17
1.2 Pronostic	19
1.2.1 Pronostic des turbines hydrauliques	21
1.2.2 Pronostic des éoliennes	23
1.2.2.1 Pronostic des éoliennes à partir des données de vibration	23
1.2.2.2 Pronostic des éoliennes à partir des données SCADA	24
1.2.3 Discussion	26
1.3 Quantification d'incertitude	27
1.4 Bilan de la littérature, lacunes et opportunités	31
 CHAPITRE 2 MÉTHODOLOGIE DE RECHERCHE	 33
2.1 Objectif général et cadre de la recherche	33
2.2 Questions de recherche	33
2.3 Sélection des cas d'application	34
2.4 Base de données	36
2.5 Hypothèses et portée de la recherche	37
2.6 Déploiement	38
2.7 Retombées scientifiques et industrielles	38
2.8 Mise en contexte des publications scientifiques	39
 CHAPITRE 3 DIAGNOSTIC OF EROSION CAVITATION IN HYDRAULIC TURBINES FROM INDIRECT MEASUREMENTS USING SCADA DATA WITH UNCERTAINTY QUANTIFICATION	 41
3.1 Abstract	41
3.2 Introduction	42
3.2.1 Indirect Measurements	44
3.2.2 Uncertainty Quantification	45
3.2.3 PHM for Erosive Cavitation	47
3.2.4 Key Gaps	48

3.2.5	Objectives and Contributions	49
3.2.6	Study Case and Data	50
3.2.7	Organization of the paper	51
3.3	Methodology	51
3.3.1	Prediction of ICR	51
3.3.2	Uncertainty Quantification	54
3.3.2.1	Heteroscedastic Modeling	55
3.3.2.2	Split Conformal Prediction	55
3.3.2.3	Conformalized Deep Ensembles	56
3.3.3	Transfer Learning	57
3.3.4	Estimation of Cumulative Cavitation Rate	58
3.3.5	Evaluation Metrics	59
3.4	Results and Discussions	61
3.4.1	Prediction of ICR	61
3.4.2	Uncertainty Quantification	64
3.4.3	Transfer Learning	68
3.4.4	Prediction of CCR	73
3.5	Conclusion	75
CHAPITRE 4 PROGNOSTICS OF THERMAL ANOMALIES IN WIND TURBINES VIA DEEP LEARNING AND CONFORMAL PREDICTION USING SCADA DATA		
4.1	Abstract	79
4.2	Introduction	80
4.2.1	Related Works	81
4.2.2	Key Gaps	85
4.2.3	Objectives	86
4.2.4	Contributions and Outline	88
4.3	Data Sources	88
4.4	Methodology	89
4.4.1	Timeline Partition and Data Preprocessing	90
4.4.2	Indirect Measurement and Uncertainty Quantification	92
4.4.3	Event Functions	98
4.4.4	Evaluation Metrics	100
4.5	Results	101
4.5.1	Case Study : Main-Bearing Threshold Exceedance Event	101
4.5.2	Sensitivity Analyses	103
4.5.3	Main Classification Performances	106
4.5.4	Historical Exogenous Input Approximation	109
4.5.5	Comparison with Existing Studies	111
4.6	Conclusion	112

CHAPITRE 5	REMAINING USEFUL LIFE PREDICTION FOR GRADUAL THERMAL ANOMALIES IN WIND TURBINES USING SCADA DATA	115
5.1	Abstract	115
5.2	Introduction	115
	5.2.1 SCADA-Based Prognostics	118
	5.2.2 Key Gaps in the Literature	119
	5.2.3 Problem Statement	120
	5.2.4 Objectives and Contributions	121
5.3	Methodology	121
	5.3.1 Health Indicator Construction and Data Preprocessing	122
	5.3.2 Transfer Function Selection	125
	5.3.3 Uncertainty Quantification	129
	5.3.4 Evaluation Metrics	131
5.4	Results and Discussion	131
	5.4.1 Case Study Description	131
	5.4.2 Transfer Function Selection	132
	5.4.3 Uncertainty Quantification	137
	5.4.4 Effect of Monotonicity Constraints on HI and RUL	139
5.5	Conclusion	140
	CONCLUSION ET RECOMMANDATIONS	143
6.1	Apports de la thèse	143
6.2	Limites de la thèse	145
6.3	Recommandations et perspectives de recherche	146
6.4	Épilogue	150
ANNEXE I	ARTICLES DE REVUE ET DE CONFÉRENCE	153
	BIBLIOGRAPHIE	157

LISTE DES TABLEAUX

	Page
Tableau 0.1	Comparaison des approches de surveillance des actifs des énergies renouvelables 5
Tableau 2.1	Publications scientifiques issues des travaux de recherche doctorale 39
Tableau 3.1	Variables used in this study 50
Tableau 3.2	Uncertainty calibration metrics 60
Tableau 3.3	Performance comparison of models on both turbines 61
Tableau 3.4	Comparison of uncertainty quantification methods on the four datasets 65
Tableau 3.5	Impact of different priors on BNN uncertainty quantification 67
Tableau 3.6	Comparison of uncertainty quantification methods on transfer scenarios 71
Tableau 4.1	Description of the key variables used in the study 89
Tableau 4.2	Summary of anomaly statistics per component from September 2022 to March 2024 89
Tableau 4.3	Matthews Correlation Coefficient (MCC) for different θ_c values across prediction horizons 105
Tableau 4.4	Matthews Correlation Coefficient (MCC) for different Δt_{train} values across prediction horizons 106
Tableau 4.5	Classification outcomes for multiple prediction horizons with optimal hyperparameters ($\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, $\theta_c = 1$) 106
Tableau 4.6	Performance metrics for multiple prediction horizons with optimal hyperparameters ($\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, $\theta_c = 1$) 107
Tableau 4.7	Classification outcomes per component across prediction horizons 108
Tableau 4.8	Confusion matrices values at different prediction horizons for optimal hyperparameters with observed exogenous data 110
Tableau 5.1	Description of the key variables used in the study 122

Tableau 5.2	Performance comparison of the tested models	134
Tableau 5.3	Sensitivity analysis of the calibration results for different numbers of folds K	137

LISTE DES FIGURES

	Page
Figure 0.1 Part de la production mondiale d'électricité des énergies renouvelables Adaptée de International Energy Agency (2023)	2
Figure 0.2 Représentation du processus PHM Adaptée de Moreira de Oliveira Filho (2024, p. 16)	3
Figure 0.3 Structure de la thèse	8
Figure 1.1 Représentation de l'approche de mesure indirecte pour un système MISO	12
Figure 1.2 Principales approches de mesure indirecte appliquées aux énergies renouvelables	13
Figure 1.3 Estimation de la RUL Adaptée de Javed, Gouriveau & Zerhouni (2017, p. 219)	20
Figure 1.4 Principales approches de pronostic appliquées aux énergies renouvelables	21
Figure 1.5 Principales méthodes de quantification de l'incertitude	29
Figure 2.1 Synthèse des objectifs de la thèse	33
Figure 3.1 Example of cavitation erosion on a blade Tirée de Clermont-Bodson (2015, p. 8)	44
Figure 3.2 Diagnostic of erosive cavitation methodology	52
Figure 3.3 Prediction of the instantaneous cavitation rate for Unit 1 on Dataset 1 ...	69
Figure 3.4 Prediction of the instantaneous cavitation rate - Unit 1 - Dataset 2	69
Figure 3.5 Prediction of the instantaneous cavitation rate - Unit 2 - Dataset 1	70
Figure 3.6 Prediction of the instantaneous cavitation rate - Unit 2 - Dataset 2	70
Figure 3.7 Transfer Learning from Unit 1 to Unit 2	72
Figure 3.8 Transfer Learning from Unit 2 to Unit 1	73

Figure 3.9	Evolution of cumulative cavitation - Unit 1 - Dataset 1 - Assuming independence	74
Figure 3.10	Evolution of cumulative cavitation - Unit 1 - Dataset 1 - Assuming perfect autocorrelation	75
Figure 3.11	Evolution of cumulative cavitation - Unit 1 - Dataset 2 - Assuming perfect autocorrelation	76
Figure 4.1	Main bearing temperature time series for a healthy and a degraded wind turbine	81
Figure 4.2	Thermal monitoring methodology for WT components	90
Figure 4.3	Proposed timeline showing training, calibration, and testing intervals ...	91
Figure 4.4	Illustration of LSTM training phase	93
Figure 4.5	Illustration of LSTM test phase	94
Figure 4.6	Illustration of LSTM calibration	96
Figure 4.7	Prediction of WT-1 main-bearing temperature	102
Figure 4.8	Evolution of IEI and CEI over time	104
Figure 5.1	Illustration of RUL prediction Adaptée de Javed <i>et al.</i> (2017, p. 219) ...	117
Figure 5.2	RUL prediction methodology for WT components	122
Figure 5.3	ΔT time series for a healthy and a degraded WT	123
Figure 5.4	Cumulative Maximum filter on ΔT	125
Figure 5.5	Schematic illustration of RUL prediction	126
Figure 5.6	Examples of degradation trajectories extracted from ΔT for different WTs	133
Figure 5.7	RUL predictions for each WT using the GRU model with LOO validation and uncertainty quantification using CQR	135
Figure 5.8	Effect of ΔT smoothness on RUL prediction	136
Figure 5.9	Uncertainty estimation before and after conformal calibration	138
Figure 5.10	Effect of monotonicity constraints on HI and RUL	140

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ACC	<i>Accuracy</i> Précision
AI	<i>Artificial Intelligence</i> Intelligence artificielle
AR	<i>Autoregressive</i> Modèle autorégressif
ARIMA	<i>Autoregressive Integrated Moving Average</i> Modèle autorégressif à moyenne mobile intégrée
Bi-GRU	<i>Bidirectional Gated Recurrent Unit</i> Réseau de neurones récurrent GRU bidirectionnel
Bi-LSTM	<i>Bidirectional Long Short-Term Memory</i> Réseau de neurones à mémoire longue à court terme bidirectionnel
BNN	<i>Bayesian Neural Network</i> Réseau de neurones bayésien
CAE	<i>Convolutional Autoencoder</i> Autoencodeur convolutionnel
CapsNet	<i>Capsule Neural Network</i> Réseau de neurones à capsules
CCR	<i>Cumulative Cavitation Rate</i> Taux de cavitation cumulé
CDE	<i>Conformalized Deep Ensembles</i> Méthode des ensembles conformalisés
CEI	<i>Cumulative Event Intensity</i> Intensité cumulée de l'événement
CI	<i>Confidence Interval</i> Intervalle de confiance

XX

CM	<i>Condition Monitoring</i> Maintenance conditionnelle
CM	<i>Cumulative Maximum</i> Maximum cumulé d'une série temporelle
CMS	<i>Condition Monitoring System</i> Systèmes de suivi des conditions
CNN	<i>Convolutional Neural Network</i> Réseau de neurones convolutionnel
CP	<i>Conformal Prediction</i> Méthodes conformes
CQR	<i>Conformalized Quantile Regression</i> Régression quantile conformalisée
DL	<i>Deep Learning</i> Apprentissage profond
ECE	<i>Expected Calibration Error</i> Erreur de calibration attendue
ELBO	<i>Evidence Lower Bound</i> Borne inférieure de la vraisemblance marginale
FN	<i>False Negative</i> Faux négatif
FNR	<i>False Negative Rate</i> Taux de faux négatifs
FP	<i>False Positive</i> Faux positif
FPR	<i>False Positive Rate</i> Taux de faux positifs
FPT	<i>First Prediction Time</i> Instant de la première prédiction

FTT	<i>Failure Threshold Time</i> Instant d'atteinte du seuil de défaillance
GB	<i>Gradient Boosted Regression Trees</i> Arbres de régression par boosting de gradient
GP	<i>Gaussian Process</i> Processus gaussien
GPR	<i>Gaussian Process Regression</i> Régression par processus gaussien
GRU	<i>Gated Recurrent Unit</i> Unité récurrente à portes
HI	<i>Health Indicator</i> Indicateur de santé
ICR	<i>Instantaneous Cavitation Rate</i> Taux instantané de cavitation
IEI	<i>Instantaneous Event Intensity</i> Intensité instantanée de l'événement
LR	<i>Linear Regression</i> Régression linéaire
LSTM	<i>Long Short-Term Memory</i> Réseau de neurones à mémoire longue à court terme
LOO	<i>Leave-One-Out</i>
MA	<i>Miscalibration Area</i> Aire de mauvaise calibration
MAE	<i>Mean Absolute Error</i> Erreur absolue moyenne
MCC	<i>Matthews Correlation Coefficient</i> Coefficient de corrélation de Matthews

ML	<i>Machine Learning</i> Apprentissage automatique
MLP	<i>Multi-Layer Perceptron</i> Perceptron multicouches
MSE	<i>Mean Squared Error</i> Erreur quadratique moyenne
MISO	<i>Multiple Input Single Output</i> Système à entrées multiples et sortie unique
NARNN	<i>Nonlinear AutoRegressive Neural Network</i> Réseau neuronal autorégressif non linéaire
NARXNN	<i>Nonlinear AutoRegressive Neural Network with eXogenous inputs</i> Réseau neuronal autorégressif non linéaire avec entrées exogènes
nMAE	<i>Normalized Mean Absolute Error</i> Erreur absolue moyenne normalisée
nRMSE	<i>Normalized Root Mean Squared Error</i> Erreur quadratique moyenne normalisée
OC	<i>Observed Coverage</i> Couverture observée
OOD	<i>Out-of-Distribution</i> Hors distribution
O&M	<i>Operation and Maintenance</i> Exploitation et maintenance
PHM	<i>Prognostics and Health Management</i> Pronostic et la gestion de santé des systèmes
PM	<i>Performance Monitoring</i> Surveillance des performances
QR	<i>Quantile Regression</i> Régression quantile

ReLU	<i>Rectified Linear Unit</i> Fonction d'activation rectifiée
RF	<i>Random Forest</i> Forêt aléatoire
RMSCE	<i>Root Mean Squared Calibration Error</i> Racine de l'erreur quadratique moyenne de calibration
RNN	<i>Recurrent Neural Network</i> Réseau de neurones récurrent
RMSE	<i>Root Mean Squared Error</i> Racine de l'erreur quadratique moyenne
RUL	<i>Remaining Useful Life</i> Durée de vie résiduelle
SCADA	<i>Supervisory Control and Data Acquisition</i> Système de supervision et d'acquisition de données
SCP	<i>Split Conformal Prediction</i> Prédiction conforme avec partition des données
SMAPE	<i>Symmetric Mean Absolute Percentage Error</i> Erreur absolue moyenne symétrique en pourcentage
SVI	<i>Stochastic Variational Inference</i> Inférence variationnelle stochastique
SVR	<i>Support Vector Regression</i> Régression par machines à vecteurs de support
TN	<i>True Negative</i> Vrai négatif
TNR	<i>True Negative Rate</i> Taux de vrais négatifs
TP	<i>True Positive</i> Vrai positif

TPR	<i>True Positive Rate</i> Taux de vrais positifs
UQ	<i>Uncertainty Quantification</i> Quantification de l'incertitude
VI	<i>Variational Inference</i> Inférence variationnelle
WT	<i>Wind Turbine</i> Éolienne

LISTE DES SYMBOLES ET UNITÉS DE MESURE

α	Niveau de risque	[-]
α^*	Niveau critique associé au dépassement du seuil	[-]
$CCR(t)$	Taux cumulé de cavitation à l'instant t	[kg]
$D(t)$	Intensité cumulée de l'événement (CEI)	[%]
d	Dimension du vecteur d'entrée	[-]
$d(t)$	Intensité instantanée de l'événement (IEI)	[%]
ΔT	Écart de température par rapport à l'état sain	[°C]
Δt	Pas de discrétisation temporelle	[temps]
Δt_{train}	Durée de la fenêtre d'apprentissage	[temps]
Δt_{test}	Durée de la fenêtre de prédiction	[temps]
$\delta T(t)$	Indicateur de santé basé sur le maximum cumulatif de ΔT	[°C]
ECE	Erreur de calibration attendue	[-]
$E_i^{(k)}$	Score de non-conformité au pli k	[temps]
FTT	Instant de dépassement du seuil critique	[temps]
FPT	Instant de première détection de l'anomalie	[temps]
GW	Gigawatt	[GW]
H_d	Hauteur d'eau aval	[m]
H_u	Hauteur d'eau amont	[m]
ICR	Taux instantané de cavitation érosive	[kg/10000 h]
I	Intensité du courant électrique	[A]
$I^{1-\alpha}(t)$	Intervalle de confiance à l'instant t	[unité de la variable prédite]
$I_{CQR}(X)$	Intervalle de confiance conforme de la RUL	[temps]
K	Nombre de plis	[-]
MA	Aire de mauvaise calibration	[-]
MAE	Erreur absolue moyenne	[unité de la variable prédite]

M	Nombre de modèles dans l'ensemble	[-]
n	Nombre d'observations	[-]
n_{calib}	Taille de l'ensemble de calibration	[-]
n_{test}	Taille de l'ensemble de test	[-]
P	Puissance active	[MW]
P_a	Puissance active	[MW]
P_e	Puissance effective de stabilité	[MW]
P_r	Puissance réactive	[MVA _r]
Q_{CQR}	Ajustement conforme des intervalles	[temps]
$q_\alpha(X)$	Quantile conditionnel d'ordre α	[unité de la variable prédite]
R^2	Coefficient de détermination	[%]
$RMSC E$	Racine de l'erreur quadratique moyenne de calibration	[-]
$RMSE$	Racine de l'erreur quadratique moyenne	[unité de la variable prédite]
$RUL(t)$	Durée de vie résiduelle réelle	[temps]
$\widehat{RUL}(t)$	Durée de vie résiduelle estimée	[temps]
s_i	Score de non-conformité	[unité de la variable prédite]
t	Temps	[temps]
T_{amb}	Température ambiante	[°C]
T_{comp}	Température du composant surveillé	[°C]
T_{ref}	Température de référence en état sain	[°C]
$T_{threshold}$	Seuil critique de température	[°C]
U	Vitesse du vent	[m/s]
$U(t)$	Incertitude associée à la prédiction instantanée	[unité de la variable prédite]
$U^{1-\alpha}$	Incertitude associée au niveau de confiance $1 - \alpha$	[unité de la variable prédite]
$U^c(t)$	Incertitude associée à la cavitation cumulée	[kg]
VO	Ouverture des directrices	[%]
y_i	Valeur cible de la durée de vie résiduelle	[temps]

$\hat{y}$	Prédiction ponctuelle du taux instantané de cavitation	[kg/10000 h]
$\hat{y}^c(t)$	Prédiction ponctuelle cumulée du taux de cavitation	[kg]
$\dot{m}$	Débit volumique de l'eau	[m ³ /s]
ρ	Coefficient d'autocorrélation	[-]
σ	Écart-type d'une prédiction	[unité de la variable prédite]
τ_0	Instant initial d'intégration	[temps]
θ_c	Seuil de l'événement	[%]
$\mathbb{R}$	Ensemble des nombres réels	
$\mathbb{R}_+$	Ensemble des réels positifs	

INTRODUCTION

La dépendance prolongée aux énergies fossiles a mis en évidence la vulnérabilité d'un modèle énergétique fondé sur des ressources limitées. Cette dépendance a été révélée par le choc pétrolier de 1973 et aggravée récemment par les changements climatiques et la croissance de la demande énergétique au niveau mondial. Cette situation rend indispensable la transition vers les énergies renouvelables, seul moyen permettant de concilier sécurité énergétique, durabilité et réduction des impacts environnementaux. Les énergies renouvelables représentent aujourd'hui environ 40 % de la capacité électrique mondiale installée (IRENA, 2023). L'hydroélectricité demeure la principale source, avec environ 1391 GW, suivie du solaire photovoltaïque (1055 GW) et de l'éolien (899 GW), tandis que la bioénergie, la géothermie et le solaire thermique contribuent à la diversification de l'offre énergétique mondiale. À noter que, depuis les années 2010, les filières solaire et éolienne connaissent une croissance particulièrement soutenue, stimulée par les progrès technologiques et la baisse continue des coûts de production, comme illustré à la Figure 0.1.

Alors que les systèmes photovoltaïques, dépourvus d'éléments mécaniques tournants, présentent peu de défaillances mécaniques, les installations hydroélectriques et éoliennes – soumises à des conditions d'exploitation sévères et à des charges dynamiques fluctuantes – demeurent sujettes à divers modes de dégradation, tels que la fatigue, la cavitation ou l'usure, tous susceptibles d'affecter leur fiabilité et leurs performances (Benabdesselam *et al.*, 2024). Face à ces phénomènes de dégradation, et dans la mesure où la durabilité des énergies renouvelables repose sur une gestion optimisée des coûts d'exploitation et de maintenance, la maintenance a évolué d'une maintenance réactive à une maintenance préventive systématique, puis conditionnelle, en surveillant l'état de santé des actifs, et enfin prédictive, en prédisant l'évolution future de l'état de santé, ouvrant ainsi la voie à une généralisation sous le concept de pronostic et de gestion de la santé des systèmes (*Prognostics and Health Management*, PHM), dans une perspective de gestion proactive des défaillances, intégrant plusieurs niveaux d'analyse, de la collecte et

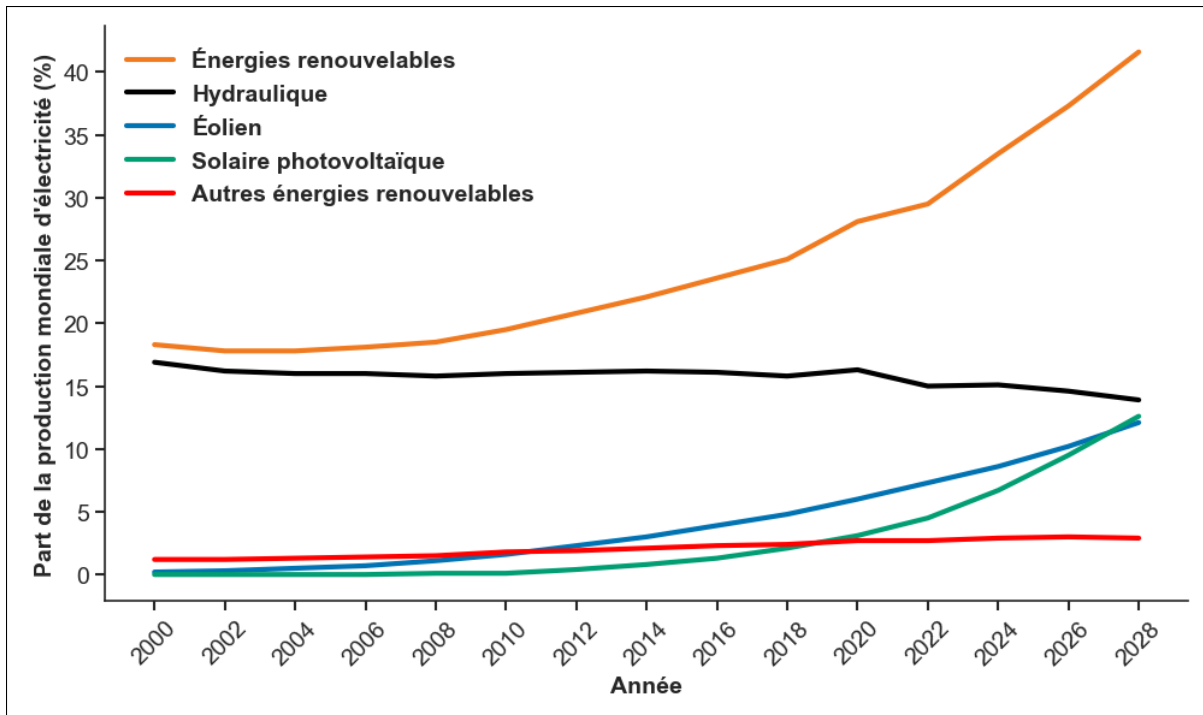


Figure 0.1 Part de la production mondiale d'électricité des énergies renouvelables

Adaptée de International Energy Agency (2023)

la consolidation des données, l'extraction des descripteurs, la construction et la validation de modèles, jusqu'à la prise de décision finale, comme l'illustre la Figure 0.2.

De manière générale, ce processus se décompose en trois (3) phases principales : la détection d'anomalies, le diagnostic, et le pronostic, détaillées comme suit (Gouriveau, Medjaher & Zerhouni, 2017) :

- **Détection d'anomalie** : cette étape vise à identifier tout écart par rapport au comportement « normal » à partir d'indicateurs de santé (*Health Indicators*, HIs), en émettant un signal d'alerte. Les performances sont donc : temporelles (détection précoce) et de précision (le moins possible de fausses alarmes).
- **Diagnostic** : le diagnostic vise d'abord à identifier et à caractériser la cause et la nature de l'anomalie détectée, laquelle peut résulter d'une dégradation réelle ou de phénomènes

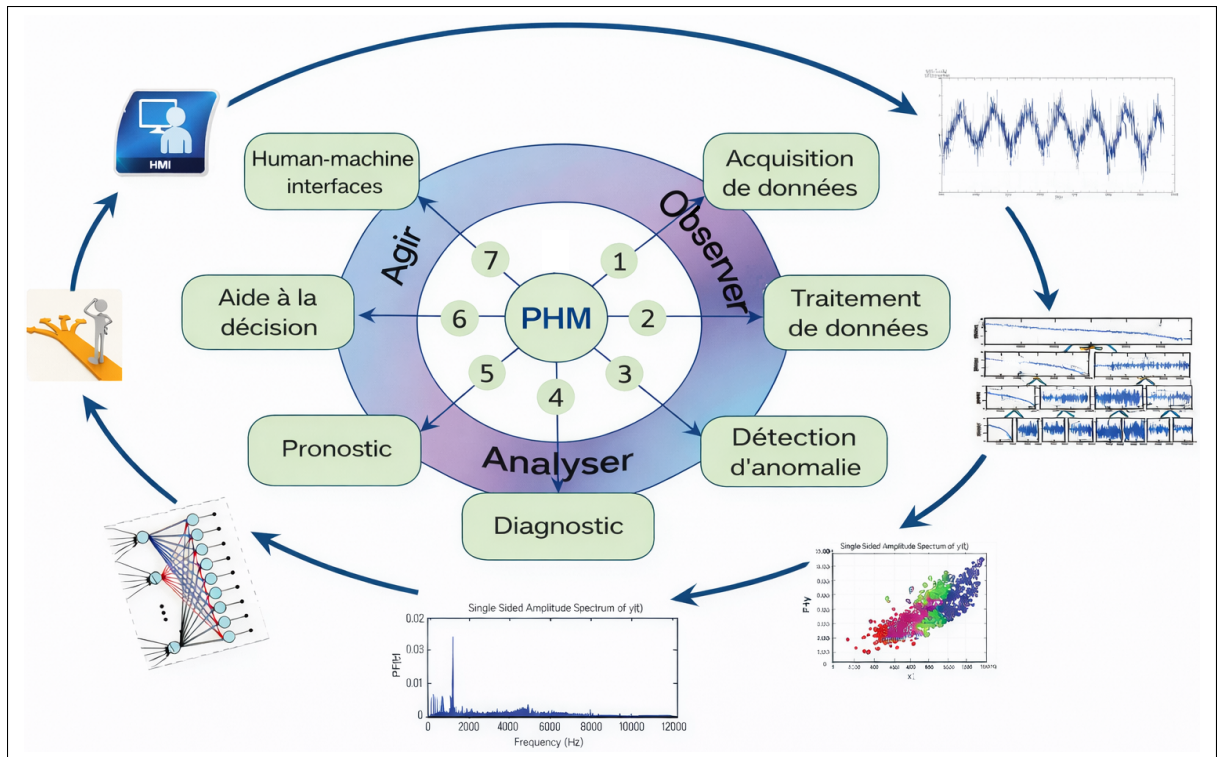


Figure 0.2 Représentation du processus PHM
Adaptée de Moreira de Oliveira Filho (2024, p. 16)

bénins, tels que des variations des conditions d'exploitation, des changements des conditions environnementales ou un défaut de capteur. Lorsque l'anomalie est confirmée comme correspondant à une dégradation effective, l'objectif devient alors de quantifier l'état de dégradation du composant concerné.

- **Pronostic** : le pronostic correspond au processus de prédiction de l'évolution future de l'état de santé d'un composant (ou d'un système). Il vise, dans un premier temps, à évaluer si la dégradation est susceptible d'évoluer vers une défaillance et, le cas échéant, à estimer le temps restant avant que celle-ci ne se produise, appelé durée de vie résiduelle (*Remaining Useful Life*, RUL), permettant ainsi de planifier les interventions de maintenance, d'optimiser l'exploitation des actifs et de réduire les coûts associés aux arrêts non planifiés.

Or, la mise en œuvre d'une démarche PHM repose sur un prérequis fondamental : la disponibilité de données. En effet, l'efficacité et la justesse des approches PHM reposent sur l'accès à des mesures fiables, représentatives de l'état de santé réel de l'actif surveillé et disponibles en continu (Gouriveau *et al.*, 2017; Javed *et al.*, 2017; Zio, 2022). Dans ce contexte, deux (2) grandes familles d'approches de suivi sont généralement distinguées : la maintenance conditionnelle (*Condition Monitoring*, CM) et le suivi des performances (*Performance Monitoring*, PM) (Rezamand *et al.*, 2020).

La CM repose sur les *Condition Monitoring Systems* (CMS), qui acquièrent des signaux à haute fréquence, un signal pouvant ici être considéré comme haute fréquence à partir de 1 Hz (Wang, Vidal & Pozo, 2026; Ibrahim *et al.*, 2023b). Ces systèmes englobent notamment les mesures de vibrations, d'acoustique et de lubrification. En raison de leur forte sensibilité aux défauts mécaniques et aux mécanismes de dégradation, ces signaux constituent une source de données particulièrement informative pour le PHM (Rezamand *et al.*, 2020). Dans les turbines hydrauliques, les CMS sont utilisés pour surveiller des modes de défaillance majeurs tels que la cavitation érosive et la fatigue (Chatenet, 2020; Gagnon, Vouligny, Cauchon & Giroux, 2022; Dollon, 2021). Sur les éoliennes, ils permettent le suivi de composants critiques (roulements, boîte de vitesses, pales) soumis à des charges cycliques et à des phénomènes de fatigue et d'érosion (Dollon, 2021; Chatenet, 2020; Dao, Kazemtabrizi & Crabtree, 2019). Malgré leurs performances, le déploiement des CMS à grande échelle demeure limité en raison des coûts élevés, de la grande quantité de données nécessitant du stockage et du traitement du signal, des besoins en calibration et en inspection, ainsi que des contraintes d'exploitation en environnements sévères, ce qui conduit à des campagnes de mesure ponctuelles plutôt qu'à une surveillance continue (Moreira de Oliveira Filho, 2024; Clermont-Bodson, 2015). En conséquence, l'adoption d'un PHM en continu fondé sur les CMS demeure marginale. La majorité des éoliennes n'en sont pas équipées et, dans les turbines hydrauliques, ces systèmes ne couvrent généralement que certains

modes de dégradation, laissant des phénomènes critiques comme la cavitation érosive rarement instrumentés (Rezamand *et al.*, 2020; Moreira de Oliveira Filho, 2024; Clermont-Bodson, 2015).

À l'inverse, le PM s'appuie sur les données issues du système de contrôle et d'acquisition de données (*Supervisory Control And Data Acquisition*, SCADA) qui est utilisé pour l'ajustement automatique du régime de fonctionnement des machines en fonction des conditions météorologiques et des exigences du réseau électrique. Ce système centralise les mesures provenant d'une multitude de capteurs – températures, courants, puissance active, vitesse du vent, etc. – acquises à différentes fréquences, puis agrégées sous forme de séries temporelles à pas de 10 minutes (Rezamand *et al.*, 2020). Bien que les systèmes SCADA n'aient pas été initialement conçus pour la surveillance de l'état de santé (mais plutôt des performances énergétiques de l'actif), de nombreuses études ont démontré que certains processus de dégradation se traduisent par des signatures visibles au sein des variables enregistrées par le système SCADA, ouvrant ainsi la voie à leur exploitation pour la surveillance de l'état des équipements (Moreira de Oliveira Filho, 2024; Yan, Hui, Jiang & Li, 2025; Maldonado-Correa, Martín-Martínez, Artigao, Honrubia-Escribano & Gómez-Lázaro, 2020). Les principaux avantages et inconvénients du CM et du PM sont présentés dans le Tableau 0.1.

Tableau 0.1 Comparaison des approches de surveillance des actifs des énergies renouvelables

Méthode de surveillance	CM	PM
Mesures utilisées	Capteurs CMS : vibrations, acoustique, lubrification	SCADA : puissance, températures, etc.
Avantage principal	Haute sensibilité aux mécanismes de dégradation	Mesures disponibles en continu, sans capteurs additionnels
Limitation	Coût élevé d'instrumentation, grandes quantités de données nécessitant traitement et stockage, et capteurs non systématiquement installés	Sensibilité limitée à la dégradation

Étant donné que les données SCADA sont déjà disponibles sur la quasi-totalité des actifs de production, qu'elles ne nécessitent pas de capteurs supplémentaires – et donc ni coûts d'acquisition ni d'installation additionnels, ni opérations de remplacement – et qu'elles sont disponibles en continu, il apparaît opportun de tirer profit de cette source de données pour remplacer les CMS, afin de mettre en œuvre un PHM en continu et de déployer des stratégies de surveillance à grande échelle. Une approche fondée sur la **mesure indirecte** peut alors être envisagée pour estimer l'état de dégradation des composants à partir de ces données opérationnelles et réaliser des fonctions de PHM, tout en quantifiant l'incertitude des prédictions, un enjeu d'actualité encore peu traité dans le cadre du PHM des systèmes d'énergies renouvelables.

Or, en raison des limitations inhérentes aux données SCADA, notamment leur sensibilité restreinte à la dégradation et leur faible résolution temporelle, certaines formes de dégradation ne peuvent pas être surveillées à l'aide de ces données. Lorsqu'il apparaît, pour un cas de dégradation donné, qu'il est possible de remplacer les CMS par les données SCADA pour assurer la surveillance et réaliser la mesure indirecte de la grandeur physique décrivant la dégradation, l'étape suivante consiste à appliquer, pour ce cas et à partir des données SCADA, le processus de PHM, allant de la détection d'anomalies au diagnostic et au pronostic. De manière générale, la détection d'anomalies constitue un domaine assez mature dans la littérature, et de nombreuses approches ont été développées, qu'elles reposent sur des données SCADA ou sur des CMS, allant des méthodes les plus simples (Cambron, Tahan, Masson & Pelletier, 2017) aux plus sophistiquées (Dollon, Labbé & Léonard, 2023; Moreira de Oliveira Filho, 2024). Le diagnostic est relativement mature dans certaines applications, notamment pour les éoliennes (Moreira de Oliveira Filho, 2024). Toutefois, il demeure plus complexe pour des actifs à faible taux de défaillance et pour lesquels les données *Run-to-Failure* sont rares, comme les turbines hydrauliques, ce qui freine le développement d'approches avancées comparables. Par ailleurs, le diagnostic fondé sur des mesures indirectes reste encore peu exploré dans la littérature. Enfin, l'étape de pronostic demeure marquée par d'importants défis scientifiques et

techniques (Moreira de Oliveira Filho, 2024), en particulier lorsque l'on s'appuie exclusivement sur des données SCADA.

0.1 Bilan de l'introduction

Face aux dégradations affectant les actifs des énergies renouvelables, en particulier dans les secteurs hydroélectrique et éolien, la surveillance continue des équipements est essentielle. Or, les CMS étant rarement déployés, les données SCADA, disponibles par défaut sur les unités de production, constituent une alternative privilégiée. Il devient ainsi nécessaire de développer des approches de mesure indirecte exploitant les données SCADA afin d'estimer des grandeurs physiques représentatives de la dégradation et de permettre le diagnostic en l'absence de CMS. Une fois cette surveillance rendue possible et le diagnostic établi à partir de mesures indirectes, le pronostic s'impose comme le second enjeu majeur de cette thèse. Il constitue un défi majeur en général, et plus particulièrement lorsqu'il repose exclusivement sur les données SCADA.

0.2 Organisation de la thèse

Cette thèse vise à contribuer au développement d'approches permettant d'exploiter les données fournies par le système d'archivage SCADA. L'objectif *in fine* est de tirer parti des données SCADA pour surveiller l'état de dégradation des actifs des énergies renouvelables par « mesures indirectes », en assurer le diagnostic, et jeter les bases d'un pronostic fondé sur ces données.

Chaque chapitre correspond à un article publié ou soumis à une revue avec un comité de lecture, à l'exception du Chapitre 1, consacré à la revue de littérature, et du Chapitre 2, dédié à exposer la méthodologie de recherche. L'ensemble des chapitres couvre les différentes étapes du travail présentées dans la Figure 0.3.

Cette thèse est structurée comme suit :

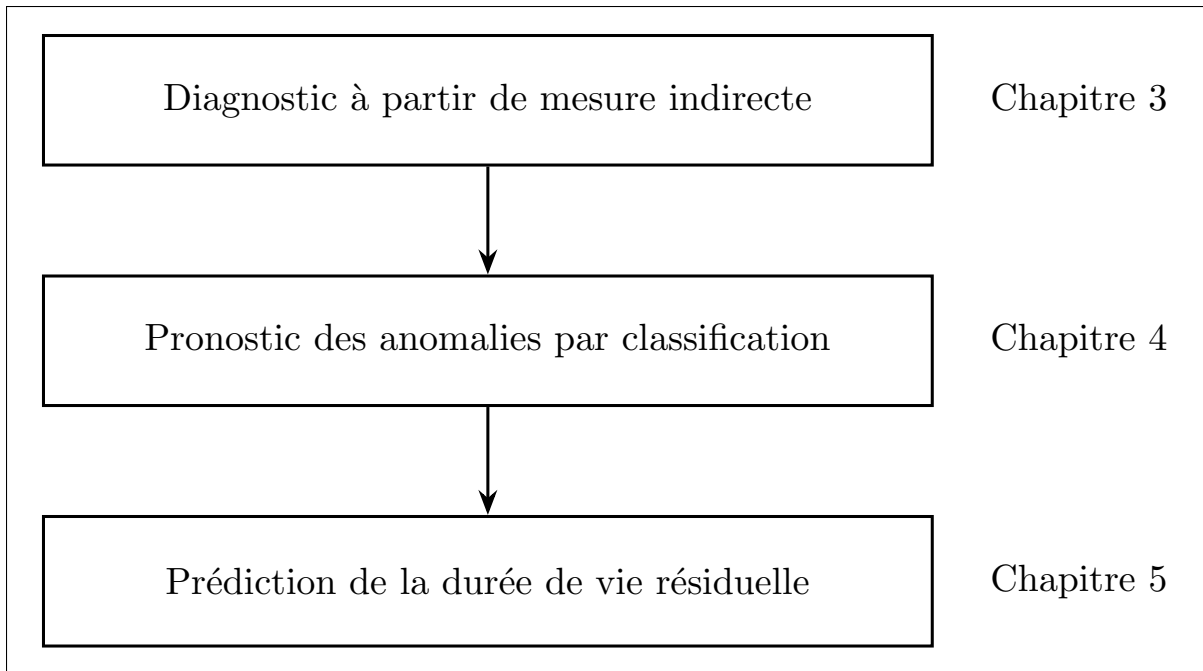


Figure 0.3 Structure de la thèse

- **Introduction** Présente le contexte et le positionnement de la recherche appliquée, met en évidence les principales lacunes identifiées dans l'état de l'art général et expose les motivations ayant conduit à la réalisation de cette étude.
- **Chapitre 1** Revue de littérature : établit le cadre théorique de la recherche et propose une synthèse structurée de l'état de l'art des approches de mesure indirecte, de pronostic et de quantification d'incertitude appliquées aux actifs des énergies renouvelables.
- **Chapitre 2** Objectifs et méthodologie : précise les objectifs spécifiques de la recherche, formule les hypothèses de travail et décrit les approches méthodologiques mises en œuvre pour répondre aux problématiques identifiées.
- **Chapitre 3** Reproduction de l'article 1, *Diagnostic of Erosive Cavitation in Hydraulic Turbines From Indirect Measurements Using SCADA Data With Uncertainty Quantification*. Ce chapitre présente une approche de mesure indirecte de la dégradation causée par la cavitation érosive dans les turbines hydrauliques, fondée exclusivement sur les données

SCADA, en remplacement des CMS, et conçue pour être transférable à d'autres turbines. Il formalise également la détermination de l'état de dégradation, permettant d'en assurer le diagnostic, tout en quantifiant l'incertitude des prédictions.

- **Chapitre 4** Reproduction de l'article 2, *Prognostics of Thermal Anomalies in Wind Turbines via Deep Learning and Conformal Predictions with SCADA Data*. Ce chapitre présente une approche de pronostic fondée sur un classificateur capable de prédire, à partir des données SCADA, si une anomalie thermique détectée sur une éolienne est susceptible d'évoluer vers une dégradation future en exploitant l'incertitude des prédictions.
- **Chapitre 5** Reproduction de l'article 3, *Remaining Useful Life Prediction for Gradual Thermal Anomalies in Wind Turbines Using SCADA Data*. Ce chapitre développe une approche de prédiction de la durée de vie résiduelle des éoliennes à partir des données SCADA avec une quantification d'incertitude.
- **Conclusion et recommandations** Fournit une vue d'ensemble des principaux résultats et contributions de la thèse, met en lumière ses limites et en souligne les perspectives, tout en présentant des pistes de recherche futures et une liste de recommandations.

CHAPITRE 1

REVUE DE LITTÉRATURE

Après avoir souligné dans l'introduction l'intérêt des données SCADA pour la surveillance et le diagnostic des actifs des énergies renouvelables via une approche de mesure indirecte, la Section 1.1 présente une revue de la littérature consacrée aux approches de mesure indirecte appliquées aux actifs des énergies renouvelables, **qui s'appuie sur un article de revue publié par le candidat en tant que premier auteur** (Benabdesselam *et al.*, 2024). Considérant que la détection d'anomalies et le diagnostic constituent aujourd'hui des domaines assez matures dans la littérature, et que le pronostic demeure un enjeu scientifique majeur et toujours largement ouvert, la Section 1.2 est consacrée à une revue critique des approches de pronostic appliquées aux actifs des énergies renouvelables. Enfin, la Section 1.3 aborde la question de la quantification de l'incertitude des prédictions, qui se situe à l'interface des mesures indirectes et du pronostic, et constitue un enjeu méthodologique transversal essentiel à l'exploitabilité des approches de PHM en contexte industriel.

1.1 Mesures indirectes

Une mesure indirecte vise à estimer une quantité latente d'intérêt $y \in \mathbb{R}$ inaccessible à partir de l'observation de n grandeurs physiques accessibles $\mathbf{x} \in \mathbb{R}^n$, parfois de nature physique différente, décrivant un système à entrées multiples et sortie unique (*Multiple Input Single Output*, MISO), qui constitue la forme la plus couramment utilisée, bien que d'autres configurations puissent être envisagées. Cette prédiction repose sur l'établissement d'une relation causale entre les quantités mesurées et la grandeur d'intérêt, appelée fonction de transfert f , telle que $y = f(\mathbf{x})$, comme illustré à la Figure 1.1. Cette fonction peut prendre diverses formes, allant d'une expression algébrique, à un système d'équations différentielles, ou encore un modèle fondé sur l'intelligence artificielle.

Les mesures indirectes sont privilégiées lorsque la mesure directe est inaccessible pour des raisons techniques ou économiques. L'exploitation des mesures indirectes repose généralement

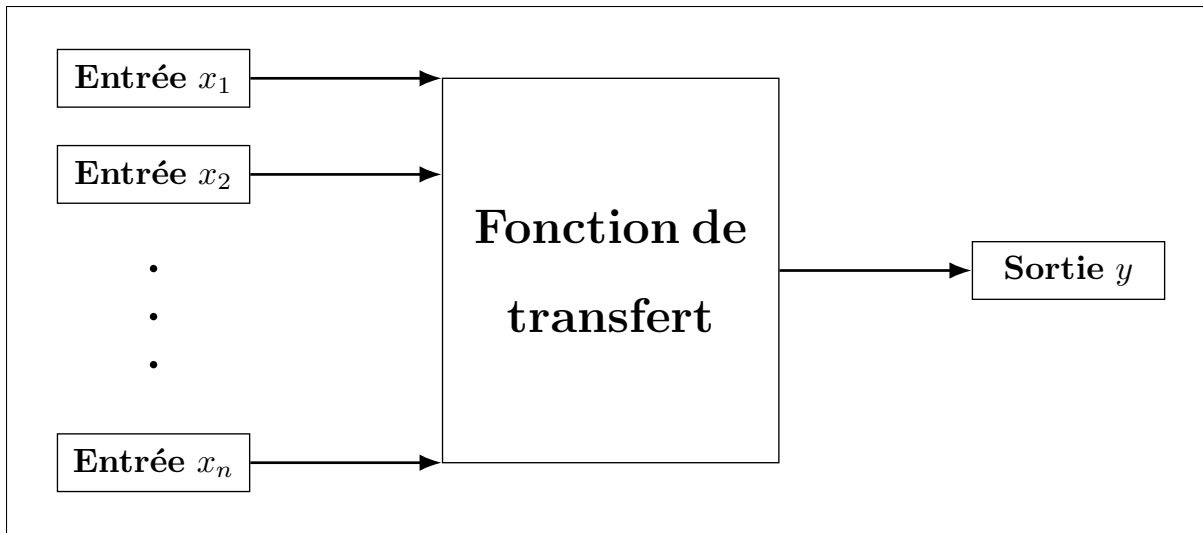


Figure 1.1 Représentation de l'approche de mesure indirecte pour un système MISO

sur une compréhension approfondie du comportement physique de la structure, en particulier sur l'analyse des relations de causalité entre les variables explicatives. Dans l'industrie, le principe des mesures indirectes est largement employé, par exemple pour évaluer la qualité des pièces ou l'usure d'outils à partir de mesures géométriques ou statistiques (Graves & Hamada, 2005; Moricz, Viharos, Németh & Szépligeti, 2018; De Paiva Silva, Bacci da Silva & De Oliveira, 2023). Dans le secteur des énergies renouvelables, ce principe est souvent utilisé sous le nom de capteur virtuel. On distingue alors trois catégories de modèles de mesure indirecte :

- **Les approches basées sur les données :** reposent sur des corrélations statistiques (pas nécessairement linéaires) et des algorithmes d'apprentissage automatique pour prédire des variables difficiles à mesurer. Flexibles et adaptatives, elles sont particulièrement utiles lorsque les processus physiques sont complexes ou mal connus. Toutefois, elles souffrent souvent d'un manque d'explicabilité et de généralisation, et leur performance dépend fortement de la qualité et de la quantité des données disponibles.
- **Les approches basées sur la physique :** s'appuient sur les lois fondamentales qui régissent le système pour en décrire le comportement. Elles offrent une compréhension fine et une

grande précision lorsque la modélisation physique est bien établie, mais deviennent limitées pour les systèmes complexes ou mal compris.

- **Les approches hybrides** : combinent la rigueur théorique des modèles physiques avec la flexibilité des approches fondées sur les données, afin d'améliorer la précision, la robustesse et la capacité d'adaptation des prédictions.

Les principales familles d'approches de mesure indirecte utilisées dans le domaine des énergies renouvelables sont illustrées à la Figure 1.2.

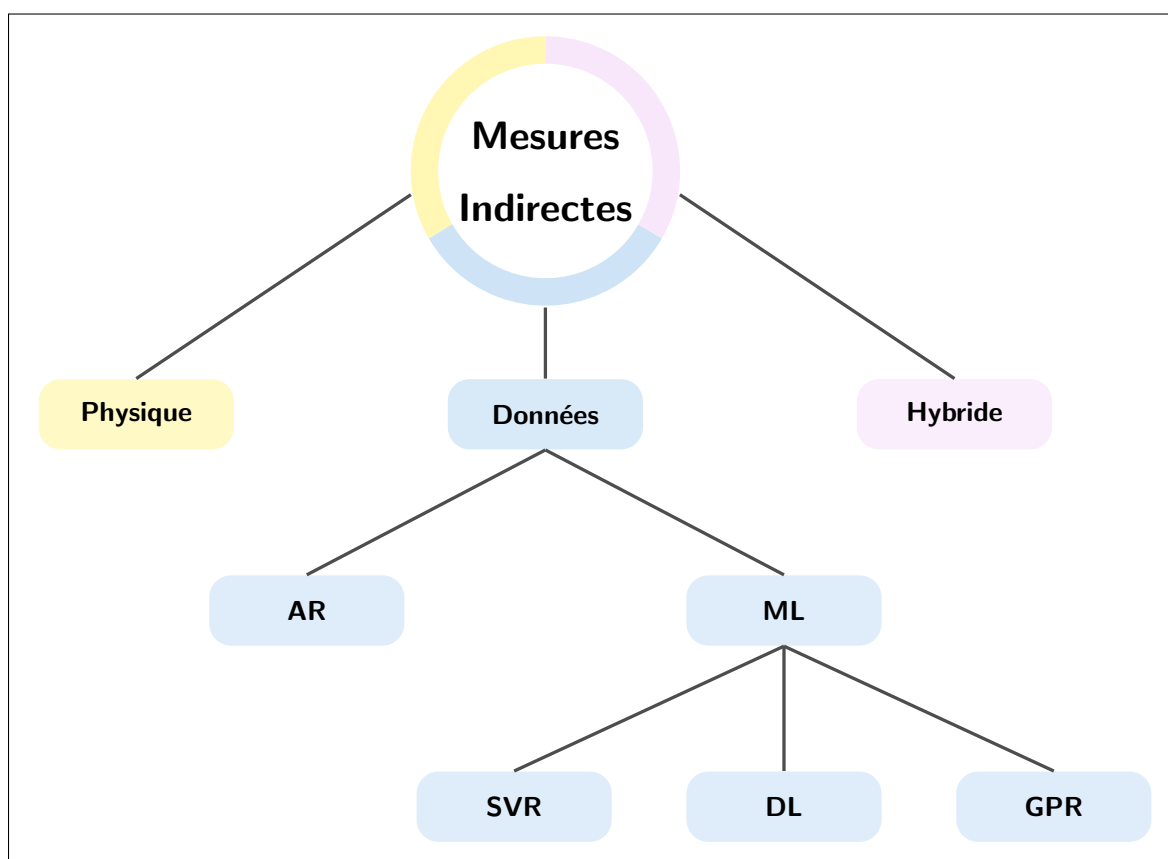


Figure 1.2 Principales approches de mesure indirecte appliquées aux énergies renouvelables

1.1.1 Les approches basées sur les données

Dans le domaine des énergies renouvelables, ces méthodes se déclinent en deux grandes catégories : les modèles autorégressifs (*Autoregressive*, AR) et les modèles d'apprentissage automatique (*Machine Learning*, ML).

1.1.1.1 Modèles autorégressifs

Les modèles AR figurent parmi les premières approches utilisées pour la modélisation dynamique des systèmes énergétiques (Ljung, 1999). Ils ont notamment été appliqués à la mesure indirecte de la déformation d'une roue de turbine hydraulique à partir de mesures issues de l'arbre (torsion, flexion, etc.) (Diagne, Gagnon & Tahan, 2016), ainsi qu'à la prédiction de variables opérationnelles telles que la vitesse du vent, la température des éoliennes et des cellules photovoltaïques (Pasari & Shah, 2020), l'irradiance solaire (Najwadi & Villarreal, 2023), ou encore la puissance produite par les éoliennes (Zhang, Li, Gao, Liu & Ren, 2021). Pour surmonter la linéarité inhérente à ces modèles, des approches hybrides combinant les modèles AR et les réseaux de neurones ont été développées (Gagnon *et al.*, 2022; Massaoudi *et al.*, 2021).

1.1.1.2 Modèles d'apprentissage automatique

Les approches de mesure indirecte fondées sur le ML appliquées aux énergies renouvelables se répartissent principalement en trois catégories : l'apprentissage profond (*Deep Learning*, DL), la régression à vecteurs de support (*Support Vector Regression*, SVR), et la régression par processus gaussien (*Gaussian Process Regression*, GPR), les autres modèles étant peu ou pas utilisés dans ce contexte.

Les approches DL ont été largement explorées en raison de leur robustesse et de leur capacité à modéliser des comportements fortement non linéaires. Les perceptrons multicouches (*Multi-Layer Perceptron*, MLP) ont été utilisés pour estimer les déformations des roues de turbines hydrauliques à partir des vibrations et du déplacement de l'arbre (Presas *et al.*, 2021a,b), ainsi que pour prédire la puissance photovoltaïque produite (Yona *et al.*, 2008; Mellit & Pavan, 2010;

Ding, Wang & Bi, 2011; De Giorgi, Congedo & Malvoni, 2014; Pedro & Coimbra, 2012). Les réseaux de neurones récurrents (*Recurrent Neural Networks*, RNN), plus particulièrement les réseaux de neurones à mémoire longue à court terme (*Long Short-Term Memory*, LSTM), se sont imposés pour les mesures indirectes lorsque la dépendance temporelle est pertinente. Les LSTM ont été employés pour prédire la puissance photovoltaïque (Wang *et al.*, 2020b; Konstantinou, Peratikou & Charalambides, 2021; Gao, Li, Hong & Long, 2019), la puissance éolienne (Han *et al.*, 2019), la vitesse du vent (Balluff, Bendfeld & Krauter, 2015) et les charges dynamiques des éoliennes (Park, You, Oh & Nam, 2022). Les réseaux de neurones autorégressifs non linéaires (*Nonlinear Autoregressive Neural Network*, NARNN), une forme de RNN exploitant la logique des modèles AR, seuls ou combinés à d'autres architectures d'intelligence artificielle, ont également été explorés pour les turbines hydrauliques (Gagnon *et al.*, 2022) afin de prédire la déformation des aubes à partir du déplacement de l'arbre, ainsi que pour la prédiction de la puissance photovoltaïque (Gensler, Henze, Sick & Raabe, 2016; Massaoudi *et al.*, 2021) et de la puissance éolienne (Garcia, Calderon, Avalos, Heras & Tshikala, 2019).

Les SVR se sont révélés efficaces pour la prédiction de la vitesse du vent, ainsi que de la puissance éolienne et photovoltaïque. Leur performance a été améliorée par diverses optimisations algorithmiques, telles que la minimisation séquentielle (Ji, Dong, Qiao & Xu, 2008) ou les colonies de fourmis (Pan *et al.*, 2020a), et confirmée dans plusieurs applications de prédiction pour les actifs des énergies renouvelables (Yamin & Giyats, 2022; Wang, Zhang, Zhang, Li & Zhang, 2022a; Wolff, Kühnert, Lorenz, Kramer & Heinemann, 2016).

Les GPR offrent une approche probabiliste et non paramétrique adaptée à la régression dans des environnements incertains. Ils ont été appliqués à la prédiction des courbes de puissance et des vitesses de vent à partir de données SCADA (Pandit, Infield & Kolios, 2020; Paiva, Nguyen, Lepreux & Bresch-Pietri, 2021), à la modélisation dynamique d'éoliennes *Offshore* (Rajiv, Verma & Subbulakshmi, 2023), ainsi qu'à l'évaluation des charges équivalentes de dommage (Haghi, Stagg & Crawford, 2024).

1.1.2 Les approches basées sur la physique

Les approches de mesure indirecte basées sur les modèles physiques dépendent fortement du cas d'application et du niveau de connaissance du système étudié. Leur précision repose sur la qualité des hypothèses physiques et des paramètres de calibration employés. Dans le domaine photovoltaïque, Dolara, Leva & Manzolini (2015) ont prédit la production électrique à partir de données météorologiques et électriques en comparant plusieurs modèles électriques équivalents comme fonctions de transfert (implicites), montrant que la complexité du modèle n'assure pas toujours une meilleure précision. Pour les éoliennes, Ibrahim *et al.* (2023a) ont développé un jumeau numérique combinant des modèles mécaniques et électriques équivalents pour estimer la vitesse du vent à partir de grandeurs SCADA, démontrant le potentiel des capteurs virtuels fondés sur la physique. Également, pour la surveillance structurelle des éoliennes *Offshore*, plusieurs travaux (Iliopoulos, Devriendt, Guillaume & van Hemelrijck, 2014; Iliopoulos, Weijtjens, Van Hemelrijck & Devriendt, 2015, 2016, 2017; Noppe, Iliopoulos, Weijtjens & Devriendt, 2016) ont utilisé l'analyse et l'expansion modale par éléments finis pour prédire les vibrations et les déformations dans les zones inaccessibles (immergées) à partir d'accéléromètres et de jauges de déformation placés sur la partie accessible (émergée) de la tour d'éolienne, démontrant une excellente capacité de reconstruction du comportement vibratoire global, bien que ces approches demeurent limitées pour les déformations quasi statiques, peu sensibles aux capteurs inertiels.

1.1.3 Les approches hybrides

Les méthodes hybrides combinent la rigueur des modèles physiques et la flexibilité des approches basées sur les données afin d'améliorer la précision et la robustesse des estimations. La majorité d'entre elles repose sur une représentation en espace d'état couplée à une mise à jour bayésienne séquentielle (filtre de Kalman et ses variantes), permettant d'intégrer en continu des données expérimentales dans un modèle physique. Le filtre de Kalman a été appliqué pour l'estimation du couple de l'arbre d'éolienne à partir de sa vitesse angulaire (Perišić, Kirkegaard & Pedersen, 2015), ainsi que pour la déformation des zones immergées des éoliennes à partir des zones émergées accessibles (Maes *et al.*, 2015; Noppe, Tatsis, Chatzi, Devriendt & Weijtjens, 2018).

Les filtres de Kalman ont également été utilisés pour estimer les charges sur les différents composants d'éolienne à partir des données SCADA (Branlard, Jonkman, Dana & Doubrawa, 2020; Mehlan, Nejad & Gao, 2022). D'autres approches hybrides, notamment les réseaux de neurones informés par la physique et les GPR informés par la physique, ont également été appliquées à la mesure indirecte dans le domaine des énergies renouvelables, bien que leur utilisation demeure encore très limitée (Tartakovsky, Ma, Barajas-Solano & Tipireddy, 2023; Bilbao, Lourens, Schulze & Ziegler, 2022; Zou, Cicirello, Iliopoulos & Lourens, 2023; Cobelli, Shukla, Nesmachnow & Draper, 2023; Li & Zhang, 2022).

1.1.4 Discussion

L'analyse des méthodes de mesure indirecte en énergies renouvelables met en évidence la prédominance des approches fondées sur les données, en particulier le DL. Si les MLP sont simples et efficaces, leur incapacité à modéliser les dépendances temporelles limite leur performance pour des grandeurs à dynamique marquée, ce qui justifie l'usage des modèles AR pour les systèmes linéaires et des RNN pour les séries temporelles présentant une dépendance temporelle. Toutefois, l'efficacité des RNN dépend fortement de la nature des données : la modélisation de longues dépendances peut introduire du bruit dans des signaux dominés par les hautes fréquences, tandis que leur omission dégrade la précision pour des dynamiques lentes. Les SVR, bien que robustes vis-à-vis des minima locaux, présentent des performances inférieures à celles des modèles DL. Les GPR suscitent un intérêt croissant en raison de leur nature probabiliste, de leur résistance au surapprentissage et de leur capacité de quantification de l'incertitude, mais cette dernière est souvent mal calibrée (comme expliqué dans la Section 1.3); de plus, ces modèles dépendent fortement du choix du noyau et s'accompagnent d'un coût computationnel élevé. Des variantes telles que les *Sparse GPR* réduisent la complexité au détriment de la précision, tandis que les *Neural Network GPR* visent à l'améliorer, sans avoir encore été explorées dans le domaine des énergies renouvelables. Une limite commune aux modèles ML réside dans leur caractère statique : en présence d'une dégradation ou d'un

changement de régime (*Distribution Shift*), les modèles deviennent rapidement obsolètes, compromettant la fiabilité des prédictions.

Les approches fondées sur la physique, bien que fondamentales, occupent aujourd’hui une place plus limitée en mesure indirecte, en raison d’un compromis défavorable entre complexité et précision : les hypothèses simplificatrices dégradent la précision, tandis que les modèles complexes sont coûteux à mettre en œuvre, bien que l’exploitation de modèles *Surrogate* pour les modèles les plus lourds demeure une option viable. Ainsi, les modèles fondés sur la physique sont souvent incomplets ou impraticables pour les systèmes complexes modernes, lorsque plusieurs mécanismes physiques sont mal caractérisés (Zio, 2022). L’analyse modale appliquée aux éoliennes constitue un exemple notable de succès, sans toutefois répondre à l’ensemble des enjeux actuels, et tend à être supplantée par des approches hybrides (notamment fondées sur le filtre de Kalman) et par des méthodes d’intelligence artificielle, plus adaptatives et généralisables.

Un intérêt croissant se manifeste ainsi pour les approches hybrides combinant modèles fondés sur les données et connaissances physiques, principalement via des mises à jour bayésiennes séquentielles ou l’intégration de modèles d’espace d’état avec des réseaux neuronaux. Des méthodes telles que les *Physics-Informed Gaussian Processes* et les *Physics-Informed Neural Networks* permettent d’incorporer explicitement les lois physiques, mais demeurent encore peu exploitées en mesure indirecte pour les énergies renouvelables et présentent une capacité de généralisation qui reste à démontrer.

Dans ce contexte, où les modèles physiques sont souvent indisponibles ou difficilement exploitables, les approches fondées sur les données, et en particulier le DL, constituent la solution la plus pragmatique, voire la seule applicable. Cependant, la diversité des architectures DL conduit fréquemment à l’adoption de modèles très complexes (ex. *Transformers*), performants dans des contextes spécifiques mais peu généralisables. Le manque de généralisabilité et de transférabilité constitue ainsi une limite majeure des approches de mesure indirecte, qu’elles soient fondées sur les données, sur la physique ou hybrides, plaidant en faveur d’architectures plus simples, intrinsèquement robustes et facilement généralisables entre actifs et cas d’application,

telles que les MLP et les LSTM, capables de couvrir la majorité des usages dans les principales filières des énergies renouvelables.

Enfin, une fois la mesure indirecte établie, peu de travaux s'intéressent à son exploitation pour le diagnostic, notamment à travers la construction de courbes de dégradation reflétant l'état du système. Cette lacune est renforcée par la quasi-absence de quantification de l'incertitude dans les modèles existants, alors que celle-ci est essentielle pour garantir la fiabilité et la confiance des prédictions en contexte industriel (Zio, 2022). L'intégration systématique de cette dimension constitue ainsi un axe de recherche prioritaire que ce travail vise à adresser.

1.2 Pronostic

Le pronostic constitue la dernière étape du processus de PHM. Il vise à prédire l'évolution future de l'état de santé d'un composant ou d'un système afin d'estimer sa RUL, pour anticiper la défaillance et orienter la planification de la maintenance préventive, comme illustré à la Figure 1.3. Cette étape s'appuie sur les données issues de la surveillance de l'état, qu'elles proviennent de CMS ou de SCADA.

On distingue principalement deux grandes familles d'approches de pronostic. Les **approches basées sur la physique** qui s'appuient sur des modèles analytiques décrivant les mécanismes de dégradation à partir de lois physiques (mécaniques, thermiques, chimiques, etc.), offrant généralement une interprétabilité directe. Leur application reste toutefois limitée par la complexité des phénomènes réels et l'incomplétude des modèles, en particulier pour les systèmes industriels modernes caractérisés par des interactions multiples et non linéaires. Les **approches fondées sur les données** apprennent directement les lois de dégradation à partir d'historiques de mesures afin d'estimer la RUL. Elles se déclinent principalement en approches par similarité, basées sur la comparaison avec des trajectoires de dégradation analogues, et en approches par régression, établissant une relation entre des HI et la RUL. Malgré leurs performances, ces méthodes restent fortement dépendantes de la qualité et de la disponibilité des données, souvent bruitées, incomplètes ou irrégulièrement échantillonnées (Moreira de Oliveira Filho, 2024; Zio, 2022).

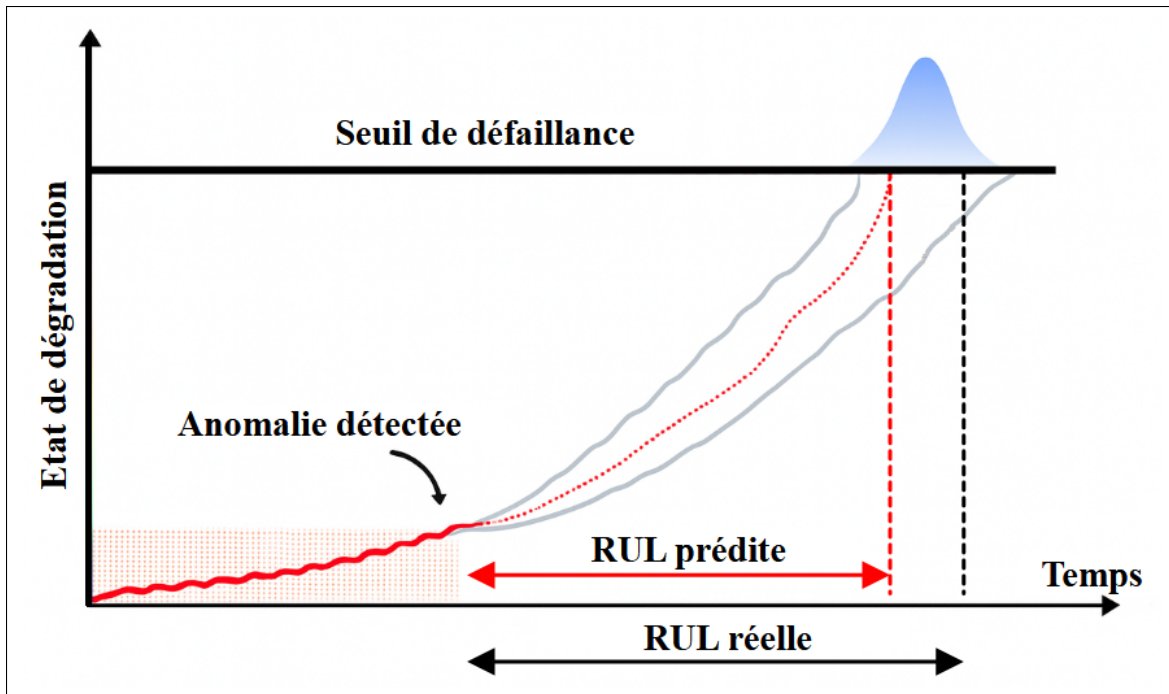


Figure 1.3 Estimation de la RUL
Adaptée de Javed *et al.* (2017, p. 219)

Dans le contexte des actifs en énergies renouvelables, ces difficultés sont amplifiées par la nature des données disponibles. Plusieurs modèles proposés dans la littérature reposent sur des informations difficilement accessibles en exploitation réelle, notamment celles issues des CMS (vibrations, acoustique, lubrification). D'autres, au contraire, s'appuient sur des données simulées ou synthétiques qui ne reproduisent pas fidèlement la variabilité et les imperfections observées sur le terrain (Moreira de Oliveira Filho, 2024). Ces limites compromettent la généralisabilité et, par conséquent, la validité des modèles de pronostic lorsqu'ils sont appliqués à des parcs éoliens ou hydroélectriques de grande échelle.

En somme, les approches de pronostic dans le domaine des actifs en énergies renouvelables demeurent très peu courantes. Le pronostic des turbines hydrauliques est présenté dans la Section 1.2.1, tandis que celui des éoliennes est présenté dans la Section 1.2.2. Les principales familles d'approches de pronostic utilisées dans le domaine des énergies renouvelables, notamment pour les turbines hydrauliques et les éoliennes, sont illustrées à la Figure 1.4.

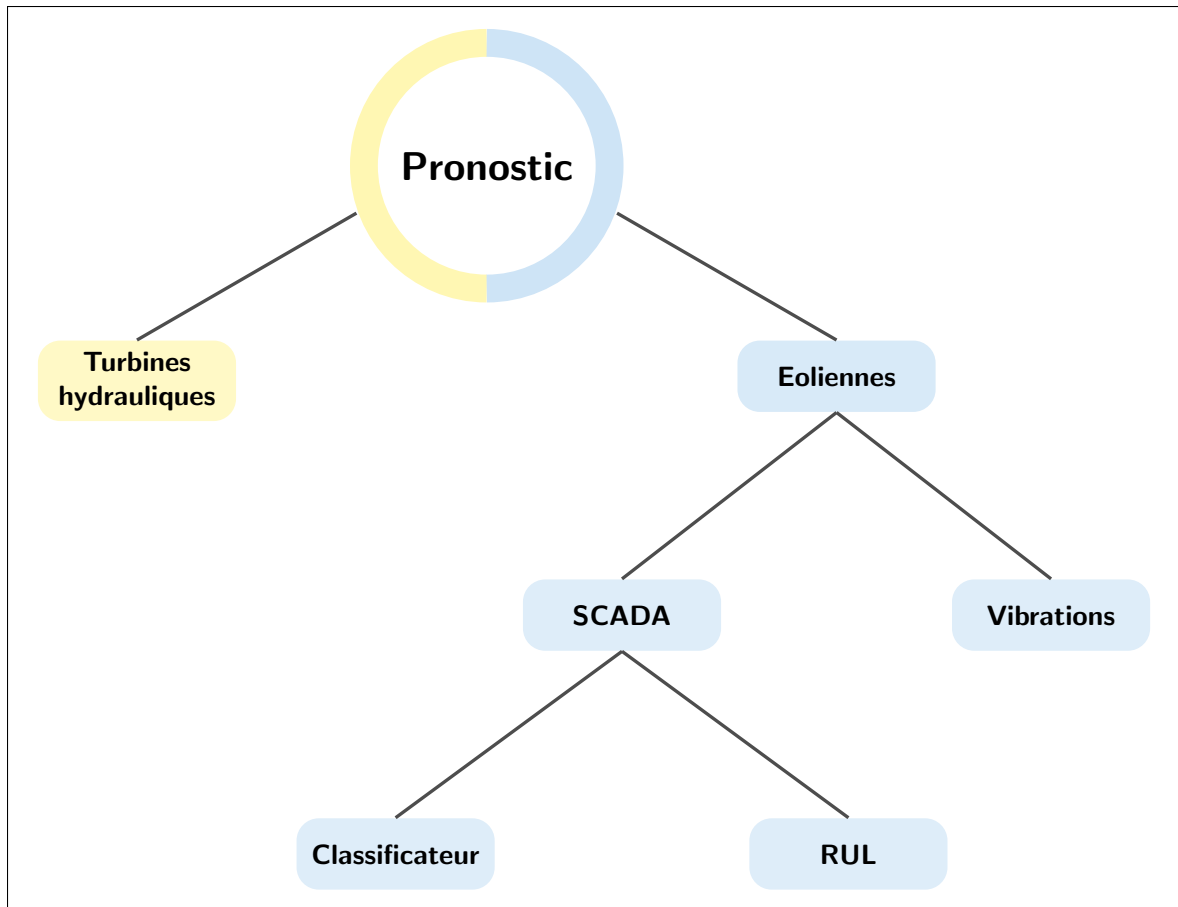


Figure 1.4 Principales approches de pronostic appliquées aux énergies renouvelables

1.2.1 Pronostic des turbines hydrauliques

Dans le secteur hydroélectrique, le pronostic demeure un domaine encore peu exploré. Contrairement à l'éolien, les systèmes hydrauliques sont confrontés à un déficit structurel de données exploitables pour l'apprentissage des modèles. La principale difficulté réside dans l'accès à des jeux de données de défaillance suffisamment riches et représentatifs. En pratique, les défaillances complètes (*Run-to-Failure*) sont extrêmement rares, car les équipements sont soumis à des politiques de maintenance préventive strictes visant à éviter toute panne. Ces interventions interrompent prématurément la trajectoire naturelle de vieillissement et empêchent l'observation continue du processus de dégradation. Cette absence de trajectoires complètes

limite fortement la possibilité d'apprendre les relations entre les conditions opérationnelles et la dynamique réelle de dégradation, condition pourtant nécessaire à la construction de modèles de prédiction de la RUL. Les approches fondées sur la modélisation physique ne constituent pas non plus une alternative simple, car elles reposent souvent sur des hypothèses simplificatrices, restent incomplètes face à la complexité des phénomènes multiphysiques, et s'avèrent coûteuses en ressources de calcul lorsqu'elles sont déployées sur de longs horizons temporels. Ces contraintes réduisent leur applicabilité en contexte industriel, où les données de calibration et les moyens de calcul sont limités.

En conséquence, le développement de modèles de pronostic fiables pour les turbines hydroélectriques demeure fortement limité par la rareté des données de défaillance, l'hétérogénéité des conditions d'exploitation et la complexité intrinsèque des mécanismes de dégradation. Ces difficultés sont accentuées par le fait que de nombreuses turbines hydroélectriques présentent des conceptions fortement spécifiques, parfois sous la forme de prototypes uniques, ce qui limite la construction de modèles généralisables à partir des données agrégées du parc. Ces verrous expliquent que les travaux actuels se concentrent majoritairement sur la détection et le diagnostic (Dollon *et al.*, 2023), c'est-à-dire les étapes amont du processus de PHM, au détriment de la prédiction à long terme et de l'estimation de la RUL. À ce jour, et à notre connaissance, aucun travail de pronostic n'a été recensé sur la roue, tandis que seules deux études récentes s'intéressent à l'alternateur, présentées ci-après.

Jose *et al.* (2024) proposent une approche de pronostic sans données *Run-to-Failure* réelles, fondée sur la reconstitution de trajectoires synthétiques de dégradation à partir de fragments issus de machines similaires. Les états de santé sont projetés dans un espace latent bidimensionnel via un autoencodeur variationnel, puis organisés sous forme d'un graphe d'évolution représentant les transitions temporelles entre états. Un *Graph Masked Autoencoder* est ensuite entraîné pour prédire les états futurs et les temps de transition associés. Appliquée à des groupes hydrogénérateurs industriels, la méthode permet de reconstruire des trajectoires complètes malgré des données irrégulières, avec des erreurs temporelles de l'ordre de quelques mois. Dans la continuité, Jose *et al.* (2025) étendent ce cadre à des environnements multimodaux et multilabels,

en modélisant la dégradation comme un graphe dynamique intégrant des données hétérogènes. Les résultats confirment la robustesse de l'approche face à la rareté et à l'incomplétude des données. Ces travaux ouvrent la voie à des modèles de pronostic basés sur l'apprentissage sur graphes, bien que la complexité des modèles et la qualité des données demeurent des freins majeurs à leur déploiement à grande échelle.

1.2.2 Pronostic des éoliennes

Le nombre de travaux consacrés au pronostic des éoliennes est nettement plus important que celui des turbines hydrauliques, en raison de la plus grande disponibilité et de l'accessibilité des données. On distingue principalement deux approches pour la prédiction de la RUL des éoliennes : la première, majoritairement représentée dans la littérature, repose sur les données de vibration et est présentée à la Section 1.2.2.1, lesquelles sont toutefois rarement disponibles dans le cas des éoliennes ; la seconde s'appuie sur les données SCADA, et est présentée à la Section 1.2.2.2.

1.2.2.1 Pronostic des éoliennes à partir des données de vibration

La prédiction de la RUL à partir des signaux de vibration repose généralement sur la construction d'un HI représentant la trajectoire de dégradation, puis sur son extrapolation jusqu'à un seuil de défaillance prédéfini.

Les premières approches reposaient sur l'extraction manuelle de caractéristiques statistiques issues des signaux vibratoires – telles que la valeur quadratique moyenne, le kurtosis ou la variance – afin de construire un (ou plusieurs) HI synthétique, ensuite exploité par des modèles prédictifs comme les RNN (Guo, Li, Jia, Lei & Lin, 2017), les filtres de Kalman (Jin, Que, Sun, Guo & Qiao, 2019; Cui, Wang, Wang & Ma, 2020; Zhang *et al.*, 2023c), les systèmes neuro-flous (Soualhi, Medjaher, Clerc & Razik, 2020), les *Extreme Learning Machines* multivariées (Pan, Meng, Chen, Gao & Shi, 2020b), les *Relevance Vector Machines* (Wang, Lei, Li & Li, 2020a) ou les SVR (Yan, Wang, Wang, Chang & Muhammad, 2020). Bien que la prédiction de la trajectoire

du HI soit conceptuellement simple, la construction de l'indicateur lui-même reste une tâche fastidieuse et dépend fortement du choix des caractéristiques extraites et de l'expertise humaine en *Feature Engineering*. Afin de surmonter cette limitation, des travaux plus récents exploitent directement les signaux vibratoires bruts pour construire des HIs monotones à l'aide de méthodes d'intelligence artificielle. Les réseaux de neurones convolutionnels (*Convolutional Neural Networks*, CNN) ont notamment été utilisés pour apprendre automatiquement des représentations pertinentes des signaux (Guo, Lei, Li, Yan & Li, 2018; Yoo & Baek, 2018; She & Jia, 2019; Ren, Sun, Wang & Zhang, 2018; Li, Zhang & Ding, 2019; Jiang *et al.*, 2022; Cheng *et al.*, 2020). Des autoencodeurs convolutionnels contraints en monotonie ont également été introduits pour générer des HI présentant une évolution strictement décroissante au cours du temps (Chen, Qin, Wang & Zhou, 2021; Milani, Zappalá & Watson, 2025b).

Cependant, il est important de rappeler que les systèmes de mesure de vibrations (ou les CMS en général) sont rarement installés sur les éoliennes, et lorsque c'est le cas, ils ne fonctionnent souvent que durant des campagnes de mesures ponctuelles. En outre, une part significative des modèles de RUL fondés sur les données vibratoires a été validée sur une même base expérimentale de référence, appelée PRONOSTIA (Nectoux *et al.*, 2012), développée pour l'étude de la dégradation accélérée de roulements à billes. Bien que cette base soit largement utilisée dans la communauté PHM, elle ne reflète pas les conditions réelles d'exploitation des éoliennes, où les charges, les régimes de vent et les contraintes mécaniques varient considérablement. Ainsi, le pronostic basé uniquement sur les mesures vibratoires demeure difficilement applicable en pratique dans le contexte des énergies renouvelables.

1.2.2.2 Pronostic des éoliennes à partir des données SCADA

Un nombre limité d'études s'est intéressé au pronostic fondé sur les données SCADA. Le pronostic est déjà une tâche complexe en soi, et le recours à ce type de données accentue encore les difficultés. Dans la littérature, on distingue principalement deux grandes familles de modèles de pronostic basés sur les données SCADA. La première regroupe les classificateurs, qui ne fournissent pas une estimation continue de la RUL, mais évaluent plutôt la probabilité qu'une

défaillance survienne dans un horizon temporel donné (parfois en la traduisant sous forme d'intervalles de durée restants). La seconde regroupe les approches qui construisent un HI, de manière explicite ou implicite, à partir des variables SCADA, puis extrapolent son évolution pour estimer la RUL (comme illustré dans les approches de pronostic fondées sur les données vibratoires).

Concernant les approches par classification, Carroll *et al.* (2019) ont utilisé un classificateur basé sur un réseau de neurones afin d'anticiper les défaillances de multiplicateurs (*Gearbox*) jusqu'à un mois avant leur occurrence à partir de données SCADA. Gomes *et al.* (2024) ont appliqué un *Random Forest* pour classer la RUL en quatre intervalles (de 7 à 60 jours), avec une précision supérieure à 80 %.

Les modèles produisant une estimation continue de la RUL adoptent des stratégies différentes. Rezamand *et al.* (2021) ont combiné des données SCADA et vibratoires dans un cadre probabiliste, où les données SCADA sont utilisées pour regrouper les états de fonctionnement en *Clusters*, tandis que les caractéristiques vibratoires définissent implicitement le HI servant à l'estimation de la RUL. Oliveira-Filho, Zemouri, Pelletier & Tahan (2021) ont employé un auto-encodeur variationnel pour projeter les données SCADA dans un espace latent, dans lequel la dégradation se manifeste sous la forme d'une trajectoire allant des régions saines vers des régions dégradées, et la RUL est alors estimée en extrapolant cette trajectoire jusqu'à l'atteinte du *Cluster* dégradé. Vieira *et al.* (2024) ont dérivé un HI en estimant la différence de température entre le roulement principal et l'air ambiant ; cette différence, la vitesse du vent et la puissance active constituent les entrées d'un modèle de régression pour la prédiction de la RUL. Shah, Tan & Kumar (2024) ont exploité des données SCADA multivariées et proposé un modèle hybride CNN-RNN intégrant un mécanisme d'attention pour prédire la RUL des éoliennes. Milani, Zappalá, Castellani & Watson (2025a) ont construit un HI à partir des données SCADA à l'aide d'un auto-encodeur convolutionnel, avec une contrainte de monotonie, pour prédire la RUL en extrapolant le HI au moyen d'un modèle quadratique jusqu'à un seuil de défaillance connu. Enfin, Hu *et al.* (2025) ont développé un modèle de dégradation fondé sur une architecture RNN avec mécanisme d'attention, qui décompose les séries temporelles SCADA en plusieurs échelles

temporelles afin de capturer la dynamique de dégradation, et utilise ces représentations pour prédire la RUL avec quantification de l'incertitude.

1.2.3 Discussion

Dans l'ensemble, ces travaux démontrent que, malgré les contraintes inhérentes aux données SCADA – faible fréquence d'échantillonnage, bruit, et agrégation temporelle –, il est possible d'en extraire des informations de santé pertinentes pour le pronostic. Ces approches marquent ainsi une évolution importante vers des méthodes de surveillance continue exploitant les données déjà disponibles sur les parcs éoliens industriels, sans nécessiter de capteurs supplémentaires. Toutefois, la prédiction de la RUL à partir des données SCADA présente encore des limitations critiques :

- **Visibilité tardive de la dégradation et horizons de pronostic limités :** les signaux SCADA ne permettent généralement de détecter les anomalies qu'à des stades relativement avancés (contrairement aux données vibratoires), ce qui rend le pronostic précoce difficile (Hu *et al.*, 2025). De nombreuses méthodologies ne prédisent les défaillances que peu de temps avant leur survenue. Dans certains cas, l'horizon de pronostic n'excède pas 2 semaines (Shah *et al.*, 2024) et peut être aussi court que 2 jours (Hu *et al.*, 2025).
- **Défis liés à la construction des HI :** les systèmes SCADA ne mesurent pas directement la dégradation, mais enregistrent des variables opérationnelles influencées par des effets saisonniers, des variations de charge et des fluctuations stochastiques. L'extraction d'un HI exploitable, monotone et représentatif de la dégradation est donc délicate (Hu *et al.*, 2025), car il doit refléter le caractère irréversible de la dégradation (Javed *et al.*, 2017). Par exemple, les températures de composants peuvent diminuer temporairement en raison des conditions ambiantes, suggérant à tort une auto-réparation alors que la dégradation se poursuit. Plus fondamentalement, tous les mécanismes de dégradation ne sont pas observables à la résolution SCADA : certains processus restent cachés et apparaissent ainsi comme des défaillances soudaines au niveau SCADA, ce qui limite la prédictibilité. Ce constat est également valable pour les mesures indirectes, pour lesquelles il a déjà été établi dans l'introduction que les

données SCADA ne peuvent pas se substituer aux systèmes CMS pour tous les types de mécanismes de dégradation.

- **Validation méthodologique limitée** : la plupart des études reposent sur un nombre restreint d'éoliennes, généralement entre 1 et 7 cas (Gomes *et al.*, 2024; Hu *et al.*, 2025; Milani *et al.*, 2025a; Rezamand *et al.*, 2021; Oliveira-Filho *et al.*, 2021; Vieira *et al.*, 2024; Shah *et al.*, 2024), principalement en raison de l'accès limité aux données SCADA pour des raisons de confidentialité industrielle, du manque d'historiques de défaillances fiables et annotés, ainsi que de la forte hétérogénéité des données entre turbines, fabricants et sites d'exploitation, ce qui limite la généralisation et le déploiement à grande échelle.
- **Quantification de l'incertitude insuffisante** : le pronostic introduit un niveau d'incertitude accru, lié à la variabilité des conditions d'exploitation, aux profils de charge fluctuants, à la nature stochastique des mécanismes de dégradation, ainsi qu'à l'accumulation des erreurs au cours du temps (Zio, 2022). Ces incertitudes rendent toute prédiction strictement déterministe inexacte et exigent des modèles capables de représenter la variabilité et de propager l'incertitude tout au long du processus de prédiction, une limitation déjà soulignée pour les méthodes de mesure indirecte.
- **Architectures complexes et faible généralisation** : plusieurs méthodes reposent sur des architectures profondes lourdes, difficiles à implémenter dans des environnements industriels (Shah *et al.*, 2024; Hu *et al.*, 2025). Le même constat a été formulé dans ce chapitre concernant les architectures de mesure indirecte.

Ces limites motivent le développement de cadres de prédiction de la RUL fiables, scalables et entièrement fondés sur les données SCADA, intégrant une quantification rigoureuse de l'incertitude et validés sur un grand nombre d'actifs.

1.3 Quantification d'incertitude

Les modèles prédictifs en PHM, qu'il s'agisse de mesures indirectes ou de pronostic, fournissent des prédictions ponctuelles déterministes, alors qu'une prise de décision efficace requiert de

manière cruciale la quantification de l'incertitude associée à ces prédictions. En l'absence d'estimations fiables de l'incertitude, les résultats prédictifs ne peuvent être traduits de manière sûre en outils d'aide à la décision pour la maintenance ou en stratégies d'exploitation intégrant le risque (Zio, 2022). Le manque de quantification de l'incertitude a ainsi été identifié comme une lacune majeure pour les deux axes de cette thèse – le diagnostic à partir de mesures indirectes et le pronostic – ce qui justifie son positionnement comme un axe méthodologique transversal. Cette lacune s'explique principalement par la complexité méthodologique des approches de quantification de l'incertitude, qui requièrent des étapes supplémentaires de modélisation et de calibration, tout en engendrant un surcoût computationnel. Elle tient également au fait que de nombreuses études privilégient encore la performance prédictive ponctuelle au détriment de la fiabilité probabiliste. Enfin, la quantification de l'incertitude étant elle-même souvent mal calibrée en pratique, certains travaux préfèrent ne pas en proposer plutôt que de fournir une information probabiliste peu fiable.

L'incertitude est classiquement décomposée en deux types : l'incertitude épistémique, liée au manque de connaissance du système ou du modèle et susceptible d'être réduite par l'acquisition de données supplémentaires, enrichissant ainsi l'apprentissage du modèle mathématique, et l'incertitude aléatoire, qui reflète la variabilité intrinsèque du système et est considérée comme irréductible (American Society of Mechanical Engineers, 2009). En pratique, ces deux sources sont souvent étroitement imbriquées et difficiles à séparer dans les systèmes industriels complexes, ce qui motive l'usage du concept plus général d'incertitude prédictive (Der Kiureghian & Ditlevsen, 2009). De nombreuses méthodes de quantification de l'incertitude ont ainsi été proposées, chaque approche mettant l'accent sur des sources d'incertitude différentes. Les principales méthodes de quantification de l'incertitude sont illustrées à la Figure 1.5.

Les méthodes ciblant principalement l'incertitude épistémique incluent notamment les réseaux de neurones bayésiens (*Bayesian Neural Networks*, BNN), qui modélisent les poids du réseau comme des variables aléatoires et infèrent leurs distributions a posteriori par inférence variationnelle, approximation de Laplace ou échantillonnage Markov-chain Monte Carlo (He, Jiang, Xiao, Xu & Li, 2025b). Bien que théoriquement bien fondés, les BNN souffrent en pratique d'hypothèses

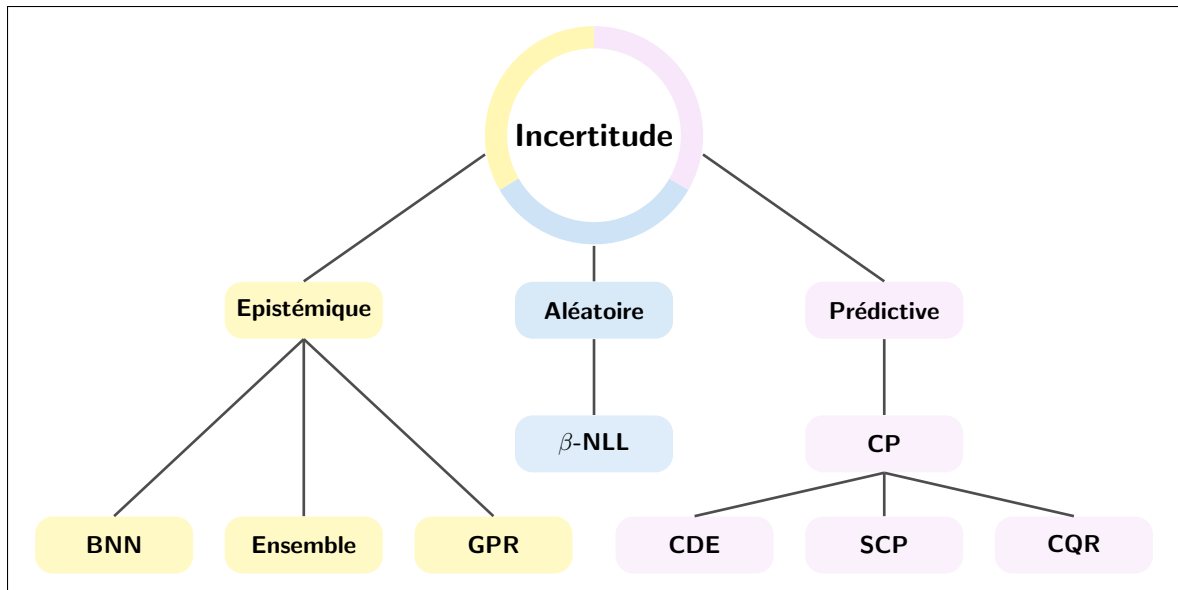


Figure 1.5 Principales méthodes de quantification de l'incertitude

paramétriques fortes (sensibilité à l'a priori), d'un coût de calcul élevé et de difficultés de calibration, ce qui limite leur robustesse en conditions opérationnelles PHM (Peng, Ye & Chen, 2020; Benker, Furtner, Semm & Zaeh, 2021; Rivas, Delipei & Hou, 2022; Basora, Viens, Arias Chao & Olive, 2025). Les méthodes d'ensembles constituent une alternative en exploitant la dispersion de prédictions issues de modèles entraînés avec différentes initialisations ou hyperparamètres (Lakshminarayanan, Pritzel & Blundell, 2017), mais restent coûteuses en calcul et nécessitent souvent d'être combinées à d'autres approches pour garantir une couverture adéquate (Basora *et al.*, 2025; Gagnon, Vouligny, Cauchon, Pham & Mai, 2024). Enfin, les GPR modélisent explicitement l'incertitude via une distribution prédictive, ce qui explique leur intérêt croissant pour les mesures indirectes en énergies renouvelables. Toutefois, ces modèles demeurent fortement sensibles au choix du noyau, qui correspond à la fonction de covariance du GP et détermine la similarité entre les entrées ainsi que la forme des prédictions. Ils s'accompagnent en outre d'un coût computationnel élevé et peuvent conduire à des intervalles de prédiction mal calibrés en conditions réelles (Leibfried, Dutordoir, John & Durrande, 2022). En particulier, la variance prédictive peut rester cohérente (notamment pour refléter une incertitude de nature épistémique), tout en produisant des intervalles de confiance peu fiables lorsque les

hypothèses usuelles du GPR (bruit additif gaussien, stationnarité) sont violées. Dans ce cas, les résidus observés peuvent être non-gaussiens (par exemple leptokurtiques ou platikurtiques), ce qui empêche d’interpréter directement des intervalles « traditionnels » comme des intervalles de couverture nominale : des résidus leptokurtiques tendent à générer un excès d’*Outliers* et donc des faux positifs, tandis que des résidus platikurtiques peuvent conduire à des faux négatifs par sous-estimation des événements extrêmes (Rasmussen & Williams, 2006; Gneiting & Raftery, 2007).

L’incertitude aléatoire est le plus souvent modélisée à l’aide d’approches hétéroscédastiques, qui estiment conjointement la moyenne et la variance prédictives (Seitzer, Tavakoli, Antic & Martius, 2022), généralement sous des hypothèses gaussiennes. Bien que simples et peu coûteuses en calcul (Benabdesselam, Pham, Gagnon & Tahan, 2025; Kahwati, Cauchon, Pham, Vouligny & Gagnon, 2025), ces approches paramétriques présentent souvent une couverture empirique insuffisante et reposent sur des hypothèses distributionnelles restrictives.

Des approches plus avancées visent à quantifier l’incertitude prédictive en tenant compte conjointement des composantes épistémique et aléatoire. Parmi elles, les méthodes conformes (*Conformal Prediction*, CP) (Vovk, Gammerman & Shafer, 2005) offrent un cadre sans hypothèse distributionnelle, avec des garanties de couverture en échantillon fini sous la seule hypothèse d’échangeabilité. Applicables en surcouche de tout modèle prédictif, elles ne nécessitent pas la spécification d’une loi paramétrique, ce qui les rend particulièrement attractives pour les approches PHM fondées sur les données. Des variantes telles que la CP par partitionnement (*Split Conformal Prediction*, SCP) (Vovk *et al.*, 2005; Angelopoulos & Bates, 2022) sont largement utilisées, tandis que des extensions récentes améliorent la calibration, notamment CP avec méthode des ensembles, dite méthode des ensembles conformalisés (*Conformalized Deep Ensembles*, CDE) (Gagnon *et al.*, 2024), la CP combinée à la théorie de l’évidence (Hoblos, Dridi, Zerhouni & Al Masry, 2025) ou encore la CP avec une régression quantile (Piao, Huang & Tsung, 2025).

Enfin, la quantification de l'incertitude n'est pertinente que si les intervalles de confiance sont correctement calibrés. En pratique, de nombreuses approches reposent sur des hypothèses simplificatrices ou des approximations propres aux modèles, sans garantie de validité statistique. La bonne calibration constitue donc une condition essentielle à l'exploitabilité de l'incertitude, en ce sens que les intervalles de confiance doivent atteindre leur niveau de couverture nominal.

1.4 Bilan de la littérature, lacunes et opportunités

Cette revue de littérature met en évidence les avancées significatives réalisées au cours des dernières années dans le domaine du PHM appliqué aux actifs des énergies renouvelables. Les données issues des systèmes SCADA, initialement conçues pour le suivi des performances, ont progressivement été reconnues comme une source d'information exploitable pour développer des approches de PHM à grande échelle et en continu, permettant, lorsque cela est possible, de remplacer les CMS sans instrumentation additionnelle, en s'appuyant sur des approches de mesure indirecte. Compte tenu de la complexité des phénomènes de dégradation, ces mesures indirectes doivent être fondées sur des approches basées sur les données (DL en l'occurrence), reposer sur des architectures simples afin d'assurer leur généralisation et leur transférabilité entre composants et actifs, et permettre la réalisation du diagnostic à partir des prédictions obtenues.

Le pronostic, qui suit naturellement le diagnostic, constitue une tâche particulièrement exigeante, rendue encore plus complexe dans le contexte des données SCADA, car certains mécanismes de dégradation ne sont pas observables à cette résolution. De plus, les approches actuelles sont souvent validées sur un nombre restreint d'actifs, présentent des horizons de prédiction limités, ou reposent sur des architectures complexes, difficilement transférables et peu adaptées à un déploiement industriel à grande échelle.

Un constat transversal concerne la quasi-absence d'une quantification rigoureuse de l'incertitude, tant pour les mesures indirectes que pour le pronostic.

Ces lacunes ouvrent plusieurs opportunités de recherche, qui structurent directement les contributions de cette thèse. Il apparaît nécessaire de développer des cadres de PHM fondés

sur les données SCADA, capables de remplacer, pour certains mécanismes de dégradation observables, les systèmes CMS traditionnels. Ces cadres doivent permettre : (i) la réalisation de mesures indirectes robustes et transférables, (ii) leur exploitation pour le diagnostic à travers des indicateurs ou des trajectoires de dégradation interprétables, (iii) leur extension vers le pronostic, malgré les limitations inhérentes aux données SCADA, et (iv) l'intégration systématique d'une quantification de l'incertitude statistiquement valide et correctement calibrée.

Dans ce contexte, cette thèse s'inscrit à l'interface entre la mesure indirecte, le PHM fondé sur les données SCADA et la quantification de l'incertitude, avec pour ambition de proposer des approches méthodologiquement rigoureuses, applicables en conditions industrielles réelles et validées à l'échelle de flottes d'actifs.

CHAPITRE 2

MÉTHODOLOGIE DE RECHERCHE

2.1 Objectif général et cadre de la recherche

L'objectif général de cette thèse est de développer un cadre de PHM fondé sur les données SCADA permettant, lorsque cela est possible, de remplacer les systèmes CMS par des approches de mesure indirecte afin d'assurer une surveillance continue à grande échelle. Plus précisément, la thèse vise à concevoir des méthodes basées sur les données – privilégiant des architectures simples, transférables et robustes – capables (i) d'estimer, par mesures indirectes, des grandeurs latentes représentatives de l'état de santé, (ii) d'exploiter ces mesures indirectes pour construire des indicateurs et des trajectoires de dégradation interprétables pour le diagnostic, (iii) de réaliser le pronostic malgré les limitations intrinsèques des données SCADA, et (iv) d'associer à chaque prédiction une quantification d'incertitude statistiquement valide et correctement calibrée afin de soutenir la prise de décision en contexte industriel. Ces objectifs sont résumés à la Figure 2.1.

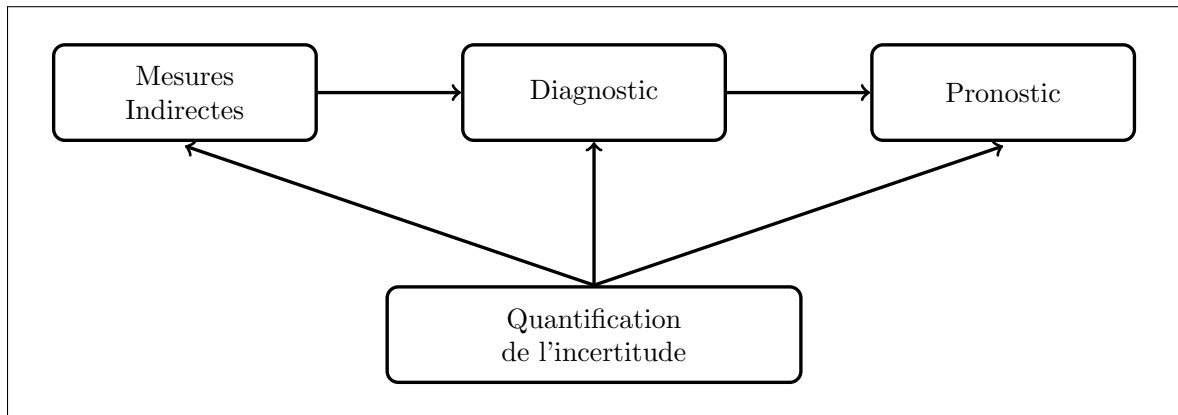


Figure 2.1 Synthèse des objectifs de la thèse

2.2 Questions de recherche

Les principales questions découlant de l'objectif général et guidant la recherche doctorale sont les suivantes :

- **QR1** : Dans quelle mesure les données SCADA peuvent-elles être exploitées pour estimer, par mesure indirecte, des grandeurs latentes représentatives de l'état de santé des composants, sans recourir à des systèmes CMS, tout en garantissant robustesse et transférabilité ?
- **QR2** : Comment exploiter les mesures indirectes issues des données SCADA pour construire des indicateurs et trajectoires de dégradation interprétables, capables de refléter l'évolution irréversible de l'état de santé et de soutenir efficacement les tâches de diagnostic en conditions industrielles réelles ?
- **QR3** : Dans quelles conditions et avec quelles limites les données SCADA peuvent-elles être exploitées pour le pronostic malgré la faible résolution temporelle (données moyennées sur 10 min), le bruit et la visibilité partielle des mécanismes de dégradation ?
- **QR4** : Quelles méthodes de quantification de l'incertitude sont les plus adaptées aux cadres *SCADA-based* de mesure indirecte et de pronostic, et comment garantir des intervalles de prédiction statistiquement valides, correctement calibrés et exploitables pour la prise de décision en PHM ?

2.3 Sélection des cas d'application

La diversité des mécanismes de dégradation rencontrés dans le secteur des énergies renouvelables rend le champ de validation des approches de PHM particulièrement vaste. Dans ce contexte, les données disponibles ne peuvent raisonnablement couvrir l'ensemble des cas d'application possibles. La présente recherche doctorale est ainsi structurée autour de deux axes complémentaires : (i) le remplacement, lorsque possible, des systèmes CMS par des données SCADA à l'aide de mesures indirectes pour le diagnostic, et (ii) le développement d'approches de pronostic fondées sur les données SCADA.

Concernant le premier axe, le remplacement de CMS par des données SCADA ne peut pas être étudié dans le cas des éoliennes, car la grande majorité d'entre elles n'est pas équipée de CMS. La recherche se concentre donc sur les turbines hydrauliques, pour lesquelles des CMS peuvent être déployés afin de surveiller certains phénomènes de dégradation. Parmi les principaux modes de dégradation figurent la fatigue et la cavitation. Si la mesure indirecte de la dégradation liée à

la fatigue dans les zones critiques a déjà été étudiée (Gagnon *et al.*, 2022; Diagne *et al.*, 2016), la cavitation érosive au niveau de la roue demeure un cas particulièrement complexe et encore peu exploré du point de vue du PHM. La majorité des travaux consacrés à la cavitation s’inscrivent dans une logique de prévention ou d’évitement, à travers l’étude de la résistance des matériaux et le développement de revêtements protecteurs (Hofmann *et al.*, 2023; Fivel, Franc & Roy, 2015; Steller & Safonov, 2024; Steller, 2020; da Silva, Ferraresi & Arencibia, 2017), ou l’identification des zones sensibles (Bark, Grekula, Bensow & Berchiche, 2009; Bilus, Hocevar, Dular & Lesnik, 2020; Mihatsch, Schmidt & Adams, 2015; Krumenacker, Fortes-Patella & Archer, 2014; Peters, Sagar, Lantermann & el Moctar, 2015; Wang, Cheng & Ji, 2022c). Or, en pratique, bien que les turbines hydrauliques soient conçues pour fonctionner sans cavitation, leur rôle croissant dans la compensation de l’intermittence des autres sources d’énergie renouvelable conduit fréquemment à des conditions d’exploitation hors régimes nominaux, favorisant l’apparition de ce phénomène (Chatenet, 2020). Dans ce contexte, accepter la cavitation et en assurer le suivi par des approches de PHM devient une nécessité opérationnelle. À ce jour, très peu de travaux ont abordé la surveillance de la cavitation en quantifiant directement la dégradation induite (Chatenet, 2020; Clermont-Bodson, 2015). Les approches existantes sont soit validées à l’échelle du laboratoire sur des horizons temporels limités, soit fondées sur des méthodes statistiques difficiles à déployer à grande échelle. En pratique, la mesure de la cavitation repose principalement sur des CMS vibratoires ou acoustiques, rarement disponibles, parfois mal calibrés et coûteux à maintenir (Clermont-Bodson, 2015; Fernández-Osete, Bermejo, Ayneto-Gubert & Escalar, 2024). Dans ce contexte, l’intérêt industriel – notamment celui d’Hydro-Québec (notre partenaire industriel) – est d’évaluer la possibilité de remplacer ces CMS par des approches fondées exclusivement sur les données SCADA afin de surveiller la cavitation érosive. Ce cas d’application, reconnu pour sa complexité, est ainsi retenu dans cette thèse comme un cas représentatif et générique pour étudier la faisabilité du diagnostic à partir de mesures indirectes.

Concernant le second objectif, relatif au pronostic, et comme établi dans la revue de littérature, le pronostic appliqué aux turbines hydrauliques ne peut être traité à ce stade en raison des limitations identifiées. Cette contrainte oriente naturellement la recherche vers les éoliennes,

pour lesquelles le pronostic constitue un enjeu opérationnel majeur, bien que les approches fondées sur les données SCADA présentent encore d'importantes lacunes. Le cas d'application retenu, motivé notamment par les besoins de notre partenaire industriel Power Factors, concerne des éoliennes commerciales modernes équipées de systèmes SCADA et de dispositifs d'alarme internes déclenchant un arrêt automatique lorsqu'un seuil de température critique, défini par le constructeur ou ajusté par l'opérateur, est atteint. La surchauffe constitue en effet une manifestation commune à plusieurs mécanismes de dégradation, ce qui permet de développer une approche générique couvrant différents phénomènes, plutôt que des modèles spécifiques à chaque mode de défaillance. Le dépassement du seuil thermique entraîne une indisponibilité immédiate de l'éolienne due à un arrêt soudain et non planifié. Dans ce cadre, la dégradation est ainsi définie comme l'indisponibilité de l'éolienne consécutive à un événement de dépassement thermique. Dès lors, lorsqu'une anomalie thermique est détectée à l'aide d'approches SCADA telles que celles proposées dans Cambron *et al.* (2017), l'objectif devient naturellement d'en assurer le pronostic.

2.4 Base de données

La recherche doctorale s'appuie sur deux ensembles de données industriels correspondant aux deux cas d'application présentés à la Section 2.3 : (i) le diagnostic à partir de mesures indirectes appliqué aux turbines hydrauliques, et (ii) le pronostic des éoliennes.

Les données du premier cas d'application ont été fournies par le partenaire industriel Hydro-Québec. Elles concernent deux turbines hydrauliques de type Francis à moyenne chute, exploitées entre 2010 et la fin de l'année 2014, et équipées de systèmes SCADA ainsi que de CMS vibratoires. Ces derniers fournissent le taux instantané de cavitation érosive (exprimé en kg/10000 h), utilisé comme variable cible. Ce taux est modélisé à partir de variables opérationnelles issues du SCADA, telles que le courant, le débit, les hauteurs amont et aval, l'ouverture des directrices, les puissances et la tension.

Le second ensemble de données regroupe 207 éoliennes industrielles issues de trois fabricants, réparties sur 20 sites. Les données SCADA couvrent la période de septembre 2022 à mars 2024 et incluent la température du composant, la température ambiante, la vitesse du vent et la puissance active. Ces données sont complétées par l'historique de l'année 2021, ainsi que par une température de référence en condition saine, définie comme la moyenne arithmétique du parc, et par le seuil de température propre à chaque éolienne tel que fixé par l'opérateur. Trois types de composants sont surveillés : le roulement principal, les roulements du générateur et le générateur. Au total, 207 anomalies thermiques ont été recensées, avec les dates de détection correspondantes, dont 102 surchauffes avérées ayant conduit à un arrêt automatique, pour lesquelles la date d'arrêt est également disponible.

2.5 Hypothèses et portée de la recherche

Les deux turbines hydrauliques étudiées sont équipées à la fois d'un système SCADA et d'un système CMS permettant de mesurer la perte de masse associée à la cavitation érosive. Les données de cavitation exploitées proviennent de mesures validées par les experts d'Hydro-Québec, partenaire industriel du projet. L'hypothèse principale retenue est que ces données sont fiables et représentatives du comportement réel des machines.

Les éoliennes considérées sont équipées de systèmes SCADA, mais ne disposent pas de CMS. En conséquence, aucune donnée vibratoire, d'émissions acoustiques ou d'analyse d'huile n'est disponible. Les modèles développés reposent donc exclusivement sur les données SCADA, excluant les approches fondées sur la modélisation physique, jugées hors de portée dans le cadre de cette recherche. Dans ce contexte, plusieurs hypothèses de travail sont formulées. Tout d'abord, les tendances annuelles des principales variables opérationnelles – température ambiante, vitesse du vent et puissance active – sont supposées relativement stables d'une année à l'autre, hypothèse étayée par des analyses empiriques présentées dans l'article correspondant (Chapitre 4). Par ailleurs, pour la prédiction de la RUL (Chapitre 5), des conditions préalables sont posées afin qu'une dégradation soit considérée comme prévisible à partir des données SCADA. En particulier, la dégradation ne doit pas correspondre à un *Sudden Fault*, et la variable

utilisée pour la construction du HI doit être directement liée au mécanisme de dégradation, en suivant une trajectoire globalement monotone et prévisible, conformément aux recommandations de la littérature, notamment celles de Javed *et al.* (2017).

2.6 Déploiement

Le langage Python a été retenu pour l'implémentation numérique des méthodes proposées, car il constitue le langage de référence pour le développement d'approches d'intelligence artificielle et répond aux normes de nos partenaires industriels. Les projets ont été développés dans l'éditeur Microsoft Visual Studio Code. Plusieurs bibliothèques Python ont été utilisées, notamment NumPy, Pandas, Matplotlib, TensorFlow, Keras et Scikit-Learn. Les architectures et leurs hyperparamètres sont détaillés dans les chapitres correspondants.

2.7 Retombées scientifiques et industrielles

Les contributions scientifiques sont les suivantes :

- Développement d'un cadre de mesure indirecte permettant d'estimer la perte de masse instantanée due à la cavitation érosive à partir des données SCADA, et assurant une surveillance continue en remplacement des CMS. Cette approche permet ensuite de quantifier l'état de dégradation et d'assurer le diagnostic, tout en étant transférable d'une turbine à une autre sans nécessiter de réentraînement, et a été validée sur des données opérationnelles réelles de turbines hydrauliques (article de journal 1).
- Analyse critique des forces et des limitations des méthodes de quantification de l'incertitude dans des contextes PHM réels, en mettant en évidence leurs impacts sur la fiabilité des prédictions et de l'aide à la décision (article de journal 1).
- Pronostic de l'évolution des anomalies thermiques des éoliennes vers un état sain ou un état dégradé (tâche de classification) à partir de données SCADA uniquement, en introduisant des indicateurs d'événement originaux dérivés de la quantification de l'incertitude (article de journal 2).

- Proposition de conditions de prédictibilité sous SCADA et développement d'un indicateur de santé permettant la prédiction de la durée de vie résiduelle des éoliennes à partir de données SCADA uniquement, avec quantification de l'incertitude associée (article de journal 3).

Les retombées industrielles sont les suivantes :

- Réduction des coûts de maintenance et des pannes non planifiées grâce à l'exploitation des données SCADA existantes, sans ajout de capteurs (sans CMS).
- Amélioration de la fiabilité de la surveillance, du diagnostic et du pronostic par les mesures indirectes et par la calibration de l'incertitude.
- Développement de modèles génériques transférables entre sites et technologies.

2.8 Mise en contexte des publications scientifiques

Le Tableau 2.1 recense les articles dont le doctorant est le premier auteur, et qui apportent des éléments de réponse aux questions de recherche motivant ce travail doctoral.

Tableau 2.1 Publications scientifiques issues des travaux de recherche doctorale

Type	Date	Journal / Conférence	Titre	QR associés	Réf
Revue	avril 2024	MDPI <i>Electronics</i>	<i>On the Use of Indirect Measurements in Virtual Sensors for Renewable Energies : A Review</i>	QR1 QR2	Annexe I

Suite à la page suivante

Type	Date	Journal / Conférence	Titre	QR associés	Réf
Conf. 1	juin 2025	PHM-2025	<i>A Multi-Layer Perceptron Approach for Indirect Measurement of Erosive Cavitation in Hydraulic Turbines</i>	QR1 QR2 QR4	Annexe I
Journal 1	avril 2026	Results in Engineering	<i>Diagnostic of Erosive Cavitation in Hydraulic Turbines From Indirect Measurements Using SCADA Data With Uncertainty Quantification</i>	QR1 QR2 QR4	Chapitre 3
Journal 2	décembre 2025 *	IEEE Access	<i>Prognostics of Thermal Anomalies in Wind Turbines via Deep Learning and Conformal Prediction Using SCADA Data</i>	QR3	Chapitre 4
Journal 3	février 2026 *	Reliability Engineering & System Safety	<i>Remaining Useful Life Prediction for Gradual Thermal Anomalies in Wind Turbines Using SCADA Data</i>	QR3	Chapitre 5

* : date de soumission

En plus de ces publications, le doctorant a été coauteur d'un article de journal soumis à *IEEE Transactions on Reliability*, présenté à la Figure I-2 (Annexe I), et a contribué à la présentation orale réalisée par notre partenaire industriel Power Factors à la conférence *Wind Europe Technology Workshop 2025*, illustrée à la Figure I-4 (Annexe I).

CHAPITRE 3

DIAGNOSTIC OF EROSIVE CAVITATION IN HYDRAULIC TURBINES FROM INDIRECT MEASUREMENTS USING SCADA DATA WITH UNCERTAINTY QUANTIFICATION

Abderraouf Benabdesselam¹, Martin Gagnon², Antoine Tahan¹

¹ Département de Génie Mécanique, École de Technologie Supérieure, 1100 Rue Notre-Dame O., Montréal, QC H3C 1K3, Canada ;

² Hydro-Québec, 75 Boulevard René-Lévesque O., Montréal, QC H2Z 1A4, Canada.

Article publié dans « Elsevier – Results in Engineering » le 10 avril 2026.

3.1 Abstract

This paper proposes a data-driven framework for monitoring and diagnosing erosive cavitation in hydraulic turbines using supervisory control and data acquisition data only, with the objective of replacing costly and sparsely deployed condition monitoring systems. A deep-learning-based indirect measurement approach is developed to estimate the instantaneous cavitation rate from operational variables and is validated on multi-year real industrial data from two Francis turbine units, while ensuring model transferability across machines. Beyond point prediction, the framework explicitly addresses uncertainty quantification to support reliable decision-making, and several methods, including heteroscedastic regression, Bayesian neural networks, split conformal prediction, and conformalized deep ensembles, are benchmarked using calibration metrics and evaluated under distribution shifts. Diagnostics is performed by aggregating instantaneous predictions into an uncertainty-bounded cumulative degradation trajectory, enabling an uncertainty-aware assessment of cavitation progression and supporting reliability- and safety-informed maintenance decisions. Results demonstrate that operational data can effectively replace condition monitoring systems for cavitation monitoring and that conformalized deep ensembles provide the most reliable uncertainty estimates under realistic industrial conditions.

3.2 Introduction

The sustainability of renewable energy systems depends on reducing operation and maintenance costs, thereby accelerating the adoption of Prognostics and Health Management (PHM) strategies (Dao *et al.*, 2019). The effectiveness of PHM relies on the availability of continuous and reliable measurements that accurately reflect the true health condition of the monitored asset (Benabdesselam *et al.*, 2024).

In practice, two main monitoring approaches are employed in renewable energy systems : Condition Monitoring (CM) and Performance Monitoring (PM) (Rezamand *et al.*, 2020). CM is typically implemented through dedicated Condition Monitoring Systems (CMS) that collect high-frequency signals such as vibration, acoustic emissions, or lubrication measurements. Because these signals are directly sensitive to mechanical defects and degradation mechanisms, CMS are widely regarded as the most informative data source for PHM. However, CMS remain sparsely deployed in renewable energy assets due to acquisition and installation costs, calibration and inspection requirements, and operational constraints that may involve maintenance interventions and production downtime. Even when CMS are available, measurements are often collected only during dedicated measurement campaigns, leaving the asset health state uncertain between campaigns and preventing continuous PHM deployment (Rezamand *et al.*, 2020; Moreira de Oliveira Filho, 2024; Yan *et al.*, 2025; Maldonado-Correa *et al.*, 2020). By contrast, PM primarily relies on Supervisory Control and Data Acquisition (SCADA) systems, which are natively embedded in all renewable energy assets to supervise operating conditions and comply with grid requirements (Moreira de Oliveira Filho, 2024). SCADA systems record a large number of operational variables, typically aggregated at low frequency (e.g., 10-minute averages) (Rezamand *et al.*, 2020). Although SCADA was not originally designed for health monitoring, numerous studies have shown that degradation processes affecting critical components can induce measurable signatures in SCADA variables (Moreira de Oliveira Filho, 2024; Yan *et al.*, 2025; Maldonado-Correa *et al.*, 2020). Due to its continuous availability, SCADA emerges as an attractive data source for large-scale PHM deployment.

Within hydraulic turbines, erosive cavitation constitutes one of the predominant degradation mechanisms (Chatenet *et al.*, 2020). It originates from the formation and collapse of vapor bubbles in low-pressure regions, generating shock waves and high-energy micro-jets that progressively damage metallic surfaces (Gregg, Steele & Van, 2020; Kim, Chahine, Franc & Karimi, 2014), as illustrated in Figure 3.1. This phenomenon is widely recognized as particularly complex to characterize and model due to its strong dependence on flow conditions, transient operating regimes, and material properties (Chatenet, 2020). Consequently, most of the existing literature focuses on cavitation avoidance through enhanced material resistance, surface treatments, or the identification of cavitation-prone regions and hotspots (Hofmann *et al.*, 2023; Fivel *et al.*, 2015; Steller & Safonov, 2024; Steller, 2020; Li *et al.*, 2024b; da Silva *et al.*, 2017; Bark *et al.*, 2009; Bilus *et al.*, 2020; Mihatsch *et al.*, 2015; Krumenacker *et al.*, 2014; Peters *et al.*, 2015; Wang *et al.*, 2022c; Wang, Li, Lv, Wang & Wang, 2025b). By contrast, very few studies address the monitoring of cavitation progression over time (Clermont-Bodson, 2015; Chatenet *et al.*, 2020; Chatenet, Tahan, Fouladirad, Gagnon & Remy, 2021). Existing works are typically validated at laboratory scale, limited in time duration (Chatenet *et al.*, 2020), or rely on the clustering of operating conditions (Clermont-Bodson, 2015), which limits their applicability to long-term, real-world PHM deployments. Such approaches typically rely on CMS capable of measuring the instantaneous mass loss induced by erosive cavitation, referred to as the Instantaneous Cavitation Rate (ICR) (Bourdon, 2000; Fernández-Osete *et al.*, 2024). While effective, these CMS are costly, not systematically installed on hydraulic turbines, and require complex calibration procedures (Chatenet, 2020; Clermont-Bodson, 2015), as previously discussed for CMS in renewable energy applications. Inspection campaigns can provide direct assessments of cavitation damage; however, they require turbine shutdowns and offer no information on the degradation state between inspections.

In this context, leveraging SCADA data to monitor erosive cavitation and perform PHM without relying on CMS represents a highly promising alternative. Such an approach avoids the costs and deployment constraints associated with CMS while enabling continuous, real-time tracking of degradation through indirect measurements (Benabdesselam *et al.*, 2024).

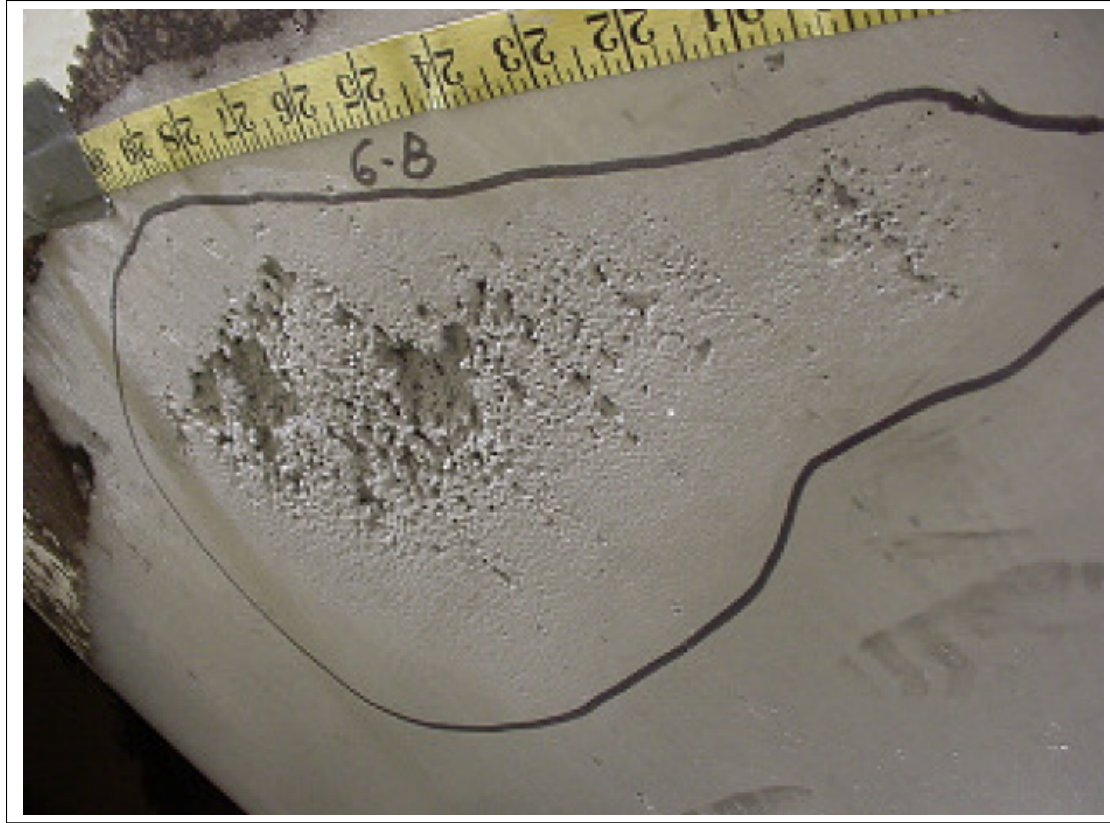


Figure 3.1 Example of cavitation erosion on a blade
Tirée de Clermont-Bodson (2015, p. 8)

3.2.1 Indirect Measurements

In this context, an indirect measurement consists of predicting an unavailable degradation-related variable, here the ICR, denoted as $y \in \mathbb{R}$, from accessible SCADA measurements $\mathbf{x} \in \mathbb{R}^n$ via a transfer function $y = f(\mathbf{x})$.

In renewable energy systems, indirect measurement has been increasingly adopted to estimate latent quantities that are otherwise unavailable or prohibitively costly to measure continuously, and a comprehensive review of such approaches is reported in Benabdesselam *et al.* (2024). A key observation from this literature is that most indirect measurement approaches in renewable energy rely on Artificial Intelligence (AI) algorithms to model the transfer function, due to the absence of reliable physics-based models for many degradation mechanisms, including erosive

cavitation (Chatenet *et al.*, 2020; Clermont-Bodson, 2015). This naturally motivates the use of AI-based models in the present study. A second observation is that many existing indirect measurement approaches employ highly complex model architectures that operate in a static or asset-specific setting. In contrast, achieving accurate predictions with relatively simple AI models is preferable, as it promotes robustness and transferability across turbines of the same type. Accordingly, the indirect measurement model should be, as much as possible, transferable to other units. Finally, it is worth noting that the majority of indirect measurement studies focus on point predictions and do not explicitly quantify uncertainty. This limitation is discussed in Section 3.2.2.

3.2.2 Uncertainty Quantification

While PHM predictive models provide deterministic point predictions, effective decision-making critically requires quantifying the associated predictive uncertainty. Without reliable uncertainty estimates, such predictions cannot be safely translated into decision support for maintenance actions or risk-aware operational strategies (Zio, 2022; Salinas-Camus, Goebel & Eleftheroglou, 2025).

Uncertainty is commonly categorized into two complementary sources : epistemic uncertainty, arising from limited knowledge about the system or the model and reducible with additional data, and aleatoric uncertainty, which reflects intrinsic variability and is considered irreducible (American Society of Mechanical Engineers, 2009). In practice, these two sources are often intertwined in complex industrial systems, motivating the use of the broader notion of predictive uncertainty (Der Kiureghian & Ditlevsen, 2009). A wide range of Uncertainty Quantification (UQ) methods has been proposed, with different approaches emphasizing different sources of uncertainty.

Methods primarily targeting epistemic uncertainty include Bayesian Neural Networks (BNNs), which model network weights as random variables and infer their posterior distributions using approximate inference techniques such as Variational Inference (VI), Laplace approximation, or Monte Carlo sampling (He *et al.*, 2025b). In this work, BNNs were included as a representative

Bayesian deep-learning approach for UQ, since they are widely used to model epistemic uncertainty in neural networks and therefore provide a relevant point of comparison with the other UQ methods considered in this study. This is particularly meaningful in PHM, where limited degradation data, operational variability, and distribution shifts can strongly affect predictive reliability. While BNNs are theoretically well grounded, their practical deployment is often hindered by strong parametric assumptions, high computational cost, and calibration challenges. Their application to PHM has been reported in Peng *et al.* (2020); Benker *et al.* (2021); Rivas *et al.* (2022); Basora *et al.* (2025), where limited robustness in operational settings has been highlighted (Basora *et al.*, 2025). Deep ensemble methods offer an alternative by training multiple models with different initializations, architectures, or hyperparameters and using prediction dispersion to quantify uncertainty. While they have demonstrated strong empirical performance (Lakshminarayanan *et al.*, 2017), they remain computationally expensive and difficult to scale to fleet-level PHM applications (Basora *et al.*, 2025), and often need to be combined with other approaches to ensure adequate coverage (Gagnon *et al.*, 2024). Other approaches, such as Gaussian Processes (GPs), explicitly account for uncertainty induced by data sparsity or distribution shifts and have been used for indirect measurements in renewable energy applications (Haghi *et al.*, 2024; Pandit *et al.*, 2020; Paiva *et al.*, 2021; Rajiv *et al.*, 2023). Despite their solid theoretical foundations and natural resistance to overfitting, GPs often suffer from overestimated uncertainty, high computational cost, and strong sensitivity to kernel choice (Leibfried *et al.*, 2022), which typically relies on expert knowledge and limits their suitability for continuous, large-scale PHM applications.

Aleatoric uncertainty is commonly modeled using heteroscedastic approaches that jointly estimate the predictive mean and variance (Seitzer *et al.*, 2022), typically under Gaussian assumptions. While simple and computationally efficient, these parametric models often exhibit poor empirical coverage and rely on restrictive distributional assumptions. Nevertheless, such approaches have been applied in PHM contexts for renewable energy systems, as reported in Benabdesselam *et al.* (2025); Kahwati *et al.* (2025).

More advanced approaches aim to quantify predictive uncertainty by jointly accounting for epistemic and aleatoric effects. Among them, Conformal Prediction (CP) (Vovk *et al.*, 2005) constitutes a general distribution-free framework for UQ rather than a predictive method. It is applied as a post-hoc calibration layer on top of an underlying predictive model to construct prediction intervals with guaranteed finite-sample coverage under the sole assumption of exchangeability. The most widely used and computationally efficient variant is Split Conformal Prediction (SCP) (Vovk *et al.*, 2005; Angelopoulos & Bates, 2022), which is commonly adopted as a baseline conformal approach. More robust extensions have subsequently been proposed, including Conformalized Deep Ensembles (CDE) (Gagnon *et al.*, 2024), which combine ensemble diversity with conformal calibration to improve the reliability of predictive uncertainty. Other variants have also been introduced, such as CP combined with evidence theory (Hoblos *et al.*, 2025) and CP with quantile regression (Piao *et al.*, 2025). These characteristics make CP-based approaches particularly attractive for data-driven PHM applications, where predictive distributions are often difficult to specify reliably.

UQ is only meaningful if the resulting Confidence Intervals (CIs) are properly calibrated. In practice, UQ often relies on simplifying distributional assumptions or model-specific approximations that do not guarantee statistical validity. Calibration is therefore a fundamental requirement for actionable uncertainty, as CIs must achieve their nominal coverage.

3.2.3 PHM for Erosive Cavitation

Once degradation, represented here by the ICR, becomes measurable through indirect measurements using SCADA data, the next step is to apply the PHM process, which consists of three main stages : anomaly detection, fault diagnosis, and health state prognosis (Zio, 2022). However, although hydraulic turbines are designed to avoid cavitation, they are frequently operated outside optimal conditions, particularly to compensate for the intermittency of other renewable energy sources (Clermont-Bodson, 2015; Chatenet, 2020). As a result, erosive cavitation is continuously present, so the objective is not primarily to detect its presence, since its existence is already established, but rather to assess its severity and monitor its progression over time.

The next step is therefore to construct the degradation trajectory in real time from the indirect measurement of the ICR by cumulatively integrating the instantaneous mass loss induced by cavitation. Once a maintenance threshold is defined, this degradation curve enables diagnosis of the current degradation state and supports maintenance planning based on the actual condition of the turbine. While cumulatively integrating the ICR to obtain a degradation curve is intuitive, propagating and aggregating the uncertainty associated with ICR predictions to derive uncertainty on the degradation state is not straightforward. This challenge is addressed later in this paper.

3.2.4 Key Gaps

Despite extensive research on cavitation phenomena, indirect measurements, PHM, and UQ, several critical gaps remain :

- Most cavitation studies emphasize cavitation avoidance or mitigation strategies rather than the continuous monitoring of degradation progression over time, which is inherently more challenging and remains largely unaddressed.
- Despite the growing interest in performing PHM using SCADA data as a cost-effective alternative to dedicated CMS, there is no evidence that SCADA-based approaches are applicable to all degradation mechanisms or operational contexts, particularly in renewable energy systems. Replacing CMS with SCADA for the monitoring of erosive cavitation therefore represents a major gap in the PHM literature, but also a highly promising opportunity, especially if the proposed models can be transferred across hydraulic turbines through transfer learning.
- More generally – and particularly in the case of erosive cavitation – a significant gap lies in the limited exploitation of indirect measurements for PHM tasks, especially for diagnostic purposes. When indirect measurements are feasible, their potential to support reliable diagnosis should be systematically explored.
- When indirect measurement models are proposed in the literature, they are often complex, and UQ is either omitted or poorly calibrated, which poses significant risks for maintenance decision-making. A systematic comparison of UQ methods on real industrial PHM datasets

is still lacking, particularly with respect to uncertainty calibration metrics. Such an analysis is essential to identify which UQ approaches should be preferred and which should be avoided in real-world PHM applications.

3.2.5 Objectives and Contributions

This paper aims to enable reliable, SCADA-only monitoring and diagnosis of erosive cavitation in hydraulic turbines, extending our preliminary work (Benabdesselam *et al.*, 2025) and addressing a relevant industrial challenge by replacing CMS-based approaches. It further serves as a methodological benchmark for uncertainty-aware, SCADA-based PHM, validated on real operational data using erosive cavitation as a representative case study. This paper makes the following contributions :

1. **SCADA-only indirect measurement of cavitation degradation :** We develop a transferable indirect measurement framework to estimate the ICR from SCADA variables, enabling continuous cavitation monitoring without relying on CMS, and validated on real-world hydraulic turbine operational data.
2. **Rigorous UQ benchmarking for PHM :** We conduct a systematic and application-driven evaluation of widely used UQ methods for SCADA-based PHM (heteroscedastic regression, Bayesian approaches, and conformal methods) using real SCADA data from operating hydraulic turbines. The goal is not to propose a new UQ algorithm, but to provide evidence-based guidance on which approaches remain reliable under realistic PHM conditions involving non-stationarity, distribution shifts, and indirect measurements.
3. **Diagnostics from indirect measurements :** The degradation trajectory is constructed from indirect measurements by cumulatively aggregating the predictions and their associated uncertainties, enabling the assessment of the degradation state and the planning of maintenance actions.

3.2.6 Study Case and Data

The monitoring of mass loss due to erosive cavitation is performed using vibration-based sensors, which estimate ICR in units of kg per 10 000 operating hours (Bourdon, 2000). ICR constitutes the target variable of this study and is obtained from the vibration-based monitoring system described in Bourdon (2000). The input SCADA variables were selected based on the exploratory analysis reported in Clermont-Bodson (2015), and all used data are summarized in Table 3.1.

Tableau 3.1 Variables used in this study

Variable	Symbol	Unit	Description
ICR	ICR	kg/10000h	Instantaneous mass loss due to erosive cavitation
Current intensity	I	A	Electric current generated by the turbine
Flow rate	$\dot{m}$	m ³ /s	Water volume entering the turbine per second
Upstream head	H_u	m	Water level at turbine inlet
Downstream head	H_d	m	Water level at turbine outlet
Guide vane opening	VO	%	Opening percentage of guide vanes
Active power	P_a	MW	Power delivered to the grid
Reactive power	P_r	MVA _r	Power due to inductive and capacitive effects
Effective stability power	P_e	MW	Maximum transient power ignoring alternator constraints
Voltage	U	kV	Electric potential difference across phases

From a physical standpoint, erosive cavitation is governed primarily by local pressure conditions inside the turbine. Accordingly, variables such as flow rate, guide vane opening, upstream level, and downstream level are the most direct predictors, since they affect the discharge, the loading, and the pressure field governing vapor formation and collapse. In addition, active power provides a compact description of the unit operating point, which is important because cavitation intensity depends strongly on departures from optimal operating conditions. In a coupled turbine-generator system, electrical variables reflect the generator operating state associated with a given mechanical and hydraulic regime. Changes in voltage, current, active power, and reactive power therefore help characterize the overall operating condition of the unit, including off-design

loading and transient regimes, which may coincide with efficiency losses or deviations from the optimal head-power operating region (Clermont-Bodson, 2015; Avellan, 2004; Yang, 2015).

Cavitation and SCADA data are available for two Francis-type hydraulic turbines (Unit 1 and Unit 2). For Unit 1, the dataset spans from 2010 to 2014, while for Unit 2 it covers the period from 2011 to 2014. These datasets originate from real industrial operations, were preprocessed and validated by experts from our industrial partner, and have been previously used in related studies (Benabdesselam *et al.*, 2025; Chatenet, 2020; Clermont-Bodson, 2015).

3.2.7 Organization of the paper

The remainder of this paper is organized as follows. Section 3.3 presents the proposed methodology, including the indirect measurement framework and the UQ approaches under consideration. Section 3.4 reports and discusses the experimental results. Finally, Section 3.5 concludes the paper and outlines possible directions for future work.

3.3 Methodology

The methodology consists of four main steps. The first step is data acquisition already described in Section 3.2.6. The second step is the indirect measurement of ICR, as presented in Section 3.3.1, and evaluated using the accuracy metrics introduced in Section 3.3.5. The third step concerns the UQ of the predictions, together with the transfer learning approach described in Sections 3.3.2 and 3.3.3, and assessed using the calibration metrics introduced in Section 3.3.5. The final step is diagnostics, presented in Section 3.3.4. The complete workflow is illustrated in Figure 3.2.

3.3.1 Prediction of ICR

The first objective of this work is to replace CMS with SCADA data for predicting the ICR through an indirect measurement approach. Based on our review study (Benabdesselam *et al.*, 2024), which surveys the architectures most commonly adopted for indirect measurement in PHM applications for renewable energy systems, three neural-network-based models were

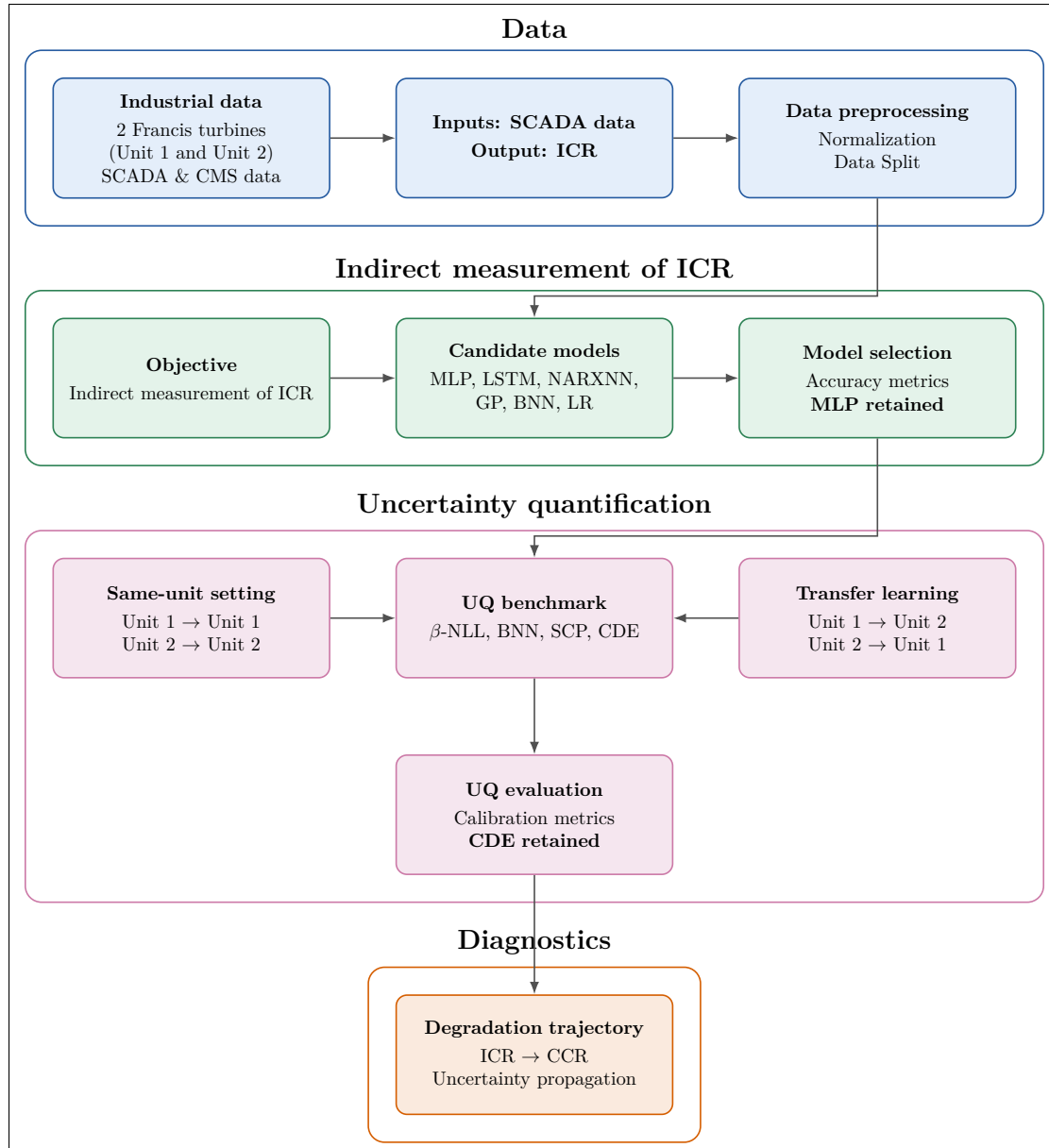


Figure 3.2 Diagnostic of erosive cavitation methodology

selected in this work : Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM), and Nonlinear AutoRegressive Neural Network with eXogenous inputs (NARXNN). These architectures were chosen because they represent widely adopted and complementary approaches for indirect measurement. More specifically, the MLP provides a nonlinear static mapping between SCADA variables and ICR, the LSTM is well suited for capturing temporal dependencies

in sequential data, and the NARXNN explicitly accounts for both exogenous inputs and past outputs, which is relevant for dynamic degradation-related signals. In addition to these models, GP and BNN, introduced in Section 3.2.2, were also considered, as they provide indirect measurement frameworks with embedded UQ.

The first model is an MLP composed of two hidden layers of 64 neurons each, followed by batch normalization and a Rectified Linear Unit (ReLU) activation. The second model is an LSTM network designed to capture temporal dependencies, where the input data are segmented into sequences of $p = 10$ time lags. It consists of a single LSTM layer with 64 hidden units, followed by a fully connected linear output layer. The third model is a NARXNN, in which the current output depends on both current inputs and past outputs. It is implemented as an MLP with an augmented input vector using the same lag length $p = 10$ for consistency with the LSTM. During testing, recursive prediction is employed, where previously predicted outputs are reused as inputs to emulate real-time inference. The MLP, LSTM, and NARXNN are trained for 150 epochs using the Adam optimizer with a learning rate of 0.001, minimizing the Mean Squared Error (MSE) loss.

The fourth model is a GP regression model, which defines a prior over functions and provides a full posterior predictive distribution for UQ (Benabdesselam *et al.*, 2024). The latent function is modeled as a zero-mean GP with a composite kernel combining a radial basis function and a periodic kernel to capture nonlinear, local, and seasonal effects in the cavitation signal. To address the cubic computational complexity of GPs (Leibfried *et al.*, 2022; Sanz-Alcaine *et al.*, 2021), a sparse GP is used with $M = 200$ inducing points selected via k-means clustering. Training is performed via variational inference by maximizing the Evidence Lower Bound (ELBO) with the Adam optimizer over 200 epochs using a learning rate of 0.001.

The fifth model is a BNN, already described in Section 3.2.2, following the same architecture as the MLP, and trained using Stochastic Variational Inference (SVI) to maximize the ELBO over 1000 epochs with the Adam optimizer and a learning rate of 0.001.

Finally, a Linear Regression (LR) model is included as a baseline. It assumes a linear relationship between the inputs and the output and serves as a simple, interpretable, and computationally efficient reference.

It should be noted that all preceding hyperparameters were chosen empirically after preliminary trials, with the goal of balancing predictive performance, architectural simplicity, and robustness/generalization. All data are normalized using Z-score standardization and split into training and test sets from the same turbine unit. The selected model must balance predictive accuracy and architectural simplicity to promote robustness and generalization in virtual sensing applications.

3.3.2 Uncertainty Quantification

Uncertainty $U_i^{1-\alpha}$ quantifies the CI $I_i^{1-\alpha}$ associated with a point prediction $\hat{y}_i$ for a confidence level $1 - \alpha$:

$$I_i^{1-\alpha} = [\hat{y}_i - U_i^{1-\alpha}, \hat{y}_i + U_i^{1-\alpha}] \quad (3.1)$$

within which the true value y_i is expected to lie with probability close to $1 - \alpha$. Uncertainty can also be derived from a standard deviation σ via :

$$U^{1-\alpha} = k \sigma \quad (3.2)$$

where k is a coverage factor determined by the confidence level $1 - \alpha$ and the assumed distribution of the predictive error.

The reliability of quantified uncertainty is assessed through coverage analysis. While the theoretical coverage $C_{th}^{1-\alpha}$ corresponds to the nominal confidence level, the observed coverage $C_{obs}^{1-\alpha}$ measures the empirical proportion of true values falling inside the predicted CIs over a dataset of size n :

$$C_{obs}^{1-\alpha} = \frac{1}{n} \sum_{i=1}^n \delta_i \quad \text{with} \quad \delta_i = \begin{cases} 1 & \text{if } y_i \in I_i^{1-\alpha} \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

and a well-calibrated uncertainty must satisfy $C_{obs}^{1-\alpha} \approx C_{th}^{1-\alpha}$.

UQ can be addressed using several methods, as reviewed in Section 3.2.2, which surveys the most commonly adopted approaches and motivates our methodological choices. In this work, we consider BNN and GP, already introduced as predictive models in Section 3.3.1, together with heteroscedastic approaches, SCP, and CDE. This selection covers the most widely used UQ paradigms in PHM applications. Although other CP variants could be considered, CDE was selected for its superior robustness compared to other CP variants (Gagnon *et al.*, 2024), rendering the exploration of additional variants unnecessary within the scope of this study.

3.3.2.1 Heteroscedastic Modeling

This approach primarily models aleatoric uncertainty by jointly predicting the conditional mean $\hat{y}_i$ and variance $\hat{\sigma}_i^2$, from which the uncertainty is computed using equation (3.2), assuming a Gaussian distribution. To this end, the classical MSE loss is replaced by a weighted Gaussian negative log-likelihood, referred to as the β -NLL loss ($\mathcal{L}_{\beta\text{-NLL}}$) (Seitzer *et al.*, 2022) :

$$\mathcal{L}_{\beta\text{-NLL}} = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i^{2\beta} \left(\frac{1}{2} \left[\frac{(y_i - \hat{y}_i)^2}{\hat{\sigma}_i^2} + \log(\hat{\sigma}_i^2) \right] \right) \quad (3.4)$$

where $\beta \in [0, 1]$ controls the influence of samples with large predicted variance (Seitzer *et al.*, 2022). For $\beta = 0$, the loss reduces to the standard NLL, while $\beta = 1$ yields MSE-like gradients for the mean while preserving variance estimation. A value of $\beta = 0.5$ provides a good compromise between variability and accuracy (Seitzer *et al.*, 2022).

3.3.2.2 Split Conformal Prediction

SCP provides distribution-free CIs under exchangeability (Angelopoulos & Bates, 2022; Vovk *et al.*, 2005). Since this assumption is commonly violated in PHM, CP methods (SCP and CDE) are used here as calibration tools whose robustness is evaluated empirically. The dataset is divided into a training set, a calibration set of size n_{calib} , and a test set. Nonconformity scores s_i

are computed on the calibration set as :

$$s_i = |y_i - \hat{y}_i|, \quad i = 1, \dots, n_{calib} \quad (3.5)$$

from which the uncertainty is obtained as the empirical quantile $q_{1-\alpha}$:

$$U^{1-\alpha} = q_{1-\alpha} (\{s_i\}) \quad (3.6)$$

3.3.2.3 Conformalized Deep Ensembles

This method combines deep ensembles and CP. Deep ensembles consist of training M neural networks with different random initializations to capture epistemic uncertainty. For an input $\mathbf{x}_i$, the ensemble prediction $\hat{y}_i$ is computed as the mean of the individual outputs $\hat{y}_{i,m}$:

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^M \hat{y}_{i,m} \quad (3.7)$$

The uncertainty associated with input $\mathbf{x}_i$ is estimated as the standard deviation σ_i of the M predictions :

$$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (\hat{y}_{i,m} - \hat{y}_i)^2} \quad (3.8)$$

To ensure well-calibrated uncertainty, a separate calibration set of size n_{calib} is used, and normalized nonconformity scores R_i are defined as :

$$R_i = \frac{|y_i - \hat{y}_i|}{\sigma_i}, \quad i = 1, \dots, n_{calib} \quad (3.9)$$

The threshold d is computed as the empirical quantile $q_{1-\alpha}$ over the calibration scores R_i :

$$d = q_{1-\alpha} (\{R_i\}) \quad (3.10)$$

The CI $I_i^{1-\alpha}$ in the test set of size n_{test} is given by :

$$I_i^{1-\alpha} = [\hat{y}_i - d \sigma_i, \hat{y}_i + d \sigma_i], \quad i = 1, \dots, n_{test} \quad (3.11)$$

This method yields coverage close to the nominal level $1 - \alpha$, provided that the test distribution matches the calibration data. Under distribution shift (out-of-distribution, OOD), the validity of the CIs may deteriorate, requiring additional mechanisms such as autoencoder-based OOD detection to assess reliability (Gagnon *et al.*, 2024). In this study, we use an ensemble of $M = 5$ models, as recommended by Lakshminarayanan *et al.* (2017).

3.3.3 Transfer Learning

To assess the robustness of indirect measurements under transfer learning, two identical Francis turbine datasets (Unit 1 and Unit 2) are considered. A single model is trained using SCADA data and cavitation rates from one unit and then directly applied to the other without retraining. A Kolmogorov-Smirnov test reveals significant differences between the cavitation-rate distributions of Unit 1 and Unit 2, which may be attributed to measurement bias or to intrinsic physical variability.

Our work (Benabdesselam *et al.*, 2025) suggests that these discrepancies are primarily caused by miscalibration of the cavitation sensor, and that applying a single calibration scaling factor is sufficient to restore predictive accuracy. This factor is determined as the ratio between the prediction and the ground truth of the target turbine over a short calibration period. Accordingly, when training on one turbine and predicting on another, only the predictions (and not the associated uncertainty) are multiplied by a calibration coefficient. If the target turbine has never been equipped with a CMS, and therefore no ground truth ICR is available, the calibration coefficient cannot be determined directly from reference measurements. One practical option would be to perform a short initial calibration campaign on the target turbine using a temporary CMS, in order to estimate this coefficient. Alternatively, future work could investigate whether the calibration factor can be derived from the differences between the distributions of the

target-unit SCADA variables and those of the source unit used for training. This perspective may open the way to the identification of hydraulic similarity laws for erosive cavitation and to more general transfer-calibration strategies across turbines.

However, the objective here is to evaluate how different UQ methods respond to this distribution shift. In particular, exposure to previously unseen operating conditions is expected to increase epistemic uncertainty, whereas methods capturing only aleatoric uncertainty are likely to become miscalibrated.

3.3.4 Estimation of Cumulative Cavitation Rate

When uncertainty is quantified, the prediction of ICR at time t can be expressed as $ICR(t) = \hat{y}(t) \pm U(t)$, where $\hat{y}(t)$ is the point estimate and $U(t)$ is the associated uncertainty. The Cumulative Cavitation Rate (CCR), representing the total accumulated mass loss from an initial time τ_0 up to time t , is expressed as $CCR(t) = \hat{y}^c(t) \pm U^c(t)$, where $\hat{y}^c(t)$ is the cumulative point estimate computed by summing individual predictions :

$$\hat{y}^c(t) = \sum_{\tau=\tau_0}^t \hat{y}(\tau) \quad (3.12)$$

However, uncertainty does not propagate linearly. Since the uncertainty $U(t)$ behaves like a standard deviation, as shown in equation (3.2), and considering two successive time steps $t = t_0$ and $t = t_1$, the cumulative uncertainty at time t_1 , denoted $U^c(t_1)$, is given by :

$$U^c(t_1)^2 = U(t_0)^2 + U(t_1)^2 + 2\rho U(t_0) U(t_1) \quad (3.13)$$

where ρ is the autocorrelation coefficient between $\hat{y}(t_0)$ and $\hat{y}(t_1)$. While ρ could in principle be estimated empirically from the residuals on a validation set, its time-dependent nature and sensitivity to operating regime changes imply that a single global value would only be representative of a limited portion of the data and could lead to misleading cumulative uncertainty estimates. As a result, the use of bounding cases ($\rho = 0$ and $\rho = 1$) is adopted to bracket

cumulative uncertainty. When there is no autocorrelation between predictions, i.e., $\rho = 0$, the cumulative uncertainty is given by :

$$U^c(t_1) = \sqrt{U(t_0)^2 + U(t_1)^2} \quad (3.14)$$

If the predictions are in fact autocorrelated, neglecting the autocorrelation term leads to an underestimation of uncertainty. When predictions are perfectly autocorrelated, i.e., $\rho = 1$, the cumulative uncertainty is given by :

$$U^c(t_1) = U(t_0) + U(t_1) \quad (3.15)$$

Although this approach is conservative, it ensures better coverage of the true values, provided that the uncertainty estimated by the prediction model is properly calibrated. Moreover, this conservative behavior is not necessarily a limitation in the present application. As the prediction horizon increases, the model operates further away from the data distribution on which it was trained, and a substantial growth in uncertainty is therefore expected. In this sense, wider cumulative uncertainty bands at long horizons are consistent with the intended interpretation of the model outputs, namely to inform the operator that predictions become increasingly uncertain as extrapolation intensifies. It should also be noted that the exact shape of the degradation trajectory is not critical in absolute terms for maintenance decision-making, since a deviation of up to about 30% may remain operationally acceptable. In practice, when the cumulative uncertainty becomes too large at long horizons, this typically coincides with the stage at which an inspection should already have been scheduled. Once such an inspection is performed, the degradation assessment is updated and the cumulative uncertainty is effectively reset. Although ρ could be considered empirically, its value would remain strongly dependent on time and operating conditions, which limits the relevance of any single intermediate estimate.

3.3.5 Evaluation Metrics

The performance evaluation relies on two complementary categories of metrics.

The first category comprises standard accuracy metrics, namely the Normalized Root Mean Squared Error (NRMSE), Normalized Mean Absolute Error (NMAE), Symmetric Mean Absolute Percentage Error (SMAPE). These metrics quantify the deviation between the predicted values $\hat{y}_i$ and the ground truth y_i .

The second category comprises uncertainty calibration metrics, presented in Table 3.2, which summarize the statistical consistency and informativeness of the CIs. These metrics are computed over a range of confidence levels, sampled uniformly between $\alpha = 0.01$ and $\alpha = 0.99$, using K discrete steps (commonly $K = 100$) (Tran *et al.*, 2020). For each step $i \in \{1, \dots, K\}$, the empirical coverage C_{obs}^i is defined as the proportion of ground truth values falling within the predicted CI at level $1 - \alpha_i$, as given in equation (3.3), while the theoretical coverage is $C_{th}^i = 1 - \alpha_i$.

Tableau 3.2 Uncertainty calibration metrics

Metric	Formula
Expected Calibration Error (ECE)	$ECE = \frac{1}{K} \sum_{i=1}^K C_{obs}^i - C_{th}^i $
Root Mean Squared Calibration Error (RMSCE)	$RMSCE = \sqrt{\frac{1}{K} \sum_{i=1}^K (C_{obs}^i - C_{th}^i)^2}$
Miscalibration Area (MA)	$MA = \int_0^1 C_{obs}^{1-\alpha} - (1 - \alpha) d\alpha$
Sharpness	$Sharpness = \frac{1}{n} \sum_{i=1}^n 2 U_i^{1-\alpha}$

The ECE quantifies the average absolute difference between theoretical and observed coverage across confidence levels. The RMSCE, computed as the root mean squared difference, penalizes larger deviations more strongly. MA corresponds to the area between the observed and theoretical calibration curves, computed via numerical integration using the trapezoidal rule over K confidence levels. A low MA (close to zero) indicates well-calibrated uncertainty estimates, where the empirical coverage closely matches the nominal level. If the empirical curve lies above the theoretical one, the model overestimates uncertainty; if it lies below, it underestimates it. Finally, sharpness reflects the average width of CIs. While narrow CIs indicate higher confidence, excessively low sharpness may suggest overconfidence at the expense of proper calibration.

3.4 Results and Discussions

3.4.1 Prediction of ICR

Using the six models described previously, we predict ICR for two application cases : Unit 1 and Unit 2. For each unit, an incremental evaluation strategy is applied. Models are successively trained on increasing time windows and tested on the following years, from 2010 to 2014 for Unit 1, and from 2011 to 2014 for Unit 2. Performances are presented in Table 3.3. All experiments were conducted on a laptop equipped with an Intel Core i5-1145G7 CPU (4 cores, 8 threads) and 16 GB of RAM, running Windows 11 Pro. The models were implemented in Python 3.12.1.

Tableau 3.3 Performance comparison of models on both turbines

Model	NRMSE (%)	NMAE (%)	SMAPE (%)	CPU-Time (s)
Unit 1 – Train 2010–2011, Test 2012–2014				
MLP	5.81	4.62	15.85	80
LSTM	5.96	4.79	16.31	300
NARXNN	7.01	5.35	17.79	70

Model	NRMSE (%)	NMAE (%)	SMAPE (%)	CPU-Time (s)
GP	6.48	5.14	17.49	1680
BNN	6.46	5.16	17.38	290
LR	7.85	6.27	21.70	50
Unit 1 – Train 2010–2012, Test 2013–2014				
MLP	5.26	5.90	18.92	100
LSTM	7.04	5.69	18.10	540
NARXNN	7.99	6.19	18.95	110
GP	8.01	6.70	21.37	2221
BNN	8.38	6.86	21.15	200
LR	10.10	8.51	25.81	53
Unit 1 – Train 2010–2013, Test 2014				
MLP	7.87	6.40	20.35	110
LSTM	7.68	6.24	19.82	960
NARXNN	8.39	6.65	19.96	120
GP	9.39	7.92	24.93	2520
BNN	8.40	6.77	21.15	190
LR	9.70	8.10	24.85	57
Unit 2 – Train 2011, Test 2012–2014				
MLP	5.25	3.86	13.93	30
LSTM	5.07	3.64	12.74	150
NARXNN	5.48	4.12	15.40	35
GP	5.93	4.17	14.41	931
BNN	5.40	3.88	14.31	210
LR	6.31	4.52	14.96	20

Model	NRMSE (%)	NMAE (%)	SMAPE (%)	CPU-Time (s)
Unit 2 – Train 2011–2012, Test 2013–2014				
MLP	5.42	3.92	13.01	40
LSTM	5.24	3.72	11.96	421
NARXNN	5.44	3.88	12.96	47
GP	6.25	4.34	13.52	1200
BNN	5.88	4.20	13.50	200
LR	6.70	4.83	15.05	20
Unit 2 – Train 2011–2013, Test 2014				
MLP	6.97	4.92	15.07	45
LSTM	7.16	5.23	16.09	682
NARXNN	7.01	5.08	15.63	49
GP	7.76	5.66	27.30	1800
BNN	7.92	5.88	17.98	200
LR	8.95	6.84	20.50	23

Table 3.3 shows that the MLP and LSTM models consistently achieve the highest predictive performance across all datasets, with comparable accuracy metrics. Although LSTM occasionally yields marginal improvements when strong temporal dependencies are present, the overall performance gap remains limited. In contrast, MLP provides a decisive computational advantage, with training times up to five times faster than those of LSTM. This efficiency becomes particularly critical when UQ is introduced, as it inherently increases computational cost. NARXNN and BNN also demonstrate competitive performance, but remain consistently inferior to MLP and LSTM. GPs, on the other hand, suffer from high computational complexity, sensitivity to kernel selection, and limited scalability, and their predictive performance is in some cases comparable to that of the LR baseline. Overall, MLP offers a favorable trade-off between accuracy, robustness,

and scalability, making it especially well suited for large-scale PHM applications with UQ requirements.

3.4.2 Uncertainty Quantification

The MLP model was retained for the remainder of the study and combined with several UQ methods for comparison, namely heteroscedastic modeling (β -NLL loss), SCP, CDE, and BNN. GPs were excluded due to insufficient accuracy and prohibitive computational cost.

For each turbine, two datasets (Dataset 1 and Dataset 2) were constructed. For Unit 1, Dataset 1 uses data from 2010–2011 for training, 25 000 observations for calibration (required by SCP and CDE), and the remaining data from mid-2012 to the end of 2014 for testing. To ensure a fair comparison, only the 2010–2011 data were used for training in the β -NLL and BNN models. Dataset 2 follows the same structure, with an extended training period from 2010–2012 and testing from mid-2013 to the end of 2014. For Unit 2, Dataset 1 uses data from 2011 for training, 25 000 observations for calibration, and data from late 2012 to the end of 2014 for testing. Dataset 2 extends the training period to 2011–2012, while maintaining the same calibration and testing protocol. Table 3.4 reports the calibration metrics obtained for each UQ method across all datasets.

The results in Table 3.4 highlight that observed coverage alone is insufficient to assess uncertainty quality, and must be jointly analyzed with calibration metrics.

The heteroscedastic β -NLL approach produces narrow but overconfident CIs, resulting in low coverage and large calibration errors. While post-hoc calibration methods (e.g., isotonic regression or conformal techniques) can partially mitigate this issue, they increase pipeline complexity and still provide no theoretical guarantee of coverage. These results suggest that this approach, when used alone and capturing only aleatoric uncertainty, is insufficient for PHM applications. In practice, epistemic uncertainty may dominate, particularly under changing operating regimes that force the model to extrapolate beyond the training distribution. In such cases, uncertainty estimates must reflect the lack of knowledge associated with extrapolation,

Tableau 3.4 Comparison of uncertainty quantification methods on the four datasets

Method	ECE	RMSCE	MA	Sharpness	OC (95%)
Unit 1 – Dataset 1					
β -NLL	0.1249	0.1372	0.1231	1.2326	81.98
SCP	0.0199	0.0284	0.0197	1.7878	97.20
CDE	0.0069	0.0079	0.0068	2.0082	95.42
BNN	0.1262	0.1416	0.1249	2.8429	98.63
Unit 1 – Dataset 2					
β -NLL	0.1801	0.1992	0.1777	1.4274	74.85
SCP	0.1582	0.1721	0.1563	1.4144	83.90
CDE	0.1198	0.1321	0.1185	1.9552	90.81
BNN	0.0321	0.0368	0.0317	2.8215	95.01
Unit 2 – Dataset 1					
β -NLL	0.2182	0.2421	0.2150	1.0969	65.00
SCP	0.0921	0.1013	0.0911	1.5394	91.30
CDE	0.0485	0.0531	0.0480	2.0296	93.38
BNN	0.2322	0.2577	0.2297	4.4378	99.37
Unit 2 – Dataset 2					
β -NLL	0.1486	0.1628	0.1465	1.3346	77.67
SCP	0.0791	0.0910	0.0780	1.4936	83.30
CDE	0.0123	0.0136	0.0121	2.2456	92.71
BNN	0.1782	0.1964	0.1763	4.1535	98.61

rather than only the intrinsic variability of the observations, in order to support reliable decision-making.

SCP achieves better calibration results than β -NLL, but still exhibits noticeable calibration errors, particularly on Unit 1–Dataset 2, due to violations of the exchangeability assumption between the calibration and test sets. Such violations are common in PHM applications and, in

degradation monitoring, typically manifest as under-coverage. This behavior primarily reflects an inadequate quantification of epistemic uncertainty induced by distribution shifts.

BNNs exhibit systematic over-coverage across datasets, reflecting overestimated uncertainty, poor calibration, and excessively wide CIs. This behavior is primarily driven by their strong sensitivity to prior specification. Although Bayesian inference is theoretically supported by the Bernstein–von Mises theorem, which guarantees posterior convergence to a normal distribution around the truth as the sample size increases, this result holds only under restrictive assumptions, namely correct model specification, finite dimensionality, and independent and identically distributed sampling (Diaconis & Freedman, 1986; Hartigan, 1983; Le Cam, 1986; van der Vaart, 1998). These conditions are rarely met in practice, and the theorem no longer applies to nonparametric models such as BNNs or GPs. In such settings, the posterior may instead be dominated by the prior, regardless of the observed data (Diaconis & Freedman, 1986), which is problematic given that BNN priors are often arbitrary and difficult to calibrate. This prior sensitivity, combined with the lack of convergence guarantees, undermines the robustness of BNNs for uncertainty-aware prediction in tasks requiring strong generalization, such as the indirect measurement of latent physical quantities. To quantify this effect, experiments were conducted on 2 datasets of contrasting complexity using a fixed BNN architecture and 6 different priors (Gaussian, $\mathcal{N}$; Laplace, $\mathcal{L}$; and Cauchy, $\mathcal{C}$, each with different variances), as reported in Table 3.5.

Table 3.5 confirms the strong influence of prior choice on BNN calibration. On Unit 1–Dataset 2, a high-variance Gaussian prior yields the best performance, whereas low-variance priors degrade the results. Conversely, on Unit 2–Dataset 2, a low-variance Laplace prior performs best. This dataset-dependent reversal highlights the strong sensitivity and instability of BNNs with respect to prior specification, with no universally optimal configuration. These findings demonstrate the limited robustness of BNNs, and more generally of Bayesian methods such as GPs, for real-world PHM applications, where the optimal prior is rarely known and may vary even within the same application due to changing operating regimes.

Tableau 3.5 Impact of different priors on BNN uncertainty quantification

Prior	ECE	RMSCE	MA	Sharpness
Unit 1 – Dataset 2				
$\mathcal{N}(0, 1)$	0.0321	0.0368	0.0317	2.8215
$\mathcal{N}(0, 0.5)$	0.0577	0.0689	0.0572	2.2029
$\mathcal{L}(0, 1)$	0.0336	0.0424	0.0342	2.4595
$\mathcal{L}(0, 0.5)$	0.0933	0.1030	0.0922	1.7651
$\mathcal{C}(0, 1)$	0.0531	0.0639	0.0529	3.0395
$\mathcal{C}(0, 0.5)$	0.1159	0.1279	0.1146	1.8005
Unit 2 – Dataset 2				
$\mathcal{N}(0, 1)$	0.1786	0.1967	0.1767	4.1543
$\mathcal{N}(0, 0.5)$	0.0962	0.1074	0.0952	2.5832
$\mathcal{L}(0, 1)$	0.1322	0.1465	0.1308	3.1172
$\mathcal{L}(0, 0.5)$	0.0127	0.0196	0.0123	1.7786
$\mathcal{C}(0, 1)$	0.1889	0.2086	0.1869	4.3932
$\mathcal{C}(0, 0.5)$	0.0157	0.0210	0.0153	1.7926

CDE consistently provides reliable and informative uncertainty estimates across all datasets, with coverage close to the 95% target, near-zero calibration errors, and controlled CI widths. Its performance is influenced by ensemble size : small ensembles promote diversity and coverage, whereas excessively large ensembles degrade calibration ; in practice, 5 models offer an effective balance between robustness, accuracy, and computational cost (Lakshminarayanan *et al.*, 2017). Importantly, CDE maintains near-nominal coverage even when the exchangeability assumption is partially violated, a situation that is common in practical PHM applications, owing to the diversity induced by deep ensembles.

It should be noted that only the training time of CDE may be relatively long. Based on Table 3.3, the MLP, for example, requires about 110 seconds for training ; therefore, using CDE with five models would result in a total training time of approximately 550 seconds. However, this cost corresponds mainly to the training phase. Once training is completed, the model is frozen and

used only for inference, for which the computation time does not exceed 10 seconds. Since the SCADA data considered in this work are sampled every 10 minutes, this inference time remains fully compatible with online deployment in the present PHM application.

Figure 3.3 and Figure 3.4 show the MLP–CDE predictions for Unit 1 with Dataset 1 and Dataset 2, respectively, while Figure 3.5 and Figure 3.6 report the corresponding results for Unit 2. Overall, the model reproduces the temporal evolution of the ICR with good accuracy in all four cases, indicating that the proposed indirect measurement framework is able to capture the main degradation dynamics from SCADA variables alone. The agreement between predictions and ground truth is particularly clear in medium- to high-cavitation regimes, where the structure of the signal is more pronounced and better represented in the training data. By contrast, slightly larger deviations can be observed in low-cavitation regions, which likely reflects the combined effect of weaker signal intensity and lower data density in these operating conditions.

In all cases, the 95% CIs adapt to local variations in noise level and data sparsity, becoming wider in regions where the prediction task is more uncertain and narrower where the dynamics are more regularly learned. Importantly, the ground truth ICR values remain enclosed within the intervals throughout the different datasets and turbines, which confirms the good calibration and robustness of the MLP–CDE combination. These visual results are consistent with the quantitative metrics reported previously and further support the relevance of combining deep ensembles with CP for uncertainty-aware PHM from SCADA data.

3.4.3 Transfer Learning

To evaluate UQ under distribution shift, models are trained on Unit 1 (2010 to early 2014) and tested on Unit 2, and vice versa. Methods that do not require calibration (such as β -NLL and BNN) use only the training data. Conformal methods (SCP and CDE) additionally use 25000 calibration samples from 2014 to construct CIs. Performance is assessed on the target unit using ECE, RMSCE, MA, sharpness, and observed 95% coverage. Table 3.6 presents the results for both transfer directions.

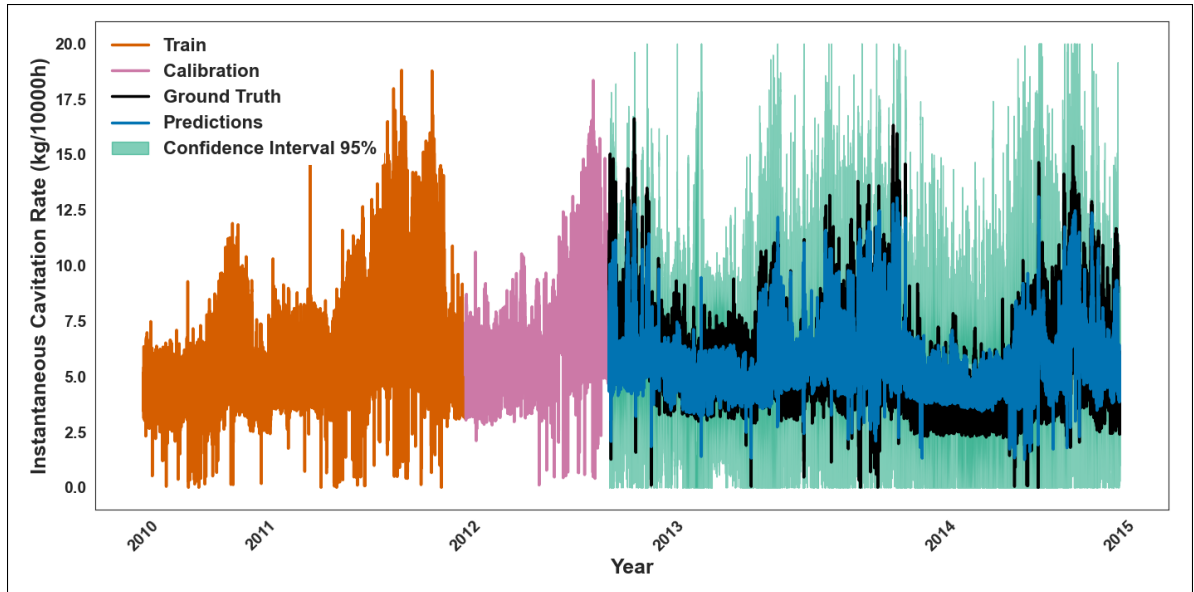


Figure 3.3 Prediction of the instantaneous cavitation rate for Unit 1 on Dataset 1

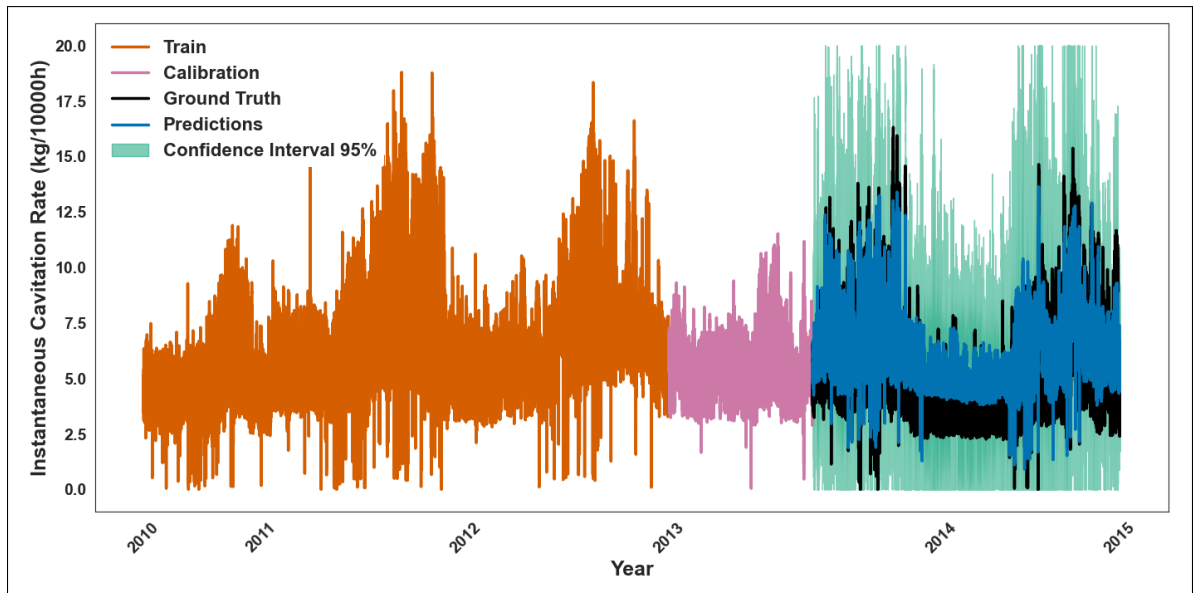


Figure 3.4 Prediction of the instantaneous cavitation rate - Unit 1 - Dataset 2

Table 3.6 reports the performance of the different UQ methods under transfer learning, where training and test distributions differ. As expected, the β -NLL approach fails to capture epistemic uncertainty, resulting in large calibration errors and severe undercoverage. BNN and SCP exhibit

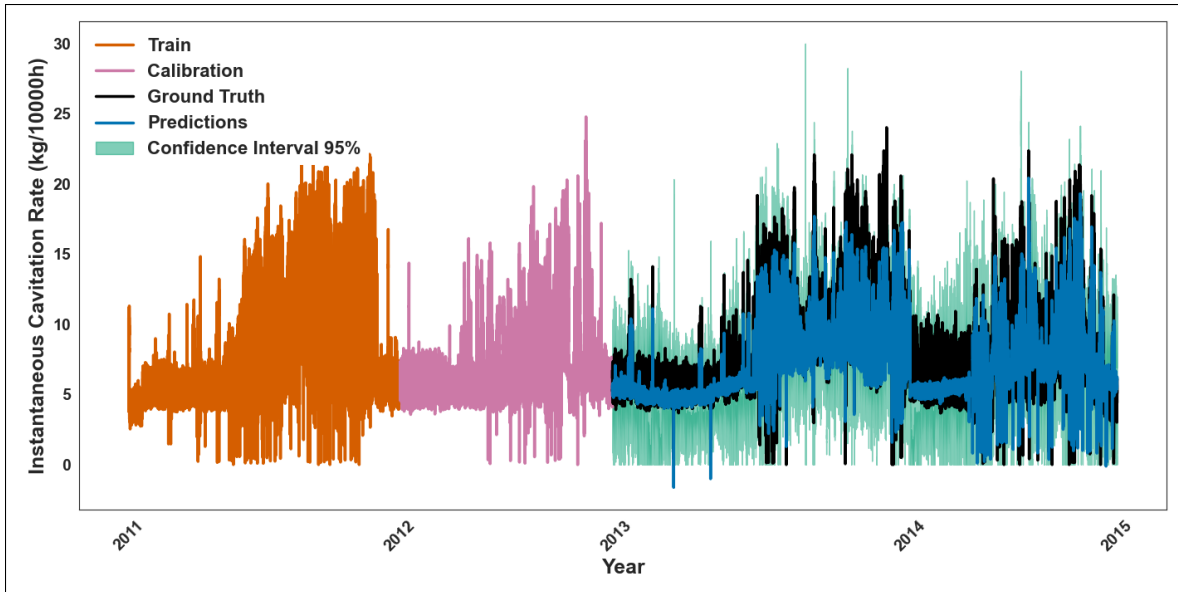


Figure 3.5 Prediction of the instantaneous cavitation rate - Unit 2 - Dataset 1

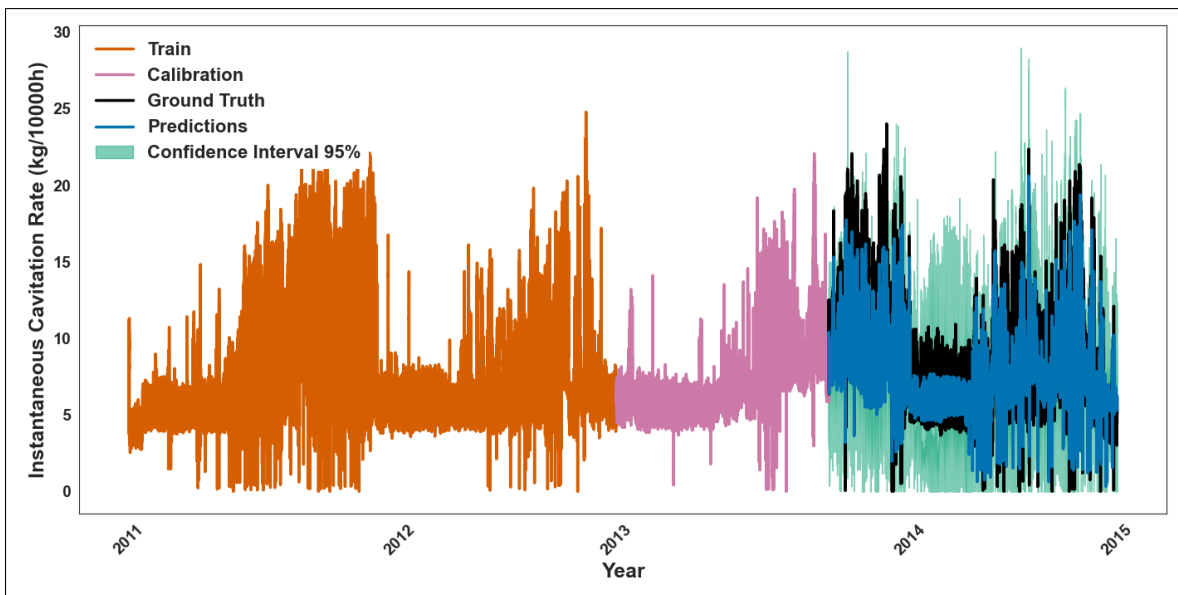


Figure 3.6 Prediction of the instantaneous cavitation rate - Unit 2 - Dataset 2

moderate performance and tend to underestimate uncertainty under distribution shift. The degradation of SCP is expected, as its exchangeability assumption is violated in this setting ;

Tableau 3.6 Comparison of uncertainty quantification methods on transfer scenarios

Method	ECE	RMSCE	MA	Sharpness	OC (95%)
Unit 1 to Unit 2					
β -NLL	0.2698	0.3075	0.2651	1.2908	48.08
SCP	0.1566	0.1976	0.1538	1.9564	64.30
CDE	0.0416	0.0561	0.0412	4.8853	96.00
BNN	0.1439	0.1723	0.1415	2.0943	62.10
Unit 2 to Unit 1					
β -NLL	0.3650	0.4143	0.3586	1.3057	32.40
SCP	0.2177	0.2452	0.2152	2.1010	71.50
CDE	0.1604	0.1795	0.1588	3.1961	88.80
BNN	0.1930	0.2192	0.1902	2.5223	59.47

nevertheless, SCP remains substantially more reliable than β -NLL and performs on par with BNN.

By contrast, even under exchangeability violations, CDE consistently delivers the most reliable uncertainty estimates, with low ECE and MA values and coverage close to the 95% target, degrading to about 88% under severe exchangeability violations, while remaining within an acceptable range. This robustness under distribution shift is explained by the high diversity of ensemble predictions in extrapolation regions under non-exchangeability, which appropriately increases predictive uncertainty dominated by epistemic effects. As a result, CDE exhibits strong generalization under distribution shift, highlighting its robustness and practical relevance for real-world PHM applications.

Figure 3.7 and Figure 3.8 further illustrate the CDE predictions obtained for the two transfer directions. In both cases, the predicted ICR remains in good agreement with the reference values, showing that the proposed framework preserves its predictive capability even when applied to a turbine that was not used during training. This is a particularly important result, as it indicates that the relationship learned between SCADA variables and cavitation-related

degradation is not restricted to a single asset, but can generalize across similar hydraulic turbines. Although some local discrepancies remain, the overall temporal evolution and amplitude of the ICR are still well captured after transfer, which confirms the relevance of the proposed indirect measurement strategy in a cross-unit setting. The corresponding 95% CIs also remain consistent with the transferred predictions and continue to enclose the ground truth well, indicating that the proposed UQ framework retains useful calibration properties under distribution shift. This point is essential in PHM applications, where deployment conditions often differ from those observed during model development. From an industrial perspective, these results suggest that SCADA data can not only replace CMS for cavitation monitoring, but also support a scalable deployment strategy in which a model trained on an instrumented turbine can be transferred to another turbine that is not equipped with CMS, without requiring retraining. Such a capability opens the way to pre-trained SCADA-based monitoring frameworks for fleet-level applications. At the same time, the results also indicate that calibration remains a necessary step to preserve reliable confidence estimates when transferring the model from one turbine to another.

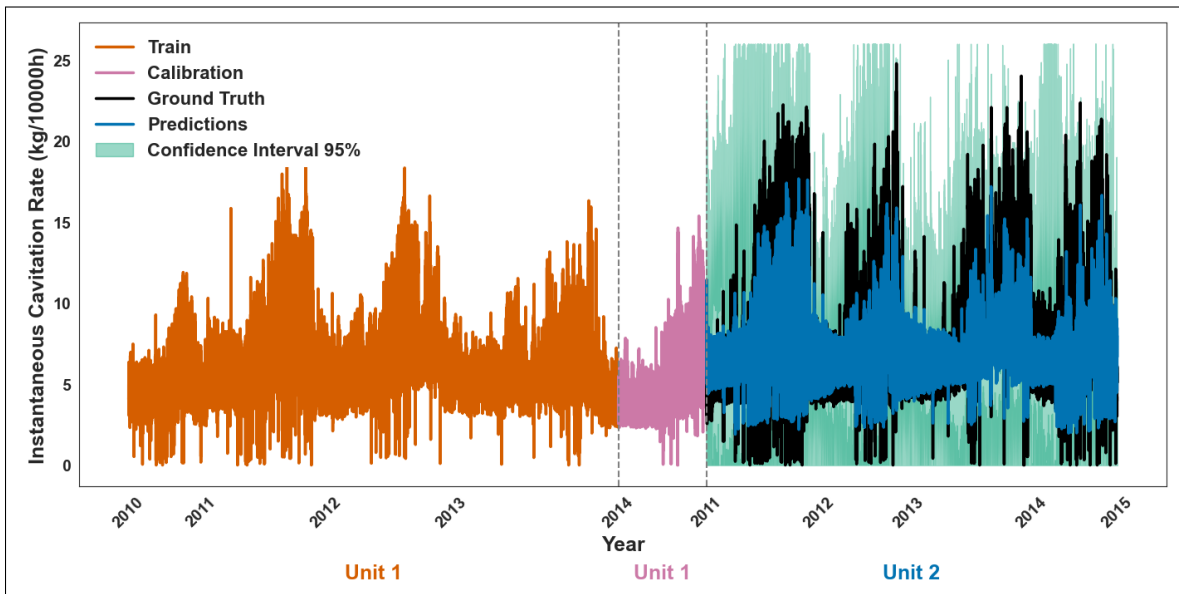


Figure 3.7 Transfer Learning from Unit 1 to Unit 2

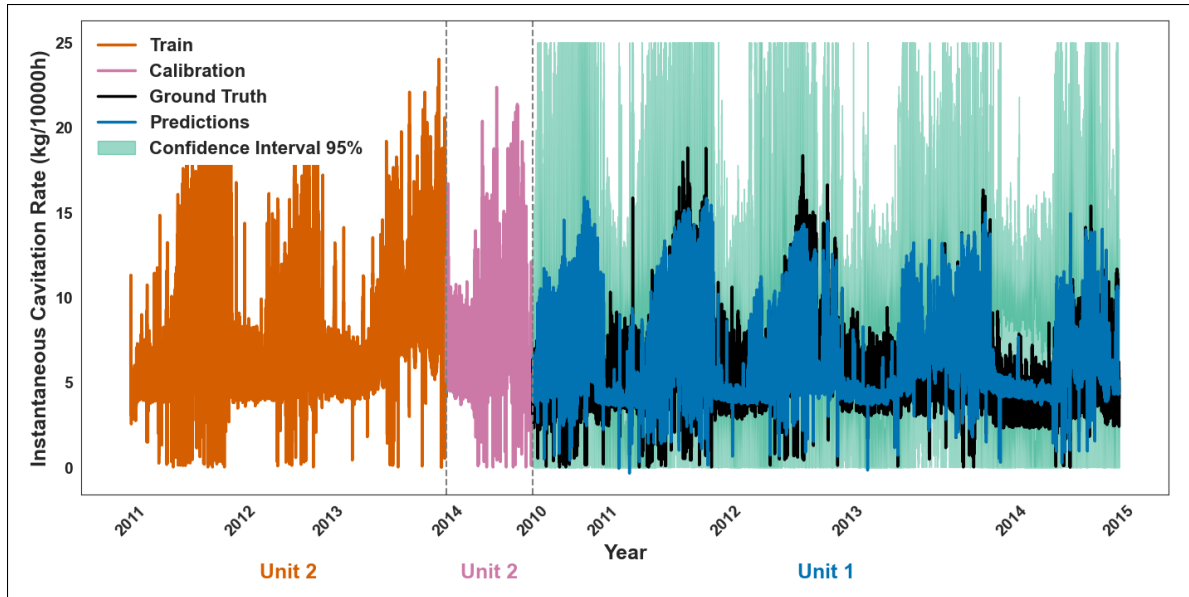


Figure 3.8 Transfer Learning from Unit 2 to Unit 1

3.4.4 Prediction of CCR

Using the MLP with CDE, the point predictions and associated 95% CIs are further aggregated to estimate the CCR. For Unit 1, two datasets are considered. Under Dataset 1, and assuming independent uncertainties as in equation (3.14), the resulting degradation curve is shown in Figure 3.9. The estimated CCR follows a smooth increasing trend, consistent with the cumulative nature of erosive cavitation. However, the associated 95% CI is extremely narrow and becomes visually indistinguishable at the scale of the figure, requiring a zoomed view to reveal its presence. This suggests that uncertainty accumulation is strongly underestimated under the independence assumption. Such a result is not physically realistic in the present context, since successive degradation predictions are derived from temporally ordered SCADA data and are therefore expected to exhibit non-negligible dependence. As a result, assuming independent uncertainties leads to overconfident cumulative estimates and an overly optimistic representation of the degradation state. This observation highlights the importance of accounting for temporal dependence when propagating uncertainty from pointwise ICR predictions to the cumulative degradation trajectory.

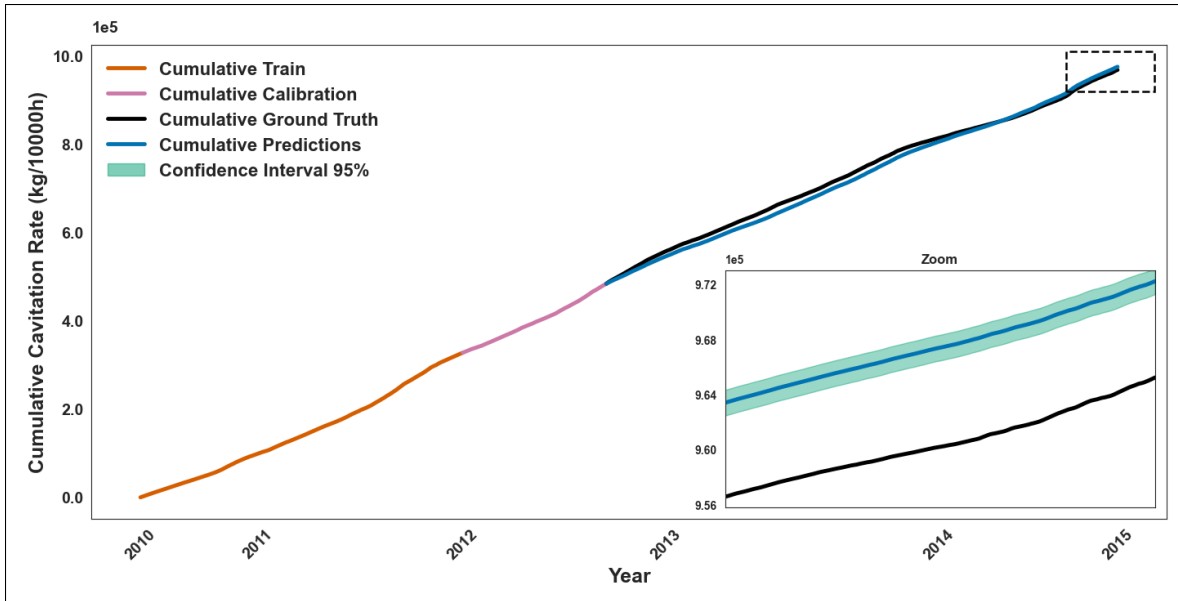


Figure 3.9 Evolution of cumulative cavitation - Unit 1 - Dataset 1 - Assuming independence

This limitation motivates a second, more realistic aggregation strategy based on the assumption of perfect autocorrelation between successive predictions, as expressed in equation (3.15). Although this assumption is more conservative, it leads to a well-defined and visually interpretable 95% CI, as shown in Figure 3.10. In contrast to the independence-based formulation, the resulting band better reflects the temporal dependence inherent to degradation trajectories derived from ordered SCADA data. The ground truth remains accurately captured within this CI, which supports the relevance of this formulation for practical diagnosis. These results suggest that, in the context of cumulative cavitation monitoring, explicitly accounting for dependence between successive predictions provides a more realistic representation of the uncertainty associated with the estimated degradation state. Importantly, when a maintenance threshold, typically defined by maintenance operators or domain experts as a critical cumulative amount of eroded mass based on industrial constraints, inspection practices, and acceptable risk levels, is specified, the proposed framework enables continuous and quantitative assessment of the degradation state, thereby supporting fault diagnosis and informed maintenance and inspection planning,

including the scheduling of turbine shutdowns. In the present case study, the threshold illustrated in Figure 3.10 is shown for illustrative purposes only.

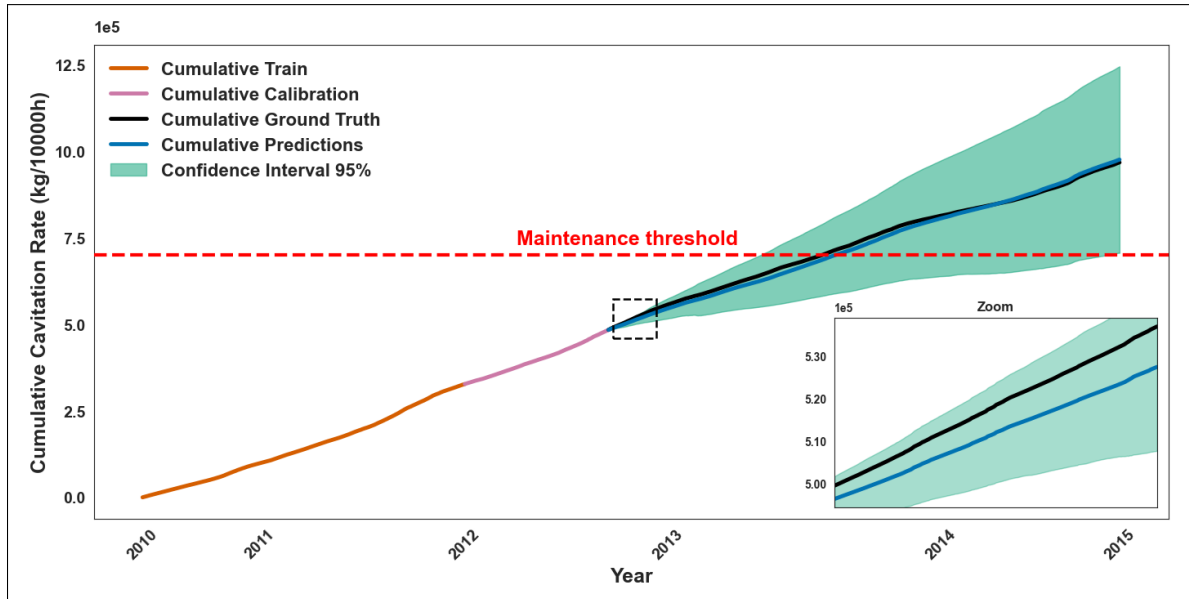


Figure 3.10 Evolution of cumulative cavitation - Unit 1 - Dataset 1 - Assuming perfect autocorrelation

Figure 3.11 presents the cumulative predictions obtained for Dataset 2 of Unit 1 using the perfect autocorrelation approach. As in the previous case, the resulting degradation trajectory remains consistent with the expected cumulative evolution of cavitation, while the associated 95% CI provides a clear and interpretable representation of uncertainty. The ground truth is well captured, which further supports the relevance of this conservative formulation for practical degradation assessment.

3.5 Conclusion

This paper addresses a practical PHM challenge for hydraulic turbines : the continuous monitoring of erosive cavitation progression without relying on dedicated CMS. Using real multi-year industrial data from two similar Francis units, we demonstrate that SCADA variables can be leveraged to indirectly estimate the ICR with competitive accuracy using simple and scalable models suitable for fleet-level deployment. Importantly, we show that the proposed model

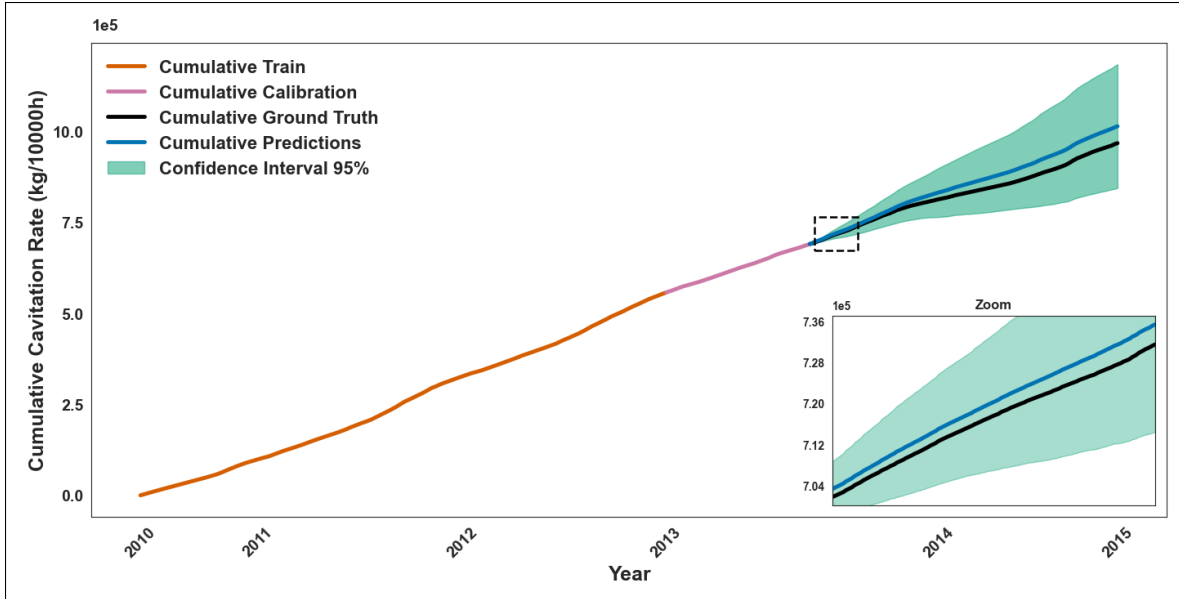


Figure 3.11 Evolution of cumulative cavitation - Unit 1 - Dataset 2 - Assuming perfect autocorrelation

is transferable across turbines without retraining, representing a significant advancement for cavitation monitoring in PHM applications.

Since point estimates alone are not actionable for maintenance decision support, we conducted a rigorous UQ benchmark tailored to PHM constraints, including non-stationarity, distribution shift, and imperfect exchangeability. Across four datasets and transfer-learning scenarios, heteroscedastic β -NLL systematically under-covered due to the lack of epistemic uncertainty modeling, while BNN uncertainty proved highly sensitive to prior choice and dataset-dependent, limiting operational robustness. SCP offered partial improvement but remained vulnerable to exchangeability violations. CDE consistently achieved the best trade-off between statistical reliability and informativeness, delivering near-nominal coverage with low calibration errors, including under distribution shift. Finally, we translated uncertainty-aware ICR predictions into an uncertainty-bounded cumulative cavitation trajectory, enabling practical diagnostics of degradation progression over time. From a reliability engineering and system safety perspective, this uncertainty-aware degradation assessment supports risk-informed, reliability-centered

maintenance decision-making and contributes to preventing unsafe operating conditions in hydraulic turbines.

Overall, the results demonstrate that SCADA can effectively replace CMS and provide evidence-based guidance on which UQ methods remain reliable for SCADA-based PHM, establishing CDE as a robust and scalable choice for uncertainty-aware virtual sensing and diagnostics of erosive cavitation under real operating conditions.

Future work should focus on enhancing UQ techniques, mainly based on CP, to better handle exchangeability violations without increasing model complexity. It should also further investigate the replacement of CMS with SCADA in PHM by considering additional application cases and promoting this paradigm. Moreover, the proposed framework naturally extends to prognostics : the cumulative degradation curve can be exploited in future work to estimate the remaining useful life, thereby closing the PHM loop.

CHAPITRE 4

PROGNOSTICS OF THERMAL ANOMALIES IN WIND TURBINES VIA DEEP LEARNING AND CONFORMAL PREDICTION USING SCADA DATA

Abderrauof Benabdesselam¹ , Monelle Comeau ² , Adaiton Oliveira-Filho¹ , Martin Gagnon³ , Antoine Tahan¹

¹ Département de Génie Mécanique, École de Technologie Supérieure, 1100 Rue Notre-Dame O., Montréal, QC H3C 1K3, Canada ;

² Power Factors, 7005 Boulevard Taschereau, Brossard, QC J4Z 1A7, Canada ;

³ Hydro-Québec, 75 Boulevard René-Lévesque O., Montréal, QC H2Z 1A4, Canada.

Article soumis pour publication à la revue « IEEE Access : Digital Twins for Energy-Related Applications » en décembre 2025.

4.1 Abstract

This study introduces a classification-based prognostic framework to anticipate temperature threshold exceedance events that lead to automatic shutdowns in wind turbines. The framework assesses whether an observed temperature anomaly is likely to evolve into such an event and relies exclusively on Supervisory Control and Data Acquisition data combined with Deep Learning and Conformal Prediction for uncertainty quantification. To characterize the temporal evolution of the exceedance tendency, we propose a novel metric, the Instantaneous Event Intensity, derived from confidence intervals produced by Conformal Prediction. By aggregating Instantaneous Event Intensity values over time, we define the Cumulative Event Intensity, which integrates the event magnitude. The Cumulative Event Intensity is then compared with a predefined event threshold to support binary decision-making on whether the component will experience a temperature threshold exceedance event. The proposed approach is validated on 207 wind turbines from 20 wind farms, covering three manufacturers and three distinct components, detecting 87.25% of temperature threshold exceedance events 15 days in advance, with a false positive rate below 21%. Notably, several events can be anticipated up to three months earlier, highlighting the robustness of the early-warning capability.

4.2 Introduction

To maintain the competitiveness of renewable energy systems, reducing Operation and Maintenance (O&M) costs has become a strategic priority (Dao *et al.*, 2019). In Wind Turbines (WTs), O&M expenses account for nearly 30% of the total levelized cost of energy (expressed in \$/MWh), while unplanned downtime is responsible for approximately 65–75% of the overall O&M budget (Santolamazza, Dadi & Introna, 2021; Qiao & Qu, 2018).

Vibration analysis remains the standard approach for detecting mechanical faults (Beretta, Julian, Sepulveda, Cusidó & Porro, 2021; Cambron *et al.*, 2017; Lu, Li, Wu & Yang, 2009; Milani *et al.*, 2025b; Salameh, Cauet, Etien, Sakout & Rambault, 2018; Vieira *et al.*, 2024), whereas acoustic analysis offers a powerful yet costly alternative (Gbashi, Madushele, Olatunji, Adedeii & Jen, 2022; Lu *et al.*, 2009; Salameh *et al.*, 2018). Both techniques, however, require the installation of additional sensors or on-site measurement campaigns to collect the data (Beretta *et al.*, 2021). Lubrication analysis, although low-cost and easy to implement, offers limited diagnostic value and is prone to frequent false alarms (Gbashi *et al.*, 2022; Wilkinson, Spinato & Tavner, 2007; Lu *et al.*, 2009). A more accessible alternative relies on condition monitoring based on Supervisory Control and Data Acquisition (SCADA) systems, which is attractive due to its low cost, data abundance, and widespread availability in wind farms. However, while some sensors operate at high sampling rates, SCADA data are typically aggregated at 10-minute intervals, limiting their suitability for high-frequency monitoring (Benabdesselam *et al.*, 2024). In this context, SCADA-based temperature measurements are used as indicators of component health, as temperature increases may signal degradation caused by friction (e.g., in bearings) or Joule heating (e.g., in generators). By analyzing deviations from historical or reference temperature profiles, early anomalies can be detected (Beretta *et al.*, 2021; Cambron *et al.*, 2017). Figure 4.1 illustrates a representative example, contrasting a healthy WT bearing with a degraded one.

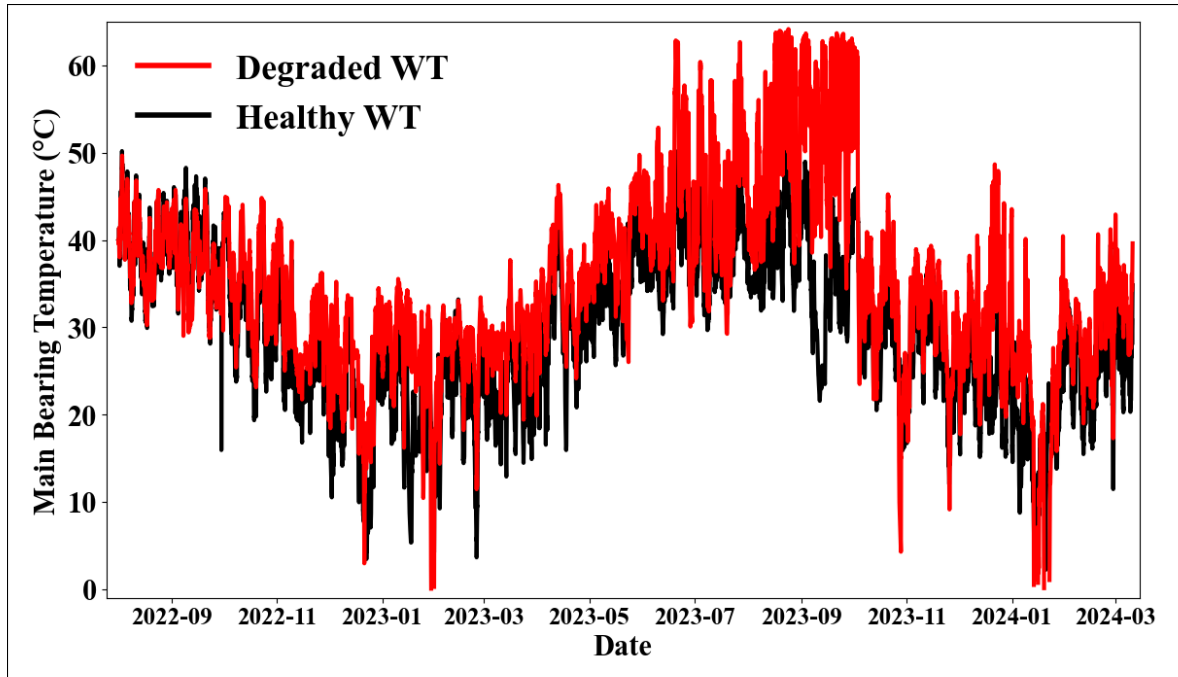


Figure 4.1 Main bearing temperature time series for a healthy and a degraded wind turbine

4.2.1 Related Works

In this context, Prognostics and Health Management (PHM) strategies have gained prominence (Benabdesselam *et al.*, 2024), and typically involve three stages :

1. **Anomaly detection** : Numerous studies have explored anomaly detection methods for WT components, relying either on vibration signals, which are generally robust (Martínez Montaña & Sanz-Bobi, 2018; Meyer, 2022; Zhang, Hu & Yang, 2022a; Jonas, Anagnostos, Brodbeck & Meyer, 2022; Lin, Pan, Xi, Wang & Jiang, 2024), or on temperature measurements, which are more widely adopted and easier to implement (Cambron *et al.*, 2017; Weissenfeld, Vanerio, Wachsenegger, Graser *et al.*, 2025; Campoverde-Vilela, Feijóo, Vidal, Sampietro & Tutivén, 2023; Beretta *et al.*, 2021; Wang *et al.*, 2025a; Liu, Corbita, N. T., Lee & Wang, 2022; Udo & Muhammad, 2021; Guo, Infield & Yang, 2012; Garan, Tidiri & Kovalenko, 2022; Jankauskas *et al.*, 2023; Santolamazza *et al.*, 2021; Qiu, Feng, Sun, Zhang & Infield, 2016; Sun, Li, Wang & Lei, 2016; Tonks & Wang, 2017; Feng, Qiu, Crabtree, Long & Tavner, 2013;

Lichtenstein, von Maltzahn & Pelka, 2024; Yucesan & Viana, 2020; Oliveira-Filho, Zemouri, Pelletier & Tahan, 2024; Oliveira-Filho, Zemouri, Cambron & Tahan, 2023). These methods can be physics-based, in cases where a real-time physical model is available, typically derived from thermal exchange equations (Cambron *et al.*, 2017), eigenvalue or modal update techniques (Iliopoulos *et al.*, 2017, 2016, 2015; Noppe *et al.*, 2016; Bhowmik, Tripura, Hazra & Pakrashi, 2019, 2020; Mucchielli, Bhowmik, Ghosh & Pakrashi, 2021; Mucchielli, Bhowmik, Hazra & Pakrashi, 2020), or state-space representations and observer-based frameworks (Perišić *et al.*, 2015). In particular, modal-based and eigen decomposition-based approaches have been investigated for real-time anomaly and downtime detection in WTs and structural systems, as they can capture system-level dynamics through changes in covariance or modal structure (Bhowmik *et al.*, 2019, 2020; Mucchielli *et al.*, 2021, 2020). However, these dynamic indicators may also vary with environmental and operational conditions, which complicates the distinction between normal variability and true degradation in practice. Moreover, such approaches mainly support real-time detection of ongoing abnormal events rather than prediction of their future evolution. Alternatively, anomaly detection can rely on data-driven approaches using deep-learning architectures (Guck, Bindl, Fernandez & Stetter, 2024; Oliveira-Filho *et al.*, 2024, 2023; Weissenfeld *et al.*, 2025; Santolamazza *et al.*, 2021).

2. **Fault diagnosis :** Fault diagnosis in WTs has traditionally relied on physics-based or rule-based models, such as fuzzy logic or Petri nets (Deng, Jiang & Liu, 2017; Zhang *et al.*, 2023c), which map observed symptoms (e.g., temperature rise, vibration patterns) to predefined failure modes. While interpretable, these approaches depend heavily on expert knowledge and struggle to cope with nonstationary operating conditions and complex failure interactions. In contrast, recent reviews highlight a shift toward machine-learning and Artificial Intelligence (AI)-based schemes, often implemented through classifiers that learn discriminative fault representations directly from SCADA and condition-monitoring data (Sun *et al.*, 2021; Xu *et al.*, 2024; Zhang, Chen, Zhu, Zhang & Li, 2023a). Deep neural-network classifiers trained on vibration or multisensor data now achieve accurate diagnosis of gearbox, bearing, and blade faults under variable regimes (Du *et al.*, 2023; Zhang *et al.*, 2023b; Han & Feng, 2023; Cheng *et al.*, 2023; Tang *et al.*, 2024; Sethi, Subba,

Faisal, Sahoo & Raju, 2024), while generative models and multisource signal fusion help mitigate data imbalance and scarce fault samples (Zhang, Zhang & Cai, 2022b). Recent self-supervised learning frameworks have further improved the characterization of early-stage device degradation and the modeling of complex operating conditions, notably through time–frequency dual-domain prediction and fusion strategies (Xu, He, Sun, Jin & Zhao, 2026; He, Xu, Sun, Jin & Zhao, 2025a). Transfer-learning strategies are increasingly used to reduce domain shifts across WTs, operating conditions, and test benches (Han & Feng, 2023; Cheng *et al.*, 2023; Tang *et al.*, 2024), but achieving robust generalization across WT types, environmental regimes, and degradation trajectories remains a major open challenge (Tang *et al.*, 2021; Alagha, Khairuddin, Haitaamar, Al-Khatib & Kanesan, 2025; Liu *et al.*, 2024; Zhang *et al.*, 2023a).

3. **Health state prognosis :** Once an anomaly is detected, the goal is to estimate the Remaining Useful Life (RUL), i.e., the time left before a critical degradation threshold is reached (Liu, Feng & Si, 2012). This remains a major scientific and industrial challenge (Hu, Tan & Tang, 2024). Physics-based prognostic models rely on fundamental physical laws and dynamic system modeling. When degradation mechanisms are well understood, they describe damage evolution through analytical formulations such as Paris’ law, Miner’s rule, or state-space representations, as demonstrated in applications beyond WTs (Daigle & Goebel, 2011; Zhang & Lee, 2011; Liu & Zio, 2017; Brownjohn, De Stefano, Xu, Wenzel & Aktan, 2011). However, in practice, the underlying dynamic model is often unknown or only partially characterized (Zio, 2022; Benabdesselam *et al.*, 2024). Even when available, such models are typically valid for relatively simple and well-defined systems (Zio, 2022). For complex systems such as WTs, physics-based approaches present significant limitations : they rely on simplifying assumptions (e.g., linear damage accumulation), depend on difficult-to-measure parameters, struggle to capture nonlinear and coupled dynamics, and lack flexibility under highly variable operating conditions (Saathoff, Rosemeier, Kleinselbeck & Rathmann, 2021; Pagitsch, Jacobs, Bosse & Duda, 2020; Guo *et al.*, 2020; van Binsbergen, Soares, Pedersen & Nejad, 2022). In addition, their deployment at fleet scale remains challenging. Consequently, purely physics-based prognostic approaches

remain scarce for WTs, and data-driven methods constitute the most viable alternative, as highlighted in Cuesta Arrillaga, Leturiondo, Vidal & Pozo (2025). Most data-driven approaches rely on vibration measurements to construct a Health Indicator (HI) that tracks degradation over time (Milani *et al.*, 2025b; Soualhi *et al.*, 2020; Yoo & Baek, 2018; Guo *et al.*, 2017, 2018; Ren *et al.*, 2018; Li *et al.*, 2019; She & Jia, 2019; Chen *et al.*, 2021; Jiang *et al.*, 2022; Zhang *et al.*, 2023c). Deep-learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), bidirectional RNNs, and attention-based architectures capture nonlinear degradation patterns and long-term dependencies, enabling more accurate RUL inference (Wang, Wen, Yang & Liu, 2018; Cao, Zhang, Meng & Wang, 2023). Capsule Neural Networks (CapsNet) further improve hierarchical representation and multimodal fusion : the sensor-aware CapsNet provides uncertainty quantification and interpretable feature attribution across sensors, outperforming conventional deep learning methods in multisensory RUL tasks (Li, Chen, Huang, Chen & Li, 2024a). In contrast, SCADA-based prognostics remain relatively scarce. Vieira *et al.* (2024) derived an HI from bearing-ambient temperature differences and used regression models on a limited number of bearings. Oliveira-Filho *et al.* (2021) projected SCADA measurements into a latent space via a variational autoencoder, interpreting degradation as a trajectory from healthy to faulty clusters. Shah *et al.* (2024) bypassed feature engineering with a hybrid CNN-RNN-attention model using multivariate SCADA data. Milani *et al.* (2025a) enforced monotonicity by constructing an HI with a constrained autoencoder and extrapolating it to a failure threshold. Hu *et al.* (2025) decoupled SCADA time series into multiple temporal scales using a multi-head RNN and predicted RUL with quantified uncertainty. Other hybrid frameworks integrate deep learning with bootstrap for prediction intervals (Huang, Huang, Li & Peng, 2021), RReliefF-based feature selection for robust vibration prognostics (Wang *et al.*, 2024), physics-informed similarity or degradation modeling from vibration-derived HIs (Le, Lee, Dinh & Park, 2024), and system-level RNN-Monte Carlo dropout strategies coupling probabilistic prognostics with inspection planning (Manna, Mitici & Dalla Vedova, 2024). Overall, most SCADA-based studies are validated on very few run-to-failure cases (typically 1-7 days) (Hu *et al.*, 2025; Milani *et al.*, 2025a; Oliveira-Filho *et al.*, 2021; Vieira

et al., 2024; Shah *et al.*, 2024), which limits robustness and generalizability. Moreover, many approaches only predict failures shortly before they occur, with prognostic horizons often below 2 weeks (Shah *et al.*, 2024) and in some cases as short as 2 days (Hu *et al.*, 2025).

While anomaly detection and fault diagnosis are well studied, a key challenge remains : distinguishing true degradation from benign events such as maintenance, oil changes, or sensor faults. Moreover, the prognostics task remains highly challenging in general, and even more so when relying solely on low-frequency SCADA data, which are preferable for PHM compared to vibration data, often unavailable.

In addition to prognostics, quantifying the uncertainty associated with predictions is essential in PHM to support decision-making, and more broadly, to transform uncertainty into actionable decisions (Zio, 2022). Uncertainty quantification remains relatively underexplored, particularly within PHM and even more so in renewable-energy applications (Benabdesselam, Gagnon & Tahan, 2026b). Several approaches to uncertainty quantification have been proposed, and Benabdesselam *et al.* (2026b) demonstrated that Conformal Prediction (CP) methods outperform alternatives such as Bayesian neural networks, while providing a principled way to construct confidence intervals without assuming a specific error distribution, as long as the data satisfy the exchangeability assumption (Vovk *et al.*, 2005). Consequently, multiple CP variants have been introduced in the literature (Barber, Candès, Ramdas & Tibshirani, 2021; Angelopoulos & Bates, 2022; Vovk *et al.*, 2005), and some have been directly applied to RUL prediction (Piao *et al.*, 2025), notably through cross-validated conformalized quantile regression. Although robust, this approach incurs a significant computational cost, unlike Split CP, the most widely used CP method due to its ease of deployment and minimal computational cost (Oliveira, Orenstein, Ramos & Romano, 2024).

4.2.2 Key Gaps

Despite significant progress, WT prognostics still face several critical limitations :

- The physical degradation model of WTs remains largely unavailable, which leads to the use of data-driven approaches, as consistently reported in the literature and highlighted in Cuesta Arrillaga *et al.* (2025).
- The majority of WT prognostic models rely on vibration data, which are not available for most industrial WTs, thereby limiting real-world deployment (Moreira de Oliveira Filho, 2024).
- Data-driven RUL approaches require the extraction of a HI, a task that is particularly challenging when relying solely on SCADA data, which are less sensitive to degradation compared to vibration measurements.
- Most SCADA-based prognostic approaches lack generalizability, rely on complex architectures, and are validated on a limited number of cases with short prognostic horizons (often below 2 weeks).
- The vast majority of WT prognostic frameworks do not explicitly quantify predictive uncertainty, limiting their reliability for operational decision-making.

4.2.3 Objectives

This study considers WTs equipped with built-in alarms that automatically shut down the WT when the predefined temperature threshold of the studied component is exceeded. In the industrial setting considered here, this shutdown is used as an operational event associated with the studied thermal degradation scenario, leading to immediate downtime. However, we acknowledge that downtime may also depend on other operational and control factors. Accordingly, the objective of this study is not to predict downtime in a general sense, but rather to predict this specific temperature-threshold exceedance event from SCADA data.

Because vibration data are not available in our case – and remain unavailable for the majority of industrial WTs – the proposed approach relies exclusively on SCADA data. However, SCADA-based prognostics present inherent limitations compared to CMS. SCADA measurements are low-frequency and indirect : they do not provide access to high-frequency or dynamic response

information, nor do they directly quantify physical damage within a system. Consequently, degradation mechanisms are only partially and indirectly observable through the limited available SCADA signals, which limits the sensitivity, reliability, and long-horizon predictive capability of the prognostic framework.

Accordingly, rather than framing the problem as a continuous regression task, this study reformulates it as a threshold-exceedance classification task. In this sense, it extends diagnostic classification frameworks previously proposed in the literature (Xu *et al.*, 2026; He *et al.*, 2025a), while going beyond diagnosis toward prognostic decision-making. The aim of this framework is prognostic, in the sense that it does not seek to detect whether the component is currently abnormal, but rather to predict whether an already detected thermal anomaly is likely to evolve into a future temperature-threshold exceedance event. This objective is supported by the following specific goals :

- **A practical forecasting approach using historical seasonal patterns :** Indirect measurement (also known as virtual sensing (Benabdesselam *et al.*, 2024)) of component temperature using a Long Short-Term Memory (LSTM) model, leveraging exogenous SCADA variables from the previous year when future data are unavailable.
- **Uncertainty quantification :** Uncertainty is quantified using CP methods to construct confidence intervals.
- **Development of event intensity functions :** Using the confidence intervals obtained from CP, the Instantaneous Event Intensity (IEI) function, denoted $d(t)$, is derived to characterize the temporal evolution of the exceedance tendency. Aggregating the IEI over time yields the Cumulative Event Intensity (CEI), denoted $D(t)$, which quantifies the overall magnitude of the exceedance event.
- **Formulation of a binary decision model :** Classification of anomalies into either an event state (associated with shutdown) or a normal state, based on the CEI and a decision threshold θ_c .
- **Extensive validation on 207 industrial cases :** To assess performance across multiple components (main bearing, generator bearings, generator) and failure types, as well as to

evaluate robustness over different prediction horizons, multiple wind farm locations, and diverse WT models and manufacturers.

4.2.4 Contributions and Outline

The proposed classifier enables asset managers to prioritize maintenance actions according to the severity and urgency of detected anomalies, optimizing preventive maintenance planning and reducing unplanned downtime. Designed for large-scale SCADA datasets, it is particularly valuable when RUL estimates are highly uncertain or unavailable due to insufficient or poor-quality data, providing a robust decision-making alternative in such contexts.

The remainder of this paper is as follows : Section 4.3 presents the data, Section 4.4 presents the proposed methodology, Section 4.5 presents the results and Section 4.6 concludes the paper.

4.3 Data Sources

This work is based on a dataset provided by our industrial partner, comprising a total of 207 WTs from three different manufacturers, distributed across 20 wind farms. All WTs are horizontal-axis, equipped with three-blade rotors and belong to the megawatt-class power category. The dataset covers the period from September 2022 to March 2024, during which numerous abnormal temperature behaviors were recorded in the main bearing, generator bearings, and generator.

The selection of relevant variables for the model was guided by a thermodynamic approach inspired by Cambron *et al.* (2017). This step aims to retain the most informative SCADA variables for health state prediction while limiting model complexity. The selected variables are listed in Table 4.1. In addition, for each component (main bearing, generator bearings, and generator), a temperature threshold $T_{threshold}$ is defined according to the WT model, which varies across manufacturers. When this threshold is exceeded, the WT's built-in alarm system automatically triggers a shutdown to prevent irreversible damage.

Tableau 4.1 Description of the key variables used in the study

Measure	Unit	Description
T_{comp}	°C	Measured temperature of the component
T_{amb}	°C	Ambient temperature
U	m/s	Wind speed
P	kW	Electrical active power

Threshold exceedance was observed in 102 of the 207 WTs, while the remaining 105 WTs maintained component temperatures below the critical threshold. Table 4.2 summarizes, for each monitored component, the number of reported anomalies, the number of threshold exceedances, and the corresponding temperature threshold intervals based on WT specifications. Complementary historical data for wind speed, ambient temperature, and active power are also available for the year 2021.

Tableau 4.2 Summary of anomaly statistics per component from September 2022 to March 2024

Component	Reported Anomalies	Overheating Events	Threshold Range (°C)
Main Bearing	132	87	[60–95]
Generator Bearing	65	11	[80–116]
Generator	10	4	[118–149]

** A unique threshold value is defined for each WT model.*

4.4 Methodology

The methodology consists of three main steps. First, data preprocessing and timeline partitioning are performed, as detailed in Section 4.4.1. Second, an indirect measurement of T_{comp} is obtained with an LSTM model, which covers training, prediction, and calibration to quantify uncertainty, as explained in Section 4.4.2. Third, event functions are constructed : CP is applied to compute confidence intervals, from which the IEI is derived and aggregated into the CEI. Based on the

CEI and an event threshold, anomalies are classified as either normal or event-related, as defined in Section 4.4.3, with performance evaluated using the metrics presented in Section 4.4.4. The complete workflow is illustrated in Figure 4.2.

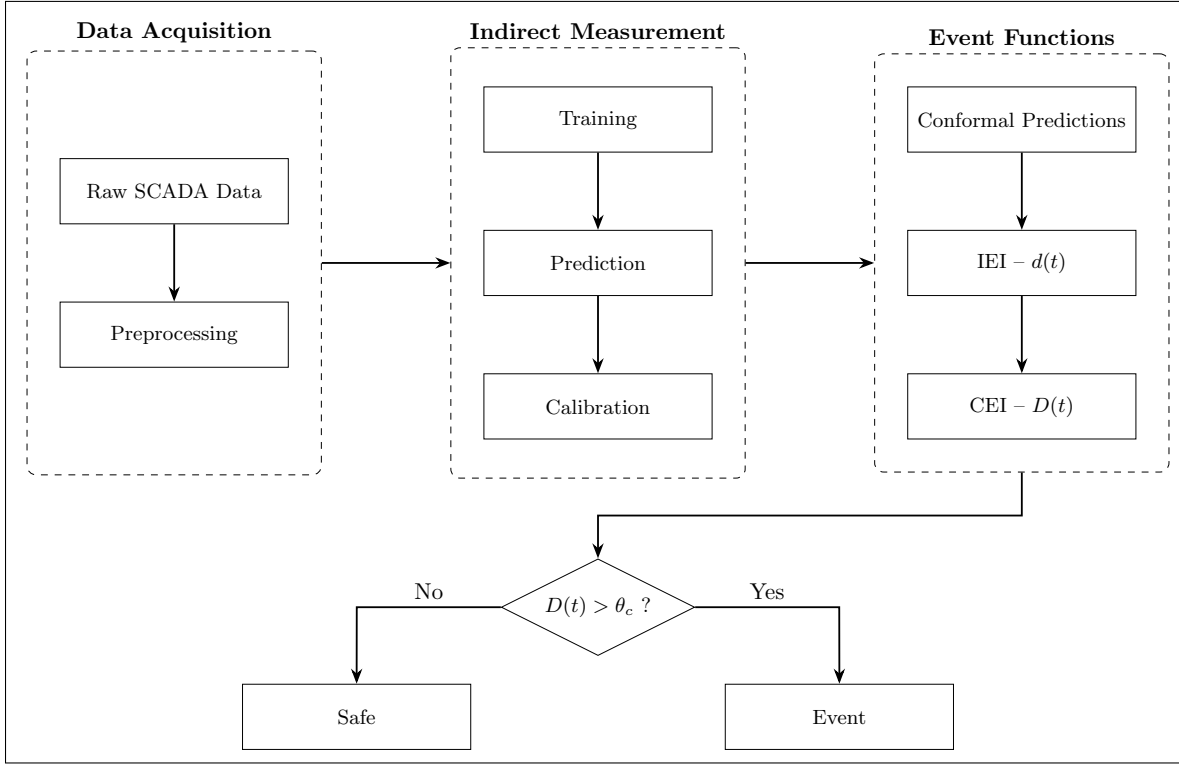


Figure 4.2 Thermal monitoring methodology for WT components

4.4.1 Timeline Partition and Data Preprocessing

Inspired by the linear physics-based model of bearing temperature introduced by Cambron *et al.* (2017), the component temperature T_{comp} is modeled as a function of ambient temperature T_{amb} , wind speed U , and active power P . From a physical standpoint, T_{amb} and U primarily characterize the convective heat transfer between the component and its surrounding environment. In contrast, P captures Joule heating effects associated with the generator, whose contribution to the overall thermal balance is typically smaller than that of convection. Unlike the original formulation, this approach accounts for potential nonlinearities, including wind-induced convection. The

exact relationship is determined empirically using a deep learning model capable of capturing complex dynamics within an indirect measurement framework.

The proposed strategy for predicting the component temperature T_{comp} is based on a temporal structure defined as follows :

- **Training period** : $[t_i^{train}, t_f^{train}]$ using both the ground truth T_{comp} and the corresponding exogenous variables (T_{amb}, U, P).
- **Prediction (test) period** : During this phase, we aim to predict T_{comp} over the test interval $[t_i^{test}, t_f^{test}]$. However, since future exogenous inputs are unavailable, we use historical exogenous data from one year earlier, i.e., from $t_i^{test-hist} = t_i^{test} - 1 \text{ year}$ to $t_f^{test-hist} = t_f^{test} - 1 \text{ year}$. The justification for this choice is provided in Section 4.4.2.
- **Calibration period** : Used for uncertainty quantification, as detailed later in the text. The calibration period corresponds to the training period $[t_i^{train}, t_f^{train}]$, but with the exogenous variables from the previous year, i.e., from $t_i^{train-hist} = t_i^{train} - 1 \text{ year}$ to $t_f^{train-hist} = t_f^{train} - 1 \text{ year}$.

The temporal structure is illustrated in the timeline shown in Figure 4.3. The training period duration $\Delta t_{train} = t_f^{train} - t_i^{train}$, and the prediction (test) period duration $\Delta t_{test} = t_f^{test} - t_i^{test}$ are tunable hyperparameters.

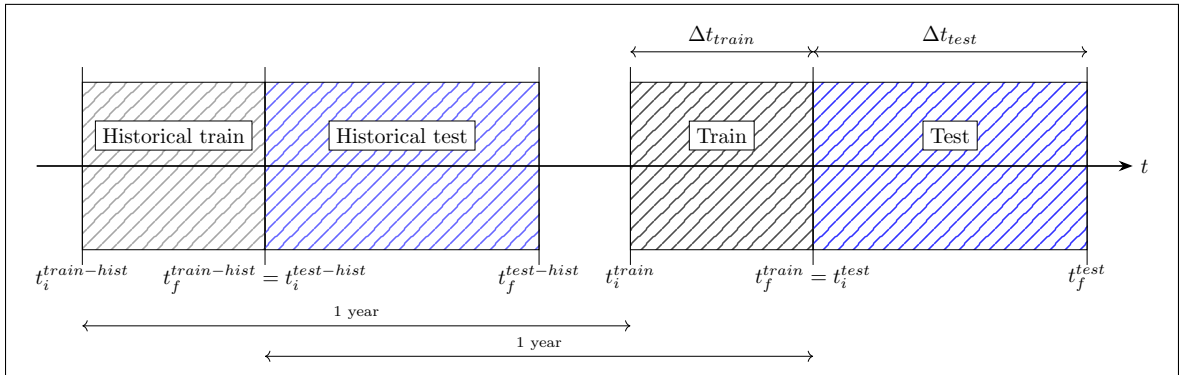


Figure 4.3 Proposed timeline showing training, calibration, and testing intervals

The data undergo preprocessing, including cleaning and filtering outliers. All time series were aggregated to a 1-h resolution by averaging all available samples within each one-hour interval. This aggregation smooths intra-hour variability, ensures temporal alignment across variables, and mitigates measurement irregularities prior to subsequent sequence generation. To prepare the data for sequence-based modeling, the continuous SCADA time series were divided into overlapping segments using a sliding window of 30 hours. In this process, each window captures 30 consecutive hourly measurements of all variables, with the window shifting forward by one time step to generate the next sequence. After segmentation, all features within each sequence were rescaled to the $[0,1]$ range using Min-Max normalization.

4.4.2 Indirect Measurement and Uncertainty Quantification

During training, sequences of exogenous variables (T_{amb} , U , P) and the component temperature T_{comp} were provided as inputs to an LSTM model. This architecture was selected for its ability to capture temporal dependencies, particularly those arising from thermal inertia. Alternative models were also assessed : the Autoregressive Integrated Moving Average (ARIMA) and the Multi-Layer Perceptron (MLP) failed to capture system dynamics, while the Nonlinear AutoRegressive Neural Network (NARNN) achieved slightly lower accuracy and required nearly three times the computation time. The LSTM therefore offered the best balance between predictive performance and computational efficiency. The architecture and hyperparameters were selected empirically based on multiple trials, aiming to balance predictive performance and model complexity. This iterative process involved testing different configurations and retaining the one yielding the lowest validation error without signs of overfitting, such as a widening gap between training and validation performance. The retained model consists of two successive LSTM layers with 30 units each, followed by dense layers with 64, 128, 32, and 16 neurons, all employing Rectified Linear Unit (ReLU) activation, and ending with a final linear output layer. Optimization is performed using the Adam optimizer, with the Mean Squared Error (MSE) as the loss function. The use of two stacked LSTM layers was motivated by the need to better capture the temporal inertia of the thermal response, where temperature evolution reflects

accumulated effects rather than instantaneous fluctuations. In practice, a single recurrent layer struggled to adequately model this delayed dynamic, particularly during slow heating or cooling phases. Conversely, the deeper MLP stack was retained because preliminary ablation experiments showed that it substantially improved the capture of nonlinear interactions between the exogenous variables and T_{comp} . Configurations with fewer dense layers systematically underperformed, especially in regimes involving rapid variations of the thermal signal. The training phase is illustrated in Figure 4.4, where T_{amb}^{train} , U^{train} , P^{train} , and T_{comp}^{train} represent the training dataset variables, corresponding to the period from t_i^{train} to t_f^{train} .

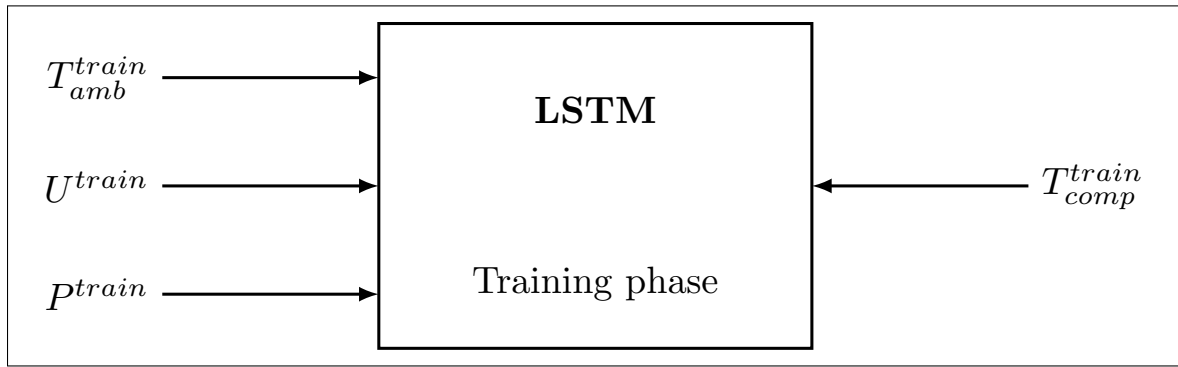


Figure 4.4 Illustration of LSTM training phase

During the test phase, a key challenge is the lack of future exogenous variables needed to predict the target component temperature T_{comp} over $[t_i^{test}, t_f^{test}]$. Several options were explored to address this limitation. One approach is to forecast each exogenous variable separately using large historical datasets or by incorporating additional explanatory variables. In practice, this becomes difficult to deploy : it requires extensive data collection, feature engineering, and preprocessing, and each additional forecast introduces further uncertainty. External meteorological forecasts represent another possibility, but they are constrained by limited spatial resolution, the need for precise WT-level localization, and inherent prediction errors. For global commercial applications, a universally available weather source would be necessary. While global weather models exist, their forecast horizons remain limited, and they do not capture local conditions that strongly influence WT thermal behavior. Integrating such forecasts would substantially increase the complexity of the monitoring system, requiring heavier pipelines and continuous

data synchronization. It would also be necessary to establish a reliable link between grid-level forecast variables, such as temperature at 10 m on a 20-40 km mesh, and the actual temperature observed at the WT. This correlation is non-trivial and highly site-dependent, making the approach technically demanding.

To address these limitations, a seasonal projection was adopted, assuming that the temporal patterns of T_{amb} , U , and P remain unchanged across years. Historical data from the previous year, over $[t_i^{test-hist}, t_f^{test-hist}]$, are reused as surrogate inputs for the current test period, as illustrated in Figure 4.5, where T_{comp}^{test} denotes the predicted component temperature over $[t_i^{test}, t_f^{test}]$, and T_{amb}^{test} , U^{test} , and P^{test} are the mapped historical values from the preceding year.

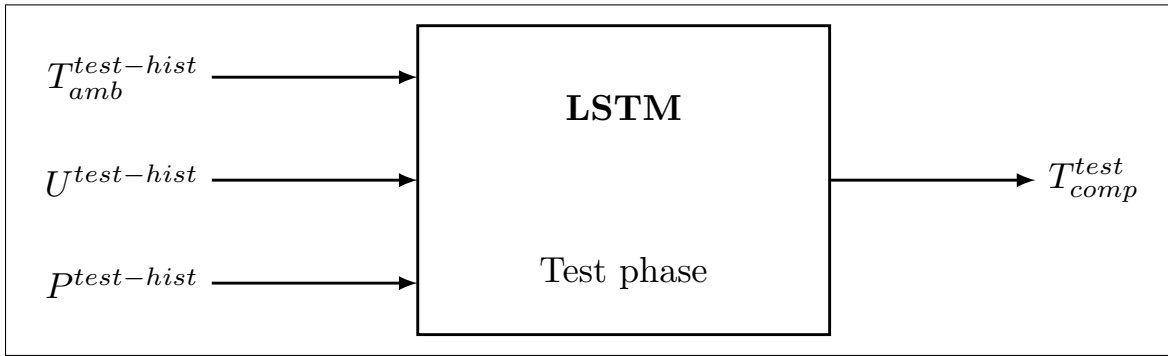


Figure 4.5 Illustration of LSTM test phase

The validity of the seasonal projection was assessed by computing, for each exogenous variable and each WT, the Normalized Mean Absolute Error (nMAE) and Normalized Root Mean Squared Error (nRMSE) between actual and projected values, and then averaging these metrics across all WTs. For ambient temperature T_{amb} , the mean nMAE was $11.28\% \pm 1.88\%$ and the mean nRMSE was $14.62\% \pm 2.85\%$. For wind speed U , the mean nMAE was $15.82\% \pm 2.92\%$ and the mean nRMSE was $20.01\% \pm 3.78\%$. Active power P showed the largest errors, with a mean nMAE of $34.00\% \pm 3.60\%$ and a mean nRMSE of $44.00\% \pm 4.13\%$. These results suggest that the projection provides a reasonable approximation for meteorological variables in the present dataset, with errors that remain acceptable for the purpose of this study. While more advanced forecasting techniques could marginally improve accuracy, the expected gains may not justify the additional complexity and computational cost in the present framework. For P , the high

variability observed during unplanned shutdowns makes precise forecasting inherently difficult. The proposed approximation still allows this variable to retain an explanatory role despite the induced discrepancy. Although this assumption is strong, it appears suitable within the scope of this study, since the objective is not to precisely forecast T_{comp} . It should also be noted that T_{comp} is substantially less sensitive to P , which is the least influential variable in predicting T_{comp} . Using a Random Forest model, P accounts on average for approximately 20% of the predictive contribution, whereas U accounts for roughly 30%, and T_{amb} for about 50%. Similar findings are reported in studies on T_{comp} prediction for WTs (Turnbull, Carroll & McDonald, 2022; Murgia, Verbeke, Tsiorkova, Terzi & Astolfi, 2023; Wang *et al.*, 2020b) (including generator components), where P consistently appears among the least influential inputs. Additionally, in the present dataset, this assumption appears more reasonable when exogenous data from the immediately preceding year are used. Using historical data that is temporally distant would introduce substantial error, as the statistical properties of the exogenous variables drift over time. Even when the previous year is used, the underlying exogenous trends must remain broadly comparable; significant seasonal shifts, curtailment events, or grid-dispatch constraints can lead to large prediction errors, as they substantially alter the operational and thermal regimes of the asset.

To apply CP for uncertainty quantification, a third dataset is required, in which the model's predictions are compared to the actual component temperatures in order to estimate nonconformity scores. Rather than using a separate dataset solely for residual estimation, we reuse the original training period $[t_i^{train}, t_f^{train}]$, but with the exogenous variables from the previous year $[t_i^{train-hist}, t_f^{train-hist}]$. Using the already trained model, we feed it with the historical exogenous variables: $T_{amb}^{train-hist}$, $U^{train-hist}$, $P^{train-hist}$, to predict the calibration component temperature T_{comp}^{calib} , from t_i^{train} to t_f^{train} , which takes into account the potential prediction error arising from the use of historical inputs. These predictions are then compared to the actual values T_{comp}^{train} (ground truth) in order to compute nonconformity scores. This process is illustrated in Figure 4.6.

By reusing the original training period while substituting the exogenous variables with those from the previous year for the calibration dataset, and subsequently forecasting the temperature

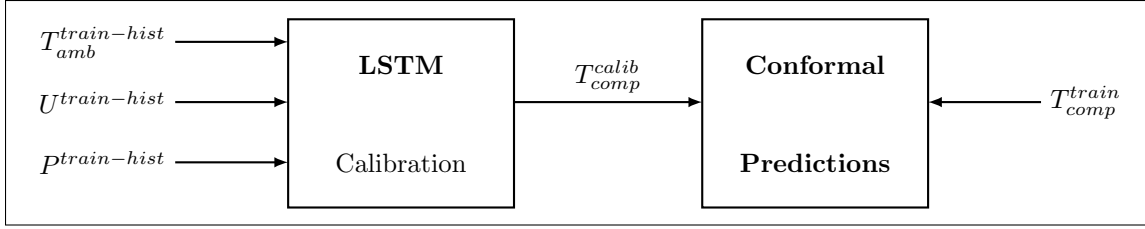


Figure 4.6 Illustration of LSTM calibration

using the previous year's exogenous inputs, the resulting uncertainty is expected to reflect part of the inter-annual variability of these variables. This is particularly relevant for active power P , which, in addition to being dependent on a meteorological variable that is difficult to predict over long horizons, also varies according to operator decisions. Consequently, the propagated uncertainty is expected to reflect part of this variability effect. Once the calibration component temperature T_{comp}^{calib} has been predicted for the entire calibration dataset, composed of n_{calib} samples (matching the training dataset in length), the nonconformity score s_i (Vovk *et al.*, 2005) is computed as shown in equation (4.1).

$$s_i = |T_{comp}^{calib}(t_i) - T_{comp}^{train}(t_i)| \quad i = 1, \dots, n_{calib} \quad (4.1)$$

The uncertainty $U^{1-\alpha}$ is defined as the empirical quantile q of order $1 - \alpha$ of the sorted nonconformity scores (Vovk *et al.*, 2005) :

$$U^{1-\alpha} = q_{1-\alpha}(\{s_i\}) \quad (4.2)$$

The confidence interval with coverage level $1 - \alpha$, denoted $I^{1-\alpha}(t)$, in the test set and associated with the component temperature $T_{comp}^{test}(t)$ at time $t \in [t_i^{test}, t_f^{test}]$ is defined as :

$$I^{1-\alpha}(t) = [T_{comp}^{test}(t) - U^{1-\alpha}, T_{comp}^{test}(t) + U^{1-\alpha}] \quad (4.3)$$

The confidence interval with nominal coverage $1 - \alpha$ guarantees this coverage only when the exchangeability assumption is satisfied (Vovk *et al.*, 2005). In practice, however, this condition is frequently violated. Many studies explicitly acknowledge that exchangeability is rarely met in real-world datasets (Oliveira *et al.*, 2024; Barber, Candès, Ramdas & Tibshirani, 2022; Benabdesselam *et al.*, 2026b). This issue is even more pronounced in PHM applications for industrial systems. Prognosis is performed after anomaly detection, which implies that degradation has already initiated and that the data exhibit a non-stationary drift. Once degradation is detected, future evolution is inherently driven by this drift, and therefore the calibration and test sets cannot be strictly exchangeable in the mathematical sense : degradation is thermodynamically irreversible and the drift is positive by definition.

Under such conditions, exchangeability is expected to be violated. Several techniques have been proposed to address non-exchangeability (Barber *et al.*, 2022; Oliveira *et al.*, 2024), although their deployment can be complex. The primary consequence of violating exchangeability is generally under-coverage, as demonstrated theoretically in Oliveira *et al.* (2024) and observed empirically in Benabdesselam *et al.* (2026b). Importantly, this should not be interpreted as a failure specific to the method, but rather as a limitation inherent to real-world prognostic settings. Even when exchangeability is violated, split CP may remain practically informative as long as the violation is not extreme. In Benabdesselam *et al.* (2026b), multiple uncertainty quantification approaches were evaluated on real degradation trajectories, including split CP. Despite the presence of drift and non-exchangeability, split CP achieved empirical coverages ranging from 83% (worst case) to 97.2% (best case) for a nominal level of 95%, even under accelerating degradation.

In the present case study, degradation is expected to progress relatively slowly due to thermal inertia (Cambron *et al.*, 2017). Accordingly, although exchangeability is not satisfied in the strict sense, the resulting confidence intervals are expected to exhibit mild under-coverage ; however, this should be interpreted as a heuristic expectation rather than a formal guarantee. Moreover, because the degradation drift is assumed to be positive in the present setting, exchangeability violation is expected to primarily manifest as under-coverage. In practice, stronger violations

may increase the likelihood of false negatives, which is consistent with the operational context considered here. If the thermal exceedance occurs near the calibration window, the calibration and test distributions are likely to remain more similar, which may facilitate detection. Conversely, if the exceedance occurs far from the end of the training period, calibration and test samples are expected to be less exchangeable, which may degrade detection performance. This behavior is consistent with the expected decrease in predictive accuracy as the degradation trajectory diverges from the calibration regime.

Finally, regarding the assumption of using historical data for exogenous variables and its influence on the CP confidence intervals, two situations must be distinguished :

- If last year's operational behavior is highly similar to the current year, the validity of the confidence interval depends primarily on the degree of exchangeability between the calibration and test datasets. The farther a sample lies from the thermal exceedance event, the larger the drift becomes, which may result in under-coverage and an increased number of false negatives. This behavior is expected for prognostic classifiers operating under progressive degradation.
- If the historical behavior differs substantially from the current year, the residuals in the calibration dataset become larger, which inflates the uncertainty. Consequently, the prediction intervals in the test set may widen and partially compensate for the structural under-coverage inherent to prognostic tasks. In the present case, behavioral discrepancies may degrade the temperature point prediction while potentially improving the conservativeness of the uncertainty estimates. Importantly, the objective is not point prediction ; only the confidence intervals are used to construct the event metrics.

4.4.3 Event Functions

The confidence interval $I^{1-\alpha}$ contains the true value with probability $1 - \alpha$ if the predictive uncertainty is well calibrated. At each time step t , confidence intervals are constructed for several confidence levels $1 - \alpha$. For a given level, if the upper bound of the interval intersects the threshold temperature $T_{threshold}$, the probability of exceeding the threshold is $\alpha/2$. A higher confidence

level $1 - \alpha$ corresponds to a smaller exceedance probability. Conversely, lower-confidence intervals (larger α) represent larger exceedance probabilities. This multi-level mechanism allows event intensity to be defined hierarchically. Starting from the most conservative interval (e.g., 95%) and progressively moving toward lower confidence levels (e.g., 90%, 80%, 70%), one identifies the largest α , denoted α^* , such that the upper bound still reaches $T_{threshold}$. This is motivated by the following : if only a high-confidence interval reaches the threshold, the exceedance probability is extremely low ; however, if a low-confidence interval still reaches it, the probability of exceedance is higher. The event therefore increases as the interval weakens.

The IEI function at time t is obtained by selecting this critical level α^* and converting its exceedance probability into a percentage, ensuring consistency regardless of the temperature threshold $T_{threshold}$ used, as presented in equation (4.4).

$$d(t) = \begin{cases} \frac{\alpha^*}{2} \cdot 100 & \text{if } \sup I^{1-\alpha^*}(t) = T_{threshold} \\ 0 & \text{if } \forall \alpha, \sup I^{1-\alpha}(t) < T_{threshold} \end{cases} \quad (4.4)$$

where :

- $I^{1-\alpha}(t) = [T_{comp}^{test}(t) - U^{1-\alpha}, T_{comp}^{test}(t) + U^{1-\alpha}]$ is the symmetric confidence interval calibrated at level $1 - \alpha$,
- $\sup I^{1-\alpha}(t) = T_{comp}^{test}(t) + U^{1-\alpha}$ denotes the upper bound of this interval,
- α^* is the largest level for which the upper bound exactly reaches $T_{threshold}$.

To capture event persistence, the IEI $d(t)$ is aggregated into the CEI $D(t)$, which reflects the cumulative intensity of the exceedance event over the prediction window. The IEI represents the time-varying exceedance tendency derived from the conformal confidence intervals, which themselves reflect the underlying thermal dynamics learned by the LSTM model. By integrating this time-dependent signal, the CEI preserves the progressive accumulation of exceedance tendency and therefore retains the temporal structure of degradation.

A single event threshold θ_c is then applied : when $D(t) > \theta_c$, an exceedance is considered likely, resulting in a binary thermal event classification. This binary decision is introduced solely for operational interpretability and does not replace the underlying dynamic event modeling. The event threshold θ_c was selected as a compromise between early detection and false alarms, maximizing the identification of known exceedance cases while limiting false positives. Although multi-region schemes (safe, gray, event) could be implemented, we adopt a strict single-threshold strategy to enable rigorous performance assessment.

The CEI is subsequently normalized by the prediction duration Δt_{test} , ensuring that the optimal event threshold remains consistent across different values of Δt_{test} . This improves the generalization of the approach and removes the need for sensitivity analysis with respect to Δt_{test} . The CEI function $D(t) \in \mathbb{R}^+$ is strictly increasing and is given by equation (4.5).

$$D(t) = \frac{1}{\Delta t_{test}} \sum_{\tau=t_i^{test}}^t d(\tau) \quad (4.5)$$

4.4.4 Evaluation Metrics

To evaluate the performance of the proposed classifier, the method is tested at multiple forecasting horizons before the actual observed temperature exceedance : 15 days, 1 month, 45 days, 2 months, 75 days, and 3 months. For positive cases, these horizons are anchored to the actual temperature-threshold exceedance time, whereas for negative cases, a reference anchor date is defined as the date at which the component temperature reaches its annual maximum, typically during summer. These horizons do not correspond to the test duration, but rather to the time interval between the end of training t_f^{train} and the actual threshold-exceedance event. This evaluation assesses the model's ability to predict whether the component will exceed the temperature threshold. Performances are summarized in a contingency table defining four outcome categories :

- True Positives (TP) : When the component actually exceeds the temperature threshold, and the model correctly predicts this exceedance.

- False Negatives (FN) : When the component actually exceeds the temperature threshold, but the model fails to detect this event.
- True Negatives (TN) : When the component does not exceed the temperature threshold, and the model correctly predicts the absence of exceedance.
- False Positives (FP) : When the component does not exceed the temperature threshold, but the model incorrectly predicts an exceedance.

The performance metrics used for the classification models include the True Positive Rate (TPR), False Negative Rate (FNR), True Negative Rate (TNR), False Positive Rate (FPR), Accuracy (ACC), F1-score, and the Matthews Correlation Coefficient (MCC), as listed in Powers (2011).

4.5 Results

The proposed methodology is evaluated on 207 WTs. To illustrate the approach, a case study of the main bearing is presented in Section 4.5.1. Sensitivity analyses of the hyperparameters are reported in Section 4.5.2, while the main classification results are summarized in Section 4.5.3. A quantitative justification of the historical exogenous input approximation is provided in Section 4.5.4, followed by a comparison with existing studies in Section 4.5.5.

4.5.1 Case Study : Main-Bearing Threshold Exceedance Event

This case study examines the main bearings of two WTs : WT-1, which exhibits an operating-temperature anomaly, and WT-2, which remains healthy. In this section, we present the temperature prediction for WT-1 only, as the same procedure applies to WT-2, followed by the computation of the IEI for both WTs. Finally, we compare the CEI of degraded WT-1 with that of healthy WT-2. For both WTs, the main-bearing temperature threshold is 63 °C, beyond which an automatic shutdown is triggered. In this case, the shutdown for WT-1 occurred on 2023-06-20. Assuming the date is 2023-06-05, which corresponds to 15 days before the temperature threshold is reached, the objective is to assess whether the temperature of the bearing is likely to exceed the temperature threshold in the future for both WTs. Here, we consider

$\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, and $\theta_c = 1$. These parameters are set a posteriori, and the sensitivity analysis leading to this choice is presented in Section 4.5.2.

For both WTs, the training dataset is selected from $t_i^{train} = 2023-04-05$ to $t_f^{train} = 2023-06-05$, and used to train the model with the corresponding values of T_{comp}^{train} , T_{amb}^{train} , U^{train} , and P^{train} . The future evolution of the component temperature is then predicted over the period from $t_i^{test} = 2023-06-05$ to $t_f^{test} = 2023-09-05$, using the exogenous variables $T_{amb}^{test-hist}$, $U^{test-hist}$, and $P^{test-hist}$ from the corresponding historical period in the previous year, that is from $t_i^{test-hist} = 2022-06-05$ to $t_f^{test-hist} = 2022-09-05$. The calibration step uses historical exogenous data $T_{amb}^{train-hist}$, $U^{train-hist}$, and $P^{train-hist}$, from $t_i^{train-hist} = 2022-04-05$ to $t_f^{train-hist} = 2022-06-05$. The trained model is applied to this period in order to predict the calibration component temperature T_{comp}^{calib} over the interval from $t_i^{train} = 2023-04-05$ to $t_f^{train} = 2023-06-05$. The predicted values of T_{comp}^{calib} are then compared to the actual temperatures T_{comp}^{train} to construct confidence intervals using CP. The training, calibration, and prediction phases are illustrated in Figure 4.7 for WT-1.

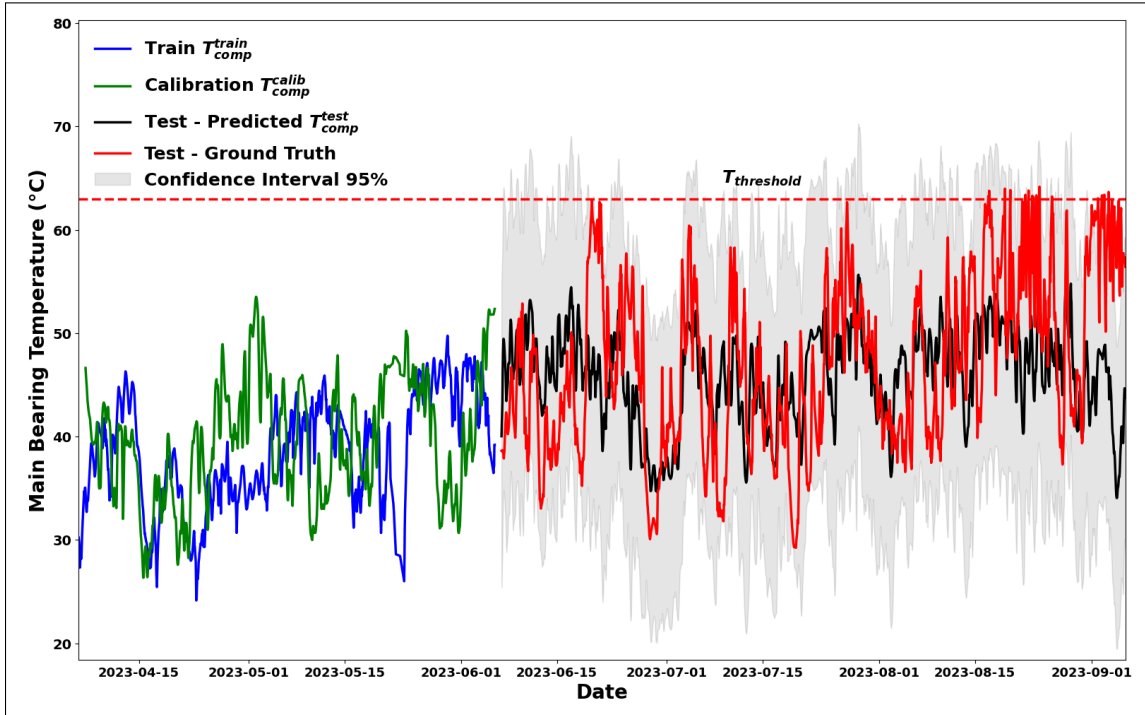


Figure 4.7 Prediction of WT-1 main-bearing temperature

The 95% confidence interval in Figure 4.7 encompasses the ground truth values despite deviations from the predicted values. These deviations arise from (i) the use of historical exogenous variables to approximate future conditions under the assumption of recurring seasonal patterns, and (ii) the difficulty of modeling a degrading system whose thermal behavior evolves over time, and the limited observability of internal dynamics from SCADA data further amplifies these challenges. Given the restricted forecast precision and weak early indicators, predicting the exact threshold-crossing time is beyond this approach's scope. Instead, the focus is placed on assessing temperature exceedance events using the IEI function derived from the confidence intervals $I^{1-\alpha}(t)$, as shown in Figure 4.8 for WT-1 and WT-2. The aggregation of the IEI function yields the CEI function, also illustrated in Figure 4.8, which compares the CEI of WT-1, representing a degraded state, with that of WT-2, representing a healthy state. For WT-1, the CEI exceeds the event threshold $\theta_c = 1$ (set a posteriori), leading to the conclusion that the main bearing is likely to exceed the critical temperature. However, for WT-2, the CEI remains well below θ_c , indicating that no temperature-threshold exceedance is expected in the future.

4.5.2 Sensitivity Analyses

In this section, we clarify the choice of the hyperparameters Δt_{train} , Δt_{test} , and θ_c . In our framework, the prediction window Δt_{test} only specifies how far into the future the user wants to assess a potential temperature exceedance. Changing Δt_{test} does not alter the classifier itself; it changes the question being asked. Checking whether a threshold will be exceeded in one, two, or three months corresponds to different prognostic tasks with different operational horizons. The CEI is normalized by the window length, so the CEI quantifies the persistence of exceedance tendencies rather than the amount of time available to accumulate them. Without this normalization, the event threshold θ_c would need to be tuned separately for each Δt_{test} . With normalization, a single θ_c remains valid across all horizons. Because each Δt_{test} defines a different target, varying it does not test any invariance of the metric – it creates a multi-horizon (effectively multi-class) classification setting. This formulation is undesirable for SCADA-based monitoring, where multi-class models have shown limited robustness and poor generalization

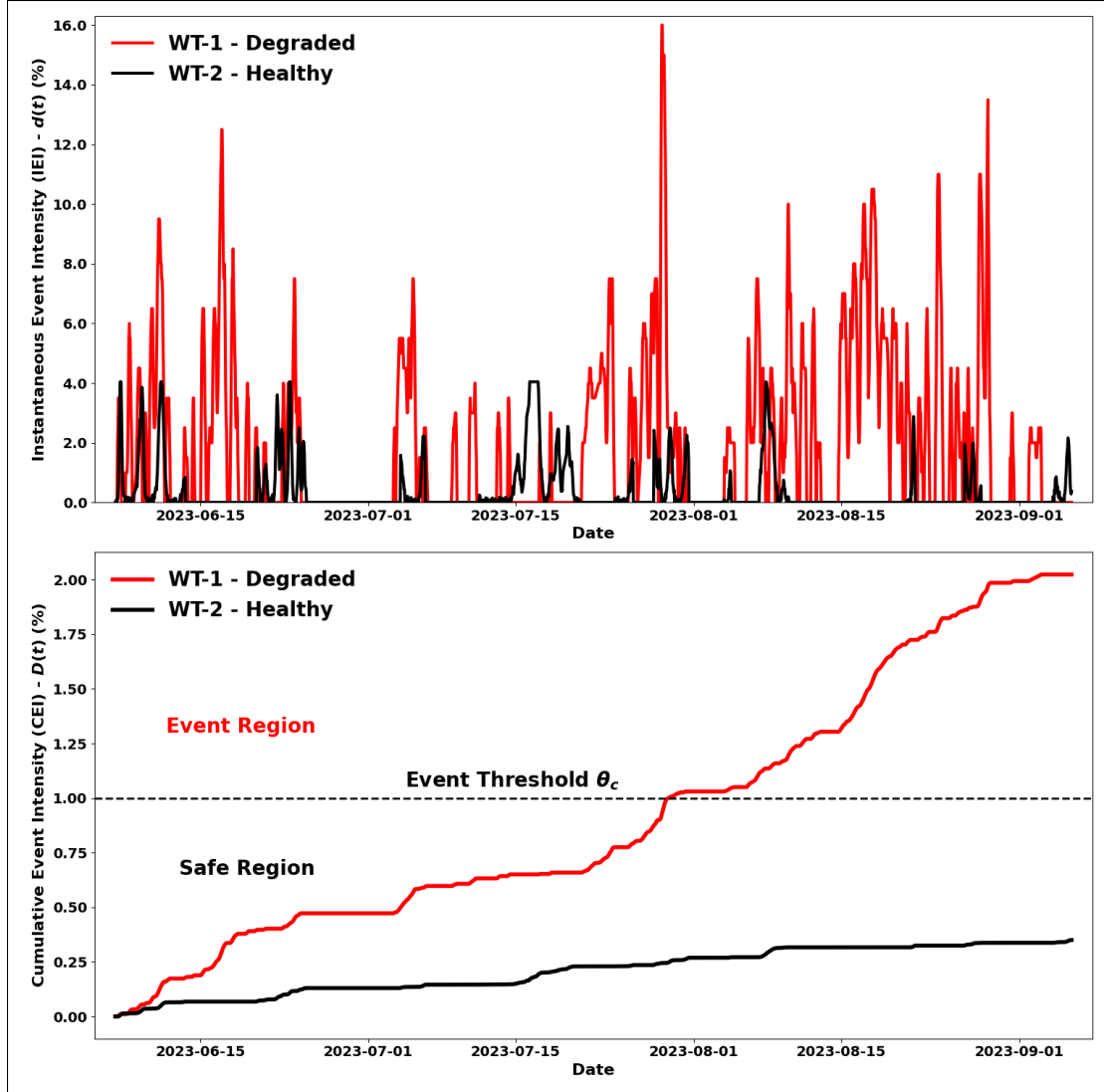


Figure 4.8 Evolution of IEI and CEI over time

across turbines and operating regimes (Carroll *et al.*, 2019). Our objective is strictly binary : determine whether the asset will exceed its critical temperature in the future, not when or at which specific horizon. To avoid mixing methodological effects with horizon redefinition, Δt_{test} is kept fixed in all experiments.

In practice, more than 67% of overheating events in our industrial dataset occur within less than three months after anomaly onset. Consequently, in this study, we fixed $\Delta t_{test} = 3$ months, as this value provides a sufficiently large assessment interval to capture the majority of events,

while ensuring that θ_c remains directly interpretable across all anticipation horizons. Detection performance is then evaluated at anticipation horizons of 15, 30, 45, 60, 75, and 90 days within this fixed window, focusing exclusively on the temporal dynamics of the event signal rather than on arbitrary changes in the prediction window length.

The second hyperparameter to be examined is the event threshold θ_c . For this analysis, we set $\Delta t_{train} = 2$ months and $\Delta t_{test} = 3$ months. For the 207 cases, the proposed methodology is applied using different θ_c values ranging from 0.5 to 2 in order to determine the optimal event threshold. The selection of this optimal value is based on a statistical optimization criterion, namely the maximization of the MCC. Specifically, for each θ_c candidate, we compute the MCC at each prediction horizon, and then derive the mean MCC associated with that θ_c . The resulting values are subsequently compared across all candidates, as presented in Table 4.3.

Tableau 4.3 Matthews Correlation Coefficient (MCC) for different θ_c values across prediction horizons

	Prediction horizon (days)						Mean
θ_c	15	30	45	60	75	90	
0.5	0.65	0.56	0.43	0.46	0.46	0.38	0.49
1.0	0.67	0.60	0.51	0.45	0.43	0.37	0.51
1.5	0.68	0.61	0.53	0.43	0.37	0.33	0.49
2.0	0.64	0.64	0.51	0.45	0.37	0.30	0.48

Referring to Table 4.3, the MCC reaches its maximum at $\theta_c = 1$; this value was therefore retained for the present study based on the available dataset.

The last hyperparameter to be optimized is the training duration Δt_{train} . Accordingly, with $\Delta t_{test} = 3$ months and $\theta_c = 1$ fixed, we vary Δt_{train} from 1 to 4 months in order to determine its optimal value through statistical optimization. For each candidate Δt_{train} , the MCC is computed at every prediction horizon (from 15 to 90 days), and the resulting values are averaged to obtain a mean MCC associated with that Δt_{train} . The mean MCCs are then compared across all

candidates, and the training duration that maximizes the MCC is selected, as presented in Table 4.4.

Tableau 4.4 Matthews Correlation Coefficient (MCC) for different Δt_{train} values across prediction horizons

	Prediction horizon (days)						Mean
Δt_{train}	15	30	45	60	75	90	
1	0.60	0.53	0.47	0.43	0.42	0.35	0.46
2	0.67	0.60	0.51	0.45	0.43	0.37	0.51
3	0.62	0.53	0.44	0.40	0.37	0.33	0.45
4	0.61	0.51	0.41	0.37	0.34	0.30	0.42

Referring to Table 4.4, the MCC reaches its maximum at $\Delta t_{train} = 2$ months ; this value was therefore retained for the present study based on the available dataset.

4.5.3 Main Classification Performances

For the optimal hyperparameters $\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, and $\theta_c = 1$, the confusion matrices at different prediction horizons are presented in Table 4.5.

Tableau 4.5 Classification outcomes for multiple prediction horizons with optimal hyperparameters ($\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, $\theta_c = 1$)

	Prediction horizon (days)					
Class	15	30	45	60	75	90
TP	89	77	68	58	54	49
TN	83	85	82	82	82	80
FP	22	18	21	21	20	21
FN	13	22	28	31	33	38

It should be noted that, as the prediction horizon increases, data availability decreases for some WTs, reducing the number of evaluable cases from 207 at 15 days to 188 at 90 days. Nevertheless, events detected at longer horizons remain detected at shorter horizons ; for example, if an

exceedance event is detected 90 days in advance, it is also detected 15 days in advance. Based on the confusion matrices reported in Table 4.5, the corresponding classification metrics across the different forecasting horizons are summarized in Table 4.6.

Tableau 4.6 Performance metrics for multiple prediction horizons with optimal hyperparameters ($\Delta t_{train} = 2$ months, $\Delta t_{test} = 3$ months, $\theta_c = 1$)

	Prediction horizon (days)					
Metric	15	30	45	60	75	90
TPR	87.25	77.78	70.83	65.17	62.07	56.32
FNR	12.75	22.22	29.17	34.83	37.93	43.68
TNR	79.05	82.54	79.61	79.61	80.39	79.21
FPR	20.95	17.46	20.39	20.39	19.61	20.79
ACC	83.10	80.20	75.37	72.91	71.95	68.61
F1-score	83.57	79.38	73.12	68.24	67.08	62.82

Table 4.6 shows a clear improvement in predictive performance as the threshold exceedance event approaches. At a 90-day horizon, the model correctly identifies 56.32% of positive events, as indicated by the TPR metric. As the prediction window shortens, detection performance steadily improves : the model captures about 60% of events two months before the temperature threshold is reached, more than 75% within the last month, and 87.25% just 15 days before failure. This means that the model misses 43.68% of positive events three months before the exceedance, as shown by the FNR, but misses only 12.75% 15 days before. The model maintains an FPR of around 20% across all horizons, which is acceptable and corresponds to a TNR of about 80%, meaning that the model correctly identifies approximately 80% of negative events and thereby limits the occurrence of false alarms. The model accuracy increases from 68.61% at a 90-day horizon to 83.10% at a 15-day horizon, reflecting the model's improved ability to correctly classify both positive and negative events as the failure date approaches. Finally, the model F1-score, which is particularly suitable for classification under class imbalance, increases from 62.82% at the 90-day horizon to 83.57% at the 15-day horizon, indicating good classification performance despite the class imbalance. This improvement in performance as the prediction horizon approaches the true event date can be explained by the fact that the calibration and test

data become increasingly exchangeable. As a result, the confidence intervals are better able to cover the true T_{comp} , thereby moving closer to the theoretical coverage. For example, for a theoretical coverage level of 95%, the average empirical coverage across all WTs was 82% at 90 days, 83% at 75 days, 84% at 60 days, 86% at 45 days, 88% at 30 days, and 91% at 15 days.

Thus, given that the dataset is strongly imbalanced, as shown in Table 4.2, and to demonstrate the generalizability of our approach, we report the classification outcomes for the three studied components in Table 4.7.

Tableau 4.7 Classification outcomes per component across prediction horizons

	Prediction horizon (days) – Main Bearing					
Class	15	30	45	60	75	90
TP	75	66	59	49	45	43
TN	36	36	35	34	35	35
FP	9	7	8	9	7	6
FN	12	18	22	25	27	29
	Prediction horizon (days) – Generator Bearing					
Class	15	30	45	60	75	90
TP	10	8	6	6	6	4
TN	42	43	42	43	42	40
FP	12	11	12	11	12	14
FN	1	3	5	5	5	7
	Prediction horizon (days) – Generator					
Class	15	30	45	60	75	90
TP	4	3	3	3	3	2
TN	5	6	5	5	5	5
FP	1	0	1	1	1	1
FN	0	1	1	1	1	2

For the main bearing, the TPR is high at short horizons (86.2% at 15 days) and decreases gradually as the horizon extends (down to 59.7% at 90 days). The FPR remains consistently low

across all horizons (between 13.0% and 14.3%), meaning the model rarely triggers false alarms even when predicting far ahead. For the generator bearing, the TPR is 90.9% at 15 days and 54.54% at 75 days, but drops to 36.4% at 90 days, meaning that at very long horizons many alarms are missed, although from 75 days onward, which still remains a long horizon, 54.5% of positive events are already detected. Thus, the FPR remains moderate at around 23%. The generator presents the most favorable alarm behavior. The TPR at 15 days is 100% and decreases with time to 75% at 75 days and reaches 50% at 90 days, which represents a long prediction horizon, while the FPR remains very low, between 0% and 16.7%.

4.5.4 Historical Exogenous Input Approximation

The proposed methodology relies on the assumption of using historical exogenous inputs during the prediction phase.

A first support for this assumption was provided in Section 4.4.2, where the discrepancies between actual and projected exogenous variables were quantified. The nRMSE was approximately 14.62% for T_{amb} , 20.01% for U , and 44.00% for P . Although the discrepancy for P is relatively large, sensitivity analysis using a Random Forest model showed that T_{comp} is substantially less sensitive to P , which contributes on average only about 20% to the prediction of T_{comp} , whereas T_{amb} and U account for approximately 50% and 30%, respectively. Similar observations have been reported in the literature for WT thermal modeling. This indicates that the structural mismatch in P has a limited direct impact on temperature prediction.

A second rationale lies in the conformal calibration procedure. The calibration dataset, used for uncertainty quantification, is constructed by substituting the exogenous variables with those from the previous year while keeping the original training targets T_{comp}^{train} . The model therefore predicts T_{comp}^{calib} using historical exogenous inputs, and the deviation between T_{comp}^{calib} and T_{comp}^{train} is used to compute the nonconformity scores. As a result, the structural discrepancy induced by the historical approximation is inherently embedded in the calibration residuals and consequently reflected in the confidence intervals. When inter-annual variability is high,

the prediction of T_{comp}^{test} may be less precise; however, the deviation between T_{comp}^{calib} and T_{comp}^{train} becomes larger, leading to increased nonconformity scores and wider confidence intervals. In this case, the uncertainty naturally compensates for the reduced pointwise accuracy. Conversely, when inter-annual variability is low, calibration residuals remain small, resulting in tighter prediction intervals and higher predictive precision. This compensation mechanism suggests that the CEI-based decision can account, at least partially, for the variability induced by the historical approximation rather than ignoring it.

The third basis supporting this assumption is presented in this section, where the confusion matrices from 15 to 90 days, shown in Table 4.8, are obtained using the observed exogenous data (ground truth) from t_i^{test} to t_f^{test} (available a posteriori) and compared with those obtained using historical data. The uncertainty quantification procedure remains unchanged and is still performed with the historical exogenous inputs.

Tableau 4.8 Confusion matrices values at different prediction horizons for optimal hyperparameters with observed exogenous data

	Prediction horizon (days)					
Class	15	30	45	60	75	90
TP	91	78	72	61	56	52
TN	86	89	87	86	86	85
FP	19	14	16	17	16	16
FN	11	21	24	28	31	35

The comparison between Table 4.5 and Table 4.8 (predictions using historical and real exogenous data, respectively) reveals only a limited performance loss when historical exogenous data are used. For a 15-day horizon, 87.25% of positive events are detected with historical inputs versus 89.21% with observed inputs, while the false positive rate is 20.95% compared to 18.09%. At 90 days, the true positive rate is 56.32% with historical inputs and 59.77% with observed inputs, with false positive rates of 20.79% and 15.84%, respectively.

The comparison between historical exogenous variables from the previous year and the true future exogenous values (as if they were available in advance) shows only a marginal decrease in performance across all horizons. This demonstrates that the historical approximation remains practically acceptable in the present context.

Since the performance improvement induced by perfect exogenous information remains limited in our experiments, more elaborate alternatives relying on exogenous prediction may provide only marginal gains in this setting. In the present study, these gains do not appear to justify the additional computational, modeling, and deployment costs required to implement such approaches in an industrial PHM context. Consequently, the approximation of exogenous variables using historical profiles appears to be a reasonable practical choice within the operational context considered in this study.

4.5.5 Comparison with Existing Studies

Direct comparison with existing studies remains challenging, as the proposed approach formulates prognostics as a classification problem, whereas the vast majority of published methods address RUL prediction as a regression task. These two problem formulations are inherently different and rely on distinct performance metrics, making quantitative comparison incompatible.

Moreover, the few classification-based approaches reported in the literature often focus on anomaly detection or fault diagnosis rather than true prognostic decision-making. Many are validated on a limited number of cases, while others adopt problem definitions that differ substantially from the threshold-triggered shutdown setting considered here. As a result, a direct benchmark against most existing studies would not provide a meaningful or fair comparison. Nevertheless, we identified three recent works with sufficiently comparable problem formulations to enable a partial evaluation.

Weissenfeld *et al.* (2025) evaluated their proposed anomaly detection approach on a dataset comprising 7 failure cases. No false negatives were reported, indicating high sensitivity. However, the approach produced 1461 false positives, resulting in a precision below 0.5% and thus severely

limiting its practical applicability. In contrast, our method achieves a better balance by preserving high sensitivity while substantially reducing false positives.

Guck *et al.* (2024) developed a classification-based anomaly detection model validated on 44 anomaly events and 51 normal events, achieving a coverage metric (generalized F1-score) of 0.46 and a TNR of 0.88. Using their coverage formula, our results range from 0.63 at a 3-month horizon to 0.81 at 15 days, while TNR values remain between 0.79 and 0.82, close to theirs. While their method produces fewer false positives, our model identifies up to twice as many true positives.

Carroll *et al.* (2019) developed an AI-based classifier to detect gear bearing and gear tooth failures, distinguishing between a critical state (failure within one month) and a healthy state (at least one year before failure or no failure) using SCADA data. In their protocol, correct predictions correspond to the proportion of correctly classified cases (TP + TN), missed failures to the proportion of FN, and false positives to the proportion of FP. Applying the same formulation to our algorithm with $\theta_c = 1$ at a one-month horizon and $\Delta t_{test} = 1$ month shows that our approach achieves 80.2% correct predictions, compared to 72.5% for gear bearing and 75.0% for gear tooth. The proportion of missed failures is also lower in our case (10.9%) than for gear bearing (20.0%) and gear tooth (13.0%). In terms of false positives, our algorithm yields 8.9%, which is 1% higher than gear bearing but still below gear tooth (12.0%).

4.6 Conclusion

This paper proposed a SCADA-based prognostic classification framework for thermal anomalies in WT components, representing a step toward SCADA-based RUL estimation, an open challenge that remains to be addressed. Combining LSTM with CP, the method derives a CEI that quantifies the exceedance event. When the CEI surpasses a defined event threshold, a binary decision is issued. The approach was validated on 207 industrial WTs, including 102 positive and 105 negative cases, across three manufacturers, 20 wind farms, and three components, substantially exceeding the scale of prior studies. Results showed strong predictive performance, with 87.25%

of overheating events correctly anticipated 15 days in advance while maintaining a false positive rate near 20%. Even at a 90-day horizon, 56.32% of critical events were predicted, demonstrating reliable early-warning capabilities. Moreover, the model maintains an FPR of around 20% across all horizons, indicating that it correctly identifies 80% of negative events and thereby limits the occurrence of false alarms.

Beyond accuracy, the framework offered three key advantages for large-scale deployment : broad applicability, as it relies exclusively on SCADA data already collected in most wind farms without requiring additional sensors ; computational simplicity, allowing training and deployment with moderate resources ; and robustness, confirmed by extensive validation on diverse fleets and components. From an operational standpoint, these advantages enable asset managers to prioritize maintenance, reduce false alarms, and anticipate interventions before unplanned downtime. This supports more efficient planning of spare parts, workforce, and outage scheduling, ultimately reducing O&M costs and mitigating production losses.

Limitations include the inability to estimate the exact date of temperature threshold exceedance, which remains an open challenge with SCADA data, and the requirement for at least one year of historical SCADA records prior to deployment. In practice, most WTs are already equipped with SCADA systems, so this constraint is mainly relevant when deploying the approach in newly commissioned wind farms, where historical measurements of ambient temperature, wind speed, and power may not yet be available. In such cases, historical weather data for the site can be used to reconstruct ambient temperature and wind speed, while numerical simulations based on the WT model can be employed to approximate the corresponding power output.

CHAPITRE 5

REMAINING USEFUL LIFE PREDICTION FOR GRADUAL THERMAL ANOMALIES IN WIND TURBINES USING SCADA DATA

Abderraouf Benabdesselam¹, Monelle Comeau², Martin Gagnon³, Antoine Tahan¹

¹ Département de Génie Mécanique, École de Technologie Supérieure, 1100 Rue Notre-Dame O., Montréal, QC H3C 1K3, Canada;

² Power Factors, 7005 Boulevard Taschereau, Brossard, QC J4Z 1A7, Canada;

³ Hydro-Québec, 75 Boulevard René-Lévesque O., Montréal, QC H2Z 1A4, Canada.

Article soumis pour publication à la revue « Reliability Engineering & System Safety » en février 2026.

5.1 Abstract

This study introduces a SCADA-based framework for predicting the remaining useful life of wind turbine components, enabling the anticipation of gradual, overheating-related degradation processes observable at SCADA resolution. To ensure physical consistency and interpretability, a health indicator is derived from temperature-rise signals using a cumulative-maximum filtering strategy that enforces monotonic behavior over time. On this basis, several artificial intelligence models are evaluated to learn the relationship between SCADA variables and remaining useful life, among which a gated recurrent unit emerges as the most effective architecture. Predictive uncertainty is subsequently quantified using conformalized quantile regression, providing calibrated confidence intervals without reliance on distributional assumptions. The proposed framework is validated on operational data from 28 main bearings spanning multiple wind farms and manufacturers. The results show an average RMSE of 26.54 days and an R^2 of 88.43%, while maintaining robust uncertainty calibration for prediction horizons exceeding three months, thereby demonstrating the practical viability of SCADA-only prognostics for fleet-scale deployment.

5.2 Introduction

Wind energy has experienced remarkable growth over the past decade, with the global installed capacity of Wind Turbines (WTs) nearly tripling (IRENA, 2023). To keep wind

power economically competitive, a major challenge lies in reducing operation and maintenance costs, which account for almost 30% of the total levelized cost of energy (Dao *et al.*, 2019; Santolamazza *et al.*, 2021). In this context, Prognostics and Health Management (PHM) has become essential for improving reliability and extending component lifetime by supporting informed maintenance decisions (Benabdesselam *et al.*, 2024). PHM typically involves 3 key tasks : anomaly detection, fault diagnosis, and health state prognosis (Javed *et al.*, 2017). While anomaly detection (Martínez Montaña & Sanz-Bobi, 2018; Meyer, 2022; Cambron *et al.*, 2017; Campoverde-Vilela *et al.*, 2023; Beretta *et al.*, 2021; Jankauskas *et al.*, 2023; Oliveira-Filho *et al.*, 2024, 2023) and fault diagnosis (Deng *et al.*, 2017; Zhang *et al.*, 2023c; Tang *et al.*, 2021; Alagha *et al.*, 2025; Zhang *et al.*, 2023a; Liu *et al.*, 2024) have reached a relatively mature state, reliable Remaining Useful Life (RUL) prediction remains a major research challenge (Zio, 2022).

Once an anomaly is detected, the aim is to estimate the RUL and its uncertainty, defined as the time remaining before the system reaches a failure threshold (Liu *et al.*, 2012), as illustrated in Figure 5.1.

Most prognostic studies focus on vibration data, as these signals capture high-frequency dynamics directly associated with degradation mechanisms. Traditional RUL prediction approaches manually extract statistical features (e.g., root mean square, kurtosis, variance) to construct a Health Indicator (HI), which is then extrapolated to a predefined failure threshold using, for example, Kalman filters (Zhang *et al.*, 2023c; Jin *et al.*, 2019; Cui *et al.*, 2020), Recurrent Neural Networks (RNN) (Guo *et al.*, 2017), multivariate extreme learning machines (Pan *et al.*, 2020b), relevance vector machines (Wang *et al.*, 2020a), or support vector machines (Yan *et al.*, 2020). More recently, deep learning models such as Convolutional Neural Networks (CNN) (Guo *et al.*, 2018; Yoo & Baek, 2018; She & Jia, 2019; Ren *et al.*, 2018; Li *et al.*, 2019; Jiang *et al.*, 2022; Cheng *et al.*, 2020), and Convolutional Autoencoders (CAE) with monotonicity constraints (Chen *et al.*, 2021; Milani *et al.*, 2025b) have been used to construct HIs directly from raw signals.

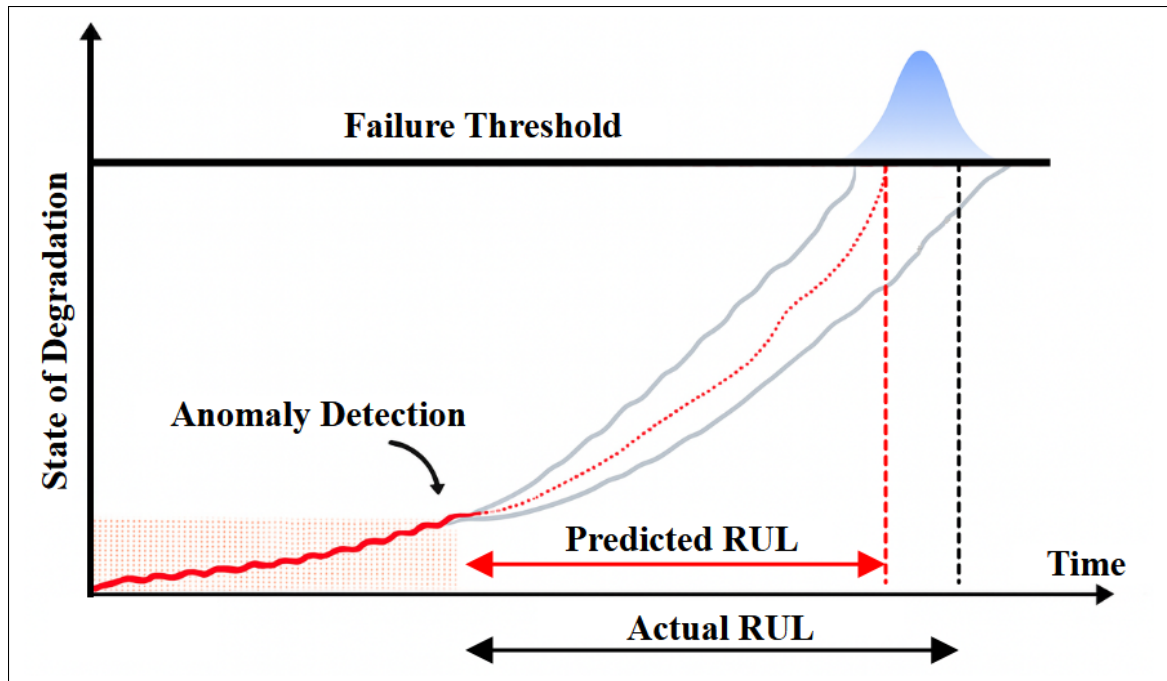


Figure 5.1 Illustration of RUL prediction
Adaptée de Javed *et al.* (2017, p. 219)

Although vibration measurements provide direct insight into degradation mechanisms, their use in WTs remains limited (Rezamand *et al.*, 2020). Condition Monitoring Systems (CMS), which provide vibration, acoustic emission, and lubrication measurements, are rarely deployed across wind farms, and when installed, vibration acquisition is often restricted to short measurement campaigns (Moreira de Oliveira Filho, 2024; Wang *et al.*, 2026). Consequently, relying solely on vibration data for condition monitoring or RUL prediction is generally impractical, limiting the real-world applicability of many vibration-based prognostic approaches (Moreira de Oliveira Filho, 2024; Yan *et al.*, 2025; Maldonado-Correa *et al.*, 2020). In contrast, Supervisory Control and Data Acquisition (SCADA) systems are systematically integrated in commercial WTs to regulate operation according to weather conditions and grid requirements (Yan *et al.*, 2025; Maldonado-Correa *et al.*, 2020). They continuously collect and store key operational variables (e.g., temperatures, currents, active power, wind speed), typically aggregated as 10-minute averaged time series (Rezamand *et al.*, 2020; Moreira de Oliveira Filho, 2024). Their widespread availability enables continuous monitoring at fleet scale without additional instrumentation,

making SCADA a cost-effective alternative to CMS. Although primarily designed for performance monitoring, several studies have also shown that the degradation of critical components can be reflected in measurable changes within SCADA variables (Yan *et al.*, 2025; Maldonado-Correa *et al.*, 2020). Therefore, SCADA data contain valuable health information, supporting effective condition monitoring and establishing SCADA-based prognostics as a practical and compelling solution for real-world deployments (Rezamand *et al.*, 2020; Moreira de Oliveira Filho, 2024).

5.2.1 SCADA-Based Prognostics

Only a limited number of studies have addressed prognostics using SCADA data, given the inherent challenges of RUL estimation and the low temporal resolution of SCADA signals (typically 10-minute averages). Existing approaches can be categorized into two groups : prognostic classifiers, which detect incipient failures or assign the RUL to discrete intervals, and continuous RUL predictors, which derive an HI, either explicitly or implicitly, before estimating the RUL.

For classifiers, Carroll *et al.* (2019) used a neural network to predict gearbox failures up to one month in advance. Gomes *et al.* (2024) applied a Random Forest to classify RUL into four ranges (7-60 days) with over 80% accuracy. We recently proposed a deep learning classifier with Conformal Predictions (CPs) (Benabdesselam, Comeau, Oliveira-Filho, Gagnon & Tahan, 2026a), predicting 87% of thermal degradations 15 days before shutdown, with 20% false positives, while maintaining predictive capability up to 3 months ahead.

For continuous RUL prediction, Rezamand *et al.* (2021) combined SCADA and vibration data within a probabilistic framework, where SCADA was used to cluster operating states and vibration features implicitly defined the HI for RUL estimation. Oliveira-Filho *et al.* (2021) employed a variational autoencoder to project SCADA data into a latent space, where degradation appears as a trajectory from healthy to degraded regions, and RUL is estimated by extrapolating this trajectory until it reaches the degraded cluster. Vieira *et al.* (2024) derived an HI by computing the temperature difference between the main bearing and the ambient air, and, together with wind

speed and active power, these SCADA variables were used as inputs to regression models for RUL prediction. Shah *et al.* (2024) used multivariate SCADA data and proposed a hybrid CNN-RNN model with an attention mechanism, bypassing feature engineering to forecast WT RUL. Milani *et al.* (2025a) constructed an HI from SCADA data using a CAE with monotonicity constraints, and predicted RUL by extrapolating the HI with a quadratic model to a failure threshold. Hu *et al.* (2025) developed a degradation model based on a multi-head RNN architecture, which decouples SCADA time series into multiple temporal scales to capture degradation dynamics and uses these as inputs to predict RUL with quantified uncertainty.

5.2.2 Key Gaps in the Literature

Despite progress, SCADA-based RUL prediction still faces critical limitations :

- **Delayed degradation visibility and limited prognostic horizons :** SCADA signals generally detect anomalies only at relatively advanced stages, making early prognostics difficult (Wang *et al.*, 2026; Hu *et al.*, 2025). Many methodologies predict failures only shortly before they occur. In some cases, the prognostic horizon does not exceed 2 weeks (Shah *et al.*, 2024) and can be as short as 2 days (Hu *et al.*, 2025).
- **Challenges in HI construction :** SCADA systems do not measure degradation directly but record operational variables influenced by seasonal effects, load variations, and stochastic fluctuations. Extracting a trendable and monotonic HI is therefore challenging (Hu *et al.*, 2025), as it must reflect the irreversible nature of deterioration (Javed *et al.*, 2017). For example, component temperatures may temporarily decrease due to ambient variations, falsely suggesting recovery despite ongoing degradation. More fundamentally, not all degradation mechanisms are observable at SCADA resolution : some processes remain hidden and therefore appear as sudden faults at the SCADA level, limiting predictability (Wang *et al.*, 2026).
- **Limited methodological validation :** Most studies rely on a small number of WTs, typically between 1 and 7 cases (Gomes *et al.*, 2024; Hu *et al.*, 2025; Milani *et al.*, 2025a; Rezamand

et al., 2021; Oliveira-Filho *et al.*, 2021; Vieira *et al.*, 2024; Shah *et al.*, 2024), which limits generalization and large scale deployment.

- **Limited uncertainty quantification** : Decision-making requires properly calibrated confidence estimates, which is rarely addressed (Zio, 2022).
- **Complex architectures with limited scalability** : Several methods rely on heavy deep learning pipelines that are difficult to implement in industrial settings (Shah *et al.*, 2024; Hu *et al.*, 2025).

These shortcomings provide ample motivation for further research on SCADA-based RUL prediction frameworks supported by uncertainty quantification with broad validation across multiple assets.

5.2.3 Problem Statement

This study focuses on WTs equipped with built-in alarm systems that automatically shut down the turbine once a predefined temperature threshold is exceeded. The threshold value is model-dependent and varies across manufacturers. Within this framework, the operational end of life (corresponding to zero RUL) is defined with respect to the first operationally critical thermal shutdown, which constitutes a relevant failure event from a PHM and maintenance decision-making perspective, even though it does not necessarily correspond to the physical end-of-life of the component.

The overheating must originate from physical degradation of the component, as the scope of this work is restricted to gradual thermomechanical degradation processes that exhibit trendable thermal drift at SCADA resolution. The method is not designed to address sudden faults, electrical anomalies, lubrication incidents, transient events, or any unpredictable degradation patterns in the SCADA data. Once an anomaly is detected and the observed thermal deviation is assessed, using the approach described in (Benabdesselam *et al.*, 2026a), as likely to lead to a future threshold exceedance, and provided that the component degradation is predictable at the SCADA time scale, the objective becomes to estimate the component's RUL.

5.2.4 Objectives and Contributions

The objective of this paper is to develop a SCADA-only framework for predicting the RUL of WT components once an anomaly has been detected, without requiring additional instrumentation or recalibration campaigns. All stages, from anomaly detection to HI construction, RUL estimation and uncertainty quantification, are compatible with real-time integration into existing asset-management pipelines. The original contributions of this work are as follows :

- Identification and formalization of the conditions under which thermal degradation processes are observable and predictable using SCADA data, contributing to a more rigorous formulation of SCADA-based PHM.
- Construction of an HI directly from temperature data, without requiring any dedicated feature extraction architecture or post-processing.
- Development of a RUL prediction tool for WT components using only SCADA data within a simple architecture.
- Anticipation of gradual thermal degradation processes more than 3 months in advance.
- Robust quantification of predictive uncertainty.
- Extensive validation was conducted across 28 different WTs, covering various failure types, wind farm locations, WT models, and manufacturers.

The remainder of this paper is organized as follows. Section 5.3 details the proposed methodology ; Section 5.4 presents and discusses the results ; and Section 5.5 concludes the paper.

5.3 Methodology

The methodology presented in Figure 5.2 consists of four main steps. First, the HI construction and data preprocessing are presented in Section 5.3.1. Second, the transfer function selection is described in Section 5.3.2. Third, the uncertainty quantification procedure is detailed in Section 5.3.3. Finally, the evaluation metrics are presented in Section 5.3.4.

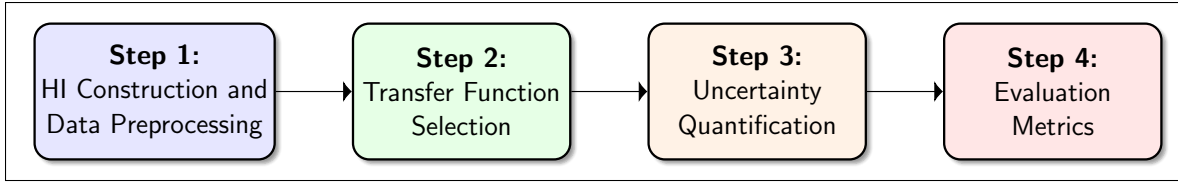


Figure 5.2 RUL prediction methodology for WT components

5.3.1 Health Indicator Construction and Data Preprocessing

Because degradation in this context manifests as overheating, the selection of variables was guided by a thermodynamic formulation of bearing temperature behavior, following the approach of Cambron *et al.* (2017). The SCADA variables considered are T_{comp} , T_{amb} , U , and P , as summarized in Table 5.1. A reference temperature T_{ref} under healthy operating conditions is also included. In this study, T_{ref} is defined as the arithmetic mean temperature across the wind farm for each component, under the assumption that the fleet is generally healthy on average. Alternatively, T_{ref} can be estimated using a model of healthy operating behavior, as proposed by Cambron *et al.* (2017). The dataset also includes the anomaly detection date, referred to as the First Prediction Time (FPT), and the threshold exceedance date, referred to as the Failure Threshold Time (FTT), the latter corresponding to the automatic shutdown event.

Tableau 5.1 Description of the key variables used in the study

Measure	Unit	Description
T_{comp}	°C	Measured temperature of the component
T_{ref}	°C	Reference temperature under healthy operating conditions
T_{amb}	°C	Ambient temperature
U	m/s	Wind speed
P	kW	Active electrical power

Anomaly detection is performed using temperature measurements. Under healthy conditions, T_{comp} remains close to T_{ref} . The anomaly indicator ΔT is defined as the temperature difference :

$$\Delta T = T_{comp} - T_{ref} \quad (5.1)$$

ΔT follows a Gaussian white noise distribution $N(0, \sigma^2)$ in the healthy state. Once degradation initiates, ΔT deviates from stationarity, indicating the presence of an anomaly, as shown in Figure 5.3.

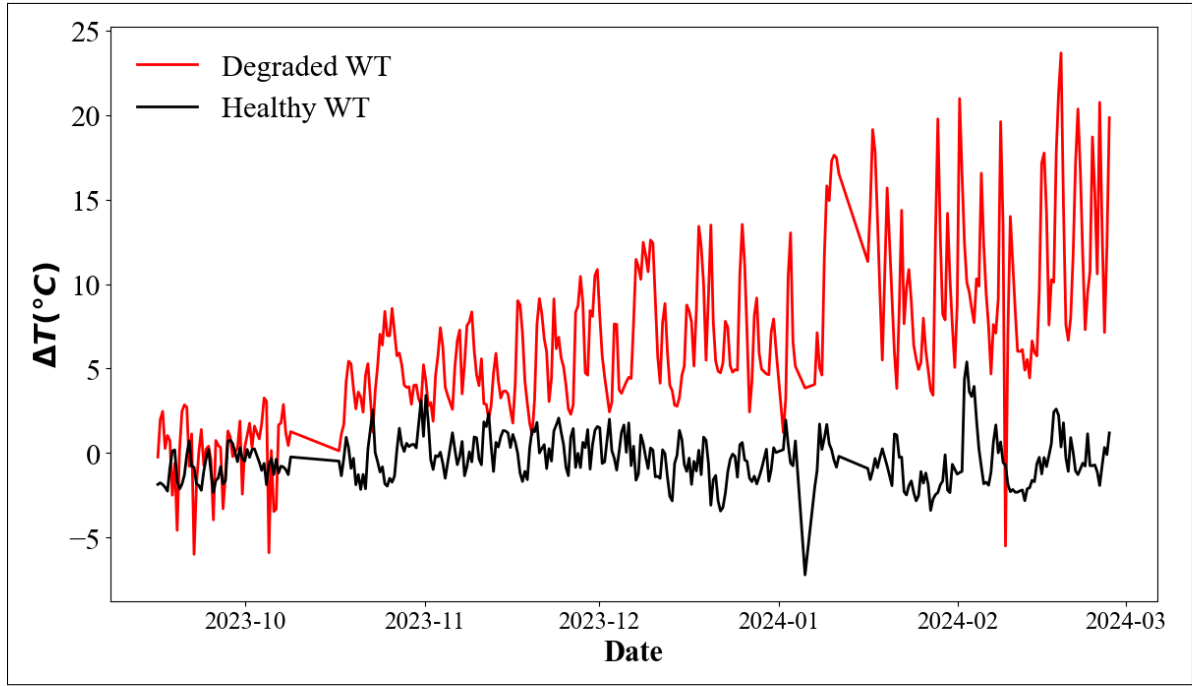


Figure 5.3 ΔT time series for a healthy and a degraded WT

The proposed method is not intended to address sudden faults, electrical anomalies, lubrication incidents, or other transient disturbances. Its applicability is limited to gradual thermomechanical degradation that induces a trendable thermal drift at SCADA resolution, as reported in (Wang *et al.*, 2026). Accordingly, reliable RUL prediction requires ΔT to first capture the irreversible nature of degradation (Javed *et al.*, 2017; Zio, 2022; Rezamand *et al.*, 2020; Wang *et al.*, 2026; Mai, Fouladirad & Xue, 2026) by exhibiting a globally increasing trend over the FPT-FTT interval. While strict pointwise monotonicity cannot be enforced due to inherent fluctuations in SCADA data, the global trend is assessed using the Sen slope, a robust non-parametric estimator of a linear trend (Sen, 1968). In addition, ΔT must reflect a smooth and predictable degradation trajectory, with no abrupt acceleration (i.e., sudden fault) near the FTT, at which ΔT is expected to reach its global maximum. This condition can be verified visually, as done in this study, or monitored in real time using second-order finite differences to detect excessive curvature.

When SCADA data fail to resolve such behaviors – due to high-frequency response to the wear processes or rapid end-of-life acceleration – the observed evolution may appear non-monotonic or unstable (Wang *et al.*, 2026), leading to unreliable RUL estimates. Proposition 1 formalizes the conditions under which degradation becomes observable and predictable at SCADA timescales, ensuring that ΔT reflects a meaningful deterioration trend.

Proposition 1. *The time evolution of ΔT between the FPT and FTT must satisfy the following conditions to allow the extraction of a monotonic HI :*

- *The overall trend must be strictly increasing, i.e., the Sen slope must be positive. If ΔT is stationary or decreasing, a reliable monotonic HI cannot be constructed.*
- *At the FTT, ΔT should not show an abrupt change in trend. The threshold exceedance must result from a smooth continuation of degradation rather than a sudden jump.*
- *In the neighborhood of the FTT, ΔT reaches its global maximum.*

Although ΔT satisfying Proposition 1 contains degradation-related information, it remains sensitive to load and ambient variations in SCADA data, which can induce transient decreases unrelated to damage progression. As widely recognized in the prognostics literature, monitoring signals are inherently noisy and must therefore be transformed into suitable HIs that are non-decreasing and trendable (Javed *et al.*, 2017). Monotonicity must therefore be enforced, but without relying on future values, as real-world monitoring systems have no access to the future evolution of ΔT . Accordingly, the HI is constructed directly from ΔT using a Cumulative Maximum (CM) transformation, which preserves the physical envelope of thermal degradation and enables real-time deployment. The CM transformation approximates an irreversible thermodynamic process, ensuring a strictly non-decreasing trend while retaining critical extrema relevant to maintenance decisions. In contrast, other methods such as isotonic regression and spline smoothing tend to attenuate such extrema and require repeated parameter updates, making them less suitable for operational prognostics. The resulting CM-based HI,

denoted $\delta T(t)$, is formally defined as :

$$\delta T(t) = \begin{cases} \Delta T(0), & \text{if } t = 0 \\ \max(\Delta T(t), \delta T(t-1)), & \text{if } t > 0 \end{cases} \quad (5.2)$$

The resulting $\delta T(t)$ is adopted in this study as the HI for RUL prediction. It preserves the long-term degradation trend while suppressing local decreases caused by environmental or operational variability, as shown in Figure 5.4.

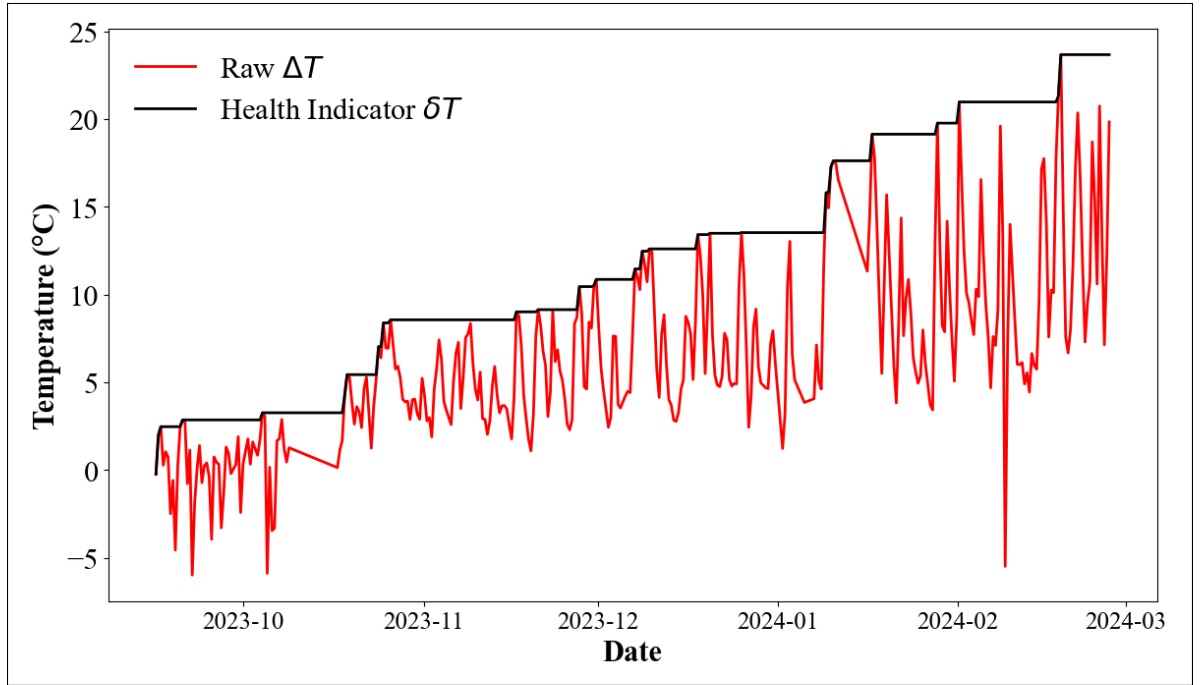


Figure 5.4 Cumulative Maximum filter on ΔT

5.3.2 Transfer Function Selection

SCADA data were extracted for each WT from FPT to FTT, including the inputs ΔT , T_{amb} , U , and P , as described in Section 5.3.1. The target RUL was computed as the time difference (in days) between each observation and the FTT. Data preprocessing consisted in discarding time steps affected by acquisition interruptions, followed by resampling all time series to a 12-hour resolution using interval-wise averaging. This step reduces measurement noise and attenuates

short-term fluctuations prior to sequence generation. The thermal deviation ΔT was then used to construct the HI δT according to equation 5.2, and all variables were normalized to the $[0,1]$ range using Min-Max scaling.

In the absence of an available physics-based degradation model to directly predict RUL in this context, an indirect measurement approach based on data-driven modeling is adopted. The relationship between the input variables and RUL is learned using AI-based transfer functions, whose architectures are deliberately kept simple to facilitate large-scale deployment, generalization, and potential industrialization. Multiple candidate models were evaluated to identify the best trade-off between predictive accuracy and computational cost, in accordance with the guidelines established in our review paper (Benabdesselam *et al.*, 2024). A schematic overview of the input-output configuration is provided in Figure 5.5.

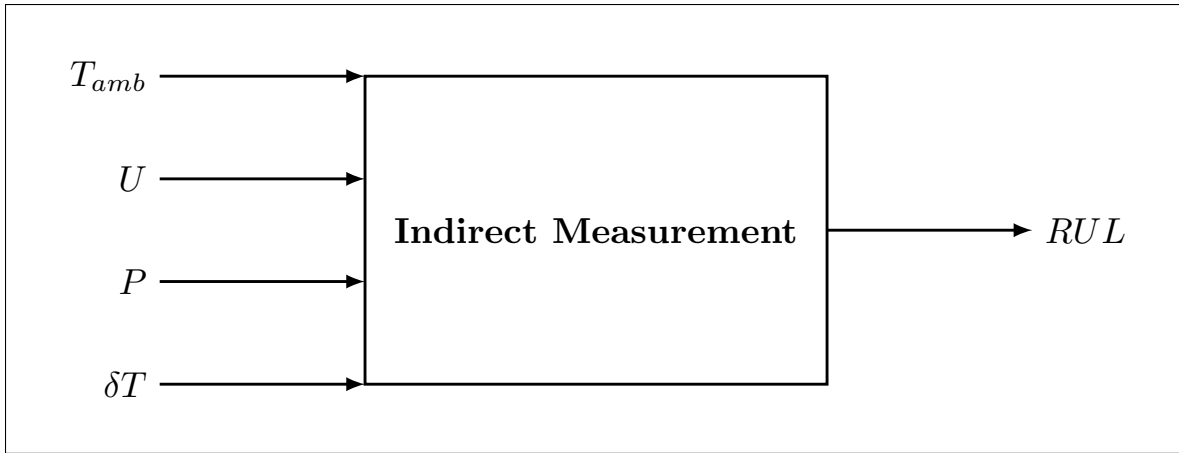


Figure 5.5 Schematic illustration of RUL prediction

The tested models include both neural and non-neural architectures. The neural models comprise the Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM (Bi-LSTM), and Bidirectional GRU (Bi-GRU). The non-neural baselines include Gradient Boosted Regression Trees (GB), Random Forest (RF), and Support Vector Regression (SVR). The hyperparameters of the evaluated models were fixed after an extensive sensitivity analysis and are as follows. For the neural networks, the MLP consists of four hidden layers with 256, 128, 64, and 32 neurons, followed by a linear output layer. Both

the LSTM and GRU models use two recurrent layers with 30 hidden units each, followed by a feedforward head identical to that of the MLP. Their bidirectional counterparts (Bi-LSTM and Bi-GRU) adopt the same configuration but with bidirectional recurrence and a window size of 4, instead of 3 for the unidirectional models. All neural models are trained using the Adam optimizer with a learning rate of 10^{-4} for 2000 epochs, minimizing the Mean Squared Error (MSE) loss. For the non-neural baselines, the GB regressor is configured with 500 estimators, a learning rate of 0.05, and a maximum tree depth of 4, while the RF regressor uses 500 estimators with no depth restriction. The SVR model employs a radial basis function kernel with a regularization coefficient of $C = 10.0$ and a tolerance margin of $\varepsilon = 0.01$, optimizing the ε -insensitive loss.

In the present context – and in many WT applications where degradation is not directly observable – the ground truth of the RUL cannot be known online. It can only be determined once failure has occurred. Specifically, at the failure time, one can retrospectively state that, for instance, the RUL was 120 days exactly 120 days before failure. During operation, however, the true RUL is inherently unknown and can only be estimated. This fundamental limitation explains why, in the literature, the RUL ground truth is commonly defined as a strictly decreasing linear function with a unit slope (Rezamand *et al.*, 2021; Hu *et al.*, 2025; Goebel *et al.*, 2017; Djeziri, Benmoussa & Sanchez, 2018; Oliveira-Filho *et al.*, 2021; Wang, Cao, Xu & Liu, 2022b; Teng *et al.*, 2020). Under this assumption, if $RUL(t)$ denotes the ground-truth RUL and t_{k-1} and t_k are two consecutive time instants, the following relationship holds :

$$RUL(t_k) = RUL(t_{k-1}) - (t_k - t_{k-1}) \quad (5.3)$$

By introducing the discretization time step $\Delta t = t_k - t_{k-1}$, this expression can be rewritten as :

$$RUL(t_k) = RUL(t_{k-1}) - \Delta t \quad (5.4)$$

In contrast to this strictly decreasing ground truth, RUL predictions may exhibit non-monotonic behavior. In particular, the estimated RUL at time t_k , denoted $\widehat{RUL}(t_k)$, may exceed the previous

estimate $\widehat{RUL}(t_{k-1})$. Such behavior is widely reported in the RUL prediction literature (Hu *et al.*, 2025; Oliveira-Filho *et al.*, 2021; Djeziri *et al.*, 2018; Wang *et al.*, 2022b; Teng *et al.*, 2020). Several explanations may account for this phenomenon : inaccurate input measurements at time t_k , model uncertainty, or changes in operating conditions that may temporarily suggest an apparent extension of RUL. Alternatively, maintenance actions or repairs could also induce an increase in the estimated RUL. However, these scenarios are not consistent with the adopted definition of the RUL ground truth as a strictly linearly decreasing function with unit slope, which explicitly excludes maintenance actions, repairs, or life-extension effects. Otherwise, such effects would manifest as plateaus or increases in the ground-truth RUL trajectory. Therefore, to ensure consistency between the predictions, the physical assumptions, and the RUL definition commonly adopted in the literature, it is necessary to enforce a monotonically decreasing behavior in the predicted RUL whenever $\widehat{RUL}(t_k) > \widehat{RUL}(t_{k-1})$, which, to the best of our knowledge, has not been previously reported in the literature. In this work, we propose a physics-based correction that constrains the RUL predictions by directly following the physical logic of equation (5.4). The corrected prediction is defined as :

$$\widehat{RUL}(t_k) = \begin{cases} \max(\widehat{RUL}(t_{k-1}) - \Delta t, 0), & \text{if } \widehat{RUL}(t_k) > \widehat{RUL}(t_{k-1}) \\ \widehat{RUL}(t_k), & \text{otherwise} \end{cases} \quad (5.5)$$

Alternative strategies could be considered, such as correcting $\widehat{RUL}(t_k)$ through interpolation based on the p previous predictions (lag-based smoothing). However, such approaches introduce additional hyperparameters and may compromise model generalization. In contrast, the proposed method remains fully consistent with the underlying physical assumptions and the standard RUL definition adopted in this study and in the broader literature, by enforcing monotonicity solely on the output trajectory at each time step (without making any assumption about future evolution) and without modifying the model parameters. It therefore acts as a physics-based temporal filter rather than a learned constraint, ensuring physically consistent RUL predictions while preserving the original predictive model.

To evaluate generalization across WT, a Leave-One-Out (LOO) method was implemented. At each iteration, data from a single WT were excluded from the training set and used as the independent test set, while the remaining WTs formed the training set. This ensures that the model is evaluated on WTs unseen during training, mimicking real-world deployment.

5.3.3 Uncertainty Quantification

A major limitation of existing prognostic models lies in the absence of calibrated uncertainty quantification. Without reliable Confidence Intervals (CIs), decision-making becomes uncertain, and models tend to degrade under out-of-distribution conditions (Zio, 2022). Several uncertainty quantification techniques exist. Some primarily capture epistemic uncertainty, such as Bayesian neural networks, Gaussian processes, and deep ensembles, while others mainly account for aleatory uncertainty, for example heteroscedastic approaches (He *et al.*, 2025b). However, most of these methods rely on strong distributional assumptions and typically model only a single type of uncertainty, whereas in real PHM applications both epistemic and aleatory uncertainties coexist (Zio, 2022). CP methods have shown strong robustness in capturing both sources of uncertainty and have demonstrated high calibration performance. Among their most effective variants, already applied in PHM, is Conformalized Quantile Regression (CQR) (Piao *et al.*, 2025), adopted in this work. CQR combines Quantile Regression (QR) with CP to provide heteroscedastic, data-adaptive CIs with distribution-free, finite-sample coverage guarantees (Romano, Patterson & Candès, 2019). This property is particularly valuable for heterogeneous WTs operating under variable environmental and loading conditions.

Let $X_i \in \mathbb{R}^d$ denote the SCADA inputs $(T_{amb,i}, U_i, P_i, \delta T_i)$, with $d = 4$, and let $y_i \in \mathbb{R}$ denote the corresponding RUL output, all observed at timestep i . For a confidence level $\alpha \in [0, 1]$, QR estimates the conditional quantiles $q_\alpha(X_i)$, and, by jointly learning $q_{\alpha/2}(X_i)$ and $q_{1-\alpha/2}(X_i)$, provides asymmetric, data-adaptive intervals. However, QR guarantees calibration only asymptotically and often yields miscalibrated intervals in industrial run-to-failure settings (Romano *et al.*, 2019).

CP methods ensure marginal coverage of $(1 - \alpha)$ independently of the model (Barber *et al.*, 2021). Split-CP requires withholding part of the training data, which may degrade performance in limited-data environments (Benabdesselam *et al.*, 2026b). Conversely, LOO CP is computationally prohibitive for deep networks. A practical compromise is K -fold Cross-Validation Conformal Prediction (CV-CP), where the training set, composed of n observations, is partitioned into K folds (Barber *et al.*, 2021). For each fold S_k , with $k = 1, \dots, K$, the subset S_k contains n_{S_k} observations, with $\sum_{k=1}^K n_{S_k} = n$. The model is trained on the other $K - 1$ folds and used to predict samples $(X_i, y_i) \in S_k$, with $i = 1, \dots, n_{S_k}$. Given the lower and upper RUL quantile bounds, the nonconformity score $E_i^{(k)}$ for fold S_k and observation i is :

$$E_i^{(k)} = \max \left(\hat{q}_{\alpha/2}^{(-k)}(X_i) - y_i, y_i - \hat{q}_{1-\alpha/2}^{(-k)}(X_i) \right) \quad (5.6)$$

where the superscript $(-k)$ indicates that fold S_k was excluded during training. Aggregating all scores from fold S_k for all its observations $i = 1, \dots, n_{S_k}$ gives $E^{(k)}$, and aggregating the scores $E^{(k)}$ for all K folds $k = 1, \dots, K$ gives E . The conformal adjustment Q_{CQR} is the empirical $(1 - \alpha)$ -quantile :

$$Q_{CQR} = \text{Quantile}_{1-\alpha}(E) \quad (5.7)$$

At inference, the new observation X_{n+1} is passed through all K models. Conservative predictive bounds are defined as :

$$\underline{q}(X_{n+1}) = \min_{k=1, \dots, K} \hat{q}_{\alpha/2}^{(-k)}(X_{n+1}) \quad \text{and} \quad \bar{q}(X_{n+1}) = \max_{k=1, \dots, K} \hat{q}_{1-\alpha/2}^{(-k)}(X_{n+1}) \quad (5.8)$$

and the calibrated CI is :

$$I_{CQR}(X_{n+1}) = \left[\underline{q}(X_{n+1}) - Q_{CQR}, \bar{q}(X_{n+1}) + Q_{CQR} \right] \quad (5.9)$$

guaranteeing marginal coverage of at least $(1 - \alpha)$ under the assumption of exchangeability. Although this assumption is frequently violated in PHM applications due to the progressive and non-stationary nature of degradation processes, the calibration folds in this study include data

from WTs over the entire degradation interval from FPT to FTT, which mitigates this violation and approximates exchangeability as closely as feasible in an operational setting.

In operational terms, the lower bound of the CQR interval represents a conservative risk scenario, while the upper bound represents the latest acceptable date for preventive maintenance scheduling. This makes the framework directly actionable within operation and maintenance workflows.

5.3.4 Evaluation Metrics

The performance evaluation relies on two complementary categories of metrics. The first category includes accuracy metrics, namely the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). The second category concerns uncertainty calibration metrics, which evaluate the agreement between theoretical and empirical confidence levels. The theoretical coverage corresponds to the nominal confidence level $1 - \alpha$ expected from the model, while the observed coverage represents the empirical proportion of true values contained within the predicted intervals. A model is considered well calibrated when the empirical coverage closely matches the theoretical one across all confidence levels. The calibration metrics, including the Expected Calibration Error (ECE), Root Mean Squared Calibration Error (RMSCE), Miscalibration Area (MA), and sharpness, are presented and discussed in detail in Benabdesselam *et al.* (2026b).

5.4 Results and Discussion

5.4.1 Case Study Description

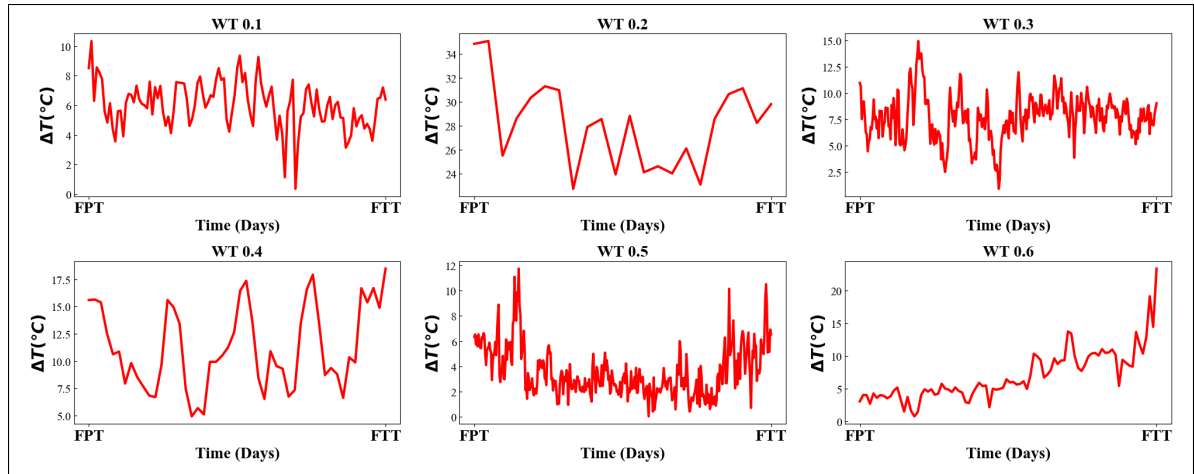
This study relies on a dataset provided by an industrial partner, comprising WTs from three manufacturers deployed across 20 wind farms. All assets are horizontal-axis, three-bladed, megawatt-class WTs. The dataset spans from September 2022 to March 2024 and includes multiple overtemperature events affecting the main bearing, generator bearings, and the generator.

An initial preprocessing step removed cases for which the FPT or the FTT was undocumented, or where SCADA data between these two events were missing or unusable.

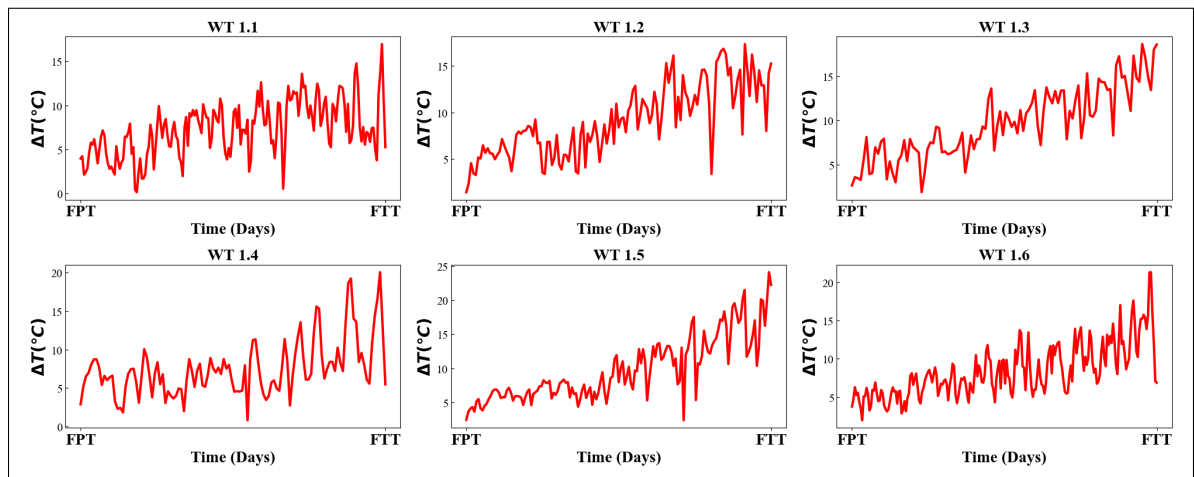
Despite the presence of overtemperature events, some WTs exhibited degradation dynamics that were not adequately captured at SCADA resolution, and CMS data would be required to characterize the underlying mechanisms. This limitation is not specific to the proposed framework but is inherent to any prognostic approach relying solely on low-frequency SCADA measurements. In this work, we explicitly characterize when SCADA-based prognostics is feasible and when it becomes fundamentally ill-posed due to lack of observability at SCADA resolution. As stated in Section 5.2.3, the scope of this study is therefore restricted to gradual thermomechanical degradation processes that produce a trendable thermal drift, while sudden faults and transient phenomena are explicitly excluded. Figure 5.6a illustrates representative discarded cases that failed to satisfy the criteria of Proposition 1 despite exhibiting overtemperature events. Some WTs show a decreasing ΔT trend, suggesting an apparent recovery, which is physically incoherent with the absence of repair between the FPT and the FTT. Others display non-monotonic or highly irregular behaviors, including plateaus, abrupt changes, or late accelerations near the FTT, all of which prevent reliable RUL inference at SCADA resolution. Out of approximately 60 screened WTs, 28 WTs satisfied Proposition 1 and provided exploitable degradation trajectories. This selection should be interpreted as a domain-of-validity characterization rather than a performance enhancement, as the proposed framework applies only when degradation is observable in the available SCADA signals. The filtering is performed independently of the prediction model and does not rely on RUL estimation performance, but solely on physical observability conditions derived from the ΔT signal. All retained cases correspond to the main bearing and involve WTs from 8 wind farms and 2 manufacturers. A representative valid ΔT trajectory is shown in Figure 5.6b, confirming compliance with Proposition 1.

5.4.2 Transfer Function Selection

Based on the models presented in the methodology, the performance was evaluated using the RMSE, MAE, and R^2 metrics, along with CPU time, under a LOO validation applied to the



a) Examples of discarded WTs



b) Examples of retained WTs

Figure 5.6 Examples of degradation trajectories extracted from ΔT for different WTs

28 WTs. The results are summarized in Table 5.2. All experiments were conducted on a laptop equipped with an Intel Core i5-1145G7 CPU (4 cores, 8 threads) and 16 Gigabyte of RAM, running Windows 11 Pro. The models were implemented in Python 3.12.1.

Among the tested models, the lowest performances were obtained with RF, GB, and SVR, confirming the limited capability of tree-based and kernel methods to capture the temporal dependencies inherent in the data. In contrast, the MLP, while not explicitly modeling temporal

Tableau 5.2 Performance comparison of the tested models

Model	RMSE (d)	MAE (d)	R^2 (%)	CPU-Time (min)
MLP	28.10	25.10	84.90	54
LSTM	26.59	24.32	88.15	72
GRU	26.54	24.19	88.43	69
Bi-LSTM	26.90	24.60	87.50	75
Bi-GRU	26.60	24.40	87.90	90
GB	29.50	27.30	80.90	15
RF	31.00	28.80	78.70	6
SVR	28.70	26.50	82.00	8

dependencies, offered a balanced compromise between predictive accuracy and computational efficiency by leveraging aggregated information from the input features, delivering consistent results with moderate complexity. Recurrent architectures further improved predictive accuracy by explicitly exploiting temporal correlations within the SCADA sequences. The GRU model achieved the best overall performance, yielding the lowest RMSE and MAE values and the highest R^2 score while maintaining a reasonable computation time. The Bi-LSTM also ranked among the top performers, confirming the advantages of recurrent neural networks for modeling time-dependent behaviors in industrial prognostics.

Using the GRU model with LOO validation, the RUL predictions for each WT are shown in Figure 5.7, where the estimated RUL is shown in red and the ground truth in black, together with the corresponding performance metrics, while the quantified uncertainty is discussed in Section 5.4.3.

It should be noted that, in an ideal case, the true RUL trajectory should follow a strictly decreasing linear function with unit slope. However, due to missing SCADA records within the FPT-FTT interval, discontinuities appear in the ground truth evolution in Figure 5.7. The predictions demonstrate a high level of accuracy, with several WTs exhibiting near-perfect alignment with the reference RUL. For WT1 to WT14, the predictions are almost indistinguishable from the

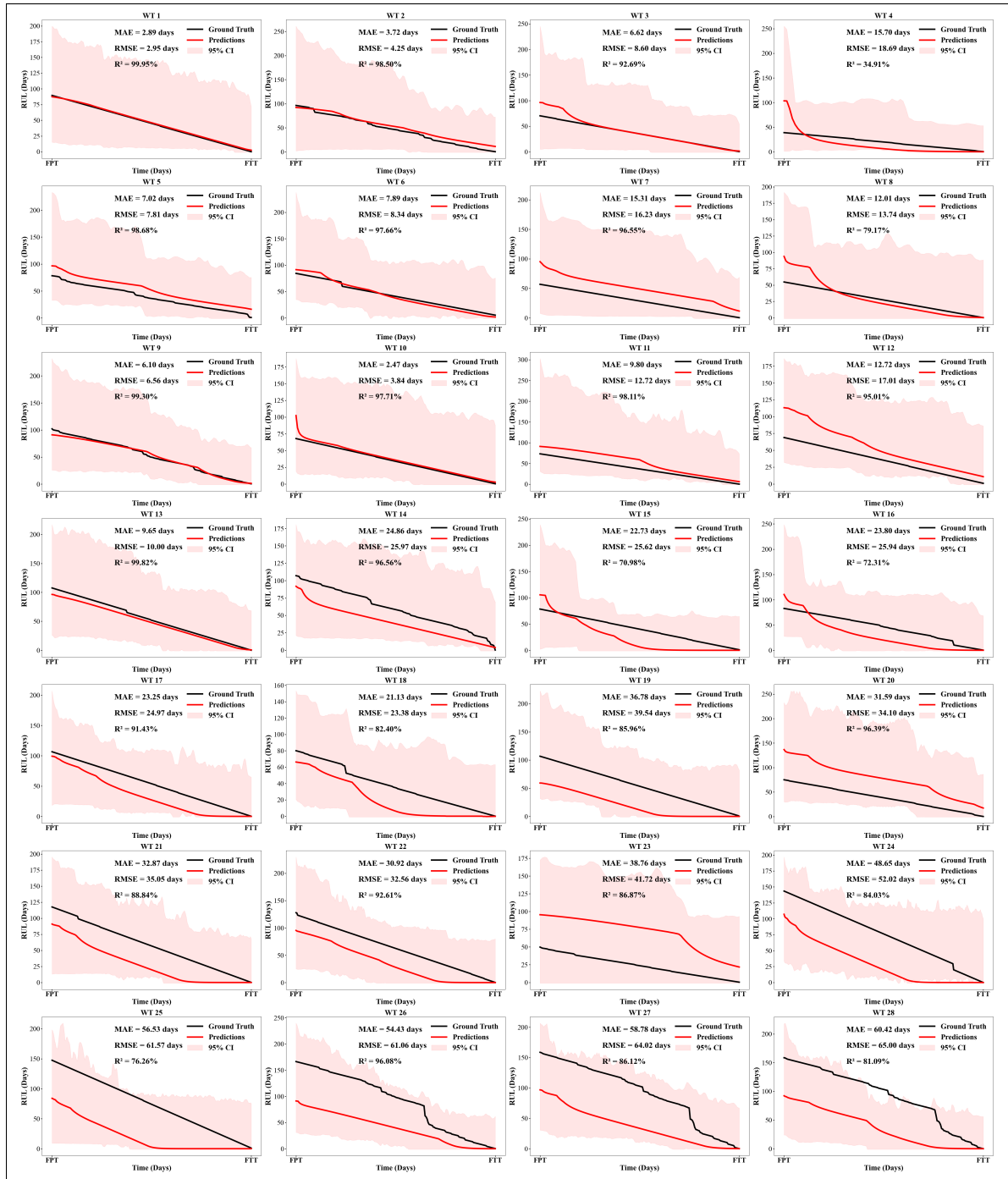


Figure 5.7 RUL predictions for each WT using the GRU model with LOO validation and uncertainty quantification using CQR

ground truth. WT15 to WT22 also show very good agreement, with only minor deviations. A slight performance degradation is observed for WT23 to WT28, although the predicted trends

remain reasonable. This reduction in accuracy is attributed to multiple interacting factors. Most notably, several WTs in this last group exhibit simulated RUL values exceeding 160 days – far beyond the typical range observed in the dataset, where most units fail within 100 days. Data quality is also a critical factor, as in any AI-based model, and RUL performance here strongly depends on the smoothness and predictability of ΔT . For well-predicted WTs, ΔT evolves smoothly and monotonically, enabling accurate RUL estimation, as illustrated by WT1 in Figure 5.8. In contrast, WTs with larger errors exhibit irregular thermal patterns – such as abrupt increases or plateaus – that disrupt degradation progression. Although WT23 satisfies the observability conditions of Proposition 1, its late-life acceleration reduces prediction accuracy.

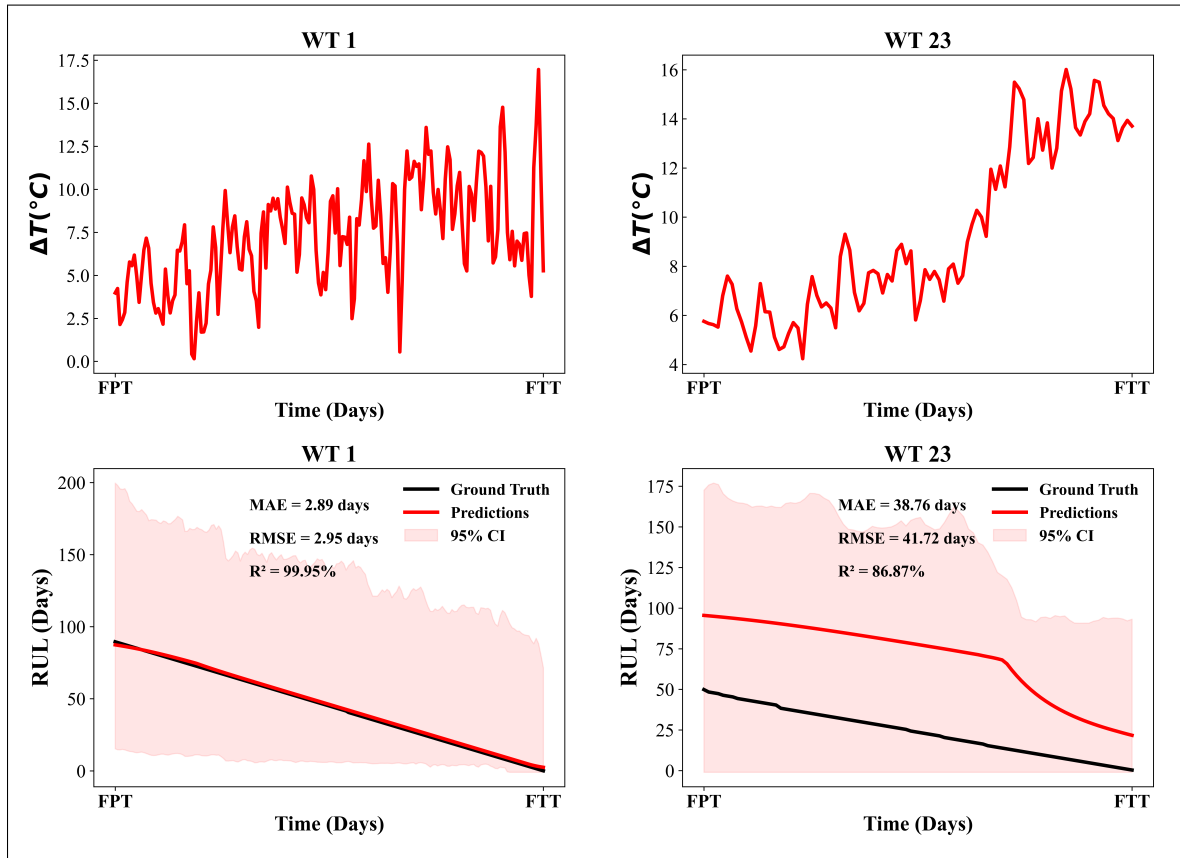


Figure 5.8 Effect of ΔT smoothness on RUL prediction

5.4.3 Uncertainty Quantification

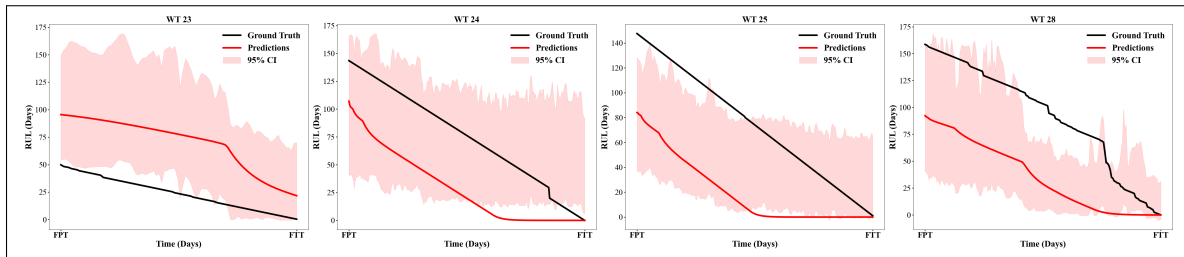
As presented in Section 5.3.3, uncertainty plays an important role in supporting decision-making. In this part, the GRU model is reused with the same LOO approach adopted for prediction. The model outputs are retained, as shown in Figure 5.7. CQR is applied to obtain well-calibrated CIs. In this process, the only hyperparameter to be set is K , representing the number of folds, and a sensitivity analysis is conducted to determine the optimal value. It should be noted that, regardless of the number of folds considered, the selection of WTs within each fold was completely random among the 28 WTs used. Several values of K were tested (5, 8, 10, 15, and 20), and the calibration metrics ECE, RMSCE, MA, and Sharpness – introduced in Section 5.3.4 and detailed in Benabdesselam *et al.* (2026b) – were computed. The results are summarized in Table 5.3.

Tableau 5.3 Sensitivity analysis of the calibration results for different numbers of folds K

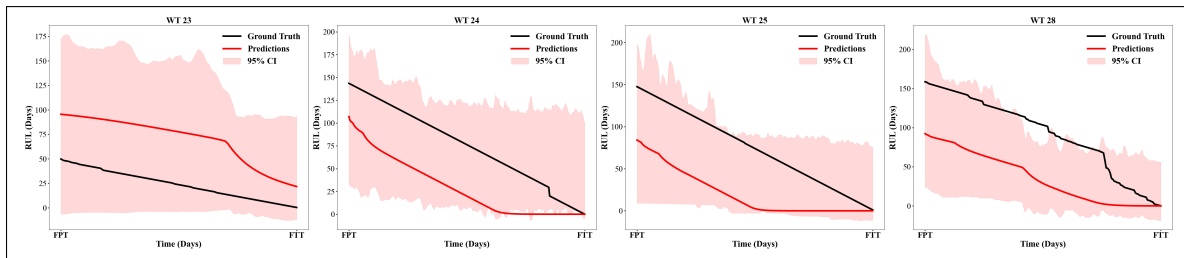
K (Number of folds)	5	8	10	15	20
ECE	0.1314	0.1285	0.1230	0.1281	0.1286
RMSCE	0.1398	0.1367	0.1312	0.1361	0.1366
MA	-0.1306	-0.1275	-0.1219	-0.1268	-0.1276
Sharpness	64.58	63.89	62.84	64.32	64.73

For the different values of K , as shown in Table 5.3, the calibration performance does not vary significantly. The miscalibration area is negative, indicating that the empirical curve lies above the nominal (theoretical) line, meaning that the model ensures at least the nominal coverage. The calibration errors are relatively small, with a slight underconfidence of the model, which can be attributed to variations in operating conditions from one WT to another. In decision-making contexts, such slight underconfidence is preferable to overconfidence. All folds achieved an empirical coverage of around 98% for a nominal coverage of 95%, which aligns with the intended objective. The optimal number of folds is around 10, which resulted in the lowest calibration error. For this configuration, and for a nominal confidence level of 95%, the uncertainty results are presented in Figure 5.7.

To demonstrate the benefits of using CQR over uncalibrated QR, Figure 5.9a illustrates the 95% CIs produced by QR for WT23, WT24, WT25, and WT28, contrasted with Figure 5.9b, which shows the corresponding intervals after conformal calibration. With QR, the CIs fail to cover the ground truth and are therefore poorly calibrated. In contrast, CQR fully addresses the undercoverage issue for WT23, WT24, and WT25. For WT28, the coverage has improved considerably compared with the results obtained using the uncalibrated model.



a) QR Uncertainty - Before conformal calibration



b) CQR Uncertainty - After conformal calibration

Figure 5.9 Uncertainty estimation before and after conformal calibration

With CQR, CIs become noticeably wider, reflecting a more realistic representation of the model's uncertainty. In many cases, the value zero is included within the interval, which can be interpreted as the model indicating that the possibility of an immediate failure cannot be entirely ruled out. This situation may arise from sudden changes in operating conditions, accelerated degradation mechanisms, or unforeseen external factors that can abruptly affect the system's health state. From a practical standpoint, this behavior is valuable for decision-making, as it warns maintenance teams that, even when the predicted RUL appears acceptable, a certain level of risk persists. The lower bound of the interval thus represents a conservative scenario,

suggesting that failure could occur sooner than expected, while the upper bound provides an estimate of the maximum number of days before the end of life is likely to occur. This upper limit can be interpreted as the latest time by which preventive maintenance should be scheduled to avoid unplanned downtime. On average, the RUL is expected to evolve around the mean prediction provided by the model, which offers a balanced estimate between optimistic and conservative scenarios.

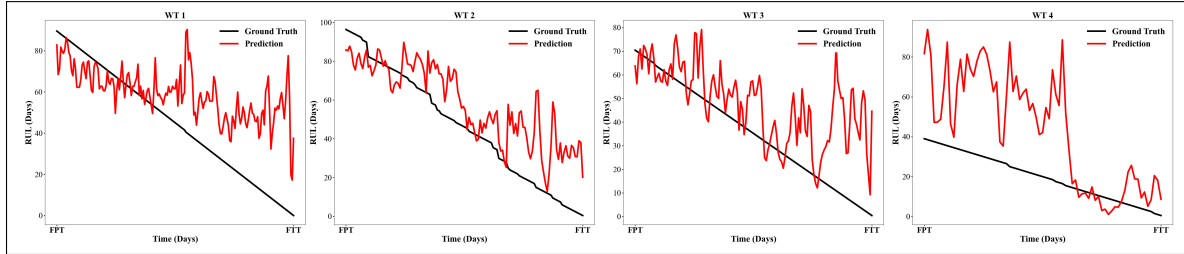
5.4.4 Effect of Monotonicity Constraints on HI and RUL

In this section, we look at the enforcement of monotonicity on the ΔT signal to construct the HI, denoted as δT , as introduced in Section 5.3.1. We further motivate the monotonic constraint imposed on the predicted RUL, described in Section 5.3.2. These constraints are introduced to preserve physical consistency and are applied in a causal, real-time manner using only past and current observations, without relying on future information or modifying the model parameters.

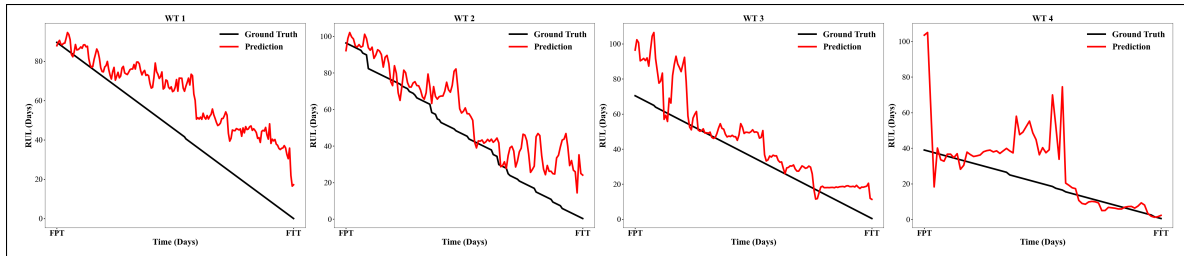
First, we present the RUL prediction results obtained for WT1, WT2, WT3, and WT4 without enforcing monotonicity on either ΔT or the RUL, as shown in Figure 5.10a. In this configuration, the predictions exhibit large fluctuations and high levels of noise, which lead to inconsistent trajectory shapes compared with those shown in Figure 5.7. Such non-physical behavior, where the RUL may increase from one time step to the next, is incompatible with the irreversible nature of degradation processes and precludes reliable interpretation. From an operational perspective, this behavior is unsuitable for practical applications, as it significantly complicates maintenance decision-making.

Next, Figure 5.10b illustrates the case where monotonicity is enforced only on ΔT using the CM filtering approach. Although the resulting trajectories are smoother than those in Figure 5.10a, residual variability remains, and upward fluctuations still occur in the predicted RUL, which has not been addressed in the literature and complicates decision-making when RUL is extended despite no adjustment or maintenance action being performed. This observation motivates the

additional enforcement of a strictly decreasing RUL evolution, as implemented in Section 5.3.2 and shown in Figure 5.7.



a) HI and RUL evolution without monotonic constraints



b) Monotonic HI and non-monotonic RUL prediction

Figure 5.10 Effect of monotonicity constraints on HI and RUL

5.5 Conclusion

This study proposes a SCADA-based framework for predicting the RUL of WT components. The framework is fully deployable across any SCADA-enabled fleet without requiring additional instrumentation or recalibration campaigns. All stages, from anomaly detection to HI construction, RUL estimation, and uncertainty quantification, are designed to operate in real time and integrate seamlessly into existing asset-management pipelines without offline post-processing or additional assumptions. This represents a significant step forward in PHM and provides wind farm operators with a practical decision-support tool when thermal anomalies or degradation patterns are detected.

The approach was validated on 28 WTs across multiple wind farms and manufacturers and demonstrated strong performance, with an average MAE of 24 days, an average RMSE of 26 days

and an R^2 value of 88%. The uncertainty quantification produced reliable and well-calibrated CIs, reaching an empirical coverage of 98% for a nominal level of 95% with low calibration errors.

Beyond predictive accuracy, this work explicitly characterizes the conditions under which SCADA-based prognostics is feasible and identifies situations in which the problem becomes fundamentally ill-posed due to lack of observability at SCADA resolution.

The principal limitation of the approach therefore lies not in the modeling strategy itself, but in the inherent inability of low-frequency SCADA measurements to capture high-frequency or abrupt degradation mechanisms, which remain invisible to SCADA signals, manifest as apparent sudden faults, and therefore cannot be reliably predicted without complementary sensing modalities.

Future work will focus on extending the validation to additional assets and on developing more robust HIs by combining multiple SCADA variables, as well as on integrating SCADA-based prognostics with CMS to overcome observability limitations and enable reliable prediction of degradation processes that remain hidden at SCADA resolution. Another possible area of development is the implementation of the current strategy on a larger-scale dataset, with automated processes to identify in real time when an asset no longer meets the criteria of Proposition 1, thereby further demonstrating the potential for automation in the field of RUL prediction.

CONCLUSION ET RECOMMANDATIONS

Cette thèse s'inscrit dans le contexte de la transition énergétique et de la nécessité de réduire les coûts d'opération et de maintenance des actifs des énergies renouvelables. Le but *in fine* est d'améliorer leur fiabilité et leur disponibilité, et par conséquent, d'optimiser les opérations de maintenance. Dans cette perspective, le PHM constitue un levier clé. Toutefois, son déploiement à grande échelle demeure freiné par deux contraintes : (i) la disponibilité inégale de systèmes de CMS sur les parcs, et (ii) la complexité intrinsèque d'un pronostic précis, particulièrement lorsque les données accessibles proviennent de systèmes SCADA à faible résolution temporelle.

6.1 Apports de la thèse

En réponse à ces enjeux, la thèse a exploré un fil directeur cohérent : rendre le PHM déployable à grande échelle en s'appuyant, lorsque possible, sur des mesures indirectes déduites à partir de données déjà disponibles (SCADA), le tout en s'assurant d'une quantification robuste de l'incertitude, permettant ainsi de transformer des prédictions en décisions opérationnelles. Trois (3) contributions originales, organisées autour des trois articles de journal à comité de lecture, ont permis d'adresser successivement (i) la surveillance et le diagnostic à partir de mesures indirectes, (ii) la décision et le pronostic sous forme de classification binaire d'événements critiques, et finalement, (iii) l'estimation continue de la durée de vie résiduelle (RUL) sous incertitude, en caractérisant explicitement le domaine de validité des approches issues des données SCADA.

(i) Mesures indirectes et diagnostic : le premier axe de notre recherche a permis de montrer que des variables SCADA, initialement conçues pour le suivi des performances énergétiques, peuvent être exploitées comme substitut aux CMS pour le phénomène de cavitation érosive, sujet encore peu traité dans la littérature et reconnu pour sa complexité. L'idée centrale consiste à construire une approche de mesure indirecte capable d'estimer une grandeur de dégradation, à

savoir *ICR*. Cette mesure est donc réalisée à partir des données SCADA, et elle s'accompagne d'une quantification de l'incertitude, opération réalisée à l'aide de différentes méthodes. Cela nous a également permis de présenter une analyse quasi complète sur les approches à retenir pour la quantification de l'incertitude qui nous semblent les plus adaptées au contexte industriel des actifs en énergies renouvelables en termes de calibration et de robustesse. L'objectif final est d'assurer que le diagnostic est précis et que la projection des prédictions et de leurs incertitudes demeure cohérente et utile pour construire la trajectoire de dégradation.

(ii) Pronostic sous forme de classification : le deuxième axe de notre recherche avait pour application les éoliennes commerciales dotées de capteurs thermiques avec des seuils préétablis déclenchant un arrêt automatique. Nous avons proposé une approche de pronostic de nature décisionnelle. Plutôt que d'estimer une métrique RUL, difficilement observable avec les données SCADA, l'approche propose de déterminer si une anomalie (de nature thermique) conduira ou non à un dépassement futur du seuil critique. La principale contribution ici réside dans la proposition d'une métrique de « l'événement » construite à partir des données SCADA et de l'incertitude. Ce formalisme nous a permis de convertir l'information probabiliste des intervalles en un signal cumulatif strictement croissant, puis en une décision binaire via un seuil de l'événement. Validée sur un large ensemble industriel, l'approche a démontré une capacité d'alerte précoce significative (jusqu'à plusieurs mois pour certains cas) tout en maintenant un faible taux de fausses alertes, compatible avec l'exploitation.

(iii) Estimation continue de la RUL et domaine de validité : le troisième axe a abordé directement la prédiction de la RUL à partir de la base de données SCADA, en reconnaissant une limite fondamentale : **toutes les dégradations ne sont pas observables à la résolution SCADA : certaines le sont, tandis que d'autres restent non détectables**. La contribution méthodologique majeure consistait ici à formaliser des conditions de validité qui ont permis d'identifier les trajectoires thermiques pour lesquelles le pronostic, en mode continu, est bien

posé. À partir du signal de hausse de température ΔT , un HI monotone a été construit par filtrage cumulatif maximum pour assurer une cohérence physique et une utilisabilité opérationnelle en temps réel. Des modèles de mesure indirecte volontairement simples ont ensuite été évalués, avec une validation de généralisation *Leave-One-Out* sur des éoliennes non vues. Enfin, nous avons proposé et intégré une quantification d'incertitude robuste via CQR, produisant une incertitude bien calibrée et directement exploitable pour la planification de maintenance.

Globalement, la thèse met en évidence deux (2) résultats transversaux. Premièrement, les approches issues des données SCADA peuvent soutenir un PHM déployable à grande échelle, à condition de poser explicitement le problème en fonction de l'observabilité des mécanismes (diagnostic, classification, pronostic ou RUL). Deuxièmement, la quantification de l'incertitude n'est pas un ajout optionnel : elle constitue le mécanisme qui relie la décision à la prédiction, en particulier dans des contextes non stationnaires où l'échangeabilité est violée, et où la dégradation induit des dérives. La thèse propose ainsi un cadre pragmatique, orienté vers une application industrielle, qui combine (i) des mesures indirectes compatibles avec les contraintes de terrain, (ii) des modèles de complexité modérée favorisant la transférabilité, et (iii) des méthodes d'incertitude non paramétriques capables de conserver une calibration utile en pratique.

6.2 Limites de la thèse

Malgré les contributions obtenues, plusieurs limites structurelles et inhérentes à notre méthodologie sont à souligner.

- **Limites d'observabilité SCADA** : une partie des mécanismes de dégradation demeure invisible à la résolution SCADA, et elle se manifeste par des défauts soudains. Dans ces cas, les approches proposées ne peuvent garantir un pronostic fiable sans capteurs complémentaires (CMS) ou sans sources exogènes supplémentaires.

- **Non-stationnarité et échangeabilité** : les méthodes conformes reposent sur l'hypothèse d'échangeabilité. Or, les données opérationnelles sont intrinsèquement non stationnaires et présentent des dérives liées à la dégradation. Bien que la thèse ait montré que les méthodes demeurent utiles sous violations modérées, une partie des écarts de couverture et des faux négatifs peut être attribuée à cette violation.
- **Définition opérationnelle de la défaillance** : dans le cas d'un mécanisme de dégradation avec pour symptôme un accroissement thermique, la défaillance est définie par un arrêt automatique sur un seuil défini a priori. Cette définition est pertinente du point de vue de la disponibilité et de la maintenance, mais elle ne coïncide pas nécessairement avec la fin de vie physique. Cela limite l'interprétation physique des prédictions en termes de dommage cumulatif.
- **Dépendance aux données historiques et aux choix de référence** : certaines approches nécessitent des historiques suffisamment longs (ex. > une année) et supposent l'existence d'un comportement de référence (modèle sain). Le déploiement de l'approche proposée dans l'article 2 ne peut donc pas s'appliquer directement sur un nouveau parc éolien ne disposant d'aucune donnée historique. Cette limitation demeure toutefois moins problématique, dans la mesure où, pour un parc nouvellement mis en service, le risque de défaillances est généralement faible et ne justifie pas un déploiement urgent de stratégies de PHM.

6.3 Recommandations et perspectives de recherche

Les travaux menés dans cette thèse ont permis plusieurs avancées en vue de consolider, étendre et industrialiser les approches de PHM fondées sur les données SCADA. Les contraintes temporelles nous obligent à mettre un terme à nos travaux et à produire un rapport. Par conséquent, nous présentons quelques pistes de réflexion pour de futurs travaux. Ces pistes ont été dictées par notre expérience et nos travaux durant cette thèse.

- **Reconnaître explicitement les limites du remplacement des CMS par les données SCADA.** Les travaux de cette thèse montrent que les données SCADA peuvent, dans certains cas et sous certaines conditions, remplacer efficacement les CMS pour le diagnostic et pour certaines tâches de pronostic (par exemple pour la cavitation érosive). Toutefois, elles ne peuvent pas couvrir l'ensemble des mécanismes de dégradation, en particulier ceux dont la signature est de haute fréquence ou faiblement corrélée aux variables SCADA. Une recommandation importante est donc de ne pas opposer SCADA et CMS, mais de définir clairement les cas où SCADA est suffisant et ceux où CMS demeure indispensable. Cela s'inscrit dans la même lignée que l'article de revue publié en 2026 dans *Reliability Engineering & System Safety* (Wang *et al.*, 2026) – journal de référence en PHM –, lequel souligne que les CMS demeurent supérieurs pour la détection fine et précoce de certains mécanismes de dégradation, et que les données SCADA ne peuvent se substituer de manière universelle aux CMS, mais constituent une alternative économique et pertinente dans des cas bien définis.
- **Renforcer la construction des indicateurs de santé à partir de données SCADA.** Si la température et son évolution cumulative se sont révélées particulièrement pertinentes pour certains mécanismes de dégradation, notamment thermomécaniques, d'autres cas présentent une observabilité limitée. Une perspective naturelle consiste à enrichir la construction des HIs en combinant plusieurs variables SCADA (courants, pertes, états de fonctionnement, consignes), afin d'améliorer l'observabilité lorsque ΔT seul n'est pas informatif. Une piste cohérente avec les travaux de cette thèse serait l'utilisation de modèles de projection latente (ex. autoencodeurs variationnels) pour combiner ces variables et extraire un HI plus robuste, tout en conservant une architecture simple et interprétable. Cette idée commence à susciter un intérêt croissant au sein de la communauté scientifique, comme en témoigne notamment le travail de Milani *et al.* (2025b), qui propose un autoencodeur convolutionnel simple imposant la monotonie du HI, sans recourir à une recherche lourde d'hyperparamètres, ou encore

en approfondissant les travaux de diagnostic et de pronostic appliqués aux éoliennes, en s'appuyant sur l'espace latent, tel qu'introduit par Moreira de Oliveira Filho (2024). Les résultats obtenus sont prometteurs, bien que plusieurs améliorations restent à envisager.

- **Formaliser et automatiser la détection des cas non prévisibles pour le pronostic.**

Les résultats montrent clairement que le pronostic n'est pertinent que lorsque certaines conditions sont satisfaites (monotonie globale, absence de saut brutal à la *FTT*, existence d'un maximum global). Une recommandation clé est d'intégrer ces conditions dans un mécanisme automatisé capable d'évaluer, en temps réel, la faisabilité d'une estimation de la RUL. Lorsque ces conditions ne sont pas respectées, le système pourrait basculer vers des stratégies alternatives, telles que la classification de l'événement, l'escalade vers des CMS ou une inspection ciblée. Cette détection pourrait également s'appuyer sur des modèles d'apprentissage (ex. autoencodeurs) entraînés à reconnaître les trajectoires jugées prévisibles, pour lesquelles une erreur de reconstruction élevée signale donc un cas non exploitable pour le pronostic.

- **Développer des architectures hybrides SCADA–CMS ciblées.** Une perspective industrielle réaliste consiste à adopter une architecture hiérarchique dans laquelle les données SCADA assurent un triage à l'échelle de la flotte (détection, priorisation, évaluation de l'événement), tandis que les CMS ne seraient activés que pour les actifs non observables ou identifiés comme critiques. Une telle approche permettrait de réduire significativement les coûts globaux de surveillance tout en conservant une capacité de diagnostic approfondie lorsque nécessaire. Cette recommandation a déjà été formulée par Helsen, Devriendt, Weijtjens & Guillaume (2015), bien qu'aucune stratégie de déploiement n'ait, à notre connaissance, été proposée depuis.

- **Améliorer la robustesse de la quantification d'incertitude en contexte non stationnaire.**

Les applications PHM, et en particulier le pronostic, violent intrinsèquement l'hypothèse d'échangeabilité en raison de la dérive irréversible induite par la dégradation. Bien que

les méthodes de *Conformal Prediction* aient montré une robustesse satisfaisante dans les cas étudiés, des travaux futurs pourraient explorer des variantes conformes adaptées aux séries temporelles et à la dérive progressive. L'objectif serait de réduire la sous-couverture observée lorsque la distribution s'éloigne du régime de calibration, tout en conservant la simplicité et la générique nécessaires à un déploiement industriel. Des travaux, tels que ceux de Barber, Candès, Ramdas & Tibshirani (2023), commencent à aborder cette problématique en proposant une théorisation des méthodes conformes lorsque l'hypothèse d'échangeabilité est violée. Toutefois, ces approches demeurent difficiles à déployer en pratique en raison de la complexité du formalisme mathématique. Cela témoigne néanmoins du fait que la communauté scientifique est consciente de cette problématique.

- **Rendre l'incertitude plus interprétable pour la prise de décision.** Si les intervalles de confiance calibrés constituent un apport méthodologique majeur, leur interprétation demeure délicate pour les gestionnaires d'actifs. Un intervalle de RUL très large (ex. 5 à 150 jours) est difficilement exploitable en l'état. Des travaux futurs devraient viser à traduire ces intervalles en informations décisionnelles plus directement actionnables, par exemple en termes de fenêtres minimales et maximales d'intervention, de priorisation multi-actifs ou de politiques de maintenance tenant compte explicitement des coûts de fausses alarmes, des coûts de panne et de l'indisponibilité.
- **Étendre les travaux de pronostic à d'autres cas d'étude.** Le pronostic appliqué à plusieurs cas d'étude, notamment aux turbines hydrauliques, demeure encore très peu étudié, principalement en raison (i) du manque de données de dégradation *Run-to-Failure* – lié aux politiques de maintenance strictes ou à la lenteur de certains mécanismes de dégradation –, (ii) de la complexité de certains mécanismes de dégradation, et (iii) de la difficulté à généraliser les approches de pronostic, les turbines hydrauliques étant souvent de conception unique. Pour surmonter ces limitations, une première recommandation consiste à encourager, lorsque le coût et le risque le permettent, des politiques de maintenance plus

tolérantes à certaines dégradations afin de collecter des données *Run-to-Failure*. Bien que coûteuse à court terme, une telle stratégie peut s'avérer bénéfique à long terme en enrichissant les bases de données nécessaires au développement et à la validation de méthodes de pronostic robustes. Une seconde recommandation, alternative ou complémentaire, consiste à approfondir les approches proposées par Jose *et al.* (2024, 2025), qui visent à reconstruire des trajectoires de dégradation à partir de fragments en exploitant des données multimodales, et à étendre leur validation à d'autres cas d'étude, notamment du point de vue mécanique pour les turbines hydrauliques. Enfin, il est à noter que les résultats obtenus pour la cavitation érosive montrent que la trajectoire cumulative de dégradation présente un comportement quasi linéaire, potentiellement favorable au pronostic, ce qui incite à initier des travaux futurs, susceptibles de constituer une contribution significative dans un domaine encore largement inexploré.

- **Améliorer et quantifier la qualité des données SCADA.** Enfin, la qualité des données apparaît comme l'un des facteurs les plus limitants pour toutes les recherches recensées. Des travaux futurs pourraient viser à définir des indicateurs ou des classes de qualité des données, par exemple sous forme de niveaux ou de codes couleur, afin d'évaluer en amont la fiabilité des données disponibles et d'adapter les méthodes de PHM en conséquence. Une telle démarche permettrait de mieux refléter la réalité industrielle et de renforcer la crédibilité des approches proposées.

6.4 Épilogue

En synthèse, cette thèse démontre que des approches PHM compatibles avec les contraintes industrielles peuvent être construites à partir des données SCADA, à condition de (i) choisir une formulation du problème alignée avec l'observabilité, (ii) imposer des contraintes physiques minimales garantissant la cohérence temporelle, et (iii) intégrer une quantification d'incertitude

robuste permettant la décision. Les contributions proposées constituent une base méthodologique pour un PHM à grande échelle, transférable et industrialisable, tout en clarifiant les limites fondamentales de la surveillance SCADA. Les perspectives identifiées, notamment l'adaptation à la non-stationnarité, l'hybridation SCADA–CMS et l'intégration explicite de critères de décision de maintenance, représentent des étapes naturelles pour consolider et généraliser ces résultats à d'autres composants, d'autres mécanismes et d'autres filières d'énergies renouvelables.

Malgré les nombreuses problématiques qui restent à résoudre et les différentes voies de recherche à explorer, telles que présentées dans les recommandations précédentes, l'approche développée dans cette thèse constitue, à notre humble avis, une base méthodologique solide et un point de départ robuste, susceptible d'accompagner nos partenaires industriels face aux défis imposés par le contexte socioéconomique mondial actuel et vraisemblablement à venir. Enfin, nous espérons que les perspectives d'amélioration esquissées dans cette section contribueront à nourrir les travaux futurs et à faire progresser l'état des connaissances en sciences appliquées, en particulier dans le domaine du PHM.

Enfin, Albert Camus rappelait que la vérité est mystérieuse, fuyante et toujours à conquérir, et que l'on doit avancer vers elle péniblement mais résolument. Cette allégorie peut, dans ce passage, être transposée à la recherche scientifique, qui demeure un chemin sans terme, exigeant, où chaque sommet atteint annonce déjà un nouvel effort, à l'image du rocher de Sisyphe. Le premier pas consiste alors à clarifier nos besoins et à reconnaître nos limites, car *« ce que l'on conçoit bien s'énonce clairement, et les mots pour le dire arrivent aisément »*.

ANNEXE I

ARTICLES DE REVUE ET DE CONFÉRENCE

1. Article de journal

La Figure I-1 reproduit la première page de l'article de revue, dont le doctorant est le premier auteur, publié dans le journal *Electronics* (Benabdesselam *et al.*, 2024). À la date de rédaction de cette thèse, cet article avait été cité quatre (4) fois (autocitations exclues). Il propose une synthèse des approches fondées sur des mesures indirectes appliquées aux systèmes d'énergies renouvelables, constituant, à notre connaissance, la première revue dédiée spécifiquement à cette thématique.

La Figure I-2 reproduit la première page de l'article de journal, dont le doctorant est coauteur, soumis à *IEEE Transactions on Reliability*.

2. Article de conférence

La Figure I-3 reproduit la première page de l'article présenté lors de la conférence PHM 2025 à Bruges (Benabdesselam *et al.*, 2025). Le doctorant y a effectué une présentation orale.

La Figure I-4 reproduit la première diapositive de la présentation sur le pronostic des éoliennes, effectuée par Mme Monelle Comeau, spécialiste des données chez Power Factors, notre partenaire industriel, lors de laquelle elle a présenté, entre autres, les travaux du doctorant, en juin 2025, à la conférence *WindEurope Technology Workshop* à Istanbul.

Review

On the Use of Indirect Measurements in Virtual Sensors for Renewable Energies: A Review

Abderrauof Benabdesselam ^{1,*} , Quentin Dollon ², Ryad Zemouri ² , Francis Pelletier ³, Martin Gagnon ²  and Antoine Tahan ¹ 

¹ Department of Mechanical Engineering, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada; antoine.tahan@etsmtl.ca

² Institut de Recherche d'Hydro-Québec, Varennes, QC J3X 1S1, Canada; dollon.quentin2@hydroquebec.com (Q.D.); zemouri.mohamedryad@hydroquebec.com (R.Z.); gagnon.martin11@hydroquebec.com (M.G.)

³ Power Factors, Brossard, QC J4Z 1A7, Canada; francis.pelletier@powerfactors.com

* Correspondence: abderrauof-ramdane.benabdesselam.1@ens.etsmtl.ca

Abstract: In the dynamic landscape of renewable energy, the primary goal continues to be the enhancement of competitiveness through the implementation of cutting-edge technologies. This requires a strategic focus on reducing energy costs and maximizing system performance. Within this framework, the continuous online monitoring of assets is essential for efficient operations, by conducting measurements that describe the condition of various components. However, the execution of these measurements can present technical and economic obstacles. To overcome these challenges, the implementation of indirect measurement techniques emerges as a viable solution. By leveraging measurements obtained in easily accessible areas, these methods enable the estimation of quantities in regions that would otherwise be inaccessible. This approach improves the monitoring process's efficiency and provides previously unattainable information. Adopting indirect measurement techniques is also cost-effective, allowing the replacement of expensive sensors with existing infrastructure, which cuts down on installation costs and labor. This paper offers a detailed state-of-the-art review by providing an in-depth examination and classification of indirect measurement techniques and virtual sensing methods applied in the field of renewable energies. It also identifies and discusses the existing challenges and limitations within this topic and explores potential future developments.



Citation: Benabdesselam, A.; Dollon, Q.; Zemouri, R.; Pelletier, F.; Gagnon, M.; Tahan, A. On the Use of Indirect Measurements in Virtual Sensors for Renewable Energies: A Review.

Electronics **2024**, *13*, 1545. <https://doi.org/10.3390/electronics13081545>

Keywords: indirect measurements; virtual sensors; renewable energies; condition monitoring; diagnosis

Figure-A I-1 Article de revue publié

Wind Turbine Remaining Useful Life Estimation Based on a Variational Autoencoder Latent Space Projecting SCADA Data

Adaiton Oliveira-Filho, *Member, IEEE*, Ryad Zemouri, Monelle Comeau, Abderraouf Benabdesselam, Antoine Tahan

Abstract—This paper introduces an approach for estimating the remaining useful life (RUL) of multiple wind turbine (WT) components using data from the supervisory control and data acquisition (SCADA) system. The proposed approach builds on prior work by the authors on WT condition monitoring and visualization using the variational autoencoder embedded with a classifier (VAEC) model. The VAEC enables projecting the high-dimensional physical description given by SCADA measurements into a low-dimensional latent space, where WT degradation modes manifest as trajectories evolving from the healthy cluster toward degraded condition clusters. The RUL estimation proposed in this paper relies on defining the end-of-life (EOL) as a subset of the latent space, thereby transforming the RUL estimation problem into determining the time remaining until the trajectory reaches the EOL region for a given condition. The RUL prediction model applies an exponentially weighted moving average (EWMA) to the non-monotonic latent variable time series to capture its trend. This trend is then combined with Monte Carlo simulation to provide a probabilistic characterization of the RUL. Case studies based on real SCADA datasets demonstrate the effectiveness of the proposed method, with the best configurations achieving a mean absolute percentage error (MAPE) of 18.3% and a cumulative relative accuracy (CRA) of 0.82. In addition to its predictive capability, the approach provides interpretable latent space visualizations of system health, supporting transparency and trust in predictive maintenance planning.

Index Terms—Wind turbines, remaining useful life, prognostics, latent variable models, SCADA data, Monte Carlo simulation.

O&M	Operation and maintenance.
PHM	Prognostics and health management.
RUL	Remaining useful life.
SCADA	Supervisory control and data acquisition.
VAE	Variational autoencoder.
VAEC	Variational autoencoder with classification.
WT	Wind turbine.

I. INTRODUCTION

WIND turbines are one of the key technological solutions driving the decarbonization of the energy sector, and their contribution to the global energy mix is expected to grow substantially in the coming decades [1]. Minimizing O&M costs remains a critical lever to sustain the downward trend in the levelized cost of wind energy, particularly as wind farms scale up globally [2].

WT components operate under harsh environmental and mechanical conditions, making them prone to degradation and failure within the typical WT lifetime of 20–30 years. In the present work, *degradation* refers to *anomalies* or *deviations* in measured variables that can be associated with root causes at the level of components. A WT may continue operating under such conditions, albeit continuing degradation, and increased risk of component failure and production losses. A *failure*, in contrast, implies the shutdown of the WT, either due to

Figure-A I-2 Première page de l'article de journal – coauteur

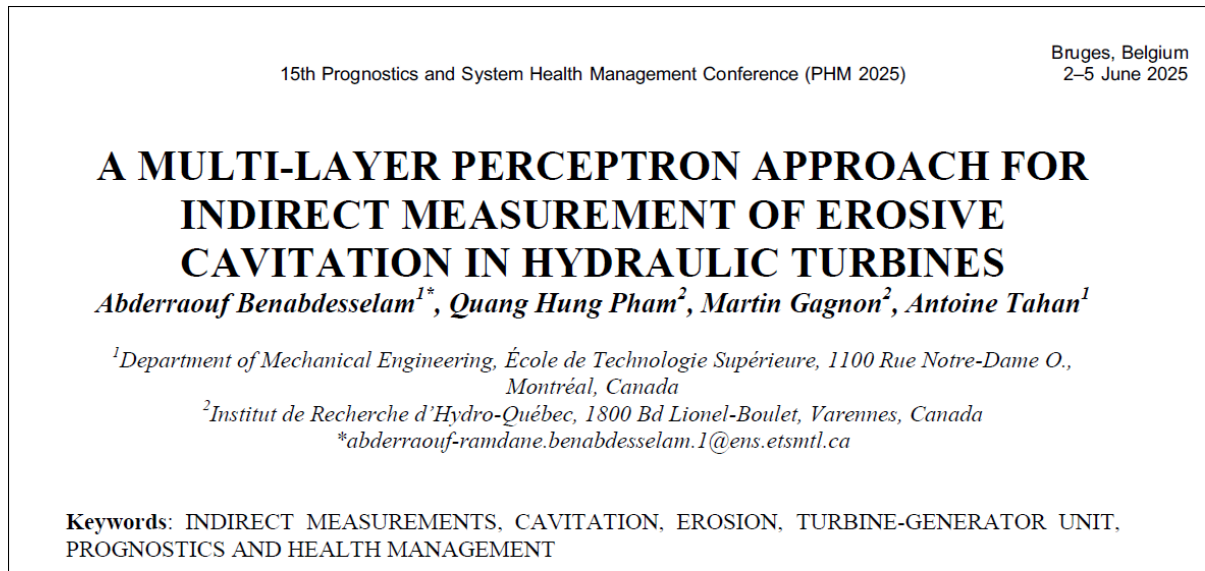


Figure-A I-3 Article de conférence publié

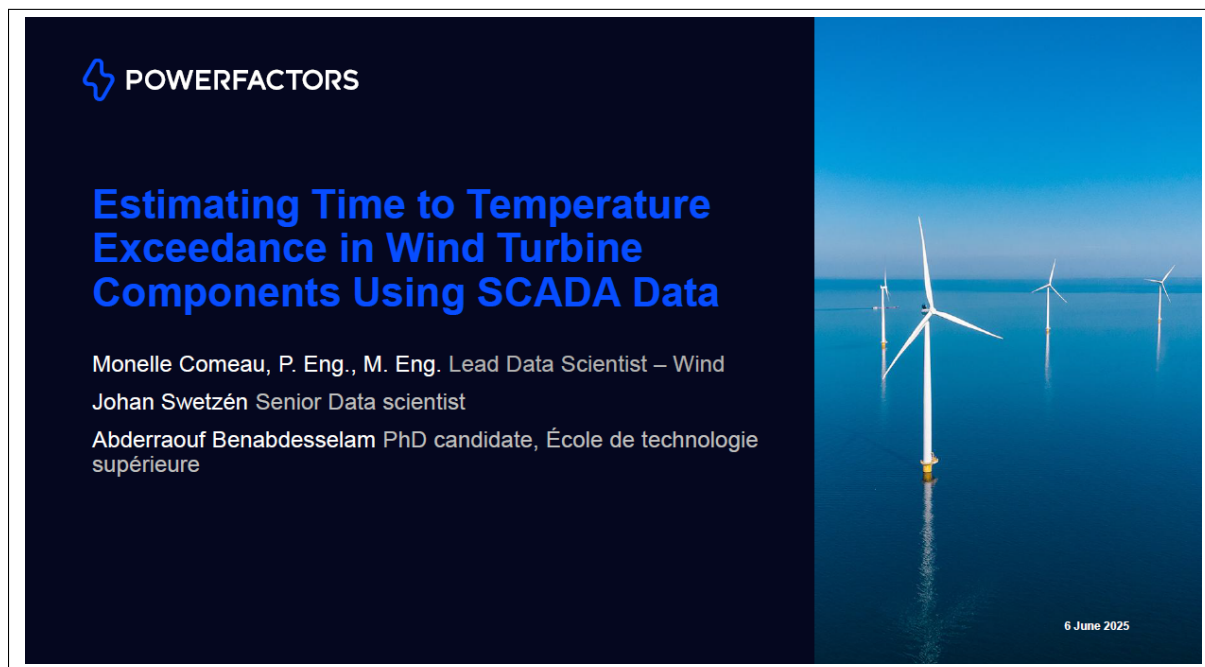


Figure-A I-4 Présentation sur le pronostic des éoliennes

BIBLIOGRAPHIE

- Alagha, N., Khairuddin, A. S. M., Haitamar, Z. N., Al-Khatib, O. & Kanesan, J. (2025). Artificial Intelligence in Wind Turbine Fault Detection and Diagnosis : Advances and Perspectives. *Energies*, 18, 1680. doi : 10.3390/en18071680.
- American Society of Mechanical Engineers. (2009). Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. New York, NY, USA. Repéré à ASMEStandardV&V20--2009.
- Angelopoulos, A. N. & Bates, S. (2022). A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification.
- Avellan, F. (2004). Introduction to Cavitation in Hydraulic Machinery. *Proceedings of the 6th International Conference on Hydraulic Machinery and Hydrodynamics*. Repéré à <https://infoscience.epfl.ch/server/api/core/bitstreams/9c4875bd-4a05-49c1-bba5-020726f89f57/content>.
- Balluff, S., Bendfeld, J. & Krauter, S. (2015). Short Term Wind and Energy Prediction for Offshore Wind Farms Using Neural Networks. *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 379–382.
- Barber, R., Candès, E. J., Ramdas, A. & Tibshirani, R. J. (2023). Conformal Prediction Beyond Exchangeability. *The Annals of Statistics*, 51(2), 816–845. doi : 10.1214/23-AOS2276.
- Barber, R. F., Candès, E. J., Ramdas, A. & Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*. arXiv :1905.02928.
- Barber, R. F., Candès, E. J., Ramdas, A. & Tibshirani, R. J. (2022). Conformal Prediction Beyond Exchangeability. *arXiv preprint*. Repéré à <https://arxiv.org/abs/2202.13415>.
- Bark, G., Grekula, M., Bensow, R. & Berchiche, N. (2009). On some physics to consider in numerical simulation of erosive cavitation. *Proceedings of the 7th International Symposium on Cavitation*.
- Basora, L., Viens, A., Arias Chao, M. & Olive, X. (2025). A benchmark on uncertainty quantification for deep learning prognostics. *Reliability Engineering & System Safety*, 253, 110513. doi : 10.1016/j.ress.2024.110513.
- Benabdesselam, A., Dollon, Q., Zemouri, R., Pelletier, F., Gagnon, M. & Tahan, A. (2024). On the Use of Indirect Measurements in Virtual Sensors for Renewable Energies : A Review. *Electronics*, 13(8), 1545.

- Benabdesselam, A., Pham, Q. H., Gagnon, M. & Tahan, A. (2025). A multi-layer perceptron approach for indirect measurement of erosive cavitation in hydraulic turbines. *Prognostics and System Health Management Conference (PHM 2025)*, 10(IET Conference Proceedings), 29–34. doi : 10.1049/icp.2025.2327.
- Benabdesselam, A., Comeau, M., Oliveira-Filho, A., Gagnon, M. & Tahan, A. (2026a). Prognostics of Thermal Anomalies in Wind Turbines via Deep Learning and Conformal Predictions with SCADA Data. *IEEE Access*. Under Review.
- Benabdesselam, A., Gagnon, M. & Tahan, A. (2026b). Diagnostic of Erosive Cavitation in Hydraulic Turbines from Indirect Measurements Using SCADA Data With Uncertainty Quantification. *Results in Engineering*, 30, 110456. doi : <https://doi.org/10.1016/j.rineng.2026.110456>.
- Benker, M., Furtner, L., Semm, T. & Zaeh, M. F. (2021). Utilizing uncertainty information in remaining useful life estimation via Bayesian neural networks and Hamiltonian Monte Carlo. *Journal of Manufacturing Systems*, 799–807. doi : 10.1016/j.jmsy.2020.11.005.
- Beretta, M., Julian, A., Sepulveda, J., Cusidó, J. & Porro, O. (2021). An Ensemble Learning Solution for Predictive Maintenance of Wind Turbines' Main Bearing. *Sensors*, 21(4), 1512. doi : 10.3390/s21041512.
- Bhowmik, B., Tripura, T., Hazra, B. & Pakrashi, V. (2019). First order eigen perturbation techniques for real time damage detection of vibrating systems : Theory and applications. *Applied Mechanics Reviews*, 71 (6), 060801. doi : 10.1115/1.4044287.
- Bhowmik, B., Tripura, T., Hazra, B. & Pakrashi, V. (2020). Real time structural modal identification using recursive canonical correlation analysis and application towards online structural damage detection. *Journal of Sound and Vibration*, 468, 115101. doi : 10.1016/j.jsv.2019.115101.
- Bilbao, J., Lourens, E.-M., Schulze, A. & Ziegler, L. (2022). Virtual Sensing in an Onshore Wind Turbine Tower Using a Gaussian Process Latent Force Model. *Data-Centric Engineering*, 3, 35.
- Bilus, I., Hocevar, M., Dular, M. & Lesnik, L. (2020). Numerical Prediction of Various Cavitation Erosion Mechanisms. *Journal of Fluids Engineering*, 142(4).
- Bourdon, P. (2000). *Détection vibratoire de l'érosion de cavitation des turbines Francis*. (Ph.D. Thesis, École Polytechnique de Montréal, Montréal, Canada).

- Branlard, E., Jonkman, J., Dana, S. & Doubrawa, P. (2020). A digital twin based on OpenFAST linearizations for real-time load and fatigue estimation of land-based turbines. *Journal of Physics : Conference Series*, 1618, 022030.
- Brownjohn, J. M. W., De Stefano, A., Xu, Y.-L., Wenzel, H. & Aktan, A. E. (2011). Vibration-Based Monitoring of Civil Infrastructure : Challenges and Successes. *Journal of Civil Structural Health Monitoring*. doi : 10.1007/s13349-011-0009-5.
- Cambron, P., Tahan, A., Masson, C. & Pelletier, F. (2017). Bearing Temperature Monitoring of a Wind Turbine Using Physics-Based Model. *Journal of Quality in Maintenance Engineering*, 23(4), 399–414.
- Campoverde-Vilela, L., Feijóo, M. C., Vidal, Y., Sampietro, J. & Tutivén, C. (2023). Anomaly-Based Fault Detection in Wind Turbine Main Bearings. *Wind Energy Science*, 8, 557–574.
- Cao, L., Zhang, H., Meng, Z. & Wang, X. (2023). A parallel GRU with dual-stage attention mechanism model integrating uncertainty quantification for probabilistic RUL prediction of wind turbine bearings. *Reliability Engineering & System Safety*, 235, 109197. doi : 10.1016/j.ress.2023.109197.
- Carroll, J., Koukoura, S., McDonald, A., Charalambous, A., Weiss, S. & McArthur, S. (2019). Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. *Wind Energy*, 22, 360–375. doi : 10.1002/we.2290.
- Chatenet, Q. (2020). *Contribution à la modélisation et à la prédiction de la durée de vie des roues de turbines hydroélectriques*. (Ph.D. Thesis, École de Technologie Supérieure, Montréal, Canada).
- Chatenet, Q., Gagnon, M., Tòn-Thât, L., Remy, E., Fouladirad, M. & Tahan, A. (2020). Stochastic modeling of cavitation erosion in Francis runner. *International Journal of Fluid Machinery and Systems*, 13(2), 400–408. doi : 10.5293/IJFMS.2020.13.2.400.
- Chatenet, Q., Tahan, A. S., Fouladirad, M., Gagnon, M. & Remy, E. (2021). Maintenance strategies comparison for Francis turbine runner subjected to cavitation. *Sustainability*, 13(19), 10838. doi : 10.3390/su131910838.
- Chen, D., Qin, Y., Wang, Y. & Zhou, J. (2021). Health Indicator Construction by Quadratic Function-Based Deep Convolutional Auto-Encoder and Its Application into Bearing RUL Prediction. *ISA Transactions*, 114, 44–56.

- Cheng, C., Ma, G., Zhang, Y., Sun, M., Teng, F., Ding, H. & Yuan, Y. (2020). A Deep Learning-Based Remaining Useful Life Prediction Approach for Bearings. *IEEE/ASME Transactions on Mechatronics*, 25, 1243–1254. doi : 10.1109/TMECH.2020.2971503.
- Cheng, Z., Li, X., Peng, G., Deng, Y., Xie, Z. & Liu, L. (2023). Transfer Deep Learning Network for Rolling Bearing Fault Diagnosis of Wind Turbines. *Journal of Physics : Conference Series*, 2503, 012095. doi : 10.1088/1742-6596/2503/1/012095.
- Clermont-Bodson, P. M. (2015). *Modélisation statistique de l'érosion de cavitation d'une turbine hydraulique selon les paramètres d'opération*. (M.Sc. Thesis, Université de Montréal, Montréal, Canada).
- Cobelli, P., Shukla, P., Nesmachnow, S. & Draper, M. (2023). Physics informed neural networks for wind field modeling in wind farms. *Journal of Physics : Conference Series*, 2505, 012051.
- Cuesta Arrillaga, J., Leturiondo, U., Vidal, Y. & Pozo, F. (2025). A review of prognostics and health management techniques in wind energy. *Reliability Engineering & System Safety*, 260, 111004. doi : 10.1016/j.ress.2025.111004.
- Cui, L., Wang, X., Wang, H. & Ma, J. (2020). Research on Remaining Useful Life Prediction of Rolling Element Bearings Based on Time-Varying Kalman Filter. *IEEE Transactions on Instrumentation and Measurement*, 69, 2858–2867. doi : 10.1109/TIM.2019.2924509.
- da Silva, H. R., Ferraresi, V. A. & Arencibia, R. V. (2017). Operational aspects of coatings welds : Erosive wear and cavitation. *ASME International Mechanical Engineering Congress and Exposition (IMECE)*, 14. doi : 10.1115/IMECE2017-70169.
- Daigle, M. J. & Goebel, K. (2011). A Model-Based Prognostics Approach Applied to Pneumatic Valves. *International Journal of Prognostics and Health Management*, 2(2), 84–99.
- Dao, C., Kazemtabrizi, B. & Crabtree, C. (2019). Wind turbine reliability data review and impacts on levelised cost of energy. *Wind Energy*, 22(12), 1848–1871.
- De Giorgi, M., Congedo, P. & Malvoni, M. (2014). Photovoltaic power forecasting using statistical methods : impact of weather data. *IET Science, Measurement & Technology*, 8, 90-97.
- De Paiva Silva, G., Bacci da Silva, M. & De Oliveira, D. (2023). Influence of Abrasive Deburring in Indirect Tool Wear Measurement in Micromilling of Inconel 718. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 45, 262.

- Deng, X., Jiang, D. & Liu, Y. (2017). Rule-Based Expert System for On-line Diagnosis of Wind Turbines. *ITM Web of Conferences*, 12, 07005. doi : 10.1051/itmconf/20171207005.
- Der Kiureghian, A. & Ditlevsen, O. (2009). Aleatory or epistemic ? Does it matter ? *Structural Safety*, 31(2), 105–112.
- Diaconis, P. & Freedman, D. (1986). On the Consistency of Bayes Estimates. *The Annals of Statistics*, 14(1), 1–26.
- Diagne, I., Gagnon, M. & Tahan, A. (2016). Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements. *IOP Conference Series : Earth and Environmental Science*, 49(7), 072014.
- Ding, M., Wang, L. & Bi, R. (2011). An ANN-based Approach for Forecasting the Power Output of Photovoltaic System. *Procedia Environmental Sciences*, 11, 1308-1315.
- Djeziri, M., Benmoussa, S. & Sanchez, R. (2018). Hybrid method for remaining useful life prediction in wind turbine systems. *Renewable Energy*, 116, 173-187. doi : <https://doi.org/10.1016/j.renene.2017.05.020>. Real-time monitoring, prognosis and resilient control for wind energy systems.
- Dolara, A., Leva, S. & Manzolini, G. (2015). Comparison of different physical models for PV power output prediction. *Solar Energy*, 115, 83-99.
- Dollon, Q., Labbé, P. & Léonard, F. (2023). Promoting Explainability in Data-Driven Models for Anomaly Detection : A Step Toward Diagnosis. *Annual Conference of the PHM Societys*, 15(1). doi : 10.36001/phmconf.2023.v15i1.3509.
- Dollon, Q. (2021). *Analyse modale opérationnelle des roues de turbines hydroélectriques par l'étude de régimes transitoires*. (Thèse de doctorat, École de Technologie Supérieure, Montréal).
- Du, Y., Cheng, X., Liu, Y., Dou, S., Tu, J., Liu, Y. & Su, X. (2023). Gearbox Fault Diagnosis Method Based on Improved MobileNetV3 and Transfer Learning. *Technical Gazette*, 30(1), 198–206. doi : 10.17559/TV-20221025165425.
- Feng, Y., Qiu, Y., Crabtree, C. J., Long, H. & Tavner, P. J. (2013). Monitoring Wind Turbine Gearboxes. *Wind Energy*, 16, 728–740.
- Fernández-Osete, I., Bermejo, D., Ayneto-Gubert, X. & Escalar, X. (2024). Review of the Uses of Acoustic Emissions in Monitoring Cavitation Erosion and Crack Propagation. *Foundations*, 4(1), 114–133.

- Fivel, M., Franc, J.-P. & Roy, S. C. (2015). Towards numerical prediction of cavitation erosion. *Interface Focus*, 5(5), 20150013. doi : 10.1098/rsfs.2015.0013.
- Gagnon, M., Vouligny, L., Cauchon, L. & Giroux, A. M. (2022). Virtual Sensors to Generate Turbine Runner Blade Strains from Indirect Measurements. *IOP Conference Series : Earth and Environmental Science*, 1079, 012067.
- Gagnon, M., Vouligny, L., Cauchon, L., Pham, Q. H. & Mai, V. (2024). Quantification d'incertitude dans un contexte de maintenance et d'évaluation des risques industriels à l'aide d'apprentissage automatique. *Congrès Lambda Mu 24*.
- Gao, M., Li, J., Hong, F. & Long, D. (2019). Short-Term Forecasting of Power Production in a Large-Scale Photovoltaic Plant Based on LSTM. *Applied Sciences*, 9(3192).
- Garan, M., Tidriri, K. & Kovalenko, I. (2022). A Data-Centric Machine Learning Methodology : Application on Predictive Maintenance of Wind Turbines. *Energies*, 15, 826. doi : 10.3390/en15030826.
- Garcia, J. L. T., Calderon, E. C., Avalos, G. G., Heras, E. R. & Tshikala, A. M. (2019). Forecast of Daily Output Energy of Wind Turbine Using sARIMA and Nonlinear Autoregressive Models. *Advances in Mechanical Engineering*, 11, 1–15.
- Gbashi, S. M., Madushele, N., Olatunji, O. O., Adedeei, P. A. & Jen, T.-C. (2022). Wind Turbine Main Bearing : A Mini Review of Its Failure Modes and Condition Monitoring Techniques. *Proc. 2022 IEEE 13th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*, pp. 127–134.
- Gensler, A., Henze, J., Sick, B. & Raabe, N. (2016). Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary*, 2858–2865.
- Gneiting, T. & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), 359–378. doi : 10.1198/016214506000001437.
- Goebel, K., Celaya, J., Sankararaman, S., Roychoudhury, I., Daigle, M. & Saxena, A. (2017). *Prognostics : The Science of Making Predictions*. CreateSpace Independent Publishing Platform.
- Gomes, G. S. P., Lopes, S. M. A., Araujo, D. C. P., Flauzino, R. A., Pinto, M. M. & Alves, M. E. G. (2024). Wind Turbine Remaining Useful Life Prediction Using Small Dataset and Machine Learning Techniques. *Journal of Control, Automation and Electrical Systems*, 35, 337–345. doi : 10.1007/s40313-024-01076-y.

- Gouriveau, R., Medjaher, K. & Zerhouni, N. (2017). *Du concept de PHM à la maintenance prédictive 1 : Surveillance et pronostic*. ISTE Group.
- Graves, T. & Hamada, M. (2005). Making Inferences with Indirect Measurements. *Quality Engineering*, 17, 555–559.
- Gregg, S. W., Steele, J. P. H. & Van, D. L. (2020). Feature selection for monitoring erosive cavitation on a hydroturbine. *International Journal of Prognostics and Health Management*, 8(1), 2528. doi : 10.36001/ijphm.2017.v8i1.2528.
- Guck, D., Bindl, S., Fernandez, F. A. & Stetter, R. (2024). CARE to Compare : A real-world dataset for anomaly detection in wind turbine dat. *arXiv*.
- Guo, L., Li, N., Jia, F., Lei, Y. & Lin, J. (2017). A Recurrent Neural Network Based Health Indicator for Remaining Useful Life Prediction of Bearings. *Neurocomputing*, 240, 98–109. doi : 10.1016/j.neucom.2017.02.045.
- Guo, L., Lei, Y., Li, N., Yan, T. & Li, N. (2018). Machinery Health Indicator Construction Based on Convolutional Neural Networks Considering Trend Burr. *Neurocomputing*, 292, 142–150.
- Guo, P., Infield, D. & Yang, X. (2012). Wind Turbine Generator Condition-Monitoring Using Temperature Trend Analysis. *IEEE Transactions on Sustainable Energy*, 3(1), 124–133. doi : 10.1109/TSTE.2011.2163430.
- Guo, Y., Sheng, S., Phillips, C. T., Keller, J., Veers, P. & Williams, L. (2020). A Methodology for Reliability Assessment and Prognosis of Bearing Axial Cracking in Wind Turbine Gearboxes. *Renewable and Sustainable Energy Reviews*, 127, 109888. doi : 10.1016/j.rser.2020.109888.
- Haghi, R., Stagg, C. & Crawford, C. (2024). Wind Turbine Damage Equivalent Load Assessment Using Gaussian Process Regression Combining Measurement and Synthetic Data. *Energies*, 17, 346.
- Han, S., Qiao, Y., Yan, J., Liu, Y., Li, L. & Wang, Z. (2019). Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. *Applied Energy*, 239, 181–191.
- Han, S. & Feng, Z. (2023). Deep Residual Joint Transfer Strategy for Cross-Condition Fault Diagnosis of Rolling Bearings. *Journal of Dynamics, Monitoring and Diagnostics*, 2, 51–60. doi : 10.37965/jdmd.2023.147.

- Hartigan, J. A. (1983). Asymptotic Normality of Posterior Distributions. Dans *Bayes Theory* (pp. 107–118). New York, USA : Springer. doi : 10.1007/978-1-4613-8242-3_11.
- He, D., Xu, Y., Sun, H., Jin, Z. & Zhao, M. (2025a). Self-supervised learning for vehicle bearing fault diagnosis based on time–frequency dual-domain contrast and fusion. *Nonlinear Dynamics*, 113, 17385–17412. doi : 10.1007/s11071-025-11101-7.
- He, W., Jiang, Z., Xiao, T., Xu, Z. & Li, Y. (2025b). A Survey on Uncertainty Quantification Methods for Deep Learning. *ACM Computing Surveys*, 37(4), Article 111. Repéré à <https://arxiv.org/abs/2302.13425>.
- Helsen, J., Devriendt, C., Weijtjens, W. & Guillaume, P. (2015). Condition Monitoring by Means of SCADA Analysis. *Proceedings of the European Wind Energy Association International Conference*.
- Hoblos, R., Dridi, N., Zerhouni, N. & Al Masry, Z. (2025). An Uncertainty Quantification Method Based on Evidence Theory and Conformal Prediction. *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pp. 870–874.
- Hofmann, J., Thiébaud, C., Riondet, M., Lhuissier, P., Gaudion, S. & Fivel, M. (2023). Comparison of acoustic and hydrodynamic cavitation : Material point of view. *Physics of Fluids*, 35(1), 013306. doi : 10.1063/5.0132085.
- Hu, C., Huang, Y., Xu, C., Sun, Q., Zhang, Z., Yang, Y. & Zhang, L. (2025). A New Two-Stage Probabilistic Remaining Useful Life Prediction Method for Wind Turbines. *Wind Energy*. doi : 10.1002/we.70049. Early View.
- Hu, X., Tan, L. & Tang, T. (2024). M2BIST-SPNet : RUL prediction for railway signaling electromechanical devices. *The Journal of Supercomputing*, 80, 16744–16774. doi : 10.1007/s11227-024-06111-y.
- Huang, C.-G., Huang, H.-Z., Li, Y.-F. & Peng, W. (2021). A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. *Journal of Manufacturing Systems*, 61, 757–772. doi : 10.1016/j.jmsy.2021.03.012.
- Ibrahim, M., Rassölkin, A., Vaimann, T., Kallaste, A., Zakis, J., Hyunh, V. K. & Pomarnacki, R. (2023a). Digital Twin as a Virtual Sensor for Wind Turbine Applications. *Energies*, 16, 6246.
- Ibrahim, R., Tahan, A., Kedjar, B., Al-Haddad, K., Zemouri, R. & Merkhoul, A. (2023b). Vibroacoustic Signal-Based Diagnosis for Rotor Faults in Large Synchronous Machines - an Overview. *2023 6th International Conference on Renewable Energy for Developing Countries (REDEC)*, pp. 72–77. doi : 10.1109/REDEC58286.2023.10208168.

- Iliopoulos, A., Devriendt, C., Guillaume, P. & van Hemelrijck, D. (2014). Continuous Fatigue Assessment of an Offshore Wind Turbine Using a Limited Number of Vibration Sensors. *EWSHM - 7th European Workshop on Structural Health Monitoring*. Repéré à <https://inria.hal.science/hal-01020445>.
- Iliopoulos, A., Weijtjens, W., Van Hemelrijck, D. & Devriendt, C. (2015). Prediction of dynamic strains on a monopile offshore wind turbine using virtual sensors. *Journal of Physics : Conference Series - 11th International Conference on Damage Assessment of Structures*, 628. doi : 10.1088/1742-6596/628/1/012108.
- Iliopoulos, A., Weijtjens, W., Van Hemelrijck, D. & Devriendt, C. (2016). Full-Field Strain Prediction Applied to an Offshore Wind Turbine. *Model Validation and Uncertainty Quantification*, 3(34), 349-357. doi : 10.1007/978-3-319-29754-5_34.
- Iliopoulos, A., Weijtjens, W., Van Hemelrijck, D. & Devriendt, C. (2017). Fatigue assessment of offshore wind turbines on monopile foundations using multi-band modal expansion. *Wind Energy*, 20(8), 1463-1479. doi : <https://doi.org/10.1002/we.2104>.
- International Energy Agency. (2023). Share of renewable electricity generation by technology, 2000–2028. Repéré le 2026-02-11 à <https://www.iea.org/data-and-statistics/charts/share-of-renewable-electricity-generation-by-technology-2000-2028>.
- IRENA. (2023). Renewable Energy Statistics 2023. Repéré le 2024-03-13 à <https://www.irena.org/Publications/2023/Jul/Renewable-energy-statistics-2023>.
- Jankauskas, M., Serackis, A., Šapurov, M., Pomarnacki, R., Baskys, A., Hyunh, V. K., Vaimann, T. & Zakis, J. (2023). Exploring the Limits of Early Predictive Maintenance in Wind Turbines Applying an Anomaly Detection Technique. *Sensors*, 23, 5695. doi : 10.3390/s23125695.
- Javed, K., Gouriveau, R. & Zerhouni, N. (2017). State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing*, 94, 214-236. doi : <https://doi.org/10.1016/j.ymssp.2017.01.050>.
- Ji, G., Dong, Z., Qiao, H. & Xu, D. (2008). SVR-Based Soft Sensor for Effective Wind Speed of Large-Scale Variable Speed Wind Turbine. *2008 Fourth International Conference on Natural Computation*, 2, 193-196.
- Jiang, C., Liu, X., Liu, Y., Xie, M., Liang, C. & Wang, Q. (2022). A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism. *Electronics*, 11, 3616. doi : 10.3390/electronics11213616.

- Jin, X., Que, Z., Sun, Y., Guo, Y. & Qiao, W. (2019). A Data-Driven Approach for Bearing Fault Prognostics. *IEEE Transactions on Industry Applications*, 55, 3394–3403. doi : 10.1109/TIA.2019.2907666.
- Jonas, S., Anagnostos, D., Brodbeck, B. & Meyer, A. (2022). Vibration Fault Detection in Wind Turbines Based on Normal Behaviour Models without Feature Engineering. *arXiv*. doi : 10.48550/arXiv.2206.12452.
- Jose, S., Zemouri, R., Nguyen, K. T., Medjaher, K., Tahan, A. & Lévesque, M. (2024). From Fragments to Futures : Construction of Synthetic Run-to-Failure Trajectories for Fault State Prognostics. *2024 Prognostics and System Health Management Conference (PHM)*, pp. 306-311. doi : 10.1109/PHM61473.2024.00062.
- Jose, S., Zemouri, R., Nguyen, K. T., Medjaher, K., Lévesque, M. & Tahan, A. (2025). Prognostics of complex machinery with sparse multilabel multimodal run-to-failure data : A graph neural network approach. *Advanced Engineering Informatics*, 65, 103361. doi : <https://doi.org/10.1016/j.aei.2025.103361>.
- Kahwati, G., Cauchon, L., Pham, Q. H., Vouligny, L. & Gagnon, M. (2025, Jun). Building a hydro-generator rotor temperature virtual sensor using machine-learning. *IET Conference Proceedings*. doi : 10.1049/icp.2025.2334.
- Kim, K.-H., Chahine, G. L., Franc, J.-P. & Karimi, A. (2014). *Advanced Experimental and Numerical Techniques for Cavitation Erosion Prediction*. Dordrecht, The Netherlands : Springer.
- Konstantinou, M., Peratikou, S. & Charalambides, A. G. (2021). Solar Photovoltaic Forecasting of Power Output Using LSTM Networks. *Atmosphere*, 12, 124.
- Krumenacker, L., Fortes-Patella, R. & Archer, A. (2014). Numerical estimation of cavitation intensity. *IOP Conference Series : Earth and Environmental Science*, 22(5), 052014. doi : 10.1088/1755-1315/22/5/052014.
- Lakshminarayanan, B., Pritzel, A. & Blundell, C. (2017). Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*.
- Le, T.-T., Lee, S.-J., Dinh, M.-C. & Park, M. (2024). Design of an Improved Remaining Useful Life Prediction Model Based on Vibration Signals of Wind Turbine Rotating Components. *Energies*, 17(1), 19. doi : 10.3390/en17010019.
- Le Cam, L. (1986). Approximately Gaussian Posterior Distributions. Dans *Asymptotic Methods in Statistical Decision Theory* (pp. 336–345). New York, USA : Springer.

- Leibfried, F., Dutordoir, V., John, S. T. & Durrande, N. (2022). A Tutorial on Sparse Gaussian Processes and Variational Inference.
- Li, D., Chen, J., Huang, R., Chen, Z. & Li, W. (2024a). Sensor-aware CapsNet : Towards trustworthy multisensory fusion for remaining useful life prediction. *Journal of Manufacturing Systems*, 72, 26–37. doi : 10.1016/j.jmsy.2023.11.009.
- Li, M., Hong, Y., Si, B., Ding, Y., Tang, Z., Wang, R. & He, X. (2024b). Numerical and experimental study of the cavitation performance of a composite propeller. *Physics of Fluids*, 36(10), 103303. doi : 10.1063/5.0228410.
- Li, X. & Zhang, W. (2022). Physics-Informed Deep Learning Model in Wind Turbine Response Prediction. *Renewable Energy*, 185, 932–944.
- Li, X., Zhang, W. & Ding, Q. (2019). Deep Learning-Based Remaining Useful Life Estimation of Bearings Using Multi-Scale Feature Extraction. *Reliability Engineering & System Safety*, 182, 208–218.
- Lichtenstein, T., von Maltzahn, V. & Pelka, K. (2024). Empirical Physics-Based Normal Behavior Modeling of Drive Train Temperatures With High-Resolution Wind Turbine Operating Data. *Wind Energy*, e2961. doi : 10.1002/we.2961.
- Lin, K., Pan, J., Xi, Y., Wang, Z. & Jiang, J. (2024). Vibration Anomaly Detection of Wind Turbine Based on Temporal Convolutional Network and Support Vector Data Description. *Engineering Structures*, 306, 117848. doi : 10.1016/j.engstruct.2024.117848.
- Liu, H., Wang, Z., Zeng, T., Wang, H., Chan, S.-C. & Ran, L. (2024). Wind Turbine Generator Failure Analysis and Fault Diagnosis : A Review. *IET Renewable Power Generation*, 18. doi : 10.1049/rpg2.13104.
- Liu, J. & Zio, E. (2017). System Dynamic Reliability Assessment and Failure Prognostics. *Reliability Engineering & System Safety*, 160, 21–36. doi : 10.1016/j.res.2016.12.003.
- Liu, J.-H., Corbita, N. T., J., Lee, R.-M. & Wang, C.-C. (2022). Wind Turbine Anomaly Detection Using Mahalanobis Distance and SCADA Alarm Data. *Applied Sciences*, 12, 8661. doi : 10.3390/app12178661.
- Liu, X. T., Feng, F. Z. & Si, A. W. (2012). Condition based monitoring, diagnosis and maintenance on operating equipments of a hydraulic generator unit. *IOP Conference Series : Earth and Environmental Science*, 15(4), 042014.
- Ljung, L. (1999). *System Identification : Theory for the User* (éd. 2). Upper Saddle River, USA : Prentice Hall PTR.

- Lu, B., Li, Y., Wu, X. & Yang, Z. (2009). A Review of Recent Advances in Wind Turbine Condition Monitoring and Fault Diagnosis. *Proc. 2009 IEEE Power Electronics and Machines in Wind Applications (PEMWA)*, pp. 1–7.
- Maes, K., De Roeck, G., Lombaert, G., Iliopoulos, A., Van Hemelrijck, D., Devriendt, C. & Guillaume, P. (2015). Continuous strain prediction for fatigue assessment of an offshore wind turbine using Kalman filtering techniques. *2015 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS) Proceedings*, pp. 44-49. doi : 10.1109/EESMS.2015.7175850.
- Mai, Y., Fouladirad, M. & Xue, J. (2026). Imperfect condition-based maintenance policy for stochastic degradation systems in dynamic environments. *Reliability Engineering & System Safety*, 265(Part B), 111583. doi : 10.1016/j.ress.2025.111583.
- Maldonado-Correa, J., Martín-Martínez, S., Artigao, E., Honrubia-Escribano, A. & Gómez-Lázaro, E. (2020). Using SCADA Data for Wind Turbine Condition Monitoring : A Systematic Literature Review. *Energies*, 13(12), 3132. doi : 10.3390/en13123132.
- Manna, D., Mitici, M. & Dalla Vedova, M. D. L. (2024). System-level Probabilistic Remaining Useful Life Prognostics and Predictive Inspection Planning for Wind Turbines. *Proceedings of the 8th European Conference of the Prognostics and Health Management Society*, pp. 802–813.
- Martínez Montaña, M. & Sanz-Bobi, M. A. (2018). Anomaly Detection Indicators of a Wind Turbine Gearbox Based on Feature Extraction from Its Vibration Performance. *Proc. Eur. Conf. Prognostics and Health Management Society*, 4(1), 1–10. Repéré à <https://www.phmsociety.org/node/1645>.
- Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., Refaat, S. S., Abu-Rub, H. & Oueslati, F. S. (2021). An Effective Hybrid NARX-LSTM Model for Point and Interval PV Power Forecasting. *IEEE Access*, 9, 36571–36588.
- Mehlan, F. C., Nejad, A. R. & Gao, Z. (2022). Digital Twin Based Virtual Sensor for Online Fatigue Damage Monitoring in Offshore Wind Turbine Drivetrains. *ASME Journal of Offshore Mechanics and Arctic Engineering*, 144, 060901.
- Mellit, A. & Pavan, A. (2010). A 24-h forecast of solar irradiance using artificial neural network : Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy*, 84(5), 807-821.
- Meyer, A. (2022). Vibration Fault Diagnosis in Wind Turbines Based on Automated Feature Learning. *Energies*, 15, 1514. doi : 10.3390/en15041514.

- Mihatsch, M. J., Schmidt, S. & Adams, N. A. (2015). Cavitation erosion prediction based on analysis of flow dynamics and impact load spectra. *Physics of Fluids*, 27(10), 103302. doi : 10.1063/1.4932175.
- Milani, A. E., Zappalá, D., Castellani, F. & Watson, S. (2025a). Simulating Run-to-Failure SCADA Time Series to Enhance Wind Turbine Fault Detection and Prognosis. *Wind Energy Science*, 10, 653–669. doi : 10.5194/wes-10-653-2025.
- Milani, A. E., Zappalá, D. & Watson, S. J. (2025b). A Hybrid Convolutional Autoencoder Training Algorithm for Unsupervised Bearing Health Indicator Construction. *Engineering Applications of Artificial Intelligence*, 139, 109477. doi : 10.1016/j.engappai.2024.109477.
- Moreira de Oliveira Filho, A. (2024). *Détection, diagnostic et pronostic pour les éoliennes : développement de modèles basés sur l'autoencodeur variationnel utilisant des données SCADA*. (Thèse de doctorat, École de Technologie Supérieure, Montréal).
- Moricz, L., Viharos, Z. J., Németh, A. & Szépligeti, A. (2018). Indirect Measurement and Diagnostics of the Tool Wear for Ceramics Micro-Milling Optimisation. *Journal of Physics : Conference Series*, 10, 102003.
- Mucchielli, P., Bhowmik, B., Hazra, B. & Pakrashi, V. (2020). Higher-order stabilized perturbation for recursive eigen-decomposition estimation. *Journal of Vibration and Acoustics*, 142, 1–11. doi : 10.1115/1.4047302.
- Mucchielli, P., Bhowmik, B., Ghosh, B. & Pakrashi, V. (2021). Real-time accurate detection of wind turbine downtime - An Irish perspective. *Renewable Energy*, 179, 1969–1989. doi : 10.1016/j.renene.2021.07.139.
- Murgia, A., Verbeke, R., Tsiporkova, E., Terzi, L. & Astolfi, D. (2023). Discussion on the Suitability of SCADA-Based Condition Monitoring for Wind Turbine Fault Diagnosis through Temperature Data Analysis. *Energies*, 16(2), 620. doi : 10.3390/en16020620.
- Najwadi, F. & Villarreal, R. (2023). Utilizing the Vector Autoregression Model (VAR) for Short-Term Solar Irradiance Forecasting. *Energy and Power Engineering*, 15, 353-362.
- Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N. & Varnier, C. (2012). PRONOSTIA : An Experimental Platform for Bearings Accelerated Degradation Tests. *Proceedings of the IEEE International Conference on Prognostics and Health Management (PHM 2012)*, pp. 1–8. doi : 10.1109/PHM.2012.6229287.

- Noppe, N., Iliopoulos, A., Weijtjens, W. & Devriendt, C. (2016). Full Load Estimation of an Offshore Wind Turbine Based on SCADA and Accelerometer Data. *Journal of Physics : Conference Series*, 753, 072025.
- Noppe, N., Tatsis, K., Chatzi, E., Devriendt, C. & Weijtjens, W. (2018). Fatigue stress estimation of offshore wind turbine using a Kalman filter in combination with accelerometers. *Proceedings of ISMA 2018 - International Conference on Noise and Vibration Engineering and USD 2018 - International Conference on Uncertainty in Structural Dynamics*, (Proceedings of ISMA 2018 - International Conference on Noise and Vibration Engineering and USD 2018 - International Conference on Uncertainty in Structural Dynamics), 4847-4855. Repéré à <https://www.isma-isaac.be/isma2018/>.
- Oliveira, R. I., Orenstein, P., Ramos, T. & Romano, J. V. (2024). Split Conformal Prediction and Non-Exchangeable Data. *Journal of Machine Learning Research*, 25, 1–38.
- Oliveira-Filho, A., Zemouri, R., Pelletier, F. & Tahan, A. (2021). Remaining Useful Life Estimation Based on the System Trajectory in a Latent Space Representation. *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pp. 1–9.
- Oliveira-Filho, A., Zemouri, R., Cambron, P. & Tahan, A. (2023). Early detection and diagnosis of wind turbine abnormal conditions using an interpretable supervised variational autoencoder model. *Energies*, 16, 4544. doi : 10.3390/en16124544.
- Oliveira-Filho, A., Zemouri, R., Pelletier, F. & Tahan, A. (2024). System condition monitoring based on a standardized latent space and the Nataf transform. *IEEE Access*, 12, 32637–32659. doi : 10.1109/ACCESS.2024.3370949.
- Pagitsch, M., Jacobs, G., Bosse, D. & Duda, T. P. R. (2020). Remaining Useful Life Determination for Wind Turbines. *Journal of Physics : Conference Series*, 1452, 012052. doi : 10.1088/1742-6596/1452/1/012052.
- Paiva, E. B. R. F., Nguyen, H.-N., Lepreux, O. & Bresch-Pietri, D. (2021). A Gaussian Process Based Approach to Estimate Wind Speed Using SCADA Measurements from a Wind Turbine. *IFAC-PapersOnLine*, 54, 65–71.
- Pan, M., Li, C., Gao, R., Huang, Y., You, H., Gu, T. & Qin, F. (2020a). Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. *Journal of Cleaner Production*, 277, 123948.
- Pan, Z., Meng, Z., Chen, Z., Gao, W. & Shi, Y. (2020b). A Two-Stage Method Based on Extreme Learning Machine for Predicting the Remaining Useful Life of Rolling-Element Bearings. *Mechanical Systems and Signal Processing*, 144, 106899. doi : 10.1016/j.ymssp.2020.106899.

- Pandit, R. K., Infield, D. & Kolios, A. (2020). Gaussian process power curve models incorporating wind turbine operational variables. *Energy Reports*, 6, 1658–1669. doi : 10.1016/j.egyr.2020.06.026.
- Park, G., You, D., Oh, K.-Y. & Nam, W. (2022). Natural Frequency Degradation Prediction for Offshore Wind Turbine Structures. *Machines*, 10, 356.
- Pasari, S. & Shah, A. (2020). Time Series Auto-Regressive Integrated Moving Average Model for Renewable Energy Forecasting. Sangwan, K., Herrmann, C. (eds) *Enhancing Future Skills and Entrepreneurship. Sustainable Production, Life Cycle Engineering and Management*.
- Pedro, H. & Coimbra, C. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7), 2017–2028.
- Peng, W., Ye, Z.-S. & Chen, N. (2020). Bayesian Deep-Learning-Based Health Prognostics Toward Prognostics Uncertainty. *IEEE Transactions on Industrial Electronics*, 67(3), 2283–2294. doi : 10.1109/TIE.2019.2907440.
- Perišić, N., Kirkegaard, P. H. & Pedersen, B. J. (2015). Cost-effective shaft torque observer for condition monitoring of wind turbines. *Wind Energy*, 18(1), 1–19. doi : 10.1002/we.1678.
- Peters, A., Sagar, H., Lantermann, U. & el Moctar, O. (2015). Numerical modelling and prediction of cavitation erosion. *Wear*, 338–339, 189–201. doi : 10.1016/j.wear.2015.06.009.
- Piao, S., Huang, R. & Tsung, F. (2025). CRULP : Reliable RUL Estimation Inspired by Conformal Prediction. *IEEE Transactions on Instrumentation and Measurement*, 74, 3505411. doi : 10.1109/TIM.2024.3522678.
- Powers, D. M. W. (2011). Evaluation : From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Presas, A., Valentin, D., Zhao, W., Egusquiza, M., Valero, C. & Egusquiza, E. (2021a). On the Use of Neural Networks for Dynamic Stress Prediction in Francis Turbines by Means of Stationary Sensors. *Renewable Energy*, 170, 652–660.
- Presas, A., Valentin, D., Zhao, W., Valero, C., Egusquiza, M. & Egusquiza, E. (2021b). Strain Prediction in Francis Runners by Means of Stationary Sensors. *IOP Conference Series : Earth and Environmental Science*, 774, 012084. 30th IAHR Symposium on Hydraulic Machinery and Systems (IAHR 2020), Lausanne, Switzerland.

- Qiao, W. & Qu, L. (2018). Prognostic Condition Monitoring for Wind Turbine Drivetrains via Generator Current Analysis. *Chinese Journal of Electrical Engineering*, 4(3), 80–89.
- Qiu, Y., Feng, Y., Sun, J., Zhang, W. & Infield, D. (2016). Applying Thermophysics for Wind Turbine Drivetrain Fault Diagnosis Using SCADA Data. *IET Renewable Power Generation*, 10(5), 661–668. doi : 10.1049/iet-rpg.2015.016.
- Rajiv, G., Verma, M. & Subbulakshmi, A. (2023). Gaussian process metamodels for floating offshore wind turbine platforms. *Ocean Engineering*, 267, 113206. doi : 10.1016/j.oceaneng.2022.113206.
- Rasmussen, C. E. & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA, USA : MIT Press.
- Ren, L., Sun, Y., Wang, H. & Zhang, L. (2018). Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network. *IEEE Access*, 6, 13041–13049. doi : 10.1109/ACCESS.2018.2804930.
- Rezamand, M., Kordestani, M., Carriveau, R., Ting, D.-K., Orchard, M. & Saif, M. (2020). Critical Wind Turbine Components Prognostics : A Comprehensive Review. *IEEE Transactions on Instrumentation and Measurement*, 69(12), 9306–9322. doi : 10.1109/TIM.2020.3030165.
- Rezamand, M., Kordestani, M., Orchard, M. E., Carriveau, R., Ting, D. S.-K. & Saif, M. (2021). Improved Remaining Useful Life Estimation of Wind Turbine Drivetrain Bearings under Varying Operating Conditions. *IEEE Transactions on Industrial Informatics*, 17, 1742–1752. doi : 10.1109/TII.2020.3014567.
- Rivas, A., Delipei, G. K. & Hou, J. (2022). Predictions of component Remaining Useful Lifetime Using Bayesian Neural Network. *Progress in Nuclear Energy*, 146, 104143. doi : 10.1016/j.pnucene.2022.104143.
- Romano, Y., Patterson, E. & Candès, E. J. (2019). Conformalized Quantile Regression. *Advances in Neural Information Processing Systems*, 32. arXiv :1905.03222.
- Saathoff, M., Rosemeier, M., Kleinselbeck, T. & Rathmann, B. (2021). Effect of individual blade pitch angle misalignment on the remaining useful life of wind turbines. *Wind Energy Science*, 6, 1079–1087. doi : 10.5194/wes-6-1079-2021.
- Salameh, J., Cauet, S., Etien, E., Sakout, A. & Rambault, L. (2018). Gearbox condition monitoring in wind turbines : A review. *Mechanical Systems and Signal Processing*, 111, 251–264.

- Salinas-Camus, M., Goebel, K. & Eleftheroglou, N. (2025). A Comprehensive Review and Evaluation Framework for Data-Driven Prognostics : Uncertainty, Robustness, Interpretability, and Feasibility. *Mechanical Systems and Signal Processing*, 237, 113015.
- Santolamazza, A., Dadi, D. & Introna, V. (2021). A Data-Mining Approach for Wind Turbine Fault Detection Based on SCADA Data Analysis Using Artificial Neural Networks. *Energies*, 14, 1845. doi : 10.3390/en14071845.
- Sanz-Alcaine, J. M., Sebastián, E., Sanz-Gorrrachategui, I., Bernal-Ruiz, C., Bono-Nuez, A., Pajovic, M. & Orlik, P. V. (2021). Online voltage prediction using gaussian process regression for fault-tolerant photovoltaic standalone applications. *Neural Computing and Applications*, 33, 16577–16590.
- Seitzer, M., Tavakoli, A., Antic, D. & Martius, G. (2022). On the Pitfalls of Heteroscedastic Uncertainty Estimation with Probabilistic Neural Networks.
- Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. doi : 10.1080/01621459.1968.10480934.
- Sethi, M. R., Subba, A. B., Faisal, M., Sahoo, S. & Raju, D. K. (2024). Fault Diagnosis of Wind Turbine Blades with Continuous Wavelet Transform Based Deep Learning Model Using Vibration Signal. *Engineering Applications of Artificial Intelligence*, 138, 109372. doi : 10.1016/j.engappai.2024.109372.
- Shah, S. S., Tan, D. & Kumar, S. C. (2024). RUL Forecasting for Wind Turbine Predictive Maintenance Based on Deep Learning. *Heliyon*, 10, e39268. doi : 10.1016/j.heliyon.2024.e39268.
- She, D. & Jia, M. (2019). Wear Indicator Construction of Rolling Bearings Based on Multi-Channel Deep Convolutional Neural Network with Exponentially Decaying Learning Rate. *Measurement*, 135, 368–375.
- Soualhi, A., Medjaher, K., Clerc, G. & Razik, H. (2020). Prediction of Bearing Failures by the Analysis of the Time Series. *Mechanical Systems and Signal Processing*, 139, 106607. doi : 10.1016/j.ymssp.2019.106607.
- Steller, J. (2020). Cavitation damage as a result of polyfractional erosion process. *Wear*, 456–457, 203369. doi : 10.1016/j.wear.2020.203369.

- Steller, J. & Safonov, V. (2024). New methodologies for assessing cavitation load and resistance of structural materials and protective coatings in nuclear power industry. Part II. Polyfractional approach. *Problems of Atomic Science and Technology*, 4(152), 126–131. doi : 10.46813/2024-152-126.
- Sun, P., Li, J., Wang, C. & Lei, X. (2016). A Generalized Model for Wind Turbine Anomaly Identification Based on SCADA Data. *Applied Energy*, 168, 550–567.
- Sun, T., Yu, G., Gao, M., Zhao, L., Bai, C. & Yang, W. (2021). Fault Diagnosis Methods Based on Machine Learning and Its Applications for Wind Turbines : A Review. *IEEE Access*, 9, 147481–147507. doi : 10.1109/ACCESS.2021.3124025.
- Tang, H.-H., Zhang, K., Wang, B., Zu, X.-j., Li, Y.-Y., Feng, W.-W., Jiang, X., Chen, P. & Li, Q.-A. (2024). Early Bearing Fault Diagnosis for Imbalanced Data in Offshore Wind Turbine Using Improved Deep Learning Based on Scaled Minimum Unscented Kalman Filter. *Ocean Engineering*, 300, 117392. doi : 10.1016/j.oceaneng.2024.117392.
- Tang, M., Zhao, Q., Wu, H., Wang, Z., Meng, C. & Wang, Y. (2021). Review and Perspectives of Machine Learning Methods for Wind Turbine Fault Diagnosis. *Frontiers in Energy Research*, 9, 751066. doi : 10.3389/fenrg.2021.751066.
- Tartakovsky, A., Ma, T., Barajas-Solano, D. & Tipireddy, R. (2023). Physics-informed Gaussian process regression for states estimation and forecasting in power grids. *International Journal of Forecasting*, 39(2), 967-980.
- Teng, W., Han, C., Hu, Y., Cheng, X., Song, L. & Liu, Y. (2020). A Robust Model-Based Approach for Bearing Remaining Useful Life Prognosis in Wind Turbines. *IEEE Access*, 8, 47133-47143. doi : 10.1109/ACCESS.2020.2978301.
- Tonks, O. & Wang, Q. (2017). The Detection of Wind Turbine Shaft Misalignment Using Temperature Monitoring. *CIRP Journal of Manufacturing Science and Technology*, 17, 71–79.
- Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E. & Ulissi, Z. W. (2020). Methods for Comparing Uncertainty Quantifications for Material Property Predictions. *Machine Learning : Science and Technology*, 1(2), 025006.
- Turnbull, A., Carroll, J. & McDonald, A. (2022). A Comparative Analysis on the Variability of Temperature Thresholds through Time for Wind Turbine Generators Using Normal Behaviour Modelling. *Energies*, 15(15), 5298. doi : 10.3390/en15145298.

- Udo, W. & Muhammad, Y. (2021). Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data. *IEEE Access*, 9, 162370–162388. doi : 10.1109/ACCESS.2021.3132684.
- van Binsbergen, D., Soares, M. N., Pedersen, E. & Nejad, A. R. (2022). A physics-, SCADA-based remaining useful life calculation approach for wind turbine drivetrains. *Journal of Physics : Conference Series*, 2265(3), 032079. doi : 10.1088/1742-6596/2265/3/032079.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, UK : Cambridge University Press.
- Vieira, J. L. d. M., Farias, F. C., Ochoa, A. A. V., de Menezes, F. D., da Costa, J. A. P., de Novaes Pires Leite, G., Vilela, O. d. C., de Souza, M. G. G. & Michima, P. S. A. (2024). Remaining Useful Life Estimation Framework for the Main Bearing of Wind Turbines Operating in Real Time. *Energies*, 17, 1430.
- Vovk, V., Gammerman, A. & Shafer, G. (2005). *Algorithmic Learning in a Random World*. New York, USA : Springer.
- Wang, B., Lei, Y., Li, N. & Li, N. (2020a). A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Transactions on Reliability*, 69, 401–412. doi : 10.1109/TR.2018.2882682.
- Wang, B., Zhou, B., Zhu, D., Zou, M., Rao, Z. & Luo, H. (2025a). Feature Alignment LSTM for Detecting Anomalies in Wind Turbine Main Bearing Temperature Based on SCADA Data. *Journal of Engineering Research*. doi : 10.1016/j.jer.2025.05.011. In Press.
- Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T. & Shie, M. (2020b). A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*, 212(112766).
- Wang, H., Zhang, Z., Zhang, W., Li, M. & Zhang, Y. (2022a). Support Vector Regression Inverse System Control for Small Wind Turbine MPPT with Parameters' Robustness Improvement Support. *Journal of Control Science and Engineering*, 2022(8), 1-14.
- Wang, J., Wen, G., Yang, S. & Liu, Y. (2018). Remaining Useful Life Estimation in Prognostics Using Deep Bidirectional LSTM Neural Network. *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 1036–1041. doi : 10.1109/PHM-Chongqing.2018.00184.

- Wang, L., Cao, H., Xu, H. & Liu, H. (2022b). A gated graph convolutional network with multi-sensor signals for remaining useful life prediction. *Knowledge-Based Systems*, 252, 109340. doi : <https://doi.org/10.1016/j.knosys.2022.109340>.
- Wang, S.-H., Kang, X., Wang, C., Ma, T.-B., He, X. & Yang, K. (2024). A Hybrid Approach for Predicting the Remaining Useful Life of Bearings Based on the RReliefF Algorithm and Extreme Learning Machine. *Computer Modeling in Engineering & Sciences*, 140(2), 1405–1429. doi : [10.32604/cmes.2024.049281](https://doi.org/10.32604/cmes.2024.049281).
- Wang, S., Vidal, Y. & Pozo, F. (2026). Recent advances in wind turbine condition monitoring using SCADA data : A state-of-the-art review. *Reliability Engineering and System Safety*, 267, 111838. doi : [10.1016/j.ress.2025.111838](https://doi.org/10.1016/j.ress.2025.111838).
- Wang, Y., Li, F., Lv, M., Wang, T. & Wang, X. (2025b). A Multi-Index Fusion Adaptive Cavitation Feature Extraction for Hydraulic Turbine Cavitation Detection. *Entropy*, 27(4), 443. doi : [10.3390/e27040443](https://doi.org/10.3390/e27040443).
- Wang, Z., Cheng, H. & Ji, B. (2022c). Numerical prediction of cavitation erosion risk in an axisymmetric nozzle using a multi-scale approach. *Physics of Fluids*, 34(6), 062112. doi : [10.1063/5.0095833](https://doi.org/10.1063/5.0095833).
- Weissenfeld, A., Vanerio, J., Wachsenegger, A., Graser, A. et al. (2025). Proactive fault detection in wind turbine generators using SCADA measurements. *IET Conference Proceedings*, 10, 128–133. doi : [10.1049/icp.2025.2344](https://doi.org/10.1049/icp.2025.2344).
- Wilkinson, M. R., Spinato, F. & Tavner, P. J. (2007). Condition Monitoring of Generators and Other Subassemblies in Wind Turbine Drive Trains. *Proc. IEEE Int. Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)*, pp. 388–393.
- Wolff, B., Kühnert, J., Lorenz, E., Kramer, O. & Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135, 197–208.
- Xu, X., Huang, X., Bian, H., Wu, J., Liang, C. & Cong, F. (2024). Total Process of Fault Diagnosis for Wind Turbine Gearbox, From the Perspective of Combination with Feature Extraction and Machine Learning : A Review. *Energy and AI*, 15, 100318. doi : [10.1016/j.egyai.2023.100318](https://doi.org/10.1016/j.egyai.2023.100318).
- Xu, Y., He, D., Sun, H., Jin, Z. & Zhao, M. (2026). Self-supervised learning for train bearing fault diagnosis based on time–frequency dual domain prediction. *Structural Health Monitoring*.

- Yamin, M. & Giyats, A. F. (2022). Support Vector Regression Approach for Wind Forecasting. *International Journal of Advanced Science in Computing Engineering*, 4, 95–101.
- Yan, M., Wang, X., Wang, B., Chang, M. & Muhammad, I. (2020). Bearing Remaining Useful Life Prediction Using Support Vector Machine and Hybrid Degradation Tracking Model. *ISA Transactions*, 98, 471–482. doi : 10.1016/j.isatra.2019.08.058.
- Yan, M., Hui, S. C., Jiang, N. & Li, N. (2025). A review on data-driven prognostics and health management for wind turbine systems. *Engineering Applications of Artificial Intelligence*, 159, 111484. doi : <https://doi.org/10.1016/j.engappai.2025.111484>.
- Yang, W. (2015). *Dynamic Processes and Active Power Control of Hydropower Plants*. (Thèse de doctorat, Uppsala University, Uppsala, Sweden). Repéré à <https://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A855346&dswid=-1021>.
- Yona, A., Senjyu, T., Saber, A., Funabashi, T., Sekine, H. & Kim, C.-H. (2008). Application of neural network to 24-hour-ahead generating power forecasting for PV system. *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA*, 1-6.
- Yoo, Y. & Baek, J.-G. (2018). A Novel Image Feature for the Remaining Useful Lifetime Prediction of Bearings Based on Continuous Wavelet Transform and Convolutional Neural Network. *Applied Sciences*, 8, 1102. doi : 10.3390/app8071102.
- Yucesan, Y. A. & Viana, F. (2020). A Hybrid Model for Wind Turbine Main Bearing Fatigue with Uncertainty in Grease Observations. *Proc. Annual Conference of the PHM Society*, 12.
- Zhang, C., Hu, D. & Yang, T. (2022a). Anomaly Detection and Diagnosis for Wind Turbines Using Long Short-Term Memory-Based Stacked Denoising Autoencoders and XGBoost. *Reliability Engineering & System Safety*, 222, 108445. doi : 10.1016/j.res.2022.108445.
- Zhang, F., Li, P. C., Gao, L., Liu, Y. Q. & Ren, X. Y. (2021). Application of Autoregressive Dynamic Adaptive (ARDA) Model in Real-Time Wind Power Forecasting. *Renewable Energy*, 169, 129–143.
- Zhang, F., Chen, M., Zhu, Y., Zhang, K. & Li, Q. (2023a). A Review of Fault Diagnosis, Status Prediction, and Evaluation Technology for Wind Turbines. *Energies*, 16, 1125. doi : 10.3390/en16031125.
- Zhang, J. & Lee, J. (2011). A Review on Prognostics and Health Monitoring of Li-Ion Battery. *Journal of Power Sources*. doi : 10.1016/j.jpowsour.2011.03.101.

- Zhang, L., Zhang, H. & Cai, G. (2022b). The Multiclass Fault Diagnosis of Wind Turbine Bearing Based on Multisource Signal Fusion and Deep Learning Generative Model. *IEEE Transactions on Instrumentation and Measurement*, 71, 3514212. doi : 10.1109/TIM.2022.3178483.
- Zhang, L., Fan, Q., Lin, J., Zhang, Z., Yan, X. & Li, C. (2023b). A Nearly End-to-End Deep Learning Approach to Fault Diagnosis of Wind Turbine Gearboxes Under Nonstationary Conditions. *Engineering Applications of Artificial Intelligence*, 119, 105735. doi : 10.1016/j.engappai.2022.105735.
- Zhang, Y., Sun, J., Zhang, J., Shen, H., She, Y. & Chang, Y. (2023c). Health State Assessment of Bearing with Feature Enhancement and Prediction Error Compensation Strategy. *Mechanical Systems and Signal Processing*, 182, 109573.
- Zio, E. (2022). Prognostics and Health Management (PHM) : Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety*, 218, 108119. doi : <https://doi.org/10.1016/j.res.2021.108119>.
- Zou, J., Cicirello, A., Iliopoulos, A. & Lourens, E. M. (2023). Gaussian Process Latent Force Models for Virtual Sensing in a Monopile-Based Offshore Wind Turbine. Dans Rizzo, P. & Milazzo, A. (Éds.), *European Workshop on Structural Health Monitoring, EWSHM 2022* (vol. 253). Cham, Switzerland : Springer.