System-level Probabilistic Remaining Useful Life Prognostics and Predictive Inspection Planning for Wind Turbines

System-level Probabilistic Remaining Useful Life Prognostics and Predictive Inspection Planning for Wind Turbines

Davide Manna[1,2], Mihaela Mitici[2], Matteo Davide Lorenzo Dalla Vedova [3]

[1] *Faculty of Science, Utrecht University, Utrecht, 3584 CC, the Netherlands*
m.a.mitici@uu.nl

[2] *Department of Mechanical and Aerospace Engineering, Politecnico di Torino, 10129, Turin, Italy*
davide.manna@studenti.polito.it; matteo.dallavedova@polito.it

ABSTRACT

Wind energy plays a crucial role in the energy transition. However, it is often seen as an unreliable source of energy, with many production peaks and lows. Some of the drivers of uncertainty in energy production are the unexpected wind turbine (WT) failures and associated unscheduled maintenance. To support an effective health management and maintenance planning of WTs, we propose an integrated data-driven framework for Remaining Useful Life (RUL) prognostics and inspection planning of WTs. We propose a Long-short term memory (LSTM) neural network with Monte Carlo dropout to estimate the distribution of the RUL of WTs, i.e. we develop probabilistic prognostics. Different from existing studies focused on prognostics for single components, we consider the simultaneous health-monitoring of multiple components of the WTs, thus seeing the turbine as an integrated system. The obtained prognostics are further included into a stochastic planning model which determines optimal moments for inspections. For this, we pose the problem of WT inspections as a renewal reward process. We illustrate our framework for four offshore WTs which are continuously monitored by Supervisory Control and Data Acquisition (SCADA) systems. The results show that LSTMs are able to estimate well the RUL of the WTs, even in the early phase of their usage. We also show that the prognostics are informative for maintenance planning and are conducive to conservative inspections.

1. INTRODUCTION

The current global environmental crisis has prompted the active shift towards renewable energy solutions. For this, as outlined in the European Green Deal, the primary objective set forth by the Global Wind Energy Council is to actively contribute to meeting, by 2030, no less than 20% of the worldwide demand for electricity through the utilization of wind energy. Furthermore, the overarching ambition extends to realizing a fully decarbonized electricity supply by 2050, positioning wind energy at the forefront of renewable sources (Apunda & Nyangoye, 2017).

The focus on wind energy is motivated by the fact that wind is a clean, sustainable and inexhaustible source of energy, it has low operational costs, and that WTs can be installed in various locations, including remote areas where higher wind speeds can result in a higher energy production. Wind energy is, however, perceived as an unreliable source of energy, with many production peaks and lows. Some of the main drivers of uncertainty in energy production are the amount of unexpected failures and associated unscheduled maintenance (Letcher, 2023).

Horizontal Axis Wind Turbines (HAWTs), currently the most promising global wind energy technology (Rezamand et al., 2020), often face accelerated degradation due to their placement in regions with harsh and variable meteorological conditions (Astolfi, Pandit, Terzi, & Lombardi, 2022). Exposed to variable aerodynamic loads and mechanical stress (Tchakoua et al., 2014), WTs necessitate continuous health monitoring and dynamic maintenance planning to achieve reliable operations (Yang, Tavner, Crabtree, Feng, & Qiu, 2014; Tautz-Weinert & Watson, 2017).

In general, a HAWT integrates several essential subsystems, including aerodynamic rotor blades, a central hub for energy transfer, a gear reducer (Tong, 2010) (typically spur, helical (Errichello & Muller, 1994), or planetary (Ragheb & Ragheb, 2010)), an electrical generator for power conversion (Wagner, 2020), a nacelle housing all critical machinery, a yaw system enabling optimal wind alignment and a towering structure (Griffith et al., 2016). The subsystems with the highest fault rates for onshore wind farms are towers, gearboxes, and rotor blades, while for offshore wind farms, the most affected

components are gearboxes, rotor blades, generators, and towers (Rezamand et al., 2020). Generally, the most critical components of WTs are the gearbox, the main bearing, and the blades (Yang, Court, & Jiang, 2013).

To boost the reliability of WTs, recent studies have developed diagnostics and prognostics for components of WTs, focusing particularly on critical components such as gearboxes, main bearings, and blades. Given the increasing availability of condition monitoring measurements, a large fraction of these studies develops data-driven approaches for diagnostics and prognostics using machine learning. For data-driven diagnostics of WTs, frequent approaches are clustering algorithms, Principle Component Analysis (PCA), and Neural Networks. For example, in (Kim et al., 2011), a data-driven, unsupervised clustering algorithm, together with PCA is developed for diagnostics of gearboxes of WTs. Anomalies due to gearbox failures are identified based on measurements related to rotor speed and power production. In (Zaher, McArthur, Infield, & Patel, 2009), a multilayer neural networks is proposed to detect anomalies of the WT gearbox. The main input of the neural network is the temperature of the gearbox. In (Garan, Tidriri, & Kovalenko, 2022), the authors estimate whether the WT will fail or not within the next 60 days using a decision tree. Here, the focus is on optimizing the data preprocessing and feature selected steps of the methodology. A regression mode is proposed in (Orozco, Sheng, & Phillips, 2018) to detect anomalies of gearboxes.

For data-driven Remaining Useful Life prognostics using machine learning, which is also the case of our analysis, existing studies have focused on supervised neural networks. Frequently, vibration and/or Supervisory Control and Data Acquisition measurements are considered as input for these neural networks. Table 1 gives an overview of the main data-driven machine learning approaches, as well as the performance achieved. We note that all these studies focus on specific WT components when developing prognostics. The main components considered for prognostics development are the gearbox and the bearings. Neural networks are a frequently employed approach, which achieves accurate prognostics at various prognostics horizons (e.g., months/days before the actual failure). A recent study (Rajaoarisoa, Randrianandraina, & Sayed-Mouchaweh, 2024) develops a recurrent neural network to estimate the RUL of WTs, following the identification of faults using autoencoders. Completementary to this work, in this paper we propose a Long-short term memory (LSTM) neural network that directly estimates the RUL of the WTs. Here, the health monitoring and generation of RUL prognostics is performed at system level, i.e., the wind turbine is seen as an integrated system. Moreover, existing studies do not consider the development of maintenance planning models for WTs based on prognostics, e.g., predictive inspection planning for wind turbines. To the best of our knowledge, we propose for the first time a maintenance planning model for WTs based on RUL prognostics that are developed using actual measurements and machine learning models.

In this paper, we propose a LSTM neural network for RUL prognostics of WT. As datasets, we consider the recordings of the SCADA systems of the EDP Wind Farm open-source dataset (EDP, 2023). Different from existing studies, our approach involves the simultaneous health monitoring of multiple WT components such as the transformer, the gearbox, the generator, the hydraulic system. Consequently, we define the end-of-life of the WT as the occurrence of the first failure among its components. We use a LSTM neural network to estimate RUL prognostics for the WT seen as an integrated system, i.e., we determine system-level prognostics. By applying Monte Carlo dropout in the testing phase of the LSTM, we quantify the uncertainty associated with these prognostics, i.e., we determine probabilistic RUL prognostics. These prognostics are updated over time, as more measurements become available. The results show that the LSTM network is effective in accurately predicting the RUL of the WTs, even in the early stages of usage. Last, taking into account the obtained probabilistic RUL prognostics, we pose the problem of WT inspections as a renewal reward process and develop a planning model for inspections. The results show that the RUL prognostics support a conservative planning of inspections. This inspection planning is adjusted over time, as prognostics are themselves updated with newly acquired measurements.

The remainder of the paper is as follows. Section 2 introduces the open-source dataset considered for prognostics development. Subsequently, in Section 3.1, the importance of these features is quantified based on their SHAP values, and the most important features are selected for prognostics development. Section 3 proposes a LSTM neural network with Monte Carlo dropout for system-level probabilistic RUL prognostics for WT. Section 4 proposes a stochastic planning model for WT inspections, which integrates the probabilistic RUL prognostics obtained. Numerical results for WT system-level RUL prognostics and WT inspection planning are presented in Section 5. Last, conclusions are provided in Section 6.

2. DATA DESCRIPTION

We consider the Energias de Portugal (EDP) open-source dataset consisting of time-series of sensor measurements recorded for five offshore WT located in the West African Gulf of Guinea in the period 1st January 2017 - 31st December 2017. The information available in the EDP dataset consists of SCADA measurements, meteorological recordings, and the logs of the WT component failures, see also the complete list of measurements (EDP, 2023). The capacity of the each WT is 10MW. The measurements are recorded every 10min. For WT09, the logs recorded concern the Gearbox noise and Pitch position error, which does not indicate a proper fault/damage.

Table 1. Overview of data-driven prognostics for WT components, where ANFIS = Adaptive Neuro-Fuzzy Inference System, (K)ELM = (Kernel) Extreme Learning Machine, NN = Neural Network, SVM= Support Vector Machine; ACC = Accuracy, MA(P)E = Mean Absolute (Percentage) Error, NE = Normalized Error, PRC= Precision, (R)MSE = (Root) Mean Squared Error, SSE= Sum Squared Error.

Reference	Component	Method	Achieved Performance
(Li, Xu, Lei, Cai, & Kong, 2022)	Gearbox	NN	RMSE = 0.0025
(Merainani, Laddada, Bechhoefer, Chikh, & Benazzouz, 2022)	Bearing	NN	RMSE = 0.0025
(Kramti et al., 2021)	Bearing	NN	graphs available
(Elasha, Shanbr, Li, & Mba, 2019)	Gearbox	NN	SSE=661.98
(Pan, Hong, Chen, Singh, & Jia, 2019)	Gearbox	ELM	RMSE=0.91, MAE=0.734, ACC=95.4%
(Carroll et al., 2019)	Gear bearing	NN; SVM	ACC=72%; ACC=60%
(Cao, Qian, & Pei, 2018)	Bearing	SVM	RMSE=16.4, MAPE=42.9%
(Herp, Ramezani, Bach-Andersen, Pedersen, & Nadimi, 2018)	Bearing	NN, GP	0.5 $\leq$PRC$\leq$1
(Kramti, Ali, Saidi, Sayadi, & Bechhoefer, 2018)	Bearing	NN	MSE=0.0023
(Teng, Zhang, Liu, Kusiak, & Ma, 2016)	Bearing	NN	NE= 12.78%
(Chen, Matthews, & Tavner, 2013), (Chen, Matthews, & Tavner, 2015)	Pitch system	ANFIS	ACC$\geq$ 78%, prognostic horizon =21days, ACC$\geq$ 80%, prognostic horizon =14days ACC$\geq$ 86%, prognostic horizon =7days
(Zhao, Liu, Jin, Dang, & Deng, 2021)	Bearing	KELM	4.68% <NE<458.14%

As such, for our analysis, we consider the remaining four WTs (WT01, WT06, WT07, WT11).

Preliminary feature selection

Feature engineering from existing studies on prognostics and diagnostics for WTs (see also Table 1), indicate temperature-related features, production power, the generator and rotor speed rotation as parameters with a high explainability power for failures. In this line, we make a preliminary selection from the available parameters, leading to the following 31 features to be analysed for RUL prognostics: Average Temperature Hydraulic Oil (°C), Max/Min/ Average/STD Generator RPM (rpm), Average Temperature Bearing/ Bearing2 (°C), Average Temperature Generator Phase 1/2/3 (°C), Average Temperature Gearbox Oil (°C), Average Temperature Gearbox Bearing (°C), Average Temperature Nacelle (°C), Max/Min/Average Rotor RPM (RPM), Average Temperature High Volt Transformer Phase1/2/3 (°C), Average Temperature Grid Inverter Phase1 (°C), Average Temperature Controller Top (°C), Average Temperature Controller Hub (°C), Average Temperature Controller VCP (°C), Average Temperature Controller VCP Chokcoil (°C), Average Temperature VCP Cooling Water (°C), Average Temperature VCP Cooling Water (°C), Average Temperature Spinner (°C), Latest Production Total Active Power (Wh), Average Temperature Generator Slip Ring (°C), Average Temperature Grid Rotor Inverter Phase1/2/3 (°C).

3. System-level RUL prognostics for wind turbines

We consider a WT consisting of multiple components. The health of each component is monitored continuously by multiple sensors. We say that the system-level RUL of the WT is the remaining time until the first failure of any one of these components. We are interested in estimating the system-level RUL of the WT based on the sensor measurements recorded. At time step d (dth day), we have available the following measurements for WT $i, i \in \{1, 2, \ldots, n\}$,

$$x_d^i = \{x_{1,d}^i, x_{2,d}^i, \ldots, x_{m,d}^i\}, \quad (1)$$

where m is the total number of considered features and $x_{j,d}^i$ is the measurement corresponding to feature $j, 1 \leq j \leq m$ recorded on day d for WT i.

Then, the actual system-level RUL of WT i at time d is:

$$RUL^a(WT_i) = \min\{\tau(c_1^i) - d, \tau(c_2^i) - d, \ldots, \tau(c_n^i) - d\}, \quad (2)$$

where $\tau(c_j^i), 1 \leq j \leq n$ is the time of failure of component c_j^i of WT i, and n the total number of components of WT i.

We are interested in estimating the system-level RUL of the four WTs in the EDP dataset at various moments (k) in time. Table 2 shows four Cases when each of the WT is the testing set, while the datasets of the remaining three WTs constitute the training and validation sets. The failure of three out of the four WTs is due to a failure of the Hydraulic group. The remaining WT fails due to a failure of the Transformer.

3.1. Feature importance using SHAP values

In Section 2, a total of 31 features has been considered. In this section we quantify the importance of these features for WT system-level RUL estimation using the Shapley additive explanations (SHAP) values (Lundberg & Lee, 2017). SHAP values quantify the impact of a feature on the RUL prognostic. SHAP values are determined as follows:

$$\phi_i = \sum_{S \subseteq Fi} \frac{|S|!(|F| - |S| - 1)!}{|F|!} |f(S \cup \{i\} - f(S))|, \quad (3)$$

Table 2. Overview of data used for testing, training and validation - EDP dataset for WT health monitoring.

	Case 1	**Case 2**	**Case 3**	**Case 4**
Testing	**WT06**	**WT07**	**WT11**	**WT01**
Training	WT01, WT07	WT01, WT06	WT06, WT07	WT06, WT07
Validation	WT11	WT11	WT01	WT11
First fault	Hydraulic Group	Hydraulic Group	Hydraulic Group	Transformer
Actual Lifetime	8 months	6 months	4 months	8 months

with F the set of all features considered for RUL prognostics, $S \subseteq F$ a subset of features obtained from the set F except feature i, and $f(S)$ the expected algorithm output given by the set S of considered features. The SHAP value quantifies the magnitude of the impact, i.e., how much a specific feature value contributes to the accurate estimation of the RUL. A large SHAP value for a given feature indicates a large importance of this feature for the RUL estimation.

For each of the four Cases, we select the 60% most important features of the the total of 31 features, i.e., we select 20 features with the highest SHAP value, see Figures 1-4. The results show that, although the WTs have various components that trigger the failure of the entire system, i.e., either the hydraulic group or the transformer, the average RPM of the generator is the feature with the highest importance for all four WTs. These confirms the findings of existing literature (see also Table 1), that the health condition of the generator is crucial for the overall operation of WTs. Most importantly, these results show that regardless of the failure mode, the WT can be seen as a system and the available measurements can support the development of system-level prognostics.

3.2. Long-short term memory (LSTM) for probabilistic RUL prognostics

Given the long-term dependencies in the measurements, as well as the high nonlinearity of the features, we propose a Long-short term memory (LSTM) with Monte Carlo dropout (Hochreiter & Schmidhuber, 1997) to estimate the distribution of the RUL (probabilistic RUL prognostics) of the WTs in each of the four Cases.

We consider a LSTM consisting of L layers, each consisting of N neurons, and LeakyReLu activation layers (Graves & Graves, 2012). The last layer of the LSTM is a Dense layer, for which a ReLu activation function is assumed. The input gate i_t, the output gate o_t, and the forget gate f_t of the LSTM are defined as follows. The forget gate f_t determines whether to consider or not the previous state c_{t-1}, i.e,

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

where x_t is the current input, h_{t-1} is the previous hidden state, W_f is a trainable weight, b_f is bias. The input gate determines whether to update the state of the LSTM using the current observation, using a sigmoid layer:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{5}$$

where x_t is the current input, h_{t-1} is the previous hidden state, W_i is a trainable weight, b_i is the bias. The output gate o_t determines whether the hidden state h_t is passed to the next iteration, i.e.,

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{6}$$

where x_t is the current input, h_{t-1} is the previous hidden state, W_o is a trainable weight, b_o is the bias. Table 3 shows the hyperparameters of the considered LSTM.

Table 3. Hyperparameters tuning - LSTM.

Number Layers	4
Neurons Layer 1	128
Neurons Layer 2	64
Neurons Layer 3	64
Neurons Layer 4	64
Dropout rate	0.5
Epochs	40
Batch size	32
Window length	3

Monte Carlo dropout for probabilistic RUL prognostics

Commonly, Monte Carlo dropout is applied in the training phase of the neural networks to avoid overfitting. To obtain the distribution of the RUL, i.e., to obtain probabilistic RUL prognostics, we also apply Monte Carlo dropout in the testing phase of the LSTM. In this line, (Gal & Ghahramani, 2016) shows that such a neural network with Monte Carlo dropout approximates a Bayesian neural network representing a deep Gaussian process.

Let X be the samples in the training set of the LSTM, and let Y be the corresponding RUL values. In a Bayesian neural network, we aim to estimate the posterior distribution $p(y|x, X, Y)$:

$$p(y|x, X, Y) = \int p(y|x, \omega)p(\omega|X, Y)d\omega \tag{7}$$

with ω the weights of the neural network, $p(y|x, \omega)$ the probability that the RUL is y, given the test sample x and the weights ω, and $p(\omega|X, Y)$ the posterior distribution of the weights, given the training samples X and Y.

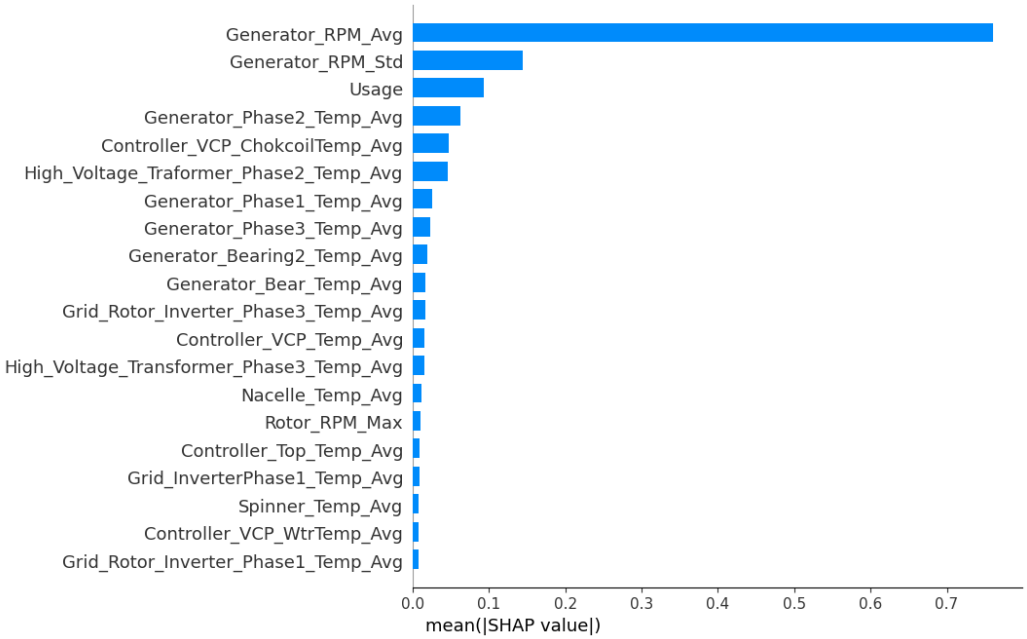

Figure 1. Case 1: WT06 - SHAP values of features.

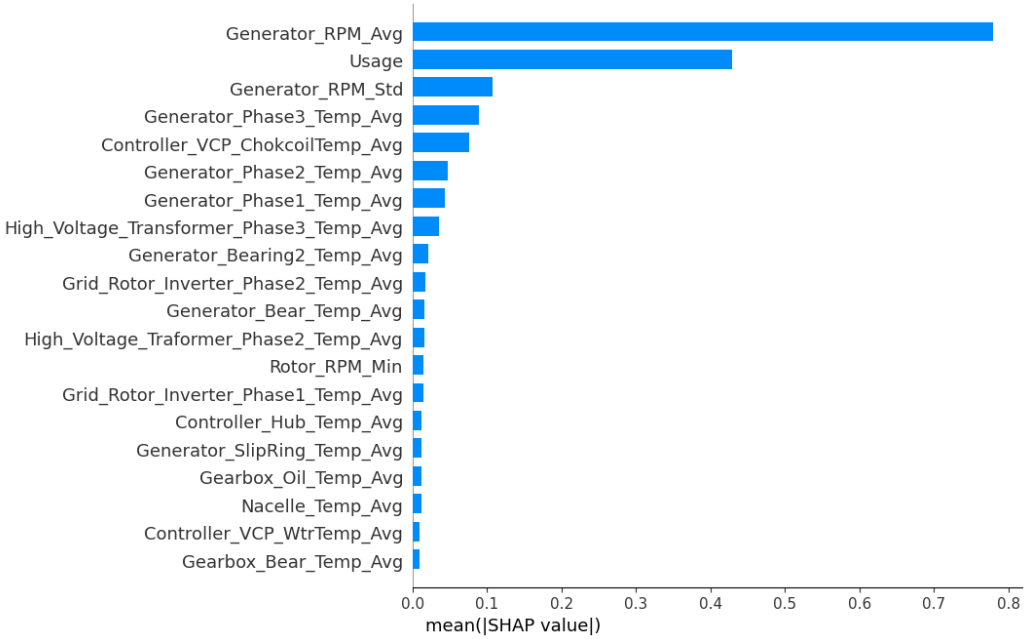

Figure 2. Case 2: WT07 - SHAP values of features.

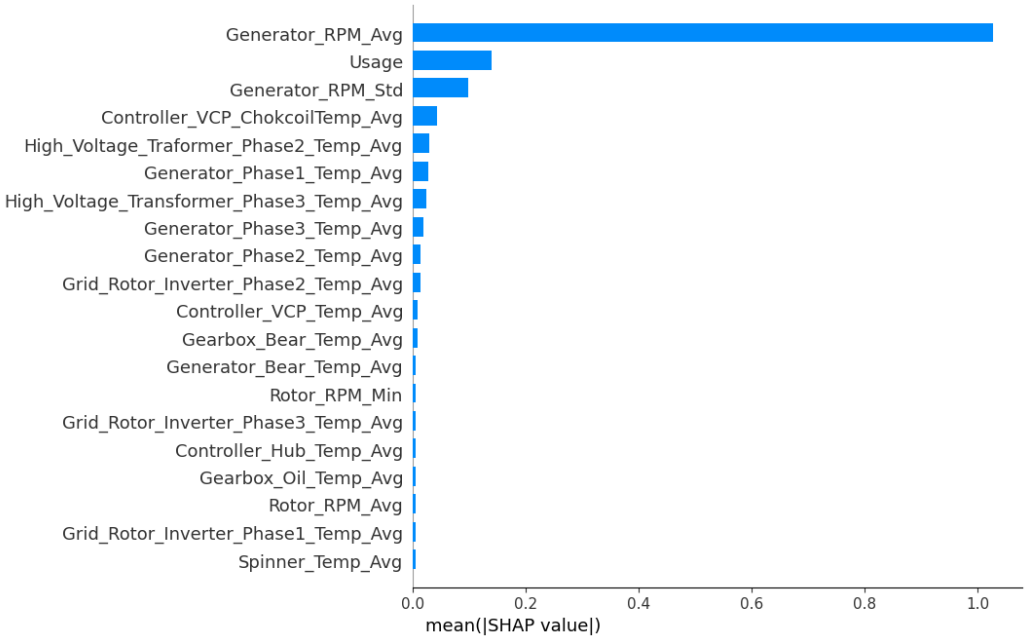

Figure 3. Case 3: WT11 - SHAP values of features.

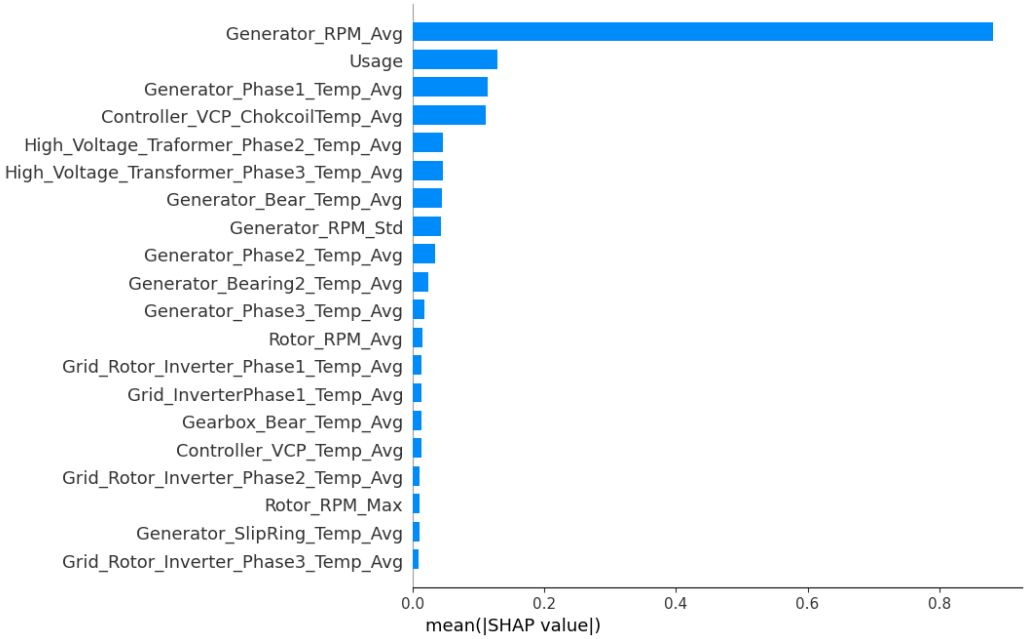

Figure 4. Case 4: WT01 - SHAP values of features.

It is computationally expensive to analyze the posterior distribution $p(\omega|X,Y)$ exactly (Gal & Ghahramani, 2016). As such, we approximate $p(\omega|X,Y)$ with a distribution $q(\omega)^*$ that minimizes Kullback–Leibler divergence KL with the true posterior distribution $p(\omega|X,Y)$, i.e. (Blei, Kucukelbir, & McAuliffe, 2017):

$$q^*(\omega) = argmin_{q(\omega)}\{KL(q(\omega|p(\omega|X,Y)))\}. \quad (8)$$

Using $q(\omega)^*$, we approximate the posterior distribution of the RUL of a test sample by:

$$q(y|x) = \int p(y|x,\omega)q^*(\omega)d\omega \quad (9)$$

where $q(y|x)$ is the approximation of $p(y|x,X,Y)$.

Lastly, we approximate the expected value $\hat{y}$ of the RUL of a test sample by:

$$\hat{y} = E_{q(y|x)}(y) = \frac{1}{M}\sum_{1}^{M}\hat{y_j}(x,\omega^j) \quad (10)$$

where M is the number of forward passes through the neural network, ω^j are the weights of the neural network belonging to the j-th forward pass (i.e., where some neurons are dropped out), and $\hat{y_j}(x,\omega^j)$ is the resulting RUL prediction from the j-th forward pass through the neural network. For the distribution of the RUL, we give each individual RUL prediction $\hat{y_j}(x,\omega^j)$ a probability $\frac{1}{M}$.

Performance metrics for RUL prognostics

To evaluate the ability of the LSTM model to predict the RUL, we consider the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Continuous Ranked Probability Score (CRPS), which are defined as follows.

$$MAE = \sum_{i=1}^{n}\frac{|RUL_i^a - R\bar{U}L_i^p|}{n}, \quad (11)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(RUL_i^a - R\bar{U}L_i^p)^2}{n}}, \quad (12)$$

with n the number of days over which predictions are made, and $R\bar{U}L_i^p$ the mean predicted RUL at day $i, 1 \leq i \leq n$.

Since we estimate the distribution of the RUL, to be able to quantify the fitness of these distributions relative to the actual RUL (a point value), we consider the Continuous Ranked Probability Score (CRPS) and the Weighted CRPS ($CRPS^W$). Here, CRPS evaluates whether the estimated RUL distribution is centered around the actual RUL of the WT and whether the variance of this distribution is low (a high sharpness of the RUL prognostic) (Mitici, de Pater, Barros, & Zeng, 2023). The Weighted CRPS applies a (larger) penalty β when overestimating the RUL then when underestimating the RUL. This is of particular importance when planning the inspections of the WTs - planning too late inspections (after the actual failure of the wind turbine) does not make effective use of the prognostics to timely identify and act upon anticipated failures of the WTs.

CRPS is defined as follows (Gneiting & Katzfuss, 2014),

$$CRPS = \frac{1}{n}\sum_{i=1}^{n}CRPS_i, \quad (13)$$

$$CRPS_i = \int_{-\infty}^{\infty}(F_{\hat{y_i}}(x) - I\{y_i \leq x\})^2 dx \quad (14)$$

with $I\{y_i \leq x\} = \begin{cases} 1 & \text{if } y_i \leq x \\ 0 & \text{if } y_i > x. \end{cases}$

The weighted CRPS ($CRPS^W$) is defined as follows (Gneiting & Katzfuss, 2014):

$$CRPS^W = \frac{1}{N}\sum_{i=1}^{N}CRPS_i^W, \quad (15)$$

$$CRPS_i^W = (2-\beta)\int_{-\infty}^{y_i}(F_{\widehat{y_i}}(x))^2 dx + \beta\int_{y_i}^{\infty}(F_{\widehat{y_i}}(x) - 1)^2 dx, 0 \leq \beta \leq 2. \quad (16)$$

4. Inspection planning of wind turbines using probabilistic RUL prognostics

In this Section we pose the problem of WT inspections as a renewal reward process (Tijms, 2003), which integrates the probabilistic RUL prognostics developed in Section 3.2. We aim to determine optimal times for WT inspections.

We consider a renewal reward process $\{N_t\}$ where the process regenerates when a wind turbine is inspected, i.e., our knowledge about the actual health condition of the wind turbine is reset upon an inspection. At day k during the life of the WT, we are interested in determining an optimal time $k + t_k^*$ to inspect the WT. At day k, using the measurements recorded up to day k and a LSTM with Monte Carlo dropout (see Section 3.2), we estimate the probability that the RUL of the WT is i days, $i \geq 0$. Let $\phi_k(i)$ denote the probability that the WT, after being used for k days, has a RUL of exactly i days. To determine an optimal time to inspect the WT, we consider the expected cost per unit of time:

$$\frac{\text{[Expected cost over the current inspection cycle]}}{\text{[Expected current inspection cycle]}}. \quad (17)$$

At day k, we are interested in finding an optimal time for

inspection t_k^* such that:

$$t_k^* := argmin_{t_k>0} \frac{\mathbb{E}[C(k,t_k)]}{\mathbb{E}[L(k,t_k)]}, \quad (18)$$

with $C(k,t_k)$ the cost of inspecting the WT at day $k+t_k$, given that this WT has already been used for k days, and $L(k,t_k)$ is the length of the inspection cycle of the WT.

If the WT is scheduled for inspection at day $k+t_k$, then a cost c_r is incurred. If, however, the WT fails at some day $j, k < j < k+t_k$ before an inspection is planned, then a failure cost c_f is incurred (corrective maintenance).

With this, the expected cost over the current inspection cycle of the WT is:

$$E[C(k,t_k)] = c_f \sum_{i=0}^{t_k-1} \phi_k(i) + c_r(1 - \sum_{i=0}^{t_k-1} \phi_k(i)). \quad (19)$$

Also, the expected current inspection cycle is:

$$E[L(k,t_k)] = k + \sum_{i=0}^{t_k-1} i\phi_k(i) + t_k(1 - \sum_{i=0}^{t_k-1} \phi_k(i)). \quad (20)$$

Eq. (18) is solved using a numerical grid search. The estimate $\phi_k(i)$ after every day k is obtained using a LSTM and the methodology in Section 3.2.

5. Numerical results

In this Section we illustrate the results obtained for the probabilistic RUL prognostics and inspection planning for the four WTs for which measurements are available at (EDP, 2023).

5.1. Probabilistic RUL prognostics for wind turbines

Table 4 shows the performance of the system-level RUL prognostics for the WTs in the four Cases considered.

Table 4. Performance - RUL prognostics using LSTM.

	MAE	$RMSE$	$CRPS$	$CRPS^W$ $\beta = 1.9$
Case 1: WT06	12.72	15.52	9.98	2.51
Case 2: WT07	11.30	13.65	7.86	9.16
Case 3: WT11	9.40	11.80	6.93	6.88
Case 4: WT01	19.35	22.42	14.68	3.11

The results show that the lowest MAE and $RMSE$ are obtained for WT11, while the highest MAE and $RMSE$ are obtained for WT01. However, when considering the prognostics as input for inspection planning, we are interested in not missing the failures. This may, however, occur when we overestimate the RUL and, based on these overestimates, we plan late inspections. The Weighted CRPS captures the tendency of the prognostics to overestimate the RUL. We consider a large penalty for RUL overestimation ($\beta = 1.9$), given our ultimate goal of planning inspections for WTs based on prognostics. In this line, we are interested in planning inspection timely, to anticipate the actual failures of the turbines rather than missing these failures.

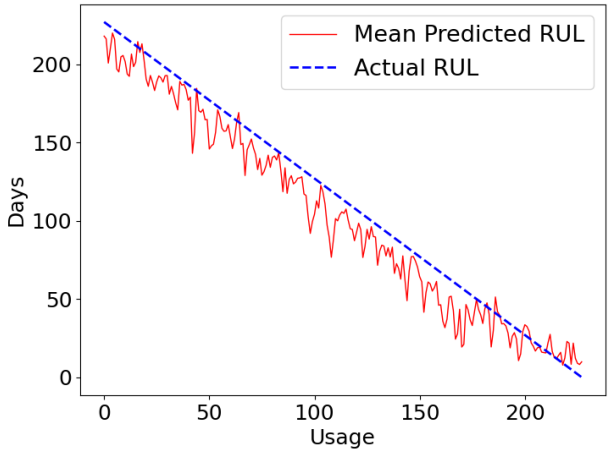

Figure 5. Case 1 - RUL estimation, WT06.

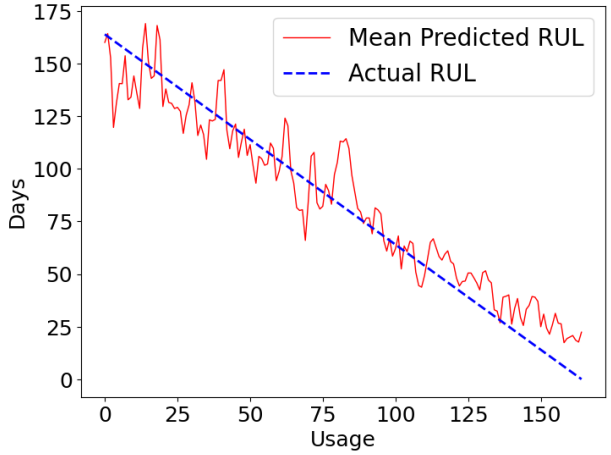

Figure 6. Case 2 - RUL estimation, WT07.

The results show that WT07 has the highest $CRPS^W = 9.16$, despite having a relatively low MAE and $RMSE$. This indicates that the RUL is predominantly overestimated and a conservative inspection planning should be considered, despite the low MAE and $RMSE$. The results also show that WT06 has the lowest $CRPS^W = 2.51$, despite having a relatively high MAE and $RMSE$ among all four turbines. This indicates that the prognostics have the least tendency to overestimate the RUL. These make the prognostics suitable for inspection planning, despite their high MAE and $RMSE$. Overall, the results show that considering MAE and $RMSE$ alone when aiming to use prognostics for maintenance planning is not sufficient. Additional metrics such as $CRPS^W$, through their ability to evaluate whether the

RUL is over/under-estimated, are particularly informative of the suitability of the prognostics for maintenance planning.

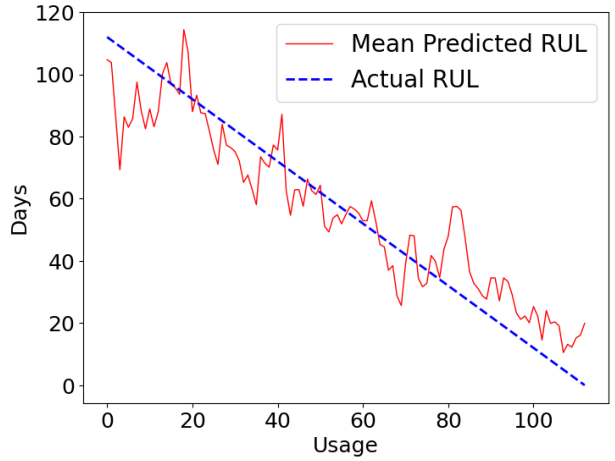

Figure 7. Case 3 - RUL estimation, WT11.

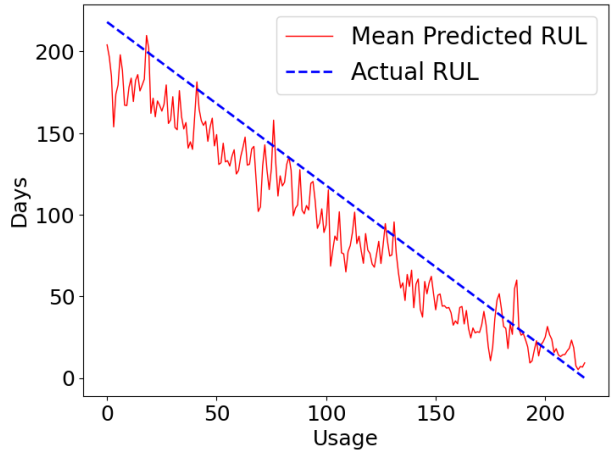

Figure 8. Case 4 - RUL estimation, WT01.

The RUL prognostics obtained over time are shown in Figures 5 - 8. The RUL of WT06 and WT01 are predominantly underestimated. The RUL of WT07 and WT11 are predominantly overestimated.

Figures 9 - 11 show the distribution of the RUL for WT06 (Case 1) at $\{202, 102, 2\}$ days before the actual failure of the WT. The results show that the sharpness of the estimated distribution increases closer to the time of failure of the WT.

5.2. Inspection planning for wind turbines using probabilistic RUL prognostics

For inspection planning, we consider $c_f = 100.000$ and $c_r = 100$. Every day k (or equivalently after k days of usage), based on the measurements collected up to this day, we develop RUL prognostics, i.e., the prognostics are updated ev-

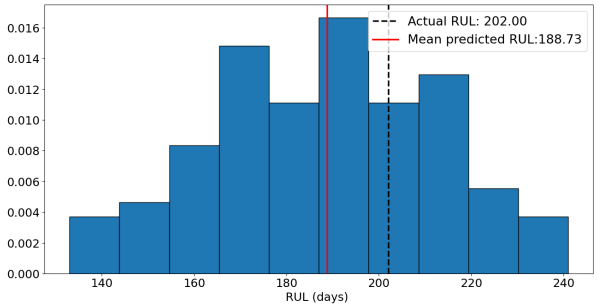

Figure 9. Estimated distribution of RUL, 202 days before the actual failure of WT06.

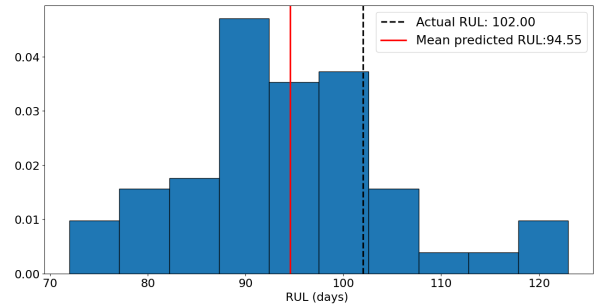

Figure 10. Estimated distribution of RUL, 102 days before the actual failure of WT06.

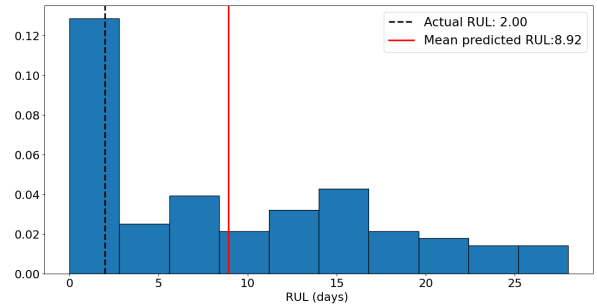

Figure 11. Estimated distribution of RUL, 2 days before the actual failure of WT06.

ery day. Based on these prognostics, every day k we determine an optimal time t_k^* to plan a WT inspection.

Figures 12 -15 show the results for the optimal inspection times of the four WTs relative to the actual RUL and the mean estimated RUL. For Case 1 - WT06, although the MAE and $RMSE$ are relatively high, the fact that $CRPS^W$ is low, i.e., the overestimation of the RUL is low, is reflected in the inspection planning - timely planning that does not miss the failure of the WT. In fact, in the last phase of the monitoring of this WT, it is consistently indicated that an optimal action

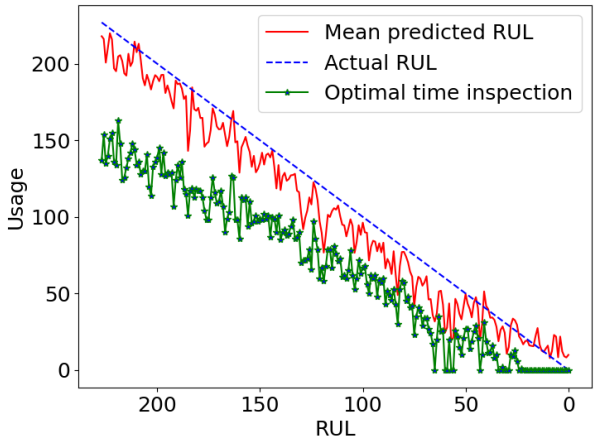

Figure 12. Case 1: WT06, Optimal time inspection.

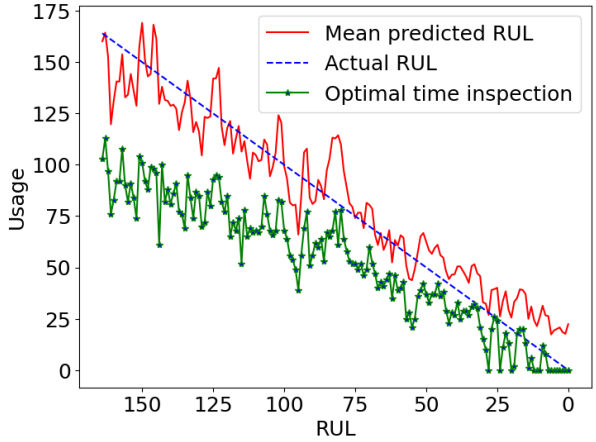

Figure 13. Case 2: WT07, Optimal time inspection.

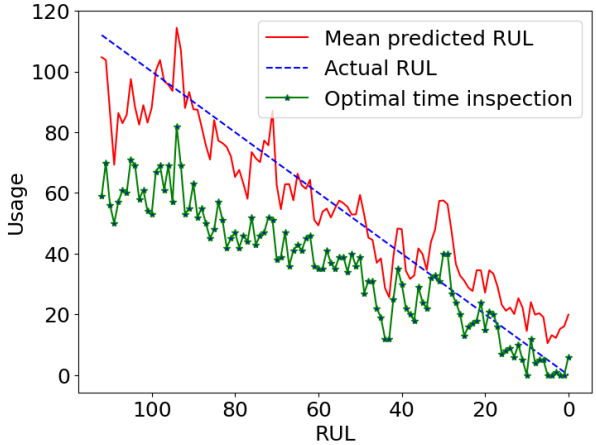

Figure 14. Case 3: WT11, Optimal time inspection.

is to plan an inspection immediately.

For Case 2 - WT07, the $CRPS^W$ is the highest, i.e., the prognostics have the tendency to overestimate the RUL, which is expected to delay the planning of the inspections leading to a potential miss of the failure. This is reflected in the inspection planning, particularly in the last phase of the monitoring of the WT, see also Figure 13. For Case 3 - WT11, the $CRPS^W$ is high, i.e. the prognostics have a tendency to overestimate the RUL. As a result, delayed inspections are planned in the last phase of the monitoring of the WT. For Case 4 - WT01, despite the lowest achieved MAE and $RMSE$, a moderate $CRPS^W$ is reflected in the inspection planning - timely inspection planning, particularly in the last phase of the WT monitoring, when an immediate inspection is consistently indicated as an optimal action (see also Figure 15). Overall, for all four cases, the planning of the inspections is conservative, where timely inspections are indicated as being optimal actions.

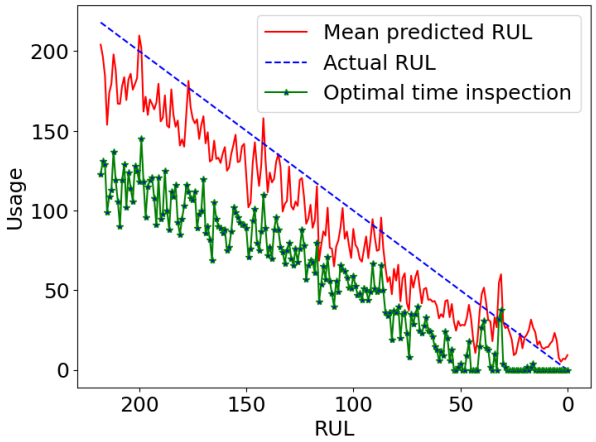

Figure 15. Case 4: WT01, Optimal time inspection.

Table 5. Optimal time for WT inspection.

	RUL^a	k	RUL^p	t_k^*
Case 1: WT06				
	200	27	193.82	139
	100	127	96.62	67
	50	177	49.96	27
	25	202	33.18	6
Case 2: WT07				
	150	14	135.44	101
	100	64	94.11	68
	50	114	38.61	37
	25	139	26.5	24
Case 3: WT11				
	100	12	86.72	53
	75	37	70.28	43
	50	62	61.64	39
	25	87	29.64	13
Case 4: WT01				
	200	18	192.69	118
	100	118	74.69	59
	50	168	28	4
	25	193	8.46	0

Table 5 shows in detail several moments throughout the monitoring of the WTs when inspections are planned (t_k^*), relative to the actual RUL (RUL^a), the usage of the WT (k), and the mean estimated RUL ($\bar{RUL}^p$).

6. Conclusions

This paper proposes a machine learning approach for system-level probabilistic RUL prognostics for WTs. In contrast with existing studies, which develop component-based prognostics, we see the WT as an integrated system and develop system-level RUL prognostics. These prognostics are further employed to determine optimal moments for inspections of the WTs, in anticipation of failures. To the best of our knowledge, this is the first study that proposes a maintenance planning model for WT based on data-driven prognostics. A LSTM with Monte Carlo dropout is developed to estimate the distribution of the RUL of the WTs, i.e., we develop probabilistic RUL prognostics. By using dropout in the test phase of the LSTM, the uncertainty associated with the RUL prognostics is quantified. To plan inspections for the WTs, a renewal reward process is proposed, which integrates these probabilistic RUL prognostics.

We illustrate our approach for four offshore wind turbines located in the West African Gulf of Guinea, and which have been monitored in the period 1st January - 31st December 2017. The results show that the proposed LSTM estimates well the RUL of the WTs, with a Mean Absolute Error ranging between 9.40 days to 19.35 days when considering all four wind turbines. Based on these RUL prognostics, inspections are planned conservatively, well ahead of the actual day of failure. The results show that, although imperfect, prognostics are informative for maintenance and support an efficient planning of inspection tasks.

As future work we aim to improve our RUL prognostics by considering additional features such as attention mechanisms integrated into the neural networks.

Acknowledgement

The research contribution of Mihaela Mitici is partially supported as part of France 2030 program ANR-11-IDEX-0003.

References

Apunda, M. O., & Nyangoye, B. O. (2017). Challenges and opportunities of wind energy technology. *International Journal of Development Research*, *9*(06), 14174–14177.

Astolfi, D., Pandit, R., Terzi, L., & Lombardi, A. (2022). Discussion of wind turbine performance based on scada data and multiple test case analysis. *Energies*, *15*(15), 5343.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

Cao, L., Qian, Z., & Pei, Y. (2018). Remaining useful life prediction of wind turbine generator bearing based on emd with an indicator. In *2018 prognostics and system health management conference (phm-chongqing)* (pp. 375–379).

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*(3), 360–375.

Chen, B., Matthews, P., & Tavner, P. J. (2013). Wind turbine pitch faults prognosis using a-priori knowledge-based anfis. *Expert Systems with Applications*, *40*(17), 6863–6876.

Chen, B., Matthews, P. C., & Tavner, P. J. (2015). Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition. *IET Renewable Power Generation*, *9*(5), 503–513.

EDP. (2023). Dataset wind turbines. *https://www.edp.com/en/innovation/open-data/data*.

Elasha, F., Shanbr, S., Li, X., & Mba, D. (2019). Prognosis of a wind turbine gearbox bearing using supervised machine learning. *Sensors*, *19*(14), 3092.

Errichello, R., & Muller, J. (1994). Design requirements for wind turbine gearboxes. *NASA STI/Recon Technical Report N*, *9*.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050–1059).

Garan, M., Tidriri, K., & Kovalenko, I. (2022). A data-centric machine learning methodology: Application on predictive maintenance of wind turbines. *Energies*, *15*(3), 826.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.

Graves, A., & Graves, A. (2012). Supervised sequence labelling.

Griffith, D. T., Paquette, J., Barone, M., Goupee, A. J., Fowler, M. J., Bull, D., & Owens, B. (2016). A study of rotor and platform design trade-offs for large-scale floating vertical axis wind turbines. In *Journal of physics: Conference series* (Vol. 753, p. 102003).

Herp, J., Ramezani, M. H., Bach-Andersen, M., Pedersen, N. L., & Nadimi, E. S. (2018). Bayesian state prediction of wind turbine bearing failure. *Renewable Energy*, *116*, 164–172.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Kim, K., Parthasarathy, G., Uluyol, O., Foslien, W., Sheng,

S., & Fleming, P. (2011). Use of scada data for failure detection in wind turbines. In *Energy sustainability* (Vol. 54686, pp. 2071–2079).

Kramti, S. E., Ali, J. B., Saidi, L., Sayadi, M., & Bechhoefer, E. (2018). Direct wind turbine drivetrain prognosis approach using elman neural network. In *2018 5th international conference on control, decision and information technologies (codit)* (pp. 859–864).

Kramti, S. E., Ali, J. B., Saidi, L., Sayadi, M., Bouchouicha, M., & Bechhoefer, E. (2021). A neural network approach for improved bearing prognostics of wind turbine generators. *The European Physical Journal Applied Physics*, *93*(2), 20901.

Letcher, T. (2023). *Wind energy engineering: a handbook for onshore and offshore wind turbines*. Elsevier.

Li, N., Xu, P., Lei, Y., Cai, X., & Kong, D. (2022). A self-data-driven method for remaining useful life prediction of wind turbines considering continuously varying speeds. *Mechanical Systems and Signal Processing*, *165*, 108315.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Merainani, B., Laddada, S., Bechhoefer, E., Chikh, M. A. A., & Benazzouz, D. (2022). An integrated methodology for estimating the remaining useful life of high-speed wind turbine shaft bearings with limited samples. *Renewable Energy*, *182*, 1141–1151.

Mitici, M., de Pater, I., Barros, A., & Zeng, Z. (2023). Dynamic predictive maintenance for multiple components using data-driven probabilistic rul prognostics: The case of turbofan engines. *Reliability Engineering & System Safety*, *234*, 109199.

Orozco, R., Sheng, S., & Phillips, C. (2018). Diagnostic models for wind turbine gearbox components using scada time series data. In *2018 ieee international conference on prognostics and health management (icphm)* (pp. 1–9).

Pan, Y., Hong, R., Chen, J., Singh, J., & Jia, X. (2019). Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion. *Mechanism and machine theory*, *137*, 509–526.

Ragheb, A., & Ragheb, M. (2010). Wind turbine gearbox technologies. In *1st international nuclear & renewable energy conference (inrec)* (pp. 1–8).

Rajaoarisoa, L., Randrianandraina, R., & Sayed-Mouchaweh, M. (2024). Predictive maintenance model-based on multi-stage neural network systems for wind turbines. In *2024 international conference on artificial intelligence, computer, data sciences and applications (acdsa)* (pp. 1–7).

Rezamand, M., Kordestani, M., Carriveau, R., Ting, D. S.-K., Orchard, M. E., & Saif, M. (2020). Critical wind turbine components prognostics: A comprehensive review. *IEEE Transactions on Instrumentation and Measurement*, *69*(12), 9306–9328.

Tautz-Weinert, J., & Watson, S. J. (2017). Using scada data for wind turbine condition monitoring–a review. *IET Renewable Power Generation*, *11*(4), 382–394.

Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T. A., & Ekemb, G. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, *7*(4), 2595–2630.

Teng, W., Zhang, X., Liu, Y., Kusiak, A., & Ma, Z. (2016). Prognosis of the remaining useful life of bearings in a wind turbine gearbox. *Energies*, *10*(1), 32.

Tijms, H. C. (2003). *A first course in stochastic models*. John Wiley and sons.

Tong, W. (2010). *Fundamentals of wind energy* (Vol. 44). WIT press Southampton, UK.

Wagner, H.-J. (2020). Introduction to wind energy systems. In *Epj web of conferences* (Vol. 246, p. 00004).

Yang, W., Court, R., & Jiang, J. (2013). Wind turbine condition monitoring by the approach of scada data analysis. *Renewable energy*, *53*, 365–376.

Yang, W., Tavner, P. J., Crabtree, C. J., Feng, Y., & Qiu, Y. (2014). Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, *17*(5), 673–693.

Zaher, A., McArthur, S., Infield, D., & Patel, Y. (2009). Online wind turbine fault detection through automated scada data analysis. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, *12*(6), 574–593.

Zhao, H., Liu, H., Jin, Y., Dang, X., & Deng, W. (2021). Feature extraction for data-driven remaining useful life prediction of rolling bearings. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–10.

BIOGRAPHIES

Davide Manna holds a B.Sc. in Aerospace Engineering from University of Naples Federico II, and an M.Sc. in Aerospace Engineering from Politecnico di Torino, Italy. He has also been a visiting scholar at Utrecht University, where he conducted research on predictive maintenance of wind turbines.

Mihaela Mitici is an Assistant Professor at Faculty of Science, Utrecht University. She has a Ph.D. in Stochastic Operations Research from University of Twente, the Netherlands, and an M.Sc. in Operations Research from University of Amsterdam. During 2016-2022 she was Assistant professor at TU Delft. Mihaela specializes in Operations Research, with a focus on stochastic processes, decision-making under uncertainty, applied probability theory, machine learning. Her application domains are predictive maintenance and mobility.

Matteo D. L. Dalla Vedova is an Assistant Professor in the Department of Mechanics and Aerospace Engineering, Po-

litecnico di Torino. He holds an M.Sc. (2003) and Ph.D. (2007) in aerospace engineering from Politecnico di Torino. His research focuses on aeronautical systems engineering, particularly, the design, analysis, and numerical simulation of onboard systems, the study of secondary flight control systems, and the development of prognostics for aerospace systems. In 2017, he joined the PhotoNext laboratory of the Politecnico di Torino, working on developing and integrating optical sensors in aerospace systems.