

Article

Assessment of Early Stopping through Statistical Health Prognostic Models for Empirical RUL Estimation in Wind Turbine Main Bearing Failure Monitoring

Jürgen Herp ^{1,*} , Niels L. Pedersen ² and Esmaeil S. Nadimi ¹ 

¹ The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, 5230 Odense, Denmark; esi@mmmi.sdu.dk

² Diagnostics, Siemens Gamesa Renewable Energy, 7330 Brande, Denmark; Niels.L.Pedersen@siemensgamesa.com

* Correspondence: herp@mmmi.sdu.dk

Received: 2 December 2019; Accepted: 20 December 2019; Published: 23 December 2019



Abstract: Details about a fault's progression, including the remaining-useful-lifetime (RUL), are key features in monitoring, industrial operation and maintenance (O&M) planning. In order to avoid increases in O&M costs through subjective human involvement and over-conservative control strategies, this work presents models to estimate the RUL for wind turbine main bearing failures. The prediction of the RUL is estimated from a likelihood function based on concepts from prognostics and health management, and survival analysis. The RUL is estimated by training the model on run-to-failure wind turbines, extracting a parametrization of a probability density function. In order to ensure analytical moments, a Weibull distribution is assumed. Alongside the RUL model, the fault's progression is abstracted as discrete states following the bearing stages from damage detection, through overtemperature warnings, to over overtemperature alarms and failure, and are integrated in a separate assessment model. Assuming a naïve O&M plan (wind turbines are run as close to failure as possible without regards for infrastructure or supply chain constrains), 67 non run-to-failure wind turbines are assessed with respect to their early stopping, revealing the potential RUL lost. These are turbines that have been stopped by the operator prior to their failure. On average it was found that wind turbines are stopped 13 days prior to their failure, accumulating 786 days of potentially lost operations across the 67 wind turbines.

Keywords: wind turbines; condition monitoring; inference; neural networks; remaining-useful-lifetime; main bearing

1. Introduction

Health prognostics in asset assessment, including remaining-useful-lifetime (RUL) estimations, are key elements in operation, and maintenance (O&M) strategies. This can help to increase production and/or reduce O&M costs. As wind power is a leading renewable energy source, availability, reliability, and lifetimes are taken incrementally into account by investors. In this work, we focus on slow developing faults, which, when not addressed, can cause unwanted or unnecessary costly downtime [1–8]. Specifically, we will focus on (i) investigate main bearing monitoring, (ii) give a full account of the underlying neural network (NN) approach presented by Herp et al. [9] on different timescales, (iii) present how this ties into O&M efforts, and (iv) compare expected main bearing RUL with model predictions. The later will give raise to a discussion on early stopping of wind turbines and potential waste of RUL (when O&M planning requires to stop the wind turbine ahead of time).

Models for assessing a systems current and future condition are often data-driven, and thus different approaches can be used depending on the type of data at hand. Trappey et al. [10] uses a combined statistical and NN approach, in which principal component analysis is used to extract relevant features. A NN was trained with back-propagation on these features to learn and predict the condition of power transformers. You et al. [11] addresses the use of NNs, by including temporal dependencies, they propose a diagnostic approach for electrical car batteries through the utilization of recurrent neural networks (RNN). On the other hand, Herp et al. [12] proposes a pure statistical approach based on the history and descriptive statistics of the observed data, using a Gaussian process to predict the future of bearing failure time-series in wind turbines. Hong et al. [13] similarly addresses the problem of bearing monitoring by extracting features and using an approach combining NN with self-organizing map approach to estimate the confidence of the bearing health states. In contrast to the aforementioned, Si [14] proposed an approach based on the time-series stochastic properties. Assuming an underlying driving Brownian motion, a closed-form predictive distribution for the time it takes the signal to reach a threshold can be calculated. Medical research uses survival analysis to understand the relations between patients' clinical features and the effectiveness of treatment options. Khan et al. [15] and others like Katzman et al. [16] have recently shown that prognostic and health management, and NNs can be combined to outperform existing state-of-the-art survival analysis models. A comprehensive overview of advances in RUL prediction can be found in Si et al. [17], Chapter 1. For a summary on wind turbine monitoring techniques, we refer to Márquez [18].

Catering to the need of health prognostics, the details of the model proposed by Herp et al. [9] are inspired by concepts of the above mentioned work. It adopts the closed-form approach by Si et al. [17] by assuming the RUL can be quantified by a closed-form expression and embeds it in a NN, to make confidence calculations of the predictions easier. A Weibull distribution has been chosen as a lifetime distribution, since literature already has shown its use in connection with NNs [19–22]. In contrast to Ranganath et al. [19] deep exponential families model, this work will limit itself to Weibull distribution's only. Yang et al. [23] and Aggarwal et al. [24] developed similar Weibull based NNs and RNNs, respectively, but focus amongst others on different disciplines than wind turbine monitoring.

For the sake of completeness, we refer the interested reader to Herp et al. [9] for a comparative study between different RUL estimation frameworks in empirical bearing fault prediction, as this study will not be concerned with the topic.

This paper is organized as follows: In Section 2, the methods for the RUL estimation are described in detail. The presented methodology is applied in Section 3 showing the potential gain from re-evaluating early stopping. A discussion on the results and the underlying assumption are provided in Section 4. The paper is closed by a final remark in Section 5.

2. Methodology

As this study is concerned with both the empirical estimation of the remaining-useful-lifetime of main bearings and the assessment of early stopping, both topics will have a dedicated section of their own. Even if treated separately here, we will later show that these methods are very much intertwined when it comes to monitoring and operation of one or more wind turbines. In Figure 1 we present how a monitoring and operation work flow might look like in order to make educated decisions based on estimations of the RUL. Fault monitoring is embedded in a back-end operations monitoring framework, which facilitates fault detection. We will not spend more thoughts on fault detection in this work and assume that one of the many fault detection approaches described in the literature can be applied. On a fault monitoring level, upon having collected data regarding the fault-to-be, we can distinguish between three main strategies for later decision making. (1) Assessing the situation at the time of detection, draw a conclusion and decide on an O&M action. Options (2) and (3) are assessments over time, where the aforementioned option derives a new assessment at given condition monitoring (CM) points spaced out over the time until failure, and the latter provides an assessment at any given time. For the remainder of this work we

refer to, (1) *Initial Assessment*, (2) *Discrete Assessment*, and (3) *Continuous Assessment*, collectively, as *Empirical Remaining-Useful-Lifetime Estimation*. The conclusion of the assessments is then facilitated again in the Operations Monitoring, while actions are left outside the monitoring framework as they are a matter of O&M management. In the conclusion step, one would estimate the benefit of continuing or stopping operations. We refer to this as *Assessment of Early Stopping*.

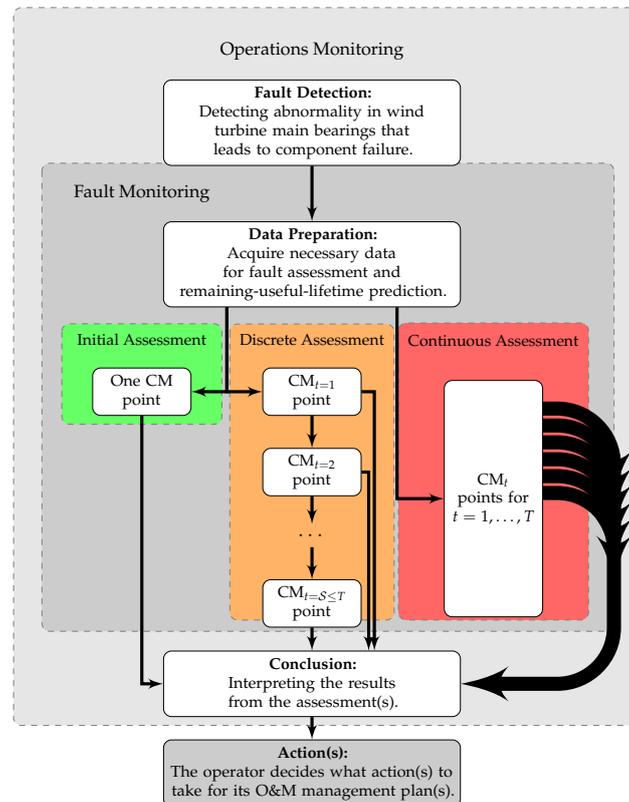


Figure 1. High-level workflow for remaining-useful-lifetime O&M management.

2.1. Data and Notation

This study is concerned with two types of data, time-series data and event data. The time-series data are composed out of a subset of SCADA data and other health indicators, while the event data contains an event description together with its start and stop time. Since a main bearing failure will be investigated as a case study, the SCADA and event data are associated with the failure (including events for the initial damage detection, failure warnings, and failure alarms) and stems from 132 wind turbines of the same type, of which 35 are operated until failure.

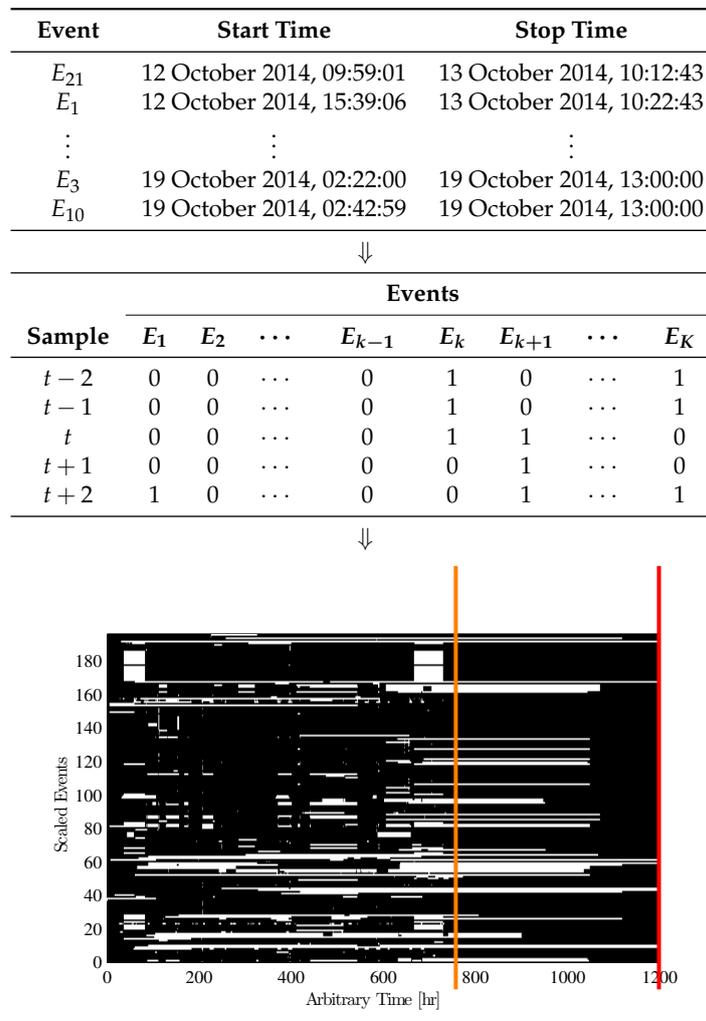
Regarding the time-series data, we let $\{x_1, \dots, x_T\}$ be a process of measurement vectors, $x_t \forall t = 1, \dots, T$, containing measurements for m variables, i.e., $x_t = [x_1, \dots, x_m]^T$. Successive samples on an interval a to b are denoted $x_{[a,b]}$. The feature space available in this study is the same as used by Bach-Andersen et al. [25] and Herp et al. [9], namely, active power, generator rpm, gear box oil temperature, ambient temperature, wind speed, nacelle temperature, and a bearing health indicator. Apart from the bearing health indicator, all features are sampled each 10 min as averages of the past 10 min interval. The bearing health indicator is based on energy bands in vibration spectra and is provided as an event-based measurement. For the remainder of this work, the SCADA and health indicator data are re-sampled to hourly timestamps. The time-series data are the foundation for the predictive models described in Section 2.2 and employed in Section 3.

The event data are a set of discrete data points where we let $\mathcal{E} = \{E_1, \dots, E_K\}$ be the set of all events $E_k \forall k = 1, \dots, K$, and we let $\mathcal{T} = \{1, \dots, t, \dots, T\}$ be the time at which an event can occur in the reference frame of the time-series data. An outtake of these data is shown in Table 1. Starting with the

initial detection of the damage, hundreds of events are recorded until failure. The failure of the wind turbine of Table 1 is indicated by the red vertical line (see graph), and the event associated with the failure is *bearing overtemperature alarm*. Wind turbines for which this event is recorded are referred to as run-to-failure wind turbines. These wind turbines also contain the event *bearing temperature warning*, shown by the orange vertical line in Table 1. Other combinations of events, different from *bearing overtemperature warning* and *bearing overtemperature alarm*, might likewise carry valuable information between the initial detection of the damage and the wind turbine failure. Considering a chain of successive and/or simultaneous events, the dependency between two sets of events $\mathcal{E}_l \Rightarrow \mathcal{E}_{l'}$ will be defined by:

$$\begin{aligned} \mathcal{E}_l &\Rightarrow \mathcal{E}_{l'} && l \neq l' && (1a) \\ \mathcal{E}_l, \mathcal{E}_{l'} &\subseteq \mathcal{E} && && (1b) \\ \mathcal{E}_l \cap \mathcal{E}_{l'} &\neq \emptyset. && && (1c) \end{aligned}$$

Table 1. Binary Event Mapping: Example of mapping from recorded event data into a binary array for identifying states, including bearing temperature warning (-) and bearing temperature alarm (-).



The found sets \mathcal{E}_k can then be used to establish a state as defined later on in Equation (15).

2.2. Empirical Remaining-Useful-Lifetime Estimation

As we are concerned with the time between detecting a failure and the failure itself, we let the RUL be defined to be positive bound, i.e., $RUL \in [0, \infty)$. Following the lead of probabilistic estimations of the remaining-useful-lifetime in literature, this section provides the missing detailed account for the underlying framework used in Herp et al. [9] for wind turbine bearing failure.

In this study, $\eta(t)$ is referred to as the hazard function and $H(t)$ is referred to as the cumulative hazard function. Following textbooks on survival analysis [26], $H(t)$ leads to a cumulative distribution function for a positive random variable RUL:

$$\mathbb{P}(RUL \leq t) = 1 - \exp\left(-\int_0^t \eta(w)dw\right) = 1 - e^{-H(t)}, \quad (2)$$

and a probability density function:

$$\mathbb{P}(t) = \frac{\partial}{\partial t} \mathbb{P}(RUL \leq t) = \frac{\partial}{\partial t} \left(1 - e^{-H(t)}\right) = e^{-H(t)} \eta(t). \quad (3)$$

As a consequence, for each cumulative distribution function, $\mathbb{P}(RUL \leq t)$, there exists $H(t)$ such that $H(t) = -\log(1 - \mathbb{P}(RUL \leq t))$. In terms of remaining-useful-lifetime, large values of the hazard function indicate a higher chance of an failure to occur up to the current time, i.e., higher failure rate in terms of fault prediction. For the remainder of the study, the focus will be on the right tail of the distribution under consideration, which will define the RUL distribution [26]:

$$L(t) = 1 - \mathbb{P}(RUL \leq t) = e^{-H(t)}. \quad (4)$$

As the time it takes to observe the remaining-useful-lifetime is the remaining lifetime of the wind turbine itself, the true remaining-useful-lifetime, \hat{RUL} , cannot be observed until it is too late. This premise is known as censored observations in survival analysis [26]. Let RUL be some positive random variable indicating the remaining-useful-lifetime. An observation of RUL is said to be censored whenever it has been observed point wise. In this framework one might distinguish between true \hat{RUL} and observed RUL. The true \hat{RUL} is always contained in the observation. When \hat{RUL} equals the observed value, the process is referred to as non-censored. More specifically we distinguish between (i): right censored data $\Delta = 1$, such that RUL is know to be above a threshold t (i.e., $RUL \in (t, \infty)$), and (ii) non-censored $\Delta = 0$, when \hat{RUL} equals the observed value (i.e., $RUL = t$). Censoring the remaining-useful-lifetime of wind turbine prohibits simply using the mean value of the already failed wind turbines, as this will lead to underestimating the RUL.

A full mathematical proof of the following description can be found in Patti et al. [27]. In Figure 2 the censoring idea is illustrated: the parametrized probability density function tightens when the failure is observed or is pushed beyond the current observed point, if the failure is censored.

Under the assumption of non-informative censoring [26], the likelihood reads:

$$\mathcal{L} = \mathbb{P}(t)^\Delta \mathbb{P}(RUL > t)^{1-\Delta} \quad (5)$$

$$= \mathbb{P}(t)^\Delta L(t)^{1-\Delta} \quad (6)$$

$$= e^{-\Delta H(t)} \eta(t)^\Delta e^{-(1-\Delta)H(t)} \quad (7)$$

$$= \eta(t)^\Delta e^{-H(t)}. \quad (8)$$

Thus, the log-likelihood can be written as:

$$\log \mathcal{L} = \Delta \log \eta(t) - H(t). \quad (9)$$

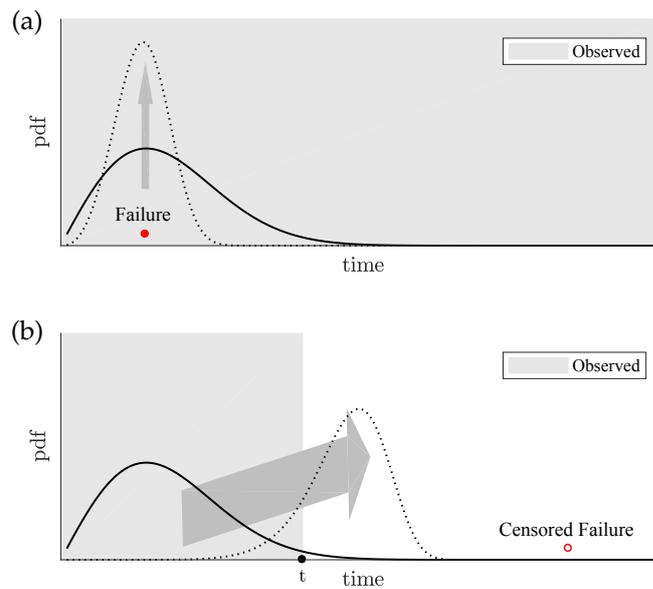


Figure 2. (a) Given the failure is observed (non-censored) it is desired to harden a distribution at the time of failure. (b) A distribution is estimated based on the measurements up to time t . As the failure is not observed *yet* (censored) the distribution is pushed beyond the point t into the unobserved future.

For non-censored and censored failures, the log-likelihood will always have a negative contribution with increasing t , penalizing as time progresses. Following Herp et al. [9] we consider a Weibull distribution, for its tractable properties, intuitive parametrization and use in other related studies [19–22]:

$$\mathbb{P}(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha} \tag{10}$$

$$\mathbb{P}(\text{RUL} \leq t) = 1 - e^{-\left(\frac{t}{\beta}\right)^\alpha} . \tag{11}$$

Here, $t \in [0, \infty)$, $\beta \in (0, \infty)$, and $\alpha \in (0, \infty)$, where α and β are referred to as the scale and shape of the distribution. Beside other features, such as analytic moments, the Weibull distribution has an analytic hazard and cumulative hazard function, which can be obtained by Equations (2) and (3):

$$\eta(t) = \left(\frac{t}{\beta}\right)^{\alpha-1} \frac{\alpha}{\beta} . \tag{12}$$

$$H(t) = \left(\frac{t}{\beta}\right)^\alpha \tag{13}$$

The cost associated with the Weibull function is then given by Equation (9) and reads:

$$\log(\mathcal{L}) = \Delta \left(\alpha \log \left(\frac{t}{\beta} \right) + \log(\alpha) - \log(t) \right) - \left(\frac{t}{\beta} \right)^\alpha . \tag{14}$$

2.2.1. Initial Assessment

The initial assessment contains only one estimation of the remaining-useful-lifetime, based on the descriptive statistics and maximizing the likelihood as described by Equation (14). Figure 3 shows the Weibull model for run-to-failure wind turbines (bearing overtemperature *alarm*), wind turbines running into bearing overtemperature *warning*, wind turbines excluded from the model construction (as they do not run into any aforementioned event), are illustrated by their cumulative distribution as a scatter plot (*non-failing*). For later comparison, a *stopped* by operator model is obtained for the non-failing

wind turbines. Furthermore, the Kaplan-Meier [28] estimation (KM) is provided for reference, as it makes no assumption with regard to the shape of the probability distribution.

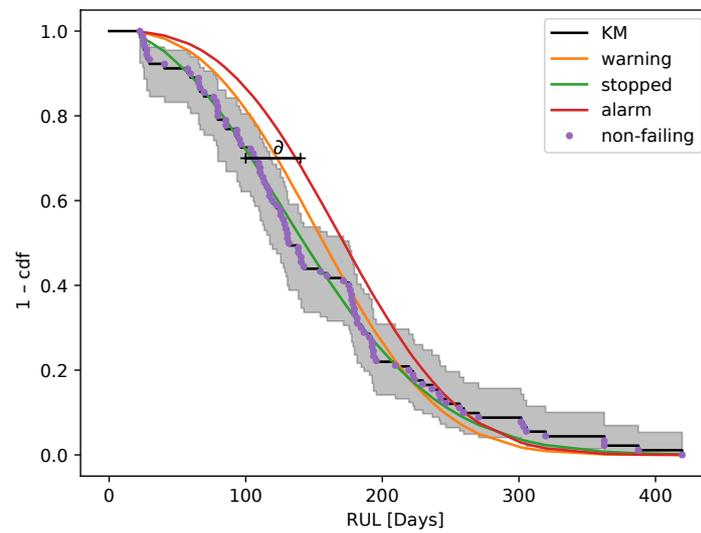


Figure 3. Initial Assessment: Comparison of different models up to different events including the cumulative distribution over wind turbines not running into an event. For comparison the Kaplan–Meier (KM) estimation is provided for the non-failing wind turbines, including confidence bound. δ is the mismatch between a non run-to-failure (non-failing) wind turbine and the bearing overtemperature alarm (alarm).

2.2.2. State Abstraction and Discrete Assessment

The same principle as in Figure 3 applies to the discrete assessment. However, in order to perform the discrete assessment, condition monitoring (CM) points will need to be taken into account. How these CM points are extracted is described in this section.

Between initial detection and failure of a damaged bearing, the bearing undergoes transitions from a minor fatigue to advanced fatigue and damage. In combination with Equation (1), this section addresses how to identify individual stages of the bearing fatigue to lay the foundation of the *discrete assessment* later on.

Consider an event E_k or set of events \mathcal{E}_k , from a library of events $\mathcal{E} = \{E_1, \dots, E_K\}$ (abstracted by Equation (1)), linked to the operation of a wind turbine at time t , to be dependent on the data collected up to time t and a hidden state variable $s^{(t)}$, $s^{(t)}$ representing the current state of the wind turbine. The probability for \mathcal{E}_k can be given by:

$$\mathbb{P}(\mathcal{E}_k | s^{(t)}, \mathbf{x}_{[1,t]}) = \frac{\mathbb{P}(\mathcal{E}_k, s^{(t)}, \mathbf{x}_{[1,t]})}{\mathbb{P}(s^{(t)} | \mathbf{x}_{[1,t]})\mathbb{P}(\mathbf{x}_{[1,t]})}, \tag{15}$$

further, S_m is referred to as the hidden states, defined by the separability of the process $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ into $\mathcal{S} \leq T$ states. The transition between those states is called state transition, where $s^{(t)}$ is the length of the current state with samples $\mathbf{x}_{[S_S,t]}$, with S_S being the time of the last state transition. Following Herp et al. [29,30] and Prescott Adam et al. [31], a state in a set of time-series can be abstracted by considering only the maximum likelihood of $\mathbb{P}(s^{(t)}, \mathbf{x}_{[1,t]})$, where

$$\mathbb{P}(s^{(t)}, \mathbf{x}_{[1,t]}) = \sum_{(s^{(t-1)})} \underbrace{\mathbb{P}(s^{(t)} | s^{(t-1)})}_{\text{conditional prior}} \underbrace{\mathbb{P}(\mathbf{x}_t | s^{(t-1)}, \mathbf{x}_{[1,t-1]})}_{\text{sample model}} \mathbb{P}(s^{(t-1)}, \mathbf{x}_{[1,t-1]}). \tag{16}$$

Here the conditional prior and sample model are implicit depending on known hyper-parameters $\beta = [\beta_c, \beta_m]$, where β_c and β_m are the parametrization of the conditional prior and the sample model. Thus, $\mathbb{P}(s^{(t)} | s^{(t-1)}) \equiv \mathbb{P}(s^{(t)} | s^{(t-1)}, \beta_c)$ and $\mathbb{P}(\mathbf{x}_t | s^{(t-1)}, \mathbf{x}_{[1,t-1]}) \equiv \mathbb{P}(\mathbf{x}_t | s^{(t-1)}, \mathbf{x}_{[1,t-1]}, \beta_m)$.

Given the nature of the data we are seeking a sample model that addresses changes in the first and second moment of the time-series. This is achieved by classifying each $s^{(t)}$ by the two moments $\mathbb{E}[\mathbf{x}]$ and $\mathbb{E}[\mathbf{x}^2]$. Given sufficient statistics for each $s^{(t)}$, a posterior distribution for \mathbf{x}_t is given for the iterative update [32,33]:

$$\mu_t = \frac{\mu_0 \kappa_0 + \mathbb{E}[\mathbf{x}_{[1,t]}]}{\kappa_0 + t}, \quad (17a)$$

$$\kappa_t = \kappa_0 + 1, \quad (17b)$$

$$\alpha_t = \alpha_0 + t, \quad (17c)$$

$$\xi_t = \sum_{j=1}^t (\mathbf{x}_j - \mathbb{E}[\mathbf{x}])^2, \quad (17d)$$

$$\gamma_t = \gamma_0 + \frac{1}{2} \xi_t + \frac{\kappa_0 t (\mathbb{E}[\mathbf{x}_{[1,t]}] - \mu_0)^2}{2(\kappa_0 + t)}, \quad (17e)$$

with $\mu_0, \kappa_0, \alpha_0, \gamma_0$ being the previous statistics. A Student's t-distribution can then be used for the posterior distribution [34] with ν degrees of freedom and μ and σ as mean and variance, it follows that:

$$\mathbb{P}(\mathbf{x}_t | s^{(t-1)}, \mathbf{x}_{(1,t-1)}) = \text{St}_{2\alpha_t} \left(\mu_t, \frac{\gamma_t \kappa_t + 1}{\kappa_t} \right). \quad (18)$$

The conditional prior is chosen as a hazard function, as motivated in Section 2.2, such that

$$\mathbb{P}(s^{(t)} | s^{(t-1)}) = \begin{cases} \eta & \text{if } s^{(t)} = 1 \\ 1 - \eta & \text{if } s^{(t)} = s^{(t-1)} + 1 \\ 0 & \text{else} \end{cases}, \quad (19)$$

here η is a hazard function of a geometric sequence and given by the elapsed time since the last state transition:

$$\eta = \frac{1}{\mathbb{E}[s^{(t)}]}. \quad (20)$$

For the remainder of this work, the algorithm is implemented with a supervised probability update as proposed by Herp et al. [29].

In order to quantify states, a piecewise constant regression for finite state numbers is performed on the cumulative probability state transitions:

$$\hat{P}(t) = \sum_{s=2}^{S+1} \alpha_s \chi_{A_s}(t), \quad (21)$$

s.t.

$$\arg \max_{\alpha_s, \chi_{A_s}} (P(t) - \hat{P}(t))^2$$

$$\text{length}(A_s) \geq 5$$

where α_s is a real number, the states cumulative probability, and A_s the interval of the s th state, i.e., $[S_{m-1}, S_m)$, χ_{A_s} is an indicator function that is 1 if $t \in A$, and 0 otherwise. The constraint on A_s comes from the delay of detecting a new state.

In detail, the state abstraction for the health indicator can be seen in Figure 4. From top to bottom the health indicator for a wind turbine is shown in (a), (b) shows the log-likelihood for the state transitions (gray-scale) and its maximum likelihood, (c) shows the state transition probability density function and cumulative density function, in addition the abstracted states in accordance with Equation (21) are provided.

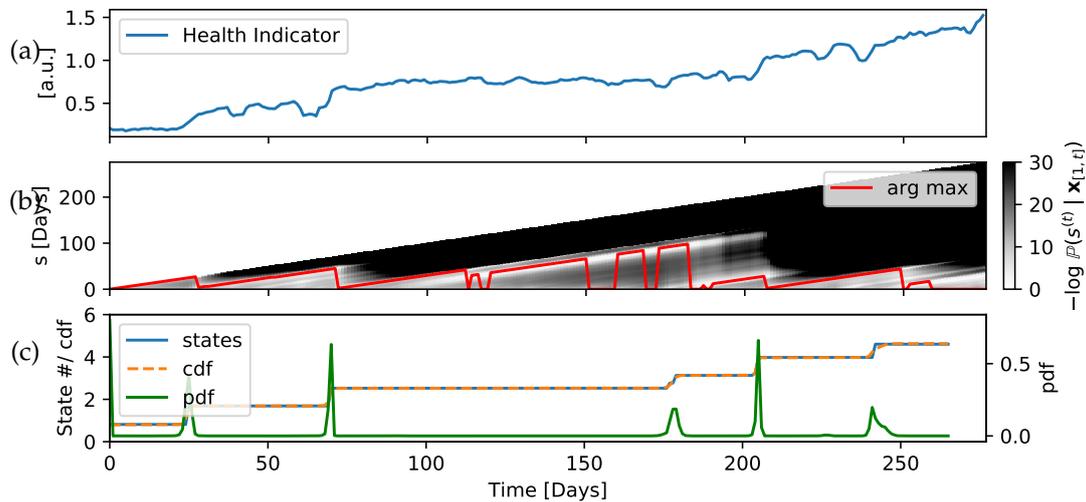


Figure 4. State Abstraction: (a) Wind turbine health indicator. (b) log-likelihood for state transition. (c) Transition probability and states.

The state abstraction give us a measure for the length of each state. Given the time for each turbine to any state transition, the RUL probabilities can be calculated equivalent to the initial assessment for any combinations of state transitions. The implementation and results of this approach are discussed in the case study of Section 3, providing the RUL probabilities from initial detection to any of the state transition. Figure 10).

2.2.3. Continuous Assessment: RUL Recurrent Neural Network

We employ a recurrent neural network (RNN) for the continuous assessment for maximizing the cost function given by Equation (14) at each t . In general, this is an optimization problem of finding a function or distribution in the space of hazard function, $F \in \mathbf{H}$, that maximizes the model's likelihood given historical data, $\mathbf{x}_{[1,t]}$. The space of all possible closed-form distributions in \mathbf{H} is too large to obtain solutions in an easy manner. Furthermore, as $\mathbf{x}_{[1,t]}$ can contain any information at time t , including information of the previous states [30] the optimization for all \mathbf{x} is computational not feasible. However, constraining the problem to a Weibull distribution and letting the Weibull distribution be dependent on $\mathbf{x}_{[1,t]}$, it then follows from Equations (12) and (13), as well as Equation (14) that:

$$\begin{aligned} \arg \max_{\alpha > 0, \beta > 0} \log \mathcal{L}(\alpha, \beta, \text{RUL}, \Delta, \mathbf{x}_{[1,t]}) &= \sum_{i=1}^t \left(\Delta_i \alpha(\mathbf{x}_{[1,i]}) \log \left(\frac{\text{RUL}}{\beta(\mathbf{x}_{[1,i]})} \right) \right. \\ &\quad \left. + \Delta_i \left[\log \left(\alpha(\mathbf{x}_{[1,i]}) \right) - \log(\text{RUL}) \right] - \left(\frac{\text{RUL}}{\beta(\mathbf{x}_{[1,i]})} \right)^{\alpha(\mathbf{x}_{[1,i]})} \right) \end{aligned} \quad (22)$$

This optimization is still not feasible, as for a large dataset, each parameter needs to be obtained through consideration of all historic data. In the context of Figure 5, θ can be written as a vector $\theta = [\alpha, \beta]^\top$. Omitting mathematical details on node level, which can be found in a wide range of textbooks such as Goodfellow et al. [35], each node in Figure 5 outputs $o_i = g(\mathbf{x}_i, o_{i-1}, \omega_0)$, for $i = 1, \dots, t$, a function of the data \mathbf{x}_i , previous output o_{i-1} , and the networks topology ω_0 . The network's output at each t will be of the form $\theta = m(\mathbf{x}_t, o_{t-1}, \omega) = [\alpha, \beta]^\top$, where $m(\cdot)$ is

the mapping for the topology of the chosen RNN. Equation (22) is thus reduced to solving the gradient by back-propagation for a set of optimal or suboptimal ω :

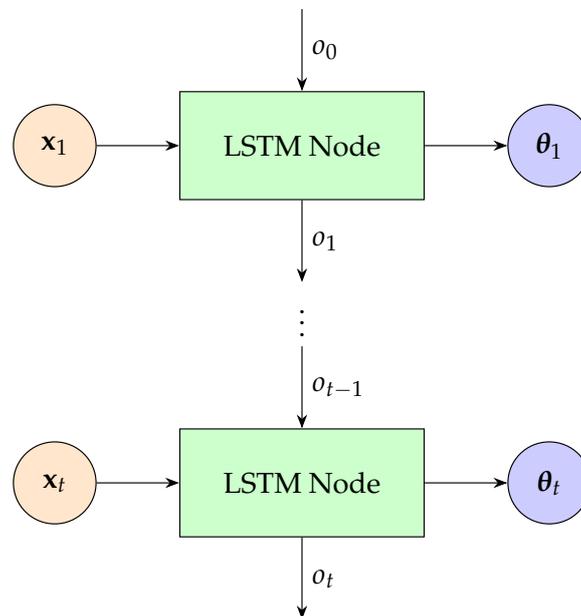


Figure 5. Schematic drawing the recurrent neural network sequence: input feature, current, previous, and future hidden states o_t .

$$\arg \max_{\omega} \log \mathcal{L}(\omega, RUL, \Delta, \mathbf{x}_{[1,t]}) = \sum_{i=1}^t \left(\Delta_i \left[\alpha_i \log \left(\frac{RUL}{\beta_i} \right) + \log(\alpha_i) - \log(RUL) \right] - \left(\frac{RUL}{\beta_i} \right)^{\alpha_i} \right) \quad (23)$$

For the remainder of this study, we use a Long Short Term Memory (LSTM) RNN [35] and maintain the topology, feature space, and optimization as in Herp et al. [9] (See Figure 6). The sequence length is set to 7 days. An example on how predictions look like can be found in Figure 7, showing the assumable \hat{RUL} and model prediction. The prediction is shown in terms of the first *moment*, *median*, and *arg max* of the Weibull distribution at each time t . The distribution itself is illustrated as shaded areas.

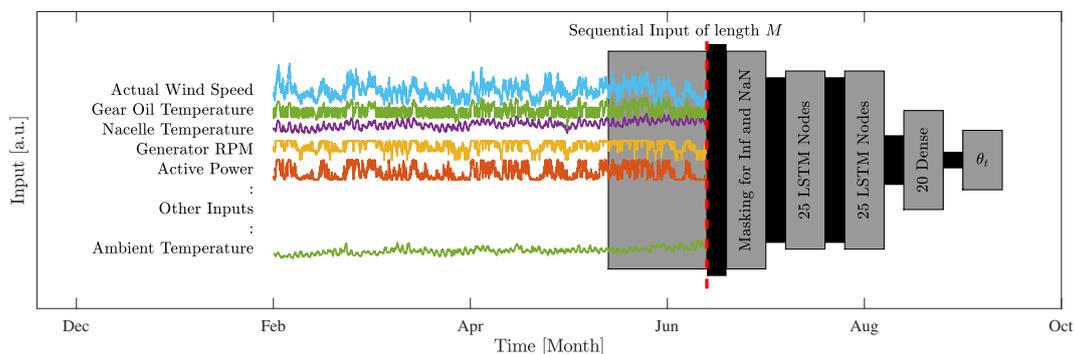


Figure 6. Illustration of the RNN used to predict the RUL. Selected SCADA time-series are shown as input [9].

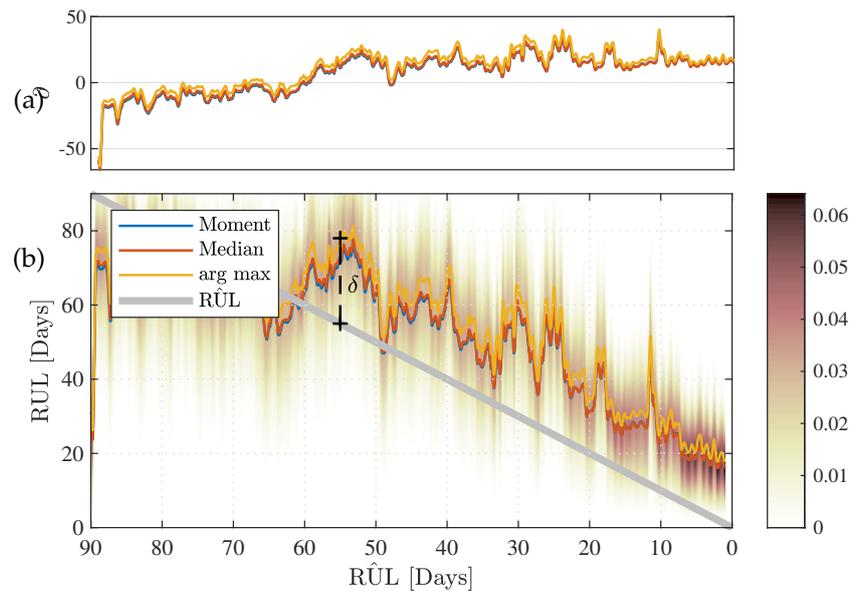


Figure 7. (a) Continuous prediction of RUL of a non run-to-failure wind turbine, compared to the expected run-to-failure \hat{RUL} . (b) Mismatch between \hat{RUL} and the predictive distributions first moment, median, and arg max.

3. Assessment of Early Stopping—A Bearing Failure Study

This section combines the afore mentioned methodology and applies it to 67 non run-to-failure wind turbines. For each assessment, all turbines are investigated with respect to early stopping, showing the lost remaining RUL after turbine operations were stopped. The remaining turbines are used for model construction.

Consider the example of Figure 3 where models for the *bearing overtemperature warning* and *bearing overtemperature alarm* are illustrated. Comparing each non run-to-failure wind turbine to any of the event models shows a mismatch δ , which is defined as the distance between any given wind turbine and the predictive models. This mismatch is interpreted as the lost remaining RUL.

Under continuous assessment, Figure 7 shows the mismatch between the presumably true \hat{RUL} and the model prediction for a non run-to-failure wind turbine. In this case we consider an average measure for each wind turbine:

$$D^{(\text{turbine})} \equiv \frac{1}{T} \sum_{t=1}^T \delta_t^{(\text{turbine})}, \quad (24)$$

where

$$\delta_t^{(\text{turbine})} \equiv RUL_t^{(\text{turbine})} - \hat{RUL}_t^{(\text{turbine})}. \quad (25)$$

$D^{(\text{turbine})}$ will then be a measure of the specific turbine's potential remaining RUL. In the following we employ the methodology of this section in a case study of mean bearing failures.

When addressing whether or not to stop a wind turbine and perform maintenance is a topic on its own. Many factors, e.g., weather condition, availability of equipment, spare parts, and manpower, have to be taken into consideration when deciding and planing O&M tasks. In the following, we simplify towards a pure time-based approach. Many wind turbines are stopped before they run into failure; based on the proposed RUL estimations, the potential remaining time until failure can be estimated.

3.1. Initial Assessment

Model construction can be facilitated up to different events, including *bearing overtemperature warning* and *bearing overtemperature alarm*. For the wind turbines in this study, the respective models are shown in Figure 3. Assuming that all turbines experience the same fault, the difference between the two models (warning and alarm) should be an offset in the predictive horizon. However, variability in environmental features and the bearing failure's development can lead to changes of the time it takes to undergo a state transition from the *bearing overtemperature warning* to the *bearing overtemperature alarm* event. This is illustrated in Figure 8, labeled warning vs. alarm. The low curvature of the graph indicates that the difference between the two models is relatively small, and relatively constant throughout the RUL space. Given the limited number of wind turbines, that are operated beyond $RUL > 250$ days and turbines that run into any event before $RUL < 100$ days, broader confidence bounds can be observed in these RUL ranges. Besides the time between warning and alarm, we can consider a model for the 67 non run-to-failure wind turbines, referred to as the *stopped* model in Figure 3. A mismatch δ for this model with respect to the warning and alarm model is shown in Figure 8. Comparing the models, turbines that have run 100 days, and were stopped too early would have had 30 days left of operations until running into failure, i.e., the *bearing overtemperature alarm*. Or if a more conservative stopping criterion is desired, 20 days until the *bearing overtemperature warning*. Variability in wind turbine operations leads to a point where the expected mismatch becomes negative; in this regime, the wind turbine is operated at the right tail of the RUL probability curve of Figure 3. An operator might interpret this as the time where the wind turbine has overextended the confidence of continuing operations, when comparing to the ensemble of already failed turbines.

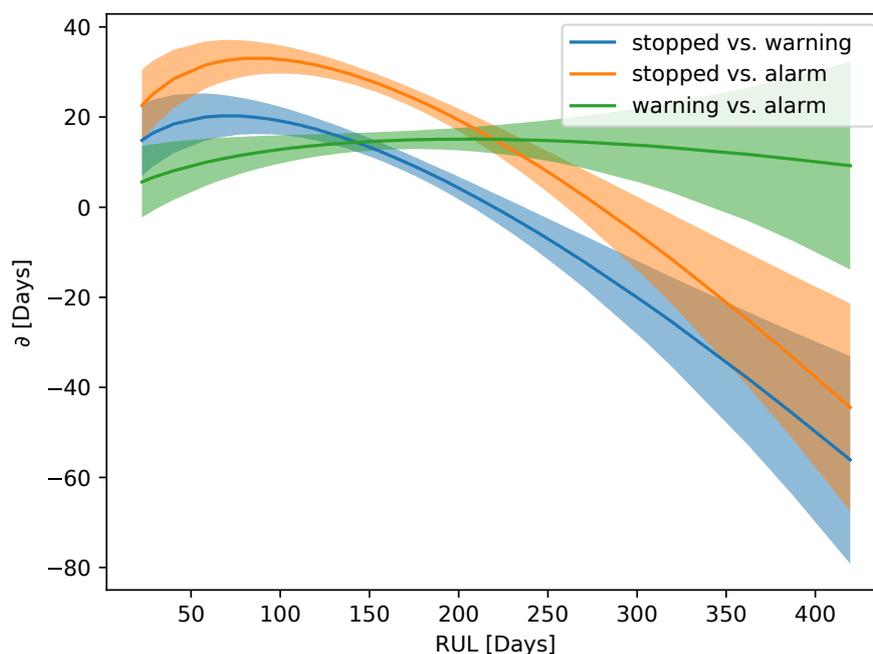


Figure 8. Comparison of initial assessment of RUL with the 67 wind turbines that did not run into either *bearing overtemperature* events, *bearing overtemperature warning*, and *bearing overtemperature alarm* models. 0.95 confidence bounds are provided.

The mismatch to failure over all individual wind turbines can be seen in Figure 9a as a histogram. The largest mismatch is 57 days, while the smallest is -40 days. The mean value and median are 11 and 15 days, respectively. The bulk of turbines is concentrated in the quantiles between 3 and 23 days. The combined mismatch accumulates to 782 days.

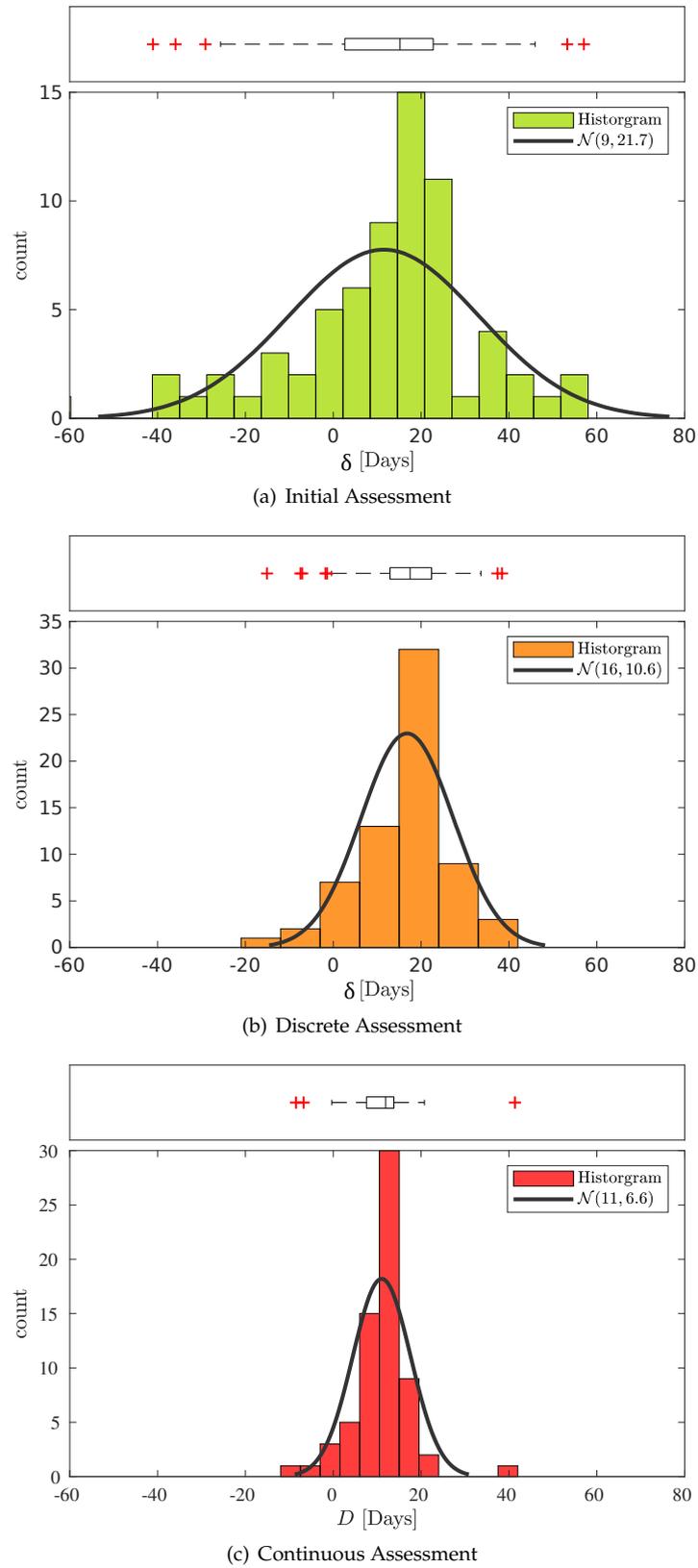


Figure 9. Distribution over mismatch for the 67 wind turbines that did not run into either *bearing overtemperature* events. (a) initial, (b) discrete, and (c) continuous assessment.

3.2. Discrete Assessment

In its core, the discrete assessment with CM points at each state transition, is the same as the initial assessment but applied at each CM point. As failure of a component can go through different stages, the operator might want to re-evaluate when the component undergoes the next state transition.

The state detection and abstraction is performed as outlined in Section 2.2.2. As the bearing failures in this case study are slowly developing faults, we impose a minimum state length of $A_s \geq 5$ days, as seen in the constraint of Equation (21). This is done in order to prevent fast switching states in case of random outliers.

In Figure 10, the the cumulative state lengths are shown for each turbine. For each state transition, the RUL for that state is provided. This gives the operator information on when to expect the next state transition of a damaged component. Remark: as the number of states increases, the models become less defined as confidence bounds increase. This is caused by the limited sample size of turbines that experience three or more states transitions.

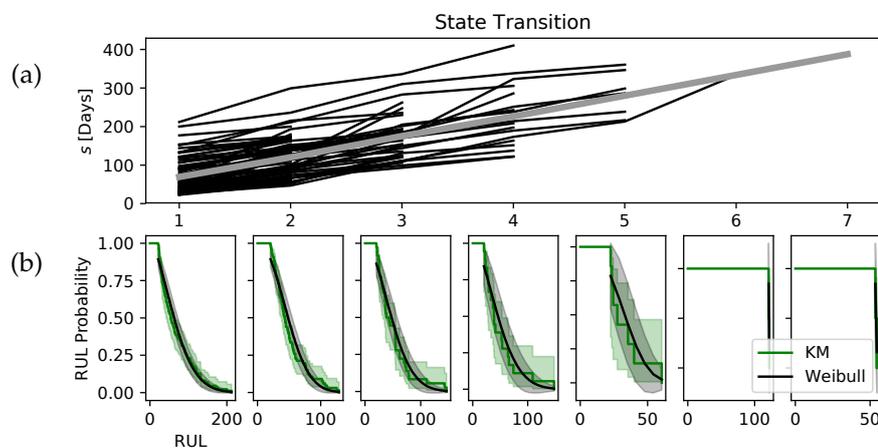


Figure 10. (a) Cumulative state length of wind turbines, including best linear fit. (b) RUL probability based on the length between each state transition.

The expected RUL is obtained by comparing the wind turbine's cumulative state length with the state length of states for wind turbines abstracted with the event pattern $\mathcal{E}_l \Rightarrow \mathcal{E}_{l'}$, where $\mathcal{E}_{l'}$ contains the *bearing overtemperature alarm*. The mismatch to failure over all individual wind turbine can be seen in Figure 9b as a histogram. The largest mismatch is 38 days, while the smallest is -15 days, the mean value and median are 16 and 17 days, respectively. The bulk of turbines is concentrated in the quantiles between 10 and 20 days. The combined mismatch accumulates to 836 days.

3.3. Continuous Assessment

As shown in Figure 7 the continuous assessment is concerned with the prediction of the RUL at each time step. Focusing on either the first moment, median, or arg max of the Weibull distribution, it becomes apparent that there is a discrepancy between the model's prediction and expected RUL. This is the aforementioned mismatch $\delta_t^{(\text{turbine})}$ and may differ for each wind turbine. Similar graphs can be drawn for other turbines, with varying degrees of $\delta_t^{(\text{turbine})}$, containing also negative values of $\delta_t^{(\text{turbine})}$. The average mismatch to failure over all individual wind turbines can be seen in Figure 9c as a histogram. The largest mismatch is 41 days, while the smallest is -8 days; the mean value and median are 11 and 12 days, respectively. The bulk of turbines are concentrated in the quantiles between 10 and 15 days. The combined mismatch accumulates to 739 days.

4. Discussion

Common for all assessments is a high count of wind turbine mismatches at the respective mean values. These peaks can contain more than twice the amount of wind turbines than their neighboring bins, indicating not only the most likely mismatch, but also a firm consistency indicating a tendency to stop too early. However, two points are worth mentioning, (i) the different assessments do not yield the same results, and (ii) the accumulative mismatch is based on naïve O&M assumption.

4.1. Discrepancy Between Assessments

The discrepancy between the different assessments stem from the underlying nature of the descriptive statistics. As the initial assessment applies for the total length of failure, no iterative adjustment occurs in the monitoring effort, thus assessing wind turbines that are further into a fault process are not presented by the initial assessment. On the other hand, as the name implies, it gives an initial assessment of the expected length of the failure state and their relative probability based on an ensemble of wind turbines.

A step closer to a real representation of the RUL is provided by the discrete assessment. While the failure state can be split further into shorter states, the RUL and probability of failure is obtained by propagating non run-to-failure wind turbines through the RUL probabilities for each state transition. Besides, the discrete assessment can provide a measure of RUL until the next state transition, aiding an operator to know when to stop a wind turbine's operation for more conservative O&M strategies. For both the initial assessment and discrete assessment, RUL of individual wind turbines and, from there, the mismatch δ , are ensemble measures, i.e., when addressing the RUL of a single wind turbine with respect to a cumulative distribution obtained by all other turbines that were undergoing the same failure.

The initial and discrete assessments are in their core curve fitting problems for one or more datasets of wind turbine RULs. When the number of CM points goes towards the number of sample vectors, $S \rightarrow T$, one can consider the discrete and continuous assessment as equivalent. While the aforementioned assessment was an ensemble measure, the continuous assessment offers a real-time monitoring approach for individual wind turbines, where the RUL is updated at each successive time step. This makes the prediction of the RUL more and more reliable the closer a wind turbine is to failure, but provides poor or misleading assessments early in the fault process, where predictions are wide [9]. The further a wind turbine is in a fault state, a continuous assessment becomes more and more desirable.

All together, when combined as proposed in Figure 1, the proposed assessment offers an operator information at desired stages of a main bearing fault.

4.2. The Naïve O&M Assumption

The cumulative, potential lost RUL has to be looked upon with caution. Firstly, naïve O&M assumption takes away any other consideration that might play a role in stopping a turbine, and secondly, wind turbines might continue to operate reliably even if models suggest otherwise. Apart from the predictive error due to model constraints, it becomes apparent that RUL, i.e., a measure of time, is a difficult measure that is not easily related to the physical parameters of a wind turbine's operation. If a wind turbine is not operated due to lack of wind or other external factors, no matter how good a model, predictions will underestimate the RUL in any of the suggested frameworks.

5. Conclusions

We have proposed three fault monitoring and assessment concepts, namely, *initial assessment*, *discrete assessment*, and *continuous assessment*, showing that in any of the frameworks, wind turbines are stopped earlier than necessary compared to a global ensemble. Based on the descriptive statistics of 67 non-run-to-failure wind turbines, wind turbines are stopped 13 days too early, prior to their failure,

accumulating, on average, 786 of potentially lost days of operations under the naïve O&M assumption made in this study. Future work will probably focus on the inclusion of supply chain and weather constraints to represent a more realistic model mismatch behavior.

Author Contributions: Conceptualization: J.H., N.L.P. and E.S.N.; Methodology: J.H.; Resources: N.L.P.; Data curation: J.H.; Writing—Original draft preparation: J.H.; Writing—review and editing, J.H.; Supervision, E.S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by Siemens Gamesa Renewable Energy and the University of Southern Denmark.

Acknowledgments: The authors would like to take the chance to thank Siemens Gamesa Renewable Energy for providing us with the necessary data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

cdf	cumulative distribution function
CM	Condition Monitoring
KM	Kaplan-Meier
NN	Neural Network
O&M	Operation and Maintenance
pdf	probability density function
RNN	Recurrent Neural Network
RUL	Remaining-Useful-Lifetime
SCADA	Supervisory Control and Data Acquisition

Nomenclature

a	scalar
\mathbf{a}	vector
\mathbf{A}	matrix
\mathbf{x}_t	sample vector
$\mathbf{x}_{[a,b]}$	samples on the interval a to b
\mathcal{E}	set of all possible events
E_k	k^{th} event
\mathcal{E}_l	subset of \mathcal{E}
RUL	random variable for the remaining-useful-lifetime
\hat{RUL}	real remaining-useful-lifetime
η	hazard function
H	cumulative hazard function
$\Delta \in \{0, 1\}$	censoring variable
\mathcal{L}	likelihood function
$\boldsymbol{\beta} = [\beta_c, \beta_m]$	hyper-parameter for conditional prior and sample model
$\mathbb{P}(s^{(t)} s^{(t-1)}) \equiv \mathbb{P}(s^{(t)} s^{(t-1)}, \beta_c)$	independent conditional prior
$\mathbb{P}(s^{(t)} \mathbf{x}_{[1,t]})$	probability distribution over the current state
$\mathbb{P}(\mathbf{x}_t s^{(t-1)}, \mathbf{x}_{[1,t-1]}) \equiv \mathbb{P}(\mathbf{x}_t s^{(t-1)}, \mathbf{x}_{[1,t-1]}, \beta_m)$	sample model
δ	mismatch between RUL and \hat{RUL}
D	average mismatch

References

1. Petersen, K.R.; Madsen, E.S.; Bilberg, A. Offshore Wind Power at Rough Sea: The need for new Maintenance Models. In Proceedings of the 20th EurOMA Conference, Dublin, Ireland, 9–12 June 2013.
2. Nielsen, J.; Sørensen, J. On risk-based operation and maintenance of offshore wind turbine components. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 218–229, doi:10.1016/j.res.2010.07.007. [[CrossRef](#)]

3. Blanco, M.I. The economics of wind energy. *Renew. Sustain. Energy Rev.* **2009**, *13*, 1372–1382, doi:10.1016/j.rser.2008.09.004. [CrossRef]
4. van Bussel, G.J.W.; Zaaijer, M.B. Reliability, Availability and Maintenance Aspects of Large-Scale Offshore Wind Farms, a Concepts Study. In Proceedings of the MAREC 2001 Marine Renewable Energies Conference, Newcastle, UK, 27 March 2001; Volume 113, pp. 119–126.
5. Wilkinson, M.; Spinato, F.; Knowles, M. Towards the Zero Maintenance Wind Turbine. In Proceedings of the 41st International Universities Power Engineering Conference (UPEC '06), Newcastle, UK, 6–8 September 2006; Volume 1, pp. 74–78, doi:10.1109/UPEC.2006.367718. [CrossRef]
6. Kim, N.; An, D.; Choi, J. *Prognostics and Health Management of Engineering Systems: An Introduction*; Springer: Berlin, Germany, 2016.
7. Sheppard, J.W.; Kaufman, M.A.; Wilmer, T.J. IEEE Standards for Prognostics and Health Management. *IEEE Aerosp. Electron. Syst. Mag.* **2009**, *24*, 34–41, doi:10.1109/MAES.2009.5282287. [CrossRef]
8. Walford, C.A. *Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs*; Technical Report SAND2006-1100; Sandia National Laboratories: Albuquerque, NM, USA, 2006. Available Online: <https://energy.sandia.gov/wp-content/gallery/uploads/SAND-2006-1100.pdf> (accessed on 12 December 2019).
9. Herp, J.; Pedersen, N.L.; Nadimi, E.S. A Novel Probabilistic Long-Term Fault Prediction Framework Beyond SCADA Data—With Applications in Main Bearing Failure. *J. Phys. Conf. Ser.* **2019**, *1222*, 012043, doi:10.1088/1742-6596/1222/1/012043. [CrossRef]
10. Trappey, A.J.C.; Trappey, C.V.; Ma, L.; Chang, J.C. Integrating Real-Time Monitoring and Asset Health Prediction for Power Transformer Intelligent Maintenance and Decision Support. In *Engineering Asset Management—Systems, Professional Practices and Certification*; Tse, P.W., Mathew, J., Wong, K., Lam, R., Ko, C., Eds.; Springer: Cham, Switzerland, 2015; pp. 533–543.
11. You, G.W.; Park, S.; Oh, D. Diagnosis of Electric Vehicle Batteries Using Recurrent Neural Networks. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4885–4893, doi:10.1109/TIE.2017.2674593. [CrossRef]
12. Herp, J.; Ramezani, M.H.; Bach-Andersen, M.; Pedersen, N.L.; Nadimi, E.S. Bayesian state prediction of wind turbine bearing failure. *Renew. Energy* **2018**, *116*, 164–172, doi:10.1016/j.renene.2017.02.069. [CrossRef]
13. Hong, S.; Zhou, Z.; Zio, E.; Wang, W. An adaptive method for health trend prediction of rotating bearings. *Dig. Signal Process.* **2014**, *35*, 117–123, doi:10.1016/j.dsp.2014.08.006. [CrossRef]
14. Si, X.S. An Adaptive Prognostic Approach via Nonlinear Degradation Modeling: Application to Battery Data. *IEEE Trans. Ind. Electron.* **2015**, *62*, 5082–5096, doi:10.1109/TIE.2015.2393840. [CrossRef]
15. Khan, S.; Yairi, T. A review on the application of deep learning in system health management. *Mech. Syst. Signal Process.* **2018**, *107*, 241–265, doi:10.1016/j.ymsp.2017.11.024. [CrossRef]
16. Katzman, J.; Shaham, U.; Bates, J.; Cloninger, A.; Jiang, T.; Kluger, Y. DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [CrossRef]
17. Si, X.; Zhang, Z.; Hu, C. *Data-Driven Remaining Useful Life Prognosis Techniques: Stochastic Models, Methods and Applications*; Springer Series in Reliability Engineering; Springer: Berlin/Heidelberg, Germany, 2017.
18. Márquez, F.P.G.; Tobias, A.M.; Pérez, J.M.P.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renew. Energy* **2012**, *46*, 169–178, doi:10.1016/j.renene.2012.03.003. [CrossRef]
19. Ranganath, R.; Tang, L.; Charlin, L.; Blei, D.M. Deep Exponential Families. *arXiv* **2014**, arxiv:1411.2581.
20. Ranganath, R.; Perotte, A.; Elhadad, N.; Blei, D. Deep Survival Analysis. *arXiv* **2016**, arXiv:1608.02158.
21. Ali, J.B.; Chebel-Morello, B.; Saidi, L.; Malinowski, S.; Fnaiech, F. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech. Syst. Signal Process.* **2015**, *56–57*, 150–172, doi:10.1016/j.ymsp.2014.10.014. [CrossRef]
22. Mazhar, M.; Kara, S.; Kaebnick, H. Remaining life estimation of used components in consumer products: Life cycle data analysis by Weibull and artificial neural networks. *J. Op. Manag.* **2007**, *25*, 1184–1193, doi:10.1016/j.jom.2007.01.021. [CrossRef]
23. Yang, Z.; Chen, C.; Wang, J.; Li, G. Reliability Assessment of CNC Machining Center Based on Weibull Neural Network. *Math. Probl. Eng.* **2015**, *2015*, 8. [CrossRef]
24. Aggarwal, K.; Atan, O.; Farahat, A.K.; Zhang, C.; Ristovski, K.; Gupta, C. Two Birds with One Network: Unifying Failure Event Prediction and Time-to-failure Modeling. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1308–1317.

25. Bach-Andersen, M.; Rømer-Odgaard, B.; Winther, O. Flexible non-linear predictive models for large-scale wind turbine diagnostics. *Wind Energy* **2016**, *20*, 753–764, doi:10.1002/we.2057. [[CrossRef](#)]
26. Kleinbaum, D.; Klein, M. *Survival Analysis: A Self-Learning Text*, 3rd ed.; Statistics for Biology and Health; Springer: New York, NY, USA, 2011.
27. Patti, S.; Biganzoli, E.; Boracchi, P. Review of the Maximum Likelihood Functions for Right Censored Data. A New Elementary Derivation. In *COBRA Preprint Series*; bepress: Berkeley, CA, USA, 2007.
28. Kaplan, E.L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481, doi:10.1080/01621459.1958.10501452. [[CrossRef](#)]
29. Herp, J.; Ramezani, M.H. State transition of wind turbines based on empirical inference on model residuals. In Proceedings of the 2015 Conference on Research in Adaptive and Convergent Systems, RACS 2015, Prague, Czech Republic, 9–12 October 2015; pp. 32–37, doi:10.1145/2811411.2811509. [[CrossRef](#)]
30. Herp, J.; Ramezani, M.H.; Nadimi, E. Dependency in State Transitions of Wind Turbines-Inference on model Residuals for State Abstractions. *IEEE Trans. Ind. Electron.* **2017**, doi:10.1109/TIE.2017.2674580. [[CrossRef](#)]
31. Prescott Adams, R.; MacKay, D.J.C. Bayesian Online Changepoint Detection. *arXiv* **2007**, arXiv:0710.3742.
32. Fisher, R.A. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1922**, *222*, 309–368, doi:10.1098/rsta.1922.0009. [[CrossRef](#)]
33. Hipp, C. Sufficient Statistics and Exponential Families. *Ann. Stat.* **1974**, *2*, 1283–1292, doi:10.1214/aos/1176342879. [[CrossRef](#)]
34. Fisher, R.A. Applications of “Student’s” distribution. *Metron* **1925**, *5*, 90–104.
35. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2006. Available online: <http://www.deeplearningbook.org> (accessed on 12 December 2019)



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).