Estimating the Remaining Power Generation of Wind Turbines—An Exploratory Study for Main Bearing Failures

Benedikt Wiese [1,*]**, Niels L. Pedersen** [2]**, Esmaeil S. Nadimi** [1] **and Jürgen Herp** [1,*]

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

[2] Diagnostics, Siemens Gamesa Renewable Energy, 7330 Brande, Denmark; nielslovmand@gmail.com

[*] Correspondence: benedikt-wiese@hotmail.de (B.W.); herp@mmmi.sdu.dk (J.H.)

Abstract: Condition monitoring for wind turbines is tailored to predict failure and aid in making better operation and maintenance (O&M) decisions. Typically the condition monitoring approaches are concerned with predicting the remaining useful lifetime (RUL) of assets or a component. As the time-based measures can be rendered absolute when changing the operational set-point of a wind turbine, we propose an alternative in a power-based condition monitoring framework for wind turbines, i.e., the remaining power generation (RPG) before a main bearing failure. The proposed model utilizes historic wind turbine data, from both run-to-failure and non run-to-failure turbines. Comprised of a recurrent neural network with gated recurrent units, the model is constructed around a censored and uncensored data-based cost function. We infer a Weibull distribution over the RPG, which gives an operator a measure of how certain any given prediction is. As part of the model evaluation, we present the hyper-parameter selection, as well as modeling error in detail, including an analysis of the driving features. During the application on wind turbine main bearing failures, we achieve prediction in the magnitude of 1 to 2 GWh before the failure. When converting to RUL this corresponds to predicting the failure, on average, 81 days beforehand, which is comparable to the state-of-the-art's 94 days predictive horizon in a similar feature space.

Keywords: wind turbines; remaining power generation; main bearing; neural network

1. Introduction

Wind energy remains the leader in renewable energy sources with expected continued growth [1]. In order to maintain and monitor the ever-increasing fleet of wind turbines, health prognostics and condition monitoring (CM) techniques are employed, covering abnormality detection, failure mode tracking, and prediction [2–7] of failures in rotating parts, in particular, the main bearing. A summary of common CM models may be found in Márquez et al. [8]. With regard to prediction, recent works have shown predictive capabilities beyond 90 days with reasonable accuracy [3]. These approaches have one thing in common: all perform predictions with respect to the remaining useful lifetime (RUL) and life cycles of an asset or component. Although these aid in operation and maintenance (O&M) efforts, they are not directly related to the physical nature of a wind turbine, i.e., power production. In terms of O&M scheduling we will consider the following thought experiment: A predictive model, such as Herp et al. [3,9], Teng et al. [5], or Si et al. [10] predicts a wind turbine to fail in the RUL = 36 days, within a confidence bound. The operator is now faced with the decision when to stop operations and when to plan maintenance tasks. This in itself is a highly complex optimization problem. However, the RUL as such is not related to any ambient or operating conditions, thus, without further information,

the best course of action would be to shut down the turbine and thereby extend the turbine's RUL to infinity, or until the socket's or tower's expected lifetime. In this scenario, the operator would lose all revenue generated by that particular turbine.

To circumvent this dilemma and provide a physical measure of the remaining operation of a failing wind turbine, we propose an approach with a focus on the remaining power generation (RPG). We draw inspiration from Teng et al. [5] and Herp et al. [3],both of which employ a neural network (NN) for learning dependencies between samples. While Teng et al. extrapolate and compare the predicted values to a threshold, Herp et al. train towards a distribution and use a recurrent neural network (RNN). We adapt the RUL prediction in Herp et al. and formulate an equivalent model for the RPG. In Section 2 we provide an overview of the data included in this study, including a preliminary investigation of the input feature. In Section 2.1.2 we highlight the difference between RUL and RPG. The proposed model is explained in Section 3. Covering the fundamental assumptions of the model, a justification of the chosen hyper-parameters is given. In addition to the loss off the neural network, we also present an additional performance measure, an aggregated error over the predictive distribution. In order to show the feasibility of the model, we consider a case study of main bearing failure (Section 4). Here we show the predictive capabilities of the model and compare the model outcome to the state-of-the-art prediction in RUL estimations, by mapping between RPG and RUL. As of the time of writing, we are not aware of any literature addressing the remaining power generation instead of remaining useful lifetime, and thus do not provide a comparative study. Finally, we conclude our work in Section 5.

2. Remaining Power Generation for Main Bearing Failures

This project is concerned with the RPG of main bearing failures and limited to turbines of the same class, i.e., three balded units composed of the same parts and rated at the same power. As such, the amount of data available for each turbine is still beyond direct processing, thus, in order to talk about the aim of this study, we start with an assessment of the availability of data sources and their relevance. Alongside we provide the notation used throughout this work.

2.1. Data and Notation

All the data used in this study come from the same class of turbines and is provided by courtesy of Siemens Gamesa Renewable Energy. The data acquired stem from three different databases: (i) meta-data containing a damage indicator for the main bearing. It is obtained by integrating over a frequency range of a fast Fourier transform of bearing vibration measurements to obtain a measure of the noise floor. The damage indicator will be treated as one of the inputs to the proposed model. These data points are event-based and contain initial detection of damage. (ii) SCADA data, sampled at 10 min averages. (iii) Wind turbine events, e.g., warning and error messages. This database is used to distinguish between run-to-failure wind turbines and non run-to-failure turbines. Here we referred to run-to-failure wind turbines, as wind turbines that are operated until failure, while turbines that are stopped by the operator before failure are referred to as non run-to-failure wind turbines. Before data preparation, 112 wind turbines, undergoing a main bearing failure, have been acquired for this study—31 of which are considered run-to-failure.

The SCADA data are chosen based on first principal considerations of the energy diffusion, selecting the same features as Bach-Andersen et al. [11], while the damage indicator is chosen to be the same as used in the study of Herp et al. [3]. After cropping the time-series at the initial damage detection and the stopping of operation at the fault, the two data sources are merged, such that $\mathbf{x}_t \in \mathbb{R}^m$ is a feature vector containing m features at time t. The merged time-series are re-sampled to daily timestamps. Table 1 shows the initial range and units of the desired features.

Table 1. SCADA and damage indicator description.

Features	Range	Unit
Main bearing temperature	$[0, 1]$	a.u. *
Nacelle temperature	$[-6, 48]$	°C
Gear oil temperature	$[0, 1]$	a.u. *
Generator revolutions	$[0.6, 1469.0]$	RPM
Power	$[0, 1]$	a.u. *
Wind speed	$[0.9, 19.0]$	m/s
Ambient temperature	$[-19, 31]$	°C
Damage indicator	$[0.0004, 0.0200]$	a.u. *

* Data is presented in arbitrary units (a.u.).

2.1.1. Main Bearing Features

Although NNs are capable of learning dependencies between input features, we present an additional temperature based feature. This feature will provide a physical interpretation of what drives the prediction in the proposed model. A simple correlation study shows, as one would expect, that the main bearing temperature is highly correlated with other temperature measures. In order to investigate trends and behavior of the main bearing, independent of other temperature sources, we create time-series of the difference between the main bearing temperature and the nacelle, ambient, and gear oil temperature, respectively. The subtraction of the ambient temperature is especially interesting as it can influence the main bearing temperature, while vice verse seems unlikely from an energy diffusion perspective—a high main bearing temperature does not necessarily indicate a broken bearing if the ambient temperature is the driving factor. Figure 1 shows the time-series towards failure, and compares them to the individual temperature measurements (dashed lines). The newly created features are indicated by solid lines. When focusing on the main bearing time-series, no changes can be seen towards the failure in the general trend. This can be explained by seasonal effects, since the ambient temperature falls and skews the main bearing temperature by lowering the temperature of the component. However, a different trend is shown when taking new features into consideration, in particular, the ones based on the ambient and nacelle temperature. These time-series show an increase in temperature towards the failure and indicate the progression of the fault. As the nacelle temperature is similarly driven by the ambient temperature, we include only the difference between the main bearing and ambient temperature as a new input feature for the proposed model.
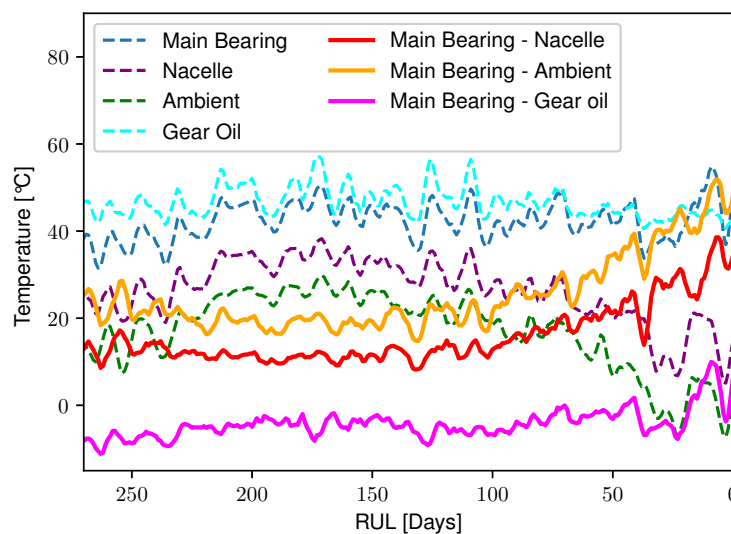
Figure 1. Modified main bearing features and their behavior.

2.1.2. Target Feature

Since related studies [2–5,10,12] in health prognostics are concerned with RUL or the remaining life-cycles of a component, we take the time to highlight the significance of the target feature presented here, namely the remaining generated power. Figure 2 shows both the RUL and RPG as a function of time, obtained from the SCADA data. Besides the physical relation to wind turbine operations, RPG and RUL are especially distinguishable in their time-series properties, when compared across other wind turbines. When comparing the RUL of all turbines (Figure 2a), it does not come as a surprise that the rate of which the RUL is decreased is the same for all turbines—since time is linear, only a linear relation can be expected. This makes the RUL a less favorable target feature for a predictive model, since any two turbines with equal RUL can have widely different RPGs, which can be seen by comparing Figuer 2b,c. When aligning the RPG with the end of each time-series, it can be seen that the rate by which the RPG is depleted towards failure, changes from one wind turbine to another. This can be explained by seasonal and local weather patterns characterizing each site, which result in different ambient conditions of the wind turbine. These local perturbations consequently lead to more variability in the target feature space.

As we later on will distinguish between run-to-failure and non run-to-failure wind turbines, run-to-failure wind turbines are indicated by dashed lines. In comparison to non run-to-failure wind turbines, they are spanning over a wide range of trends and absolute RPG, allowing the proposed model to cover a variety of different scenarios.
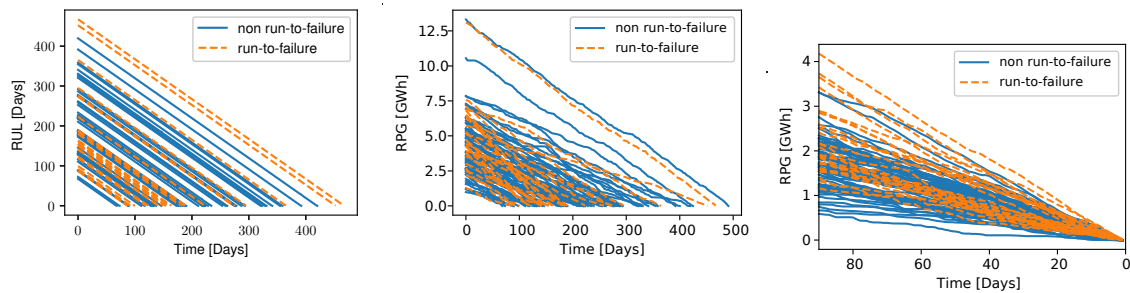
(**a**) Remaining useful lifetime (RUL) for all 112 turbines. (**b**) Remaining power generation (RPG) for all 112 turbines. (**c**) RPG for all 112 turbines aligned.

Figure 2. Target feature visualization.

3. Predictive Modeling of Remaining Power Generation

We facilitate the prediction of the RPG in the setting of survival analysis [13], with censoring as a key concept. In short, the concept of censoring is that the time-series ends before the desired event. More specifically, we distinguish between right-censored data, such that the RPG is known to be above some time t, and non-censored data, when the RPG equals the current time. The hazard (rate of change of RPG towards failure) expressed at the current time, can be written as a hazard function $\eta(t)$ and cumulative hazard function $H(t)$. For a positive random variable RPG, it follows [13] for the cumulative distribution:

$$F(\text{RPG} \leq t) = 1 - \exp\left(-\int_0^t \eta(\omega)d\omega\right) = 1 - e^{-H(t)}, \tag{1}$$

and for the probability density function:

$$\begin{aligned}\mathbb{P}(t) =& \frac{\partial}{\partial t}\mathbb{P}(\text{RPG} \leq t) = \frac{\partial}{\partial t}(1 - e^{-H(t)}) \\ =& e^{-H(t)}\eta(t).\end{aligned} \tag{2}$$

Furthermore, for each cumulative distribution function, $\mathbb{P}(\text{RPG} \leq t)$ there exists $H(t)$ such that $H(t) = -\log(1 - \mathbb{P}(\text{RPG} \leq t))$. Large values of the hazard function are an indication of a high risk of failure, i.e., vanishing RPG. In the remainder of this study we are concerned with the right tail of the RPG distribution:

$$\mathbb{P}(t < \text{RPG}) = 1 - \mathbb{P}(\text{RPG} \leq t) = e^{-H(t)}. \tag{3}$$

Combining both the censored and non-censored data, the likelihood function for the RPG in terms of non run-to-failure (C censored) and run-to-failure (F non-censored) wind turbines can be written as:

$$\mathcal{L} = \prod_{f \in F} \mathbb{P}(t_f) \prod_{c \in C} \mathbb{P}(\text{RPG} > t_c). \tag{4}$$

Following Patti et al. [14], and introducing Δ as a health indicator, such that $\Delta = 0$ indicates an observed failure, and $\Delta = 1$ otherwise, the likelihood reads:

$$\mathcal{L} = \mathbb{P}(t)^{\Delta} \mathbb{P}(\text{RPG} > t)^{1-\Delta} = \eta(t)^{\Delta} e^{-H(t)}. \tag{5}$$

It follows that the log likelihood is given by:

$$\log \mathcal{L} = \Delta \log \eta(t) - H(t). \tag{6}$$

Equation (6) becomes the basis for the cost function of the RNN proposed later on. Here η and H can be obtained through the Equations (1) and (2) when choosing an appropriate distribution. In this study, following common practice in survival analysis [13] and its use in RUL estimation for main bearing failure in wind turbines [3], we chose the Weibull distribution for its versatile properties:

$$f(\text{RPG}) = \frac{\alpha}{\beta} \left(\frac{\text{RPG}}{\beta} \right)^{\alpha - 1} \exp\left(-\left(\frac{\text{RPG}}{\beta} \right)^{\alpha} \right) \tag{7}$$

$$F(\text{RPG}) = 1 - \exp\left(-\left(\frac{\text{RPG}}{\beta} \right)^{\alpha} \right), \tag{8}$$

where $\alpha \in]0, \infty)$ is the distributions shape parameter, and $\beta \in]0, \infty)$ its scale.

We will focus on RNNs with gated recurrent units (GRUs), letting the Weibull distribution be parameterized by $\boldsymbol{\theta} = [\alpha, \beta]^{\top}$. Omitting mathematical details, for which the interested reader is referred to textbooks, such as Goodfellow et al. [15], the parametrization is obtained by the RNN mapping $m(\cdot)$ of the current data $\mathbf{x}_t$, NN weights $\boldsymbol{\theta}$, and former layer output o_{t-1}:

$$\boldsymbol{\theta} = m(\mathbf{x}_t, o_{t-1}, \boldsymbol{\omega}). \tag{9}$$

Combining Equations (6) with the hazard and cumulative hazard function of Equations (7) and (8) with the RNN, the problem to solve is the following:

$$\arg\max_{\boldsymbol{\omega}} \log \mathcal{L}(\boldsymbol{\omega}, \text{RPG}, \Delta, \mathbf{x}_{[1,t]}) = \sum_{i=1}^{t} \left(\Delta_i \left[\alpha_i \log\left(\frac{\text{RPG}}{\beta_i} \right) + \log(\alpha_i) - \log(\text{RPG}) \right] - \left(\frac{\text{RPG}}{\beta_i} \right)^{\alpha_i} \right). \tag{10}$$

Model Optimization

The training and validation data for the proposed model are split 0.75/0.25 into sets of full turbine time-series. Besides the regular techniques for model evaluation through the loss function of the RNN model, we propose an error measure of the prediction itself with respect to the true remaining power generation, $\hat{\text{RPG}}$. As the estimated RPG comes in terms of a parameterization of a probability function, the error will be a modification of the $|\cdot|_{\ell=1}$ norm between RPG and $\hat{\text{RPG}}$, such that:

$$
\begin{aligned}
E(t \mid \boldsymbol{\theta}) = & E(t \mid \alpha, \beta) \\
= & \int_0^{\infty} |\text{RPG} - \text{R}\hat{\text{P}}\text{G}|_1 \mathbb{P}(\text{RPG} \mid \alpha, \beta) d\text{RPG}.
\end{aligned}
\tag{11}
$$

Based on the validation loss, we optimized the topology of the RNN by exploring the depth and width as listed in Table 2. Based on this investigation, the topology was fixed to two layers of 75 GRU nodes. The remaining hyper-parameters were set in accordance with general practice [16]: The dropout rate after each GRU layer was 20%, the batch size was 128 samples, and the learning rate was initialized at 0.0001. Where the dropout rate was selected by a discreet uniform search on $[0, 40]\%$ dropout rates, see Table 3 for the loss values of the RNN.

Table 2. Topology optimization in terms of loss.

		Depth			
		1	**2**	**3**	**4**
	12	0.078	0.081	0.080	0.72
	25	0.077	0.071	0.073	0.070
Width	50	0.074	0.072	0.069	0.075
	75	0.070	0.068	0.068	0.071
	100	0.072	0.070	0.069	0.071

Table 3. Dropout optimization in terms of loss for the optimal network of Table 2.

		Avg. E (RPG)	**Epoch at Min E (RPG)**
	0%	1.9 GWh	131
	10%	0.9 GWh	145
Dropout Rate	20%	0.6 GWh	242
	30%	0.9 GWh	238
	40%	1.1 GWh	226

In addition, we investigated the dependency of the sequence slicing for the input of the RNN. Changing the sequence length from 2 to 30 days, we evaluated the optimal sequence length by minimizing the error presented in Equation (11). Figure 3 shows the predictive error as a function of the true RPG. Despite all sequence lengths performing with a similar error, it can be noticed that a sequence length of 10 days yielded the best results in terms of predictive error and variability. Sequence slicing beyond 10 days did not improve the predictive error and became increasingly computationally expensive when averaging over all turbines. The proposed model thus captured up to weekly dynamics in each sequence.
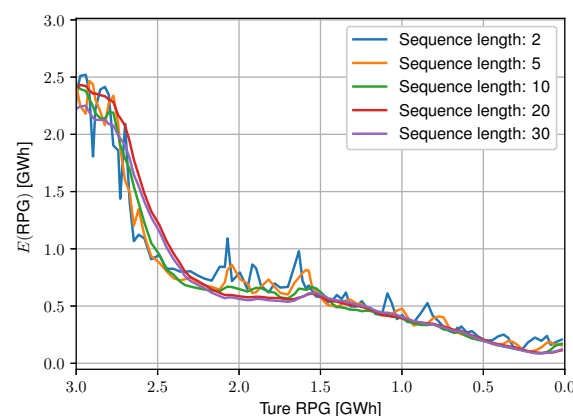

Figure 3. Sequence length optimization in terms of model error for one wind turbine.

4. Bearing Failure Study

The final model for this study is illustrated in Figure 4. After training, the model was applied to the remaining run-to-failure wind turbines. Figure 5 shows a common RPG prediction pattern in Figure 5a and a fast developing fault in Figure 5b. The first mentioned shows prediction in agreement with the expected RPG up to 60 days before the failure. Considering only the damage indicator, it is evident that predictions beyond 60 days for this particular wind turbine were difficult, since there was no evidence for robust estimations. As more data became available, and the closer the wind turbine was to failure, the predictions became more and more accurate. Consequently, the confidence bound around the prediction became narrow, which is indicated by the shaded area. In comparison to point-wise estimates, this shows why predicting a distribution is an advantage; it provides an operator with a measure of how reliable the predictions are.
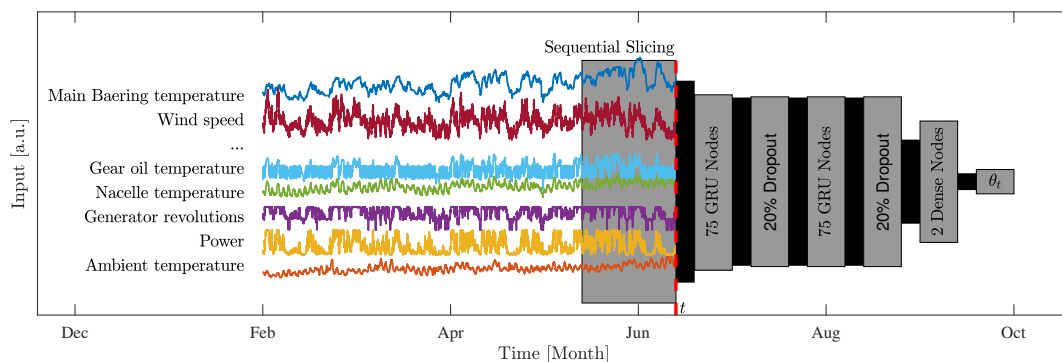

Figure 4. Model illustration.

While the behavior shown in Figure 5a can be observed for most turbines, Figure 5b serves as an example of how the model performs when encountering unforeseen events. The predictions of the RPG were off by 2 to 3 GWh for most of the duration of the failure process, only converging in the last couple of days. This behavior arose from a lack of fast-developing faults in the training data set, biasing the model towards more common failure patterns. As the damage indicator did not pick up prior to 20 days before the failure, the model had no indication that allowed it to converge towards a prediction. This opens up for a discussion on what are the driving features behind the model. We will address this in the next section.
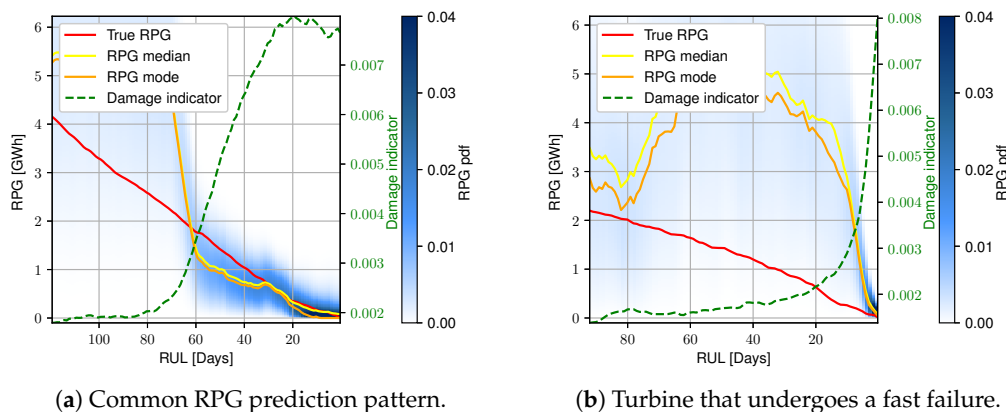

(**a**) Common RPG prediction pattern.　(**b**) Turbine that undergoes a fast failure.

Figure 5. RPG prediction.

In order to reduce good and bad predictions by chance, we investigated the ensemble performance of the proposed model by means of a five-fold cross-validation approach to see how the models performed on different test data. We illustrated this by considering one test turbine for all folds. Figure 6 shows the predictive error, where all folds started with different initial values, but they nevertheless

converged rapidly towards accurate predictions at the 2 GWh mark before failure. From here on, they showed identical behavior providing consistent and increasingly accurate results. When averaging over all test wind turbines and folds, predictions became accurate at 1.8 GWh (The averaged prediction for all folds are shown in Table 4). For now, this provides the upper bound of what the proposed model is capable of predicting.
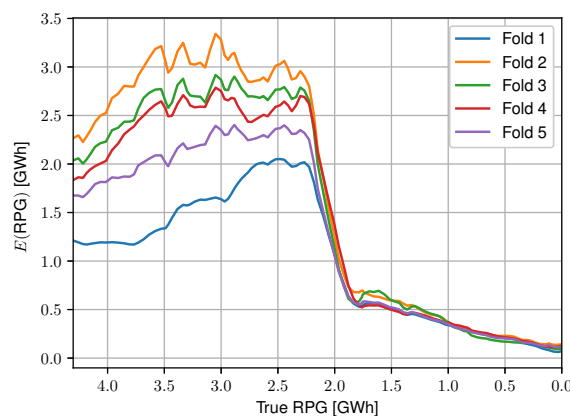

Figure 6. Predictive five-fold error for one wind turbine.

Table 4. Prediction horizon per fold.

		Avg. Prediction [GWh]	Variance [GWh]	min and max [GWh]
	1	1.9	0.4	[2.6, 0.5]
	2	1.6	0.3	[2.0, 0.6]
Fold	3	1.8	0.3	[2.2, 0.5]
	4	2.0	0.4	[2.6, 1.0]
	5	1.7	0.2	[2.1, 0.7]

4.1. Feature Importance

As mentioned earlier, when the proposed model did not have sufficient information from the input feature space, predictions could not be considered accurate (See confidence bound in Figure 5). We again chose the turbine in Figure 5a to be representative of common fault behavior patterns. Further, we constrained the time scale under investigation to 120 days before the fault. Figure 7 shows relation between individual features and predictions. For illustrative purposes, the features were normalized individually, thus two equal values between features did not imply equal measurement in the feature space. However, the true RPG and mode of the predicted RPG were scaled to ensure direct comparison and maintain their relations as shown in Figure 5a.

The impact on the predictive capabilities of the proposed models from temperature measurements and the damage indicator is shown in Figure 7a. For the first 50 days neither the main bearing temperature, the difference between the main bearing temperature and the ambient temperature, nor the damage indicate suggested any abnormality. On this scale, the predictions of the proposed model were misaligned with the true RPG. After 50 days, features rapidly picked up, most notably the temperature based features to begin with. In this time period, the predicted RPG was getting in alignment with the true RPG. After further 20 days, the initial increase in the features was replaced by variations around some asymptote. As the predicted RPG remained to follow the true RPG, other features must drive the predictive behavior. Next, we took the generator revolution into account. Figure 7b contains the main bearing temperature, generator revolution, and their difference. When passing the 60-day mark before failure, the generator revolution had yet to show significant changes, and it was first in the remaining 20 days that changes could be observed. This change, though, did not show any trend. This changed when comparing the generator revolution with the main bearing

temperature, from 40 days before the fault, and onwards, a steady increase of the combined feature could be observed. This is believed to drive predictions towards the end of the wind turbines RPG.

Summarizing the trends of Figure 7, the damage indicator was driving most of the prediction at the beginning of the fault, while temperature measurements aided in forcing the predictions to more accurate predictions. 20 to 30 days before the wind turbine failure, a combination of temperature and rotational features (i.e., generator revolution) seemed to drive predictions. In the case of fast-developing faults, such as shown in Figure 5b, none of the presented trends could be found. Thus, the model did not indicate that the wind turbine should undergo any state changes.

(**a**) Impact of temperature and damage indicator.   (**b**) Impact including generator revolution.
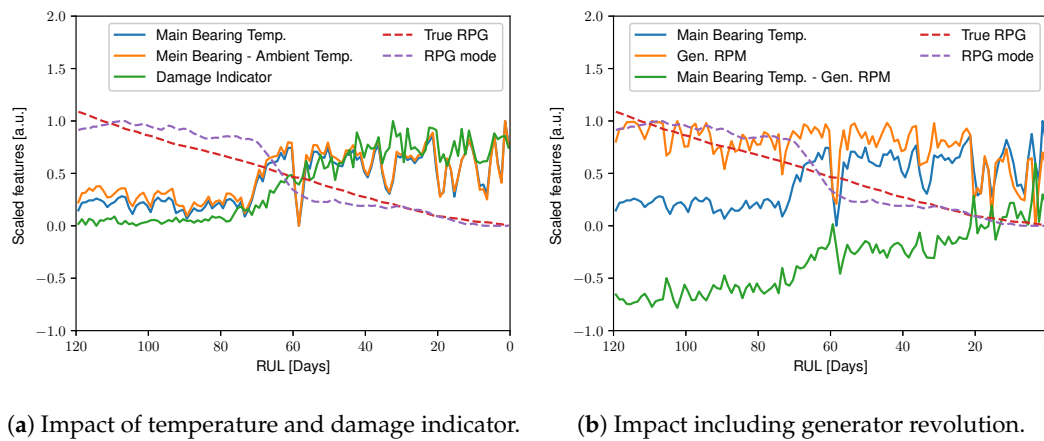
Figure 7. Relations between feature space and predictions for the same turbine as in Figure 5a.

4.2. Ties to Remaining Useful Lifetime

As of the time of writing, we are not aware of any work addressing the prediction of RPG or similar. To compare the model to other approaches, we dedicate this section to explore the mapping from RPG to RUL and comparing the results to RUL prediction when employing the model by Herp et al. [3]. The mapping to RUL was done individually for each wind turbine. This was achieved by including how much power a wind turbine produced since the beginning of the time-series. Alternatively, one could fix an interval, e.g., 90 days before the current timestamp. Depending on the length of the interval, seasonal effects could be covered. The rate of power remaining could then be defined over a time Δt:

$$r = \frac{\Delta t}{P} \left[\frac{\text{Days}}{\text{GWh}} \right]. \tag{12}$$

It follows that the mapping from RPG to RUL is given by:

$$\text{RUL} = r\text{RPG}. \tag{13}$$

Figure 8 shows the mapped RUL from the RPG displayed in Figure 5a. The RUL median and mode, mapped from the RPG median and mode, showed accurate prediction around 60 days before the failure, slightly underestimating the RUL compared to the true RUL. Mapping RPG to RUL for other turbines that were representative for the common fault behavior, systematically underestimate the RUL, averaging 81 days, over the internal 50 to 100 days. Comparing to Herp et al. [3] average predictions of 94 days, this approach fell short of 13 days.
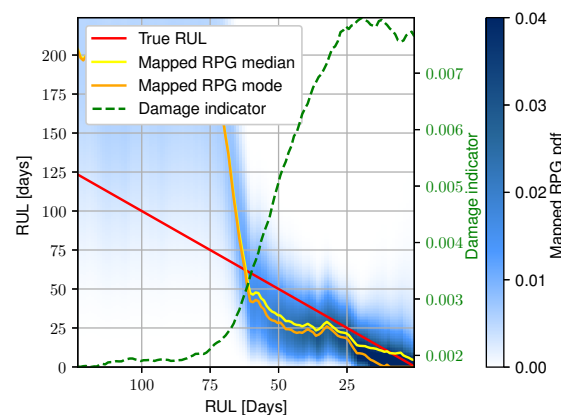
Figure 8. Mapping between RPG and RUL for the same turbine as in Figure 5a.

5. Conclusions

The purpose of this work was to create and motivate a model capable of predicting the remaining power generation (RPG) with application in wind turbine bearing failure. Furthermore, we have pointed out the similarities and differences between the RUL and the RPG. We have provided a description of the underlying model and how to tune its hyper-parameters. In the implementation of main bearing failure, we archived predictive horizons in the order of 1 to 2 GWh. In a five-fold cross-validation over multiple turbines, we showed that predictions become accurate at an upper bound of 1.8 GWh.

When converting RPG to RUL, we achieved predictive horizons of 50 to 100 days, averaging 81 days overall wind turbines. For a preliminary conversion, we consider this a satisfying result as the state-of-the-art [3] reports 94 days, on average.

During the course of this work, it becomes apparent that future research can follow up on the feature space presented, filtering or identifying more driving features that eventually will foster better predictions. Besides, as the proposed model performs differently on individual turbines with different failure behavior, one could consider following Aggarwal et al. [17] in allowing classification into different failure modes as en additional output of the RNN.

References

1. REN21 Secretariat. *Renewables 2019 Global Status Report*; Technical Report; REN21: Pris, France, 2019.
2. Deutsch, J.; He, D. Using Deep Learning-Based Approach to Predict Remaining Useful Life of Rotating Components. *IEEE Trans. Syst. Man, Cybern. Syst.* **2018**, *48*, 11–20. [CrossRef]
3. 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. [CrossRef]
4. Singleton, R.K.; Strangas, E.G.; Aviyente, S. The Use of Bearing Currents and Vibrations in Lifetime Estimation of Bearings. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1301–1309. [CrossRef]

5. Teng, W.; Zhang, X.; Liu, Y.; Kusiak, A.; Ma, Z. Prognosis of the Remaining Useful Life of Bearings in a Wind Turbine Gearbox. *Energies* **2017**, *10*, 32. [CrossRef]

6. Kusiak, A.; Li, W. The prediction and diagnosis of wind turbine faults. *Renew. Energy* **2011**, *36*, 16–23. [CrossRef]

7. Kusiak, A.; Verma, A. Analyzing bearing faults in wind turbines: A data-mining approach. *Renew. Energy* **2012**, *48*, 110–116. [CrossRef]

8. 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. [CrossRef]

9. 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**. [CrossRef]

10. 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.

11. Bach-Andersen, M.; Rømer-Odgaard, B.; Winther, O. Flexible non-linear predictive models for large-scale wind turbine diagnostics. *Wind Energy* **2016**. [CrossRef]

12. You, G.W.; Park, S.; Oh, D. Diagnosis of Electric Vehicle Batteries Using Recurrent Neural Networks. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4885–4893. [CrossRef]

13. Kleinbaum, D.; Klein, M. *Survival Analysis: A Self-Learning Text*, 3rd ed.; Statistics for Biology and Health; Springer: New York, NY, USA, 2011.

14. Patti, S.; Biganzoli, E.; Boracchi, P. Review of the Maximum Likelihood Functions for Right Censored Data. A New Elementary Derivation. In *Collection of Biostatistics Research Archive, COBRA Preprint*; Bepress: Berkeley, CA, USA, 2007.

15. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 30 June 2020).

16. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

17. 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.