Hybrid physics-informed neural networks for main bearing fatigue prognosis with visual grease inspection

Yigit A. Yucesan*, Felipe A.C. Viana

Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL 32816-8030, USA

A R T I C L E   I N F O

Article history:
Received 17 August 2020
Received in revised form 27 October 2020
Accepted 11 December 2020

Keywords:
Physics-informed neural networks
Scientific machine learning
Uncertainty quantification

A B S T R A C T

Field failures of wind turbine main bearings yield to undesired downtime and significant maintenance costs. Fatigue failure is a dominant mode for legacy turbines, which can be expressed with physics-informed models to some extent. However, these models often inherent large uncertainties due to unknown lubricant degradation mechanism. Therefore, periodical assessment of the grease state plays a crucial role in calibration of bearing fatigue models. As opposed to detailed laboratory analysis, grease visual inspection can lead to large uncertainties in characterization of grease condition (although visual inspection can be cost and time effective). In this paper, we introduce a hybrid model for main bearing fatigue damage accumulation and calibrated using only visual grease inspections. In our hybrid model, bearing fatigue damage portion consists of known physics formulations, and unknown grease degradation is represented with deep neural networks. In addition, we introduce a custom tailored classifier that enables the model to map from damage scale to visual rankings. Results showed that the bearing fatigue prognosis model can be successfully calibrated, even with limited and noisy field observations. Moreover, the model can help optimizing park reliability by suggesting turbine-specific regreasing intervals. The source codes and links to the data can be found in the following GitHub repository https://github.com/PML-UCF/pinn_wind_bearing.

1. Introduction

Unscheduled main bearing failure is a major operation and maintenance issue, as it can cost hundreds of thousands of dollars. Sethuraman et al. (2015) support this statement and illustrate main bearing failures rates observed across 50 different wind farms. This information can be rearranged to depict fleet unreliability, as in Fig. 1. Common failure modes for wind turbine bearings include, but are not limited to, micro-pitting, white etching crack, electrical erosion, and contact fatigue or spalling (Hornemann and Crowther, 2013). Nevertheless, for a large portion of the installed fleet, time since commissioning has already passed the 10 year mark; and therefore, bearing fatigue has become a concern. As early failures are mitigated over time (e.g., through retrofitting), fleet moves towards fatigue failure (nearing design curves).

Physics-based models for field prediction of bearing fatigue (as opposed to design) have been extensively studied. Watanabe and Uchida (2015) utilized standard bearing life formulations given in ISO 281 to build a prognosis model for wind turbine rear bearing. The authors tested their model on an actual wind farm located in Japan. The physics-based prognosis model proposed by the authors predicted approximately 12 years of life for a sample turbine, while the actual field failure occurred in 12.7 years. They also showed ways to make use of their model for life extension purposes through curtailment operation. We should note that the referred paper considers wind turbines with four-point mounting setting, where two main bearings exist (front and rear), and focus on rear main bearings. For comparison, while in four-point mounting setting two bearings share most of the load, in three-point mounting setting single main bearing carry most incoming load. In another study on main bearing, Walker and Coble (2018) combined the adaptive sampling and order tracking approaches to use vibration signals for anomaly detection. Authors showcased their method on three wind turbine main bearings, and they detected an outer race fault in one of the machines, and they validated this failure with post-mortem examinations of the asset.

The challenges in prognosis and remaining useful life estimation have motivated researchers to look at machine learning. Shao et al. (2018) proposed continuous deep belief networks with locally linear embedding for bearing fault detection. In this two-step approach, they first determined a comprehensive feature

* Corresponding author.

E-mail addresses: yucesan@knights.ucf.edu (Y.A. Yucesan), pml@ucf.github.io, viana@ucf.edu (F.A.C. Viana).

https://doi.org/10.1016/j.compind.2020.103386
0166-3615/© 2020 Elsevier B.V. All rights reserved.
index with locally linear embedding to gauge the performance degradation of the bearing. The continuous deep belief network consists of multiple continuous restricted Boltzmann machines built to model temporal vibration output. Authors showcase the advantages of their proposed approach against traditional methods through experimental data. Zhao et al. (2019) also aimed to tackle the fault diagnosis problem for bearings with machine learning. The authors utilize deep convolutional neural networks to classify the bearing fault type using vibration signals. The novelty of the proposed approach is that its ability to automatically recognize the fault type without artificial processing of fault characteristics, even though there might exist interference between different faults. With an analysis of a planetary gearbox data, the overall classification accuracy of the model is evaluated as 98.3%. Zhang et al. (2019) proposed to model bearing performance degradation with long short-term memory recurrent neural networks. They also make use of bearing vibration signals as degradation indicator for the component. Authors introduced a novel degradation indicator as an input for their model and optimize the parameters of the data-driven time-dependent network through particle swarm optimization method. The resultant model is able to comprehensively assess the state of the bearing degradation, that is shown with experimental results. The common point of these contributions is that they use vibration signals, which are indicative for short-term fault detection but not necessarily for long-term degradation predictions for wind turbine bearings.

Previously discussed contributions (Shao et al., 2018; Zhao et al., 2019; Zhang et al., 2019) mainly use vibration signals to train their deep learning models. Unfortunately, for legacy turbine platforms (mostly under 2 MW), vibration data is not always available and modeling is constrained by the supervisory control and data acquisition (SCADA) system along with maintenance and operation records (e.g., grease inspection and regreasing intervals). Under such circumstances, predictive models used for prognosis are limited by the quality and completeness of such data. For example, while SCADA systems track operation over long periods of time; only sensor statistics (not the raw data) are recorded in 10 min intervals. Grease quality is poorly tracked, with visual inspection reported occasionally. Hence, a rigorous framework for combining such sources of information is strongly needed to reduce uncertainty in remaining useful life prediction. In this paper, we propose a hybrid approach for modeling bearing fatigue using physics-informed kernels within neural networks. This hybrid model uses only turbine-specific SCADA data as inputs and predicts turbine-specific bearing and grease damage over time. Moreover, grease damage is modeled as a deep neural network and calibrated using only noisy visual inspection information. The visual inspection data we use is limited to ranks assigned by technicians as they perform routine maintenance in the wind turbines.

The use of physics-informed neural networks in science and engineering has received growing attention. For example, Raissi (2018) introduced a method where one can wield two separate neural networks to approximate the solution of partial differential equations for a fluid mechanics problem. The author used one network as a prior to the solution, and the other to adjust the steady state of the solution. Alternatively, researchers have also explored hybrid approaches where physics-informed kernels are used alongside neural networks. This approach has been showing promising results for cumulative damage modeling (Yucesan and Viana, 2020; Dourado and Viana, 2020; Nascimento and Viana, 2019).

In this paper, we propose customizing a recurrent neural network to perform bearing fatigue and grease damage accumulation. The bearing fatigue model uses the physics formulation found in the ISO 281 standards (ISO, 2007). We build an artificial wind farm damage history for both bearing fatigue and grease using physics-models and manufacturer catalogs to use as a ground true of our case study. Similarly to Yucesan and Viana (2020), we assume that no reliable physics-informed model for grease degradation is available. Therefore, we use a data-driven kernel within the recurrent neural network to model grease damage accumulation. However, as opposed to Yucesan and Viana (2020), the grease inspection is not limited to highly accurate grease damage indicators (such as those obtained in laboratory analysis). Here, we study the case in which grease condition is monitored through visual inspection. As a consequence, a significant advantage of our proposed approach is that the grease damage accumulation model is trained directly with noisy visual grease inspection data. In order to accomplish this task, we use a hybrid physics-informed neural network and introduce a novel ordinal classifier that we call discrete ordinal classifier (DorC).

The remaining of the paper is organized as follows. Section 2 gives an overview on main bearing fatigue damage accumulation models and challenges associated with grease degradation. Section 3 details our proposed physics-informed neural networks model. Section 4 describes the case study with regards to the machine specifications, wind park, operational data, and inspection campaigns. Section 5 presents and discusses the numerical results. Finally, Section 6 concludes the paper by summarizing significant remarks, and providing insight on potential future studies. Appendices are provided with discussions on neural networks hyperparameter initialization and benchmark studies.

2. Background: main bearing fatigue modeling and challenges with grease degradation

2.1. Main bearing fatigue modeling

We establish our main bearing fatigue model based on standardized formulas and manufacturer catalogs. Across this study, we consider a 1.5 MW rated power and 80 m hub height wind turbine with a spherical roller bearing in the three-point mounting setting (GE-Contributors, 2009). Given that bearings operate at dif-

---

Table 1: Bearing fatigue life adjustment factor (Skf-contributors, 2007).

<table>
<thead>
<tr>
<th>Reliability level (%)</th>
<th>90</th>
<th>95</th>
<th>97</th>
<th>98</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.00</td>
<td>0.62</td>
<td>0.44</td>
<td>0.33</td>
<td>0.21</td>
</tr>
</tbody>
</table>

---

1 The interested reader is referred to literature on white etching cracking (Luyckx, 2012; Stadler et al., 2014; Bruce et al., 2015). Within the wind energy industry, this is an emerging topic of research as it can lead to further understanding of bearing fatigue (although literature has not reached consensus yet).
different load levels and rotational speeds, we model fatigue damage, $a_{BRC}$, using (ISO, 2007; SKF-Contributors, 2007):

$$\frac{da_{BRC}}{dt} = \frac{1}{c_1 c_2(t)} \left( \frac{P(t)}{C} \right)^{10/3}, \quad P(t) = f_1(V_W(t)), \quad c_2(t) = f_2(P(t), \eta_c(t), v(t)),$$

(1)

$$\eta_c(t) = f_3(v(t), \eta_{GRS}(t)) \quad \text{and} \quad v(t) = f_4(T_{BRC}(t), a_{GRS}(t)),$$

where $c_1$ is a life modification factor based on reliability level (see Table 1); $c_2$ is an adjustment factor accounts for lubricant condition; $P$ is the equivalent dynamic bearing load; $C$ is the design load rating; $\eta_c$ is the grease contamination factor; $v$ is the grease viscosity; $V_W$ is the wind speed; $T_{BRC}$ is the bearing temperature; $a_{GRS}$ is an indicator of grease degradation; and $f_1$, $f_2$, $f_3$, $f_4$ are functions defining the models for different components of the bearing damage.

Independent variables wind speed and bearing temperature contributes to the variation of $P$ and $c_2$ over time. Fig. 2a shows how bearing loads are affected by wind speed (i.e., $f_1$ in Eq. (1)), as these results are obtained through high-fidelity multi-body physics analysis (Sethuraman et al., 2015). Fig. 2b shows how grease related parameters $\eta_c$, $v$, and $c_2$ are evaluated as a function of temperature, loads, and grease condition. The implication for our model is that bearing temperature affects viscosity $v$, and therefore, the viscosity ratio. In turn, the contamination factor $\eta_c$ varies with viscosity ratio. Finally, both viscosity and contamination factor influence the life modification factor $c_2$. In addition, both viscosity and contamination factor curves change as grease degrades (Fig. 2b).

2.2. Challenges with grease degradation

Lubricant condition impacts main bearing fatigue, as per adjustment factors in Eq. (1). Often, grease suppliers provide life curves for lubricants they sell (Klueter-Contributors, 2011). Unfortunately, multiple safety factors are used alongside these life curves, which increases conservatism (and applicability) associated with these models. Modeling lubricant degradation through physics is a major challenge and researchers have been trying to tackle it. Zhu et al. (2013) studied lubricant condition assessment based on different variables. Authors considered lubricant viscosity and dielectric constant outputs and used particle filtering for accurate useful life predictions. They also validated the method with experiments, showing that using only dielectric constant sensor output yields to a better performing prognosis model for lubricant.

Unfortunately, sensors dedicated to monitor grease condition in each turbine individually can be costly for wind park operators. Alternatively, periodic and detailed laboratory analysis can be used as a monitoring mechanism. Data about lubricant in terms of viscosity level and several other indexes of grease degradation and particle contamination can be monitored over time. Drawbacks of these laboratory tests are cost and time (both in terms of cost and time associated with the laboratory analysis as well as the extended time needed for careful collection of grease samples), as well as difficulties associated with reliably collecting and analyzing grease samples (it is hard to collect a representative sample while avoiding problems associated with extra grease contamination, region of the bearing that sample comes from, sample-to-sample variation, etc.).

Under such limitations, wind park operators often choose regreasing in fixed time intervals and visually monitor main bearing condition (associated with routine maintenance of turbines). In visual inspection, trained technicians assess the state of the lubricants based on visual characteristics of the lubricant and then assign a discrete rank to the grease condition. These visual hints may vary from the color to texture, and even to particle contamination that are visible to the naked eye. A potential ranking system is illustrated in Fig. 3. A discrete scale from 1 to 5 is adopted, where 1 indicates pristine grease, and 5 is contaminated and highly degraded grease. As opposed to detailed laboratory analysis, visual inspection campaigns are affordable and allow instant monitoring. However, visual inspection involves large uncertainties due to human error factor (both in terms of technician bias and technician-to-technician variation).

In this paper, we propose using physics coupled with machine learning to use grease visual inspection in modeling of bearing fatigue. The challenges introduced by grease inspection are: (i) variability in technician readings, (ii) inconsistency and conservatism of inspection, and (iii) limited number of available observations. In the next section we will show how we use such information to build a cumulative damage model.

3. Proposed physics-informed neural network for bearing fatigue and grease degradation

3.1. Hybrid physics-informed neural network

In this paper, we propose a hybrid physics-informed neural network approach to model main bearing fatigue damage
accumulation along with the grease degradation. Since damage accumulation is a time-dependent problem, we leverage recurrent neural networks, which are specially suitable for dynamical systems (Pearlmutter, 1989; Aussem, 1999; Goodfellow et al., 2016). Recurrent neural networks are machine learning models, where a state transition is carried out throughout a sequence (Goodfellow et al., 2016):

\[ a_t = f(x_t, a_{t-1}) \]  

where \( t \in [0, \ldots, T] \) represent the temporal discretization, \( a_t \) are the states at time \( t \), and \( x_t \) is the vector of input variables at time \( t \). \( f(\cdot) \) defines the transition between time steps (function of input variables and previous states).

Sophisticated recurrent neural network cells have been proposed to overcome training difficulties associated with “vanishing or exploding gradients” (Goodfellow et al., 2016). Fig. 4 illustrates the well-known long short-term memory cell (Hochreiter and Schmidhuber, 1997), detailing how the inputs, outputs, and auxiliary states pass through multiple non-linear nodes within the cell. In Appendix C, we compare our hybrid approach with pure data-driven long short-term memory cell.

The solution of the ordinary differential equation given in Eq. (1) is usually obtained through numerical integration (Press et al., 2007). In this work, we propose using the knowledge about the application to choose the integration method and then implement it as a deep neural network. The vector of states \( a_t = [a_{BG}, a_{GRS}] \) is formed by the bearing fatigue and grease damage, and that the input vector \( x_t = [V_{1}, T_{BG}] \) is formed by the wind speed and bearing temperature at time \( t \). Therefore, we can use the Euler cell (Nascimento and Viana, 2019) presented in Fig. 5a for numerical integration. The bearing damage increment, \( \Delta a_{BG,t} \), is quantified through physics-informed models discussed in Section 2.1. We compensate our uncertainty about the physics of grease degradation by modeling the grease damage increment, \( \Delta a_{GRS,t} \), with a multi-layer perceptron (MLP). This way, the transition function that models the incremental damage at time \( t \) is:

\[
\begin{align*}
    a_t &= a_0 + \sum_{k=1}^{t} \Delta a(x_t, y_{t-1}), \\
    \Delta a_{BG,t} &= \frac{1}{C_1 C_2} \left( \frac{P_1}{C} \right)^{10} \text{ and } \Delta a_{GRS,t} = \text{MLP} \left( V_{1,t}, T_{BG,t}, a_{GRS,t-1}; \mathbf{w}, \mathbf{b} \right).
\end{align*}
\]

Lubricant viscosity and the contamination level are the two parameters affecting main bearing fatigue life. In our work, we assume that both parameters are represented with one upper bound (acts as the state where grease is pristine), and one lower bound (where grease is fully degraded), as depicted in Fig. 2b. In order to account for grease damage, we scale these parameters with respect to current grease damage value \( a_{GRS} \):

\[
\begin{align*}
    y_t &= a_{GRS,t}(v_{degraded} - v_{pristine}) + v_{pristine} \text{ and } \eta_{c,t} = a_{GRS,t}(\eta_{c-degraded} - \eta_{c-pristine}) + \eta_{c-pristine},
\end{align*}
\]

where \( y_t \) and \( \eta_{c,t} \) are viscosity and contamination factor of the grease respectively.

The recurrent neural network cell implementing Eq. (3) is shown in Fig. 6a. The recurrent neural network cell takes in the wind speed and bearing temperature coming from the SCADA system (every 10 min). Physics-based nodes inside the cell model the bearing fatigue step-by-step. First, the wind speed is converted to dynamic bearing load (Fig. 2a), and the bearing temperature is used to evaluate grease related properties along with current grease damage (Fig. 2b). The information is used to obtain life adjustment factor \( C_2 \) for bearing fatigue. Finally, incremental bearing damage is evaluated as given in Eq. (1).

3.2. Handling grease visual inspection with a stacked recurrent neural network

As we debated in detail in Section 2, with visual inspection, we only have discrete ranks from \( R = 1 \) to \( R = 5 \) for grease state (with 1 meaning pristine and 5 meaning maximum acceptable degraded grease). We also know that the actual grease damage state is a continuous variable. Pristine grease has damage \( a_{GRS} = 0.0 \), which increases monotonically over time. This grease damage metric can be normalized such that the maximum allowable value is \( a_{GRS} = 1.0 \). Hence, grease damage relates but is not necessarily equivalent to the visual inspection ranking.

We model the mapping between the continuous grease damage into a discrete ranking as a classification problem. This way, we extend the model detailed in Section 3.1 by implementing a stacked recurrent neural network in which the base physics-informed neural network cell is stacked with a classifier. Fig. 6 illustrates the resulting model.

There are many options for implementing the classifier in Fig. 6. As shown in Appendix B, many off-the-shelf classifiers perform poorly in our application. We attribute this poor performance to the noisy nature of grease visual inspection (noisy labeled data in machine learning terminology) on top of the fact that grease rankings are ordinal (e.g., \( R = 1 \) precedes \( R = 2 \)). In order to overcome limitations of these classifiers, in this paper, we introduce a novel ordinal classifier that we call discrete ordinal classifier (DOrC).

DOrC consists of multiple sequences of switches, depicted in Fig. 7a. The first switch takes the continuous scale damage and activated with a unit output, if the scale satisfies the threshold that specific switch. The output of these switches sharply transition from 0 to 1 as they are activated. The output of a switch is multiplied by the damage and passed to the next switch. The DOrC output (which is the estimated rank) is the sum of the output of all switches. Depending on the target ranking system, a constant \( b \) can shift the final output to the desired lower bound. The final input–output relation resembles a staircase, as illustrated by Fig. 7b. This design preserves the sense of order of ranks and is also able to generate non-linear relations without assuming any pre-specific form (e.g., quadratic, cubic, etc.). Furthermore, in our application, the DOrC thresholds are optimized during the training of the stacked model.

---

\(^{2}\) Interested reader can refer to Yucean and Viana (2020), Dourado and Viana (2020), Nascimento and Viana (2019) for utilization of the Euler integration cell for damage accumulation across different application fields.
4. Case study: onshore wind farm with 1.5 MW turbines

4.1. Turbine specifications and operational data

In this study, we chose a 1.5MW wind turbine with 80 m hub height as our reference machine. This turbine model is equipped with a spherical main bearing in the three-point mounting setting. Table 2 provides some of the important parameters and specifications of the wind turbine and the bearing used throughout this work.

We consider a wind park consists of 120 wind turbines. On-site data is extracted from NREL database (Draxl et al., 2015), which includes hourly environmental data at different altitudes between 2007 and 2013 for more than 100,000 spatial data points across United States. We picked a certain location in Cooke County, TX, where an actual wind park is located. It should be noted that the data does not come from actual SCADA systems, however, we underline that using data from a location where an actual wind park exists would complement the representation of the input data.

In addition, raw data extracted from the database is preprocessed with the methods given in Yucesan and Viana (2020). The data is first augmented to 10 min resolution (in order to imitate SCADA system), and afterwards expanded to 30 years for long term prediction capabilities. Furthermore, an analytical relation that utilizes produced power and rotational speed is used to map the ambient temperature (provided by database) to bearing temperature. Detailed information about input preprocessing can be found in Appendix D. Fig. 8 shows the wind speed and bearing temperature recorded every 10 min over 7 years for single turbine.
4.2. Grease damage and visual inspections

We recognize that in real world applications, grease damage might not be fully observed (unless detailed and accurate laboratory analysis is performed). In fact, in this paper, we advocate that modern machine learning can allow us to still use grease visual inspection while developing models for bearing fatigue. However, in order to fully test our hybrid physics-informed neural network approach, we will use the manufacturer (Klueber-Contributors, 2011) grease lifing curves to build a model for grease damage (as detailed in Appendix E). We illustrate two different mappings between actual grease damage and visual inspection ranking. This allows us to study aspects such as conservatism level and variability in the ranking system – not practical without synthetic data. In addition, we assume that main bearings are fully degreased every six months. Again, only the synthetic visual inspections are used in the training of our proposed models. This synthetic data is generated by taking the grease degradation model elaborated in Appendix E as a baseline, and sampling from this ground true data using the distributions given in Fig. 9.

Ranking levels between 1 and 5 are assigned at inspection depending on actual (but unknown) grease damage. We established two scenarios to illustrate conservatism and variability in grease visual inspections, as depicted in Fig. 9. Fig. 9a illustrates “baseline inspection,” where the diagonal crosses approximately the 50th percentile of each distribution. This distribution is not symmetric and tends to be above 45° line, modeling minor conservatism. Fig. 9b illustrates “conservative inspection” since the distribution is heavily skewed towards higher ranks.

We assume that grease visual inspection campaigns are conducted on 10 out of 120 turbines of the wind park. The grease of these turbines is visually inspected monthly over a period of six months. For illustration only, the actual (but unknown) grease damage for the observed turbines is given in Fig. 10a. This data is never used by our models, but only used to generate visual inspection samples. In this figure, we also illustrate a shaded region that represent the degradation of entire fleet. Based on this illustration we show that the training turbines uniformly represent the distribution of the farm. Fig. 10b illustrates examples of the inspection rankings for one turbine. Even though grease damage increases monotonically, the visual inspections may not, illustrating the challenge associated with the variability in visual inspection.

4.3. Physics-informed neural networks configuration

We consider that wind speed and main bearing temperature from SCADA systems of every turbine across the farm, and visual grease inspections at every month for six months straight for 10 turbines within the park is available. We proceed to construct our hybrid model for bearing fatigue elaborated in Section 3. In this framework, bearing fatigue is physics-based, grease damage increment, ΔtGRS, is data-driven multi-layer perceptron, and the classification of grease damage to visual rankings is carried out through our custom designed discrete ordinal classifier (DOrC). The architecture of the ΔtGRS multi-layer perceptron is given in Table 3. The configuration in terms of number of layers, number of neurons in each layer, and activation functions was arbitrary and empirically refined. Further discussion on how to design the multi-layer perceptron is outside the scope of this paper. Depending on computational cost associated with application, we even encourage the interested reader to pursue neural architecture search (Kandasamy et al., 2018; Liu et al., 2018; Elsken et al., 2019) for optimization of the data-driven portions of the model. In a different application, Yucesan and Viana (2020) present an empirical study on the different architectures to model grease degradation. The inputs for this multi-layer perceptron models are scaled between zero and one so that the order of magnitudes of each input do not dominate one another.

Given that the visual grease inspections provide discrete ranks between 1 and 5, DOrC is designed with four switches. We implement the DOrC switches as sigmoid functions:

\[ \text{switch}_i(x) = \frac{1}{1 + \exp(a(x - \lambda_i))} \]

where \( i \in [1 \ldots 4] \), \( \lambda \) is the set of trainable hyperparameters (acting as thresholds between ranks), and \( a = -50 \) is chosen to make the function steep enough to maintain a binary transition (while smooth enough to avoid discontinuities during training of the deep neural network). This way, by adjusting each threshold, the classifier maps continuous damage index into discrete rankings. Even though we let thresholds to be learned by the model, we imposed the following bounds to constraint the parameters during the training:

\[ \lambda_1 \in [0.0, 0.3], \quad \lambda_2 \in [0.2, 0.5], \]
\[ \lambda_3 \in [0.4, 0.8], \quad \text{and} \quad \lambda_4 \in [0.8, 2.0]. \]

The optimization of the resulting network hyperparameters (1,251 from the multi-layer perceptron – see Table 3 – and 4 DOrC thresholds) uses the mean squared error as loss function:

\[ \text{Loss} = \frac{1}{N_0} \sum_j \sum_i \left( R_{ij}^{GRS} - \bar{R}_j \right)^2, \]

where \( R_{ij}^{GRS} \) is the ith observed grease visual inspection ranking for turbine j, \( \bar{R}_j \) respective rank predicted by our hybrid model, and \( N_0 \) is the total number of observations.

The hyperparameter optimization of these recurrent neural networks can be a daunting task. Initial parameters away from optimal would most likely induce divergence or long time of training procedure. In order to overcome this issue, we initialize the multi-layer perceptron parameters with the strategy presented in Yucesan and Viana (2020) and summarized in the Appendix A. We would like to underline that we are using the same initialization technique discussed in Yucesan and Viana (2020), although the parameters are not necessarily the same. After weights are initialized, we used RMSprop set with learning rate 0.0005 and 2500 epochs.

---

<table>
<thead>
<tr>
<th>Wind turbine</th>
<th>Main bearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rated power</td>
<td>Designation</td>
</tr>
<tr>
<td>1.5 MW</td>
<td>SKF 230/600 CAW3</td>
</tr>
<tr>
<td>Cut-in wind speed</td>
<td>Basic dynamic load rating C</td>
</tr>
<tr>
<td>3.5 m/s</td>
<td>6000N</td>
</tr>
<tr>
<td>Cut-out wind speed</td>
<td>Fatigue load limit ( P_f )</td>
</tr>
<tr>
<td>12 m/s</td>
<td>750 kN</td>
</tr>
<tr>
<td>Cut-out wind speed</td>
<td>Mass</td>
</tr>
<tr>
<td>25 m/s</td>
<td>405 kg</td>
</tr>
<tr>
<td>Maximum rotor speed</td>
<td>Mean diameter ( d_m )</td>
</tr>
<tr>
<td>20 rpm</td>
<td>735 mm</td>
</tr>
<tr>
<td>Hub height</td>
<td>80 m</td>
</tr>
</tbody>
</table>

---

Table 2: Wind turbine and main bearing specifications (adopted from GE-Contributors (2009) and SKF-Contributors (2007), respectively.)
Fig. 9. Mapping between grease damage and visual inspection ranks. Bars show the probability associated with ranks.

Fig. 10. Grease damage propagation and visual inspection example for the 10 turbines used in the training of the physics-informed neural network.

Table 3
Multi-layer perceptron architecture for grease damage increment, Δa(t). Total number of trainable parameters is 1251.

<table>
<thead>
<tr>
<th>Layer</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
</tr>
</thead>
<tbody>
<tr>
<td>#neurons/activation function</td>
<td>40/sigmoid</td>
<td>20/elu</td>
<td>10/elu</td>
<td>5/elu</td>
<td>1/sigmoid</td>
</tr>
<tr>
<td>#trainable parameters</td>
<td>160</td>
<td>820</td>
<td>210</td>
<td>55</td>
<td>6</td>
</tr>
</tbody>
</table>

4.4. Replication of Results

Our implementation is done in TensorFlow (version 2.0.0-beta1) using the Python application programming interface. In order to replicate our results, the interested reader can download codes and data. First, install the PINN package (base package for physics-informed neural networks used in this work) available at Viana et al. (2019). Then, clone the “pinn_wind_bearing” repository found in Yucesan and Viana (2019) and go to folder “probabilistic_grease_inspection/comind_2020”. This repository includes three python scripts where the first one samples visual grease inspections based on ground true grease data, the second one trains the recurrent neural network using a pretrained multi-layer perceptron model with fixed initial weights, and the last script predicts the fatigue damage accumulation of the wind turbine main bearing for 20 years. The reason we limited the time frame to 20 years and not 30 years as we used in the paper, that the size limitation of the database we use to share our data. The data used in this work is publicly available in Yucesan (2020). Download the data and extract folders inside “visual_inspection_dataset_2020” to the directory where the “pinn_wind_bearing/probabilistic_grease_inspection/comind_2020” repository is cloned. For reference, our simulations were conducted using a laptop configured with an Intel Core i7-8650U CPU at 1.90GHz, 32GB of RAM, and NVIDIA Quadro P500 GPU running Windows 10.

5. Results and discussion

We start by analyzing the training of the proposed physics-informed neural network model. The training data (wind speed, bearing temperature, and rankings out of grease visual inspection for 10 turbines) is used to simultaneously optimize the 1251 multi-layer perceptron hyperparameters and the 4 DORC thresholds. The computational cost for each training process is approximately 7 hours for 10 wind turbines. The prediction takes approximately 8 min per turbine, since we forecast up to 30 years.

Table 4 shows the confusion matrices out of the predictions at the test turbines (which are selected from the same farm other than training turbines) when the model is trained with either the baseline or conservative inspection data. Given that the training of the neural network uses the mean square error as loss function, in both cases, the networks will result in unbiased predictors. However, these are predictions of rankings out of grease visual inspection. Unfortunately, as we previously discussed, grease visual inspection is prone to large bias and variance. Should we compare the predicted and actual grease damage, we would see the manifestation of such uncertainty in predictions.
We proceed using our physics-informed neural network model to estimate both grease degradation and main bearing fatigue damage. Fig. 13 illustrates the results for one turbine of the park that was not in the training set. Fig. 13a shows the estimated and actual grease damage over time. The damage is reset back to zero every six months as we assume regreasing in fixed time intervals. The conservatism in grease visual inspection used to train the models is reflected in grease damage predictions. Fig. 13b presents the estimated and actual bearing fatigue damage over time. Even though there are different degrees of conservatism in grease damage estimation, the bearing fatigue damage estimation is in relatively good agreement. Bearing fatigue damage is only marginally overestimated, despite conservatism in grease models. The reason for such behavior is the regreasing policy. Thus far, we assumed that bearings are fully regreased every six months. Therefore, discrepancies in grease damage do not accumulate for long enough to largely influence bearing fatigue damage estimation.

Fig. 14 summarizes the study of bearing fatigue estimation with the time-to-failure for the wind park (i.e., time needed for fatigue damage to reach 1.0). As expected, the degree of conservatism is related to the grease visual inspection rankings. Nevertheless, the other important observation is that expected useful lives are very long (easily surpassing the anecdotal 20 year limit). Given that our physics-informed neural network model predicts both grease degradation and fatigue damage, we can use it for optimizing maintenance costs (trade-off with the useful life).

We illustrate the use of our physics-informed neural network model so that we find the optimum regreasing time on a turbine-by-turbine basis such that bearing fatigue life stays near 20 years. These model-based regreasing intervals take into consideration the wind speed, bearing temperature, and degradation dynamics particular to a given turbine. Therefore, regreasing intervals are expected to be different across the wind park. From a practical perspective, we keep the optimized regreasing intervals discretely between 3 and 18 months. Fig. 15a shows a scatter plot of regreasing interval versus predicted bearing fatigue life across the 120 turbines of the wind park. Interestingly, while for most turbines predicted fatigue life is near the 20 year target; the regreasing intervals vary

The data used in this research allows us to compare the estimated grease damage (out of our proposed recurrent neural network cell) with actual grease damage. In fact, one of the benefits of our stacked physics-informed neural network is that, the trained model can be used to estimate grease damage, as opposed to only estimating the rankings (even though actual grease damage was never observed and our model was trained with only grease visual inspection). Fig. 11 presents the prediction results at the training set. As expected, the comparison with Fig. 10a reveals that the conservatism in grease visual inspection rankings is translated into conservatism in grease damage estimation.

Fig. 12 illustrates how the predicted grease damage compares with actual (but unknown) grease damage across the entire wind park. The model trained with the baseline inspection results is significantly less conservative than the one trained with the conservative inspection results. In our opinion, most practitioners expect that visual inspection rankings are noisy and bias in such data contaminates the resulting models. While the less contaminated the data is the better the resulting models are; our proposed approach shows promising results even under such limited data.

Table 4
Confusion matrices obtained on the test turbine set (predicted ranks versus ranks given by visual inspection).

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Baseline visual inspection</th>
<th>Conservative visual inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Fig. 11. Grease damage predictions for the 10 turbines used in the training our physics-informed neural networks.

Fig. 12. Grease damage predictions versus actual grease damage for the wind park (120 turbines).
dramatically across the park. This result is driven by the differences in input conditions (wind speed and bearing temperature). Fig. 15b shows the comparison of the bearing unreliability over time for the model-based optimized results against the two reference fixed intervals (3 and 18 months). While solid lines are obtained with model predictions, dashed lines are the actual unreliability levels. Adopting fixed regreasining intervals is sub-optimum and bounded by the red (18 months) and green (3 months) curves. We can also acknowledge these red and green curves as a representation of farm operating with pristine grease versus operation under bad grease condition. Therefore, the gap in between can be perceived as the influence of grease damage on the bearing fatigue, from a farm level time-to-failure perspective. On the other hand, our model was able to optimize regreasining intervals given the target useful life.

Finally, Fig. 16 illustrates how input conditions drive regreasining intervals. Fig. 16a shows the distribution of grease damage at the time of regreasing for two turbines when regreasing is obtained with the model-based optimization. For some turbines regreasing is required even when grease is not close to failure, and others can be left to operate with bad grease. When turbine is regreased every 3 months, the grease damage is considerably low. On the other hand, when turbine is regreased every 18 months, the grease damage is considerably high. This difference in regreasining interval compensates for the differences in input conditions (wind speed and bearing temperature) such that useful life is expected to be around 20 years for both turbines. Fig. 16b provides a perspective in terms of bearing temperature for the turbines of the farm with expected life around 20 years (vertical dotted line of Fig. 15a). Main bearings operating under high temperatures require frequent regreasining, while main bearings exposed to mild temperatures can go up to 18 months without regreasining.

6. Summary and closing remarks

In this paper, we proposed a hybrid physics-informed neural network approach to handle the simultaneous modeling of bearing fatigue and grease damage accumulation. Our proposed recurrent neural network cell is a hybrid that combines estimation of bearing fatigue damage increment through physics-informed kernels with estimation of grease damage increment through a multi-layer perceptron. The grease model is then tuned with the help of grease visual inspection data, which provides a ranking system indicating the grease condition. Therefore, the proposed recurrent neural network is stacked with a classifier that maps grease damage into the grease ranking.

With this proposed framework, we hope to address two main challenges in wind turbine bearing fatigue estimation. Firstly, we provide an estimation for grease degradation with the physics-informed recurrent neural network cell. In fact, this is equivalent to performing model-form uncertainty quantification (as we assume that physics of grease degradation is unknown). Secondly, we address the tuning of the neural network hyperparameters with discrete grease ranking coming from visual inspection. In order to illustrate the performance of our framework, we considered a case study in which (a) a wind park of 120 turbines, out of which 10 were inspected every month for 6 months, and (b) rankings out of grease visual inspections had different levels of uncertainty. Results showed that our hybrid physics-informed neural network can simultaneously learn the grease damage accumulation and the classification.

Additionally, we confirmed that conservatism in ranking system leads to conservatism in the hybrid physics-informed neural network. Nevertheless, under mildly conservative ranking, we can use the resulting model to optimize regreasining intervals for a desired useful life target. With that, operators will be able to trade-off maintenance and operations cost with overly exceeding bearing life.

Finally, in the light of the results obtained thus far, we hope the proposed approach can help practitioners and wind park operators modeling bearing fatigue and grease degradation. Given that we propose using turbine operation data as input and grease visual inspection as output in a manner that is accurate and cost effective. Nevertheless, we recognize that, in general, the degree of conservatism in grease visual inspection is not known. Therefore, we envision that future research could involve quantification of visual inspection uncertainty with data coming from grease laboratory analysis, for example. Even though our approach uses turbine-specific data, a major issue with any predictive model that relies on SCADA is the quality and completeness of such data. In the future, it would be interesting to study ways to mitigate problems with such data by, for example, sharing information across turbines that are spatially close to one another. In addition, Other topics of interest may include uncertainty in the loads model and material capabil-
ities, as well as optimal strategies for fleet recommissioning and bearing replacement.

**Authors’ contribution**

Yigit A. Yucesan: methodology, software, formal analysis, investigation, data curation, writing, visualization. Felipe A.C. Viana: conceptualization, methodology, validation, formal analysis, investigation, writing, supervision, funding acquisition.

**Appendix A. Guided initialization of neural network**

Training large number of parameters within deep neural networks can be challenging, especially when the output is highly nonlinear. Therefore, we promote the initialization of multi-layer perceptron parameters prior to recurrent neural networks training in an effort to improve optimization convergence. Here, we initialize the grease damage increment model with a linear plane:

$$\Delta a_{GRS} = \alpha_0 + \alpha_1 \times T_{BRC} + \alpha_2 \times V_W + \alpha_3 \times a_{GRS},$$  

(8)

where $\Delta a_{GRS}$ is the grease damage increment, $T_{BRC}$ is the main bearing temperature, $V_W$ is the wind speed, and $a_{GRS}$ is predicted cumulative grease damage.

The coefficients, $\alpha_i$, are determined through engineering intuition-based judgement. For instance, we should expect increasing $\Delta a_{GRS}$ with increasing bearing temperature, which constrains $\alpha_1$ to be positive. Within these logical limitations, $\alpha_i$ can be randomly perturbed. Interested reader can find the ablution study on the effect of different initial conditions on model performance in Yucesan and Viana (2020). Fig. 17 illustrates the linear plane approximation. In this figure, we see how the output grease damage increment $\Delta a_{GRS}$ is changed with wind speed and bearing temperature (note that the third input variable $a_{GRS}$ is fixed at 0.5 for the sake of three dimensional illustration). While the orange surface represents the actual (but unknown) input-output relationship, the blue surface is the linear plane used for network initialization.

**Appendix B. Comparison of classifiers used in the stacked recurrent neural network**

In Section 3, we proposed the novel Discrete Ordinal Classifier (DOc) in order to map grease damage scale to visual grease inspections. Here, we show a comparison between DOc and other two popular classification methods. One of the simplest methods to encode discrete variables is called “one-hot” encoding. Output classes are encoded as a binary vector, such that for each output time step the ground true vector becomes 1 for the class that is active, and 0 for all the other classes that are inactive. For the 1–5
discrete scale we have for visual grease rankings, for example if the visual ranking is 4, the output vector becomes:

\[ R_{GRS} = 4 \text{ one-hot encoding} \{1 1 1 1 0\}. \] (9)

The other popular approach we tested is the ordinal encoding (Cheng et al., 2008). The binary encoding is applied; however, each output is encoded as a vector whose length is one less than the number of classes. For a class to be activated, all the prior elements have to be 1 (up to the class of interest). For instance, for our 1–5 discrete scale, if the visual ranking is 4, the output vector becomes:

\[ R_{GRS} = 4 \text{ ordinal encoding} \{0 0 0 1 0\}. \] (10)

Fig. 18 shows the performance of the resulting trained model with each classification method. Baseline visual inspection campaign data is used for training of all models. One-hot encoding yields to very poor performance for this application, mostly because the encoding does not provide an ordinal sense. The ordinal encoding performs better than the one-hot encoding. Unfortunately, predictions tend to surprisingly underestimate grease damage, even though ranking data used for training is slightly conservative. DORC outperforms both conventional methods yielding the expected conservative predictions.

**Appendix C. Conventional recurrent neural networks**

In this appendix, we compare our hybrid physics-informed neural network approach against a pure data-driven model: long short-term memory (LSTM) recurrent neural network cell (Hochreiter and Schmidhuber, 1997). We chose two different complexity levels for LSTM cells. One is a single layer architecture, that we call “shallow LSTM”, and the other one consists of multiple layers, we call “deep LSTM”. Table 5 summarizes the architectural details for these models. The reason behind we picked these two configurations is to have two LSTM models that will provide fair comparison, one shallow network with fewer number of parameters to be trained and the other deep network with more parameters than our proposed hybrid model (1251 from multi-layer perceptron and 4 from Discrete Ordinal Classifier). Another interesting approach here can be applying neural architecture search that might provide the optimum configuration best fits to the problem. We encourage interested readers to compare the physics-informed neural networks model to other fully data-driven models for this application.

We trained both models with the same optimization settings and data used to train our hybrid model discussed in Section 4.3. Table 6 present the confusion matrices after training of both models. At face value, these LSTM models exhibit good performance in predicting the noisy visual inspection rankings (the network depth does not make any difference in the prediction performance). However, we suspect these models tend to fit the data by disregarding the ordinal nature of the problem. In fact, Fig. 19 illustrates this point with a time history prediction of LSTM models against our physics-informed neural network approach for a single turbine within the training set. Unfortunately, the LSTM models perform poorly to approximate time history prediction of grease visual inspection ranks. Without the injected physics, the LSTM models are not robust to the variability in the data and predictions tend to go up and down (as opposed to monotonically increase). On the other hand, the hybrid approach we proposed with the physics-informed recurrent network stacked with the DORC classifier performs well in prediction and preserves the ranking evolution over time, reflecting damage accumulation.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Long short-term memory (LSTM) network designs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
<td>Layers</td>
</tr>
<tr>
<td>Shallow LSTM</td>
<td>1</td>
</tr>
<tr>
<td>Deep LSTM</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 6**

Confusion matrices obtained on the testing set (predicted ranks versus ranks given by visual inspection).

<table>
<thead>
<tr>
<th></th>
<th>Shallow LSTM Baseline visual inspection</th>
<th>Deep LSTM Baseline visual inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
| 5              | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1
Appendix D. Input data preprocessing

In wind turbines, supervisory control and data acquisition (SCADA) systems are usually available on board. SCADA systems record data from sensors and control system every 10 min. In this study, we assume wind speed and main bearing temperature are provided through SCADA system for every turbine of the fleet. However, the data we can extract from NREL database (Draxl et al., 2015) is wind speed and ambient temperature at 80 m altitude recorded every hour. In order to represent SCADA data, we bootstrapped data obtained from NREL database. Each day is represented by eight bins of three hours segments and each bin aggregates a week worth of data. In other words, each bin has 21 data points coming from the same 3 hours of the day across a week. We then sample at random (with replacement) from this pool to fill in the extra 5 points per hour needed within each bin. This process is repeated with a sliding weekly window throughout the year so that seasonality is preserved. While the NREL database covers 7 years, some of our simulations needed data for up to 30 years. To overcome this limitation and also to provide a mechanism for forecasting damage accumulation. Again, we bootstrapped from the previously augmented data binning it at every 10 min by time of the day and day of the year across the seven years. We calculated the mean and standard deviation of each bin and assuming normal distribution, we sampled data points for the same time stamp of the forecasted year.

As we mentioned before, the NREL database provides ambient temperature, however our model requires main bearing temperature. In order to preprocess the temperature data, we used the model proposed by Cambron et al. (2017). In essence, the main bearing temperature is described by a recursive model as a function of previous bearing temperature, nacelle temperature, angular velocity, and generated power.

Appendix E. Baseline grease degradation data

Grease degradation is a complex phenomenon to model. In this paper, we adopted a simplified model found in Klueber-Contributors (2011) to form our baseline ground true data for grease degradation. The model relates grease life with bearing temperature and a number of adjustment factors:

\[ L_{nm}^{GRS} = L_{nm}^{GRS} \cdot \kappa \cdot K_N \cdot K_B \cdot F_1 \cdot F_2 \cdot F_3 \cdot F_4 \cdot F_5 \cdot F_6 \]  

(11)

Fig. 20a illustrates how grease service life varies with temperature. Most adjustment factors are given in Table 7. \( F_3 \) is a factor that accounts for dynamic load variation and it is shown in Fig. 20b. As stated by Lucht (2009), the bearing life is commonly expressed in terms of \( L_{10} \) life (as a safety factor to account for the variation in grease properties).

Table 7
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Accounts for</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_N )</td>
<td>7.69</td>
<td>Bearing design</td>
</tr>
<tr>
<td>( K_B )</td>
<td>0.15</td>
<td>Spherical bearing design</td>
</tr>
<tr>
<td>( F_1 )</td>
<td>0.8</td>
<td>Dust and humidity</td>
</tr>
<tr>
<td>( F_2 )</td>
<td>0.9</td>
<td>Shock, vibration, and oscillation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Accounts for</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_3 )</td>
<td>1.0</td>
<td>Air flow</td>
</tr>
<tr>
<td>( F_5 )</td>
<td>1.0</td>
<td>Rotating outer ring</td>
</tr>
<tr>
<td>( F_6 )</td>
<td>1.0</td>
<td>Vertical shaft arrangement</td>
</tr>
</tbody>
</table>

Declaration of Competing Interest

The authors report no declarations of interest.

References


