A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation

Yigit A. Yucesan, Felipe A.C. Viana

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

**Abstract**

Fatigue life of a wind turbine main bearing is drastically affected by the state of the grease used as lubricant. Unfortunately monitoring the grease condition through predictive models can be a daunting task due to uncertainties associated with degradation mechanism and variations in grease batch quality. Eventually, discrepancies in the grease life predictions caused by variable grease quality may lead up to inaccurate bearing fatigue life predictions. The convoluted nature of the problem requires a novel solution approach; and in this contribution, we propose a new hybrid physics-informed neural network model. We construct a hybrid model for bearing fatigue damage accumulation embedded as a recurrent neural network cell, where reduced-order physics models used for bearing fatigue damage accumulation, and neural networks represent grease degradation mechanism that quantifies grease damage that ultimately accelerates bearing fatigue. We outline a two-step probabilistic approach to quantify the grease quality variation. In the first step, we make use of the hybrid model to learn the grease degradation when the quality is the median of the distribution. In the second step, we take the median predictor from the first step and track the quantiles of the quality distribution by examining grease samples of each wind turbine. We finally showcase our approach with a numerical experiment, where we test the effect of the random realizations of quality variation and the number of sampled turbines on the performance of the model. Results of the numerical experiment indicate that given enough samples from different wind turbines, our method can successfully learn the median grease degradation and uncertainty about it. With this predictive model, we are able to optimize the regreasing intervals on a turbine-by-turbine basis. 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

Approximately 30% of the installed wind turbine fleet in the United States have already been operating for more than 10 years, and have 1.5 to 2.0 MW rated power [1]. For this legacy fleet, fatigue of the main bearing is a major issue that impacts the operation and maintenance costs. Additionally, old machines are not equipped with high end condition monitoring systems, and rely on 10 min averaged supervisory control and data acquisition (SCADA) system data. In the literature, researchers have built prognosis models for main bearing fatigue using SCADA data. Butler et al. [2] considered parameters provided by SCADA system such as main shaft rotational speed, hydraulic brake temperature, hydraulic brake pressure, and blade pitch position as well as a compensation for ambient temperature. Combining the information from these parameters and analyzing the deviation from normal operation, they...
predicted a failure indication with a 30 day lead time, that is verified with actual failure. Watanabe and Uchida [3] estimate wind turbine rear bearing fatigue using standard bearing life calculations found in ISO 281 [4]. Their model processed SCADA data and showed good agreement with failures observed within a wind energy park in Japan. While collected field data concluded 90% reliability 12.7 years, the model predicted 12 years. Authors also showcased benefits of their model for life extension through curtailment. SCADA data is also extensively used for prognosis of all types of bearings within wind turbine drive-train, either with a physics-based approach [5], or purely data-driven prognostics [6]. Different than SCADA data, the analytics on bearing prognosis can also be improved if vibration data exist [7,8]; however, vibration response is usually utilized for short term anomaly detection, rather than long term life prediction.

State of the grease that surrounds a main bearing is a major factor for the fatigue life of the component [9]. There are two main challenges for engineers while modeling grease service life: complex mechanism of the degradation and uncertainties associated with the grease (i.e. variable quality). Eventually, an accurate model that is robust to the uncertainties and can predict when grease will fail is a need, not only for the grease but also indirectly for the bearing fatigue life estimation. The experts have also thoroughly investigated the grease degradation phenomenon. Zhu et al. [10] proposed a methodology for estimating the remaining useful life of lubricant using viscosity and a dielectric constant sensor output and integrating these parameters as an observation function by particle filtering technique to predict the remaining useful life of the lubricant. Their proposed model was validated by laboratory experiments. Results of the conducted case study show that the single observation on dielectric constant sensor gives the best accuracy on the life prediction of grease. Iyer et al. [11] proposed a method for the early detection of lubrication anomalies in oil-lubricated bearings. Authors investigated two types of anomalies: lack of lubricant and presence of contamination. In their study, they used acoustic emissions and vibration signals. Through experiments, they showed that these techniques not only detect the anomaly, but also provide an insight on the level of the anomaly. Even though these methods proven to be effective for grease life modeling, they are insufficient to capture quality variation. For a lifting model that can be useful for prognosis, we require a grease degradation model that can quantify grease quality variation, and differentiate site to site, and even turbine to turbine variation.

Hybrid models that mix physics and machine learning are attractive due to the potential to reduce required data to train data-driven nodes with the guidance of fundamental physics relations. While known physics is used to establish a baseline for the model, machine learning is used to compensate for the gap due to unknown portion. Specifically for prognosis and health management area, different hybrid methods found themselves various applications. Liu et al. [12] adopted a sequential hybrid approach for remaining useful life estimation of a jet engine auxiliary power unit. Their approach first utilized Wiener process to formulate degradation data of the system, then this data is used to train a data-driven long short-term memory cell for remaining useful life prediction. Chao et al. [13] discussed how hybrid physics-informed and machine learning models can be used together for prognostics of complex safety critical systems. The authors use physics-based models to infer features cannot be observed, and feed them into a deep learning framework along with other inputs. They evaluated the performance of their framework on a case study that covers degradation data from nine turbofan engines, and proved an extension of prediction horizon approximately by 127% compared to pure data-driven counterparts, by requiring even less data. Although there is popularity within the literature towards hybrid models, there is still a gap with respect to a method that uses sparse and noisy data to model time series degradation for grease.

In our recent work, we have utilized hybrid models to uncover the unknown grease degradation mechanism [14], and prove the robustness of the same model by implementing noisy and biased visual inspection routines [15]. In this article, we substantially expand the research and address the variable grease quality problem by building a hybrid physics informed neural network [16] model for wind turbine main bearing fatigue damage accumulation, where reduced-order physics sub-models will form the bearing fatigue portion, and a neural network node will represent the grease degradation that directly affects bearing fatigue. Using this hybrid model, we aim to establish a probabilistic two-step method in order to quantify the uncertainty in grease parameters caused by the variation in quality. First, we will train our hybrid model with field grease samples (collected from a few turbines in a short period of time) to learn the median behavior. Second, we will extract the quantiles of the quality distribution using samples from individual machines.

The remaining of the paper is organized as follows. Section 2 establishes core models essential for bearing fatigue modeling, and underlines the manifestation of grease quality variation on field. Section 3 presents the formulation and implementation of the hybrid physics-informed neural network model and outlines how it is utilized to model quality uncertainty. Section 4 summarizes the case study picked to showcase the performance of proposed model. Section 5 details the results of the numerical experiment and discuss the performance of the model from several different perspectives. Section 6 closes the paper recapitulating salient points and presenting conclusions and future work. Finally, Appendix A summarizes concepts about neural networks used in this paper and Appendix B provides results for data-driven model applied to the problem at hand.

2. Bearing fatigue prognosis with grease quality variation

2.1. Bearing fatigue prognosis

In design phase of a bearing, engineers mainly rely on accurate high-fidelity fatigue models (i.e. detailed finite element models) to evaluate and optimize the lifespan of the component. While high-fidelity models work for design, their computational hinder their application as prognosis digital twins for the component.
A convenient alternative is to make use of reduced-order models for prognosis, that can capture physics-of-failure to some extent, and can be calibrated and corrected without requiring days of run time. For that purpose, we choose to adopt a standardized bearing life formula found in ISO 281 [4] and express it in terms of incremental damage in the form of an ordinary differential equation:

$$\frac{da_{BRG}}{dt} = \frac{1}{c_1 c_2(t)} \left( \frac{P(t)}{C} \right)^{10},$$

where $P$ is the equivalent dynamic bearing load (see Fig. 1(a)); $C$ is the design load rating; $c_1 = 1.0$ is a reliability level factor; and $c_2$ is an adjustment factor based on grease properties (see Fig. 1(b)).

The model presented in Eq. (1) is provided by bearing manufacturers and widely accepted and practiced in the industry, as it relates the component life to both dynamic loads and grease properties through factor $c_2$. One should also recognize that grease surrounds the bearing also degrades over time. Degradation of the grease naturally yields to lower viscosity, higher contamination, and eventually accelerates the fatigue degradation of the bearing. Life adjustment factor $c_2$, given in Eq. (1) enables us to account for the effect of grease with quantifiable properties such as viscosity and contamination ratio. As seen in Fig. 1(b), these properties can be parameterized to reflect the actual state of grease, by linearly interpolating between curves denoted as degraded and pristine based on the grease damage value:

$$v_i(t) = a_{GRS}(t)(v_{deg} - v_{prs}) + v_{prs}, \quad \eta_i(t) = a_{GRS}(t)(\eta_{deg} - \eta_{prs}) + \eta_{prs} \quad (2)$$

Available grease degradation models provide life of the grease as a function of temperature. Figs. 1(c) and 1(d) illustrate two example for such models. Fig. 1(c) depicts two life curves provided by the grease manufacturer Klueber. These curves indicate two distinct reliability levels ($L_{10}$ and $L_{50}$). In this case multiple reliability levels can be interpreted as different quality of the grease batch that causes uncertainty in grease life estimations. In a similar model, Lugt [18] provides several curves for $L_{10}$ life of grease, where each curve represents different design and operation condition for the bearing. All in all, these models are useful to provide insight on degradation of grease, however they include large uncertainties due to variable environment conditions, unaccounted extreme events, and mostly because of the quality variation.

All the models discussed above require inputs from field observations for calibration. Unfortunately, environmental factors and turbine operation can highly contaminate the samples, therefore makes it really difficult to build principle-based models. Eventually, when we are dealing with a very complex degradation mechanism like grease, and have uncertain observation data, we need a unique solution towards all these challenges.

---

1 See [17] for all reliability levels.
Fig. 2. First two figures show the impact of variation in the grease quality in the grease and bearing damage accumulation for two representative turbines. The "mild" and "aggressive" turbine denomination is based on operating conditions. Last illustration depicts grease damage accumulation for different samples of grease for 10 turbines. Individual turbines are color-coded to highlight the impact of grease quality in grease damage accumulation.

2.2. Grease quality variation

In a wind farm, the quality of the grease used in each wind turbine's main bearing can change drastically (even if the grease is from the same manufacturer). This batch-to-batch variation is very difficult to keep track of in practical terms, and may lead to large uncertainties in bearing fatigue life estimation.

Fig. 2(a) illustrates how the spread in grease quality interacts with operating conditions and affects grease degradation. As expected, the more aggressive the operating conditions, the higher the damage accumulation rate. Here we assume that once a turbine is loaded with a grease, it operates with that grease until the machine is regreased, which happens regularly every 6 months. As a consequence, every time the turbine is regreased it receives a grease behavior of which falls in the distribution shown in Fig. 1(c) (effectively, a random sample from that distribution). Fig. 2(b) shows the variation in the fatigue life as a result of the uncertainty in grease quality for the two representative turbines.

We hypothesize that every time a turbine is greased, it actually receives a random sample of grease from the distribution illustrated in Fig. 1(c). Therefore, it is virtually impossible to assign ahead of time a grease sample to a specific turbine. In other words, we assume that the case in which the most aggressive turbine receives a poor grease sample is practically possible (and operators cannot avoid it). Fig. 2(c) illustrates realizations of grease damage accumulation for 10 different turbines. For the sake of illustration, the curves shown in both subfigures of Fig. 2(c) are obtained using exactly the same operating conditions. The color code is used to mark a specific turbine so that we can visualize the effect of grease quality in grease damage accumulation.

3. Hybrid physics-informed neural networks for damage accumulation

Recurrent neural networks are specially suitable for modeling dynamical systems [19–21]. They extend traditional feed forward networks to handle time-dependent responses. Recurrent neural networks have been used to model time-series [22], speech recognition [23], diagnosis and prognosis [24–27], and many other applications. A recurrent neural network [19] repeatedly apply transformations to given states in a sequence such that

\[ y_t = f(x_t, y_{t-1}). \]

where \( t \in [0, \ldots, T] \) represent the time discretization, \( y \in R^n \) are the states representing the sequence, \( x \in R^m \) are input variables, and \( f(\cdot) \) defines the transition between time steps (function of input variables and previous states). In the recurrent neural network terminology, different implementations of \( f(\cdot) \) are referred to as cells.

In Fig. 3(a) we see one of the popular architectures, the long short-term memory cell [28,29]. This cell exhibits complex implementations, with inputs, outputs (and even auxiliary states) passing through multiple non-linear nodes within the cell. We also applied this pure data-driven approach to our problem in order to provide a comparison with our hybrid solution in Appendix B. Following the same logic, we can also design a cell that can carry out numerical integration over a fixed time step. In this case, we leverage the Euler cell (illustrated in Fig. 3(b)) introduced in [30,31]. Similar to operations within a long short-term memory cell, we can tailor the increment function \( \Delta y \) that can contain data-driven or reduced order physics-based sub-models, or even both as a hybrid model. In the next subsection, we will present the main bearing fatigue and grease degradation implementation as a hybrid Euler cell.

3.1. Main bearing fatigue damage accumulation

Here, we model bearing fatigue (including grease degradation mechanism) with the concept of hybrid physics-informed neural networks. To summarize this approach, a graph model represent the input–output relationship such that different nodes in the graph can be either physics-based or data-driven nodes. In our implementation, this graph model also represents a deep neural network.
Knowing that we are interested in time-dependent damage accumulation for both bearing fatigue and grease as a function of turbine operation data, we use recurrent neural networks [19] repeatedly apply transformations to given states in a sequence
\[ \mathbf{a}_t = f(\mathbf{x}_t, \mathbf{a}_{t-1}) \tag{4} \]
where \( \mathbf{a}_t = [a_{BRG,t}, a_{GRS,t}] \) are the bearing and grease damage states, \( \mathbf{x}_t \) is the vector of input variables (wind speed and bearing temperature).

In this work, we expand the Euler integration cell to implement numerical integration of Eq. (1) while also accounting for grease uncertainty. We encapsulate both bearing fatigue damage and the grease damage models in to a recurrent neural network cell in order to estimate damage accumulation at each cycle (Fig. 4(a)). The recurrent neural network takes wind speed and bearing temperature at each time step, which would come from SCADA data. Within the cell, there are physics-informed nodes modeling bearing contact fatigue. As we walk through the numbers on Fig. 4(a), node 1 depicts the loads model, also detailed in Fig. 1(a). This model comes from multi-body physics simulations that map wind speed to equivalent dynamic bearing load (provided in Sethuraman et al. [32]), specific to the spherical bearing used in this study. Node 2 is the grease property model given by manufacturer specification catalog [17], and depicted in the first two plots of Fig. 1(b). This node takes the current value of grease damage, bearing temperature, and wind speed, and outputs quantifiable grease properties, viscosity and contamination factor. Node 3 is a grease-related bearing fatigue life coefficient model also given by bearing manufacturers [17], shown in the last plot of Fig. 1(b). Sole purpose of this model is to lump grease characteristics into a life modification factor. Finally, node 4 calculate incremental bearing fatigue damage based on loads and grease properties. One can also state that this node is the graph implementation of Eq. (1).

Even though most grease manufacturers provide life curves for the grease [33], these models often include several reduction factors due to large uncertainties associated with experiments that yield these curves. Given the poorly understood physics of grease degradation, we use a data-driven node to model grease damage increment. As we model the grease degradation, we consider the grease damage index \( a_{GRS} \) is the quantification of the grease state, that is accumulated as a function of wind speed (or loads), bearing temperature, and the damage index value at the previous time stamp (for nonlinear degradation). In this paper, we implement this grease damage increment node as a multi-layer perceptron (see Appendix A for the fundamentals of a multi-layer perceptron).

![Fig. 3. Hybrid physics-informed neural network modeling using recurrent neural networks.](image)

### 3.2. Modeling grease quality uncertainty by tracking quantiles

The cumulative damage model depicted in Fig. 4(a) has previously been proved to quantify model-form uncertainty (due to unknown grease degradation mechanism) [14] and observation uncertainty (visual grease inspections) [15]. The substantial difference between the previous studies and the challenge we are dealing with now is the fact that the quality of grease is uncertain, as it is sampled from a distribution.

For the sake of completion of our numerical study, we choose the grease service life distribution shown in Fig. 1(c) as our ground truth model. In order to have a concrete understanding of how this model is utilized to generate ground truth data, let us consider the \( L_{50} \) curve from Fig. 1(c). For a given operating bearing temperature at any time stamp, \( t \), we can obtain a life value for the grease:
\[ L_{50,GRS}(t) = f(T_{BRG}(t)) \tag{5} \]

Using Palmgren–Miner’s rule, we can also calculate an incremental damage. Assuming that the degradation mechanism is rather nonlinear, we adopt a quadratic relation between the life and damage for grease:
\[ \Delta a_{50,GRS}(t) = \left( \frac{1}{L_{50,GRS}(t)} \right)^2. \tag{6} \]
By the nature of damage accumulation, we express the cumulative damage at any given total time of $T$ as follows:

$$a_{S0,GRS}(T) = \sum_{t=0}^{T} \Delta a_{S0,GRS}(t).$$

We should underline that Eq. (5) through (7) are valid only for the 50th quantile, also known as the median, of the distribution. One can realize that the 10th quantile curve of grease life is only a shift of 50th quantile curve in the log-scaled $y$-axis, as shown in Fig. 1(c). We denote this shift as the quantile ratio, $C$. With that said, we can now express any $k$th quantile curve (also the cumulative damage) as a function of the median grease curve and the quantile ratio:

$$L_{k,GRS}(t) = C_k L_{S0,GRS}(t), \quad \Delta a_{k,GRS}(t) = \left( \frac{1}{C_k L_{S0,GRS}(t)} \right)^2, \quad a_{k,GRS}(T) = \sum_{t=0}^{T} \left( \frac{1}{C_k L_{S0,GRS}(t)} \right)^2.$$

Since the quantile ratio $C_k$ is a constant value for every $k$th quantile, and we are interested in mapping the damage rather than the life, the operations applied to this coefficient given in Eq. (8) does not make any difference in the formulation if we take the value out as a multiplier (if one is interested in reconstructing the life curves, the model form for $C_k$ can also be preserved):

$$a_{k,GRS}(T) = C_k \sum_{t=0}^{T} \left( \frac{1}{L_{S0,GRS}(t)} \right)^2 = C_k a_{S0,GRS}(T).$$

Eq. (9) implies that, all we need to know to quantify the uncertainty in quality is to learn the median behavior of grease, and distribution of the quantile ratios. At this point, we propose a two step approach. First, we will use our hybrid physics-informed neural networks model described in Section 3.1 to learn the median grease damage accumulation using periodic grease observations sampled from multiple wind turbines. Second, we will infer the quantile ratio with a simple optimization utilizing the median predictor we obtained in the first step, and grease sample data.

3.3. Training of the hybrid model

In the previous studies [14,15], mean squared error (MSE) is used as the loss function to train the hybrid model (Fig. 4(a)). The benefit of MSE is convergence to the mean of sampled data, regardless of the distribution. On the contrary, this benefit can become a disadvantage when we are dealing with non-symmetrically distributed data. The remedy for these kind of data is using mean absolute error (MAE) as the loss function. As opposed to MSE, MAE tends to favor median instead of mean. Therefore, MAE can be utilized as loss if we aim to estimate the 50th quantile of any given distribution. The formulation for MAE as loss is given as:

$$\text{Loss} = \frac{1}{N_T N_O} \sum_{j=1}^{N_T} \sum_{i=1}^{N_O} \left| a_{GRS,ij} - \hat{a}_{S0,GRS,ij} \right|,$$

where $N_T$ is the number of turbines within the training set, $N_O$ is the number of observations for a single turbine, $a_{GRS,ij}$ is the $i$th observation of grease damage (from sample results) for $j$th turbine, and $\hat{a}_{S0,GRS,ij}$ is the median grease damage prediction for the $i$th grease sample of the $j$th turbine.

For this model we used a 5 layered architecture to represent grease damage increment. Layers from input to output consists of 40, 20, 10, 5, and 1 neurons respectively, with activation functions of sigmoid, elu, elu, elu, and sigmoid (see Appendix A for the
closed forms of activation functions); and in overall embodies 1,251 parameters to be trained. The model is trained with grease samples analyzed by detailed laboratory analysis (continuous damage scale) obtained by the end of every month for six months duration. Note that we will also investigate the effect of the number of sampled turbines. For this training we used RMSprop, with a learning rate 0.0005 and 200 epochs.

After training our model to learn the median behavior of grease quality (i.e. 50th quantile), for every single machine in our sampled fleet, we can use the discrepancy in between observations and estimated median to calculate a multiplication factor (i.e. quantile ratio) that will provide a mapping from the median to a quantile. Since we recognize that in a realization, turbines are equipped with a grease with certain quality, if we solve for a ratio for a single turbine, that ratio should provide a single sample from the grease quality distribution. Extending this procedure across the entire training set, we can reconstruct a distribution of quantile ratios.

We can also express the calculation of the quantile ratio as a simple optimization problem, where we find the multiplication factor that will minimize the MSE of observed grease damage and estimated median grease damage at observation time stamp. For every turbine in the training set we find \( C_k \) by solving the following optimization problem:

\[
\min_{C_k} \frac{1}{N_O} \sum_{i=1}^{N_O} \left( a_{GRS} - C_k \hat{a}_{GRS} \right)^2 , \quad \text{s.t. } C_k > 0
\]

where \( C_k \) is the quantile ratio for the given turbine, \( N_O \) is the number of observations for a single turbine, \( a_{GRS} \) is the \( i \)th observation of grease damage (from sample results), and \( \hat{a}_{GRS} \) is the predicted median grease damage for the \( i \)th grease sample.

One should notice that in a given realization as the number of turbines sampled are increased, more sample points we get to rebuild grease quality distribution. In addition to the validation of our modeling strategies, we will investigate the effect of number of sampled turbines on the model accuracy in Section 5.

4. Case study: on shore wind farm with 1.5 MW turbines

4.1. Turbine specifications and operational data

In this study, we chose a 1.5 MW wind turbine with 80 meters hub height as our reference machine. This turbine model is equipped with a spherical main bearing in the three-point mounting setting.

Table 1 provides some of the important parameters and specifications of the wind turbine and the bearing used throughout this work.

<table>
<thead>
<tr>
<th>Wind turbine</th>
<th>Main bearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rated power</td>
<td>1.5 MW</td>
</tr>
<tr>
<td>Cut-in wind speed</td>
<td>3.5 m/s</td>
</tr>
<tr>
<td>Cut-out wind speed</td>
<td>25 m/s</td>
</tr>
<tr>
<td>Maximum rotor speed</td>
<td>20 rpm</td>
</tr>
<tr>
<td>Hub height</td>
<td>80 m</td>
</tr>
</tbody>
</table>

4.2. Grease data

In this paper, we consider a thorough sampling and evaluation procedure that operators could conduct to quantify state of grease. Operators could organize inspection campaign, where grease that surrounds the main bearing of a number of machines are sampled, delivered to a laboratory to extract properties such as viscosity, humidity penetration, and foreign particle contamination. It is possible to lump these values into a degradation index, or as we call it here, grease damage.
In practical terms, operators do not have access to actual grease degradation data at the same frequency that SCADA data is recorded. Therefore, in this study, we generate grease degradation data using empirical life curves provided by manufacturers (see Fig. 1(c)). These curves are obtained through extensive coupon tests to provide an estimation for the grease. However, it should be noted that these curves are provided with different levels of confidence, which makes them probabilistic. The variability on these curves can easily be interpreted as quality variation. Ultimately, when an inspection campaign is conducted, the operators should expect such variability to influence the damage accumulation. Fig. 2(c), for instance, depicts two potential realizations of grease quality on same machines under same loading conditions. This variation is indeed what we will capture with our hybrid approach.

4.3. Replication of results

Simulations were conducted using a laptop configured with an Intel Core i7-8650U CPU at 1.90 GHz, 32 GB of RAM, and NVIDIA Quadro P500 GPU running Windows 10. Our implementation is all done in TensorFlow using the Python application programming interface (version 2.0). 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 [37]. Then, clone the “pinn_wind_bearing” repository found in [38] and go to folder “probabilistic_grease_inspection/mssp_2022”. The data used in this work is publicly available in [39]. Download the data and extract folders to the directory where the “pinn_wind_bearing/probabilistic_grease_inspection/mssp_2022” repository is cloned.

5. Results and discussion

First of all, we train our hybrid physics-informed neural networks model (Fig. 4(a)) with MAE loss, using grease samples obtained from 100 turbines every month for six months duration. We used 100 turbines to validate our modeling approach. In addition, we obtained samples for three different realizations to study the effect of different variations across the farm on our model.

In Fig. 5 we demonstrate the performance of our median predictions on different turbines. Fig. 5(a) is an example where grease quality used is better than median grease. Our model can successfully predict the median even though the observations indicate low grease damage. In Fig. 5(b), we observe a turbine with average loads equipped with a grease that has very poor quality. This is a good example to illustrate the high accuracy of our model on predicting the median, even for an extreme machine. Fig. 5(c) provides a turbine with very low loads but with poor grease condition, and the median prediction is almost perfect.

As the next step, we extract the quantile ratio distribution for all realizations using the technique detailed in this section. Fig. 6 depict the true empirical cumulative distribution function of the quantile ratios for given realization, along with the prediction of our method. For the realization # 1, we can clearly see a major agreement of the true distribution with the predicted distribution. In realizations # 2 and # 3, our model also provide high accuracy estimations, however they also indicate it is possible to marginally overestimate or underestimate the distribution, depending on the how well the grease quality distribution in the realization reflects the true distribution.

So far, we have demonstrated that if we provide enough data points, we can accurately predict the median behavior, and then rebuild the distribution for the quantile ratio. Now we will investigate the sensitivity of our approach to the number of sampled turbines. In Fig. 7 we showcase our median predictor model trained with 10, 20, 50, and 100 turbines on two different turbines that are not within our training set. From the aggressive turbine case in Fig. 7 we can clearly infer that as the number of turbines increased, the accuracy of median prediction is increased as expected. However, if we take a look at a turbine with mild loading conditions, we observe that all the models are on point regardless of the sampled data used to train those models. This can be simply explained by the flatness of the true median grease degradation that helps the model to learn almost linear behavior. We should also note that from a maintenance point of view, high accuracy damage estimation for aggressive turbines are more important.

---

Fig. 5. Performance of median predictor is illustrated for three different turbines for a realization.
Fig. 6. Empirical cumulative distribution of the quantile ratios predicted with models trained with samples from 100 turbines and using 3 different grease quality realization.

Fig. 7. Estimating median with models trained with different number of sampled turbines. Note that same realization is used for the same model across turbines.

Table 2
RMSE of the median grease damage predictions for the model trained with different number of sampled turbines, with multiple realization of the grease quality. Recall that grease quality realizations are different for each model (i.e. Realization #1 of the model trained with 10 turbines is different than the Realization #1 of the 100 turbine model).

<table>
<thead>
<tr>
<th>Sampled turbines</th>
<th>Realization #1</th>
<th>Realization #2</th>
<th>Realization #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.038</td>
<td>0.167</td>
<td>0.031</td>
</tr>
<tr>
<td>20</td>
<td>0.025</td>
<td>0.023</td>
<td>0.063</td>
</tr>
<tr>
<td>50</td>
<td>0.051</td>
<td>0.022</td>
<td>0.049</td>
</tr>
<tr>
<td>100</td>
<td>0.021</td>
<td>0.032</td>
<td>0.052</td>
</tr>
</tbody>
</table>

In Table 2 we again report the root mean squared error (RMSE) for median predictions obtained for three different realizations, this time for models trained with different number of sampled turbines. Even though the best performance is obtained with the most number of turbines (0.021 RMSE for Realization # 1 of model trained with 100 turbines), and the worst performance is observed with the least number of turbines (0.167 RMSE for Realization # 2 of model trained with 10 turbines), this does not necessarily prove that it is not possible to get accurate models with low number of sampled turbines. Naturally, an important parameter in a realization is how the quality is distributed with the loading conditions of each turbine (i.e. aggressive turbine getting poor quality grease, or vice versa), and unfortunately there is no practical way to know and keep track of it. Therefore, natural engineering intuition is to have as much samples as possible to increase the reliability of the model.

We also illustrate how the sampled turbine count affects the quantile ratio distribution prediction. Fig. 8 shows the evolution of the prediction of the empirical cumulative distribution function with the increasing number of sampled data points for a single realization. As expected and observed from Fig. 8(a), more data points we have to train our model, better we can rebuild the true distribution for the quantile ratio. It is also intuitive that more samples allow us to have more refined discretization of the probability distribution. Intuitively, the number of available samples impacts the ability to accurately estimate the cumulative distribution curve as a function quantile ratio (i.e., low systematic error). Moreover, it also affects the ability to provide precise estimators (i.e., low spread). In Fig. 8(b), we estimated the spread of the quantile ratio distribution using subsets of available turbines for sampling by using bootstrapping (drawing 90% subset of points with replacement many times over). For instance, on top subfigure, we used 18 out of 20 machine models and carried out the prediction for quantile ratios for all combinations of 18 out of 20 machines. Same procedure is also carried out for 50 machine model. The resulting curves estimate the uncertainty caused by the number of points. Practitioners could use convergence of such uncertainty to decide how many points are needed in the final model.
One should also realize that frequently sampling from large number of turbines within a wind farm can be impractical for operators. We suggest an efficient inspection campaign where grease state of recently regreased turbines are monitored for a short period of time (6 months), keeping in mind that number of turbines sampled have implications towards the accuracy of the model.

Finally, we combine our median predictor and quantile ratio distribution to estimate the 95% confidence interval of the grease damage propagation for a single validation turbine. In Fig. 9, we show two realizations for each model trained with different number of turbines. The bottom row of Fig. 9 illustrates the predictions of models trained with the realization that provided lowest RMSE in Table 2. We observe that almost all models can accurately estimate the median, however the confidence interval predictions for 10 and 20 turbine models are visibly poorer than 50 and 100 turbine models. The top row of Fig. 9 illustrates the predictions of models trained with the realization that provided highest RMSE in Table 2. Bad realizations yield to poor median predictions. It also applies to confidence interval predictions. One exception is the model trained with 100 turbines, which provides decent approximation to true confidence intervals. Therefore, we can confidently state that even with a bad distribution of grease quality across the farm, sampling enough number of turbines can provide better uncertainty quantification for grease quality variation.

6. Summary and closing remarks

The hybrid physics-informed neural networks approach presented in this contribution proved to be accurate with a low amount of data used for training.

Table 3 compares our approach with other methods that utilize machine learning, such as supervised machine learning and physics-informed machine learning. In Appendix B, we used a pure data-driven supervised machine learning method to model grease degradation. Results show that the long short-term memory cell was inadequate to approximate the actual degradation mechanism, due to strong unbalance of dataset (many more input than output observations) and unawareness of physical constraints. On the other hand, physics-informed machine learning, as a method to solve partial differential equations by using collocations points across the solution space, is incompatible to solve our problem since governing equations for grease degradation are unknown.

In this paper, we demonstrated a two-step approach to quantify uncertainty in a grease degradation model. Our first step was to utilize a hybrid physics-informed neural network model for main bearing fatigue to learn median behavior of grease using a grease samples collected from a number of turbines. We observed that modifying loss function to favor the median of the samples, our model can accurately estimate the median grease degradation for each machine. Then, we used trained model as median predictor and estimated a quantile ratio that maps the median grease into a certain quantile in grease quality distribution. With that, we were able to successfully rebuild the quality distribution.

We studied the effect of randomization in different realizations of quality distribution. We have showed that depending on the realization used for training, reconstructed distribution can marginally overestimate or underestimate true distribution.
Moreover, we analyzed the effect of number of inspected turbines on the performance of model. We have displayed that, more machines sampled indicate more data for the framework, and naturally implies improved performance on median prediction. The same conclusion can also be drawn for quality distribution predictions, since more sampled wind turbines would suggest a finer discretization of the probability distribution for quantile ratios.

Here, we addressed the quantification of grease quality uncertainty with the hybrid model. We firmly believe that this study can be extended for further exploration of the benefits of the framework towards other sources of uncertainties. For instance, the loads model used in this study consider normal operation, and disregard extreme conditions or complex loading schemes (i.e. startups and shut-downs, emergency brakes, yaw misalignment). On the field, operators can make use of information obtained from proximity sensors to compensate for limitations in the load model, which can also be provided to our framework to improve the fidelity of the loads model. In addition, we know the model is also robust towards inspection noise [15]. Therefore, we can explore the simultaneous quantification of multiple uncertainties (i.e. grease quality and sample noise). In this contribution, we focused on machines installed on a single site. This implies that conditions such ambient temperature and humidity can be similar across the park (although wind speed might differ due to topology and turbine placement). Provided that the data is gathered such that mechanical loads and bearing temperatures cover the range of operation and that enough samples are collected to account for the interplay between operation and grease quality, we believe that the proposed models would be useful. This seems to be true, at least at the site level, as shown in results. In addition, as future research, we believe the method can be generalized and extended to multiple wind farms that have different environmental conditions.

Table 3
Comparison of different approaches that utilize machine learning.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Supervised machine learning</th>
<th>Physics-informed machine learning</th>
<th>Hybrid physics-informed neural networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory</td>
<td>Data-driven methods that map input to labeled data.</td>
<td>Collocation methods to solve partial differential equations.</td>
<td>Graph models where physics and data-driven nodes can co-exist.</td>
</tr>
<tr>
<td>Data needs</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Less accurate</td>
<td>Accurate</td>
<td>Accurate</td>
</tr>
<tr>
<td>Computational cost</td>
<td>Expensive</td>
<td>Intermediate</td>
<td>Cheap</td>
</tr>
</tbody>
</table>

Fig. 9. Grease damage prediction evolution with the number of training turbines (per columns of the subplot). Shown predictions are for a single turbine that is not in the training set. This additional turbine was not used in the training of the hybrid model. Black lines show true median and 95% confidence interval behaviors, and colored lines are predictions for median and 95% confidence intervals. We also illustrate one bad and one good realization for each model (per rows).
Table 4
Long short-term memory (LSTM) network designs.

<table>
<thead>
<tr>
<th>Design</th>
<th>Layers</th>
<th>Neurons</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow LSTM</td>
<td>1</td>
<td>8</td>
<td>360</td>
</tr>
<tr>
<td>Deep LSTM</td>
<td>3</td>
<td>8</td>
<td>1,448</td>
</tr>
</tbody>
</table>

Fig. 10. Time history prediction performance of LSTM models on grease damage propagation.

CRediT authorship contribution statement

Yigit A. Yucesan: Methodology, Software, Formal analysis, Investigation, Data curation, Writing, Visualization. Felipe A.C. Viana: Conceptualization, Methodology, Validation, Software, Formal analysis, Investigation, Writing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Multi-layer perceptrons

In a multi-layer perceptron, each layer can have one or more perceptrons (nodes in the graph). A perceptron applies a linear combination to the input variables followed by an activation function

\[ v = f(z) \quad \text{and} \quad z = w^T u + b, \]

where \( v \) is the perceptron output; \( u \) are the inputs; \( w \) and \( b \) are the perceptron hyperparameters; and \( f(.) \) is the activation function. Throughout this paper, we used the hyperbolic tangent (tanh), sigmoid and the exponential linear unit (elu) activation functions (although others could also be used, such as the rectified exponential linear unit):

\[
\begin{align*}
\text{tanh}(z) &= \frac{e^z - e^{-z}}{e^z + e^{-z}}, \\
\text{sigmoid}(z) &= \frac{1}{1 + e^{-z}}, \\
\text{elu}(z) &= \begin{cases} 
  z & \text{when } z > 0, \\
  e^z - 1 & \text{otherwise}.
\end{cases}
\end{align*}
\]

The choice of number of layers, number of neurons in each layer, and activation functions 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 [40–42] for optimization of the data-driven portions of the model.

Appendix B. Conventional recurrent neural network results

The long short-term memory (LSTM) cell [28,29] illustrated in Section 3 and Fig. 3(a) carries out state transition and used to model time-series problems with a pure data-driven framework. In this study, we use an LSTM cell to model main bearing fatigue damage accumulation and estimate grease degradation to compare against the novel hybrid approach proposed in this article.

We designed two different networks with different complexity levels. While the first model is a single layer architecture, which we call “shallow LSTM”; the second model has multiple layers and we call “deep LSTM”. Table 4 summarizes the architectural details for these models. We use these two configurations to provide an assessment of the number of parameters in the prediction of the models. Fig. 10 illustrates the performance of LSTM models after training is performed using the training set of 10 machines.
Unfortunately, predictions are neither aligned with the observations nor are monotonic, as one would desire.

We observe that using a deeper architecture only slightly improves the results. Unfortunately, the LSTM models perform poorly and do not properly approximate the time history of grease damage.

References


