

# An improved-accuracy method for fatigue load analysis of wind turbine gearbox based on SCADA



Eduardo J. Alvarez <sup>a, b</sup>, Adrijan P. Ribaric <sup>a, \*</sup>

<sup>a</sup> Research and Development, Sentient Science Corporation, Idaho, United States

<sup>b</sup> Mechanical Engineering Department, Brigham Young University Idaho, Idaho, United States

## ARTICLE INFO

### Article history:

Received 30 November 2016  
Received in revised form  
18 July 2017  
Accepted 14 August 2017  
Available online 30 August 2017

### Keywords:

Wind turbine  
Gearbox failure  
Load duration distribution  
Torque  
SCADA  
Physics-based prognostic

## ABSTRACT

In spite of their increasing popularity, managing the use of wind turbines has been exceptionally challenging. Through computational prognostics, Sentient Science determined that current operating lifetime for a large number of turbines is only between five to thirteen years. Initial estimates indicate that savings of \$150,000 per turbine per gearbox replacement can be achieved using physics-based long-term prognostics, leading to a substantial return of investment for wind farm operators. However, long-term prognostics require a precise determination of the loads in all six degrees of freedom occurred on the drive-train. One of these loads—torque—can be directly estimated in situ from the historical data provided by the Supervisory Control and Data Acquisition (SCADA) system. In many cases, the historical data only provides 10-min statistical values, and a common practice of reliability analysts is the calculation of torque using only 10-min averages. Disregarding the load fluctuation within 10-min intervals of recorded SCADA introduces a loss of accuracy in the resulting torque histogram that is indeed meaningful for an accurate life prognostic. This paper introduces a novel improved-accuracy method for calculation of torque histograms based on SCADA. Using 10-min distributions of power output and rotor speed, this method is able to successfully reconstruct the distribution of instantaneous torque in between 10-min intervals of recorded SCADA. The method predicts a high-torque region more dispersed than the current method used in the industry, which introduces substantially different results when used in life prognostics. Using this method in the lifing of a GE 1.5 SLE wind turbine, it is shown that the error in predicted L50 is reduced by 10.1%.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Wind energy continues to grow in popularity as a source of renewable energy. The U.S. Energy Information Administration reports that, in 2016, 34% of all new U.S. energy generating capacity was attributed to wind turbines [1]. However, managing the use of wind turbines has been exceptionally challenging. Through computational prognostics, Sentient Science determined that current operating lifetime for a large number of turbines is only between five to thirteen years. This is much lower than the initial expectation of lasting between 20 and 30 years. One of the major components that experience premature failure is the gearbox, with replacement costs that can range from \$200,000 to \$700,000 per failure [2,3]. This means that wind farm operators can readily incur

costs up to millions of dollars dealing with turbine failures.

In order to avoid critical failure and extend the life of the asset, a significant research effort has been undertaken in the development of methods for diagnostics and prognostics of wind turbine components. A common practice in the wind industry is the use of a diagnostics-only technology known as Condition-Based-Monitoring (CBM) to evaluate the current as-is health state of their assets. The methodology of CBM revolves around the concept of fault detection by identifying patterns in sensor signal associated to specific failure modes [4,5]. For instance, Kusiak and Verma [6] used a data mining approach on temperature data to predict bearing fault 1.5 h before occurrence; for gearbox diagnosis, several vibration analysis techniques have been developed to detect fault on sun and planet gears [7–9]. Providing a diagnostic of the current health state, CBM systems warn operators of the need of maintenance only after fault has initiated. Some CBM systems include data-driven prognostic technology as well [10], but these systems are able to predict the health state of a turbine up to a three-month time frame only.

\* Corresponding author. 1000 Riverwalk Drive, Suite 375, Idaho Falls, ID 83402, United States.

E-mail address: [aribaric@sentientscience.com](mailto:aribaric@sentientscience.com) (A.P. Ribaric).

Nomenclature	
<i>Physical parameters</i>	
$\tau$	Torque (kNm)
$\mathbf{P}$	Distribution of power (kW) over a 10-min interval
$\mu_P$	Mean power (kW) over a 10-min interval
$\sigma_P$	Standard deviation of power (kW) over a 10-min interval
$\omega$	Distribution of rotor speed (RPM) over a 10-min interval
$\mu_\omega$	Mean rotor speed (RPM) over a 10-min interval
$\sigma_\omega$	Standard deviation of rotor speed (RPM) over a 10-min interval
$\rho$	Correlation between $\mathbf{P}$ and $\omega$
$D_{res}(p \parallel q)$	Residual between profiles $p$ and $q$
<i>Mathematical parameters</i>	
$z, w$	Jointly normal variables
$\mu_z, \mu_w$	Mean values of distributions $z$ and $w$
$\sigma_z, \sigma_w$	Standard deviations of distributions $z$ and $w$
$\rho$	Correlation between $z$ and $w$
$x, y$	Independent standard normal variables
$a, b, s$	Translation parameters
$r$	Rescaling parameter
$U = \frac{z}{w}$	Distribution of the ratio of normal variables
$T = \frac{a+x}{b+y}$	Distribution of the translated and rescaled ratio of normal variables
$f_U, f_T$	Density functions of $U$ and $T$
$\phi(t) = \frac{t}{r} + s$	Counter-rescaling, counter-translating function

In contrast to the fleeting predictions done with CBM, long-term predictions can be achieved using physics-based approaches. Physics-based prognostics (also called model-driven prognostics) combine physics of failure with measured data to calculate damage accumulation and predict remaining useful life [11,12]. The general methodology can be described as follows:

- 1. System analysis:** The system is defined by detailing each component's design and material specifications. Based on failure reports, theory, and engineering experience, most-likely failure modes are identified and corresponding components are chosen for subsequent lifing.
- 2. Determining load spectra:** Loads driving each failure mode need to be determined. In the design stage of a wind turbine, loads are taken from standardized design load cases. For asset-specific prognostics, loads in all degrees of freedom can be accurately determined using aeroelastic models that consider the historical wind and operation data recorded at the installed location. Alternatively, the torque—which drives gearbox failure—can be directly estimated in situ from historical SCADA data.
- 3. Damage model:** For each component and corresponding failure mode, the rate of damage accumulation due to loads is calculated through a damage model that describes system dynamics and damage kinetics.
- 4. Life prognostic:** Using a Weibull reliability model, damage accumulation on each failure mode is converted into failure probability, and reliability values are used to estimate the expected life of each component.

In past years, physics-based models for common failure modes in a variety of rotating machinery have been developed. Li et al. developed models of micro-pitting [13] and scuffing [14] in lubricated point contacts and applied them to a ball-on-disk contact. Watson et al. [15] used a wear model to predict life of high-power clutch systems. Kacprzyński et al. [16] built a crack initiation model to predict remaining useful life in a H-60 helicopter's gear. Marble and Morton [17] used a model to compute spall growth in an aircraft's engine thrust bearing, concluding a window of about 100 flight hours to critical failure after initial fault; to this prognostic method Bolander et al. [18] added diagnostics to detect the presence of a spall and reduce uncertainty in remaining useful life. The epitome of physics-based prognostics technology is DigitalClone™, a material-science-based simulation engine that predicts the earliest point in time when cracks initiate in the microstructure of rotating mechanical components [19,20].

In wind turbines, using physics-based prognostics technology it is possible to identify the specific component and point in time where and when failure will initiate, allowing to implement the operation strategy or provide timely maintenance that will extend the life of the turbine. Sentient's initial estimates indicate that savings of \$150,000 per turbine per gearbox replacement can be achieved using physics-based long-term prognostics, leading to a substantial return of investment for wind farm operators. However, long-term prognostics require a precise determination of the loads occurred on the drive-train.

Determining the actual loads experienced by the turbine can be challenging. In the design stage of a turbine, load spectra are taken from standardized design load cases—as defined by International Electrotechnical Commission standards (IEC 61400-22) and/or the guidelines and standards defined by Det Norske Veritas and Germanischer Lloyd (DNVGL-ST-0437)—to certificate the turbine. However, these standardized load cases cannot be used for asset-specific life prediction since they are valid only as a generalization and may not represent the environmental conditions that the turbine is exposed to after commissioning. Once the turbine is installed, the actual loads in all six degrees of freedom can be determined using aeroelastic models that consider the turbine historical wind and operation data captured at the installed location. However, this approach is often avoided by reliability analysts due to its complexity. Alternatively, one of these loads—torque—can be directly estimated in situ from historical SCADA data.

Torque is a major damage driver and therefore can be used as a first start for lifing of the gearbox and other components. The torque can be derived from instantaneous measurements of rotor speed and power output. These measurements are recorded by the turbine's Supervisory Control and Data Acquisition (SCADA) system. For storage reasons, it is a common practice on SCADA systems to record only 10-min-based statistical data, making available mean, standard deviation, and min/max values of rotor speed and power output every 10-min intervals. The historical distribution of torque can be empirically built by agglomerating large volumes of SCADA-based torque estimations into a histogram, resulting in a distribution as the one shown in Fig. 1, which was built on six years of SCADA collected in a GE 1.5 SLE turbine. In the torque histogram, the endurance limit defines a threshold of loads that contribute to failure in gearbox components. Theoretically, loads under the endurance limit do not contribute to component failure, meanwhile loads over the endurance limit incur in damage accumulation that eventually exhausts the life of the component; thus, accurately assessing the frequency a turbine operates in the high-torque

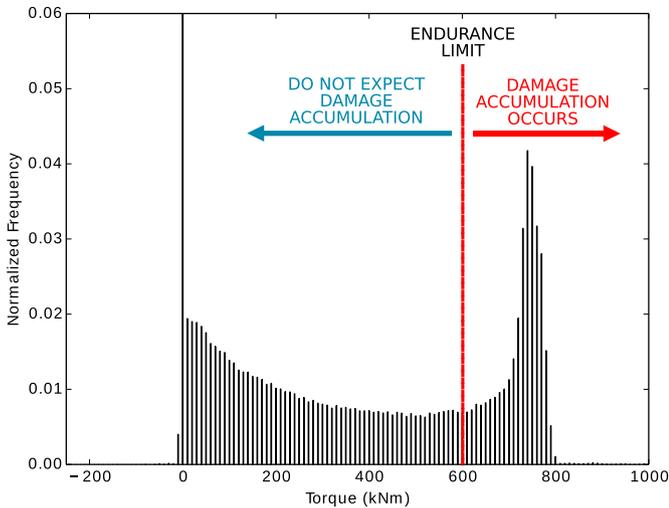


Fig. 1. Torque histogram of a GE 1.5 SLE turbine built on six years of SCADA data using 10-min averages.

region is critical for an accurate life prediction.

A common practice in industry and academia is the calculation of torque by dividing mean values of power output and rotor speed as recorded in SCADA every 10 min. For instance, Gray and Watson [21] used this method to calculate the load in a gearbox bearing damage model, and Al-Tubi et al. [22] built a torque histogram through this method, which was used to derive probabilistic distributions of shaft torque of the gearbox HSS pinion gear and assess risk of micro-pitting. Even though practical, this approach for estimating torque disregards the fact that the information in SCADA represents statistical distributions rather than instantaneous measurements. The physical measurements being recorded fluctuate in between the 10 min of recorded SCADA, and the instantaneous torque can be expected to fluctuate accordingly. Disregarding this fluctuation introduces a loss of accuracy in the resulting torque histogram that is indeed meaningful for an accurate life prognostic.

Considering the importance of accuracy in the loads used for physics-based prognostics, this paper introduces an improved-accuracy method for calculation of torque histograms based on historical 10-min-based SCADA data. In doing so, the current method used in the industry—hereafter called Method 1—will first be described. Such method disregards any fluctuations in between 10-min intervals of recorded SCADA and calculates the instantaneous torque by dividing mean power output and mean rotor speed. The method proposed in this paper—hereafter called Method 2—uses both 10-min distributions of power and rotor speed to reconstruct the 10-min distribution of instantaneous torque. Considering the torque as the ratio of normal variables power and rotor speed, a statistical model will be developed based on Marsaglia’s method for ratios of normal variables [23,24]. Validation will be provided by comparing resulting torque histograms to a histogram built on instantaneous measurements obtained from high-frequency SCADA sampled at 1 Hz. Finally, the improved lifing accuracy will be shown by independently using these three histograms (Method-1, Method-2, and 1-Hz histograms) in a life prognostic on gearbox components of a GE 1.5 SLE wind turbine.

## 2. Methods

### 2.1. Method 1: simplified torque

Disregarding any fluctuations in between each 10 min of

recorded SCADA, the instantaneous torque is calculated as

$$\tau = \frac{\mu_p}{2\pi\mu_\omega/60} \quad (1)$$

with  $\mu_p$  and  $\mu_\omega$  mean power (kW) and mean rotor speed (RPM) over such interval. The torque histogram is then built by agglomerating each  $\tau$  value calculated over each SCADA entry. Fig. 1 shows a histogram built using this method over six years of historical SCADA data collected on a GE 1.5 SLE turbine. This method is widely used in industry for reliability analysis, and it can be seen implemented in the work of Gray and Watson [21], and Al-Tubi et al. [22], as previously mentioned.

### 2.2. Method 2: distributed torque

The fluctuation of torque in between each 10 min of recorded SCADA can be determined dividing the 10-min distribution of power output by the distribution of rotor speed, resulting in a probability distribution of torque values for each 10-min interval of recorded SCADA. This is depicted conceptually as

$$\tau = \frac{P}{\omega} = \tau$$

Thus, this method calculates the torque as the ratio of random variables  $P$  and  $\omega$ . The following paragraphs are spent formalizing this concept.

In 1965, Marsaglia encountered an application in medicine in which estimating the life span of red cells depended on the calculation of a ratio of normal variables, and he developed a statistical model we will now adapt for the calculation of the 10-min torque distribution. As proposed by Marsaglia [23,24], the ratio  $\frac{z}{w}$  of two jointly normal variates  $z, w$  with correlation  $\rho$ —after translation and rescaling—can be described by the ratio  $(a + x)/(b + y)$ , with  $x, y$  independent standard normal variables and  $a, b$  non-negative constants. The proposition is stated as follows:

$$“r\left(\frac{z}{w} - s\right) \text{ is distributed as } \frac{a+x}{b+y} \text{ and } \frac{z}{w} \text{ is distributed as } \frac{1}{r}\left(\frac{a+x}{b+y}\right) + s” \quad [24]$$

with

$$\begin{aligned} \bullet r &= \frac{\sigma_w}{\pm\sigma_z\sqrt{1-\rho^2}}, s = \frac{\rho\sigma_z}{\sigma_w} \\ \bullet b &= \frac{\mu_w}{\sigma_w}, a = \pm\frac{\mu_z/\sigma_z - \rho\mu_w/\sigma_w}{\sqrt{1-\rho^2}} \end{aligned}$$

Letting  $T, U$  be  $(a + x)/(b + y)$  and  $\frac{z}{w}$ , respectively, Marsaglia determines the density function of  $T$  to be

$$f_T(t) = \frac{e^{-\frac{1}{2}(a^2+b^2)}}{\pi(1+t^2)} \left( 1 + qe^{\frac{1}{2}q^2} \int_0^q e^{-\frac{1}{2}x^2} dx \right) \quad (2)$$

with  $q = \frac{b+at}{\sqrt{1+t^2}}$ .

Defining  $u = \phi(t) = \frac{t}{r} + s$  and  $t = \phi^{-1}(u) = (u - s)r$ , and applying the theorem of change of variable to  $f_T(t)$ , the density function of  $\frac{z}{w}$  (named  $f_U(u)$ ) is determined to be

$$f_U(u) = |r|f_T((u - s)r) \quad (3)$$

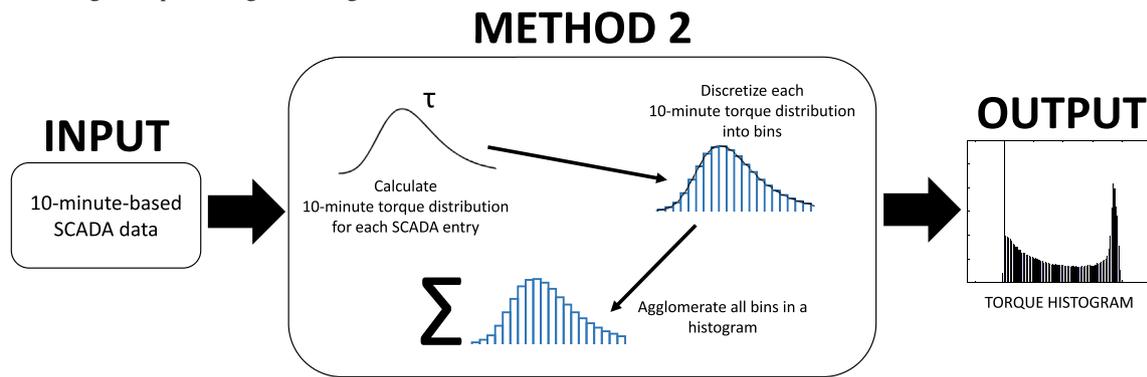
Thus, the 10-min torque distribution will be calculated as the ratio of normal random variables power (kW) and rotor speed (RPM)— $\mathbf{P}$  and  $\omega$ , respectively—as

$$\tau = \frac{\mathbf{P}}{2\pi\omega/60} \quad (4)$$

On the underlying assumption of  $\mathbf{P}$  and  $\omega$  to be jointly normally distributed,  $\mathbf{P}$  and  $2\pi\omega/60$  will take the place of  $z$  and  $w$ , respectively, when calculating parameters  $a, b, r, s$ . The torque probability density function  $f_\tau$  will then be calculated as  $f_u$  in Eq. (3) (methods for numerical implementation are described in the Appendix), with the following parameters obtained from historical SCADA data:

- $\mu_P$ : Mean power (kW) over a 10-min interval
- $\mu_\omega$ : Mean rotor speed (RPM) over a 10-min interval
- $\sigma_P$ : Standard deviation of power (kW) over a 10-min interval
- $\sigma_\omega$ : Standard deviation of rotor speed (RPM) over a 10-min interval
- $\rho$ : Correlation between  $P$  and  $\omega$  in such interval

Finally, the torque histogram is built by calculating the 10-min torque distribution over each SCADA entry, discretizing each distribution in bins as shown in Fig. 2, and agglomerating all the bins in a histogram. The following flowchart summarizes the overall procedure for calculating a torque histogram using Method 2:



The limitations of this method reside in the assumption that power output and rotor speed are jointly distributed forming a bivariate distribution, that both are normally distributed in between each 10-min interval of recorded SCADA, and that their correlation  $\rho$  can be accurately estimated. The validity of these assumptions will be evidenced later in this paper upon comparing the resulting histogram with high-frequency SCADA data (Section 3.2).

2.2.1. Power and rotor speed correlation

Since the correlation  $\rho$  between random variables  $P$  and  $\omega$  in Eq. (2) is unknown, different assumptions will here be explored. This parameter represents the mechanical relation between power output and rotor speed throughout operation of the turbine. Constructing a bivariate histogram of power output and rotor speed evidences that the correlation between these two variables changes during different stages of operation. Fig. 3 shows the bivariate histogram of two years of SCADA data obtained from a GE 1.5 SLE turbine (bins of idle operation have been omitted). Three stages of operation with distinctive correlation values can be identified: A region of uncorrelation ( $\rho = 0$ ) while operating close to the cut-in wind speed ( $P < 150$  or  $\omega < 11.25$ ), a region of direct correlation ( $\rho \approx 1$ ) while ramping-up passed the cut-in wind speed ( $150 > P > 800$  or  $11.25 < \omega < 17.75$ ), and a transition back to uncorrelation while approaching rated performance ( $P > 800$  or

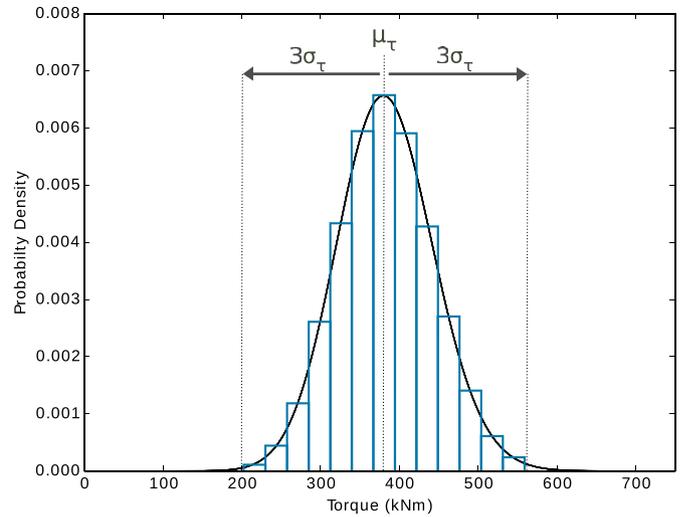


Fig. 2. Discretization of a 10-min torque distribution.

$\omega > 17.75$ ). Thus,  $\rho$  is dynamic and its current value needs to be determined in order to calculate each 10-min torque distribution. For each SCADA entry,  $\rho$  can be determined by identifying the

current operation stage of the turbine, which requires building and exploring the bivariate power-speed histogram of the turbine being

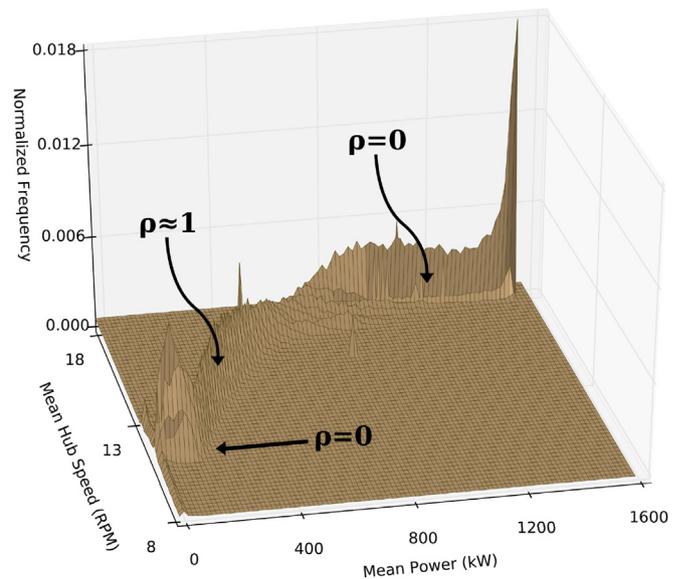


Fig. 3. Speed-power histogram of a GE 1.5 SLE turbine.

assessed. This process could be avoided by identifying a constant  $\rho$  value that produces a similar torque histogram than when  $\rho$  is calculated dynamically.

In order to identify a constant  $\rho$  value that could replace the need for calculating  $\rho$  dynamically, three constant  $\rho$  values (0, 0.5, and 0.85) and a dynamic  $\rho$  were independently used to build torque histograms on the same SCADA dataset, and resulting histograms can be seen in Fig. 4 ( $\rho = 0.5$  is omitted). Mean torque values for cases 0, 0.5, 0.85, and dynamic, were respectively 326.3, 329.6, 331.4, and 326.1 kNm, differing from the dynamic case on 0.1% for  $\rho = 0$ , 1.1% for  $\rho = 0.5$ , and 1.6% for  $\rho = 0.85$ . Using the  $D_{res}$  criteria (explained in Section 2.3), it was determined a profile discrepancy of 10.8% between the dynamic case and  $\rho = 0.85$ , contrasted to a discrepancy of only 0.8% for  $\rho = 0.0$ . Thus, it is concluded that  $\rho = 0$  models the dynamic case with sufficient accuracy. This is explained by noticing from Fig. 3 that a turbine is found in the ramp-up region ( $\rho \approx 1$ ) only sporadically, operating most frequently either close to the cut-in wind speed, or approaching rated performance. Subsequently, in this paper all evaluations of Eq. (2) will be done with  $\rho = 0$ .

### 2.2.2. Sensitivity analysis

In order to understand how parameters  $\mu_P$ ,  $\mu_\omega$ ,  $\sigma_P$ ,  $\sigma_\omega$ , and  $\rho$  will affect the 10-min torque distribution given by Eq. (2), a sensitivity analysis is here performed. Defining the default case  $\mu_P = 600$  kW,  $\mu_\omega = 15$  RPM,  $\sigma_P = 90$  kW,  $\sigma_\omega = 0.8$  RPM,  $\rho = 0$ , the parameters were varied in the following ranges:  $200 < \mu_P < 1400$  kW,  $9 < \mu_\omega < 21$  RPM,  $50 < \sigma_P < 130$  kW,  $0.8 < \sigma_\omega < 3.2$  RPM, and  $0 < \rho < 0.90$ . The resulting torque distributions are shown in Fig. 5. In Figs. 5a, 5c, and 5d, the dispersion increased with larger  $\mu_P$ ,  $\sigma_P$ , and  $\sigma_\omega$  values, and in Figs. 5b and 5e the dispersion increased as  $\mu_\omega$  and  $\rho$  decreased. The distribution deviated from a normal distribution attaining a positive skewness for high values of  $\sigma_\omega$  (see Fig. 5d), and the same is hinted for low values of  $\mu_\omega$  (see Fig. 5b). The most relevant observation is that the torque distribution gets more dispersed at greater values of mean torque, as seen in Figs. 5a and 5b. This observation will later on explain a major difference in the high-torque region between histograms obtained through Method 1 and Method 2, which will be relevant for lifing analysis.

### 2.3. Comparing histograms

In order to quantify the discrepancy between histograms built through the different methods, the following distance measure  $D_{res}$  will be used:

$$D_{res}(p \parallel q) = \frac{1}{2} \sum_i |p_{\tau_i} - q_{\tau_i}| \quad (5)$$

where  $p$  and  $q$  are the histograms being compared,  $i$  the index of each bin, and  $p_{\tau_i}$  is the normalized frequency of the bin corresponding to torque  $\tau_i$  in  $p$ .  $D_{res}(p \parallel q)$ , named "residual between  $p$  and  $q$ ", provides a scale 0 to 1 of discrepancy between profiles, where a residual of 0 indicates identical profiles (or 0% discrepancy), and 1 indicates perfectly dissimilar profiles (or 100% discrepancy).

## 3. Results

### 3.1. Comparison of method 1 and method 2

Method 1 and Method 2 were independently used to construct the torque histograms of four GE 1.5 SLE turbines over 24 months of SCADA data. The resulting histograms are compared in Fig. 6 and

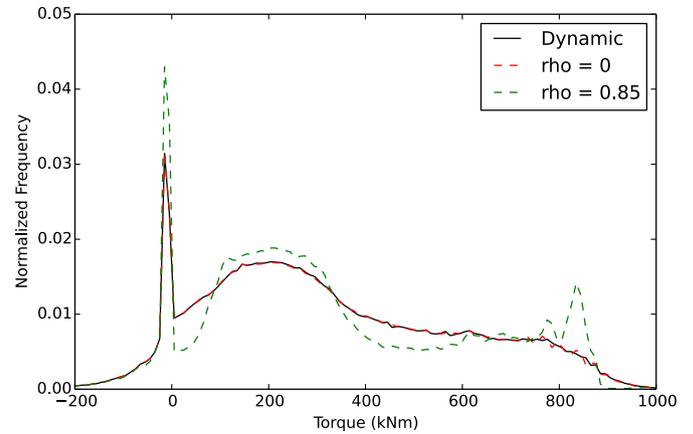


Fig. 4. Comparison of torque histograms obtained through different correlation  $\rho$  values using Method 2.

Table 1. The histograms are almost identical in low and mid-torque regions, but a dispersed high-torque region is observed in histograms calculated through Method 2. This can be explained by the fact that Method 2 accounts for fluctuations of torque in each 10-min interval, and that 10-min torque distributions tend to be more dispersed with a higher mean torque, as previously evidenced in the sensitivity analysis (Section 2.2.2). Method 2 resulted in mean torques 7.2%–9.0% higher than Method 1, and profile discrepancies ( $D_{res}$ ) between 5.4% and 6.0%, which is attributed to the high-torque region. A second dataset obtained from a different site (three GE 1.5 SLE turbines with 32 months of SCADA) displayed similar dispersions in the high-torque region, resulting in mean torques 4.8%–9.6% higher with Method 2, and profile discrepancies between 9.0% and 14.8%.

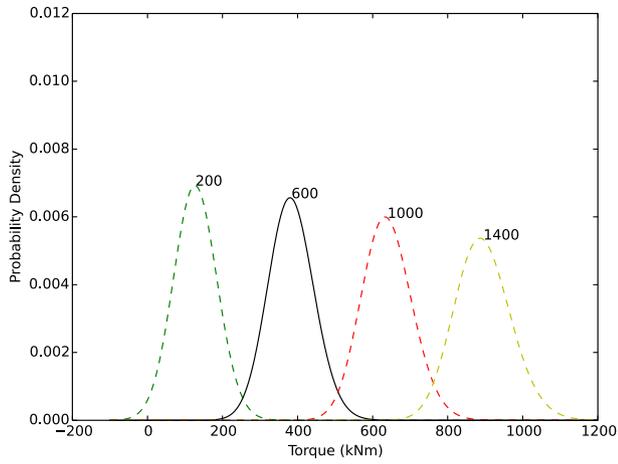
Method 2 conclusively shows a substantial difference characterizing the high-torque region. The damage accumulated in gearbox components is directly related to the frequency a turbine operates in the high-torque region; therefore, accurately assessing this region is critical for an accurate life prediction. In consequence, if Method 2 correctly models the true torque distribution, the high-torque dispersion captured through this method gives a strong reason to prefer it in lifing.

### 3.2. Validation

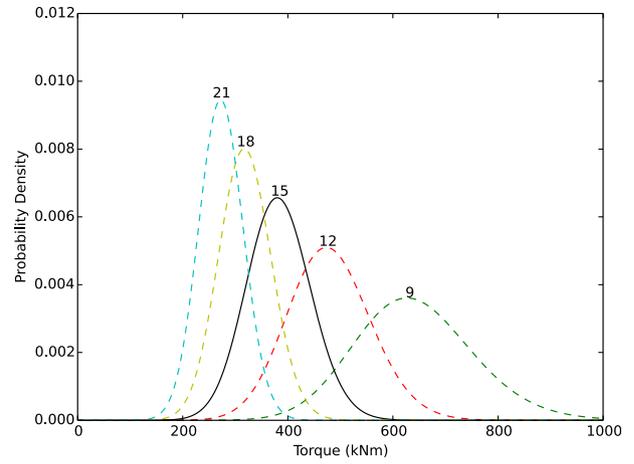
In order to validate the statistical model that Method 2 implements, instantaneous measurements of power output and rotor speed on a GE 1.5 SLE turbine<sup>1</sup> were recorded every 1 s for twelve months along with 10-min-based SCADA data. With this data, the true torque distribution was built directly from the empirical instantaneous torque. Fig. 7 compares the true torque distribution to the histograms obtained through Method 1 and Method 2, showing that the dispersed high-torque region predicted by Method 2 is ratified by the true distribution. With a profile discrepancy of 4% between Method 2 and the true distribution,<sup>2</sup>

<sup>1</sup> This turbine had been updated to 1600 kW throughout the twelve months of operation.

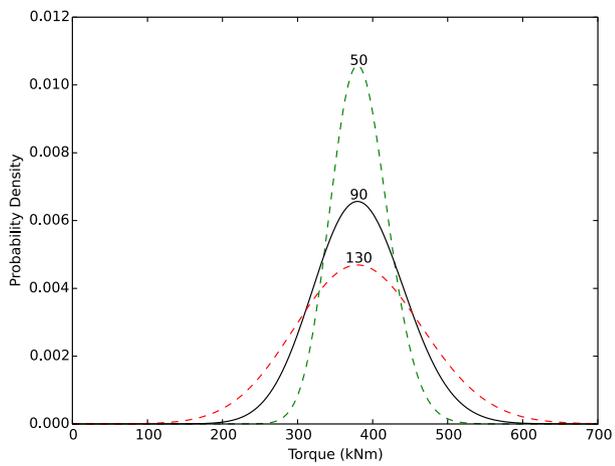
<sup>2</sup> Mean torque values of the true distribution, Method 1, and Method 2 were respectively 446, 463, and 467 kNm, differing from the true distribution on 3.8% for Method 1 and 4.6% for Method 2. This overshoot of mean torque is attributed to an underpredicted low-torque region in both methods. Method 2 resulted in a profile discrepancy of 4% to the true distribution due to the disagreement in the low-torque region, meanwhile Method 1 resulted in a discrepancy of 10% due to disagreement in both low and high-torque regions.



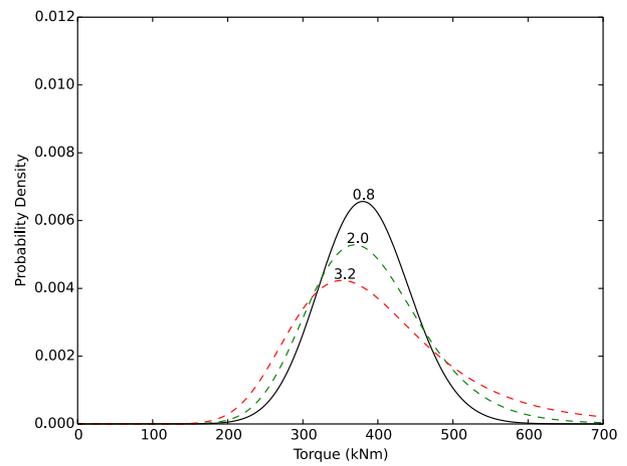
(a)  $200 < \mu_P < 1400$  kW



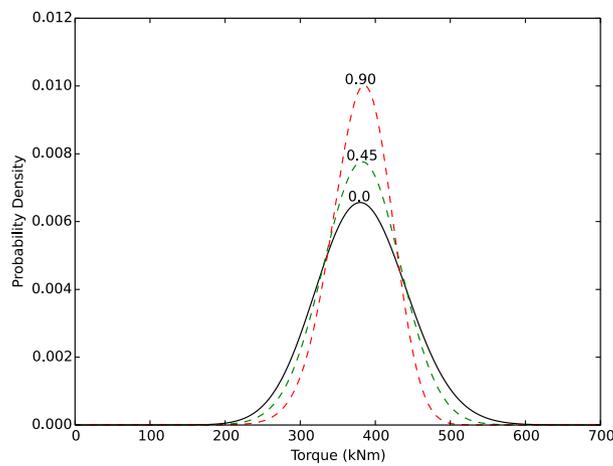
(b)  $9 < \mu_\omega < 21$  RPM



(c)  $50 < \sigma_P < 130$  kW



(d)  $0.8 < \sigma_\omega < 3.2$  RPM



(e)  $0 < \rho < 0.90$

**Fig. 5.** Effects of parameters  $\mu_P$ ,  $\mu_\omega$ ,  $\sigma_P$ ,  $\sigma_\omega$ , and  $\rho$  on the 10-min torque distribution. Solid black line corresponds to the default case.

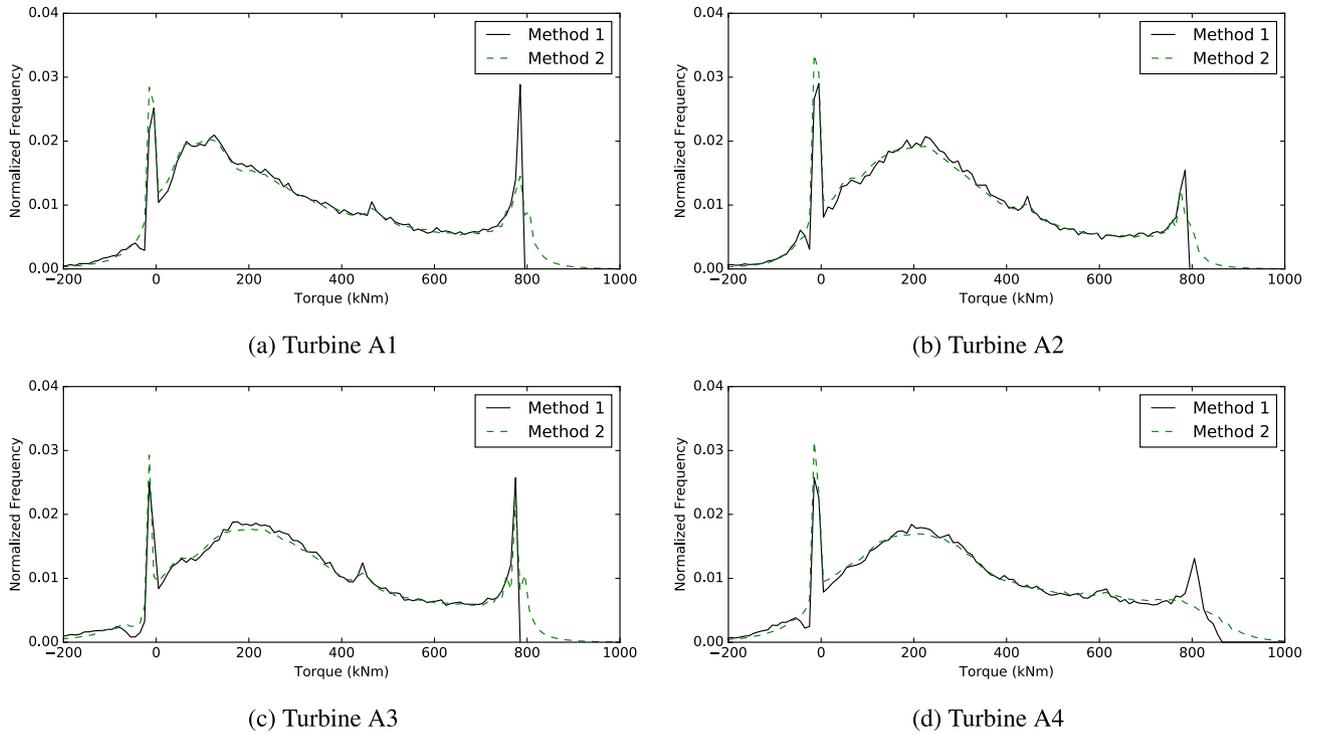


Fig. 6. Comparison of torque histograms obtained through Method 1 and Method 2 over 24 months of SCADA data collected on GE 1.5 SLE turbines.

Table 1  
Comparison of torque histograms obtained through Method 1 and Method 2.

	Mean Torque (kNm)		Mean Torque Difference	Profile Discrepancy ( $D_{res}$ )
	Method 1	Method 2		
Turbine A1	278	301	7.3%	6.0%
Turbine A2	271	293	8.1%	5.4%
Turbine A3	293	319	9.0%	6.0%
Turbine A4	304	326	7.2%	5.8%

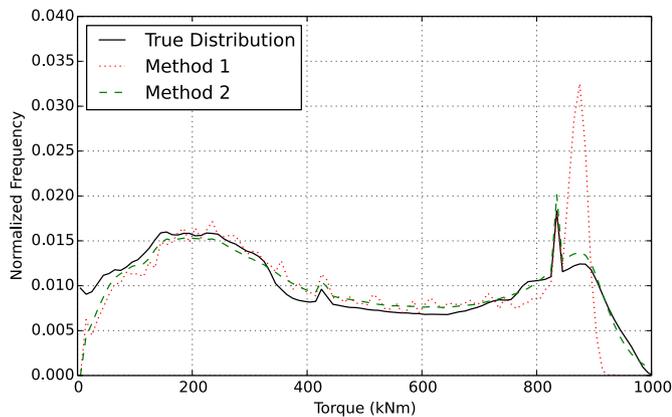


Fig. 7. Comparison of torque histograms obtained through Method 1 and Method 2 to the true torque distribution (values of idle operation have been discarded).

this agreement proves that the statistical model is able to recover information that was lost in the process of packing instantaneous measurements into 10-min-based SCADA data. Thus, it is concluded that Method 2 satisfactorily models the true torque distribution, successfully capturing the fluctuation of torque in each interval of recorded SCADA.

### 3.3. Improvement in lifing accuracy

In order to illustrate the accuracy that is gained when Method 2 is used in physics-based prognostics, a lifing analysis based on ISO 281:2007 was performed on two bearings of the GE 1.5 SLE turbine mentioned in Section 3.2 using the torque histograms obtained from Method 1, Method 2, and the true torque distribution. Since the derivation and detailed analysis of the lifing calculation is outside the scope of this paper, only L50 values will here be reported and compared.

Using Miner's rule to model cumulative damage, the three torque histograms shown in Fig. 7 were independently used to calculate L50 of the two bearings. The bearing supporting the low speed intermediate shaft at the rotor side (LSIS-RS bearing) was predicted to have a L50 of 14.03 years with Method 1, 12.74 with Method 2, and 12.67 with the true torque distribution, meanwhile the bearing supporting the planet shaft at the rotor side (PS-RS bearing) was predicted a L50 of 6.36 years with Method 1, 5.96 with Method 2, and 5.91 with the true distribution. Assuming that the damage model used in this analysis accurately models the damage kinetics of the system, Method 1 introduced an error<sup>3</sup> of 10.7% in the predicted L50 of the LSIS-RS bearing, and 7.6% for the PS-RS

<sup>3</sup> Relative to the L50 calculated using the true distribution.

bearing, meanwhile Method 2 introduced errors of only 0.6% and 0.8% for LSIS-RS and PS-RS, respectively. Thus, by using Method 2 the error was reduced in a 10.1% for LSIS-RS and 6.6% for PS-RS.

#### 4. Conclusion

A novel method to determine the torque histogram based on SCADA has been introduced, and its application in physics-based prognostics evidenced a greater accuracy over the method currently used in the industry. Using 10-min distributions of power output and rotor speed, the method was able to successfully reconstruct the distribution of instantaneous torque in between 10-min intervals of recorded SCADA, predicting large fluctuations of torque in intervals of high mean torque leading to a very dispersed high-torque region. This prediction was validated with high-frequency SCADA data. The use of the proposed method in the lifing of a gearbox showed a L50 error only as high as 0.8%, meanwhile it was shown that the method currently used in the industry introduced an error as high as 10.7%. Using the method proposed in this paper, the max error was reduced by 10.1%. In conclusion, the new method proposed in this paper is recommended for lifing analysis due to its greater accuracy. The old method can still be used as a preliminary estimation of torque, but it must be kept in mind that this method undercuts the high-torque region, overestimating life in gearbox components.

Since the validity of the new method resides in the assumption that power output and rotor speed are jointly distributed forming a bivariate distribution, and that both are normally distributed in between each 10-min interval of recorded SCADA, it is proposed for a future research the study of operation conditions when these assumptions may not be valid.

#### Acknowledgements

The authors would like to thank NextEra Energy for providing the SCADA data used in this study. The development of the methods exposed in this paper was solely funded by Sentient Science Corporation as part of the company's efforts to provide the wind energy industry with accurate durability prognostics of their assets.

#### Appendix A. Approaches for Evaluating Eq. (2)

Since the second term in the parenthesis of Eq. (2) involves an exponential factor that can easily overflow a floating point precision, multiplied by an integral factor requiring numerical evaluation, the evaluation of Eq. (2) is not trivial. In search for an approach that can ease the computation of torque histograms in an intensive workflow, three different methods were explored. For a comparison on computation efficiency and numerical accuracy, the three approaches were tested on a data set consisting of 24 months of SCADA data collected from a single GE 1.5 SLE turbine, and results are here below reported.

##### A.1. Direct numerical integration

The integral factor can be directly evaluated numerically. To do so, Romberg's method of integration was implemented in a Python script, with a tolerance of 1%. In order to evaluate the exponential factor, it was determined that any value of  $q > 37$  would overflow the floating point precision; thus, for any  $q$  higher than 37 the scheme was switched to a symbolic computation scheme through the package Sympy.

This method resulted in a computation time unfeasible for any intensive workflow, indicating the need for a more effective approach. The histogram resulted in a mean torque of 321.77 kNm.

##### A.2. Transformation to erf

Dealing with pixel noise in astronomical measurements, Melchior and Viola [25] implemented Marsaglia's distribution noticing that the integral can be indirectly evaluated by transforming it into an error function as follows:

$$\int_0^q e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{\pi}{2}} \frac{2}{\sqrt{\pi}} \frac{1}{\sqrt{2}} \int_0^q e^{-\frac{1}{2}x^2} dx \quad (\text{A.1})$$

$$= \sqrt{\frac{\pi}{2}} \frac{2}{\sqrt{\pi}} \int_0^{\frac{q}{\sqrt{2}}} e^{-\theta^2} d\theta, \text{ with } \theta = \frac{x}{\sqrt{2}} \quad (\text{A.2})$$

$$= \sqrt{\frac{\pi}{2}} \text{erf}\left(\frac{q}{\sqrt{2}}\right) \quad (\text{A.3})$$

By using an efficient erf evaluation method (Scipy package), the computation was reduced to only 5% of the total time used through the direct integration approach, proving to be 20 times faster. The calculated torque histogram had a mean torque of 321.84 kNm, and a discrepancy of 0.013% to the direct integration approach.

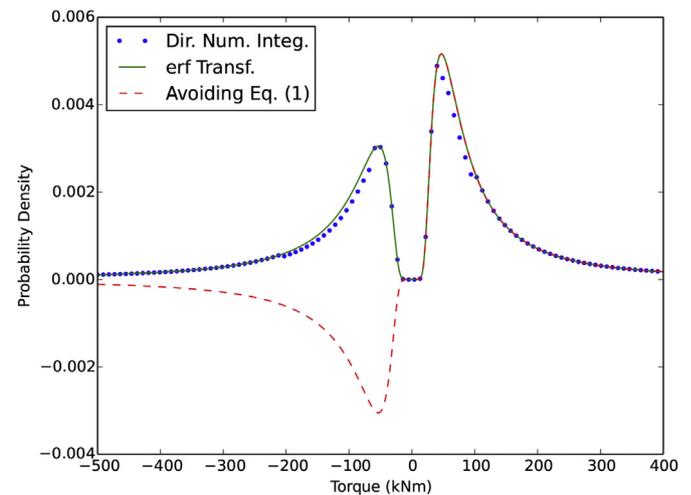


Fig. A.1. Comparison of approaches for evaluating Eq. (2) over extreme case [ $\mu_p, \mu_w, \sigma_p, \sigma_w, \rho$ ] = [600, 15.0, 90, 80.0, 0].

##### A.3. Avoiding Eq. (2)

Marsaglia proposed in Ref. [24] that the  $\frac{a+x}{b+y}$  transformation can be avoided by simply assuming  $w > 0$ . With such assumption, the PDF of the ratio of normal random variables  $U = \frac{z}{w}$  is given directly by

$$f_U(u) = \phi \left( \frac{\mu_w u - \mu_z}{\sqrt{\sigma_z^2 - 2\rho\sigma_z\sigma_w u + \sigma_w^2 u^2}} \right) \frac{\mu_w \sigma_z^2 - \rho\sigma_z\sigma_w \mu_z + (\mu_z \sigma_w^2 - \rho\sigma_z\sigma_w \mu_w) u}{(\sigma_z^2 - 2\rho\sigma_z\sigma_w u + \sigma_w^2 u^2)^{\frac{3}{2}}} \quad (\text{A.4})$$

where  $\phi$  is the standard normal density.

It is important to notice on Eq. (A.4) that, even though in most situations the PDF resembles the true density distribution given by Eq. (2), in extreme cases it produces odd results as exemplified in

Fig. A.1. This is due to the numerator  $\mu_w \sigma_z^2 - \rho \sigma_z \sigma_w \mu_z + (\mu_z \sigma_w^2 - \rho \sigma_z \sigma_w \mu_w) u$  becoming negative when  $u < \frac{-(\mu_w \sigma_z - \rho \sigma_w \mu_z)}{\mu_z \sigma_w - \rho \sigma_z \mu_w}$  with  $\mu_z \sigma_w - \rho \sigma_z \mu_w > 0$ , and in the negative when not.

This method achieved a similar time than the erf transformation, resulting in a mean torque of 324.18 kNm, and a discrepancy of 0.563% to the direct numeral integration approach, and 0.558% to the erf transformation approach.

## References

- [1] U.S. Energy Information Administration, U.S. electric generating capacity increase in 2016 was largest net change since 2011. URL <https://www.eia.gov/todayinenergy/detail.php?id=30112>.
- [2] A. Ghosh, Mitigate Costly main Bearing Failures with New Digitalclone Live Capabilities, 2017.
- [3] GCube, GCube gets to grips with grinding gearboxes. URL <http://www.gcube-insurance.com/en/press/gcube-gets-to-grips-with-grinding-gearboxes/>.
- [4] W. Liu, B. Tang, J. Han, X. Lu, N. Hu, Z. He, The structure healthy condition monitoring and fault diagnosis methods in wind turbines: a review, *Renew. Sustain. Energy Rev.* 44 (2015) 466–472, <http://dx.doi.org/10.1016/j.rser.2014.12.005>.
- [5] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T.A. Tameghe, G. Ekemb, Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges, *Energies* 7 (2014) 2595–2630, <http://dx.doi.org/10.3390/en7042595>.
- [6] A. Kusiak, A. Verma, Analyzing bearing faults in wind turbines: a data-mining approach, *Renew. Energy* 48 (2012) 110–116, <http://dx.doi.org/10.1016/j.renene.2012.04.020>.
- [7] Z. Feng, M. Liang, Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time frequency analysis, *Renew. Energy* 66 (2014) 468–477, <http://dx.doi.org/10.1016/j.renene.2013.12.047>.
- [8] Z. Feng, M. Liang, Complex signal analysis for wind turbine planetary gearbox fault diagnosis via iterative atomic decomposition thresholding, *J. Sound Vib.* 333 (2014) 5196–5211, <http://dx.doi.org/10.1016/j.jsv.2014.05.029>.
- [9] L. Hong, J.S. Dhupia, S. Sheng, An explanation of frequency features enabling detection of faults in equally spaced planetary gearbox, *Mech. Mach. Theory* 73 (2014) 169–183, <http://dx.doi.org/10.1016/j.mechmachtheory.2013.10.014>.
- [10] A.K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mech. Syst. Signal Process.* 20 (2006) 1483–1510, <http://dx.doi.org/10.1016/j.ymsp.2005.09.012>.
- [11] A. Cubillo, S. Perinpanayagam, M. Esperon-Miguez, A review of physics-based models in prognostics: application to gears and bearings of rotating machinery, *Adv. Mech. Eng.* 8 (2016) 1–21, <http://dx.doi.org/10.1177/1687814016664660>.
- [12] R.F. Orsagh, J. Sheldo, C.J. Klenlie, Prognostics/diagnostics for gas turbine engine bearings, in: *Proceedings of the 2003 IEEE Aerospace Conference*, 2003, pp. 3095–3103.
- [13] S. Li, A. Kahraman, A physics-based model to predict micro-pitting lives of lubricated point contacts, *Int. J. Fatigue* 47 (2013) 205–215, <http://dx.doi.org/10.1016/j.ijfatigue.2012.09.002>.
- [14] S. Li, A. Kahraman, N. Anderson, L. Wedeven, A model to predict scuffing failures of a ball-on-disk contact, *Tribol. Int.* 60 (2013) 233–245, <http://dx.doi.org/10.1016/j.triboint.2012.11.007>.
- [15] M. Watson, C. Byington, D. Edwards, S. Amin, Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems, *Tribol. Trans.* 48 (2005) 208–217.
- [16] G. Kacprzynski, A. Sarlashkar, M. Roemer, A. Hess, W. Hardman, Predicting remaining life by fusing the physics of failure modeling with diagnostics, *JOM* 56 (2004) 29–35, <http://dx.doi.org/10.1007/s11837-004-0029-2>.
- [17] S. Marble, B.P. Morton, Predicting the remaining life of propulsion system bearings, *Proc. 2006 IEEE Aerosp. Conf.* (2006) 1–8, <http://dx.doi.org/10.1109/AERO.2006.1656121>.
- [18] N. Bolander, H. Qiu, N. Eklund, E. Hindle, T. Rosenfeld, Physics-based remaining useful life prediction for aircraft engine bearing prognosis, *Annu. Conf. Progn. Health Manag. Soc.* (2009) 1–10.
- [19] R. V. Pulikollu, N. Bolander, S. Vijayakar, M. D. Spies, Analytical modeling and performance prediction of remanufactured gearbox components, 2nd World Congress on Integrated Computational Materials Engineering, doi:10.1002/9781118767061.ch13.
- [20] R. V. Pulikollu, N. Bolander, T. Shen, S. Vijayakar, M. D. Spies, Microstructure-based Fatigue Life Prediction Tool for Rotorcraft Spiral Bevel Gears, *AHS International Forum* 69.
- [21] C.S. Gray, S.J. Watson, Physics of failure approach to wind turbine condition based maintenance, *Wind Energy* 13 (2010) 395–405, <http://dx.doi.org/10.1002/we.360>.
- [22] I. Al-Tubi, H. Long, P. Tavner, B. Shaw, J. Zhang, Probabilistic analysis of gear flank micro-pitting risk in wind turbine gearbox using supervisory control and data acquisition data, *IET Renew. Power Gener.* 9 (2015) 610–617, <http://dx.doi.org/10.1049/iet-rpg.2014.0277>.
- [23] G. Marsaglia, Ratios of normal variables and ratios of sums of uniform variables, *J. Am. Stat. Assoc.* 60 (1965) 193–204, <http://dx.doi.org/10.2307/2283145>.
- [24] G. Marsaglia, Ratios of normal variables, *J. Stat. Softw.* 16 (2006) 1–10, <http://dx.doi.org/10.18637/jss.v016.i04>.
- [25] P. Melchior, M. Viola, Means of confusion: how pixel noise affects shear estimates for weak gravitational lensing, *Mon. Not. R. Astronomical Soc.* 424 (2012) 2757–2769, <http://dx.doi.org/10.1111/j.1365-2966.2012.21381.x>.