IOP Conf. Series: Journal of Physics: Conf. Series **1037** (2018) 032005 doi:10.1088/1742-6596/1037/3/032005

Assessing the Utility of Early Warning Systems for Detecting Failures in Major Wind Turbine Components

L Colone[1,*], M Reder[2], N Dimitrov[1], D Straub[3]

[1] Technical University of Denmark, Frederiksborgvej 399, 4000 Roskilde, Denmark
[2] CIRCE - University of Zaragoza, C/ Mariano Esquillor 15, 50018, Zaragoza, Spain
[3] Technical University of Munich, Theresienstr. 90, 80333 Munich, Germany

E-mail: lcol@dtu.dk

Abstract. This paper provides enhancements to normal behaviour models for monitoring major wind turbine components and a methodology to assess the monitoring system reliability based on SCADA data and decision analysis. Typically, these monitoring systems are based on fully data-driven regression of damage sensitive-parameters. Firstly, the problem of selecting suitable inputs for building a temperature model of operating main bearings is addressed, based on a sensitivity study. This shows that the dimensionality of the dataset can be greatly reduced while reaching sufficient prediction accuracy. Subsequently, performance quantities are derived from a statistical description of the prediction error and used as input to a decision analysis. Two distinct intervention policies, replacement and repair, are compared in terms of expected utility. The aim of this study is to provide a method to quantify the benefit of implementing the online system from an economic risk perspective. Under the realistic hypotheses made, the numerical example shows for instance that replacement is not convenient compared to repair.

1. Introduction

The significant maintenance cost associated with failures of large wind turbine (WT) components calls for improved operation and maintenance (O&M) decision support systems. Such systems are increasingly drawing the attention of operators, especially in the offshore sector [1]. Maintenance actions are necessary to ensure a certain level of reliability of a machine throughout its lifetime. Maintenance strategies can be broadly categorised into corrective and preventive [2,3]. The second type can in turn be divided into scheduled and condition based maintenance (CBM), with the latter being a predictive policy. CBM is helpful to avoid early replacement of healthy components while identifying critically worn-out components. Wind energy, especially offshore, is still a maturing industry and the scenario around CBM poses some new and unique challenges. The lack of extensive datasets containing run-to-failure data often does not allow the adoption of supervised learning techniques [4]. Moreover, realistic physics-based deterioration models are not generally available due to the complexity of the machine behaviour under operation and its dependence on multiple physical variables. To deal with this problem some studies have suggested to derive a fully data-driven normal behaviour model (NBM) of specific damage-sensitive features [5]. This process refers to building a model characterizing the behaviour of a system directly from measured data. Such NBMs are regularly used to detect anomalies in wind turbine behaviour, as demonstrated in [6,7]. Since the model variables

are typically continuous, prediction models are based on linear or nonlinear regression. This approach seems to be well tailored for main components, thanks to its simplicity of application with regards to the data requirements and scalability [8]. From an operator's perspective it is important to understand the economic benefit of implementing predictive systems. This can be achieved by defining a performance measure to quantify the reliability of early detection warnings, followed by a cost-benefit analysis. Often, when dealing with rare events, a sufficient number of failure observations is not available and thus not allowing to establish performance statistics, as for instance used in [8]. Hence, this indicates the need for alternative metrics. In the present study, the lead time to failure events against the probability of false alarms (PFA) is analysed, in order to assess the system performance from an economic perspective. Considerations about the nature of failure are taken into account to identify suitable detection thresholds. A case study of main bearing failures from an onshore wind farm is carried out and analysed. The system output is condensed into a scalar anomaly measure to track the component deterioration. Furthermore, a sensitivity study selecting the most important input variables is carried out, in order to enhance NBM. Each section explains the methodology adopted and uses information previously derived. From section 3 on, results are presented in a progressive order without a dedicated result section. The paper concludes with a discussion on the contribution and future research in the area.

2. Model

The analytical model used herein has the structure of a hierarchical NBM between input vectors $\mathbf{x}$ and output $\hat{\mathbf{y}}$. For modelling the main bearing temperature in the form $\hat{\mathbf{y}} = g(\mu(\mathbf{x})) + \mathbf{e}$, a generalised linear model (GLM) with mean μ, a Gaussian error distribution $e_i \sim \mathcal{N}(0, \sigma_{e,i})$ and an identity link function $g(\cdot)$ is used [9], so that, for the i^{th} output variable, $g(\mu_i) = \mu_i = E(\hat{y}_i)$, where E is the expected value operator. The model is expressed in scalar form as

$$E(\hat{y}_i) = \hat{\alpha}_{0i} + \sum_{j=1}^{K} \hat{\alpha}_{ij} x_j + e_i, \tag{1}$$

where $\hat{\alpha}$ is assumed to be $\hat{\alpha} \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha)$, with $\bar{\alpha}$ and σ_α being respectively mean and standard deviation of the model parameters and K the number of explanatory variables, or covariates. The error between the model output $\hat{\mathbf{y}}$ and the measurements $\mathbf{y}$ has zero mean, and the appearance of bias indicates deviation from the normal behaviour. The parameter estimation is based on a least-squares optimization with a least absolute shrinkage and selection operator (Lasso) [10]. This regularisation method uses a penalisation on the $\mathcal{L}_1$ norm, and is an effective technique for subset selection in high dimensional multivariate models. The Lasso solves the following optimisation problem, where the penalisation parameter λ is introduced:

$$\hat{\alpha}_{lasso}(\lambda) = \arg\min_{\hat{\alpha}^*} \|\mathbf{y} - \mathbf{x}\hat{\alpha}^*\|_2^2 + \lambda \|\hat{\alpha}^*\|_1 , \tag{2}$$

where $\mathbf{x}$ is the vector of model covariates and λ is the penalisation parameter. The remaining terms are the same as in Eq.1. Lasso is applied here by virtue of its ability to achieve high prediction quality with the minimum number of covariates. The relative importance of covariates is measured for the training with the coefficient of determination (R-squared), and for the predictions with the mean absolute error (MAE) and root mean squared error (RMSE). In this manner, the herein used techniques have practical utilization.

3. Training

A set of SCADA data from an onshore wind farm with more than 10 turbines is available over a period of 5 years. During this period 3 turbines, WT1, WT2 and WT3, experienced main bearing failures. This component has been identified in previous studies as one of the most critical WT components in terms of failure frequency and downtime [11]. The full SCADA dataset consists of 74 channels given as 10 minutes average values. Operational data only are selected from the dataset by removing unnecessary observations, corresponding to power less or equal than zero as well as measurement errors.

The training dataset is built by selecting normal behaviour data from 6 turbines including the 3 having experienced failures. In these latter, NBM data are selected until 6 months before the failure event, in order to not introduce degradation information into the training dataset. From the selected training pool of normal behaviour data, random down-sampling is applied in order to reduce its size while conserving the variability of the data coming from different turbines and their operational conditions. In total, $N = 3 \cdot 10^5$ samples are employed for training, corresponding to the equivalent of 6 years of data from 1 turbine. All the SCADA variables are normalised between 0 and 1. The turbine validation demonstrates that the NBM has more universal applicability and ensures that the failure detection does not rely on conditions specific to the turbines that have experienced the event. The GLM was trained with a 10-fold cross validation.

4. Variable Selection and sensitivity study

Model reduction is an important step in improving the model speed and usability when dealing with large datasets. Although the GLM is computationally efficient, reducing the number of covariates will lead to further benefits as it will reduce the data storage requirements and data processing. A certain degree of model reduction is ensured by the Lasso approach, as the penalization function leads to some model coefficients being reduced towards zero, which allows their elimination from the model. This is performed through a variable importance analysis by comparing the standardised coefficient magnitudes for each input variable obtained through the estimation process. The model predicts the main bearing temperature using the full set of variables in the SCADA dataset. The 15 most important variables obtained are displayed in Fig.1. Blue colour indicates a positive effect on the model result, which means that with increasing values for the coefficient of this covariate the model response also increases. Red bars indicate the inverse relationship.

Interestingly, the first top 10 covariates are temperature measurements, which may suggest a correlation between abnormal bearing temperature rise, which eventually leads to failure, and high operational demand, as shown in Fig.1. Rising brake temperature, lower external temperature and lower non-drive-end (NDE) generator bearing temperatures contributed to the increasing main bearing temperature. At first glance, this might seem contradictory. However, lower external temperatures are usually related to higher wind speeds, which affect the failure bahaviour of certain components [12,13]. At this step, further analysis could be carried out such as Chi-square tests or ANOVA, in order to further reduce the number of input parameters, [14].

Fig.2 displays the model RMSE, the MAE and R-squared performance metrics as function of the number of model covariates. To obtain the figure, the model is first trained using the full set of variables, then retrained subsequently by adding one variable at the time from the most important to least important variable to establish the variable importance. The x-axis indicates the number of input covariates ordered according to the standardized model coefficients from the most important (number 1) to the least important variable (number 74). Including more covariates to a model implies higher model complexity and longer model evaluation times. Thus, a compromise between the number of model covariates and model accuracy has to be found. It is shown in the graph that by including the 15 most important covariates the R-squared reaches

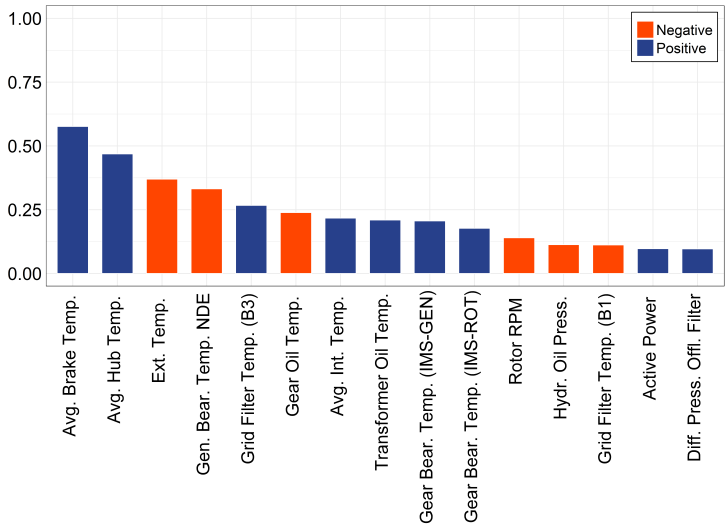


Figure 1: Top-15 standardised regression coefficient with respect to main bearing temperature from Lasso

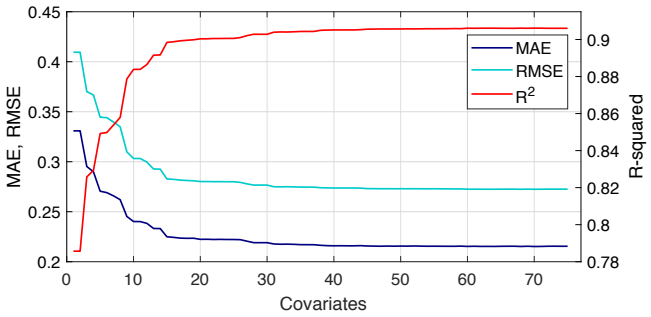


Figure 2: The R-squared of the model and the prediction errors MAE and RMSE using different numbers of inputs.

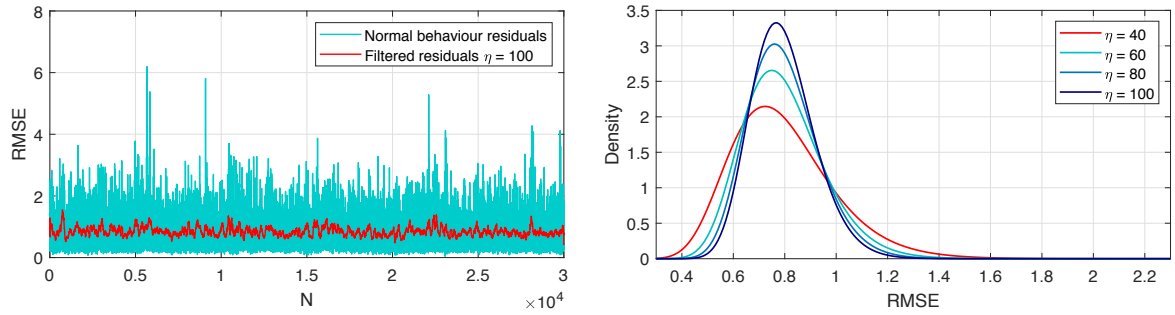
a value of approximately 0.91, along with a sufficiently low prediction error.

5. Damage detection

The set of covariates obtained in the previous section is selected for building the model and testing it on the damaged turbines. In the case of a single model output (e.g. inner-ring temperature), the deviation function is a univariate measure, which is here selected as the RMSE, namely RMS of 10 consecutive samples of the model residuals, corresponding to an effective operating time $\Delta t = 100$ minutes. Previous studies e.g. [6], suggested to average the discordance measure over 3 days in order to reduce the occurrence of false alarms. In the present study, however, shorter time periods are achieved by applying a low-pass filter, a centered moving average, to the model residuals. In Eq.3, the parameter η is the window size of the filter.

$$e_f(i) = \frac{1}{2\eta + 1} \sum_{j=0}^{2\eta} h(e(i + \eta - j)) \tag{3}$$

In this way, a more robust performance is ensured. The raw and filtered RMSE in normal behaviour are displayed in Fig.3a, while Fig.3b shows their distributions at different filtering levels, where a log-normal distribution is fitted, which is the distribution that resulted in the best fit. As can be noticed, the uncertainty decreases with the level of filtering. This information is useful to quantify the reliability of the monitoring system.



(a) (b)

Figure 3: *a*) Raw and filtered RMSE from the normal behaviour dataset. *b*) Log-normal distributions of the filtered RMSE from normal behaviour data.

In general a multivariate discordance metric can be used as degradation function, if for instance the output set is composed of a correlated set of damage-sensitive features. For instance, Fig.4 shows the main bearing vertical acceleration RMS, temperature, and tower-top acceleration RMS in an example case of failure. The interaction between these variables could be exploited to obtain earlier and more robust predictions. However, this study only focuses on SCADA data, leaving out the vibration analysis.

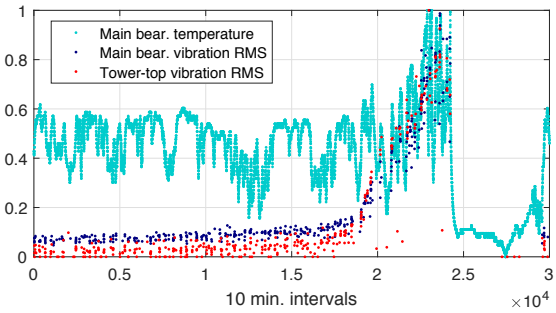
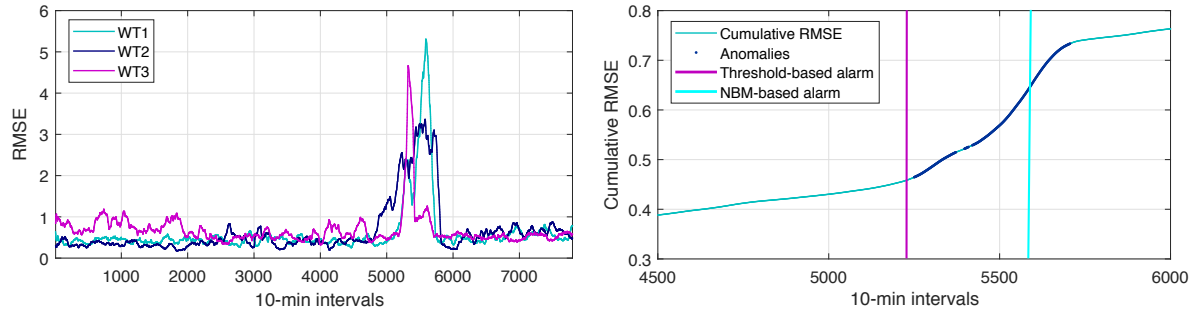


Figure 4: Main bearing temperature, acceleration RMS and tower-top acceleration RMS close a main bearing failure event (WT1).

Fig.5a and 5b show the trend of the filtered RMSE for the damaged dataset on the three turbines. The damage progression can be readily identified and an NBM-based alarm is issued when the failure threshold is crossed.

The lead time is here defined as the time lag between the first warning issued by the model and the first threshold-based alarm from the SCADA system. The selection of the NBM-based alarm threshold is a tradeoff between lead time and probability of false alarms (PFA), where



(a) (b)

Figure 5: *a*) Filtered RMSE of the damaged (testing) dataset. *b*) Normalised cumulative RMSE and alarms issued by threshold-based and NBM-based system (WT2)

for a given reference period the PFA equals the probability of exceeding the threshold under normal, stationary conditions. The PFA can thus be found as $P[\text{RMSE} > T]$, where T is a threshold.

6. Decision analysis

The decision analysis has the aim to assess the economic advantage of using the system based on field data. Since the analysis is based on simple cost considerations, it is best suitable for preliminary assessments. A utility function is derived by associating a cost to a false alarm and a saving to detecting a failure early. The NBM-based threshold is chosen such that the lead time is long enough to cover the entire mobilization time, needed for preparing the crew and hiring the crane vessel. Only consequences associated with direct financial losses are considered, i.e. human injuries, environmental effects and similar are not taken into account. The decision tree is sketched in Fig.6, which shows all possible alternatives originating from an NBM-based alarm A. In the simplest case, when an alarm is issued, the turbine keeps running and a site inspection is performed, which can result in false case ($\bar{E}$) or a true case (E). It is assumed that the inspections are perfect. As shown in Fig.6, if the inspection reveals a developing failure, an intervention a is performed, corresponding to a repair r or full component replacement R and their associated costs. In case of repair $C_a = C_r$, and in case of replacement $C_a = C_R$. These two cases are studied separately, as if the adopted policy is constrained to be only one of them.

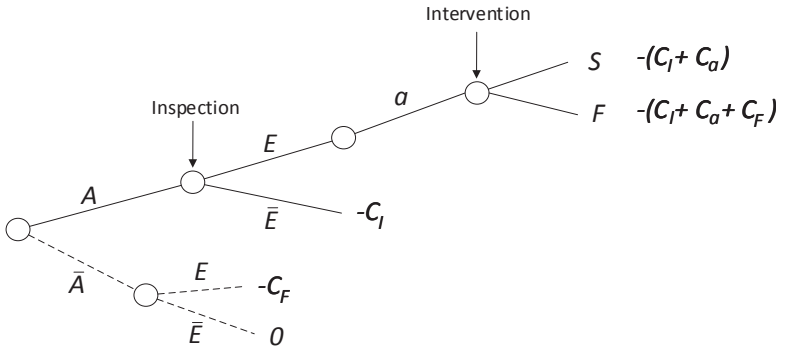


Figure 6: Decision tree associated to the detection of rare events and corresponding utility functions for each outcome.

Two discrete states: failure (F) and functional, or safe, (S) are considered after the intervention is performed. In order to model the possibility that a repair fails to eliminate the component damage, an efficiency coefficient $\psi \in [0, 1]$ is introduced as formerly developed in [15]. The possible outcomes of the maintenance action are considered binary, i.e., the component is either restored to its original undamaged state, or its state remains equivalent to its current damaged state. The efficiency term specifies the probability that the repair action is successful and the component health is fully restored, as

$$P(F \mid E \cap A \cap a) = P(F \mid E)(1 - \psi) = (1 - \psi), \tag{4}$$

where the assumption $P(F \mid E) = 1$ means that a true case is always associated with a developing failure, regardless of the maintenance policy adopted. Therefore, the repair action influences the probability of failure of the component by making it decrease linearly with the increasing repair efficiency. In real applications, a similar function could be derived from historical data. With reference to Fig.6, the discrete probabilities can be readily written as

$$\begin{aligned} P(F \cap E \cap A \cap a) &= P(F \mid E \cap A \cap a)P(E \mid A)P(A) = (1 - \psi)\lambda_0 \\ P(S \cap E \cap A \cap a) &= [1 - P(F \mid E \cap A \cap a)]P(E \mid A)P(A) = \psi\lambda_0 \end{aligned} \tag{5}$$

which make use of the following quantities expressed as functions of PFA and efficiency, derived from the Bayes rule

$$P(E \mid A) = \frac{P(A \mid E)P(E)}{P(A)} = \frac{\lambda_0}{P(A)} \tag{6}$$

$$P(\bar{E} \mid A) = \frac{P(A \mid \bar{E})P(\bar{E})}{P(A)} = \frac{\mathrm{PFA}(1 - \lambda_0)}{P(A)}. \tag{7}$$

Note that in Fig.6, the alternative $\bar{A}$ (dashed) is associated with zero utility, because $U(\bar{E} \mid \bar{A}) = 0$ and $P(E \mid \bar{A}) = 0$. $P(A \mid E) = 1$ because it is assumed the system always issues an NBM-based alarm in case a failure is present. This is equivalent to assuming a probability of detection equal to 1. Further information on this probability could not be inferred from available data.

The tree includes the utilities u_i associated to each possible outcome. Note that the quantity $P(A) \neq 0$ does not need to be explicitly determined, since it cancels out when Eq.6 is inserted into Eq.5.

The analysis is performed on a hourly basis, namely all the probabilities are expressed with respect to 1 hour. A constant event rate $P(E) = \lambda_0$ is assumed, derived as a frequentistic rate from historical data, where 3 events on three different turbines were reported over the period analysed of 5 years. Note that the hourly PFA is derived from the PFA computed for 100-minute sampling windows by scaling with the ratio between the length of the sampling periods (see Fig.3). To perform a repair comes at a cost C_r, while the inspection costs are C_I. These quantities were retrieved from available literature [16].

The cost of replacement is equal to the material cost plus the crew costs (which means the possibility that the lead time is shorter than the mobilization time for crew and equipment is neglected), while the cost of failure comprises the cost of replacement plus the power loss during the mobilization time and replacement time. With reference to Tab.1, these quantities are expressed as

Table 1: Numerical quantities employed in the decision analysis.

Parameter	Value	Description
C_I [€]	2500	Cost of inspection
C_r [€]	$10 \cdot 10^3$	Cost of repair
$\bar{P}$ [MWh/h]	1.05	Hourly average production
C_m [€]	$15 \cdot 10^4$	Material cost
N_w [€]	10	Number of workers
C_{el} [€/MWh]	100	Price of electricity
C_l [€/h]	33	Cost of labour
MTTR [h]	240	Mean time to replacement
MT [h]	170	Mobilization time

$$\begin{aligned} C_R &= C_m + \text{MTTR}\ C_l\ N_w \\ C_F &= C_R + (\text{MTTR} + \text{MT})\ \bar{P}\ C_{el}. \end{aligned} \tag{8}$$

The current electricity price from offshore wind was taken from [17]. Due to the absence of cost parameters and time to repair for this specific case, the quantities found in [1] were adopted, corresponding to the replacement of an equivalent main component, the gearbox. The total expected utility is found as

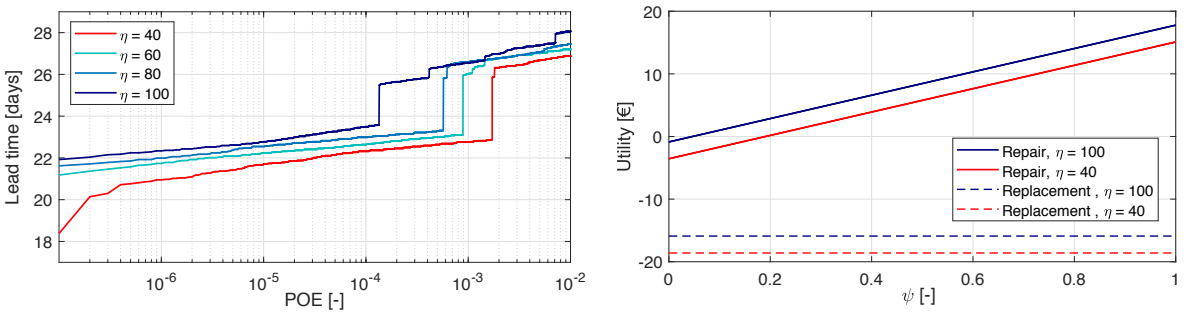
$$U = \sum_i P_i u_i - U_0, \tag{9}$$

where $U_0 = \lambda_0 C_F$ is the utility associated with not implementing the system. This comprises the average power production losses, replacement or repair costs and labour costs, part of which could be saved when a sufficient lead time and repair efficiency can be provided by the system. Fig.7a shows the lead time as a function of the PFA where the jumps are due to the small sample size. The changes in lead time and PFA are a direct function of the chosen error threshold, T. The figure shows that achieving longer lead time requires lowering the error threshold and hence increasing PFA. Conceptually, for a specifically chosen error threshold the lead time would be a random variable, and there is a non-zero likelihood for the lead time to be shorter than the required mobilization time. This is associated with a certain cost, which has a trade-off with the cost of false alarms, as reducing the threshold will mean increasing the lead time but also increasing the false alarm rate. It is possible to search for an optimal threshold that results in the lowest overall cost, however this is not considered in the present study. Instead, a suitable error threshold level is chosen such that the average lead time equals 23 days for filtering level $\eta = 100$. Based on this choice, the corresponding PFA is obtained (Fig.7).

Fig.7b displays the expected utility in case of repair $U_a(\psi)$, which is positive already for a large span of the efficiency domain. The plot displays that low-pass filtering the error deviations lead to lower PFA and thus higher benefit, besides providing positive utility at lower efficiency values. In this numerical example, the replacement policy is not cost-effective compared to repair, with $U_a = -15.9$ € and $U_a = -18.6$ € respectively for the maximum and minimum levels of error filtering as shown in case of repair in Fig.7b).

7. Discussion

The positive standardized coefficient of the brake temperature could be due to thermal conduction from the bearing to the brake, possibly indicating a cross-correlation between



(a) (b)

Figure 7: *a*) Lead time against probability of exceedance (POE) averaged over 3 failure cases, for different filter levels. POE is equivalent to PFA. *b*) Repair utility as a function of the efficiency of intervention.

the bearing temperature, the brake temperature and failure. For example, a high bearing temperature combined with high brake temperature might mean normal operation under heavy load. The NDE generator bearing located on the opposite end of the drive train compared to the main bearing, is able to maintain lower temperatures. The increasing hub temperature might instead be a consequence of the main bearing temperature raise, by thermal conduction through the main shaft.

Main bearing failures as the ones analysed here, which show a clear temperature rise over time, may be related to high friction, probably due to lack of proper grease lubrication. This justifies the utilization of a repair policy.

More detailed cost models should be used in the decision analysis to include site accessibility, parameter uncertainty [18], discount rate, updated probability of failure. Furthermore, advanced models of inspections would be desirable [16], which will trigger more decision alternatives. In this paper, perfect inspection was assumed. However these details are case-specific and thus showing their value mainly in real applications.

Regarding the probability of failure in case of replacement, more detailed models would allow for an updated probability of event or failure based on condition, rather than considering the same measured event rate after replacement as done in this work. This is to account for the increased probability of failure due to usage, as opposed to new.

A further improvement would be to incorporate lifetime prediction models as for instance proposed in other structural components [19], in order to give an estimate of the remaining life of the component. In this case, the latter can be achieved by for instance setting a maximum temperature threshold at which the component is considered to be in failed state. In this way, information about progressive failures can be obtained and included in the decision model, which in real cases is expected to be more advanced.

8. Conclusions

This research has highlighted improvements in NBM of WT components and the definition of a performance metric in case of rare events was defined. Thus, a model of the normal main bearing inner-ring temperature is trained by using turbines other than the ones experiencing the damage analysed (turbine validation). A linear model is used and a sensitivity study is performed to select the most important variables. The Lasso-based model reduction successfully eliminated unimportant input variables and provided an understanding of the failure mechanisms related parameters. Brake temperature, hub temperature, external temperature and generator bearing

(NDE) temperature were found to be the most important model covariates. The decision analysis has shown that a replacement policy would not result in cost-effective solutions, while a repair policy would be more suitable. These types of problems should be addressed by wind farm operators in order to advance in maintenance management and make the wind energy market more competitive with conventional energy generation. The approach can be implemented for any kind a major wind turbine subsystems such as gear box, generators or blades. Thus, future research will need to cover failure detection of main mechanical components and quantify the benefit of using multivariate over single parameter approaches.

Acknowledgements
This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 642108 (AWESOME). Furthermore, the authors wish to thank Vattenfall for the support provided.

References

[1] Carroll J, McDonald A and McMillan D 2016 *Wind Energy* **19** 1107–1119
[2] Randall R B 2011 *Vibration-based condition monitoring: industrial, aerospace and automotive applications* (John Wiley & Sons)
[3] Rausand M and Arnljot H 2004 *System reliability theory: models, statistical methods, and applications* vol 396 (John Wiley & Sons)
[4] Bishop C M 2006 *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Secaucus, NJ, USA: Springer-Verlag New York, Inc.) ISBN 0387310738
[5] Tautz-Weinert J and Watson S J 2016 Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection *Journal of Physics: Conference Series* vol 753 (IOP Publishing) p 072014
[6] Bangalore P and Tjernberg L B 2015 *IEEE Transactions on Smart Grid* **6** 980–987
[7] Mazidi P, Bertling Tjernberg L and Sanz Bobi M A 2017 *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **231** 121–129
[8] Bach-Andersen M, Winther O and Rømer-Odgaard B 2015 Scalable systems for early fault detection in wind turbines: a data driven approach *Annual Conference of the european Wind Energy Association*
[9] Nelder J A and Baker R J 1972 *Generalized linear models* (Wiley Online Library)
[10] Tibshirani R 1994 *Journal of the Royal Statistical Society, Series B* **58** 267–288
[11] Reder M, Gonzalez E and Melero J J 2016 *Journal of Physics: Conference Series* **753** 072027 ISSN 1742-6588
[12] Reder M and Melero J J 2017 *Journal of Physics: Conference Series* **926** 012012 ISSN 1742-6588
[13] Reder M, Yürüşen N Y and Melero J J 2018 *Reliability Engineering & System Safety* **169** 554–569 ISSN 09518320
[14] Friedman J, Hastie T and Tibshirani R 2001 *The elements of statistical learning* vol 1 (Springer series in statistics New York)
[15] Colone L, Dimitrov N and Straub D *(Unpublished)*
[16] Nielsen J J and Sørensen J D 2010 *Proceedings of the Reliability and Optimization of Structural Systems, München, Germany* 7–10
[17] Dinwoodie I, Endrerud O E V, Hofmann M, Martin R and Sperstad I B 2015 *Wind Engineering* **39** 1–14
[18] Seyr H and Muskulus M 2016 *Journal of Physics: Conference Series* **753** 092009
[19] Nielsen J S and Sørensen J D 2017 *Energies* **10**