Adaptive Data-driven Approach for Fault Prognostics based on Normal Conditions - Application to Shaft Bearings of Wind Turbine

Koceila Abid[1], Moamar Sayed-Mouchaweh[2], and Laurence Cornez[3]

[1] *CEA Tech, Hauts de France, 59000, France*
koceila.abid@cea.fr

[1,2] *IMT Lille Douai, Lille University, 59500, France*
koceila.abid@imt-lille-douai.fr
moamar.sayed-mouchaweh@imt-lille-douai.fr

[3] *CEA List, Paris Saclay, 91191, France*
laurence.cornez@cea.fr

ABSTRACT

Prognostics can enhance the reliability and availability of industrial systems while reducing unscheduled faults and maintenance cost. In real industrial systems, data collected from the normal operation conditions of system is available, but historical degradation data is often unavailable. Hence, this paper proposes a general data-driven prognostic approach dealing with the lack of degradation data in the offline phase. First, features are computed on the collected raw signal, then One Class Support Vector Machine (OCSVM) is used to detect the degradation, this anomaly detection method is trained using only normal operation data. Then, features are ranked according to the selection criteria. The feature having the highest score is chosen as Health Indicator (HI). Finally an adaptive degradation model is applied for the prediction of the degradation evolution over time and Remaining Useful Life (RUL) estimation. The proposed approach is validated using run-to-failure vibration data collected from a high speed shaft bearings of a commercial wind turbine.

1. INTRODUCTION

Prognostics and Health Management (PHM) or predictive maintenance can provide an advanced maintenance strategy compared to the traditional one (e.g., curative and preventive maintenance). PHM can enhance the reliability and the availability while reducing unscheduled faults and maintenance cost of industrial systems such as Wind Turbines (WTs). The production of energy using WT has an increasing trend, this entails a growing evolution in the number and size of WTs which increase the Operation and Maintenance (O&M) cost. Prognostic is a part of the PHM strategy and focus to define the remaining time to a failure occurrence, named the Remaining Useful Life (RUL). It is estimated by using the HI, where the HI represents the evolution over time of the system performances or conditions, it is used to predict the degradation evolution and estimate the RUL. RUL is estimated based on two main approaches (Abid, Sayed-Mouchaweh, & Cornez, 2018): Experience based (reliability, similarity), and degradation modeling based (model, data-driven) approaches. Experience based approaches require a large datasets about the degradation dynamics, which is not often available in industrial systems. It is difficult to apply model based approaches in real complex systems, due to the interaction between different components and the environmental variations. Despite the lack of interpretation, data-driven approaches are the most suitable methods for the prognostics of industrial complex systems. These methods offer a trade-off in terms of applicability, precision, implementation, and cost.

Prognostic of bearings is widely investigated in the literature, for the reason that bearings are most subject to failure due to the wear and tear during operation. For this application, prognostic methods are mainly data-driven approaches. HI construction in this category of methods can be based on a single feature or it can be based on a fusion of multiple features that can improve the representation of the system degradation. For example, In (Benkedjouh, Medjaher, Zerhouni, & Rechak, 2013) HI is built from a reduction of eight features computed on a vibration signal by using the Wavelet Packet Decomposition (WPD). HI is contructed using Mahalanobis distance in (Wang, Peng, Zi, Jin, & Tsui, 2016), where the HI is the distance between normal and evolving class (i.e., degradation).

In (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012), features are extracted from vibration signal using WPD, which can analyses the higher frequency domains of a signal.

Some works evaluate the extracted features using criteria and select the feature with high score as HI. The best features can also be fused in order to enhance the selection criteria. Monotonicity and trendability selection criteria are widely employed in the literature. Monotonicity evaluates the increasing or decreasing trend, where the trendability evaluates the correlation between features and the operating time. In (Javed, Gouriveau, Zerhouni, & Nectoux, 2015), classical feature and proposed trigonometric features are ranked using monotonicity and trendability criteria (Coble, 2010). A combination of the monotonicity and trendability criteria are applied in (Guo, Li, Jia, Lei, & Lin, 2017) for features selection, the best features are fused using a Reccurent Neural Network (RNN) in order to construct the HI. These criteria are also applied in (Saidi, Ali, Bechhoefer, & Benbouzid, 2017) to select the best features as HI among different statistical and frequency domain features.(Atamuradov, Medjaher, Dersin, Zerhouni, & Camci, 2018) proposed a hybrid feature selection method for point machine sliding chair monitoring, where the affinity matrix and inherent features (monotonicity, trendability, and robustness) are combined in order to improve the selection.

In industry, data from different normal operating conditions can be found, otherwise there is a lack of past occurred degradation data or it is often unavailable (e.g., new machine). Also generating degradation data in the laboratory is very costly and the degradation dynamics do not behave as the reality due to the change in environmental and operating conditions (e.g., wind speed variability for wind turbine). Hence, applying a prognostic approach is challenging due to several reason. i) it is difficult to select the most sensitive features in order to detect and predict the degradation using only normal conditions data, ii) it is arduous to construct the HI online; most of the works in the literature select the best features as HI using evaluation criteria in the offline phase (using run-to-failure degradation data). In order to address these challenges, a prognostic approach dealing with the lack of degradation data is proposed in this work. First of all, a general library is constructed in order to include as much as possible features sensitive to different degradation dynamics. Secondly, anomaly detection method is used since only normal conditions data is available. Anomaly detection or novelty detection is the identification of new patterns which differ significantly from the patterns generated during normal operation conditions. Hence, One Class Support Vector Machines is applied in order to detect the degradation and trigger the RUL estimation without using degradation data for training. Thirdly, features are ranked online according to meaningful evaluation criteria (monotonicity and trendability), where the feature with high score is selected as HI and used for the degradation prediction. Finally, adaptive degradation model is used in the aim to follow the degradation evolution over time and estimate RUL.

This paper is organized as follows. Section 2 explains the proposed approach for degradation detection and RUL estimation. Section 3 describes the run-to-failure high speed shaft bearings data and presents the results of the proposed approaches. Conclusion and perspectives are given in Section 4.

2. PROPOSED APPROACH

A data-driven based approach is proposed (Figure 1) for the degradation detection and prognostics. The proposed approach uses data about the normal conditions of the system as a priori knowledge. It can be decomposed into three parts.

2.1. Feature definition

The first part aims to choose the most sensitive features to the degradation. The selected features can be used for early and reliable detection of the degradation, and can also be used for the characterization of the degradation evolution with time. Degradation result in a change of a measured signal which can be a system performance signal (e.g., produced energy of a wind turbine) or it can be related to the operation conditions of a system (e.g., vibration and temperature). In order to detect the change in the measured signal, a library of statistical features is used in this work. Table 1 presents the 12 statistical features defined in the library, which are widely used in the litterature for the degaradation detection and prognostics (Saidi et al., 2017)(Wang et al., 2016). These different features are sensitive to different degradation dynamics. For example, the standard deviation can measure the dispersion of the signal, skewness and kurtosis examine the probability density function (PDF) of the signal, where the skewness measures the asymmetry of the PDF and the kurtosis measures the flatness degree. The shape factor is defined as the ratio of the Root Mean Square (RMS) value to the mean absolute. RMS describes the signal strength (e.g., sensitive to the bearing's wear). This standard library is defined in the offline phase and the included features can be modified, suppressed, or enriched by feedback or human experts.

2.2. Degradation detection

In this work, One Class Support Vector Machines (OCSVM) is chosen among the different anomaly detection methods for several reasons:

- No assumption about data distribution
- Capability of dealing with high dimensional data
- Deal with complex problem (Non linear decision boundary)

OCSVM separates all the data points from the origin in a

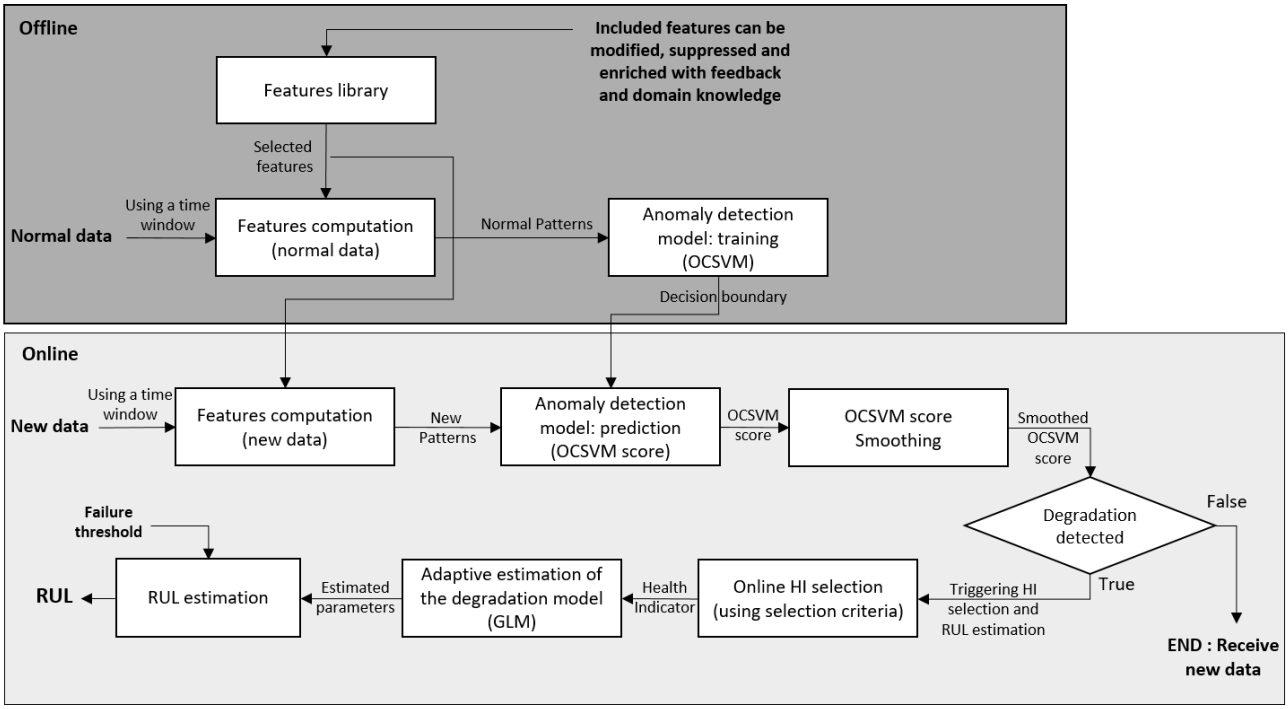

Figure 1. Flowchart of the proposed prognostic approach.

high dimensional feature space. The objective is to find an optimal hyperplan that maximizes the distance from it to the origin. It can be formulated in a quadratic programming problem (Schölkopf, Williamson, Smola, Shawe-Taylor, & Platt, 2000).

$$\min_{\omega,\xi,\rho} \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu n}\sum_i \xi_i - \rho \tag{1}$$

Subject to

$$(\omega \cdot \Phi(x_i)) \geq \rho - \xi_i, \xi_i \geq 0 \tag{2}$$

Where n is the number of training samples, ω is the normal vector that separating hyperplane, ρ is the offset of the desired hyperplane, $\xi = [\xi_1 ... \xi_n]$ is a vector of errors, and $\Phi(.)$ is a kernel function. $\nu \in [0, 1]$ is an upper bound on the fraction of training samples outside the decision boundaries and a lower bound on the fraction of support vectors.

After solving the quadratic programming problem, the final decision function is :

$$f(x') = \text{sgn}(\sum_i \alpha_i K(x_i, x') - \rho) \tag{3}$$

$$K(x_i, x') = exp[-\|x_i - x'\|^2/2\sigma^2] \tag{4}$$

Where: K is the kernel function, Radial Basis Function (RBF) is used (Eq. (4)), α_i is the observation coefficient, σ^2 is the variance, and $\|.\|$ is the Euclidean norm.

The function $f(x')$ returns +1 if the observation x' belong to the known regions and -1 elsewhere. A OCSVM score can also be computed using Eq. (3) which is positive in the normal class, negative outside, and 0 in the boundary. OCSVM doesn't take in consideration the sequentiality of points, for this reason a moving median is applied to the score. Moving median is used for smoothing the OCSVM score by removing outliers score, which can be interpreted as including the sequentiality between the observation at the present time and past observations. The degradation can thus be detected by using the smoothed OCSVM score when it is below the boundary 0.

2.3. HI-selection

The detection of the degradation triggers the HI-selection and RUL estimation. The HI selection method start in a blind way, where no apriori knowledge about the best HI is known. The best HI is selected online for each time sample among the different features defined in the library according to selection criteria. Selection criteria allow combining the most pertinent HI evaluation metrics which are monotonicity and trendability (Coble, 2010; Abid et al., 2018; Atamuradov et al., 2018).

Monotonicity

The monotonicity evaluates the negative or positive trend of the HI, with the assumption that the system cannot self-heal. Monotonicity is measured by the absolute difference between

Table 1. Features library.

Feature	Equation
Mean	$X_{ma} = \frac{1}{n}\sum_{i=1}^{n} y_i$
Standard Deviation	$X_{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})}$
Root Mean Square	$X_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}$
Minimum	$X_{min} = \min(y)$
Maximum	$X_{max} = \max(y)$
Kurtosis	$X_{kur} = \frac{1}{X_{rms}^4}\sum_{i=1}^{n}(y_i - \bar{y})^4$
Skewness	$X_{kur} = \frac{1}{X_{rms}^3}\sum_{i=1}^{n}(y_i - \bar{y})^3$
Peak to Peak	$X_{ptp} = \max(y) - \min(y)$
Energy	$Xen = \sum_{i=1}^{n} y_i^2$
Shape factor	$X_{sf} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} \Big/ \frac{1}{n}\sum_{i=1}^{n}\lvert y_i\rvert$
Crest factor	$X_{cf} = \max(y) \Big/ \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}$
Impulse factor	$X_{if} = \max(y) \Big/ \frac{1}{n}\sum_{i=1}^{n}\lvert y_i\rvert$

negative and positive derivative of HI as indicated in the following equation

$$Monot = \left| \frac{\text{Nb of } (d/dx > 0)}{n-1} - \frac{\text{Nb of } (d/dx < 0)}{n-1} \right| \quad (5)$$

where d/dx represents the derivative of the HI, n represents the number of observations, $Monot \in [0, 1]$, where 1 represents the highest monotonicity.

Trendability

Trendability is related to time and represents the correlation between the degradation trend and the operating time of a component, and it is calculated as follow (Javed et al., 2015)

$$Trend = \frac{\left|n(\sum_{i=1}^{n} x_i t_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} t_i)\right|}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2\right]\left[n\sum_{i=1}^{n} t_i^2 - (\sum_{i=1}^{n} t_i)^2\right]}} \quad (6)$$

$Trend \in [0; 1]$ represents the correlation coefficient between the value of HI for a pattern x at time t. $Trend$ approaches 1, when the HI has a strong positive linear correlation with time.

Selection criteria

The two metrics are confined in the range $[0, 1]$ and are positively correlated with the performance of the best feature. Consequently they are suitable for the selection criterion, which can be computed by a linear combination of the two metrics as follow:

$$Crit = \frac{Trend + Monot}{2} \quad (7)$$

The selection criteria is computed online using the points from degradation detection until present time. According to this criterion, the features defined in the library are ranked online (at each time sample), then the best one is selected as HI and used in the next step of remaining useful life estimation.

2.4. Remaining useful life estimation

Remaining useful life estimation consists in predicting the evolution of the HI through time. The remaining time is the difference between the present time and the time of End Of Life t_{EOL}, where the degradation evolution reaches a specified Failure Threshold (FT).

For this work a Generalized Linear Model (GLM) is used in predicting the HI evolution, which is a flexible generalization of the standard linear regression (McCullagh, 2018). This model is chosen because it can be used to fit different degradation dynamics (e.g., linear or exponential). GLM allows for response variable that has error distribution model other than a normal distribution, which belongs to the exponential family (e.g., the normal, binomial, Poisson). GLM generalizes linear regression by allowing the linear model $(\alpha + X\boldsymbol{\beta})$ to be related to the response variable via a link function $(g(\mu) = (\alpha + X\boldsymbol{\beta}))$. Accordingly, the regression model is given by

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\alpha + X\boldsymbol{\beta}) \quad (8)$$

The coefficient estimation of the GLM is achieved using the method of maximum likelihood. The development of a GLM model can be viewed as choosing the response distribution and the link function with the linear model, where the selection of the appropriate parameters is often dependent on the application. The degradation model is adapted by updating the model parameters (i.e., α and β) online. Hence, the RUL is corrected with each new sample collected.

The predicted RUL must be evaluated a posteriori using suitable and meaningful metrics. Root Mean Square Error

(RMSE) and Mean Absolute Percentage Error (MAPE) are used for RUL evaluation. RMSE is the standard deviation of the prediction errors, where MAPE is a relative error that can expresses the accuracy as a percentage.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(RUL_r(i) - RUL_p(i))^2} \quad (9)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{RUL_r(i) - RUL_p(i)}{RUL_r(i)}\right| \quad (10)$$

Where n is the number of observations, i is the index, RUL_r represents the real RUL, and RUL_p represents the predicted RUL. The prediction accuracy is computed after failure occurrence, due to the dependency on the real RUL.

3. EXPERIMENTATION

3.1. Description of the recorded data and sequences generation

The vibration are collected from a real high speed shaft bearing installed in a real commercial wind turbine with a 2MW power output provided by the Green Power Monitoring Systems in USA (Bechhoefer, Van Hecke, & He, 2013). After the last day of recording, an inspection of the bearing showed that the inner race was cracked (Figure 2).

Figure 2. Cracked inner race of the high speed shaft bearing after the last day of recording.

Vibration data is measured each day for 6 seconds at high sample rate (97656 samples per second), this measure is repeated for 50 days. The unit of measurement is in "g", where 1g is the earth gravitational acceleration. Figure 3 shows the collected run to failure vibration signal over 50 days, where each part of the signal with different color has a length of 585936 samples (97656 samples times 6 seconds).

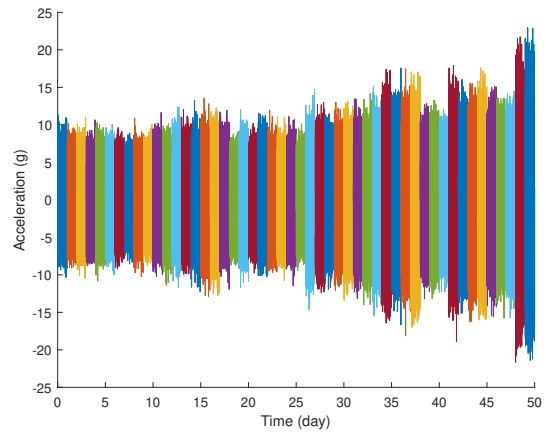

Figure 3. Collected vibration signal.

The collected data are composed of one sequence run to failure vibration measure where the first 10 days of measurement are considered in this paper as normal conditions and for the remaining days, the system is considered under degraded conditions. This one sequence of degradation is not sufficient for testing the robustness of our proposed approach. For this reason, different degradation sequences are generated based on this vibration measurement. A white Gaussian noise is added to the vibration signal in order to increase the number of degradation sequences. Two levels of white Gaussian noise have been added with 0 mean and different std ($N(0, 0.5)$ and $N(0, 1)$). In order to test our approach, 5 generated sequences of noise are added to the signal, and this is done for each level of noise. The obtained vibration noisy signals are noted: vibration signal with low level of noise ($N_L^1 ... N_L^5$), and vibration signal with high level of noise ($N_H^1 ... N_H^5$).

3.2. Feature computation

The 12 statistical features predefined in the library are computed using a window length of 6 seconds, and the library is expanded with other frequency features. For bearings monitoring, spectral analysis is useful. More precisely Spectral Kurtosis (SK), which is the kurtosis of the spectral components. It can detect impulsive bearing signatures since they could be masked by other source of vibration (e.g., gears, shafts, mechanical misalignment...)(Saidi et al., 2017; Ali, Saidi, Harrath, Bechhoefer, & Benbouzid, 2018). It is computed as follow (Antoni, 2006):

$$K(f) = \frac{\langle S(t,f)^4 \rangle}{\langle S(t,f)^2 \rangle^2} - 2 \quad (11)$$

Where $K(f)$ is the spectral kurtosis around frequency f, $\langle . \rangle$ denotes the averaging over time, and S represents the Short-Time Fourier Transform (STFT).

Figure 4 shows the computed SK on the original collected

data, where a window length of 128 time samples is chosen. As shown in this figure, the SK value increases from yellow color to red color as the degradation severity increases. Hence, it can be used as an indicator of the degradation severity. 4 statistical features (i.e., mean, std, peak-to-peak, kurtosis) are computed on the SK value for each day as indicated in (Saidi et al., 2017; Ali et al., 2018). The features library is enriched this 4 frequency features.

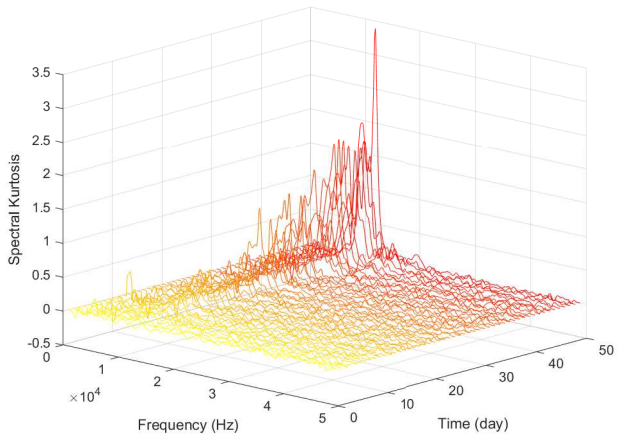

Figure 4. Spectral Kurtosis.

Figure 5 shows some features examples computed on the SK (sk-mean,sk-std, and sk-peak2peak) and some features computed on the vibration signal (kurtosis, peak2peak, and shapefactor).

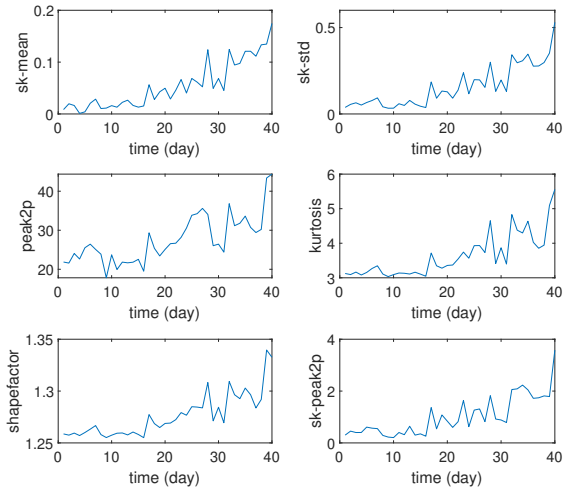

Figure 5. Examples of computed features.

3.3. Degradation detection

The data provided by the article (Bechhoefer et al., 2013) are labeled as normal for the first 10 days of the original vibration, where the remaining days from 11-50 are labeled as degradation data. 10 days corresponds to 10 samples of features which is not sufficient for training the OCSVM, features are computed on a moving window of 6s with a moving step of 1.2s in order to augment the normal data samples for the training. Hence 50 samples of normal data corresponding to the 10 first days are computed for each feature and used for the OCSVM training. For both the training and prediction, features are normalized (z-score normalization) using the mean and standard deviation of normal data. The predicted OCSVM score is then smoothed using the moving median with a window of 3 samples. The anomaly is detected when the score is below 0 (outside the boundary).

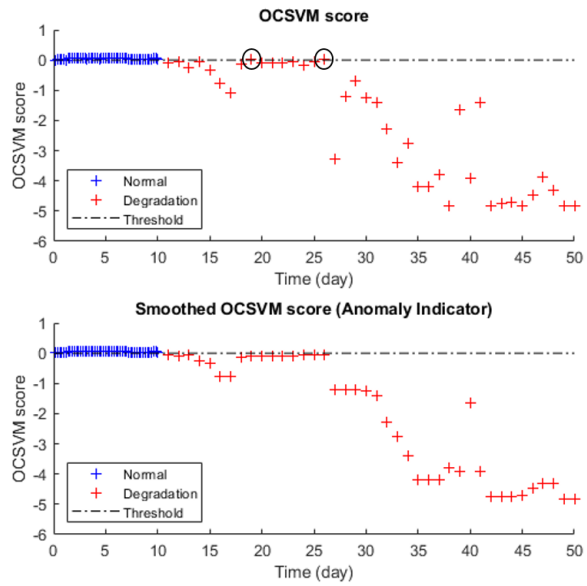

Figure 6. OCSVM score of the original vibration data before and after smoothing.

The computed OCSVM score is shown in Figure 6. It shows that the days 19 and 26 are considered as normal before the smoothing and are corrected after smoothing which is considered as the anomaly indicator. The anomaly indicator shows that the degradation can be detected at the day 11 using only the normal conditions data. So the proposed approach has an early degradation detection capabilities. Degradation detection trigger the HI selection and RUL estimation step.

3.4. HI selection

The online HI selection is triggered after the degradation is detected. The computed features are ranked according to their selection criterion computed with Eq. (7). The computed on-

line selection criterion for each features is presented in Figure 7. The feature with the best score of selection criterion is selected as HI at each day (Table 2), it can be seen that for the first period of degradation (from day 11 to day 22), the selected feature changes among std, sk-std, skewness, mean due to the low degradation severity. Where, the sk-mean remains after the day 23 the best feature until the last day (failure). Because of sk-mean represents the mean of the SK at each day, where the SK is proportional to the degradation severity.

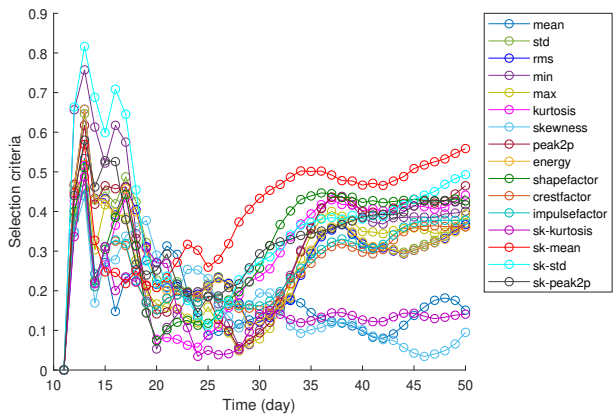

Figure 7. Selection criteria computed online for each day of the degradation.

Table 2. Selected features as HI for each day.

Selected features	std	sk-std	skewness	mean	sk-mean
Time(day)	11	12-18	19-20	21-22	23-50

Table 3 shows the features ranking, according to the selection criterion for the last day of degradation, where the 6 best features (sk-mean, sk-std, peak2p, kurtosis, shapefactor, and sk-peak2p) are plotted in Figure 5.

3.5. Remaining useful life estimation

The feature selected as HI for each time sample (day) can characterize the degradation evolution. The aim is to predict the HI evolution over time until it reaches a failure threshold. HI evolution is predicted by applying an adaptive GLM model where its coefficients are updated according to new sample (day), and the GLM parameters are chosen depending on the domain knowledge. It is reported that the bearing degradation has commonly an exponential growth over time (Gebraeel, Lawley, Liu, & Parmeshwaran, 2004). Hence, a GLM model is used with a log link function and Poisson distribution as defined in Eq. (12). Failure threshold can be computed by using some available failure data or it can be set by a domain expert. In this case, features computed on the last day of the degradation are used as failure threshold,

Table 3. Ranked features at the last day of degradation.

Rank	Features	Monot	Trend	Crit
1	sk-mean	0,231	0,887	0,559
2	sk-std	0,128	0,858	0,493
3	peak2p	0,179	0,75	0,465
4	kurtosis	0,077	0,805	0,441
5	shapefactor	0,026	0,828	0,427
6	sk-peak2p	0,026	0,808	0,417
7	min	0,026	0,775	0,401
8	max	0,077	0,708	0,392
9	impulsefactor	0,077	0,679	0,378
10	std	0,128	0,613	0,371
11	rms	0,128	0,606	0,367
12	energy	0,128	0,595	0,362
13	crestfactor	0,077	0,646	0,361
14	mean	0,077	0,225	0,151
15	sk-kurtosis	0,077	0,205	0,141
16	skewness	0,026	0,164	0,095

Table 4. RUL prediction accuracy for the different degradation dynamics.

Degradation dynamics	N	RMSE	MAPE
Original degradation data	-	16.484	42.908
Degradation data with low level noise	N_L^1	16,182	46,68
	N_L^2	15,402	61,41
	N_L^3	17,481	50,235
	N_L^4	16,289	51,745
	N_L^5	16,264	54,794
Degradation data with high level noise	N_H^1	15,906	62,94
	N_H^2	14,296	44,01
	N_H^3	17,502	63,979
	N_H^4	13,601	51,646
	N_H^5	14,165	44,543

then HI is normalized between 0 and 1, where 1 is the failure threshold, after that the time of End Of Life t_{EOL} (time of failure) is computed from Eq. (13) as follow

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = e^{(\alpha + X\boldsymbol{\beta})} \tag{12}$$

$$t_{EOL} = \frac{ln(\mu) - \alpha}{\beta} \tag{13}$$

Where μ is the failure threshold equal to 1 after normalization, α and β are the regression model coefficients. RUL

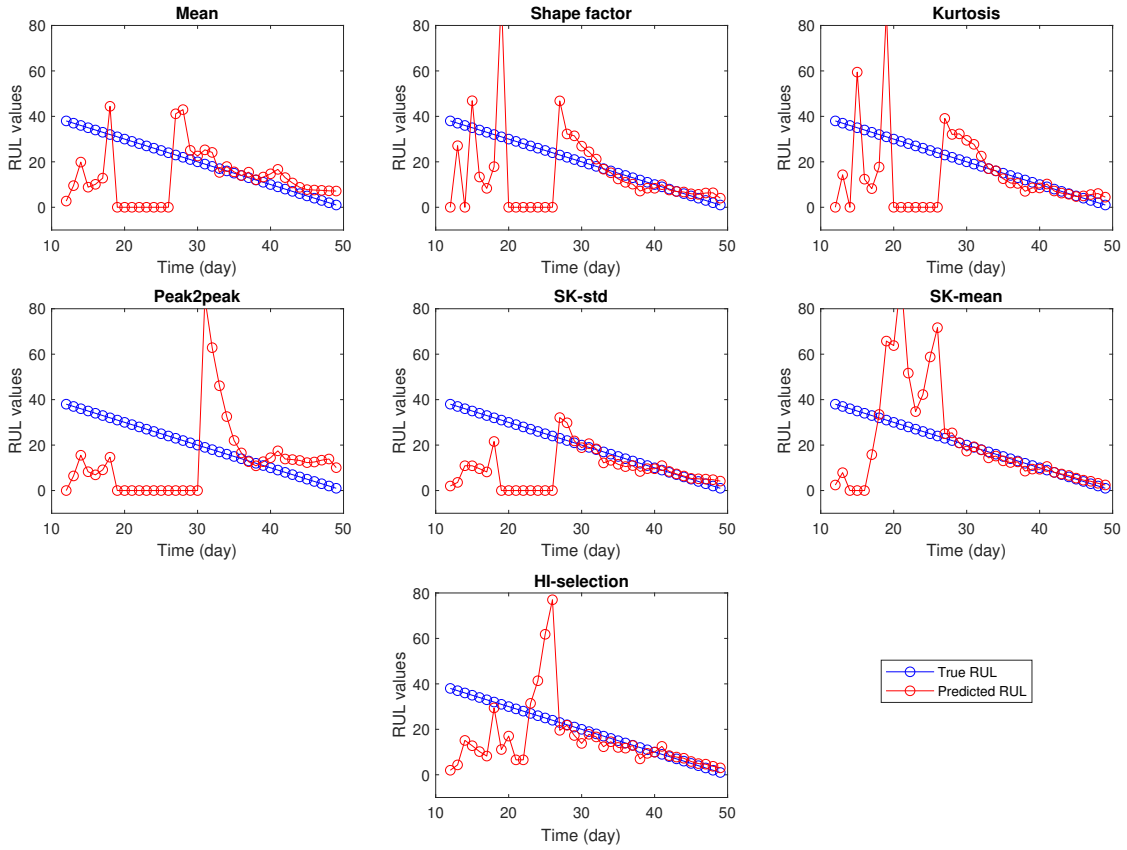

Figure 8. Predicted RUL based on the collected vibration data using different Health indicators.

Table 5. RUL prediction accuracy using different HI and regression model.

HI	Linear Regression		Exponential Model		GLM Model	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
sk-peak2p	42,158	453,665	17,064	86,027	17,235	80,534
shapefactor	36,675	372,112	16,927	72,177	19,286	65,375
kurtosis	36,215	377,605	17,591	75,129	19,271	66,424
peak2p	29,069	335,871	21,915	136,879	23,647	139,926
sk-std	37,417	370,55	17,438	70,794	17,233	56,53
sk-mean	36,804	304,617	22,214	58,207	21,112	50,514
HI-selection	36,252	298,243	16,691	48,238	**16,484**	**42,908**

is deduced by the difference time between (t_{EOL}) and the present time at index i (t_i).

Figure 8 shows the true and the predicted RUL of the collected vibration data over the 40 days of degradation. In this figure, The RUL is estimated using the defined GLM model. The latter is applied by two manners in order to compare the performance (RUL prediction accuracy) of the proposed online HI-selection and the traditional use (offline) of HIs. In the former, the selection of the best HI according to the selection criterion Eq. (7) is achieved online after the reception of each new vibration data sample. While in the latter, the performance of each HI is calculated using all the sequence of vibration data. For simplicity, Figure 8 shows the performance of the best 6 HIs defined in Table 3. It can be seen

that the predicted RUL deviates from the true RUL in the beginning until 50% of degradation length due to the lack of data. It becomes more precise when more data points are collected. RUL prediction accuracy for the original collected data and for the generated data (noisy data) are presented in Table 4. The computed RUL errors RMSE and MAPE using the HI-selection method indicate that the proposed approach is robust to the different levels of noise (low and high levels).

In Table 5, a comparison is presented according to RUL prediction accuracy. RUL is computed using 6 HI selected offline and the proposed online HI-selection method. It is also computed using different regression models: linear regression $y = at + b$, exponential model $y = ae^{bt}$, and the GLM model represented by Eq. (12). It is shown that the GLM model outperforms the linear regression model and it slightly exceeds the exponential model. It is noticed also that the proposed online HI selection method performs better than the selected offline HI (6 best HI). It is even better than the sk-mean selected offline which is the best feature at the end of degradation.

4. CONCLUSION

In this paper, a general data-driven prognostic approach is proposed, which can deal with the lack of degradation data in the offline phase. First, a library of statistical features is defined, then the smoothed OCSVM score is used to detect the degradation as early as possible. The degradation detection trigger the HI selection and RUL estimation steps. In the HI selection step, features defined in the library are ranked according to the evaluation criteria, where the best feature is selected as HI. In the RUL estimation step, an adaptive degradation model is used where its parameters are re-estimated online. This model can predict the degradation evolution (HI evolution) over time, where the RUL is corrected online. The proposed approach is validated with real degradation data collected from a high speed shaft bearings of a commercial wind turbine. Experimentation result shows that the proposed approach is able to detect the degradation early at day 11 and can predict a precise RUL after collecting about 50% of the degradation data.

As future work, it is planned to improve the RUL estimation by computing RUL confidence, where the confidence increases while collecting more data. Then, the proposed approach will be validated on other degradation data-sets related to other wind turbine components, where the system is impacted by the environmental variation (e.g., wind speed) and the change in operating conditions.

ACKNOWLEDGMENT

This work is supported by the European Union - European Regional Development Fund.

REFERENCES

Abid, K., Sayed-Mouchaweh, M., & Cornez, L. (2018). Fault prognostics for the predictive maintenance of wind turbines: State of the art. In *Joint european conference on machine learning and knowledge discovery in databases* (pp. 113–125).

Ali, J. B., Saidi, L., Harrath, S., Bechhoefer, E., & Benbouzid, M. (2018). Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning. *Applied Acoustics*, *132*, 167–181.

Antoni, J. (2006). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical systems and signal processing*, *20*(2), 282–307.

Atamuradov, V., Medjaher, K., Dersin, P., Zerhouni, N., & Camci, F. (2018). A new adaptive prognostics approach based on hybrid feature selection with application to point machine monitoring.

Bechhoefer, E., Van Hecke, B., & He, D. (2013). Processing for improved spectral analysis. In *Annual conference of the prognostics and health management society, new orleans, la, oct* (pp. 14–17).

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Engineering Applications of Artificial Intelligence*, *26*(7), 1751–1760.

Coble, J. B. (2010). Merging data sources to predict remaining useful life–an automated method to identify prognostic parameters.

Gebraeel, N., Lawley, M., Liu, R., & Parmeshwaran, V. (2004). Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Transactions on industrial electronics*, *51*(3), 694–700.

Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, *240*, 98–109.

Javed, K., Gouriveau, R., Zerhouni, N., & Nectoux, P. (2015). Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Transactions on Industrial Electronics*, *62*(1), 647–656.

McCullagh, P. (2018). *Generalized linear models*. Routledge.

Saidi, L., Ali, J. B., Bechhoefer, E., & Benbouzid, M. (2017). Wind turbine high-speed shaft bearings health prognosis through a spectral kurtosis-derived indices and svr. *Applied Acoustics*, *120*, 1–8.

Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., & Platt, J. C. (2000). Support vector method for novelty detection. In *Advances in neural information processing systems* (pp. 582–588).

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on reliability*, *61*(2), 491–503.

Wang, Y., Peng, Y., Zi, Y., Jin, X., & Tsui, K.-L. (2016). A two-stage data-driven-based prognostic approach for bearing degradation problem. *IEEE Transactions on Industrial Informatics*, *12*(3), 924–932.

BIOGRAPHIES

Koceila Abid received his Master Degree in Industrial Process Automation from the university of Science and Technology USTHB of Algiers - Algeria in 2016. Then he obtained a second Master Degree in Human Machine System Engineering from the university of Lorraine, Metz - France in 2017. Currently, he is a Ph.D student at The French Alternative Energies and Atomic Energy Commission (CEA) in collaboration with the Department of Automatic Control and Computer Science of the Mines Telecom Institut (IMT Lille - Douai). His research interest include Health monitoring, Fault Diagnostics, and Prognostics of Industrial Systems using Machine Learning techniques.

Moamar Sayed-Mouchaweh received his Master Degree from the University of Technology of Compiegne – France in 1999. Then, he received his Ph.D. degree from the University of Reims – France in December 2002. He was nominated as Associated Professor in Computer Science, Control and Signal processing at the University of Reims – France in the Research center in Sciences and Technology of the Information and the Communication (CReSTIC). In December 2008, he obtained the Habilitation to Direct Researches (HDR) in computer science, control and signal processing. Since September 2011, he is working as a Full Professor in the Mines Telecom Institut (Institu Mines Télécom IMT Lille - Douai) at the Department of Automatic Control and Computer Science (Informatique & Automatique IA).

Laurence Cornez graduated from SUPAERO (French National School for Aeronautics) where she obtained her PhD degree in Computer Science in 2007. She currently works at The French Alternative Energies and Atomic Energy Commission (CEA List) as a research engineer since 2007. Her work focuses on modeling human expertise with fuzzy expert system and their integration in complex systems and on machine learning techniques.